



Veröffentlichungen der DGK

Ausschuss Geodäsie der Bayerischen Akademie der Wissenschaften

Reihe C

Dissertationen

Heft Nr. 861

Hamza Alkhatib

**Fortgeschrittene Methoden und Algorithmen
für die computergestützte geodätische Datenanalyse**

München 2020

Verlag der Bayerischen Akademie der Wissenschaften

ISSN 0065-5325

ISBN 978-3-7696-5273-4

**Diese Arbeit ist gleichzeitig veröffentlicht in:
Wissenschaftliche Arbeiten der Fachrichtung Geodäsie und Geoinformatik der Leibniz
Universität Hannover**

ISSN 0174-1454, Nr. 367, Hannover 2020



Veröffentlichungen der DGK

Ausschuss Geodäsie der Bayerischen Akademie der Wissenschaften

Reihe C

Dissertationen

Heft Nr. 861

Fortgeschrittene Methoden und Algorithmen für die computergestützte geodätische Datenanalyse

Von der Fakultät für Bauingenieurwesen und Geodäsie
der Gottfried Wilhelm Leibniz Universität Hannover

zur Verleihung der Lehrbefähigung

für das Fachgebiet "Geodätische Auswertemethoden und Ingenieurgeodäsie"

genehmigte Habilitationsschrift

Vorgelegt von

Dr.-Ing. Hamza Alkhatib

Geboren am 05.06.1969 in Damaskus

München 2020

Verlag der Bayerischen Akademie der Wissenschaften

ISSN 0065-5325

ISBN 978-3-7696-5273-4

Diese Arbeit ist gleichzeitig veröffentlicht in:
Wissenschaftliche Arbeiten der Fachrichtung Geodäsie und Geoinformatik der Leibniz Universität Hannover
ISSN 0174-1454, Nr. 367, Hannover 2020

Adresse der DGK:



Ausschuss Geodäsie der Bayerischen Akademie der Wissenschaften (DGK)

Alfons-Goppel-Straße 11 • D – 80 539 München
Telefon +49 – 331 – 288 1685 • Telefax +49 – 331 – 288 1759
E-Mail post@dgk.badw.de • <http://www.dgk.badw.de>

Habilitationskommission:

Vorsitzender: Prof. Dr.-Ing. Steffen Schön

Hauptberichter: Prof. Dr.-Ing. Ingo Neumann

Mitberichter: Prof. Dr.-Ing. habil. Hansjörg Kutterer (Karlsruher Institut für Technologie)
Prof. Dr.-Ing. Nico Sneeuw (Universität Stuttgart)
Prof. Dr.-Ing. Michael Beer (Leibniz Universität Hannover)

Tag der Einreichung: 20.12.2019

Tag des Habilitationskolloquiums: 30.10.2020

Kurzfassung

Die fortschreitende Digitalisierung mit ihren innovativen Technologien stellt zunehmende Anforderungen an Wirtschaft, Gesellschaft und Verwaltungen. Digitale Daten gelten als Schlüsselressource, die hohe Ansprüche u. a. an die Datenverarbeitung stellt, wie z. B. hohe Geschwindigkeit und Zuverlässigkeit. Besondere Bedeutung sind digitalen Daten mit Raumbezug beizumessen. Digitale Daten stammen im Bereich der Geodäsie und Geoinformatik von Multi-Sensor-Systemen, Satellitenmissionen, Smartphones, technischen Geräten, Computern oder von Datenbanken unterschiedlichster Institutionen und Behörden. „Big Data“ heißt der Trend und es gilt die enormen Datenmengen so breit und so effektiv wie möglich zu nutzen und mit Hilfe von computergestützten Tools, beispielsweise basierend auf künstlicher Intelligenz, auszuwerten. Um diese großen Datenmengen statistisch auszuwerten und zu analysieren, müssen laufend neue Modelle und Algorithmen entwickelt, getestet und validiert werden. Algorithmen erleichtern Geodätinnen und Geodäten seit Jahrzehnten das Leben – sie schätzen, entscheiden, wählen aus und bewerten die durchgeführten Analysen.

Bei der geodätisch-statistischen Datenanalyse werden Beobachtungen zusammen mit Fachkenntnissen verwendet, um ein Modell zur Untersuchung und zum besseren Verständnis eines datengenerierenden Prozesses zu entwickeln. Die Datenanalyse wird verwendet, um das Modell zu verfeinern oder möglicherweise ein anderes Modell auszuwählen, um geeignete Werte für Modellterme zu bestimmen und um das Modell zu verwenden, oder um Aussagen über den Prozess zu treffen. Die Fortschritte in der Statistik in den vergangenen Jahren beschränken sich nicht nur auf die Theorie, sondern umfassen auch die Entwicklung von neuartigen computergestützten Methoden. Die Fortschritte in der Rechenleistung haben neuere und aufwendigere statistische Methoden ermöglicht. Eine Vielzahl von alternativen Darstellungen der Daten und von Modellen können untersucht werden.

Wenn bestimmte statistische Modelle mathematisch nicht realisierbar sind, müssen Approximationsmethoden angewendet werden, die oft auf asymptotischer Inferenz basieren. Fortschritte in der Rechenleistung und Entwicklungen in der Theorie haben die computergestützte Inferenz zu einer praktikablen und nützlichen Alternative zu den Standardmethoden der asymptotischen Inferenz in der traditionellen Statistik werden lassen. Die computergestützte Inferenz basiert auf der Simulation statistischer Modelle.

Die vorliegende Habilitationsschrift stellt die Ergebnisse der Forschungsaktivitäten des Autors im Bereich der statistischen und simulationsbasierten Inferenz für die geodätische Datenanalyse vor, die am Geodätischen Institut der Gottfried Wilhelm Leibniz Universität Hannover während der Zeit des Autors als Postdoktorand von 2009 bis 2019 publiziert wurden. Die Forschungsschwerpunkte in dieser Arbeit befassen sich mit der Entwicklung von mathematisch-statistischen Modellen, Schätzverfahren und computergestützten Algorithmen, um raum-zeitliche und möglicherweise unvollständige Daten, welche durch zufällige, systematische, ausreißerbehaftete und korrelierte Messabweichungen charakterisiert sind, rekursiv sowie nicht-rekursiv auszugleichen. Herausforderungen bestehen hierbei in der genauen, zuverlässigen und effizienten Schätzung der unbekannten Modellparameter, in der Ableitung von Qualitätsmaßen der Schätzung sowie in der statistisch-simulationsbasierten Beurteilung der Schätzergebnisse. Die Forschungsschwerpunkte haben verschiedene Anwendungsmöglichkeiten in den Bereichen der Ingenieurgeodäsie und der Immobilienbewertung gefunden.

Stichworte: Klassische und robuste Parameterschätzung, Bayessche Inferenz, rekursive Zustandschätzung, Georeferenzierung, Monte Carlo Techniken, Bootstrapping, Student-Verteilung, B-Spline Modelle, EM-Algorithmus, Guide to the Expression of Uncertainty in Measurement (GUM)

Abstract

The advancing digitalization with its innovative technologies places increasing expectations on economy, society and public authorities. Digital data is considered a key resource, which requires high demands on data processing, such as high speed and reliability. Special importance is attached to digital data with spatial reference. In the fields of geodesy and geoinformatics, digital data stem from multi-sensor systems, satellite missions, smartphones, technical devices, computers, or databases from various institutions and governmental agencies. “Big Data” is called the trend to use these gigantic amount of data as widely and effectively as possible, and to analyze them with the aid of computerized tools, such as artificial intelligence. To process and analyze the large data volume captured, new models and algorithms are being constantly developed, tested and validated. Algorithms have made the lives of geodesists easier for decades - they estimate, decide, select, and evaluate.

In geodetic-statistical data analysis observations are used jointly with expert knowledge in order to develop a model for investigation and for a better understanding of a data-generated process. Data analysis is employed to optimize a model or possibly to select a different one, to determine adequate values for the terms of the model, and to use the model to make statements about the process. The advances of statistics throughout the previous years are not limited only to theory, but include also the development of innovative computerized methods. The advances of computational performance enabled newer and more laborious statistical methods. Many alternative representations of the data and many different models can be investigated.

When certain statistical models cannot be realized mathematically, approximative methods, which are often based on asymptotic inference, must be applied. Progress in computational performance and advances of theory have turned computerized inference into a practical and useful alternative to the standard methods of asymptotic inference in traditional statistics. Computerized inference is based on the simulation of statistical models.

The habilitation thesis presents the results of my research activities in the field of statistical and simulation-based algorithms for geodetic data analysis that I published at the Geodetic Institute of the Gottfried Wilhelm Leibniz Universität Hannover during my time as a post-doctoral researcher between 2009 and 2019. The main research topics in this exposition deal with the development of mathematical-statistical models, estimation procedures and computerized algorithms for the purpose of a recursive or non-recursive adjustment of spatio-temporal and possibly incomplete data that are characterized by random, systematic, outlier-afflicted and correlated measurement errors. Here, challenges exist in connection with the accurate, reliable and efficient estimation of the unknown model parameters, with the derivation of quality measures of the estimation, as well as with statistical-simulation-based evaluation of the estimation results. These main research topics have found different possibilities of application in the fields of surveying engineering and real-estate valuation.

Keywords: Classical and Robust Parameter Estimation, Bayesian Inference, Recursive State Estimation, Georeferencing, Monte Carlo Techniques, Bootstrapping, Student Distribution, B-Spline Models, EM-Algorithm, Guide to the Expression of Uncertainty in Measurement (GUM)

Inhaltsverzeichnis

Kurzfassung	1
Abstract	3
Abbildungsverzeichnis	7
Liste der Abkürzungen	9
1 Einleitung	11
1.1 Motivation	11
1.1.1 Merkmale von geodätischen Daten	11
1.1.2 Statistische Modelle	12
1.1.3 Computergestützte und simulationsbasierte geodätische Datenanalyse	14
1.2 Ziele der Arbeit	15
1.3 Übersicht der originalen Publikationen zur kumulativen Habilitation	15
I TEIL 1: Synopse der Habilitation	17
2 Optimierungsverfahren und Parameterschätzung in linearen und nicht-linearen Modellen	19
2.1 Einführung in die Parameterschätzung in linearen Regressionsmodellen	19
2.2 Vorstellung der ausgewählten nicht-linearen Regressionsmodelle	20
2.3 MODELL I: Robuste B-Spline Regressionsmodelle	24
2.3.1 Mathematisches Modell und Algorithmen zur Optimierung von Knotenvektoren	24
2.3.2 Anwendungsfälle zum MODELL I	27
2.4 MODELL II: Regressionsmodelle mit autoregressiven Prozessen und t-verteilterm Messrauschen	29
2.4.1 Mathematisches Modell und GEM-Algorithmen zur Parameterschätzung . . .	29
2.4.2 Anwendungsfälle zum MODELL II	32
2.5 MODELL III: Bayessche robuste Regressionsmodelle mit t-verteilterm Messrauschen	34
2.5.1 Mathematisches Modell	34
2.5.2 Markov-Chain-Monte-Carlo (MCMC) Algorithmen zur Parameterschätzung .	35
2.5.3 Anwendungsfälle zum MODELL III	37
2.6 Zusammenfassung des Kapitels	40
3 Filterungs- und Optimierungsverfahren für die Zustandsschätzung im nicht-linearen Zustandsraum	41
3.1 Einführung zur verschiedenen Filterungstechniken und zur sequentiellen Bayesschen Zustandsschätzung	41
3.2 Vorstellung nicht-linearer Filterungstechniken	44
3.3 FILTERUNGSANSÄTZE I: Stochastisch-analytische Ansätze	46
3.4 FILTERUNGSANSÄTZE II: Monte-Carlo-Simulationsbasierte Ansätze	54
3.5 FILTERUNGSANSÄTZE III: Kombinierte stochastisch-deterministische Ansätze . .	57
3.6 Zusammenfassung des Kapitels	61

4	Simulationsbasierte Methoden für die Inferenzstatistik	63
4.1	Einführung in Monte-Carlo- und Bootstrapping-Methoden	63
4.2	Vorstellung der simulationsbasierten Methoden für die Inferenzstatistik	64
4.3	Simulationsbasierte-Techniken zur Bestimmung von statistischen Momenten und Konfidenzintervallen	65
4.3.1	BEREICHSSCHÄTZUNG I: Monte Carlo-Methoden zur GUM Unsicherheitsmodellierung	65
4.3.2	BEREICHSSCHÄTZUNG II: Bootstrapping zur Schätzung des Konfidenzbereichs für den EM-Algorithmus	68
4.4	Statistische Beurteilung von Schätzergebnissen mit simulationsbasierten Hypothesentests	71
4.4.1	TESTPROBLEM I: Simulationsbasierter Ansatz für die Modellwahl zur Oberflächenmodellierung von 3D-Punktwolken	72
4.4.2	TESTPROBLEM II: Bootstrap-Ansatz für das Testen auf Zeitvariabilität eines AR-Prozesses	74
4.5	Zusammenfassung des Kapitels	78
5	Zusammenfassung und Ausblick	81
5.1	Zusammenfassung	81
5.2	Ausblick	83
II	TEIL 2: Eigene Veröffentlichungen	85
6	Liste der eigenen Publikation	87
6.1	Veröffentlichungen zur Kapitel 2	87
6.2	Veröffentlichungen zur Kapitel 3	87
6.3	Veröffentlichungen zur Kapitel 4	88
6.4	Beiträge der Autoren	88
7	Originalpublikationen	91
7.1	Originalpublikationen zur Kapitel 2	91
7.2	Originalpublikationen zur Kapitel 3	173
7.3	Originalpublikationen zur Kapitel 4	266
	Literaturverzeichnis	333
	Danksagung	345
	Lebenslauf	347

Abbildungsverzeichnis

1.1	Gesamtbild Inferenz/Statistik	13
1.2	Übersicht der originalen Publikationen zur kumulativen Habilitation	16
2.1	Links: Approximationsergebnis Testfunktion 1 mit $p = 3$ und $n = 11$, rechts: Approximationsergebnis Testfunktion 2 mit $p = 3$ und $n = 8$	28
2.2	Die Approximationsergebnisse der realen Daten mit $p = 3$ und $n = 60$	29
2.3	Histogramm aus der Monte-Carlo-Simulation für die Koeffizienten a_0 (oben links), a_1 (Mitte links), a_2 (unten links) und für die AR Koeffizienten α_1 (oben rechts), α_2 (mitte rechts), α_3 (unten rechts) für $n = 100000$ Beobachtungen der Fourier-Modell mit AR(100) Abweichungen und weißes Rauschen nach der $t_2(0, 0.001^2)$ -Verteilung (gelb), der $N(0, 0.001^2)$ -Verteilung (rot) und das $0.6 \cdot N(0, 0.001^2) + 0.4 \cdot N(0, 0.008^2)$ -Verteilung (blau). Dicke schwarze Linien zeigen die Position des wahren Wertes an.	33
2.4	Links: Prototypische Realisierung des für das Experiment verwendeten MSS. Rechts: Das geschätzte Periodogramm der dekorrelierten Residuen für die Nord- und Ostkomponenten-AR(15)-Modell (blau und grün) und für das Hochkomponenten-AR(18)-Modell (magenta) in Bezug auf das theoretische Periodogramm mit theoretisch weißem Rausches (schwarz) und 99% Signifikanzgrenzen (rot).	34
2.5	Verteilung der 10.000 RMSE-Ergebnisse für die drei Ansätze GMM (blau), Bayessche nicht-informative Wahrscheinlichkeitsdichtefunktion (PDF) (grün) und Bayessche informative PDF (rot) (Dorndorf et al., 2019)	38
2.6	Links: RMSE für die Reduktion auf das Minimum der Originalstichprobe und die Nutzung von Gutachten als a-priori Information. Rechts: RMSE für die Reduktion auf das Maximum der Originalstichprobe und die Nutzung von Gutachten als a-priori Information; Originalabbildung aus Weitkamp und Alkhatib (2014)	39
3.1	Schematischer Ablauf des rekursiven Filteransatzes zur Georeferenzierung eines kinematischen MSS. Die Schritte des IEKF (grau) werden mit möglichen gesuchten Zustandsparametern in gelb, verfügbaren Beobachtungen in grün und vorhanden a-priori Informationen in blau dargestellt. Entsprechende Unsicherheitsinformationen werden durch rote Kreise gekennzeichnet; Abbildung wurde übersetzt aus Vogel et al. (2019).	48
3.2	Die Grundidee des Ansatzes IEKF-GHM (Bureick et al., 2019c, modifiziert)	50
3.3	Links: Gesamtansicht des kinematischen MSS, welches die verwendete Sensoren und die unterschiedlichen Koordinatensysteme veranschaulicht, rechts: durchgeführtes Experiment im Untergeschoss des Geodätisches Institut Hannover (GIH); Originalabbildungen aus Vogel et al. (2019)	51
3.4	Präzise IMU: Zeitlicher Verlauf des Medians des RMSE für die Position (links) und ME für die Orientierung (rechts) mittels 500 Wiederholungen für entsprechende Kombinationen der verwendeten Zustandsrestriktionen. Die römischen Ziffern beziehen sich auf die jeweiligen Zustandsrestriktionen, die in Bezug auf die Tabelle 3 aus Vogel et al. (2019) angewendet werden.	52
3.5	Schematischer (links) und tatsächlicher Aufbau des UAV (rechts); Originalabbildung aus Bureick et al. (2019c).	53

3.6	Mittlere Root Mean Square Error (RMSE)-Werte für jede Epoche über die gesamte Monte-Carlo (MC)-Simulationen für alle geschätzte 6 -DoF: Oben links für die Translation in Richtung x ; Oben rechts für die Translation in Richtung y , Mitte links für die Translation in Richtung z ; Mitte rechts für die Orientierung ω ; unten links für die Orientierung Φ ; unten rechts für die Orientierung κ	54
3.7	RMSE-Werte für die nicht-linearen Filterungsergebnisse mit dem neuen entwickelten Erweiterte Partikel-Kalmanfilter (EKPF)-Algorithmus (blau) verglichen zu dem generischen Partikelfilter (PF) (magenta) und zum Erweiterte Kalmanfilter (EKF) (grün). Insgesamt wurden 500 MC-Durchläufe wiederholt; Originalabbildung aus Alkhatib (2015).	56
3.8	Ergebnisse von EKF- und EKPF-Algorithmen; oben: Residuen, die mittels einer linearen Regression zur Bestimmung der Orientierung gewonnen wurden; unten: die gefilterten Neigungen. Der EKPF-Ansatz zeigt eine signifikante Verbesserung der Filterwirkung für die gefilterte Neigung; Abbildung übersetzt aus Alkhatib et al. (2012).	57
3.9	Links: RMSE-Vergleich von EGKF und EKF in 1000 Durchläufen; unten: mit EKF und Ellipsoidisch-Gaußsches Kalmanfilter (EGKF) geschätzte 2D Trajektorie, Originalabbildungen aus Sun et al. (2019).	60
3.10	Gegenüberstellung der aus EKF- und EGKF geschätzten Positionen. Ca. 99,95% der EKF-Punkte befinden sich in den EGKF-Intervallen. Oben: Geschätzte Position in x - und z -Richtung. Unten: Geschätzte 3D-Trajektorien anhand des EGKF; Originalabbildungen aus Sun et al. (2018).	60
4.1	Statistische Momente zweiter bis vierter Ordnung der Zielgrößen (hier z -Koordinaten der 2D-Profilmessungen) aus Monte-Carlo-Simulationen für die Einflussgrößen Z_1 bis Z_4 aus dem realen Datensatz. Für die Standardabweichung (oben), die Schiefe (Mitte) und die Kurtosis (unten) werden jeweils die Ergebnisse für drei Klassen (Aggregation von 1, 5 und 10 Punkten pro Klasse) der 2D-Profilpunkte gezeigt (Alkhatib und Kutterer, 2013).	69
4.2	Statistische Momente der Zielgröße (hier z -Koordinaten der 2D-TLS-Profilmessungen) aus dem realen Datensatz für die Einflussgrößen Z_1 bis Z_4 . Für die Standardabweichung (oben), die Schiefe (Mitte) und die Kurtosis (unten) werden jeweils die Ergebnisse für drei Klassen (Aggregation von 1, 5 und 10 Punkten pro Klasse) der 2D-Profilpunkte gezeigt (Alkhatib und Kutterer, 2013).	69
4.3	Vergleich von Polynom- (a,c) und B-Spline-Oberflächenmodellen (b,d) in Bezug auf die Unterschiede der 1. und 13. Epoche in Segment I (Zhao et al., 2018)	75
4.4	Vergleich des geschätzten $\widehat{pv}^{(i)}$ -Wertes (Ablehnung der Nullhypothese, wenn $\widehat{pv}^{(i)} < 0.05$) unter dem Autoregressiv (AR)(1)-Modell mit zeitvariablen Parametern im Wertebereich $y_2 = [0 : 0.0001 : 0.003]$ für parametrisches und nicht-parametrisches Bootstrapping; Übersetzung der Originalabbildung aus Alkhatib et al. (2019)	78

Liste der Abkürzungen

AIC	Akaike-Informationskriterium
AR	Autoregressiv
BIC	Bayessches Informationskriterium
CSA	Clonal Selection Algorithm
DoF	Degrees of Freedom
ECM	Expectation Conditional maximization
ECME	Expectation Conditional Maximization Either
EGA	Elitärer Genetischer Algorithmus
EGKF	Ellipsoidisch-Gaußsches Kalmanfilter
EKF	Erweiterte Kalmanfilter
EKPF	Erweiterte Partikel-Kalmanfilter
EM	Expectation Maximization
EMC	Elitäre Monte-Carlo
EA	Elitärer Algorithmus
EnKF	Ensemble Kalmanfilter
GA	Genetische Algorithmen
GEM	Generalized Expectation Maximization
GHM	Gauß-Helmert-Modell
GIH	Geodätisches Institut Hannover
GMM	Gauß-Markov-Modell
GNSS	Global Navigation Satellite System
GUM	Guide to the Expression of Uncertainty in Measurement
IEKF	Iterative Erweiterte Kalmanfilter
IMU	Inertial Measurement Unit
IRLS	Iterativ Regewichtete Kleinste Quadrate, engl. iteratively reweighted least squares
KF	Kalmanfilter
KQ	Kleinste Quadrate

kTLS	Kinematic Terrestrial Laser Scanning
LQ	Likelihood-Quotienten
LS	Least-Squares
MAP	Maximum-a-posteriori
MC	Monte-Carlo
MCMC	Markov-Chain-Monte-Carlo
MH	Metropolis-Hastings
MkQ	Methode der kleinsten Quadrate
ML	Maximum-Likelihood
MSS	Multi Sensor System
NURBS	Non-Uniform Rational B-Spline
PDF	Wahrscheinlichkeitsdichtefunktion
PF	Partikelfilter
RMSE	Root Mean Square Error
SIS	Sequential Importance Sampling
SMC	Sequential Monte-Carlo
TLS	Terrestrisches Laserscanning
UAV	Unmanned-Aerial-Vehicle
UBB	Unknown But Bounded Uncertainty
UKF	Unscented Kalmanfilter
VKM	Varianz-Kovarianz-Matrix

1 Einleitung

1.1 Motivation

1.1.1 Merkmale von geodätischen Daten

Die fortschreitende Digitalisierung mit ihren innovativen Technologien stellt zunehmende Anforderungen an Wirtschaft, Gesellschaft und Verwaltungen. Digitale Daten gelten als Schlüsselressource, die hohe Ansprüche u. a. an die Datenverarbeitung stellt, wie z. B. hohe Geschwindigkeit und Zuverlässigkeit. Besondere Bedeutung ist digitalen Daten mit Raumbezug beizumessen. Digitale Daten stammen im Bereich der Geodäsie und Geoinformatik von Multi-Sensor-Systemen, Satellitenmissionen, Smartphones, technischen Geräten, Computern oder von Datenbanken unterschiedlichster Institutionen und Behörden. „Big Data“ heißt der Trend und es gilt die enormen Datenmengen so breit und so effektiv wie möglich zu nutzen und mit Hilfe von computergestützten Tools, beispielsweise basierend auf künstlicher Intelligenz, auszuwerten. Um diese großen Datenmengen statistisch auszuwerten und zu analysieren, müssen laufend neue Modelle und Algorithmen entwickelt, getestet und validiert werden. Algorithmen erleichtern Geodätinnen und Geodäten seit Jahrzehnten das Leben – sie schätzen, entscheiden, wählen aus und bewerten die durchgeführten Analysen.

Neben den Modellen und der Algorithmik spielt die Beurteilung der Qualität der Daten und der Ergebnisse eine entscheidende Rolle. Gerade vor dem Hintergrund der notwendigen Automatisierung bei der Behandlung von „Big Data“ gilt es wichtige Qualitätsmerkmale auch automatisch aus den Daten ableiten zu können, insbesondere vor dem Hintergrund, dass die erfassten Daten sehr unterschiedliche Charakteristika aufweisen können. Die wichtigsten Merkmale, die die Qualität von Daten beschreiben, sind:

- Genauigkeit: Dieses Merkmal beschreibt die Genauigkeit der erfassten Messungen unter Berücksichtigung von zufälligen und systematischen Messabweichungen.
- Zuverlässigkeit und Konsistenz: Viele Multi-Sensor-Systeme (MSSs) in heutigen Anwendungen erheben Messungen für dieselben Objekte in einer gemeinsamen Umgebung. Unabhängig davon, welche Sensoren die Daten erfasst haben, können sie einem Messwert widersprechen, der von anderen Sensoren oder einem anderen Sensor-System stammt. Es sollte stabile und zuverlässigen Methoden geben, die Daten ohne Widersprüche auswerten und analysieren.
- Vollständigkeit der Daten: Unvollständige Daten können als unsichere Daten klassifiziert werden. Fehlende Daten oder Datenlücken können zu Problemen bei Auswerte- und Analyseansätzen führen.
- Datenmenge: Durch die kontinuierlichen technologischen Fortschritte der letzten Jahre hat sich der Messdatenerfassungsprozess grundlegend verändert. So stehen dem Anwender beispielsweise nach der Vermessung eines Objekts mittels Laserscanning innerhalb kürzester Zeit erhebliche Datenmengen in Form von 3D-Punktwolken zur Verfügung.

Bevor die Daten genutzt werden können – sei es für Vermessungsprojekte, für die technische Planung oder für den Einsatz in einem geografischen Informationssystem – müssen sie verarbeitet werden. Einer der wichtigsten Aspekte dabei ist die Tatsache, dass Beobachtungen stets Unsicherheiten aufweisen. In geodätischen Beobachtungen können Unsicherheiten in verschiedenen Größenordnungen auftreten und je nach Aufgabenstellung, Sensorik und Umgebungsbedingungen unterschiedliche Charakteristika aufweisen. Die Abweichungen werden in „Ausreißer“, systematische, und zufällige

Abweichungen eingeteilt. Es kann also uneingeschränkt festgestellt werden, dass (1) keine Messung exakt ist, (2) jede Messung Abweichungen beinhaltet, (3) der wahre Wert einer Messung nie bekannt ist, und somit (4) die genauen Größen der vorhandenen Abweichungen immer unbekannt sind (siehe Ghilani und Wolf, 2010).

Insbesondere von dem Hintergrund, dass die neuen Sensortechnologien die Charakteristik der gesamten Messunsicherheiten verändern, sind neue Mess- und Auswertemodelle zu entwickeln, weil z. B. (i) systematisch wirkende Abweichungen im Beobachtungsmaterial immer bedeutender gegenüber den stochastischen Abweichungen werden, (ii) der Ausreißeranteil bzw. zumindest die absolute Anzahl an Ausreißern in den Daten weiter zunimmt und (iii) die Verteilung der zufälligen Abweichungen nicht als normalverteilt angenommen werden kann. Demzufolge kann die Qualität der erfassten Daten stark variieren. Die Möglichkeit, aus Daten mit eingeschränkter Qualität aufgrund eines oder mehrerer der oben genannten Merkmale aussagekräftige und statistisch einwandfreie Ergebnisse zu liefern, ist eine große Herausforderung. Eine typische Herausforderung ist beispielsweise die Analyse großer Datenmengen wie Punktwolken, die durch Laserscanning entstehen. Um diese Herausforderung erfolgreich zu bewältigen, werden im Rahmen der Arbeit fortgeschrittene **computergestützte Ansätze zur geodätischen Datenanalyse** entwickelt. Der Begriff “computergestützte Ansätze zur geodätischen Datenanalyse” wird im Kapitel 1.1.3 näher erläutert.

1.1.2 Statistische Modelle

Offensichtlich bezieht sich die computergestützte geodätische Datenanalyse auf die traditionelle Disziplin der Statistik. Bevor die eigentliche computergestützte geodätische Datenanalyse definiert werden kann, ist es daher notwendig, ein Gesamtbild zu geben, was mit dem Bereich der traditionellen Statistik gemeint ist. Auf der einfachsten Ebene beschäftigt sich die Statistik mit der Transformation von Rohdaten in Fachwissen (Wegman, 1988). Wenn Anwendungsbereiche betroffen sind, die die Analyse von Rohdaten erfordern, muss sich jeder Wissenschaftler/Anwender mit folgenden Fragen befassen:

- Welche Daten sollten zur Beantwortung der Fragen in der Analyse verwendet werden?
- Sollen die erhobenen Daten für die Modellierung und die Analyse direkt verwendet werden (direkte Beobachtungen) oder soll zunächst transformiert (indirekte Beobachtungen) werden?
- Wird ausschließlich von zufälligen Messabweichungen ausgegangen (charakterisiert z. B. durch eine Varianz-Kovarianz-Matrix (VKM)) oder sind weitere systematische wirkende Anteile oder gar Ausreißer enthalten?
- Welche Schlussfolgerungen sollen aus den Daten gezogen werden?
- Inwieweit können diese Schlussfolgerungen vertrauenswürdig sein?

Vor einer Diskussion über Inferenzmethoden ist das Gesamtbild für eine bestimmte Perspektive zu betrachten (siehe Abbildung 1.1). Wenn ein vorliegendes Modell sorgfältig ausgewählt wurde, dann sollte es auch für die Praxis im Hinblick auf Prädiktion oder Entscheidungen relevant sein. Zur Frage, ob es sich um ein zutreffendes Modell handelt, ist es notwendig, die Daten zu sondieren, die von der realen Welt erhoben wurden und diese dann für die Modellierung aufzubereiten. Im Mittelpunkt steht das Feld der Inferenz und Statistik, das in den letzten Jahren einen radikalen Wandel vollzogen hat.

Funktionales und stochastisches Modell

Nach der Voruntersuchung und nach der Überprüfung der Datenqualität startet die formale Analyse mit einem vorläufigen Modell, welches in der Regel zu bestimmende Parameter enthält. Diese Parameter, deren Werte durch die Beobachtungen zu schätzen sind, sind geometrische, physikalische

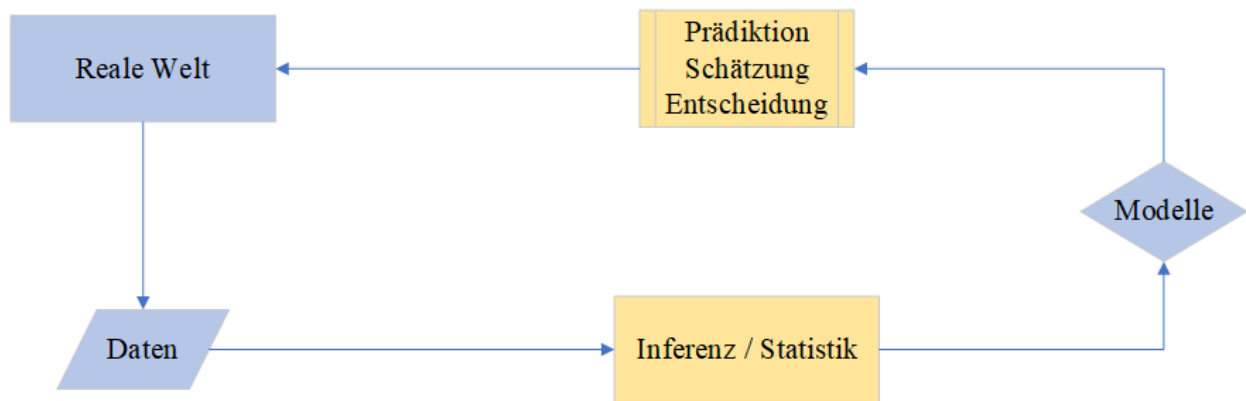


Abbildung 1.1: Gesamtbild Inferenz/Statistik

oder stochastische Größen, die eine genau festgelegte Aufgabe innerhalb des Modells aufweisen. Die mathematischen Beziehungen zwischen den Beobachtungen und den wahren Parametern sind dabei oft nicht-linear. Die funktionalen Beziehungen zwischen den Beobachtungen und den Parametern werden überlagert von zufälligen ϵ und systematischen Messabweichungen sowie ggf. Ausreißern. Im rein stochastischen Modell sind sämtliche Annahmen hinsichtlich der statistischen Eigenschaften dieser Abweichungen enthalten. Oft werden diese zufälligen Messabweichungen als normalverteilt mit $\epsilon \sim \mathcal{N}(\mathbf{0}, \Sigma_{II})$ angenommen. Wie jedoch bereits einleitend erläutert, greift dieser Ansatz zu kurz. Es müssen sowohl systematische Abweichungen als auch Ausreißer sowie nicht-normalverteilte zufällige Abweichungen in die Beschreibung der Unsicherheiten einbezogen werden.

Parameterschätzung und Hypothesentests

Liegt ein funktionales Modell und ggf. ein stochastisches Modell vor, können die Parameter anhand einer Schätzung ermittelt werden. Die Methode der kleinsten Quadrate (MkQ) stellt eine weit verbreitete Methode zur Parameterschätzung dar, bei der die (ggf. die gewichtete) Quadratsumme der Residuen minimiert wird. Im Gegensatz zur MkQ erfordert die Maximum-Likelihood (ML)-Methode die Festlegung einer Familie von Wahrscheinlichkeitsverteilungen, die durch Wahrscheinlichkeitsdichtefunktionen definiert und deren unbekannte Parameter geschätzt werden. Wenn die Beobachtungen normalverteilt sind, dann stimmen die Schätzergebnisse beider Methoden überein. Liegen a-priori Informationen hinsichtlich der zu schätzenden Parameter in Form einer a-priori PDF vor, kann die Bayessche Schätzung verwendet werden. Der Bayessche Ansatz ermöglicht eine intuitive Methode zur Ableitung von Statistiken. Drei Regeln, die durch logisches und konsistentes Denken abgeleitet werden, reichen für die Ableitung aus (siehe, z. B. Koch, 2018). Sie führen zu a-posteriori PDFs, Punkt- und Bereichsschätzungen, sowie Hypothesentests. Die Punktschätzung ergibt eine zahlenmäßige Realisierung für den unbekannten Parametervektor aus gegebenen Messungen, während die Bereichsschätzung ein Intervall oder einen Bereich liefert, in dem sich die unbekannten Parameter mit einer festgelegten Wahrscheinlichkeit befinden. Bei einem Hypothesentest wird geprüft, ob die zu testenden Parameter in einer bestimmten Teilmenge des Parameterraums liegen („Nullhypothese“) oder nicht.

Die drei genannten Inferenzmethoden (Punkt-, Bereichsschätzer und Hypothesentest) können sowohl im Batch-Processing (d. h. Auswertung aller Beobachtungen in einer Gesamtausgleichung) als auch rekursiv eingesetzt werden. Letzteres ist relevant in zahlreichen Problemstellungen der geodätischen Praxis, z. B. in den Bereichen der Deformationsanalyse und der Georeferenzierung von MSS, wobei vorhandene Lösungen mit aktuellen Beobachtungen zu kombinieren sind. In solchen Fällen ist es wünschenswert, nach jeder Hinzunahme neuer Beobachtungen die berechneten Statistiken (z. B. Parameterschätzwerte und Testgrößen) zu aktualisieren. Insbesondere besteht der Prozess der rekursiven Schätzung darin, nach jeder Epoche eine Korrektur der Lösung zu ermitteln,

die aus den aktuellen Beobachtungen ermittelt wird.

Mit Hilfe von Hypothesentests wird versucht, anhand von Stichprobendaten Schlussfolgerungen über die wahre PDF der Messgrößen zu ziehen. Hypothesentests sind somit geeignet zur Detektion von Inkonsistenzen zwischen den Daten und dem verwendeten Modell. Hypothesentests werden in allen Phasen des Analyseprozesses angewandt, insbesondere zur Detektion von Ausreißern, zur Modellauswahl, zur Feststellung systematischer Abweichungen im Modell oder lediglich zur Überprüfung der Korrektheit der Implementierung eines Schätzverfahrens anhand simulierter Daten, siehe u. a. Förstner und Wrobel (2016) sowie Koch (1999).

1.1.3 Computergestützte und simulationsbasierte geodätische Datenanalyse

Viele klassische statistische Methoden (Regression, lineare Parameterschätzung, Hypothesentests, Konfidenzintervalle, usw.), die in den letzten Jahrzehnten entwickelt wurden, sind der Wissenschaft wohl vertraut und werden in vielen geodätischen Teildisziplinen verwendet (siehe z. B. Koch, 1999, 2000; Niemeier, 2008; Caspary, 2013; Jäger et al., 2016; Ghilani und Wolf, 2010; Mikhail und Gracie, 1981). Die eingangs erwähnten Charakteristika der Beobachtungen erfordern jedoch eine neue Sichtweise auf die Datenanalyse, die sowohl neue methodische als auch numerische Vorgehensweisen erfordert.

Die im Rahmen der Arbeit durchgeführten Neuentwicklungen sollen unter dem Dach der computergestützten statistischen Datenanalyse einsortiert werden. Wegman (1988) definiert computergestützte Statistik als eine Zusammenstellung von Techniken, die einen starken Fokus auf der Nutzung von Computern bei der Entwicklung neuer statistischer Methoden haben. Viele dieser Methoden wurden nach der Entwicklung preiswerter Computer-Hardware seit den 1980er Jahren ermöglicht. Die Computerrevolution hat es Wissenschaftlern und Ingenieuren gestattet, große Datenmengen zu speichern und zu verarbeiten. Diese Daten werden jedoch in der Regel ohne eine klare Vorstellung davon erhoben, wofür sie in einer entsprechenden Untersuchung verwendet werden sollen. Z. B. werden in der modernen Datenanalyse oft Daten erhoben und erst im Nachgang wird eine Studie entworfen, um sinnvolle Informationen daraus zu gewinnen.

Da die Speicherung und Sammlung von solchen Daten kostengünstig ist, sind die Datensätze, mit denen sich die Analyse heute auseinandersetzen muss, sehr groß und hochdimensional (z. B. Daten aus Laserscanning-Punktwolken). In Situationen wie diesen sind viele der klassischen Methoden der Statistik unzureichend. Efron und Tibshirani (1991) bezeichnen computergestützte Statistik als „*Computer intensive statistische Methoden*“. Typische Beispiele für diese Methoden sind Expectation Maximization (EM)- und heuristische Optimierungsmethoden zur Minimierung oder Maximierung einer Schätzfunktion, MC-Methoden (Kroese et al., 2011; Gentle, 2003), Bootstrapping (Efron, 1979), parametrische und nichtparametrische Regressionsanalyse (Martinez und Martinez, 2015) sowie Klassifizierung und Regressionsbäume (Bishop, 2009).

Ein wichtiger Bestandteil moderner computergestützter Statistik sind simulationsbasierte Datenanalysen. Sie nutzen taktile und rechnerische Simulationen, um inferentielle Techniken, wie Bootstrapping für Konfidenzintervalle, und simulationsbasierte Hypothesentests durchzuführen. Ähnlich wie bei realen Daten können computersimulierte Daten viele verschiedene Formen annehmen. Im Gegensatz zu realen Daten sind Simulationsdaten jedoch streng reproduzierbar, da das Simulationsmodell vollständig bekannt ist. Darüber hinaus wird die Datenmenge, die mit Hilfe eines Simulationsmodells gewonnen werden kann, nur durch die Zeitdauer der Simulationsdurchläufe und durch die Speicherkapazität des Computers beschränkt. Die durch ein Simulationsexperiment generierten Daten können als Ergebnisse von Zufallsvariablen, Zufallsvektoren, Zeitreihen oder stochastischen Prozessen betrachtet werden. Im Allgemeinen besteht das Ziel der Simulation darin, Rückschlüsse auf verschiedene Eigenschaften dieser zufälligen Objekte zu ziehen, wie z. B. deren Erwartungswerte, Verzerrung der Schätzer, Korrelationen und Verteilungsfunktionen.

1.2 Ziele der Arbeit

Sowohl die Modellierung der Qualität als auch die Schätzungsproblematik sind abhängig von der Charakteristik der Unsicherheiten und von den zu modellierenden funktionalen Abhängigkeiten zwischen Beobachtungen und Parametern. Erst die Berücksichtigung beider Aspekte ermöglicht die Aufstellung optimaler Analyseverfahren. Hier setzen die Ziele der vorliegenden Arbeit an. Der erste Themenblock widmet sich neuen Modellen und Modellierungsansätzen in der statistischen Inferenz (insb. in ML- und Bayesverfahren) für die klassische Parameterschätzung und für Filterungstechniken im Zustandsraum. Insbesondere sollen auf dieser Weise geeignete Modelle aufgestellt werden, die den aktuellen Charakteristika des Beobachtungsmaterials gerecht werden. So sind sowohl zufällige als auch systematische Abweichungen sowie Ausreißer im Unsicherheitsbudget der Beobachtungen abzubilden. Eine rein zufällige Modellierung greift aufgrund der notwendigen Berücksichtigung von Sensorik, Erfassungsprozess und Umgebungsbedingungen heutzutage zu kurz. Des Weiteren sind die entsprechenden funktionalen Modelle zur Beschreibung der Abhängigkeit zwischen den unsicheren Beobachtungen und den gesuchten Parametern als Konsequenz daraus zu erweitern. Dies schließt ausdrücklich auch die Aufstellung neuer robuster Schätzer mit ein.

Der zweite Themenblock widmet sich der (numerischen) Optimierung. Aufgrund der hoch nicht-linearen funktionalen Zusammenhänge sowie nicht-normalverteilten und ausreißerbehafteten Daten sind die klassischen und oftmals lokalen Optimierungsverfahren für die Schätzung der (Zustands)Parameter nicht in der Lage, realistische Ergebnisse zu liefern. Dies ist insbesondere darin begründet, dass die neu aufgestellten funktionalen Modelle im Zusammenspiel mit den Unsicherheiten der Beobachtungen eine globale Optimierung für die Schätzung der (Zustands)Parameter erfordern. Darüber hinaus erfordert die Berücksichtigung von Ausreißern eine iterative Vorgehensweise, um reale Fragestellungen mit großen Datenmengen effizient lösen zu können. Die für die Arbeit insbesondere relevanten Optimierungsmethoden sind (i) linearisierungsbasierte (analytische) Techniken wie EM-Algorithmus und Iterativ Regewichtete Kleinste Quadrate, engl. iteratively reweighted least squares (IRLS), (ii) kombinatorische Techniken wie genetische Algorithmen sowie (iii) simulationsbasierte MC und sequentielle MC Methoden.

Der dritte und letzte Themenblock der Arbeit widmet sich der induktiven Inferenz-Statistik. Im Vordergrund stehen dabei die Bewertung der Schätzergebnisse anhand von Hypothesentests mittels MC-Verfahren und anhand von Bootstrapping-Techniken. Durch diese numerischen Verfahren gelingt es, den nicht-normalverteilten und ggf. systematisch verfälschten Zielgrößen ein adäquates Unsicherheitsbudget zuzuweisen und somit realistische Bereichsschätzer für die (Zustands)Parameter zu bestimmen. Alle innovativen Modellansätze wurden anhand von neuen entwickelten Algorithmen realisiert und in Simulationsexperimenten getestet, ausgewertet und validiert. Sie sind über die Geodäsie hinaus anwendbar, wurden aber insbesondere anhand von Anwendungen im Bereich der Ingenieurgeodäsie und der Immobilienbewertung untersucht und kritisch beurteilt.

1.3 Übersicht der originalen Publikationen zur kumulativen Habilitation

Im weiteren Verlauf gliedert sich die Habilitation in zwei Teile:

Synopse der Habilitation

Nach der Einleitung stellt das **Kapitel 2** die neu entwickelten Algorithmen zur Durchführung von robusten Parameterschätzungen basierend auf ML- und Bayesschen Schätzern aus redundanten Messungen im Kontext des Gauß-Markov-Modells (GMM) zusammen. Zur Demonstration verschiedener entwickelter Optimierungsverfahren werden dafür drei unterschiedliche parametrische Regressionsmodelle ausgewählt und ausführlich beschrieben: B-Spline Regressionsmodelle zur Kurvenapproximation (*MODELL I*), Regressionsmodelle mit autoregressiven Prozessen und student-verteilter Messrauschen (*MODELL II*) und Bayessche robuste Regressionsmodelle mit t-verteilterm

Messrauschen (*MODELL III*). **Kapitel 3** gibt eine ausführliche Erläuterung der optimalen Gaußschen approximativen rekursiven Bayesschen Schätzung an. Hierfür werden drei verschiedenen Filterungsansätze zur Schätzung von Zustandsparametern entwickelt: Stochastische analytische Filterungstechniken (*FILTERUNGSANSÄTZE I*), stochastische simulationsbasierte Filterungsansätze (*FILTERUNGSANSÄTZE II*) und kombinierte stochastische und mengenbasierte Filterungstechniken (*FILTERUNGSANSÄTZE III*). In **Kapitel 4** werden eine Reihe von neu entwickelten simulationsbasierten Methoden für die Inferenzstatistik im Bereich der Unsicherheitsmodellierung, der Konfidenzbereichsschätzung und der Hypothesentests vorgestellt (*BEREICHSSCHÄTZUNG I*). Bei der Konfidenzbereichsschätzung wird gezeigt, wie Messunsicherheit von nicht-linearen Funktionen behandelt und wie anhand von Bootstrapping-Techniken die Kovarianzmatrizen von geschätzten Parametern bestimmt werden (*BEREICHSSCHÄTZUNG II*). Im zweiten Teil dieses Kapitels wird auf die Durchführung von statistischen Hypothesentests anhand simulationsbasierter Algorithmen fokussiert. Der simulationsbasierte Hypothesentest-Ansatz wird anhand von zwei unterschiedlichen Testproblemen demonstriert: Monte-Carlo-basierte Ansätze für die Modellwahl zur Oberflächenmodellierung von 3D-Punktwolken (*TESTPROBLEM I*) und Bootstrapping-basierte Ansätze für das Testen auf Zeitvariabilität eines AR-Prozesses (*TESTPROBLEM II*). In **Kapitel 5** werden die wichtigsten Ergebnisse dieser Arbeit diskutiert und zusammengefasst, Schlussfolgerungen gezogen und mögliche Richtungen für die zukünftige Forschung aufgezeigt.

Originalveröffentlichung

Die in dieser kumulativen Habilitationsschrift relevanten eigenen Publikationen des Autors wurden in peer-reviewed Medien veröffentlicht und auf zahlreichen Konferenzen und Workshops präsentiert. Eine Auflistung dieser Veröffentlichungen ist in Kapitel 6 zu finden. Die wesentlichen Abhängigkeiten zwischen den Kapiteln und den eigenen Publikationen sind in der Abbildung 1.2 dargestellt. In grün gekennzeichnete Kreise und Rechtecke deuten drauf hin, dass die Publikationen in begutachteten Zeitschriften erschienen sind. Die orangene Farbe bedeutet, dass die Beiträge in begutachteten Konferenzbeiträgen veröffentlicht wurden.

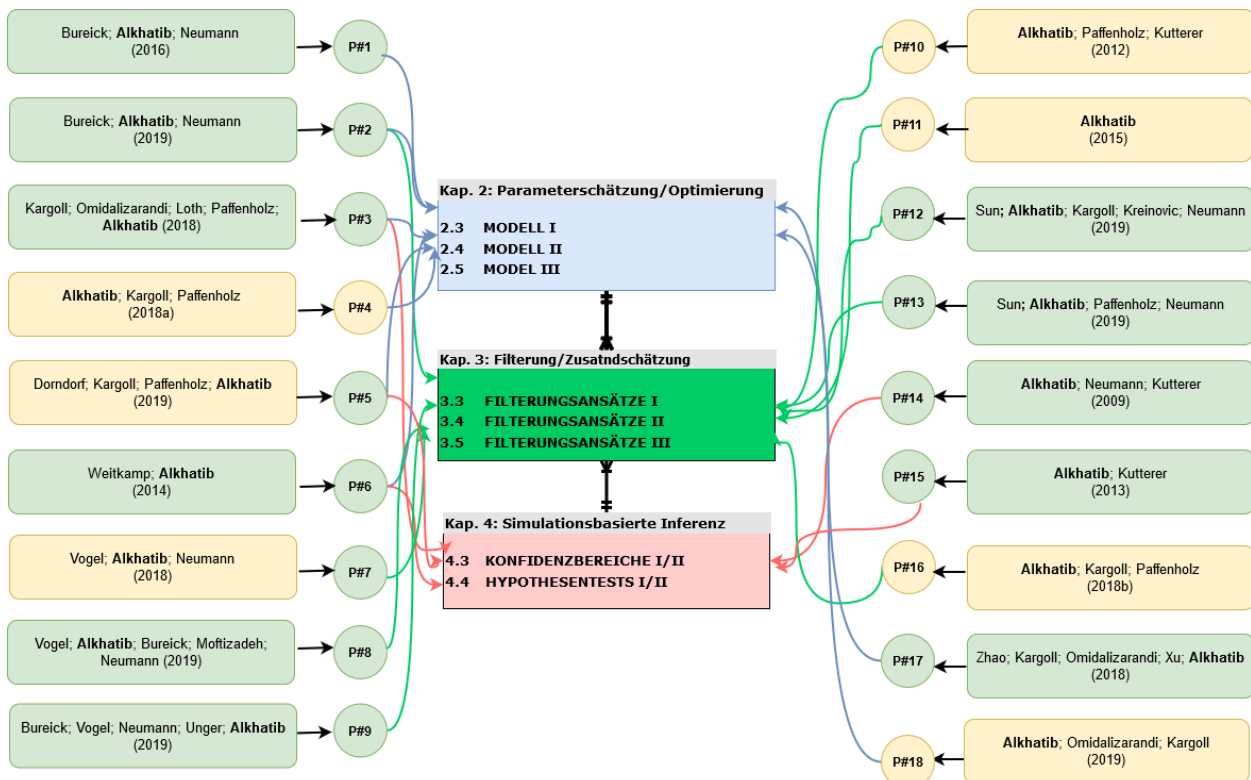


Abbildung 1.2: Übersicht der originalen Publikationen zur kumulativen Habilitation

Teil I

TEIL 1: Synopse der Habilitation

2 Optimierungsverfahren und Parameterschätzung in linearen und nicht-linearen Modellen

2.1 Einführung in die Parameterschätzung in linearen Regressionsmodellen

Die Schätztheorie bildet den Rahmen für die Methodik zur Schätzung unbekannter Parameter aus gegebenen verrauschten Beobachtungen. Es wird von einem mathematischen parametrischen Modell ausgegangen, das aus einem funktionalen und einem stochastischen Modell des Beobachtungsprozesses besteht. Das funktionale Modell spezifiziert die angenommenen Zusammenhänge zwischen den Beobachtungen und den unbekannten Parametern. Die Beobachtungen können als Stichproben aus einer Folge von Zufallsvariablen $\mathcal{L} = [\mathcal{L}_1, \dots, \mathcal{L}_n]$ interpretiert werden. Die Gültigkeit dieses funktionalen Modells kann in der Regel anhand von Hypothesentests überprüft werden (siehe Kapitel 4). Das stochastische Modell spezifiziert die probabilistischen Eigenschaften des Beobachtungsprozesses, d. h. die mit den Beobachtungen zusammenhängenden Zufallsvariablen und eventuell die Vorinformationen über die unbekannten Parameter. Diese Spezifikationen sind nicht in jedem Fall vollständig. Das stochastische Modell selbst kann also unbekannte Parameter enthalten, die ebenfalls aus den gegebenen Beobachtungen geschätzt werden müssen. Da die Verteilung der stochastischen Variablen oft unbekannt ist, kann es sinnvoll sein, nicht die gesamte Verteilung, sondern nur einige ihrer Eigenschaften, wie z. B. das zweite Moment, und damit Varianzen und Kovarianzen anzugeben. Selbst eine einfache relative Gewichtung der Beobachtungen kann eine sinnvolle Wahl sein. Bei dem Schätzungsprozess ist die besondere Art des Wissens über die Unsicherheit der Beobachtungen zu berücksichtigen. Wie für das funktionale Modell kann auch die Gültigkeit eines stochastischen Modells anhand von Hypothesentests überprüft werden (siehe Kapitel 4). Ausgehend von einem gewählten mathematischen Modell kann die Parameterschätzung nach unterschiedlichen Schätzprinzipien erfolgen, basierend auf den verfügbaren Informationen. Die zumeist verwendeten Schätzer in der geodätischen Datenanalyse sind: Kleinste Quadrate (KQ)-Schätzer, ML-Schätzer und der Bayessche Schätzer (Maximum-a-posteriori (MAP)) (Koch, 1999, 2000; Förstner und Wrobel, 2016; Jäger et al., 2005).

Eine gängige mathematische Modellstruktur, die in vielen Disziplinen, einschließlich der Geodäsie, verwendet wird, weist die Form eines GMMs auf: Beobachtungen $\mathcal{L} \in \mathbb{R}^n$ werden hierbei durch (linear oder nicht-linear) Funktionen unbekannter Parameter $\Theta \in \mathbb{R}^u$ sowie additivem Rauschen ϵ in der Form

$$\mathcal{L} = f(\Theta) + \epsilon \tag{2.1}$$

beschrieben. Das additive Rauschen ϵ wird oft als normalverteilt angenommen, symbolisch $\epsilon \sim \mathcal{N}(\mathbf{0}, \Sigma_u)$, wobei Σ_u die VKM der Beobachtungen symbolisiert. Von der Normalverteilung abweichende Annahmen werden gelegentlich genutzt, z. B. für den Fall von auftretenden Ausreißern. Ausreißer sind ein unvermeidliches Problem bei der Erfassung von geodätischen Daten, die mit modernen geodätischen Sensoren gemessen werden. Sie verzerren die Schätzungen der parametrischen Modelle auf unterschiedliche signifikante Weise. Die für eine geodätische Datenanalyse am häufigsten betrachteten robusten Schätzer sind M -Schätzer mit einer beschränkten Verlustfunktion; hier insbesondere die L_1 - und L_p -Norm-Schätzer (siehe Nowel und Kamiński, 2014; Marx, 2013;

Götzelmann et al., 2006; Junhuan, 2005; Marshall, 2002), Hubers M-Schätzer (siehe Huber, 1964; Bureick et al., 2016a; Koch, 2013; Guo et al., 2010; Götzelmann et al., 2006; Chang und Guo, 2005; Hekimoglu und Berber, 2003) und robuste ML-Schätzer, basierend auf einer langschwänzigen-Verteilung (z.B. Wiśniewski, 2014).

Aus den genannten Gründen existiert in der Regel keine offensichtliche optimale Lösung für das Schätzproblem, vielmehr kann es mehrere Lösungen geben. Nach der Modellformulierung im Gauß-Markov-Modell (GMM) kann ein Optimierungsalgorithmus (meist mit Hilfe von computergestützten Algorithmen) zur Lösung des Schätzproblems (Optimierung der ML- bzw. MAP-Schätzung) beitragen. Es gibt keinen universellen Optimierungsalgorithmus, sondern eine Reihe von Algorithmen, die jeweils auf eine bestimmte Art von Optimierungsproblem zugeschnitten sind. Eine umfassende Beschreibung verschiedener Optimierungsalgorithmen finden sich z. B. in Antoniou und Lu (2007); Nocedal und Wright (2006) sowie in Givens und Hoeting (2013). Die Entscheidung, den für eine bestimmte Anwendung geeigneten Algorithmus zu wählen, liegt oft beim Anwender. Diese Wahl ist unabdingbar, denn sie kann bestimmen, ob das Optimierungsproblem schnell oder langsam gelöst wird und ob die optimale Lösung überhaupt erreicht wird. Zu den verschiedenen Optimierungsalgorithmen sind Methoden zu zählen, die im Allgemeinen in drei Klassen unterteilt werden können:

- *Ableitungsbasierte* Optimierungsansätze, die die Nichtlinearität der Funktionen im funktionalen Modell in einem iterativen Schema behandeln (die bekanntesten Optimierungsalgorithmen sind der Gauß-Newton-Algorithmus, der Newton-Raphson-Algorithmus und der EM-Optimierungsalgorithmus),
- *simulationsbasierte* Optimierungsansätze, die anhand von MC Methoden die MAP-Schätzung numerisch approximieren,
- *kombinatorische* Optimierungsansätze, wie z. B. Genetische Algorithmen (GA) und Simulated Annealing.

Im Rahmen der Habilitation werden Aspekte aus drei parametrischen Modellen zur Erhaltung der bestmöglichen Anpassung an die Daten ausgewählt und diskutiert. Dabei werden eine Auswahl der statistischen Analysen, in denen Optimierungsprobleme auftreten, sowie eine Vielzahl von Methoden zu ihrer Lösung beschrieben. Sämtliche vorgestellte parametrische Modelle sind robust gegen Ausreißer:

- Zuerst wird das B-Spline Regressionsmodell zur Kurvenapproximation eingeführt. Es wird insbesondere auf das Problem der optimalen Knotenvektorauswahl eingegangen. Die robuste Kontrollpunktschätzung basiert auf der klassischen robusten M-Schätzung.
- Danach wird ein nicht-lineares raumzeitliches Regressionsmodell mit einem alternativen Unsicherheitsmodell, der Student-Verteilung (t-Verteilung), vorgestellt. Die Parameterschätzung erfolgt mittels EM-Algorithmus, welcher auf einer ML-Schätzung basiert.
- Anschließend wird ein nicht-lineares Bayessches Modell mit t-Verteilung vorgestellt. Der simulationsbasierte Optimierungsalgorithmus basiert auf MCMC zur Approximation der MAP-Schätzung.

2.2 Vorstellung der ausgewählten nicht-linearen Regressionsmodelle

MODELL I: B-Spline Regressionsmodelle für die Kurvenapproximation

Die Approximation von Punktwolken mittels Freiformkurven ist seit einigen Jahrzehnten eine wichtige Aufgabe in unterschiedlichen Ingenieursdisziplinen, wie z. B. in den Bereichen des Reverse Engineering, der Automobilindustrie oder im Bereich der geometrischen Modellierung in der Bildverarbeitung. Für die Erfassung von Punktwolken hat sich der Laserscanner als geeignete Messtechnik für

verschiedenste Anwendungen in der Geodäsie und in anderen Fachdisziplinen, wie z. B. dem Bauingenieurwesen, erwiesen. Das Laserscanning stellt beispielsweise die Datengrundlage für die 3D-Oberflächenmodellierung mittels Non-Uniform Rational B-Splines (NURBS) und B-Splines (Koch, 2009a; Harmening und Neuner, 2015a; Bureick et al., 2016b; Wu et al., 2000) zur Verfügung. Das Laserscanning wird auch für die Überwachung verschiedener Strukturen wie Schienen, Gewölbe oder Tunnel (Bureick et al., 2016a; Xu et al., 2017a,b, 2018) eingesetzt. Laserscanner liefern jedoch Punktwolken, die von unvermeidlichem Rauschen betroffen sind und Datenlücken aufweisen können. Darüber hinaus kann das Messobjekt in Form von Spitzen, Kanten oder Sprüngen eine komplexe Geometrie aufweisen. Freiformkurven, insbesondere B-Spline-Kurven, sind im Allgemeinen gut geeignet, um solche Herausforderungen im Approximationsprozess zu bewältigen. Neben einer Glättung der Daten können B-Spline-Kurven leicht an komplexe Objekte angepasst werden, ohne dass die allgemeine Funktionsbeziehung geändert werden muss. Die resultierende B-Spline-Kurve benötigt deutlich weniger Speicherplatz als die Rohpunktwolke und kann für weitere Auswertungsschritte einfach in Modellierungssoftware, wie z. B. Computer-Aided Design-Software (CAD), überführt werden. Außerdem werden in der Regel durch die Approximation Ausreißer aufgedeckt.

Jedoch müssen einige Anforderungen berücksichtigt werden, um eine gute Approximation durch eine B-Spline-Kurve zu ermöglichen. Die Anordnung der Knoten in der B-Spline-Approximation, bekannt als Knoten-Optimierungs-Problem, welches ein multimodales und multivariates kontinuierliches nicht-lineares Optimierungsproblem (Dierckx, 1993; Gálvez et al., 2015) darstellt, kann einen signifikanten Einfluss auf die endgültige Schätzung der Kurve bzw. der Oberfläche haben. Da es keinen analytischen Ansatz für optimale Knotenpositionen gibt, wurden in der Fachliteratur verschiedene Methoden zur Auswahl und Optimierung von Knotenvektoren diskutiert. Einige deterministische Methoden verwenden beispielsweise die Position (Piegl und Tiller, 1997) oder die Krümmung (Park und Lee, 2007) der Punktwolke, um den Knotenvektor zu bestimmen. Bei komplexen Punktwolken sind die Ergebnisse jedoch weit von der optimalen Lösung entfernt. Schmitt und Neuner (2015) stellten eine gemeinsame Schätzung der Knoten mit den Kontrollpunkten vor, durch die sich ein hochgradig nicht-lineares Gleichungssystem ergibt. Zur Stützung des nicht-linearen Gleichungssystems werden Bedingungsgleichungen und verbesserte Näherungswerte eingeführt. Alternativ liefern metaheuristische Methoden wie künstliche Immunsysteme (Gálvez et al., 2015), Genetische Algorithmen (Yoshimoto et al., 2003) oder Schätzung von Verteilungen (Zhao et al., 2011) Knotenvektoren, die sehr nahe am Optimierungsziel liegen, aber nur langsam konvergieren und daher Zeit und Rechenleistung benötigen. Darüber hinaus wird die Leistungsfähigkeit solcher Algorithmen durch das Auftreten von Ausreißern und Datenlücken erheblich beeinträchtigt. Ausreißer verzerren die Schätzung der Kontrollpunkte auf unterschiedliche Weise.

Kurze Vorstellung des innovativen Beitrags – MODELL I

In Abschnitt 2.3 wird die robuste B-Spline Regression als Approximationsmethode ausgewählt. In diesem Zusammenhang wurden zwei Verfahren **P#1** in Bureick et al. (2016a) und **P#2** Bureick et al. (2019a) entwickelt, welche die optimale Beschaffenheit der Knotenvektoren anhand heuristischer Optimierungsmethoden (MC-simulationsbasiert anhand einer MC-Methode sowie kombinatorische Optimierungsverfahren anhand einer GA-Optimierung) bestimmen. Außerdem wurde die ML-Schätzung für unterschiedliche M-Schätzer (Huber- und Hampel-Schätzer), mit der Methode der IRLS eingesetzt, um den Einfluss der möglicher Ausreißer zu minimieren.

MODELL II: Nicht-lineare Zeitreihenmodelle mit t-verteilter Messrauschen

Die Normalverteilung spielt in der geodätischen Ausgleichstheorie traditionell aus praktischen und theoretischen Gründen eine bedeutende Rolle (cf. Koch, 1999). Wie Helmert (1907) feststellte, ist es erfahrungsgemäß bekannt, dass die Gaußsche Verteilungsannahme dem Auftreten zufälliger Messabweichungen in der Regel sehr gut entspricht; die Form des Unsicherheitsgesetzes kann jedoch

nur durch Beobachtungen bestimmt werden. Inzwischen liegen zunehmend Indizien dafür vor, dass zufällige Abweichungen in vielen Fällen tatsächlich eine nicht-Gaußsche (vgl. Box und Andersen, 1955; Orlov, 1991), langschwänzige Wahrscheinlichkeitsverteilung besitzen. So diskutiert Lehmann (2015) Beobachtungsunsicherheitsgesetze neben der Laplace-Verteilung und der Verteilung basierend auf der Huber-Funktion (Huber, 1964) (die die probabilistischen Grundbausteine der L_1 -Norm und Hubers M-Schätzer sind) auch mit der skalierten t-Verteilung. Insbesondere die t-Verteilung, die auch im vorliegenden Beitrag als Unsicherheitsgesetz verwendet wird, wurde in einer Reihe von Forschungsarbeiten (vgl. Nadarajah, 2009; Geweke, 1993; Fraser, 1976) und mit einer Reihe von nützlichen mathematischen Eigenschaften, die sich aus sinnvollen Anforderungen ergeben (vgl. Alkhatib et al., 2017a), erfolgreich angewendet.

Die Fokussierung auf die Ränder einer Verteilung ist bedeutend, da die Ränder mit Ausreißern zusammenhängen, sofern diese durch einen Schwellwert definiert werden (beispielsweise durch die bekannte 3σ -Regel (vgl. Lehmann, 2013)). Zwei große Nachteile weisen die für eine geodätische Datenanalyse am häufigsten betrachteten robusten Schätzer auf. Zum einen wurden diese Schätzer unter einer bestimmten Bedingung von Gewichtungen als nicht robust nachgewiesen (Xu, 2005), zum anderen sind diese Schätzer ungenau, wenn die wahre Verteilung vom angenommenen probabilistischen Modell abweicht (z. B. bei Anwendung des L_1 -Norm-Schätzers auf annähernd normalverteilte Beobachtungen). Robuste Schätzer, welche freie Verteilungsformparameter oder Tuningkonstanten (z. B. die Konstante k für Hubers M-Schätzer) beinhalten, sind in dieser Hinsicht offenbar flexibler. Solche Tuning-Konstanten müssen jedoch vor der tatsächlichen Parameterschätzung und damit meist in unzureichendem Kenntnisstand über die Ausreißereigenschaften festgelegt werden.

Im Folgenden soll die robuste Schätzung basierend auf der zuvor erwähnten skalierten t-Verteilung ausführlicher beschrieben werden, die neben einem Skalierungsparameter auch einen datenadaptierbaren Freiheitsgrad beinhaltet, um die Form der Ränder der definierten Dichtefunktion zu steuern. Lange et al. (1989) behandelt in einer ersten Veröffentlichung die Anwendung und den Nutzen der skalierten t-Verteilung in einer robusten ML-Schätzung für Regressionsmodelle. Wie bereits durch Dempster et al. (1977) dargelegt, kann diese Art der ML-Schätzung in einer computergestützten Form als iterativ regewichtete kleinste Quadrate (IRLS) ausgedrückt werden, wobei die Gewichte verwendet werden, um die Varianzen der zufälligen Abweichungen entsprechend ihrer Lage unter der PDF zu skalieren. Damit ist es möglich, den Freiheitsgrad der zugrunde liegenden t-Verteilung neben den Regressionsparametern und dem Skalenparameter zu schätzen und in einen sogenannten (teilweise) adaptiven Schätzer umzuwandeln.

In einem multivariaten Regressionsmodell wird jede Beobachtung als Zufallsvektor modelliert, der durch eine vektorwertige (möglicherweise nicht-lineare) deterministische Regressionsfunktion und einen Vektor der zufälligen Abweichungen definiert wird. Liu (1997) nahm eine multivariate t-Verteilung mit unbekanntem Skalenfaktor und unbekanntem Freiheitsgrad für jeden Vektor von zufälligen Abweichungen an und untersuchte verschiedene Formen des EM-Algorithmus zur Schätzung der unbekannten Modellparameter. Es wurde bereits zuvor in Liu und Rubin (1994) und Meng und Rubin (1993) gezeigt, dass die Varianten der Expectation Conditional maximization (ECM) und Expectation Conditional Maximization Either (ECME) die Konvergenz des EM-Algorithmus erheblich beschleunigen können. Um mit Modellen umzugehen, die keine geschlossene Lösungsgleichungen des EM zulassen, wurde das Optimierungsprinzip der Generalized Expectation Maximization (GEM) von Dempster et al. (1977) vorgeschlagen.

Die Idee ist hierbei, sich dem gesuchten Maximum der Likelihoodfunktion in jedem EM-Schritt anzunähern, anstatt zu versuchen, dieses vollständig zu erreichen. Häufig wurden GEM-Algorithmen mit Newton-Raphson-Schritten angewendet (siehe McLachlan und Krishnan, 2008). Ein GEM-Algorithmus kann insbesondere zur Auswertung nicht-linearer Regressionsmodelle eingesetzt werden. In dieser Konstellation wurde ein IRLS Algorithmus mit Gauß-Newton-Schritten als geeignete Methode (Phillips, 2002; Lange et al., 1989) festgestellt.

Neben langschwänzigen Verteilungen, multivariaten und nicht-linearen Beobachtungsmodellen liegt

ein weiterer Aspekt, der die Parameterschätzung deutlich erschwert, in der auftretenden Autokorrelation der zufälligen Abweichungen. So zeigen beispielsweise viele Arten von Sensordaten wie Inertialsensordaten, Satellitengravitationsgradiometriedaten und Global Navigation Satellite System (GNSS)-Daten ausgeprägte farbige Rauscheigenschaften (siehe z. B. Park und Gao, 2008; Schuh, 2003; Luo, 2013).

Kurze Vorstellung des innovativen Beitrags – MODELL II

Üblicherweise enthalten Zeitreihen mehrere Ausreißer, so dass ein robustes Schätzverfahren im Allgemeinen anzustreben ist. Um mit dieser Situation umzugehen, wurde der oben erwähnte teilweise adaptive Schätzer für Regressionsmodelle auf Basis der skalierten t-Verteilung in Kargoll et al. (2018b) **P#3** um AR zufällige Abweichungen erweitert, wobei die Komponenten des weißen Rauschens des AR-Prozesses unabhängig voneinander und identisch t-verteilt sind. Eine Limitierung dieser Methodik ist jedoch, dass die Beobachtungen lediglich eine univariate Zeitreihe mit einem nicht-linearen Regressionsmodell beschreiben. Das Ziel des Beitrags in Alkhatib et al. (2017b) **P#4** war es, das bestehende univariate, lineare Modell auf ein multivariates und nicht-lineares (differenzierbares) Regressionsmodell zu erweitern. Bezüglich des Aufbaus des AR-Modells wurde der Fall untersucht, dass jeder Zeitreihenkomponente ein univariater AR-Prozess individueller Ordnung zugeordnet ist, unabhängig von den AR-Prozessen der übrigen Komponenten. Die Modellierung von Abhängigkeiten in Form von Kreuzkorrelationen wurde daher ausgeschlossen; es handelt sich um eine Aufgabe, die den Einsatz von Vektor AR (VAR)-Prozessen erfordern würde und die über den Rahmen der vorliegenden Arbeit hinausgeht.

MODELL III: Lineare und nicht-lineare Bayesche Regressionsmodelle mit t-verteiletem Messrauschen

Die Bayessche Methodik bietet eine vielseitige und natürliche Möglichkeit zusätzliche Informationen aufzunehmen. Dabei werden verrauschte Daten ergänzt, indem die unbekannten Parameter auch als Zufallsvariablen modelliert werden. Die Darstellung der Lösung in Form einer a-posteriori PDF bietet eine Vielzahl von Möglichkeiten, um nützliche Schätzwerte zu ermitteln. Die Bayessche Regression eröffnet die Möglichkeit, Vorwissen hinsichtlich der unbekannten Parameter und Daten zu kombinieren. Die Bayessche Statistik basiert auf fundamentalen Wahrscheinlichkeitsgesetzen: Die Schätzung unbekannter Parameter eines Modells, die Schätzung der Konfidenzregionen, der Vergleich unterschiedlicher kausaler Modelle (Hypothesentests) und spezielle Filterungs- und Prädiktionsaufgaben werden aus den gleichen bedingten Wahrscheinlichkeitsgesetzen abgeleitet (Koop, 2010; Gelman, 2013; Koch, 2000). Ein charakteristisches Merkmal der Bayesschen Inferenz ist die Einführung von a-priori-PDF der unbekannten Parameter, bevor die Daten in Betracht gezogen werden. Die Bayessche Statistik wird seit Jahrzehnten ebenfalls für verschiedene geodätische Anwendungen verwendet (Koch, 2018; Schaffrin, 1987; Yuanxi, 1991; Zhu et al., 2005). Außerdem wurde in Alkhatib und Weitkamp (2012) sowie in Weitkamp und Alkhatib (2012) eine lineare multiple Bayessche Regression zur Integration von Expertenwissen für die Immobilienbewertung entwickelt. Alkhatib und Weitkamp (2013) sowie Weitkamp und Alkhatib (2014) erweiterten diesen Ansatz, um eine zuverlässige Auswertung auch in Lagen mit wenigen Kauffällen zu ermöglichen. Zu diesem Zweck wurde ein robustes Bayessches Regressionsmodell von Geweke (1993) eingeführt, welches die unabhängige Student-t-Verteilung im linearen Modell verwendet, sodass es auch in Fällen angewendet werden kann, in denen signifikante Abweichungen von den optimalen Annahmen auftreten. Im Abschnitt 2.5 wird die Bayessche Inferenz in Regressionsmodellen betrachtet, bei denen die unbekannten Parameter auf intrinsisch nicht-lineare Weise eintreten. Die a-posteriori PDF zeigt in diesem Fall keine geschlossene Form und ist somit im robusten Ansatz analytisch nicht lösbar – im Gegensatz zur klassischen Bayesschen Regression (unter Annahme der Normal-Gamma-Verteilung) nach Koch (2000). Damit ist eine Lösung des Optimierungsproblems nur numerisch möglich; diese erfolgt als MCMC-Algorithmus (Kroese et al., 2011; Gelman, 2013), indem eine stochastische Kette

generiert wird, welche zu der gesuchten Verteilung der zu schätzenden Parameter konvergiert. Aus dieser Kette können die verschiedenen Momente der Verteilungsfunktion ermittelt werden.

Kurze Vorstellung des innovativen Beitrags – MODELL III

Abschnitt 2.5 beschäftigt sich mit der Entwicklung eines robusten Bayesschen Schätzers für die Anwendung sowohl in linearen, als auch in nicht-linearen parametrischen Modellen. Die Robustheit des Bayesschen Ansatzes wird durch die Verwendung der t-Verteilung, welche die Normalverteilung in der Likelihood-Funktion ersetzt, realisiert. Die resultierende a-posteriori-PDF wird mit einer MCMC approximiert. Im Abschnitt 2.5 werden die verwendeten MCMC-Algorithmen beschrieben: Der Gibbs-Sampler und der Metropolis-Hastings-Algorithmus. Die konkrete Umsetzung des neuen entwickelten Algorithmus wurde in Dorndorf et al. (2019) **P#5** und in Weitkamp und Alkhatib (2014) **P#6** dargestellt. Der entwickelte Algorithmus wurde anhand von zwei Anwendungen mit realen und simulierten Datensätzen getestet und validiert.

2.3 MODELL I: Robuste B-Spline Regressionsmodelle

2.3.1 Mathematisches Modell und Algorithmen zur Optimierung von Knotenvektoren

Ein auf einer B-Spline-Kurve liegender 3D-Kurvenpunkt $\mathbf{C}(\bar{u})$ wird aus der Summe der Linearkombinationen der Basisfunktionen $N_{i,p}(\bar{u})$ und der 3D-Kontrollpunkte $\mathbf{x}_i$ gebildet:

$$\mathbf{C}(\bar{u}) = [x(\bar{u}), y(\bar{u}), z(\bar{u})]^T = \sum_{i=0}^n N_{i,p}(\bar{u}) \mathbf{x}_i \text{ mit } \mathbf{x}_i = [x_i, y_i, z_i]^T. \quad (2.2)$$

Formal tragen $n + 1$ Linearkombinationen und damit die gleiche Anzahl von Kontrollpunkten und Basisfunktionen zu $\mathbf{C}(\bar{u})$ bei. p gibt den Grad der Basisfunktion und $p + 1$ die Ordnung der B-Spline-Kurve an. Lediglich $p + 1$ Kontrollpunkte und Basisfunktionen beeinflussen praktisch einen bestimmten Kurvenpunkt $\mathbf{C}(\bar{u})$. Diese Eigenschaft wird als lokale Unterstützung bezeichnet (siehe Piegler und Tiller (1997)). Der Ortsparameter $\bar{u}$ legt die Position des Kurvenpunktes auf der B-Spline-Kurve fest. Der erste Ortsparameter wird in der Regel mit 0 und der letzte Ortsparameter mit 1 festgelegt; dieses Regelschema wird in der vorliegenden Arbeit übernommen.

Zur Berechnung der Basisfunktionen $N_{i,p}(\bar{u})$ wird die von Cox (1972) und de Boor (1972) eingeführte rekursive Formel (siehe Gl. 2.3 und 2.4) verwendet.

$$N_{i,0}(\bar{u}) = \begin{cases} 1 & \text{falls } u_i \leq \bar{u} < u_{i+1} \\ 0 & \text{sonst} \end{cases} \quad (2.3)$$

$$N_{i,p}(\bar{u}) = \frac{\bar{u} - u_{i+p}}{u_{i+p} - u_i} N_{i,p-1}(\bar{u}) + \frac{u_{i+p+1} - \bar{u}}{u_{i+p+1} - u_{i+1}} N_{i+1,p-1}(\bar{u}) \quad (2.4)$$

Zusätzlich zum Grad der Basisfunktion p , der Anzahl der Kontrollpunkte $n + 1$ und des Ortsparameters $\bar{u}$ beeinflussen die Knoten u_i mit $i \in \{0, \dots, m\}$ die Berechnung der Basisfunktionen in Gl. 2.3 und 2.4. Die $m + 1$ Knoten werden im Knotenvektor $\mathbf{U}$ in einer monoton steigenden Reihenfolge gespeichert (siehe Gl. 2.5):

$$\mathbf{U} = [u_0, \dots, u_m] \text{ mit } u_i \leq u_{i+1}, i \in \{0, \dots, m - 1\}. \quad (2.5)$$

Die Knoten im Knotenvektor werden in externe und interne Knoten unterteilt. Die ersten $p + 1$ Knoten erhalten den Wert 0, die letzten $p + 1$ Knoten erhalten den Wert 1. Diese Knoten werden als externe Knoten bezeichnet. Eine Basisfunktion beginnt oder endet bei jedem externen Knoten. Die übrigen Knoten werden als interne Knoten bezeichnet. Bei jedem internen Knoten endet eine

Basisfunktion und eine weitere beginnt. Die Gesamtzahl der Knoten $m + 1$ steht in einem Zusammenhang mit der Anzahl der Kontrollpunkte und dem Grad der Basisfunktion. Die Berechnung von m wird durch $m = n + p + 1$ realisiert.

Bei der Anwendung von B-Spline-Kurven zur Approximation einer 3D ($d = 3$) Punktwolke $\mathbf{I}$ mit

$$\mathbf{I}_{[r,d]} = \begin{bmatrix} l_{x,1} & l_{y,1} & l_{z,1} \\ \vdots & \vdots & \vdots \\ l_{x,r} & l_{y,r} & l_{z,r} \end{bmatrix}, \quad (2.6)$$

bestehend aus r Punkten, müssen in der Regel vier Schritte durchgeführt werden: Modellauswahl, Parametrisierung, Knotenvektorbestimmung und Kontrollpunkt-Schätzung.

Schritt 1 - Modellauswahl: Im ersten Schritt wird der Grad der Basisfunktion p und die Anzahl der Kontrollpunkte $n + 1$ festgelegt. Ein üblicher Ansatz zur Bestimmung der optimalen Anzahl von $n + 1$ und p ist die Anwendung eines Informationskriteriums, insbesondere des Akaike-Informationskriterium (AIC) und des Bayessches Informationskriterium (BIC) (siehe z. B. Gálvez et al., 2015; Harmening und Neuner, 2016). Harmening und Neuner (2016) geben eine detaillierte Beschreibung des theoretischen Hintergrunds von AIC und BIC. Darüber hinaus stellen die Autoren einen neuen Ansatz vor, indem sie die aus der Theorie des statistischen Lernens stammende strukturelle Risikominimierung bei der Modellauswahl für B-Spline-Kurven anwenden.

Schritt 2 - Parametrisierung: Der zweite Schritt ist die Bestimmung eines Ortsparameters $\bar{u}$ für jeden Punkt in der Punktwolke. Diese Parametrisierung der Punktwolke wird häufig durch den Einsatz verschiedener deterministischer Methoden realisiert: Gleichmäßiger Abstand, Sehnenlänge oder Zentripetal (z. B. Piegler und Tiller (1997)). Für diese Verfahren sind sortierte Punktwolken erforderlich; dies ist meist der Fall, wenn die Punktwolke von einem Laserscanner erfasst wurde. Im Falle einer unsortierten Punktwolke kann die Methode von Ma und Kruth (1995) verwendet werden. Die Lageparameter werden iterativ verbessert, indem die Punkte der Punktwolke auf eine Basiskurve projiziert werden, die die B-Spline-Kurve der vorherigen Iteration abbildet. Anschließend werden die Lageparameter durch die der projizierten Punkte auf der Basiskurve aktualisiert.

Schritt 3 - Knotenvektorbestimmung: In diesem Absatz werden zwei Algorithmen zur Knotenvektorbestimmung vorgestellt, welche das Problem der Knotenvektorbestimmung effizienter und genauer lösen als bestehende Ansätze. Die beiden entwickelten Algorithmen basieren auf simulationsbasierten (MC-) und kombinatorischen (GA)-Optimierungstechniken und wurden in Bureick et al. (2016a) **P#1** bzw. in Bureick et al. (2019a) **P#2** veröffentlicht.

(a) Knotenvekturoptimierung mittels Elitären Monte-Carlo (EMC)-Algorithmus

Die Grundidee des in Bureick et al. (2016a) **P#1** vorgeschlagenen Elitären Monte-Carlo (EMC)-Algorithmus besteht darin, die Knotenvektoren, die die besten Ergebnisse liefern, zu nutzen, um eine Verteilung für die Knoten zu schätzen. Neue Knoten werden anhand der Gewichte erzeugt, die sich aus der geschätzten Verteilung ableiten. Dieser Vorgang wird so oft wiederholt, bis ein bestimmtes Abbruchkriterium erreicht ist. Der Algorithmus 1 beschreibt die Vorgehensweise der adaptierten MC-Methoden. Zu Beginn des Algorithmus wird die gesamte Knotenspanne, in der Regel von 0 bis 1, an jedem Ortsparameter $\bar{u}$ in Intervalle unterteilt. Insgesamt resultieren somit $r - 1$ Intervalle. Jedem Intervall ist eine bestimmte Wahrscheinlichkeit w_j zugeordnet. Die Menge der Gewichte wird nachfolgend als Wahrscheinlichkeit $\mathbf{R}$ oder $\mathbf{R}_{eq}$ bezeichnet. $\mathbf{R}_{eq}$ setzt sich aus gleichen Wahrscheinlichkeiten zusammen. In den ersten Schritten sind $\mathbf{R}$ und $\mathbf{R}_{eq}$ gleich. Beim Eintritt in die while-Schleife wird zunächst abgefragt, ob die Anzahl der Iterationen $iter$ den Wert N_{chance} überschreitet. Dann wird $\mathbf{R}$ in Abhängigkeit von der Population P_{best} – die Population ist

eine Teilmenge von Lösungen in der aktuellen Generation – berechnet, andernfalls wird $\mathbf{R}$ gleich gewichtet.

Die Wahrscheinlichkeit $\mathbf{R}$ wird in Abhängigkeit von der Population P_{best} berechnet. P_{best} enthält die N_b Individuen mit den bisher besten Fitnesswerten. Die Fitnesswerte werden mit Hilfe von einer Fitnessfunktion ermittelt. Die Fitnessfunktion bewertet, wie nahe eine bestimmte Lösung an der optimalen Lösung des gewünschten Problems liegt. Sie bestimmt, wie geeignet eine Lösung ist. Die Knoten sämtlicher Individuen von P_{best} werden in einem Vektor $\mathbf{kv}$ der Länge r_{kv} gespeichert und zur Schätzung einer Kerndichtefunktion $\hat{f}_h$ an einer beliebigen Position t verwendet:

$$\hat{f}_h(t) = \frac{1}{r_{kv} \cdot bw} \sum_{j=1}^{r_{kv}} K\left(\frac{t - kv_j}{bw}\right). \quad (2.7)$$

Die Kerndichte-Schätzung wurde ursprünglich von Rosenblatt (1956) und Parzen (1962) eingeführt. Für die Schätzung von $\hat{f}_h$ wird die MATLAB[®]-Routine *ksdensity* mit einer normalverteilten Kerndichtefunktion K an Position t_a verwendet:

$$K(t_a) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t_a^2}. \quad (2.8)$$

Der wichtigste Parameter für die Schätzung der Kerndichte ist die Bandbreite bw . Dabei wird bw mit der mittleren Breite der Intervalle gleichgesetzt:

$$bw = \frac{1}{r - 1}. \quad (2.9)$$

Die Bandbreite bw steuert die Glättung der geschätzten Kerndichte. Grundsätzlich ist die Wahl der Bandbreite bw eine entscheidende Stellschraube der Elitären Monte-Carlo (EMC)-Methode. Ziel ist es, klare Maxima in der Kerndichte zu erhalten, um in der Folge möglichst viele gute Knotenvektoren zu generieren. Gleichzeitig soll die Kerndichte in der direkten Nachbarschaft dieser Maxima ebenfalls leicht erhöht werden, um eine weitere Evolution der Knotenvektoren zuzulassen. Zudem soll bw für jegliche zu approximierende Punktwolken gelten. Die genannten Anforderungen werden erfüllt, wenn bw mit der durchschnittlichen Intervallbreite gleichgesetzt wird.

(b) *Knotenvektoroptimierung mittels Elitären genetischem Algorithmus (EGA)*

Die zweite heuristische Methode zur Optimierung der Knotenvektorbestimmung wurde auf der Basis von einem Elitären Algorithmus (EA) durchgeführt. Der Unterschied zwischen traditionellen Algorithmen und EA besteht darin, dass EA nicht statisch, sondern dynamisch sind, da sie sich mit der Zeit verändern können. Im Allgemeinen weisen EA drei Hauptmerkmale auf:

- **Populationsbasiert:** EA sollen einen Prozess optimieren, in dem die aktuellen Lösungen der Knotenvektorbestimmung nicht optimal sind, um neue, bessere Lösungsmengen zu generieren. Der Satz der aktuellen Lösungen, aus denen neue Lösungen generiert werden sollen, wird als Population bezeichnet.
- **Fitness-orientiert:** Jeder einzelnen Lösung ist ein Fitnesswert zugeordnet, der aus einer Fitnessfunktion berechnet wird. Dieser Fitnesswert spiegelt wider, wie gut die Lösung ist.
- **Variationsgesteuert:** Wenn es keine optimale Lösung in der aktuellen Population gibt, werden die einzelnen Individuen in der Population verändert, um so neue, bessere Lösungen zu generieren.

Die Idee für den entwickelten Elitären Genetischen Algorithmus (EGA) entstand durch den Vergleich des GA mit dem Clonal Selection Algorithm (CSA), siehe Gálvez et al. (2015). Der CSA löst

Algorithmus 1: Optimierung von Knotenvektoren in B-Spline mit Hilfe von Elitäre Monte-Carlo (EMC)

Input : Anzahl der maximalen Iterationen $iter_{\max}$;
 Anzahl der Iterationen N_{chance} mit gleichgewichtetem $\mathbf{R}$;
 Anzahl der Individuen N ;
 Anzahl der Individuen N_r ausgewählt aus gleichgewichtetem $\mathbf{R}_{\text{eq}}$;
 Anzahl der Individuen N_b in der Population P_{best} ;
Output: Bestes/Fittestes Individuum Ind_{best}

- 1 Teile die gesamte Knotenspanne an jedem Ortsparameter in Intervalle;
- 2 Weise jedem Intervall die gleiche Wahrscheinlichkeit zu und ordne sie $\mathbf{R}_{\text{eq}}$ zu;
- 3 Setze $\mathbf{R}_{\text{eq}}$ zu $\mathbf{R}$;
- 4 Setze die Laufvariable $iter = 1$;
- 5 **while** $iter \leq iter_{\max}$ **do**
- 6 **if** $iter > N_{\text{chance}}$ **then**
- 7 | Berechne $\mathbf{R}$ in Abhängigkeit von der Population P_{best} ;
- 8 Wähle $N - N_r$ Individuen zufällig in Abhängigkeit von $\mathbf{R}$ aus und ordne sie P_N zu;
- 9 Wähle N_r Individuen zufällig in Abhängigkeit von $\mathbf{R}_{\text{eq}}$ aus und ordne sie P_N zu;
- 10 Ermittle einen Fitnesswert Ω_i für jedes Individuum in P_N ;
- 11 **if** $iter = 1$ **then**
- 12 | Wähle die N_b Individuen mit dem kleinsten Ω_i in P_N aus und ordne sie P_{best} zu;
- 13 **else**
- 14 | Wähle N_b Individuen mit dem kleinsten Ω_i von P_N und alle Individuen in P_{best} und weise sie P_{all} zu;
- 15 | Wähle die N_b Individuen mit dem kleinsten Ω_i in P_{all} und weise sie P_{best} zu;
- 16 Lösche P_N und P_{all} ;
- 17 $iter = iter + 1$;
- 18 Wähle das Individuum Ind_{best} mit dem kleinsten Ω_i in P_{best} ;

das Problem der Knotenvektorbestimmung, welches eine leistungsfähige globale Optimierungsmethode voraussetzt, mit Hilfe von evolutionären Rechentechniken. Der EGA erlaubt im Gegensatz zum CSA das Auftreten von Crossover, welches die Möglichkeit der Kombination von lokal gut approximierenden Knotenvektorteilen zulässt und somit die Ergebnisdiversität erhöht. Der Berechnungsablauf ist in Algorithmus 2 zusammengefasst. Eine detaillierte Erläuterung der Berechnungsschritte ist in Bureick et al. (2019a) **P#1** zu finden.

Schritt 4 - Robuste-Kontrollpunktschätzung: Die Schätzung von Kontrollpunkten bei B-Splines führt zu Ergebnissen, die durch das Auftreten von Ausreißern stark verzerrt werden. Das Problem der Parameterschätzung kann in diesem Fall gelöst werden, indem anstatt der KQ-Schätzung robuste Schätzer, wie M-Schätzer, eingesetzt werden. Verfahren der robusten ML-Schätzung ersetzen die Optimierungsfunktion der quadrierten Summe der gewichteten Verbesserungen.

2.3.2 Anwendungsfälle zum MODELL I

Um die Leistungsfähigkeit der entwickelten MC- und EGA-Algorithmen zur Knotenvektorbestimmung zu evaluieren, werden diese Algorithmen auf zwei Testfunktionen angewendet und mit dem gesamten CSA-Algorithmus verglichen. In beiden Testfunktionen wird der Grad der Basisfunktion auf $p = 3$ gesetzt. Außerdem wird angenommen, dass sowohl die Punktwolke als auch die Kontrollpunkte eindimensional sind. Abbildung 2.1 (links) zeigt für die erste Testkurve einen einzelnen Simulationslauf, bei dem der CSA zu einem lokalen Optimum konvergiert ist. Es ist deutlich zu

Algorithmus 2: Optimierung von Knotenvektoren des B-Splines mittels EGA**Input** : Anzahl der Iterationen $iter_{max}$;Anzahl der Individuen N ;Anzahl der zu selektierenden Individuen N_b ;Anzahl der elitären Individuen N_e ;Anzahl der zufälligen Individuen N_r ;Rate für zufälliges Einzelpunkt-Crossover r_{rc} ;**Output:** Fittestes Individuum $\mathcal{I}_{best}$;

- 1 Generiere zufällig $N - 1$ Individuen und weise sie $\mathcal{P}_N$ zu;
- 2 Erzeuge 1 Individuum mit deterministischer Methode und weise es $\mathcal{P}_N$ zu;
- 3 Berechne Fitnesswert Ω_i für jedes Individuum in $\mathcal{P}_N$;
- 4 Setze die Laufvariable $iter = 1$;
- 5 **while** $iter \leq iter_{max}$ **do**
- 6 Wähle die N_b Individuen aus $\mathcal{P}_N$ mit der kleinsten Ω_i ;
- 7 Klone die N_b Individuen nach ihrer Ω_i und weise sie $\mathcal{P}_c$ zu;
- 8 Weise zufällig $r_{rc} \cdot N_c$ aus $\mathcal{P}_c$ $\mathcal{P}_{rc}$ zu;
- 9 Führe zufälliges Einzelpunkt-Crossover in $\mathcal{P}_{rc}$ durch;
- 10 Ordne $\mathcal{P}_{rc}$ und die verbleibenden Individuen von $\mathcal{P}_c$ $\mathcal{P}_M$ zu;
- 11 Führe Mutation in $\mathcal{P}_M$ durch;
- 12 Berechne Ω_i für jedes Individuum in $\mathcal{P}_M$;
- 13 Wähle die N_e Individuen aus $\mathcal{P}_N$ mit den kleinsten Ω_i und teile diese $\mathcal{P}_e$ zu;
- 14 Lösche $\mathcal{P}_N$;
- 15 Selektiere die $N - (N_e + N_r)$ Individuen aus $\mathcal{P}_M$ mit der kleinsten Ω_i und weise sie $\mathcal{P}_N$ zu;
- 16 Generiere zufällig N_r Individuen, berechne deren Ω_i und weise sie $\mathcal{P}_N$ zu;
- 17 Füge $\mathcal{P}_e$ zu $\mathcal{P}_N$;
- 18 $iter = iter + 1$
- 19 Selektiere das Individuum $\mathcal{I}_{best}$ mit den kleinsten Ω_i aus $\mathcal{P}_N$;

erkennen, dass sich die Kurvenapproximation durch den CSA signifikant von den Daten unterscheidet, insbesondere bei der Stützstelle $\omega > 0,6$. Abbildung 2.1 (rechts) zeigt einen Simulationslauf der zweiten Testfunktion, bei dem sämtlichen Verfahren in das globale Optimum konvergiert sind. Auch die Approximation der Spitze ($\omega = 0,5$) unterscheidet sich nicht wesentlich. Beide Verfahren

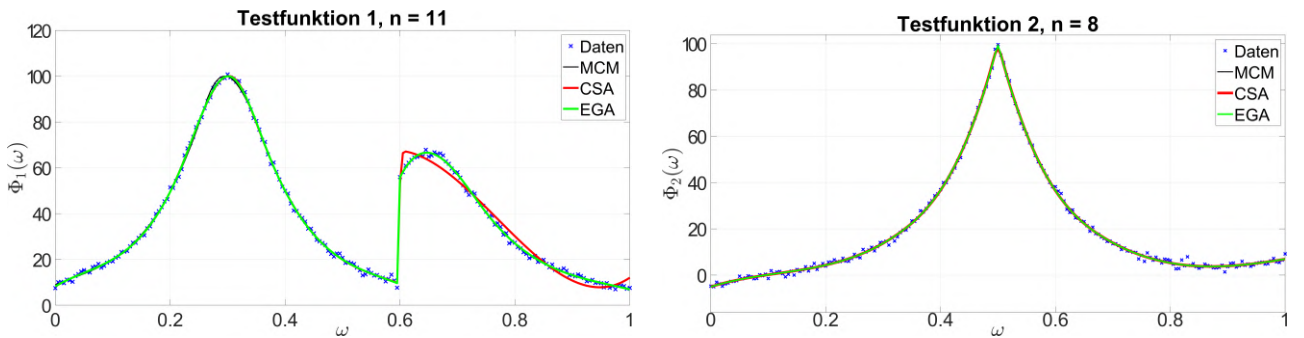


Abbildung 2.1: Links: Approximationsergebnis Testfunktion 1 mit $p = 3$ und $n = 11$, rechts: Approximationsergebnis Testfunktion 2 mit $p = 3$ und $n = 8$

werden auf einen realen Datensatz angewendet, der vom Advanced Rail Track Inspection System (ARTIS) erfasst wurde (für weitere Informationen siehe Dennig et al. (2017)). ARTIS ist unter anderem mit zwei Profil-Laserscannern (PLS) ausgestattet, die die Schienengeometrie abtasten. Aufgrund der Messkonfiguration enthält die Punktwolke zwei große Datenlücken im verdeckten Be-

reich unterhalb des Schienenkopfes. Die aus 1144 zweidimensionalen Punkten bestehende Punktwolke und die Approximationsergebnisse sämtlicher Verfahren sind in Abbildung 2.2 dargestellt. Die EGA lieferte in allen Fällen die besten Ergebnisse. Die Ergebnisse der CSA sind numerisch instabil, wenn Datenlücken auftreten. Weitere Betrachtungen sind in Bureick et al. (2019a) zu finden.

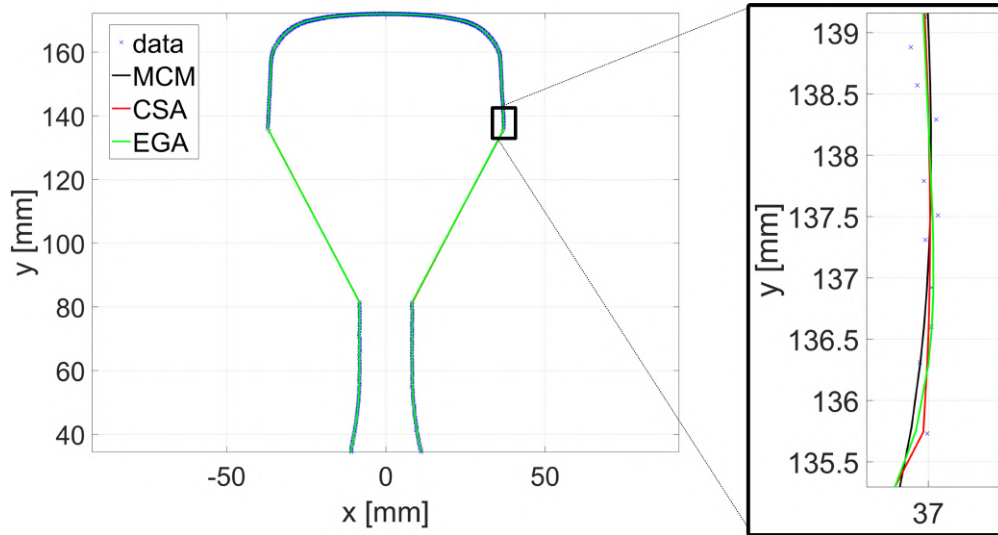


Abbildung 2.2: Die Approximationsergebnisse der realen Daten mit $p = 3$ und $n = 60$.



Innovation kurz zusammengefasst | Es wurden in diesem Abschnitt Methoden zur B-Spline-Approximation entwickelt, die in der Lage sind, kurvenhafte Punktwolken von komplex geformten Objekten mit Ausreißern und Datenlücken optimal zu approximieren. Speziell wurden heuristische Methoden zur Berechnung des Knotenvektors weiterentwickelt und optimiert. Deren Umsetzung auf simulierte und reale Daten demonstrierte eine deutliche Verbesserung der Approximationsgüte gegenüber etablierten Verfahren zur Knotenvektorberechnung.

2.4 MODELL II: Regressionsmodelle mit autoregressiven Prozessen und t-verteilter Messrauschen

2.4.1 Mathematisches Modell und GEM-Algorithmen zur Parameterschätzung

Gegeben seien die N -dimensionalen Beobachtungen $\mathbf{Y}_t = [Y_{1,t} \cdots Y_{N,t}]^\top$, welche an äquidistanten Zeitpunkten $t = 1, \dots, n$ gemessen werden. Die Aufgabe besteht darin, die entsprechenden Messergebnisse $\mathbf{y}_1, \dots, \mathbf{y}_n$ durch eine nicht-lineare Funktion $\mathbf{h}_t(\boldsymbol{\xi}) = [h_{1,t}(\boldsymbol{\xi}) \cdots h_{N,t}(\boldsymbol{\xi})]^\top$ unbekannter Parameter $\boldsymbol{\xi} = [\xi_1, \dots, \xi_m]^\top$ zu modellieren. Die Unsicherheiten des Messprozesses werden mittels zufälliger Abweichungen $\mathbf{E}_t = [E_{1,t} \cdots E_{N,t}]^\top$ zwischen den Beobachtungen und dem funktionalen Modell modelliert, so dass für die Beobachtungsgleichungen die Form

$$\mathbf{Y}_t = \mathbf{h}_t(\boldsymbol{\xi}) + \mathbf{E}_t \quad (t = 1, \dots, n) \quad (2.10)$$

gilt. Hier wird angenommen, dass jede der N Komponenten der zufälligen Abweichungen Autokorrelationen in Form eines kovarianzstationären AR-Modells unterliegen, d. h.

$$E_{k,t} = \alpha_{k,1} E_{k,t-1} + \dots + \alpha_{k,p_k} E_{k,t-p_k} + U_{k,t} \quad (k = 1, \dots, N; t = 1, \dots, n). \quad (2.11)$$

Mit der Einführung der Lag-Operator-Notation $L^j e_t = e_{t-j}$ und dem Lag-Polynom $\alpha_k(L) = 1 - \alpha_{k,1}L - \dots - \alpha_{k,p_k}L^{p_k}$ kann die rechte Seite von Gl. 2.11 durch $\alpha_k(L)e_{k,t}$ abgekürzt werden. Hier wird $\alpha_k(L)$ als Dekorrelationsfilter betrachtet. Die Zufallsvariablen $U_{k,1}, \dots, U_{k,n}$ für jedes $k = 1, \dots, N$ werden als unabhängig und identisch t-verteilt gemäß

$$U_{k,t} \sim t_{\nu_k}(0, \sigma_k^2) \quad (k = 1, \dots, N; t = 1, \dots, n). \quad (2.12)$$

angenommen. Somit wird es ermöglicht, dass jede Zeitreihe des weißen Rauschens $U_{k,1}, \dots, U_{k,n}$ ihr individuelles Rauschverhalten (gesteuert durch den komponentenabhängigen Skalenparameter σ_k^2) und ihren eigenen Freiheitsgrad ν_k der t-Verteilung besitzt. Die Kenngrößen werden neben den AR-Koeffizienten als zusätzliche unbekannte Parameter gemeinsam mit den funktionalen Parametern ξ geschätzt. Die PDF der skalierten t-verteilten weißen Rauschkomponenten $U_{k,t}$ wird somit definiert durch

$$f(u_{k,t}) = \frac{\Gamma\left(\frac{\nu_k+1}{2}\right)}{\sqrt{\nu_k \pi \sigma_k^2} \Gamma\left(\frac{\nu_k}{2}\right)} \left[1 + \left(\frac{u_{k,t}}{\sigma_k} \right)^2 / \nu_k \right]^{-\frac{\nu_k+1}{2}}. \quad (2.13)$$

Eine ML-Schätzung der unbekannten Modellparameter $\xi, \alpha_1, \dots, \alpha_N, \sigma_1^2, \dots, \sigma_N^2$ und $\nu_1, \dots, \nu_N$ basierend auf der gemeinsamen PDF $f(\mathbf{u})$ oder seinem natürlichen Logarithmus als Log-Likelihood-Funktion

$$\begin{aligned} \log \mathcal{L}(\theta; \mathbf{y}) &= \log f(\mathbf{u}) = \log [f(\mathbf{u}_1) \cdots f(\mathbf{u}_N)] = \sum_{k=1}^N \left(n \log \left[\frac{\Gamma\left(\frac{\nu_k+1}{2}\right)}{\sqrt{\nu_k \pi \sigma_k^2} \Gamma\left(\frac{\nu_k}{2}\right)} \right] \right. \\ &\quad \left. - \frac{\nu_k+1}{2} \sum_{t=1}^n \log \left[1 + \left(\frac{\alpha_k(L)(y_{k,t} - h_{k,t}(\xi))}{\sigma_k} \right)^2 / \nu_k \right] \right) \end{aligned} \quad (2.14)$$

bei gegebenen Messergebnissen $\mathbf{y}$ erfordern eine numerische Optimierung, da eine geschlossene Form des Schätzers nicht verfügbar ist.

EM-Optimierungsproblem: In Kargoll et al. (2018b) und Alkhatib et al. (2017b) (**P#3** und **P#4**) wurde ein generalisierter GEM-Algorithmus entwickelt, der sich aus einem E- und einem M-Schritt zusammensetzt. Die im E-Schritt benötigte Q -Funktion ist definiert als die bedingte Erwartung der obigen Log-Likelihood-Funktion bei gegebenen Beobachtungsvektor $\mathbf{y}$ und Parametervektor $\theta^{(i)}$ aus dem i -ten Iterationsschritt, d. h.

$$Q(\theta | \theta^{(i)}) = E_{\mathbf{W} | \mathbf{y}, \theta^{(i)}} \{ \log \mathcal{L}(\theta; \mathbf{y}, \mathbf{W}) \}. \quad (2.15)$$

Für die Durchführung des E-Schritts (erforderlich für die Bestimmung der Gewichte $\mathbf{W}$ innerhalb der IRLS) werden Initialparameterwerte benötigt. Zusätzlich wird innerhalb des ersten Iterationsschritts die einheitliche Gewichtung $w_{k,t}^{(0)} = 1$ für jede Komponente der multivariaten Zeitreihe gewählt. Um den M-Schritt durchzuführen, wird die Q -Funktion maximiert (siehe Gl. 2.15), indem die ersten partiellen Ableitungen der Q -Funktion in Bezug auf die unbekannten Parameter $\xi, \alpha_k, \sigma_k^2$ und ν_k (zusammengefasst in θ) bestimmt werden; anschließend werden diese Ableitungen gleich Null gesetzt. Der aktuelle Iterationsschritt wird mit $(i+1)$ bezeichnet und führt zur Parameterlösung $\theta^{(i+1)}$, die die Lösung $\theta^{(i)}$ des vorherigen Iterationsschritts ersetzt. Da die Bestimmung der Ableitungen in Bezug auf die Parameter ξ die Linearisierung der Funktionen $h_{k,t}$ beinhaltet, wird das Maximum durch einen Gauss-Newton-Schritt (der Länge γ) approximiert. Somit steht ein GEM-Algorithmus zur Verfügung, dessen Berechnungsablauf als Algorithmus 3 zusammengefasst wird.

Die Aktualisierung der Lösung $\xi^{(i+1)}$ (siehe Zeile 10 in Algorithmus 3) wird ganz oder teilweise der Probe-Lösung hinzugefügt (im Sinne eines Gauß-Newton-Schritts mit der Schrittweite $\gamma \in (0, 1]$).

Algorithmus 3: GEM-Algorithmus**Input** : $y_{k,t}, h_{k,t}(\xi), p_k, \text{itermax}$ ($k = 1, \dots, N; t = 1, \dots, n$)**Output:** $\hat{\xi}, \hat{\sigma}_k^2, \hat{\nu}_k, \hat{e}_k, \hat{u}_k, \hat{\mathbf{W}}_k, \hat{\alpha}_k$ 1 Initialisierungsschritt: $\xi^{(0)}; \nu_k^{(0)}; \sigma_k^{(0)}; \alpha_k^{(0)}$ 2 **for** $i = 0 \dots \text{itermax}$ **do**3 Führe E-Schritt durch:

$$4 \quad w_{k,t}^{(i)} = \frac{\nu_k^{(i)} + 1}{\nu_k^{(i)} + \left(\frac{\alpha_k^{(i)}(L)(y_{k,t} - h_{k,t}(\xi^{(i)}))}{\sigma_k^{(i)}} \right)^2} \rightsquigarrow \mathbf{W}_k^{(i)} = \begin{bmatrix} w_{k,1}^{(i)} & 0 & \cdots & 0 \\ 0 & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & w_{k,n}^{(i)} \end{bmatrix}$$

5 Führe M-Schritt durch:

$$6 \quad \Delta y_{k,t}^{(i)} = y_{k,t} - h_{k,t}(\xi^{(i)}), \mathbf{A}_{k,t}^{(i)} = \frac{\partial h_{k,t}(\xi^{(i)})}{\partial \xi}$$

$$7 \quad \overline{\Delta y}_{k,t}^{(i)} = \alpha_k^{(i)}(L) \Delta y_{k,t}^{(i)}, \quad \overline{\mathbf{A}}_{k,t}^{(i)} = \alpha_k^{(i)}(L) \mathbf{A}_{k,t}^{(i)}$$

$$8 \quad \Delta \xi^{(i+1)} = \left(\sum_{k=1}^N \frac{1}{(\sigma_k^{(i)})^2} \overline{\mathbf{A}}_k^{(i),T} \mathbf{W}_k^{(i)} \overline{\mathbf{A}}_k^{(i)} \right)^{-1} \sum_{k=1}^N \frac{1}{(\sigma_k^{(i)})^2} \overline{\mathbf{A}}_k^{(i),T} \mathbf{W}_k^{(i)} \overline{\Delta \mathbf{y}}_k^{(i)}$$

$$9 \quad \gamma = 1$$

$$10 \quad \xi^{(i+1)} = \xi^{(i)} + \gamma \Delta \xi^{(i+1)} \text{ (halbiere } \gamma \text{ falls nötig)}$$

$$11 \quad e_{k,t}^{(i+1)} = y_{k,t} - h_{k,t}(\xi^{(i+1)}) \rightsquigarrow \mathbf{E}_k^{(i+1)} = \begin{bmatrix} e_{k,0}^{(i+1)} & \cdots & e_{k,1-p_k}^{(i+1)} \\ \vdots & & \vdots \\ e_{k,n-1}^{(i+1)} & \cdots & e_{k,n-p_k}^{(i+1)} \end{bmatrix}$$

$$12 \quad \alpha_k^{(i+1)} = \left(\mathbf{E}_k^{(i+1),T} \mathbf{W}_k^{(i)} \mathbf{E}_k^{(i+1)} \right)^{-1} \mathbf{E}_k^{(i+1),T} \mathbf{W}_k^{(i)} \mathbf{e}_k^{(i+1)}$$

13 (stabilisiere $\alpha_k^{(i+1)}$ falls nötig)

$$14 \quad u_{k,t}^{(i+1)} = e_{k,t}^{(i+1)} - \alpha_{k,1}^{(i+1)} e_{k,t-1}^{(i+1)} - \dots - \alpha_{k,p_k}^{(i+1)} e_{k,t-p_k}^{(i+1)} = \alpha_k^{(i+1)}(L) e_{k,t}^{(i+1)}$$

$$15 \quad (\sigma_k^{(i+1)})^2 = \frac{1}{n} \sum_{t=1}^n w_{k,t}^{(i)} \left(u_{k,t}^{(i+1)} \right)^2 = \frac{\mathbf{u}_k^{(i+1),T} \mathbf{W}_k^{(i)} \mathbf{u}_k^{(i+1)}}{n}$$

$$16 \quad 0 = 1 + \log \nu_k^{(i+1)} - \psi \left(\frac{\nu_k^{(i+1)} + 1}{2} \right) + \psi \left(\frac{\nu_k^{(i+1)} + 1}{2} \right) - \log \left(\nu_k^{(i+1)} + 1 \right)$$

$$17 \quad + \frac{1}{n} \sum_{t=1}^n \left(\log \frac{\nu_k^{(i+1)} + 1}{\nu_k^{(i+1)} + \left(u_{k,t}^{(i+1)} / \sigma_k^{(i+1)} \right)^2} - \frac{\nu_k^{(i+1)} + 1}{\nu_k^{(i+1)} + \left(u_{k,t}^{(i+1)} / \sigma_k^{(i+1)} \right)^2} \right)$$

18 **if** $\max_{j,k} (|\xi_j^{(i)} - \xi_j^{(i+1)}|, |(\sigma_k^{(i)})^2 - (\sigma_k^{(i+1)})^2|) < \varepsilon$ **und** $\max_k (|\nu_k^{(i)} - \nu_k^{(i+1)}|) < \varepsilon_\nu$ **then**

19 | break

Wird die Q-Funktion um die aktuelle Schrittlänge verringert (die in jedem Iterationsschritt mit $\gamma = 1$ initialisiert wird), wird die Schrittlänge halbiert und die Q-Funktion bei den aktuellen Schätzungen erneut ausgewertet. Die Schrittlänge wird reduziert, bis die Q-Funktion zunimmt und sich damit dem Maximum nähert, wie es durch den GEM-Algorithmus gefordert wird. Die Schätzungen der Freiheitsgrade $\nu_1^{(i+1)}, \dots, \nu_N^{(i+1)}$ der t-Verteilung (siehe Zeilen 16 und 17 in Algorithmus 3) ergeben sich als die Nullstellen dieser Gleichungen, die numerisch zu finden sind. Für eine zuverlässige

Nullstellensuche wird die eindimensionale Intervall-Newton-Methode verwendet (siehe Hargreaves, 2002).

2.4.2 Anwendungsfälle zum MODELL II

Anwendung des Algorithmus 3 in simulierten Fourier-Reihen

In Kargoll et al. (2018b) **P#3** wurden zwei verschiedene lineare Regressionsmodelle in der MC-Simulation betrachtet. Aus Platzgründen wird an dieser Stelle auf das Fourier-Regressionsmodell eingegangen. Das Modell

$$\mathcal{L}_t = \frac{a_0}{2} + \sum_{j=1}^{12} a_j \cos(2\pi f_j x_t) + b_j \sin(2\pi f_j x_t) + \mathcal{E}_t \quad (2.16)$$

($t = 1, \dots, n$) setzt sich aus Sinus- und Kosinusbasisfunktionen mit unbekannten (“Fourier”) Koeffizienten $a_0, a_1, \dots, a_{12}$ und $b_1, \dots, b_{12}$ zusammen, die im Parametervektor ξ zusammengefasst sind. Die Fourier-Frequenzen werden als fehlerfreie Werte behandelt. Die Anzahl der Beobachtungen beträgt jeweils $n = 100, n = 1.000, n = 10.000$ oder $n = 100.000$. Die zufälligen Abweichungen in (2.16) erfüllen die AR-Modellgleichung

$$\mathcal{E}_t = \sum_{j=1}^p \alpha_j \mathcal{E}_{t-j} + \mathcal{U}_t, \quad (t = 1, \dots, n), \quad (2.17)$$

mit $p \in \{0, 1, 10, 100\}$; somit stellen die zufälligen Abweichungen entweder weißes Rauschen (im Falle von $p = 0$) oder farbiges Rauschen dar. Die Komponenten des weißen Rauschens $\mathcal{U}_1, \dots, \mathcal{U}_n$ werden unabhängig voneinander generiert, entweder durch: (i) die zentrale t-Verteilung mit $\mathcal{U}_t \stackrel{\text{ind}}{\sim} t_2(0, 0.001^2)$, (ii) die zentrale Normalverteilung mit $\mathcal{U}_t \stackrel{\text{ind}}{\sim} N(0, 0.001^2)$ oder (iii) die zentrale Mischnormalverteilung $\mathcal{U}_t \stackrel{\text{ind}}{\sim} 0.6 \cdot N(0, 0.001^2) + 0.4 \cdot N(0, 0.008^2)$. Die Rauschstichproben werden im Anschluss der wahren Beobachtungen hinzugefügt. Der Algorithmus 3 wurde angewendet, um jede generierte MC-Stichprobe für die Beobachtungen auszugleichen. So werden aus dem letzten GEM-Iterationsschritt die geschätzten Parameter $\hat{\xi}, \hat{\alpha}, \hat{\sigma}$ und $\hat{\nu}$ für jedes AR-Modell, jede Rauschverteilung und jeden Monte-Carlo-Lauf erhalten. Abbildung 2.3 zeigt die Histogramme für die ersten drei geschätzten Parameter des Fourier-Regressionsmodells. Es zeigt sich, dass die Schätzungen der funktionalen Parameter unbeeinflusst über die drei Sampling-Verteilungen für das jeweils betrachtete Regressionsmodell sind. Darüber hinaus ist die Streuung der Histogramme im Allgemeinen bei der Misch-Normalverteilung am höchsten und bei der t-Verteilung am geringsten. Aus der Simulation ist es auch festzustellen, dass die Approximation der wahren Parameterwerte mit zunehmendem Stichprobenumfang n deutlich verbessert wird, die Verzerrung scheint bei $n = 100.000$ geringfügig zu sein.

Anwendung des Algorithmus 3 in simulierten GNSS-Zeitreihen

In einer spezifischen Anwendung (siehe Alkhatib et al., 2018a, **P#8**) wird ein Multisensorsystem (MSS) verwendet, welches aus einem Laserscanner und zwei fest angebrachten GNSS-Antennen besteht (siehe Paffenholz, 2012). Zweck des MSS ist es, 3D-Punktwolken mittels 3D-Koordinaten in einem übergeordneten Koordinatenrahmen effizient georeferenzieren zu können. Zum besseren Verständnis wurden lediglich die erhaltenen 3D-Koordinaten von einer GNSS-Antenne gemäß Gl. (2.18) für n beobachtete 3D-Punkte betrachtet. Diese Punkte beschreiben einen Kreis in 3D durch die Drehung des Laserscanners um seine vertikale z -Achse. Für die Georeferenzierung des Laserscanners müssen die unbekannten Kreisparameter geschätzt werden. Die Parametrisierung eines dreidimensionalen Kreises ist gegeben durch $\Theta = [C_x, C_y, C_z, r, \Omega, \Phi]^T$. Der Parameter C ist der Mittelpunkt des Kreises für die x -, y - und z -Koordinate, r ist der Radius, die Winkel ω und φ beschreiben die Orientierung des Kreises in 3D durch die Drehungen um die x - und y -Achse. Mit den Parametern

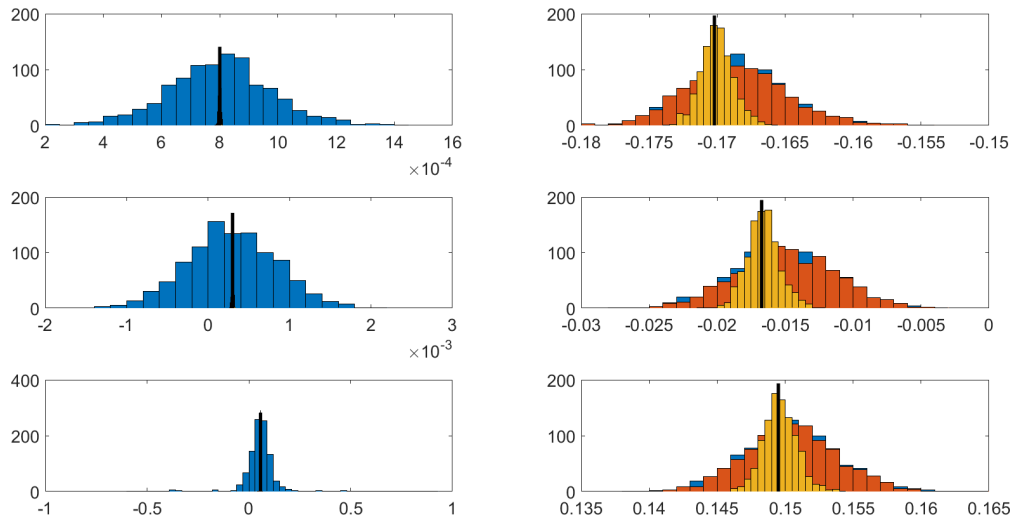


Abbildung 2.3: Histogramm aus der Monte-Carlo-Simulation für die Koeffizienten a_0 (oben links), a_1 (Mitte links), a_2 (unten links) und für die AR Koeffizienten α_1 (oben rechts), α_2 (mitte rechts), α_3 (unten rechts) für $n = 100000$ Beobachtungen der **Fourier-Modell** mit $AR(100)$ Abweichungen und weißes Rauschen nach der $t_2(0, 0.001^2)$ -Verteilung (gelb), der $N(0, 0.001^2)$ -Verteilung (rot) und das $0.6 \cdot N(0, 0.001^2) + 0.4 \cdot N(0, 0.008^2)$ -Verteilung (blau). Dicke schwarze Linien zeigen die Position des wahren Wertes an.

Θ kann das funktionale Modell für die verschiedenen Koordinatenkomponenten wie folgt aufgebaut werden:

$$\begin{pmatrix} X_t \\ Y_t \\ Z_t \end{pmatrix} = \begin{pmatrix} -r \cos(T_t) \sin(\Phi) + r \sin(T_t) \cos(\Omega) \cos(\Phi) + C_x \\ r \cos(T_t) \cos(\Phi) + r \sin(T_t) \cos(\Omega) \sin(\Phi) + C_y \\ -r \sin(T_t) \sin(\Omega) + C_z \end{pmatrix} + \begin{pmatrix} E_{1,t} \\ E_{2,t} \\ E_{3,t} \end{pmatrix} \quad (2.18)$$

Bezüglich der zufälligen Abweichungen $\mathbf{E}_t$ wurden drei verschiedene AR-Modelle bestimmt. Um ein korrektes und rechenzeitgünstiges Modell für den gegebenen realen Datensatz zu identifizieren, wurde Algorithmus 3 für verschiedene AR-Ordnungen bezüglich der drei Komponenten verwendet. Um zu prüfen wie stark sich jede Rauschreihe $\hat{u}_{k,1}, \dots, \hat{u}_{k,n}$ von dekorrelierten Residuen von theoretischem weißem Rauschen unterscheidet, wurde der in Kargoll et al. (2018b) **P#7** beschriebene Periodogramm-Test verwendet. Auf diese Weise wurde für die Nord- und Ostkomponente jeweils ein AR-Modell mit Ordnung 15 (kurz $AR(15)$) und für die Up-Komponente ein $AR(18)$ -Modell bestimmt. In Abbildung 2.4 werden die geschätzten kumulierten Periodogramme der dekorrelierten Residuen für die Nord- und Ostkomponente dargestellt. Die hierbei festgestellte Überschreitung der Signifikanzgrenzen ist geringer als bei anderen AR-Modellordnungen und insgesamt nur geringfügig, so dass die gewählten AR-Modellordnungen als angemessen beurteilt werden können.



Innovation kurz zusammengefasst | Die Innovation der entwickelten Methodik in diesem Abschnitt lag darin, eine simultane Schätzung von parametrisierten funktionalen und stochastischen Modellen auf Grundlage von raumzeitlichen Messreihen innerhalb einer Gesamtausgleichung zu ermöglichen. Das stochastische Modell besteht hierbei aus einem kombinierten autoregressiven Korrelationsmodell und einem stochastischen Ausreißermodell basierend auf der t-Verteilung, deren Formparameter bei der Ausgleichung mit Hilfe neu entwickelter GEM-Algorithmen bestmöglich an die gegebenen Daten adaptiert werden.

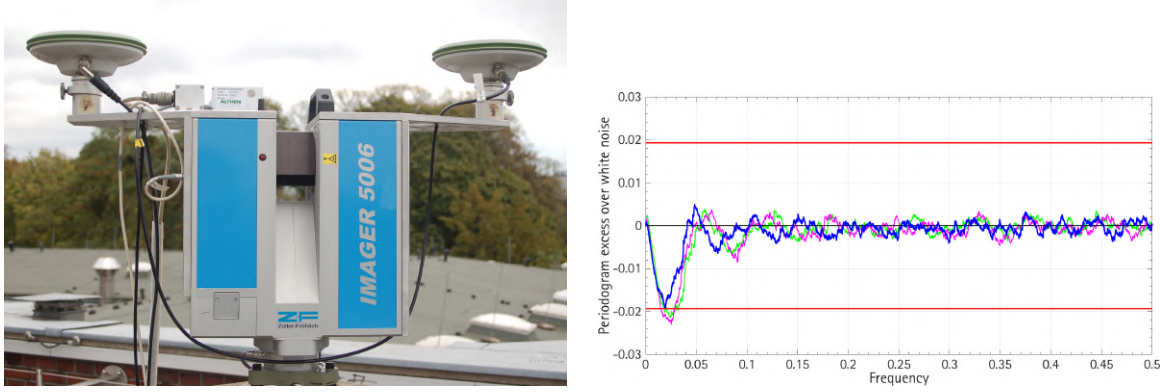


Abbildung 2.4: Links: Prototypische Realisierung des für das Experiment verwendeten MSS. Rechts: Das geschätzte Periodogramm der dekorrelierten Residuen für die Nord- und Ostkomponenten-AR(15)-Modell (blau und grün) und für das Hochkomponenten-AR(18)-Modell (magenta) in Bezug auf das theoretische Periodogramm mit theoretisch weißem Rausches (schwarz) und 99% Signifikanzgrenzen (rot).

2.5 MODELL III: Bayessche robuste Regressionsmodelle mit t-verteiletem Messrauschen

Bayessche Regressionsmodelle, die auf der Normalverteilung basieren, sind bekanntlich nicht robust gegenüber Ausreißern (Gelman, 2013). Wie schon in Abschnitt 2.3.2 erläutert, wurde als robuste Verteilungsfamilie die t-Verteilung mit niedrigen Freiheitsgraden vorgeschlagen. Dieser Abschnitt behandelt Ansätze, um mittels der Bayesschen Inferenz t-verteilte Abweichungen in einem nicht-linearen parametrischen Modell mit unabhängigen Beobachtungen zu lösen. Zur Approximation der resultierenden a-posteriori PDF werden MCMC-Sampler verwendet. Dieser entwickelte Algorithmus wird auf die GNSS-Zeitreihe (beschrieben im Abschnitt 2.4.2) und auf einen realen Datensatz im Bereich der Immobilienbewertung angewendet.

2.5.1 Mathematisches Modell

Der Zusammenhang zwischen Beobachtungsvektor und unbekanntem Parametervektor wird mit Hilfe des funktionalen Modells $\mathcal{L} = \mathbf{f}(\Theta) + \epsilon$ mit $\epsilon_i \sim t_\nu(0, s_i^2)$ beschrieben. Als Erweiterung dieses Ansatzes verwendet die Bayessche Inferenz PDFs, um die unbekannten Parameter eines Modells zu bestimmen, welche auf dem Bayesschen Theorem basieren:

$$p(\Theta|\mathbf{l}) \propto p(\Theta) \cdot p(\mathbf{l}|\Theta). \quad (2.19)$$

Hier ist $p(\Theta|\mathbf{l})$ die sogenannte a-posteriori PDF, aus der sich die unbekannten Modellparameter bedingt durch den Beobachtungsvektor $\mathbf{l}$ ableiten lassen. Die a-priori PDF $p(\Theta)$ hängt nicht von den Daten ab und beschreibt die zusätzlichen Informationen über die unbekannten Parameter. Dementsprechend enthält sie alle nicht datenbezogenen Informationen, die über Θ verfügbar sind. Die a-priori PDF $p(\Theta)$ kann z. B. aus den Ergebnissen einer früheren Ausgleichung oder einem Herstellerdatenblatt abgeleitet werden. Die a-priori PDF für die robuste Bayessche Regression stellt sich wie folgt dar:

$$p(\Theta, s_t, \nu) \propto p(\Theta)p(s_t)p(\nu). \quad (2.20)$$

Gl. 2.20 liegt die Annahme zugrunde, dass der Skalierungsfaktor der t-Verteilung, der Freiheitsgrad und die funktionalen Modellparameter als stochastisch unabhängig betrachtet werden. Die Vorkenntnisse über die Modellparameter sind der Erwartungswert θ der multivariaten Normalverteilung und die zugeordnete VKM Σ_Θ . Sie kontrolliert u. a. die Streuung der Verteilung, kurz

$\Theta \sim \mathcal{N}(\theta, \Sigma_\Theta)$. Die Likelihood-Funktion $p(\mathbf{l}|\Theta)$ beinhaltet die Informationen der Beobachtungen in Abhängigkeit von den unbekannten Parametern. $p(\mathbf{l}|\Theta)$ wird durch das funktionale Modell $\mathbf{f}(\Theta)$ und das stochastische Modell für die Residuen bestimmt. Im Falle der Bayesschen Inferenz basiert der Lösungsansatz auf Rand- und bedingten Dichten, wobei zwischen konjugierten und nicht-konjugierten a-priori PDFs unterschieden wird (siehe Koch, 2000).

Das beschriebene t-Verteilungsmodell kann auf einfache und äquivalente Weise neu formuliert werden (siehe Gelman, 2013) als

$$\begin{aligned} \epsilon_i &\sim \mathcal{N}(0, \alpha^2 W_i), \\ W_i &\sim \text{Inv-}\chi^2(\nu, \tau^2), \\ \text{mit } s_i^2 &= \alpha^2 \tau^2. \end{aligned} \tag{2.21}$$

Diese Äquivalenz wird durch die Einführung zusätzlicher Gewichte W_i gewährleistet, die sich aus einer skalierten, inversen Chi-Quadrat-Verteilung ergeben. Während ν in der $\text{Inv-}\chi^2$ -Verteilung dem Freiheitsgrad der t-Verteilung entspricht, wird der Skalierungsfaktor s_i in den Faktor α in Bezug auf die Normalverteilung und den Parameter τ der $\text{Inv-}\chi^2$ -Verteilung faktorisiert. Da das Modell in Gl. 2.21 zu einer nicht konjugierten a-priori PDF führt und das funktionale Regressionsmodell nicht-linear ist, werden MCMC-Methoden zur Berechnung der a-posteriori PDF herangezogen. Für Bayessche lineare Modelle mit t-verteilten Abweichungen existieren zahlreiche Lösungsansätze mit MCMC-Methoden, insbesondere mittels Gibbs-Samplern (siehe z. B. Geweke, 1993; Gelman, 2013).

2.5.2 MCMC Algorithmen zur Parameterschätzung

Die Approximation der a-posteriori PDF des robusten Bayesschen Regressionsmodells kann mittels Markov-Ketten erfolgen. Zur Generierung der Ketten der t-verteilten Abweichungen wird am häufigsten das Konzept des Gibbs-Samplers und des Metropolis-Hastings (MH)-Algorithmus (siehe Gelman, 2013) verwendet. In Dorndorf et al. (2019) **P#10** wurde ein MCMC-basierter Algorithmus entwickelt, der die a-posteriori PDF der funktionalen Parameter sowie Verteilungsparameter eines nicht-linearen Regressionsmodells mit t-Verteilung approximiert. Der Algorithmus wurde exemplarisch zur Schätzung von funktionalen Parametern aus 3D-Zeitreihen $\mathbf{l}_i = [x_i, y_i, z_i]$ ($i = 1, \dots, n$) entwickelt. Um die Flexibilität des Bayesschen Modells der Auswahl der PDFs zu demonstrieren, wurde davon ausgegangen, dass Ausreißer ausschließlich in der z -Koordinate auftreten. Den Residuen dieser Koordinatenkomponente wird eine t-Verteilung zugeordnet. Den Residuen der beiden übrigen Koordinatenkomponenten werden als normalverteilt angenommen. Die Residuen aller drei Koordinatenkomponenten können unterschiedliche Varianzniveaus aufweisen. Die stochastischen Abhängigkeiten zwischen den Koordinatenkomponenten wurden hierbei als unkorreliert angenommen. Das stochastische Modell lautet somit:

$$\epsilon_{x_i} \sim \mathcal{N}(0, \sigma_x^2), \quad \epsilon_{y_i} \sim \mathcal{N}(0, \sigma_y^2), \quad \epsilon_{z_i} \sim t_\nu(0, s_z^2). \tag{2.22}$$

Um die Ketten mittels Gibbs-Sampler zu approximieren, werden die bedingten a-posteriori PDFs der oben genannten Parameter benötigt. Diese werden aus der folgenden Likelihood-Funktion

$$\begin{aligned} p(\mathbf{l}|\Theta, \sigma_x, \sigma_y, \alpha, \mathbf{W}, \nu, \tau) &= \prod_{i=1}^n \frac{1}{\sqrt{(2\pi)^k \det \Sigma_{\mathbf{l}_i}}} \times \exp \left(-\frac{(\mathbf{l}_i - \mathbf{f}_i(\Theta))^T \Sigma_{\mathbf{l}_i}^{-1} (\mathbf{l}_i - \mathbf{f}_i(\Theta))}{2} \right) \\ \text{mit } \Sigma_{\mathbf{l}_i} &= \begin{bmatrix} \sigma_x^2 & 0 & 0 \\ 0 & \sigma_y^2 & 0 \\ 0 & 0 & \alpha^2 W_i \end{bmatrix} \text{ und } k = 3 \end{aligned} \tag{2.23}$$

und den a-priori PDFs der gesuchten Parameter berechnet. Für die Modellparameter Θ wurde

eine multivariate Normalverteilung als a-priori PDF $p(\Theta)$ verwendet, während für die restlichen Parameter eine nicht informative a-priori PDF angesetzt wurde. Zur Reduktion der Komplexität des Algorithmus wurde der Freiheitsgrad der t-Verteilung als bekannt angenommen, weswegen für ν keine bedingte PDF berechnet werden muss. Hieraus ergeben sich die folgenden bekannten a-posteriori PDFs (siehe Gelman, 2013):

- *Bedingte a-posteriori PDF für ein Gewicht W_i* : Die Gewichte für die z -Koordinate sind abhängig von τ , α , ν , Θ und l_{z_i} . Gemäß

$$W_i | \tau^2, \alpha^2, \Theta, l_{z_i} \sim \text{Inv-}\chi^2 \left(\nu + 1, \frac{\nu \tau^2 + (l_{z_i} - f_{z_i}(\Theta))^2 / \alpha^2}{\nu + 1} \right), \quad (2.24)$$

sind die Gewichte $\text{Inv-}\chi^2$ -verteilt, wobei der Parameter α , ν sowie τ die Robustheit steuern.

- *Bedingte a-posteriori PDF für τ* : Der Skalierungsparameter τ ist abhängig von den Gewichten $\mathbf{W}$ sowie den Modellparametern Θ , dem Skalierungsparameter α und den Beobachtungen l_{z_i} . Hieraus folgt

$$\tau^2 | \mathbf{W}, \alpha^2, \Theta, \mathbf{l}_z \sim \text{Gamma} \left(\frac{n\nu}{2}, \frac{\nu}{2} \sum_{i=1}^n \frac{1}{W_i} \right). \quad (2.25)$$

Für die Generierung der Gamma-verteilten Zufallsvariablen τ werden die Parameter Θ und α sowie die Beobachtungen $\mathbf{l}_z$ nicht direkt verwendet. Diese Werte sind indirekt durch die Gewichte $\mathbf{W}$ und die Anzahl der Beobachtungen n einbezogen.

- *Bedingte a-posteriori PDF für α* : Der Skalierungsfaktor α wird aus einer $\text{Inv-}\chi^2$ -Verteilung generiert:

$$\alpha^2 | \mathbf{W}, \tau^2, \Theta, \mathbf{l}_z \sim \text{Inv-}\chi^2 \left(n, \frac{1}{n} \sum_{i=1}^n \frac{(l_{z_i} - f_{z_i}(\Theta))^2}{W_i} \right). \quad (2.26)$$

- *Bedingte a-posteriori PDF für σ_x^2* : Die bedingte PDF der Varianz σ_x^2 unterscheidet sich zu der von α in (2.26) nur durch das Ersetzen der z_i -Komponenten durch die x_i -Komponenten unter Verzicht auf das Gewicht W_i :

$$\sigma_x^2 | \Theta, \mathbf{l}_x \sim \text{Inv-}\chi^2 \left(n, \frac{1}{n} \sum_{i=1}^n (l_{x_i} - f_{x_i}(\Theta))^2 \right). \quad (2.27)$$

Die bedingte PDF der Varianz σ_y^2 entspricht der bedingten PDF von σ_x^2 , abgesehen davon, dass die Komponenten y_i - anstelle von x_i verwendet werden.

- *Bedingte a-posteriori PDF für Θ* : Im Gegensatz zu den zuvor behandelten Parametern ist die bedingte PDF des Parametervektors Θ aus der Literatur nicht bekannt. Grund ist die Verwendung eines beliebigen nicht-linearen funktionalen Modells. Dies stellt allerdings kein Problem dar, da die bedingte PDF aus Zufallszahlen approximiert werden kann, die mit Hilfe eines MH-Algorithmus (Hastings, 1970) generiert werden. Zu diesem Zweck werden die a-priori PDFs von Θ und die Likelihood-Funktion benötigt. Durch die Kombination des Gibbs-Samplers und des MH-Algorithmus können auf diese Weise für jedes Θ_i die bedingten PDF gelöst werden. Der verwendete MH-Algorithmus ist in Schritt 6 von Algorithmus 4 beschrieben.

Aus den bedingten Dichten wurde der Algorithmus 4 zur Erstellung der Markov-Ketten in Dorndorf et al. (2019) **P#10** entwickelt. Dieser beruht auf dem Konzept des Gibbs-Samplers und des MH-Algorithmus und setzt sich aus sechs Schritten zusammen. Neben den Input-Parametern werden

weiterhin Startwerte für $\Theta^{(0)}$, $\alpha^{(0)}$ und $\tau^{(0)}$ benötigt. Sind diese bekannt, startet der Algorithmus mit Schritt 1 bis eine Markov-Kette mit der Länge m für die unbekannten a-posteriori Parameter erstellt ist. Aus den resultierenden Markov-Ketten können die a-posteriori Ergebnisse für alle unbekannten Parameter (Modellparameter, Varianzkomponenten, Gewichte und Skalenparameter der t-Verteilung) approximiert werden. Die ersten Iterationen der generierten Ketten werden als burn-in Phase verworfen. Die restlichen Stichproben dienen zur Schätzung der Momente der approximierten a-posteriori PDF, u. a. Mittelwerte und VKM (siehe z. B. Alkhatib und Schuh, 2007; Kroese et al., 2011).

Algorithmus 4: MCMC Algorithmus

Input : $l_x, l_y, l_z, f(\Theta)$, # Stichproben m

Output: $\sigma_x^{2(j)}, \sigma_y^{2(j)}, \sigma_z^{2(j)}, \mathbf{W}^{(j)}, \Theta^{(j)}, j = 1 \dots m$

```

1 Initialisierungsschritt:  $\Theta^{(0)}; \tau^{(0)}; \alpha^{(0)}$ 
2 for  $j = 1 \dots m$  do
3   Schritt 1: Aktualisiere  $\mathbf{W}^{(j)}$  mit  $\tau^{2(j-1)}, \alpha^{2(j-1)}, \Theta^{(j-1)}$  und  $l_z$  anhand Gl. 2.24
4   Schritt 2: Aktualisiere  $\tau^{2(j)}$  mit  $\mathbf{W}^{(j)}, \Theta^{(j-1)}$  und  $l_z$  anhand Gl. 2.25
5   Schritt 3: Aktualisiere  $\alpha^{2(j)}$  mit  $\tau^{2(j)}$  mit  $\mathbf{W}^{(j)}, \Theta^{(j-1)}$  und  $l_z$  anhand Gl. 2.26
6   Schritt 4: Aktualisiere  $\sigma_x^{2(j)}$  mit  $\Theta^{(j-1)}$  und  $l_x$  anhand Gl. 2.27
7   Schritt 5: Aktualisiere  $\sigma_y^{2(j)}$  mit  $\Theta^{(j-1)}$  und  $l_y$  anhand Gl. 2.27
8   Schritt 6: Generiere  $\Theta^{(j)}$  mit Hilfe des MH-Algorithmus:
9   for  $i = 1 \dots u$  do
10    1. Generiere:  $\Theta_i^{\text{new}} \sim N(\Theta_i^{(j-1)}, \lambda_{\Theta_i})$ 
        2. Stelle auf:  $\Theta^{\text{new}} = [\theta_1^{(j)}, \theta_2^{(j)}, \dots, \theta_i^{\text{new}}, \dots, \theta_u^{(j-1)}]^T$ 
                      $\Theta^{\text{old}} = [\theta_1^{(j)}, \theta_2^{(j)}, \dots, \theta_i^{(j-1)}, \dots, \theta_u^{(j-1)}]^T$ 
        3. Berechne:  $\Psi = \min \left[ 1, \frac{p(\Theta^{\text{new}}) \cdot p(\Theta^{\text{new}} | l, \sigma_x, \sigma_y, \alpha, \mathbf{W}, \nu, \tau)}{p(\Theta^{\text{old}}) \cdot p(\Theta^{\text{old}} | l, \sigma_x, \sigma_y, \alpha, \mathbf{W}, \nu, \tau)} \right]$ 
        4. Akzeptiere oder verwirfe:
            · Generiere:  $\rho \sim U(0, 1)$ 
            · falls  $\rho \leq \Psi$ :  $\theta_i^{(j)} = \theta_i^{\text{new}}$ 
            · falls  $\rho > \Psi$ :  $\theta_i^{(j)} = \theta_i^{(j-1)}$ 

```

2.5.3 Anwendungsfälle zum MODELL III

Anwendung des Algorithmus 4 in simulierten GNSS-Zeitreihen

In der MC-Simulation wird ein MSS verwendet, welches bereits in Abschnitt 2.4.2 vorgestellt wurde. Der Parameter c ist der Mittelpunkt des Kreises für die x -, y - und z -Koordinate und r ist der Radius. Die Winkel ω und φ beschreiben die Ausrichtung des Kreises in 3D durch die Drehungen um die x - und y -Achse. Mit dem Parameter Θ kann das funktionale Modell für die verschiedenen Koordinatenkomponenten wie folgt aufgebaut werden:

$$\begin{aligned}
 f_{x_i}(\Theta) &= r \cos(t_i) \cos(\varphi) + c_x, \\
 f_{y_i}(\Theta) &= r \cos(t_i) \sin(\varphi) \sin(\omega) + r \sin(t_i) \cos(\omega) + c_y, \\
 f_{z_i}(\Theta) &= -r \cos(t_i) \sin(\varphi) \cos(\omega) + r \sin(t_i) \sin(\omega) + c_z, \\
 \text{mit } \mathbf{f}_i(\Theta) &= [f_{x_i}(\Theta), f_{y_i}(\Theta), f_{z_i}(\Theta)].
 \end{aligned} \tag{2.28}$$

Die Aufteilung des funktionalen Modells in die drei Koordinatenkomponenten entspricht einem nicht-linearen Regressionsmodell, welches die Schätzung der unbekannten Parameter mittels eines

GMM ermöglicht. Die Variable t_i in Gl. 2.28 ist der Drehwinkel des Laserscanners um seine z-Achse. Um das Modell zu vereinfachen, wird davon ausgegangen, dass der Drehwinkel $\mathbf{t}$ bekannt und fehlerfrei ist. Neben den simulierten GNSS-Beobachtungen liegen Vorkenntnisse hinsichtlich des Radius aus Kalibriermessungen mit einem Lasertracker vor. Somit sind der Wert r und die entsprechende Varianz σ_r^2 bekannt. Für die übrigen Modellparameter $\mathbf{c}$, ω und φ wird eine nicht-informative a-priori PDF angenommen. Die Untersuchung des vorgestellten Bayesschen Schätzverfahrens in Algorithmus 4 basiert auf einer MC-Simulation und einem realen Datensatz. In Dorndorf et al. (2019) **P#10** wird die Generierung der MC-Simulation und der Ergebnisse ausführlich beschrieben.

Mit dem funktionalen Modell in Gl. (2.28) und Θ_{true} wurden 50 gleichverteilte Beobachtungen auf dem 3D-Kreis berechnet. Anschließend wird ein zufällig normalverteiltes Rauschen generiert und zu diesen Beobachtungen addiert. Zu den normalverteilten Abweichungen werden zusätzlich 15% der z -Beobachtungen als Ausreißer hinzugefügt, die aus der Gleichverteilung generiert wurden. In jeder Iteration der Simulation wurden das normalverteilte Rauschen und die Ausreißer zufällig erzeugt und die Simulation 10.000 mal wiederholt. Für die Schätzung von $\hat{\Theta}^{(j_{\text{CL}})}$ werden drei verschiedene Ansätze verwendet: (1) Ein linearisiertes, nicht robustes GMM (vgl. Koch, 1999), (2) der vorgestellte robuste Bayessche Ansatz mit nicht-informativer a-priori PDF und (3) der dargestellte robuste Bayesschen Ansatz mit a-priori Informationen für den Radius. In jeder Iteration der MC-Simulation wird eine Markov-Kette mit $m = 7.000$ und einer Warm-up Periode von $o = 3.500$ generiert. Die entsprechenden Verteilungen der 10.000 RMSE-Ergebnisse sind in Abbildung 2.5 dargestellt. Der durchschnittliche RMSE für das GMM (blau) ist größer als die Werte für die Bayesschen nicht-informativen (grüne) und informative (rote) Ansätze. Der einzige Unterschied zwischen den Bayesschen Ergebnissen ist bei der Schätzung des Radius zu erkennen. Im informativen Fall reduziert das Vorwissen über den Radius die Standardabweichung des geschätzten Radius, jedoch ist der Unterschied sehr gering.

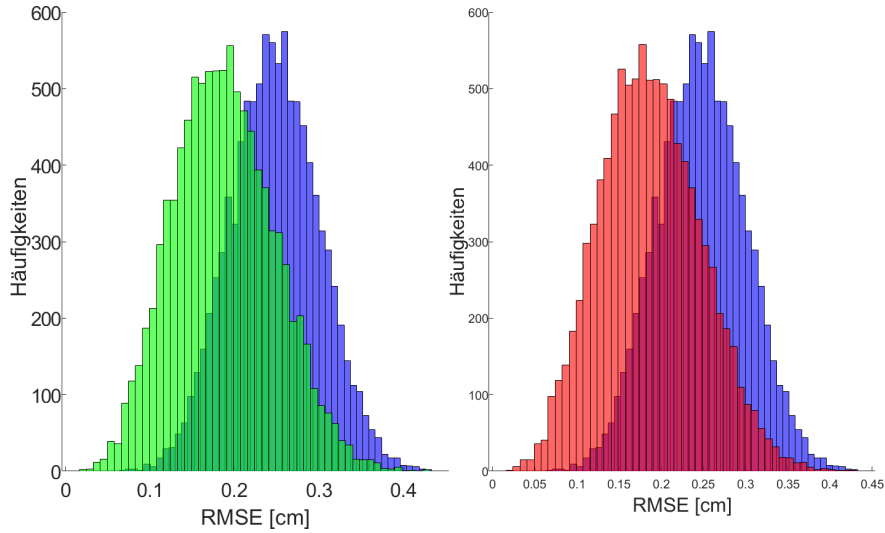


Abbildung 2.5: Verteilung der 10.000 RMSE-Ergebnisse für die drei Ansätze GMM (blau), Bayessche nicht-informative PDF (grün) und Bayessche informative PDF (rot) (Dorndorf et al., 2019)

Anwendung des Algorithmus 4 in der Immobilienbewertung

Der zweite Anwendungsfall zielt darauf ab, einen Ansatz zu entwickeln, welcher eine zuverlässige Bewertung auch in Lagen mit wenigen Kauffällen ermöglicht. Zu diesem Zweck wurde der robuste Bayessche Ansatz (präsentiert in Algorithmus 4) eingeführt, welcher in einer vereinfachten Form für die lineare Regression adaptiert wurde. Durch den linearen funktionalen Zusammenhang ist der MH-Schritt in Algorithmus 4 nicht notwendig, da die bedingte PDF für die funktionalen Parameter

analytisch gebildet werden kann. Als a-priori Information wurde Expertenwissen in das datengetriebene Modell, hier der multiplen linearen Regressionsanalyse, integriert, die auf einer kleinen Datenstichprobe (kaufpreisarme Lagen) beruhen. Auf der Grundlage eines Bayesschen Ansatzes werden Daten und Expertenwissen in einem umfassenden Modell aufbereitet. Dieses Modell verwendet die unabhängige t-Verteilung im linearen Modell. Die hier dargestellte Untersuchung wurde im räumlichen Teilmarkt Osnabrück durchgeführt (siehe Weitkamp und Alkhatib, 2014, **P#9**). Es handelte sich hierbei um einen sachlichen Teilmarkt, welcher Ein- und Zweifamilienhäuser, Doppelhäuser und Reihenhäuser beinhaltet. Das funktionale Modell der linearen multiplen Regression ergibt sich zwischen der Zielgröße Wohnflächenpreis (€/m²) und den Einflussgrößen Grundstücksfläche (m²), Bodenrichtwert (€/m²), Baujahr (Jahr), Wohnfläche (m²) und Ausstattungsstandard (einheitslos). Das a-priori Vorwissen stammt aus Experteninterviews bzw. aus der Analyse von Gutachten. Infolgedessen wurden 270 bzw. 75 Pseudofälle erzeugt. Zur Simulation von kaufpreisarmen Lagen werden die Daten systematisch reduziert. Die Stichprobe wird zunächst zufällig auf $n = 30$ Kauffälle reduziert und das Experiment 1.000 mal wiederholt (Monte-Carlo-Wiederholungen). Die Ergebnisse werden durch eine vorab selektierte Stichprobe validiert. Hierzu wurden 250 Käufe zur Überprüfung der Ergebnisse zurückgelegt und im Rahmen einer Kreuz-Validierung verwendet (siehe z. B. Bishop, 2009). Am besten passen sich die Funktionen in Bereichen an, in denen Daten vorliegen. Abbildung 2.6 (links) veranschaulicht das Verhalten für die Reduktion auf das Minimum und Abbildung 2.6 (rechts) für die Reduktion auf das Maximum jeweils der Originalstichprobe. Die beiden Bayesschen informativen Verfahren lösen auch außerhalb der Daten den funktionalen Zusammenhang vergleichsweise gut. Die klassische Bayessche Regression löst die Regressionsfunktion unwesentlich besser als das robuste Verfahren, versagt aber häufig numerisch aufgrund der geringen Anzahl an Daten. Ein signifikanter Unterschied zwischen den informativen Bayesschen Verfahren ist allerdings nicht zu verzeichnen. Von sämtlichen Verfahren zeigt sich die robuste Regression von Huber als die instabilste. Bis zu 20 % der MC-Durchläufe können aufgrund numerischer Probleme nicht gelöst werden.

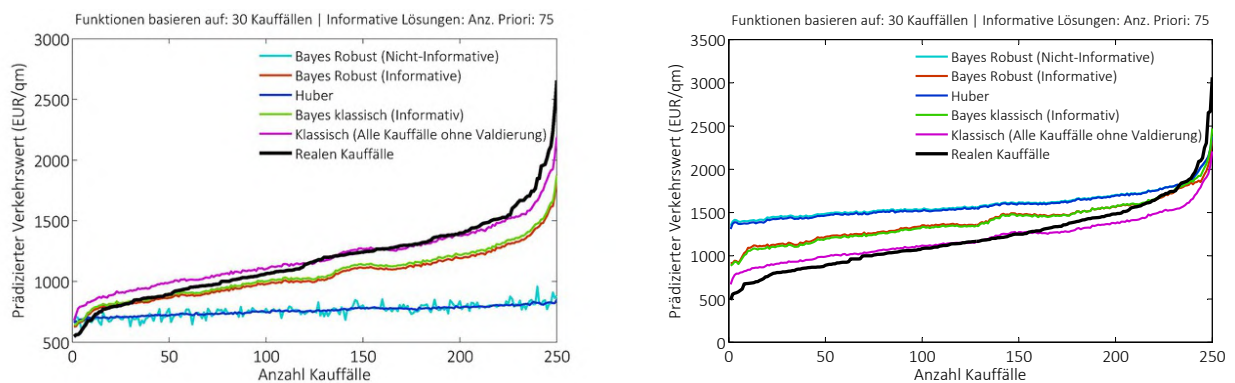


Abbildung 2.6: Links: RMSE für die Reduktion auf das Minimum der Originalstichprobe und die Nutzung von Gutachten als a-priori Information. Rechts: RMSE für die Reduktion auf das Maximum der Originalstichprobe und die Nutzung von Gutachten als a-priori Information; Originalabbildung aus Weitkamp und Alkhatib (2014)



Innovation kurz zusammengefasst | Basierend auf der Bayesschen Statistik wurden innovative Ausgleichsalgorithmen entwickelt, welche eine robuste und zuverlässige Schätzung der funktionalen Parameter sowie der Parameter der Messunsicherheitsverteilung (t-Verteilung) unter Einbeziehung von Vorinformation hinsichtlich der Modellparameter erlaubt. Eine gewinnbringende Besonderheit dieser Algorithmen liegt in der Verwendung von MCMC Methoden zur Gesamtausgleichung unter Vermeidung einer Linearisierung der nicht-linearen Modelle.

2.6 Zusammenfassung des Kapitels

Dieses Kapitel stellt die Algorithmen zur Durchführung von Parameterschätzungen aus redundanten Messungen im Rahmen von parametrischen Modellen im Kontext des GMM zusammen. Der Fokus liegt auf Optimierungsverfahren zur Parameterschätzung aus einem ML-Algorithmus bzw. der Bayesschen Inferenz. Drei unterschiedliche parametrische Regressionsmodelle werden ausführlich behandelt: B-Spline Regressionsmodelle zur Kurvenapproximation (*MODELL I*), Regressionsmodelle mit autoregressiven Prozessen und t-verteilter Messrauschen (*MODELL II*) sowie Bayessche robuste Regressionsmodelle mit t-verteilter Messrauschen (*MODELL III*).

In Abschnitt 2.3 wurden neue Ansätze zur Bestimmung des optimalen Knotenvektors in der Approximation mittels B-Spline-Kurven entwickelt. Die Knotenvektorberechnung basiert auf heuristischen Optimierungsverfahren, MC-Methoden und GA. Darüber hinaus wurde der robuste M-Schätzer zur Schätzung der Kontrollpunkte angewendet. Durch den Einsatz solcher Schätzer wurde das resultierende nicht-lineare System durch die Verwendung eines IRLS-Algorithmus gelöst. Die durchgeführten Anwendungen zeigten, dass beide Algorithmen, EGA und EMC, trotz auftretender Ausreißer und Datenlücken zu genauen und zuverlässigen robuste Approximationen von komplexen Kurven geführt haben. Dadurch werden die besten Approximationsergebnisse erzielt, insbesondere bei der Knotenvektorbestimmung. Zusätzlich konvergieren beide Algorithmen schnell in Richtung des globalen Optimums.

Der Abschnitt 2.4 beinhaltet die Entwicklung einer innovativen Theorie der robusten Ausgleichung nach Parametern im Kontext der Zeitreihenanalyse. Die Parameter können sich sowohl auf ein lineares, als auch auf ein nicht-lineares raumzeitliches Regressionsmodell beziehen. In der betrachteten Theorie werden Ausreißer stochastisch durch die Annahme t-verteilter zufälliger Abweichungen modelliert und die Parameterschätzung mittels EM-Algorithmen effizient umgesetzt. Die resultierenden Algorithmen sind selbstoptimierend in dem Sinne, dass aus einer vom Freiheitsgrad der t-Verteilung abhängigen Familie von unendlich vielen Schätzfunktionen die beste entsprechend der tatsächlichen Ausreißercharakteristik automatisch ausgewählt wird. Als weitere Datencharakteristik neben Ausreißern werden vorhandene, aber vorab unbekannte Korrelationsmuster in Form von farbigem Messrauschen durch AR-Prozesse modelliert und mitgeschätzt. Erfolgt eine gleichzeitige und unabhängige Messung mehrerer Größen als Zeitreihen, dann werden für jede Komponente eine eigene t-Verteilung und ein eigener AR-Prozess angepasst. Die verschiedenen Zeitreihen werden auf Ebene eines raumzeitlichen Regressionsmodells miteinander verknüpft. Der für diese Modellklasse entwickelte generalisierte GEM-Algorithmus wurde auf simulierte Fourier-Reihen und auf reale GNSS-Beobachtungen eines 3D-Kreises angewendet.

In Abschnitt 2.5 wurde ein robuster Bayesscher Ansatz zur Ausgleichung eines nicht-linearen funktionalen Modells mit t-verteilter Unsicherheitsmodell präsentiert. Ein empirischer Ansatz zur Berechnung der unbekannten a-posteriori PDF basierend auf MCMC wurde entworfen. Der entwickelte Algorithmus basiert auf dem Gibbs-Sampler mit einem zusätzlichen MH-Schritt. Dies ermöglicht es, Ketten für die funktionalen Parameter ohne Bildung der bedingten Verteilung zu generieren. Abschließend wurden zwei Anwendungen vorgestellt: Die Schätzung eines 3D-Kreises von GNSS Zeitreihen und die Schätzung der Regressionskoeffizienten in simulierten kaufpreisarmeren Lagen. Die Ergebnisse beider Anwendungen verdeutlichten, dass die eingeführte t-Verteilung im Bayesschen Modell den Einfluss von Ausreißern auf die geschätzten Parameter signifikant reduziert.

3 Filterungs- und Optimierungsverfahren für die Zustandsschätzung im nicht-linearen Zustandsraum

3.1 Einführung zur verschiedenen Filterungstechniken und zur sequentiellen Bayesschen Zustandsschätzung

Die nicht-lineare Filterung im Zustandsraum ist für nahezu alle Bereiche der Ingenieur- und Naturwissenschaften von Bedeutung. Jede Disziplin, die die mathematische Modellierung ihrer Systeme fordert, benötigt eine Zustandsschätzung; dazu sind (neben zahlreichen weiteren Anwendungsgebieten) insbesondere die Fachdisziplinen Geodäsie, Elektrotechnik, Maschinenbau, Chemieingenieurwesen, Luft- und Raumfahrttechnik, Robotik und die Steuerung dynamischer Systeme zu zählen. Das Kernproblem besteht darin, den Zustand eines dynamischen Systems unter Verwendung einer Folge von Messungen sequentiell zu schätzen. In dieser Arbeit wird sich lediglich auf die diskrete zeitliche Formulierung des Zustandsproblems zur Modellierung dynamischer Systeme fokussiert. Bei der Zustandsschätzung ist zu berücksichtigen, dass Messungen unsicher sind, d. h., die Messergebnisse enthalten Messunsicherheiten – z. B. aufgrund unvermeidlicher Imperfektion von Messsensoren oder aufgrund unkontrollierbarer Effekte, wie beispielsweise atmosphärischen Einflüssen. Weiterhin ist das Wissen über die Systemdynamik in der Regel nur grob bekannt und deswegen nur annähernd physikalisch modellierbar. Neben den internen Faktoren (die durch den Systemzustand beschrieben werden) und den bekannten externen Faktoren beeinflussen in der Regel auch viele weitere Faktoren das System und stellen somit aus Sicht des Modells das Rauschen dar. System- und Messrauschen beeinflussen die Fähigkeit, das Systemverhalten exakt zu präzisieren und zu steuern. Es ist daher anzustreben, den Einfluss der System- und Messunsicherheiten zu minimieren – insbesondere die Minimierung dieser Unsicherheiten auf die sequentielle Zustandsschätzung. Um das Problem der Zustandsschätzung präzise zu formulieren, ist ein mathematisches Modell des tatsächlichen Systems erforderlich sowie ein Fehlermodell des Rauschens, welches die System- und Messunsicherheiten beschreibt. Zur Auswahl der bestmöglichen nicht-linearen Filterungstechnik ist es zudem notwendig, ein numerisches Maß zur Beschreibung der verbleibenden Unsicherheit der Zustandsschätzung zu quantifizieren. Die Qualität der resultierenden Zustandsschätzungen hängt davon ab, wie gut das ausgewählte System- und Beobachtungsmodell das tatsächliche System und das tatsächliche Unsicherheitsniveau beschreiben.

Die probabilistische Zustandsraumformulierung und die Notwendigkeit der Aktualisierung von Informationen nach Eingang neuer Messungen sind für den Bayesschen Ansatz prädestiniert: Sie bietet einen allgemeinen Rahmen für dynamische Zustandsschätzungsprobleme. Rekursive Bayessche Schätzverfahren nutzen Vorkenntnisse – typischerweise in Form eines mathematischen Modells des Systems – sowie verfügbare Messungen, um die bedingte a-posteriori PDF des Zustandsvektors zu erhalten (siehe z. B. Thrun et al., 2006; Ristic et al., 2004; Candy, 2016). Um das Problem der nicht-linearen Filterung zu definieren, wird der Zustandsparametervektor $\mathbf{x}_k \in \mathbb{R}^{n_x}$ eingeführt, wobei n_x die Dimension des Zustandsvektors symbolisiert; $\mathbb{R}$ stellt die Menge der reellen Zahlen dar, $k \in \mathbb{N}$ den Zeitindex und $\mathbb{N}$ die Menge der natürlichen Zahlen. Der Index k wird einer kontinuierlichen Zeitepoche t zugeordnet. Der Systemzustand entwickelt sich nach dem folgenden diskreten

stochastischen Systemmodell:

$$\mathbf{x}_k = \mathbf{f}_{k-1}(\mathbf{x}_{k-1}, \mathbf{u}_{k-1}) + \mathbf{w}_{k-1}, \quad (3.1)$$

wobei $\mathbf{f}_{k-1}$ eine bekannte, möglicherweise nicht-lineare Funktion der Zustandsparameter $\mathbf{x}_{k-1}$ ist, $\mathbf{w}_{k-1}$ der Prozessrauschvektor und $\mathbf{u}_{k-1}$ ein bekannter deterministischer Stellvektor ist. Das Prozessrauschen ist auf fehlerhafte Modellierungseffekte oder unvorhergesehene Störungen im Bewegungsmodell zurückzuführen. Das Ziel der nicht-linearen Filterung ist es, $\mathbf{x}_k \in \mathbb{R}^{n_x}$ aus den Messungen $\mathbf{y}_k$ rekursiv zu schätzen. Die Messungen stehen über die Messgleichung mit dem Zustandsparametervektor im Zusammenhang:

$$\mathbf{y}_k = \mathbf{h}_k(\mathbf{x}_k) + \mathbf{v}_k, \quad (3.2)$$

wobei $\mathbf{h}_k$ eine bekannte, möglicherweise nicht-lineare Funktion des Zustandsparametervektors $\mathbf{x}_k$ ist und $\mathbf{v}_k$ den Messrauschvektor symbolisiert. Ein solches Modell nach Gl. 3.2 wird im Allgemeinen als GMM bezeichnet. Für die Rauschvektoren $\mathbf{w}_k$ und $\mathbf{v}_k$ wird oft weißes Rauschen mit bekannter PDF unabhängig voneinander angenommen. Außerdem wird davon ausgegangen, dass der Anfangszustand eine bekannte PDF $p(\mathbf{x}_0)$ aufweist und zudem unabhängig von $\mathbf{w}_k$ und $\mathbf{v}_k$ ist.

Ziel der stochastischen Filterung ist die Beschreibung des aktuellen Systemzustands über seine PDF hinsichtlich sämtlicher verfügbaren Informationen. Gesucht werden gefilterte Schätzungen des Zustandsparametervektors $\mathbf{x}_k$, welche auf der Sequenz sämtlicher verfügbaren Messungen $\mathbf{Y}_k = \{\mathbf{y}_i, i = 1, \dots, k\}$ bis zur Zeitepoche k basieren. Der Bayessche Ansatz zur Zustandsschätzung für dynamische Systeme ermittelt die PDF des unbekannten Systemzustands. Die benötigten Informationen stammen zum Teil aus einem Systemmodell (Gl. 3.1) und teilweise aus früheren Messungen. Die Zustandsschätzung wird mittels eines rekursiven Filters durchgeführt, der zwischen einer Zeitaktualisierung, die den Zustand über das Systemmodell prädiziert (Prädiktionsschritt), und einer Messungsaktualisierung (Filterungsschritt), die die Schätzung mit der aktuellen Messung korrigiert, abwechselt.

Im Prädiktionsschritt wird die PDF $p(\mathbf{x}_k | \mathbf{Y}_{k-1})$ des aktuellen Zustands unter der Berücksichtigung sämtlicher vorangegangener Messwerte $\mathbf{Y}_{k-1}$ über das Chapman-Kolmogorov-Integral

$$p(\mathbf{x}_k | \mathbf{Y}_{k-1}) = \int p(\mathbf{x}_k | \mathbf{x}_{k-1}) \cdot p(\mathbf{x}_{k-1} | \mathbf{Y}_{k-1}) d\mathbf{x}_{k-1} \quad (3.3)$$

bestimmt (Candy, 2016). Dabei wird davon ausgegangen, dass die a-posteriori PDF des Zustandsvektors aus der vorangegangenen Epoche $p(\mathbf{x}_{k-1} | \mathbf{Y}_{k-1})$ bekannt ist. Zur Durchführung des ersten Prädiktionsschrittes ist in diesem Fall eine Annahme (Vorinformation) bezüglich der a-priori PDF des Anfangszustands $p(\mathbf{x}_0)$ erforderlich. Die PDF $p(\mathbf{x}_k | \mathbf{x}_{k-1})$ beschreibt das Systemmodell und das entsprechende Systemrauschen. Diese Dichte ist als Transitionsdichte bekannt und gibt die Übergangswahrscheinlichkeit von vorangegangem Zustand in den aktuellen Zustand an. Die Transitionsdichte für das nicht-lineare Systemmodell kann über die PDF des Systemrauschens $p_{\mathbf{w}_{k-1}}$ und die nicht-lineare Systemfunktion $\mathbf{f}_{k-1}$ angegeben werden:

$$p(\mathbf{x}_k | \mathbf{x}_{k-1}) = p_{\mathbf{w}_{k-1}}(\mathbf{x}_k - \mathbf{f}_{k-1}(\mathbf{x}_{k-1}, \mathbf{u}_{k-1})). \quad (3.4)$$

Der Filterungsschritt kann allgemein über das Bayes-Theorem bestimmt werden:

$$p(\mathbf{x}_k | \mathbf{y}_k, \mathbf{Y}_{k-1}) = \frac{p(\mathbf{y}_k | \mathbf{x}_k) p(\mathbf{x}_k | \mathbf{Y}_{k-1})}{p(\mathbf{y}_k | \mathbf{Y}_{k-1})}. \quad (3.5)$$

Die a-posteriori PDF $p(\mathbf{x}_k | \mathbf{y}_k, \mathbf{Y}_{k-1})$ wird aus der Faltung der aus der Prädiktionsdichte ermittelten PDF (Gl. 3.3) und der Likelihood-PDF $p(\mathbf{y}_k | \mathbf{x}_k)$ gebildet. Der Nenner $p(\mathbf{y}_k | \mathbf{Y}_{k-1})$ in Gl. 3.5 wird als unabhängige Normierungskonstante betrachtet. Die Likelihood-PDF $p(\mathbf{y}_k | \mathbf{x}_k)$ kann äquivalent zu

$p(\mathbf{x}_k|\mathbf{x}_{k-1})$ mit Hilfe der PDF des Messrauschens $p_{\mathbf{v}_k}$ und der nicht-linearen Beobachtungsfunktion $\mathbf{h}_k$ aufgestellt werden:

$$p(\mathbf{y}_k|\mathbf{x}_k) = p_{\mathbf{v}_k}(\mathbf{y}_k - \mathbf{h}_k(\mathbf{x}_k)). \quad (3.6)$$

In Filterungsschritt 3.5 wird die Messung zum Zeitpunkt k verwendet, um die Priori-Dichte aus 3.4 zu modifizieren. Daraus ergibt sich die gewünschte a-posteriori PDF des aktuellen Zustands $p(\mathbf{x}_k|\mathbf{y}_k, \mathbf{Y}_{k-1})$. Die Darstellung im Prädiktionsschritt und im Filterungsschritt erlaubt die Bayesche rekursive Aktualisierung der a-posteriori PDF des Systemzustands zu beliebigen Zeitpunkten. Die Kenntnis der a-posteriori-PDF $p(\mathbf{x}_k|\mathbf{y}_k, \mathbf{Y}_{k-1})$ ermöglicht es, eine optimale Zustandsschätzung zu ermitteln. Durch Bildung des bedingten Erwartungswertes folgt die MMSE (Minimum Mean-Square Error)-Schätzung:

$$\hat{\mathbf{x}}_{k|k}^{\text{MMSE}} = \mathbb{E}\{\mathbf{x}_k|\mathbf{Y}_k\} = \int \mathbf{x}_k p(\mathbf{x}_k|\mathbf{Y}_k) d\mathbf{x}_k, \quad (3.7)$$

während die maximale a posteriori (MAP)-Schätzung durch das Maximum von $p(\mathbf{x}_k|\mathbf{Y}_k)$ gegeben ist

$$\hat{\mathbf{x}}_{k|k}^{\text{MAP}} = \operatorname{argmax}_{\mathbf{x}_k} p(\mathbf{x}_k|\mathbf{Y}_k). \quad (3.8)$$

In ähnlicher Weise kann ein Maß für die Genauigkeit einer Zustandsschätzung (durch die Angabe der VKM des geschätzten Zustandsvektors) auch aus $p(\mathbf{x}_k|\mathbf{Y}_k)$ gewonnen werden.

Die rekursive Fortpflanzung der a-posteriori PDF, gegeben durch Gl. 3.4 und Gl. 3.5, stellt lediglich eine rein konzeptionelle Lösung in dem Sinne dar, dass sie im Allgemeinen nicht analytisch bestimmt werden kann. Die Umsetzung der konzeptionellen Lösung erfordert die Speicherung der gesamten (nicht-Gauss'schen) PDF, welche im Allgemeinen einem unendlichen Dimensionsvektor entspricht. Nur in einer begrenzten Anzahl von Spezialfällen kann die a-posteriore Dichte genau und vollständig durch eine suffiziente Statistik charakterisiert werden. Einer der bekanntesten Spezialfälle ist das zeit-diskrete lineare Kalmanfilter (KF). Das KF geht davon aus, dass die a-posteriori PDF zu jedem Zeitschritt normalverteilt ist und somit genau und vollständig durch zwei Parameter charakterisiert wird, ihren Mittelwert und ihre VKM. Dies geschieht unter der Voraussetzung, dass bestimmte Annahmen gelten (Bar-Shalom et al., 2001):

- $\mathbf{w}_{k-1}$ und $\mathbf{v}_{k-1}$ sind normalverteilt verteilt mit bekannten Verteilungsparametern,
- $\mathbf{f}_{k-1}(\mathbf{x}_k, \mathbf{u}_k)$ und $\mathbf{h}_k(\mathbf{x}_k)$ sind bekannte lineare Funktionen.

Eine analytische Lösung von Gl. 3.4 und Gl. 3.5 ist wegen (i) Nichtlinearität der System- und Messgleichungen oder (ii) wegen Nichtstationarität des Zustandssystems oder (iii) wegen Nichtnormalverteiltheit des Prozess- und Messrauschens in den meisten praktischen Situationen nicht realisierbar, so dass approximative oder suboptimale Filterungsalgorithmen verwendet werden müssen. Im Laufe der Jahre wurde eine große Anzahl von approximativen nicht-linearen Filtern entwickelt.

Im Rahmen der Habilitation werden Aspekte aus drei nicht-linearen Filterungstechniken zur Schätzung der best-möglichen Zustandsparameter ausgewählt und diskutiert: stochastische analytische Filterungstechniken, stochastische simulationsbasierte Ansätze und kombinierte stochastische und mengenbasierte Filterungstechniken. Zu diesem Zweck werden Algorithmen vorgestellt, die die Optimierungsprobleme mit unterschiedlichen Techniken (ableitungsbasiert und simulationsbasiert) lösen (siehe auch Kapitel 2).

- Zunächst werden stochastische analytische Filterungstechniken eingeführt. Dabei wird das nicht-lineare System anhand einer Taylor-Reihe linearisiert. Insbesondere wird auf das EKF fokussiert.

- Im Anschluss werden stochastische simulationsbasierte Filterungstechniken, wie das PF vorgestellt. Das PF führt sequentielle Monte Carlo (SMC)-Schätzungen basierend auf der Darstellung der a-posteriori PDF anhand von Partikeln durch.
- Abschließend wird der KF-Ansatz verallgemeinert, um zwei Arten von Unsicherheiten – stochastische und mengenbasierte – simultan zu verarbeiten.

3.2 Vorstellung nicht-linearer Filterungstechniken

FILTERUNGSANSÄTZE I: Stochastisch-analytische Filterungstechniken

Die am häufigsten verwendeten Filterungstechniken zur Zustandsschätzung sind *stochastisch* und beruhen auf der Annahme, dass die PDFs des zu modellierenden System- und Messrauschens bekannt sind. In der überwiegenden Anzahl der Techniken wird angenommen, dass das System- und Messrauschen normalverteilt ist; es werden darüber hinaus jedoch auch weitere Verteilungen herangezogen. Die populärsten stochastischen Schätztechniken für lineare Systeme unter der Annahme der Normalverteilung sind zeit-diskrete lineare KF-Techniken (Kalman, 1960). Durch die allgegenwärtige Verwendung nicht-linearer Systeme werden mehrere stochastische Filterungstechniken für den nicht-linearen Fall entwickelt. Die am häufigsten verwendeten stochastischen analytischen Filterungsansätze für nicht-lineare Systeme sind die EKF-Techniken: EKF 1. Ordnung, das EKF 2. Ordnung und Iterative Erweiterte Kalmanfilter (IEKF). Dabei wird das nicht-lineare System anhand einer Taylor-Reihe linearisiert. Bei dem EKF 1. Ordnung wird die Taylor-Reihe nach dem linearen Glied abgebrochen, während beim EKF 2. Ordnung zusätzlich der quadratische Term in der Taylor-Reihe gebildet wird (Gelb, 1974). Beim IEKF wird versucht, die Linearisierungsfehler bei der Bildung der Taylor-Reihe des nicht-linearen Beobachtungsmodells minimal zu halten, indem die Schätzung des a-posteriori Zustandsvektors iterativ durchgeführt wird (Simon, 2006). Die Idee des EKF wurde ursprünglich von Stanley Schmidt vorgeschlagen, um das lineare KF auf nicht-lineare Navigationsprobleme von Raumfahrzeugen anwenden zu können.

Oft enthalten Optimierungsprobleme als wesentlichen Bestandteil Bedingungen und Einschränkungen bzw. Restriktionen, welche erlaubte Wertebereich des Zustandsparametervektors festlegen. Restriktionen beschränken Zustände vieler dynamischer Systeme aufgrund physikalischer Gesetze, technologischer Grenzen und geometrischer Bedingungen, wie z. B. Ebenheit von Gebäudefassaden. Solche Einschränkungen, die auch als Nebenbedingungen oder Zwangsbedingungen bezeichnet werden, können durch Gleichungen oder durch Ungleichungen gegeben sein. Die harten Restriktionen führen zu Nichtlinearitäten in Form von Abschneiden (Truncation) der Wahrscheinlichkeitsverteilungen und verkomplizieren die analytische Variabilität des nicht-linearen EKF zusätzlich. Im Allgemeinen sollte jedoch durch die Einbeziehung von Restriktionen oder a-priori Vorkenntnissen in das Filter die Zustandsparameterschätzung verbessert werden (Chiang et al., 2002; Yang et al., 2005; Simon, 2010). Weiterhin kommen im Filterungsproblem Zwangsbedingungen zwischen Beobachtungen und Zustandsparametern in Form impliziter Gleichungen vor. Bisher werden in vielen wissenschaftlichen Arbeiten lediglich explizite Messgleichungen in Form von Gl. 3.2 im Filterungsschritt berücksichtigt. Die Einführung impliziter Zwangsbedingungen in rekursiven Filterungstechniken wurde in einigen Studien entwickelt und diskutiert (Dang, 2007; Steffen und Beder, 2007).

Kurze Vorstellung des innovativen Beitrags – FILTERUNGSANSÄTZE I

Die erste Gruppe (siehe Abschnitt 3.3), die nicht-linearen stochastischen analytische Filtertechniken, ist auf nicht-lineare Zusammenhänge zwischen Ein- und Ausgangssignalen mit reinen stochastischen System- und Messrauschen beschränkt. Die analytischen stochastischen Filterungstechniken beruhen auf dem Prinzip der iterativen Approximation mit Linearisierungstechniken, d. h. IEKF-basierte Techniken anhand von Taylorreihen. Der in **P#7** und **P#8** (Vogel et al., 2018, 2019) neu entwickelte Algorithmus kann erstmalig nicht-lineare Gleichungs- und Ungleichungsrestriktionen im Rahmen eines IEKF mit impliziten oder expliziten Messgleichungen berücksichtigen. Diese Methode

wird zur Georeferenzierung eines laserscanner-basierten kinematischen Multisensorsystems in einer simulierten und realen Indoor-Umgebung (Vogel et al., 2018, 2019, **P#7** und **P#8**) sowie zur Georeferenzierung eines Unmanned-Aerial-Vehicle (UAV) (Bureick et al., 2019b, **P#9**) angewendet und evaluiert.

FILTERUNGSANSÄTZE II: Stochastische simulationsbasierte Filterungstechnik

Eine weit verbreitete simulationsbasierte Filtertechnik, die auf nicht-lineare Systeme anwendbar ist, ist das Unscented Kalmanfilter (UKF), das zuerst in Julier und Uhlmann (1997) vorgeschlagen wurde. Das UKF basiert auf der Nutzung einer speziellen nicht-linearen Transformation der Daten (genannt Unscented Transformation (UT)), welche die tatsächlichen Wahrscheinlichkeitsverteilungen in einfacher zugängliche Verteilungen transformiert. Das UKF versucht die gesamte Verteilung des Zustandsvektors anhand einiger weniger deterministisch generierter Sigma-Punkte zu approximieren. Der entscheidende Vorteil des UKF besteht darin, dass UKF-Techniken im Gegensatz zu EKF nicht davon ausgehen, dass die erforderlichen nicht-linearen Transformationen differenzierbar sind. Dadurch entfallen für UKF zeitaufwändige Berechnungen der entsprechenden ersten und zweiten partiellen Ableitungen - d. h. die Jacobi- und Hesse-Matrizen.

Eine Monte-Carlo-simulationsbasierte Filterungstechnik, die erstmalig in der geophysikalischen Literatur von Evensen (1994) vorgestellt wurde, ist das Ensemble Kalman Filter (EnKF). Im Gegensatz zum UKF propagiert, speichert und aktualisiert das EnKF mittels einer großen Anzahl von normalverteilten Stichproben (auch als Ensembles bezeichnet), was der Verteilung des Zustandsvektors entspricht. Diese Ensembles-Darstellung ist eine Form der Dimensionsreduktion des Zustandsraums, indem nur ein kleines Ensemble anstelle der gemeinsamen Verteilung propagiert wird, welches die vollständige Kovarianzmatrix des Zustandsvektors beinhaltet (Katzfuss et al., 2016).

Eine weitere verbreitete stochastische simulationsbasierte Filterungstechnik für nicht-lineare, nicht-normalverteilte und nicht-stationäre Systeme ist die PF-Technik (Gordon et al., 1993; Doucet et al., 2001). Der Kerngedanke dieser Technik ist es, die Posteriori-PDF anhand von mehreren generierten Stichproben (sogenannten Partikeln) zu approximieren. PF stellen die Verteilung sämtlicher relevanten Zufallsvariablen in der Bayesschen Rekursion durch empirische Punktmassenapproximationen dar und aktualisieren rekursiv den Zustandsvektor anhand von MC-Techniken, wie Sequential Importance Sampling (SIS) (Doucet et al., 2001). Diese Filterungstechnik bietet allgemeine Lösungen für viele Probleme, bei denen Linearisierung und Gauß'sche Approximationen nicht möglich sind oder schlechte Ergebnisse liefern würden. Nicht-Gauß'sche Annahmen des Prozess- und Messrauschens und die Einführung von Restriktionen für die Zustandsparameter können ebenfalls auf natürliche Weise durchgeführt werden. Darüber hinaus sind PF-Techniken sehr flexibel, einfach zu implementieren, parallelisierbar und in universellen komplexen Umgebungen einsetzbar.

Kurze Vorstellung des innovativen Beitrags– FILTERUNGSANSÄTZE II

In der zweiten Gruppe (siehe Abschnitt 3.4) werden entwickelte simulationsbasierte Filterungstechniken aus Alkhatib et al. (2012) **P#10** und aus Alkhatib (2015) **P#11** vorgestellt. Dieser entwickelte effiziente Ansatz basiert auf der Kombination von PF und EKF. EKPF wird anhand von simulierten Szenarien zur Verfolgung eines Ziels auf einer hoch nicht-linearen Trajektorie, welche typischerweise bei der Navigation auftritt, validiert. Außerdem wurde der EKPF-Algorithmus eingesetzt, um Transformationsparameter für die direkte Georeferenzierung von Terrestrisches Laserscanning (TLS)-basierten MSS abzuleiten.

FILTERUNGSANSÄTZE III: Kombinierte stochastische und mengenbasierte Filterungstechniken

Obwohl die oben erwähnten stochastischen Filterungstechniken in vielen praktischen Anwendungen verlässlich funktionieren, haben sie eine gemeinsame Einschränkung: Sie setzen voraus, dass

eine hinreichend detaillierte Kenntnis über die zugehörigen PDFs vorliegen muss. In der Regel sind diese Informationen nicht verfügbar. In manchen Fällen werden approximative Verteilungen festgelegt, jedoch kann sich das tatsächliche Rauschniveau stark von der Vorhersage der approximativen Verteilungen unterscheiden. Dies ist häufig der Fall bspw. bei Messinstrumenten, für die der Hersteller lediglich eine Obergrenze für die systematisch wirkenden Unsicherheitskomponenten (oder sogar für die Gesamtmessabweichung) ohne Informationen über die Wahrscheinlichkeiten verschiedener Werte innerhalb der vorgegebenen Grenzen bereitstellt (Milanese und Vicino, 1991). Ein interessanter, alternativer Ansatz, der als Set-Membership oder Unknown But Bounded (UBB)-Unsicherheitsbeschreibung bezeichnet wird, wurde durch die Arbeiten von Witsenhausen und Schweppe in den späten 60er Jahren (Witsenhausen, 1968; Schweppe, 1968) entwickelt. Bei diesem Ansatz wird die Unsicherheit durch einen additiven Anteil beschrieben, für die Intervalle bekannt sind. Motivation für den UBB-Ansatz ist die Tatsache, dass die UBB-Unsicherheitsbeschreibung in vielen Praxisfällen realistischer und weniger anspruchsvoll ist als die statistische Beschreibung durch die Festlegung einer PDF. Es gibt mehrere Techniken, um mit einer solchen Unsicherheit umzugehen: Die bekanntesten unter ihnen verwenden generische Polytope (Vicino und Zappa, 1996; Walter und Piet-Lahanier, 1989), Zonotope (Le et al., 2013; C., 2005; Schön, 2003) und Ellipsoide (Polyak et al., 2004; Durieu et al., 2001). Die Ellipsoid-basierte UBB-Technik ist im Vergleich zu den übrigen Techniken (Polytope und Zonotope) sehr effizient; sie wurde speziell für hoch-dimensionale Anwendungsfälle entwickelt. Werden Ellipsoide zur Beschreibung der UBB-Unsicherheitsbeschreibung verwendet, sind einige Probleme leicht zu lösen: Ist z. B. ein Ellipsoid in einem linearen System, in dem sich sein Initialzustand befindet, bekannt, dann ist der Satz sämtlicher möglicher Zustände in der nächsten Zeitepoche ebenfalls ein Ellipsoid. Andere Probleme sind hingegen schwieriger zu lösen. Sind beispielsweise zwei verschiedene Rauschmodelle vorhanden und jedes von ihnen durch ein Ellipsoid beschreibbar, dann ist der Satz möglicher Werte des Gesamtbudgets der Unsicherheit kein Ellipsoid. Die oben genannten Filterungstechniken berücksichtigen lediglich die Situationen, in denen entweder alle korrespondierenden PDF (stochastische Filterungstechniken) oder nur die Grenzen (obere und untere Grenze) der UBB-Unsicherheit bekannt sind. In der Praxis sind oftmals nur die Wahrscheinlichkeitsinformationen über einige Rauschkomponenten (stochastische Unsicherheitskomponenten) und die Grenzen für andere Rauschkomponenten (UBB-Unsicherheitskomponenten) bekannt. Für solche Fälle ist eine Filterungstechnik zu entwickeln, die beide Arten von Unsicherheiten berücksichtigen kann. Für den linearen Fall wurden solche Filterungstechniken von Noack (2014) und Neumann (2009) entwickelt.

Kurze Vorstellung des innovativen Beitrags – FILTERUNGSANSÄTZE III

In der dritten Klasse (siehe Abschnitt 3.5) wird zunächst der EKF-Algorithmus verallgemeinert, um stochastische und mengenbasierte Unsicherheiten simultan verarbeiten zu können. Die entwickelte Filterungstechnik EGKF wird mit bekannten Ansätzen verglichen und auf nicht-lineare Schätzprobleme angewendet. In Sun et al. (2019) **P#12** wird das EGKF hergeleitet und auf simulierte Benchmark-Probleme getestet und validiert. In Sun et al. (2018) **P#13** wird der EGKF-Algorithmus auf einen realen Datensatz angewendet, der aus einem Experiment zur Georeferenzierung eines MSS gewonnen wurde.

3.3 FILTERUNGSANSÄTZE I: Stochastisch-analytische Ansätze

Im Rahmen der Habilitationsschrift wird auf eine detaillierte Beschreibung des entwickelten IEKF und eine umfangreiche Darstellung des Standes der Forschung verzichtet. Diese Beschreibung würde zum einen den gegebenen Rahmen überschreiten, zum anderen ist sie bereits umfangreich in Vogel et al. (2018), Vogel et al. (2019) und in Bureick et al. (2019b) dargestellt.

In Vogel et al. (2018, 2019) **P#7** und **P#8** wird die grundlegende Anwendung des IEKF zur Georeferenzierung von MSS allgemein eingeführt. Dabei wird erläutert, wie explizite (Gl. 3.2) und

implizite Messgleichungen in der Form

$$\mathbf{h}_k(\mathbf{y}_k + \mathbf{v}_{k-1}, \mathbf{x}_k) = \mathbf{0} \quad (3.9)$$

sowie (nicht-)lineare Restriktionen in der Form der Gleichungen und Ungleichungen

$$\mathbf{g}(\mathbf{x}_k) = \mathbf{b}_k; \quad \text{bzw.} \quad \mathbf{lb} \leq \mathbf{g}(\mathbf{x}_k) \leq \mathbf{ub} \quad (3.10)$$

in eine rekursive Parameterschätzung zur Georeferenzierung integriert werden können. In Gl. 3.10 symbolisieren $\mathbf{g}$ realwertige Funktionen, $\mathbf{b}_k$ einen bekannten Vektor, sowie $\mathbf{lb}$ und $\mathbf{ub}$ die oberen und unteren Grenzen der Ungleichungen.

IEKF-Algorithmus für die Georeferenzierung von Multi-Sensor-Systemen

Mit zunehmender Leistung und fortschreitender Automatisierung von Hardware und Sensoren ist es möglich, eine Vielzahl von MSS zu realisieren, die perfekt auf die jeweilige Messaufgabe abgestimmt sind. MSS werden in der Ingenieurgeodäsie häufig zur effizienten Erfassung der Umgebung eingesetzt (siehe z. B. Dennig et al., 2017; Paffenholz, 2012; Hartmann et al., 2018). Dabei ist es in den überwiegenden Fällen notwendig, die gewonnenen Daten in ein übergeordnetes Koordinatensystem zu transformieren. Die Bestimmung der notwendigen Transformationsparameter wird als Georeferenzierung bezeichnet. Die Transformationsparameter sind von der Position und Orientierung des Sensors im übergeordneten Koordinatensystem abhängig. Die Position und die Orientierung werden im Allgemeinen durch drei Translationen und drei Rotationen beschrieben und zusammen als sechs Freiheitsgrade (6DoF) bezeichnet. Eine Georeferenzierung wird im Allgemeinen durch drei verschiedene Ansätze realisiert, die auf den verfügbaren Sensoren und Umgebungsbedingungen basieren. Diese Ansätze werden als *direkte*, *indirekte* und *datengesteuerte Georeferenzierung* bezeichnet (Paffenholz, 2012; Vogel et al., 2016). Bei der *direkten Georeferenzierung* werden Transformationsparameter einer Messplattform direkt aus den vorhandenen Sensoren, z. B. anhand einer GNSS-Antenne (Paffenholz, 2012), einer IMU oder eines externen Sensors (Dennig et al., 2017), beispielsweise eines Lasertrackers (Hartmann et al., 2017) oder einer Totalstation (Keller, 2013), abgeleitet. Der Ansatz ist jedoch in hohem Maße von den Umgebungsbedingungen abhängig (z. B. durch Sichtunterbrechungen in komplexen Innenräumen oder durch fehlende GNSS-Beobachtungen im Außenbereich). Bei der *indirekten Georeferenzierung* werden Beobachtungen von anderen auf der Plattform verfügbaren Sensoren wie Laserscannern oder Kameras berücksichtigt. Dabei werden die 6DoF des MSS anhand von Messungen zu definierten Zielen bestimmt. Die Ziele können flache Marker mit einem bestimmten Muster (Abmayr et al., 2008) oder einfache 3D-Geometrien wie Zylinder oder Kugeln sein (Elkhrachy und Niemeier, 2006). *Datengesteuerte Georeferenzierung* ist der *indirekten Georeferenzierung* grundsätzlich ähnlich. Dabei werden die Datensätze anstelle von bekannten Zielmarken auf Referenzdatensätze abgestimmt. Diese Referenzdatensätze können Punktwolken sein, die bereits georeferenziert sind (Soloviev et al., 2007; Glira et al., 2015), digitale Oberflächenmodelle oder 3D-Stadtmodelle (Hebel et al., 2009; Unger et al., 2017).

Um die Genauigkeit und Zuverlässigkeit der Georeferenzierung eines kinematischen MSS zu verbessern, werden Kombinationen der oben genannten Ansätze empfohlen. Die Datenfusion wird üblicherweise durch Entwicklung eines Filteransatzes im Systemzustand realisiert. Mit derartigen rekursiven Filter-Ansätzen lassen sich große Datenmengen verarbeiten, die mit aktuellen und zukünftigen Multisensortechnologien einhergehen. Darüber hinaus sind sie für Online-Applikationen geeignet und erfordern in der Regel weniger Speicher- und Rechenaufwand als Batch-Algorithmen, die auf herkömmlichen Ausgleichungsansätzen beruhen.

Abbildung 3.1 stellt eine schematische Übersicht über den universellen rekursiven Filteransatz für die Georeferenzierung eines kinematischen MSS dar; zusätzlich werden korrespondierende Zusammenhänge zwischen möglichen Zustandsparametern, Beobachtungen, vorhandenen a-priori Informationen und zugehörigen Komponenten eines IEKF gezeigt.

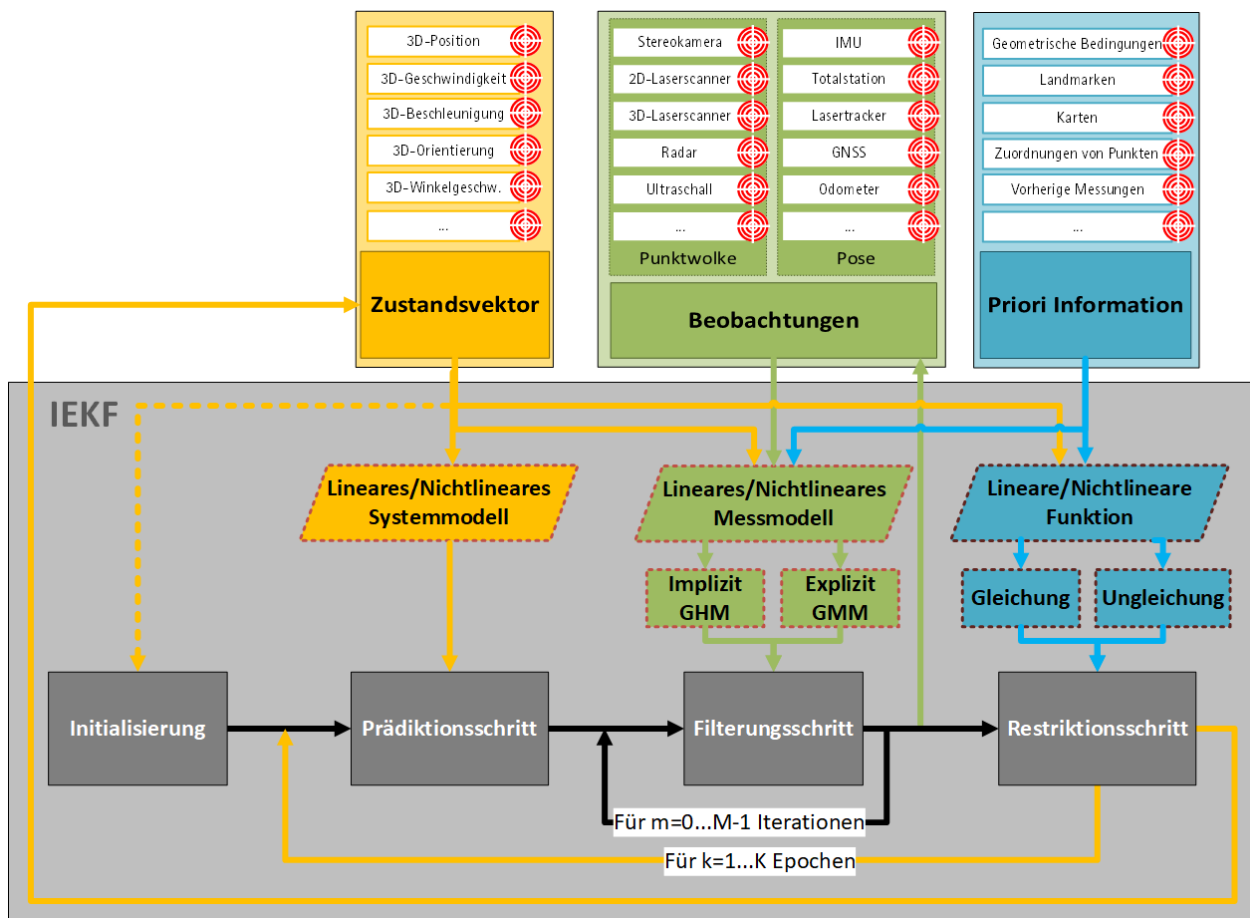


Abbildung 3.1: Schematischer Ablauf des rekursiven Filteransatzes zur Georeferenzierung eines kinematischen MSS. Die Schritte des IEKF (grau) werden mit möglichen gesuchten Zustandsparametern in gelb, verfügbaren Beobachtungen in grün und vorhandenen a-priori Informationen in blau dargestellt. Entsprechende Unsicherheitsinformationen werden durch rote Kreise gekennzeichnet; Abbildung wurde übersetzt aus Vogel et al. (2019).

Der in Vogel et al. (2018, 2019) **P#7** und **P#8** vorgestellte Ansatz für die informationsbasierte Georeferenzierung eines MSS basiert auf einem IEKF. Im IEKF wird die nicht-lineare Messgleichung anhand von Taylor-Reihenentwicklung (erster oder zweiter Ordnung) linearisiert. Die Verfeinerung des Filterungsschritts mit zusätzlichen Iterationen wird verwendet, um den Linearisierungsfehler im EKF für hochgradig nicht-lineare Systeme zu minimieren. Das benannte IEKF wurde im Allgemeinen für explizite Messgleichungen entwickelt. Wie jedoch in nachfolgenden Anwendungen gezeigt wird, erfolgt ebenfalls eine Behandlung von impliziten Messgleichungen aus Gl. 3.9, bei denen es nicht möglich ist, Beobachtungen und Parameter zu trennen. Daher wird ein IEKF genutzt, welches sich mit impliziten Messgleichungen beschäftigt. Dieses Filter wird im Folgenden aufgrund des zugrunde liegenden Gauß-Helmert-Modells (GHM) als IEKF-GHM bezeichnet.

Um die Ergebnisse der Georeferenzierung zu verbessern und damit die Genauigkeit der geschätzten Transformationsparameter der Georeferenzierung zu steigern, ist es neben der Adaption eines Filteransatzes, wie dem IEKF-GHM, grundsätzlich sinnvoll, die Szeneninformationen in die gesamte Auswertung einzubeziehen. Jede Szene hat ihre eigenen Eigenschaften, die sie exklusiv und einzigartig machen. Markante Merkmale (wie parallele oder zwei zueinander senkrechte Ebenen) sind Beispiele für solche Informationen, die bei der Datenanalyse als vorgegebene Randbedingungen oder Restriktionen verwendet werden können. In Vogel et al. (2018) wird das IEKF mittels impli-

ziter Messgleichungen (GHM) und nicht-linearer Gleichungsrestriktion zur Georeferenzierung eines kinematischen MSS in einer Indoor-Posenbestimmung verwendet, woraus gute Ergebnisse resultieren. Der finale IEKF-GHM-Algorithmus ist in Algorithmus 5 zusammengefasst. Der Algorithmus 5 verschafft einen Überblick über den Initialisierungsschritt, den Prädiktionsschritt, den Filterungsschritt sowie über die Einführung nicht-linearer Gleichungs- und Ungleichungsrestriktionen.

In Vogel et al. (2019) wird der Algorithmus erweitert, damit nicht-lineare Ungleichungsrestriktionen (Gl. 3.10) mittels abgeschnittener Wahrscheinlichkeitsdichtefunktionen (*truncated PDF*) einbezogen werden können. In diesem Ansatz wird angenommen, dass der resultierende Zustandsparameter nach dem Filterungsschritt normalverteilt ist. Diese Gauß'sche PDF wird an der oberen und unteren Grenze der Ungleichungsrestriktionen in Gl. 3.10 abgeschnitten. Die Schätzung des eingeschränkten Zustands ist gleich dem Mittelwert der abgeschnittenen PDFs. Diese Methode erweist sich als komplizierter, wenn die Zustandsdimension mehrdimensional ist. In diesem Fall wird die Zustandschätzung so normiert, dass ihre Komponenten statistisch unabhängig voneinander sind (Simon und Simon, 2010).

Algorithmus 5: Das Iterative Erweiterte Kalman-Filter mit nicht-linearer impliziter Messgleichung und nicht-linearen Gleichungsrestriktionen (IEKF-GHM).

```

1 Systemmodell  $\mathbf{x}_k = \mathbf{f}(\mathbf{x}_{k-1}, \mathbf{u}_{k-1}, \mathbf{w}_{k-1})$ ,  $\mathbf{w}_{k-1} \sim N(\mathbf{0}, \Sigma_{ww})$ 
2 Beobachtungsmodell  $\mathbf{h}(\mathbf{l}_k + \mathbf{v}_k, \mathbf{x}_k) = \mathbf{0}$ ,  $\mathbf{v}_k \sim N(\mathbf{0}, \Sigma_{vv})$ 
3 Initialisierungsschritt
4 Anfangszustandsparametervektor und seine VKM:  $\hat{\mathbf{x}}_0^+ = \mathbf{x}_0$ ,  $\Sigma_{\hat{\mathbf{x}},0}^+ = \Sigma_{xx,0}$ ,  $k = 1$ 
5 while  $k < K$  do
6   Prädiktionsschritt
7    $\mathbf{F}_{x,k} = \partial \mathbf{f} / \partial \mathbf{x} |_{\hat{\mathbf{x}}_{k-1}^+, \mathbf{u}_{k-1}, \mathbf{w}_{k-1}}$ 
8    $\mathbf{F}_{w,k} = \partial \mathbf{f} / \partial \mathbf{w} |_{\hat{\mathbf{x}}_{k-1}^+, \mathbf{u}_{k-1}, \mathbf{w}_{k-1}}$ 
9    $\hat{\mathbf{x}}_k^- = \mathbf{f}(\hat{\mathbf{x}}_{k-1}^+, \mathbf{u}_{k-1}, \mathbf{w}_{k-1})$ 
10   $\Sigma_{xx,k}^- = \mathbf{F}_{x,k} \Sigma_{xx,k-1}^+ \mathbf{F}_{x,k}^T + \mathbf{F}_{w,k} \Sigma_{ww} \mathbf{F}_{w,k}^T$ 
11  Filterungsschritt
12   $\check{\mathbf{l}}_{k,0} = \mathbf{l}_k$ ,  $\check{\mathbf{x}}_{k,0} = \hat{\mathbf{x}}_k^-$ 
13  for  $m = 0 \dots M-1$  do
14     $\mathbf{H}_{x,k,m} = \partial \mathbf{h} / \partial \mathbf{x} |_{\check{\mathbf{l}}_{k,m}, \check{\mathbf{x}}_{k,m}}$ ,  $\mathbf{H}_{l,k,m} = \partial \mathbf{h} / \partial \mathbf{l} |_{\check{\mathbf{l}}_{k,m}, \check{\mathbf{x}}_{k,m}}$ 
15     $\mathbf{K}_{k,m} = \Sigma_{xx,k}^- \mathbf{H}_{x,k,m}^T (\mathbf{H}_{x,k,m} \Sigma_{xx,k}^- \mathbf{H}_{x,k,m}^T + \mathbf{H}_{l,k,m} \Sigma_{vv} \mathbf{H}_{l,k,m}^T)^{-1}$ 
16     $\check{\mathbf{x}}_{k,m+1} = \check{\mathbf{x}}_k^- - \mathbf{K}_{k,m} \cdot (\mathbf{h}(\check{\mathbf{l}}_{k,m}, \check{\mathbf{x}}_{k,m}) + \mathbf{H}_{l,k,m} \cdot (\mathbf{l}_k - \check{\mathbf{l}}_{k,m}) + \mathbf{H}_{x,k,m} \cdot (\hat{\mathbf{x}}_k^- - \check{\mathbf{x}}_{k,m}))$ 
17     $\mathbf{G}_{k,m} = \Sigma_{vv} \mathbf{H}_{l,k,m}^T (\mathbf{H}_{x,k,m} \Sigma_{xx,k}^- \mathbf{H}_{x,k,m}^T + \mathbf{H}_{l,k,m} \Sigma_{vv} \mathbf{H}_{l,k,m}^T)^{-1}$ 
18     $\check{\mathbf{l}}_{k,m+1} = \mathbf{l}_k - \mathbf{G}_{k,m} \cdot (\mathbf{h}(\check{\mathbf{l}}_{k,m}, \check{\mathbf{x}}_{k,m}) + \mathbf{H}_{l,k,m} \cdot (\mathbf{l}_k - \check{\mathbf{l}}_{k,m}) + \mathbf{H}_{x,k,m} \cdot (\hat{\mathbf{x}}_k^- - \check{\mathbf{x}}_{k,m}))$ 
19   $\hat{\mathbf{x}}_k^+ = \check{\mathbf{x}}_{k,M}$ 
20   $\hat{\mathbf{l}}_k^+ = \check{\mathbf{l}}_{k,M}$ 
21   $\mathbf{L}_k = \mathbf{I}_{j \times j} - \mathbf{K}_{k,M-1} \mathbf{H}_{x,k,M-1}$ 
22   $\Sigma_{\hat{\mathbf{x}},k}^+ = \mathbf{L}_k \Sigma_{xx,k}^- \mathbf{L}_k^T + \mathbf{K}_{k,M-1} \mathbf{H}_{l,k,M-1} \Sigma_{vv} \mathbf{H}_{l,k,M-1}^T \mathbf{K}_{k,M-1}^T$ 
23   $\mathbf{U}_k = \mathbf{G}_{k,M-1} \cdot \mathbf{H}_{x,k,M-1}$ 
24   $\Sigma_{\hat{\mathbf{v}},k}^+ = \Sigma_{vv} + \mathbf{G}_{k,M-1} \mathbf{H}_{l,k,M-1} \Sigma_{vv} \mathbf{H}_{l,k,M-1}^T \mathbf{G}_{k,M-1}^T - \mathbf{U}_k \Sigma_{xx,k}^- \mathbf{U}_k^T$ 
25  Gleichungsrestriktionsschritt
26   $\mathbf{D} = \mathbf{g}'(\hat{\mathbf{x}}_k^-)$ 
27   $\mathbf{d} = \mathbf{b} - \mathbf{g}(\hat{\mathbf{x}}_k^-) + \mathbf{g}'(\hat{\mathbf{x}}_k^-) \cdot \hat{\mathbf{x}}_k^-$ 
28  Set  $\mathbf{W} = \mathbf{I}_{j \times j}$ 
29   $\tilde{\mathbf{x}}_k^+ = \hat{\mathbf{x}}_k^+ - \mathbf{W}^{-1} \mathbf{D}^T (\mathbf{D} \mathbf{W}^{-1} \mathbf{D}^T)^{-1} (\mathbf{D} \hat{\mathbf{x}}_k^+ - \mathbf{d})$ 
30   $\Sigma_{\tilde{\mathbf{x}},k}^+ = \Sigma_{\hat{\mathbf{x}},k}^+ - \Sigma_{\hat{\mathbf{x}},k}^+ \mathbf{D}^T (\mathbf{D} \Sigma_{\hat{\mathbf{x}},k}^+ \mathbf{D}^T)^{-1} \mathbf{D} \Sigma_{\hat{\mathbf{x}},k}^+$ 
31  Setzte  $\hat{\mathbf{x}}_k^+ = \tilde{\mathbf{x}}_k^+$  und  $\Sigma_{\hat{\mathbf{x}},k}^+ = \Sigma_{\tilde{\mathbf{x}},k}^+$ 

```

Anwendungsfälle zum IEKF-GHM

Im Folgenden sollen die Grundidee des Ansatzes aus Vogel et al. (2018, 2019) **P#7** und **P#8** und Bureick et al. (2019b) **P#9** und die daraus resultierenden Messgleichungen vorgestellt werden. Die Grundidee ist graphisch in Abbildung 3.2 dargestellt. Im rechten Teil der Abbildung ist eine Ebene dargestellt; mathematisch kann diese Ebene durch den Normalenvektor $\mathbf{n} = [n_x, n_y, n_z]^T$ und den Abstandparameter d in der Hesseschen Normalform repräsentiert werden. Voraussetzung dafür ist, dass $\mathbf{n}$ auf die Länge 1 normiert ist, also $|\mathbf{n}| = \sqrt{n_x^2 + n_y^2 + n_z^2} = 1$ erfüllt ist.

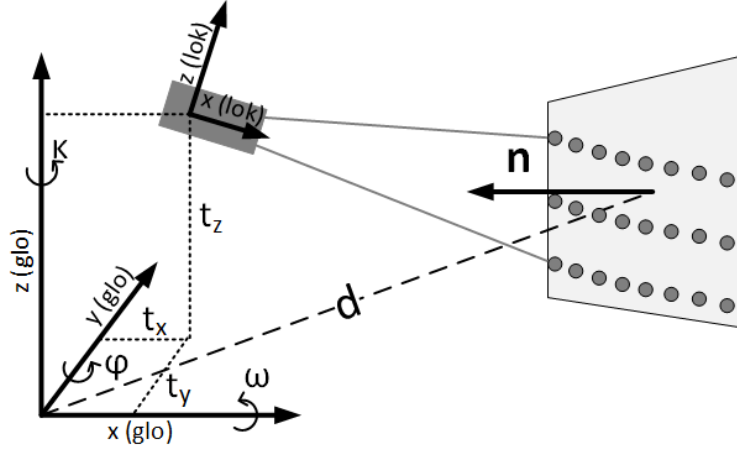


Abbildung 3.2: Die Grundidee des Ansatzes IEKF-GHM (Bureick et al., 2019c, modifiziert)

Der Laserscanner, welcher in Abbildung 3.2 als dunkelgraue, rechteckige Fläche gekennzeichnet ist, misst 3D-Punkte, die auf der Ebene (z. B. Wand eines Gebäudes) liegen sollen. Diese Punkte sind in Abbildung 3.2 als dunkelgraue Kreise auf der oben beschriebenen Ebene dargestellt. Die Punkte liegen in der Regel im lokalen Koordinatensystem des Laserscanners vor. Um sie vom lokalen Koordinatensystem (lok) in das übergeordnete Koordinatensystem (glo) transformieren zu können, ist die Bestimmung der 6 DoF $t_x, t_y, t_z, \omega, \varphi$ und κ erforderlich.

Die Grundidee des Ansatzes zur Georeferenzierung ist es, die Transformationsparameter mit den Ebenenparametern ($\mathbf{n}, d$) und den Messungen des Laserscanners in einem funktionalen Zusammenhang zu verknüpfen.

Der funktionale Zusammenhang ist in impliziter Form (Gl. 3.9) dargestellt:

$$\mathbf{n}^T \times \underbrace{\left[[t_x, t_y, t_z]^T + \mathbf{R}(\omega, \varphi, \kappa) \times [x_{lok}, y_{lok}, z_{lok}]^T \right]}_{[x_{glo}, y_{glo}, z_{glo}]^T} - d = 0. \quad (3.11)$$

Ein beliebiger vom Laserscanner gemessener Punkt $[x_{lok}, y_{lok}, z_{lok}]^T$ liegt im lokalen Koordinatensystem des Laserscanners vor. Dieser kann mithilfe der sechs unbekannten Transformationsparameter $t_x, t_y, t_z, \omega, \varphi$ und κ in das übergeordnete Koordinatensystem transformiert werden. $\mathbf{R}(\omega, \varphi, \kappa)$ in Gl. 3.11 symbolisiert die Rotationsmatrix. Der in das übergeordnete Koordinatensystem transformierte Punkt $[x_{glo}, y_{glo}, z_{glo}]^T$ wird in die Hessesche Normalform der Ebene, die durch den Normalenvektor und den Abstandparameter beschrieben werden kann, eingesetzt. Bei der Einordnung der Ebenenparameter ($\mathbf{n}, d$) ist es möglich, diese als bekannt oder als Parameter (unbekannt) einzuführen.

Anwendung 1: Georeferenzierung eines MSS in einer Indoor-Umgebung:

Innerhalb der ersten Anwendung wird der vorgeschlagene IEKF-GHM (vgl. Algorithmus 5) auf ein kinematisches MSS angewendet, das in Hartmann et al. (2017, 2018) ausführlich beschrieben

wurde. Das kinematische MSS besteht aus einem 3D-TLS, einem Lasertracker und einer speziellen T-Probe (Kombination aus einem Reflektor und zehn LEDs zur 6 DoF Schätzung hinsichtlich des Lasertrackers). Es handelt sich beim TLS um einen Zoller + Fröhlich Imager 5016, welcher für diese Anwendung in einem 2D-Profilmodus verwendet wurde (siehe Abbildung 3.3 links). Die Datenerfassung für diese Fallstudie erfolgt im Untergeschoss des Geodätischen Instituts in einem ausgewählten Abschnitt eines Korridors (vgl. Abbildung 3.3, rechts). Da dieses MSS über keinen Sensor (Inertial Measurement Unit (IMU)) zur Erfassung der Orientierungsänderungen für die Georeferenzierung verfügt, wurden die IMU-Poseninformationen für zwei unterschiedliche IMUs simuliert (moderate und präzise IMU). Um die Unabhängigkeit von simulierten IMU-Poseninformationen zu gewährleisten, sind die Ergebnisse in diesem Abschnitt bezogen auf den Mittelwert von 500 Wiederholungen im Rahmen einer umfassenden MC-Simulation, welche zufällig aus der Normalverteilung generiert wurden. Eine detaillierte Beschreibung der durchgeführten MC-Simulation ist in Vogel et al. (2019) zu finden.

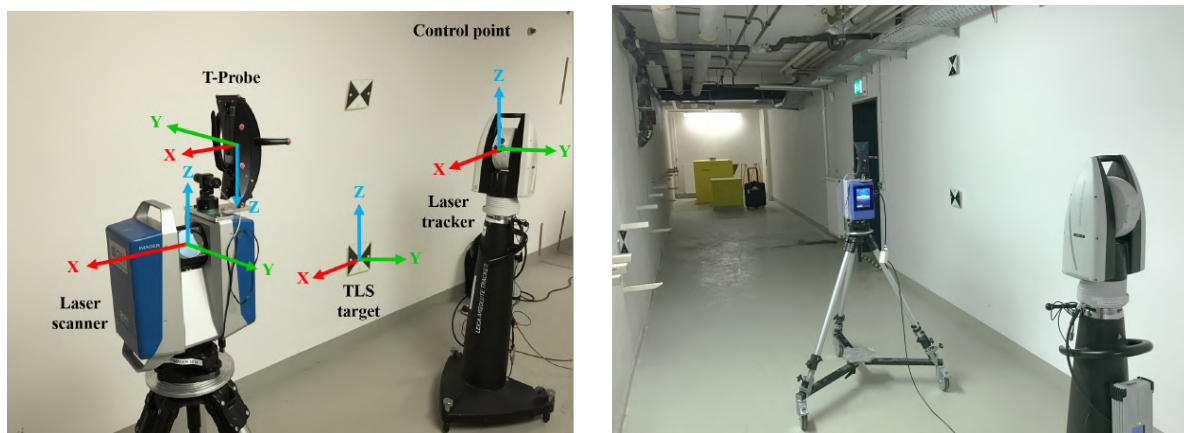


Abbildung 3.3: Links: Gesamtansicht des kinematischen MSS, welches die verwendete Sensoren und die unterschiedlichen Koordinatensysteme veranschaulicht, rechts: durchgeführtes Experiment im Untergeschoss des GIH; Originalabbildungen aus Vogel et al. (2019)

Unterschiedliche Methoden zur Berücksichtigung von geometrischen Kombinationen von Restriktionen in Form von Gleichungen und Ungleichungen wurden untersucht und validiert. Basierend auf den Poseninformationen kann der RMSE für die kombinierte Position in X-, Y-, Z-Richtung berechnet werden. Um einen Qualitätsparameter für die kombinierte Orientierung zu geben, wird die Transformation von der Rotationsmatrix $\mathbf{R}_k$ zur Achsen-Winkel-Repräsentation durch einen normierten Vektor $\mathbf{r}_k = [r_1, r_2, r_3]$ und Drehwinkel Θ_k durchgeführt. Anschließend wird der mittlere Fehler (ME) des repräsentativen Winkels zwischen Schätzung Θ_k^{MSS} und durch den Lasertracker verfügbare Referenzdaten Θ_k^{GT} berechnet und verwendet.

In allen Kombinationen (siehe Abbildung 3.4, links) nimmt der RMSE innerhalb der ersten Epochen stark zu. Nach dem Einlaufeffekt des Filters nehmen die RMSE-Werte schnell ab und setzen sich mit der Zeit unterschiedlich fort. Über alle Epochen hinweg führt die Kombination II zu einem signifikant größeren RMSE und hat den größten Anstieg. Die Kombinationen III, V und VIII sind sich sehr ähnlich und führen zu den präzisesten Ergebnissen (bei etwa 1,5 cm). Die restlichen Kombinationen haben einen größeren Anstieg und erreichen am Ende Werte zwischen 2 und 3 cm. Das zeitliche Verhalten des ME wird in Abbildung 3.4 (rechts) für die geschätzten Orientierungen dargestellt. Die ME-Werte weisen einen steilen Anstieg innerhalb der ersten Epochen auf. Es liegen leichte Abweichungen zwischen allen Kombinationen vor. Die Kombination III weist die geringsten ME-Werte unter sämtlichen Kombinationen auf und ist bis zu Epoche 400 der IMU-Lösung am

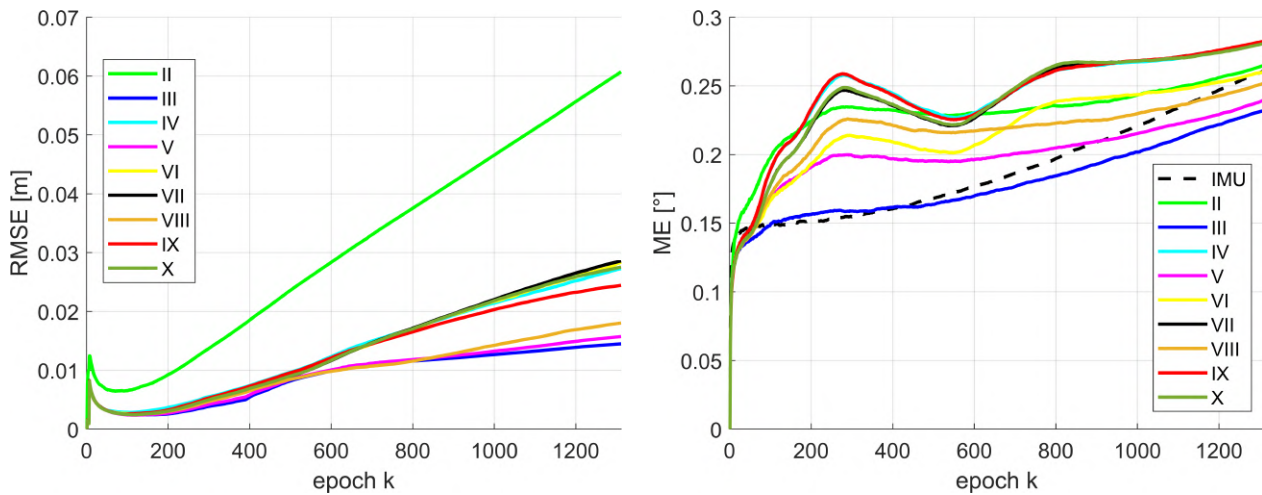


Abbildung 3.4: Präzise IMU: Zeitlicher Verlauf des Medians des RMSE für die Position (links) und ME für die Orientierung (rechts) mittels 500 Wiederholungen für entsprechende Kombinationen der verwendeten Zustandsrestriktionen. Die römischen Ziffern beziehen sich auf die jeweiligen Zustandsrestriktionen, die in Bezug auf die Tabelle 3 aus Vogel et al. (2019) angewendet werden.

ähnlichsten; danach steigt die IMU-Lösung bis zum Wert $[ME = 0,25^\circ]$ an.

Zusammenfassend wurde festgestellt, dass die Berücksichtigung geeigneter Restriktionen zwischen den Zustandsparametern sinnvoll ist. Die Verwendung von zusätzlichen Ungleichungsrestriktionen bietet signifikante Verbesserungen der Genauigkeit der geschätzten Trajektorie (Vogel et al., 2019). Dies begründet die allgemeine Berücksichtigung von Ungleichungsrestriktionen bei der Georeferenzierung eines kinematischen MSS. Für eine ausführliche Beschreibung des durchgeführten Experiment und die Analyse der Ergebnisse wird auf Vogel et al. (2019) **P#8** verwiesen.

Anwendung 2: Georeferenzierung eines UAV

In der zweiten Anwendung handelt es sich um eine Georeferenzierung eines simulierten UAV anhand des entwickelten Algorithmus 5. Abbildung 3.5 zeigt den schematischen Aufbau des UAV bestehend aus einem Laserscanner, einem GNSS-Empfänger und einer IMU. Als Laserscanner wird der VLP-16 der Firma Velodyne verwendet. Durch insgesamt 16 Scanlinien erreicht dieser Laserscanner ein Sichtfeld (*Field of view*) von $360^\circ \times 30^\circ$ und damit eine flächenhafte Erfassung. Beim VLP-16 korrespondiert die Auflösung innerhalb der Scanlinie mit der Messfrequenz. Mit der geringsten Messfrequenz von 5 Hz erreicht der VLP-16 eine Auflösung von $0,1^\circ$, mit der höchsten Messfrequenz von 20 Hz reduziert sich die Auflösung auf $0,4^\circ$.

In Bureick et al. (2019b) **P#9** wurde gezeigt, dass die Bestimmung der UAV-Trajektorie anhand des entwickelten Algorithmus besser als 5 cm in der Position (Medianwert) und $0,08^\circ$ in der Orientierung (Medianwert) bestimmt werden kann. Eine kontinuierliche Verfügbarkeit von GNSS- und IMU-Beobachtungen für das implementierte IEKF ist nicht zwingend erforderlich. Die Trajektorie wird hauptsächlich durch Laserscanner-Messungen von Gebäudefassaden bestimmt, die als Ebenen in einem 3D-Stadtmodell modelliert werden können. Die Laserscanner-Messungen und die Ebenen des 3D-Stadtmodells werden innerhalb des IEKF durch implizite Messgleichungen und nicht-lineare Gleichungsrestriktionen kombiniert. Um die Funktionalität und Leistungsfähigkeit des implementierten IEKF zu demonstrieren, wurde eine umfassende MC-Simulation realisiert. Dafür wurden zwei Szenarien simuliert, in denen die Laserscanner- und IMU-Messungen unter bestimmten Annahmen bezüglich Messgenauigkeit und Verzerrung unterschiedlich erzeugt wurden. Die Simulation wurde für beide Szenarien 500 mal wiederholt. In Szenario 1 wurde bei den Messungen des Laserscanners, der Position (die den GNSS-Empfänger repräsentiert) und der Orientierung (die die IMU repräsentiert) ausschließlich normalverteiltes Rauschen hinzugefügt. In Szenario 2 wurden

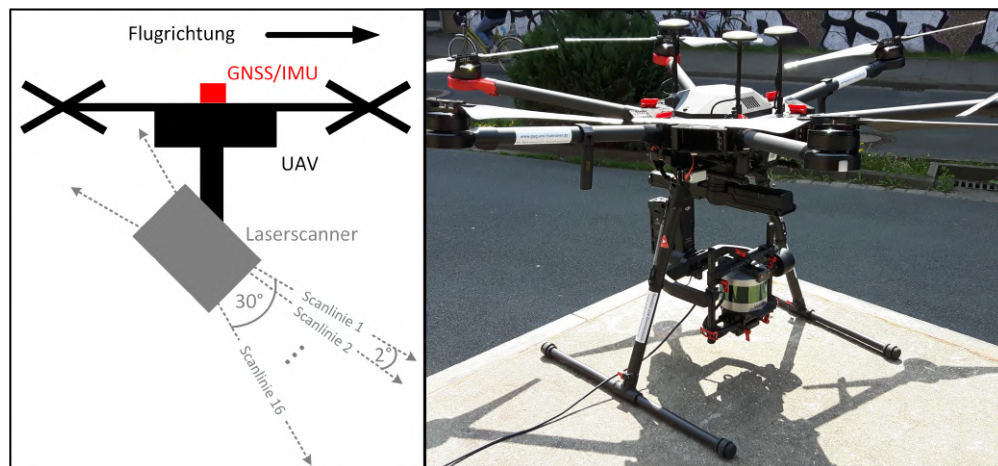


Abbildung 3.5: Schematischer (links) und tatsächlicher Aufbau des UAV (rechts); Originalabbildung aus Bureick et al. (2019c).

diejenigen Messungen des Laserscanners systematisch verfälscht, die auf die Fenster des simulierten Gebäudes treffen. Zusätzlich wurde das Rauschniveau dieser Messungen erhöht. Die Ergebnisdarstellungen nach Abbildung 3.6 unterstützen im allgemeinen die in Anwendungsfall 1 getroffenen Interpretationen. Der RMSE der vom IEKF bestimmten Translationsparameter t_x und t_z nimmt mit zunehmender Anzahl von Epochen ab. In der letzten Epoche sind die $RMSE_{t_x}$ und $RMSE_{t_z}$ des IEKF deutlich kleiner als der RMSE, der durch den klassischen linearen KF in beiden Szenarien erhalten wurde. Im Gegensatz dazu unterscheidet sich der $RMSE_{t_y}$ und steigt nach Epoche 10 mit zunehmender Anzahl von Epochen an. In beiden Szenarien, insbesondere im Szenario 2, ist der durch den IEKF erhaltene $RMSE_{t_y}$ sogar größer als der des linearen KF in den letzten Epochen. Der Grund dafür ist, dass Zusatzinformation in Form von Laserscanner-Messungen in Richtung t_y fehlt. Die RMSE-Werte der Orientierungsparameter weisen unterschiedliche Eigenschaften auf, sind aber an sich recht ähnlich.

Es wurde in diesen Simulationen festgestellt, dass mit dem IEKF gute Ergebnisse erzielt werden können, auch wenn systematisch verfälschte Beobachtungen vorliegen. Eine ausführliche Beschreibung der durchgeführten MC-Simulation und eine Analyse der Ergebnisse findet sich in Bureick et al. (2019b) **P# 9**.



Innovation kurz zusammengefasst | Um eine genaue Georeferenzierung von MSSs in schwierigen Indoor- oder Outdoor-Umgebungen zu gewährleisten, realisiert dieses entwickelte Filter einen informationsbasierten Ansatz mit rekursiver Zustandsschätzung. Dies wird durch eine optimale Zusammenführung der gemessenen Objektrauminformationen mit zuverlässigen Vorinformationen, z. B. geometrische Bedingungen, realisiert. Diese Restriktionen basieren auf unabhängigen geometrischen Informationen (z. B. Gesamttoleranzen im Bauwesen unter Verwendung von Normen). Dies führt zu einem iterativen erweiterten Kalman-Filter (IEKF) mit nichtlinearen Gleichheits- und Ungleichheitszustandsbedingungen. Die Fähigkeit, sowohl explizite als auch implizite Formulierungen des funktionalen Zusammenhangs zwischen Zuständen und Beobachtungen in Kombination mit Zustandsbedingungen zu verwenden, ist ebenfalls eine innovative Besonderheit. Dies ermöglicht es, die Integrität (Maß des Vertrauens in technische Systemen) von MSS dauerhaft zu garantieren und die Genauigkeit und Zuverlässigkeit der erhaltenen Pose zu erhöhen.

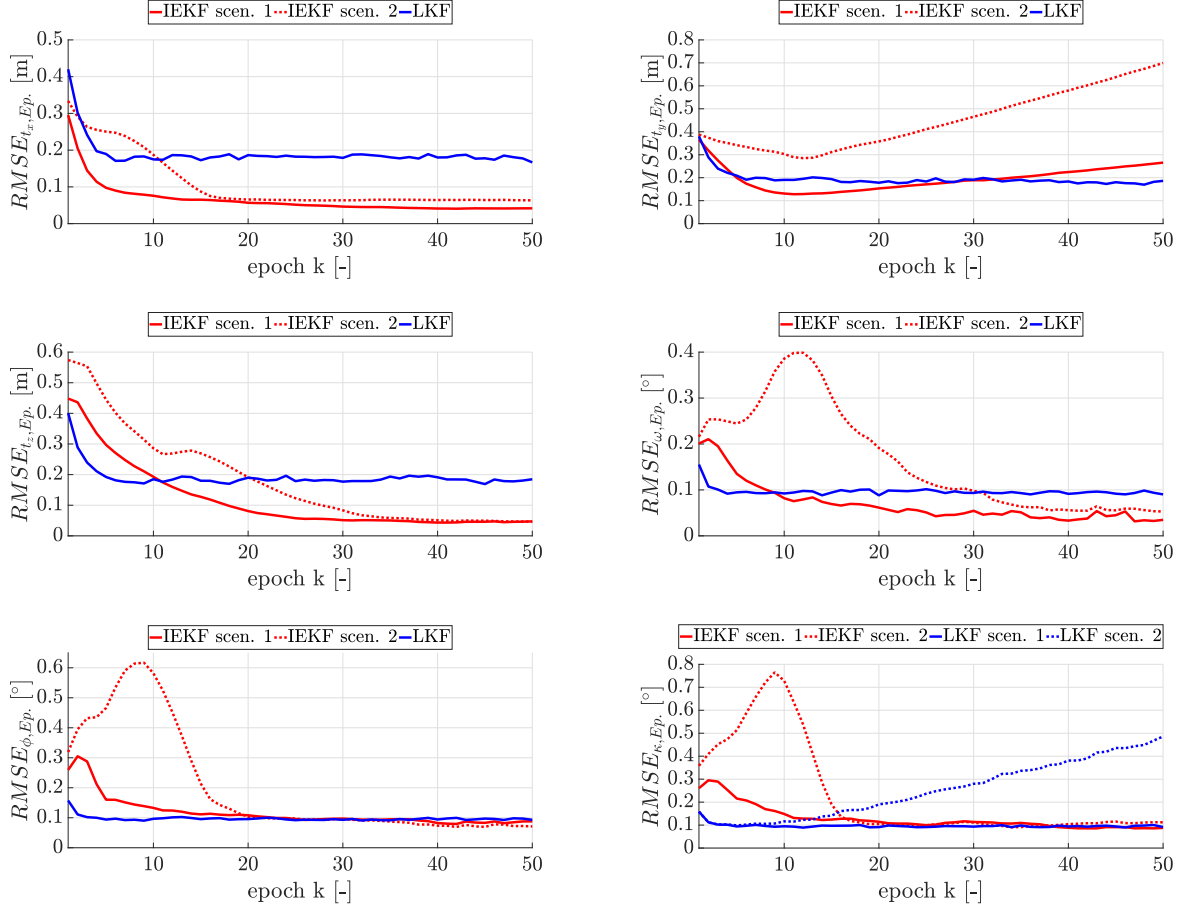


Abbildung 3.6: Mittlere RMSE-Werte für jede Epoche über die gesamte MC-Simulationen für alle geschätzte 6 -DoF: Oben links für die Translation in Richtung x ; Oben rechts für die Translation in Richtung y , Mitte links für die Translation in Richtung z ; Mitte rechts für die Orientierung ω ; unten links für die Orientierung Φ ; unten rechts für die Orientierung κ

3.4 FILTERUNGSANSÄTZE II: Monte-Carlo-Simulationsbasierte Ansätze

In den letzten Jahrzehnten hat sich eine weitere Gruppe von simulationsbasierten Lösungen zur numerischen Approximation des Bayesschen Schätzers (siehe Gl. 3.3 und Gl. 3.5) etabliert, die als Sequential Monte-Carlo (SMC)-Filter oder PF bekannt sind (nicht nur in der Theorie, sondern auch als praktikable Lösungen). Dabei bildet der SIS-Algorithmus die Grundlage für die meisten in den letzten Jahrzehnten entwickelten SMC-Filter. Die Grundidee des SMC basiert auf der Generierung einer großen Menge von Stichproben (sogenannte Partikel), die die posteriori PDF (Gl. 3.5) durch einen Satz gewichteter Stichproben approximiert. Eine explizite Annahme über die Form der a-posteriori PDF ist nicht erforderlich und kann daher in nicht-linearen, nicht-Gaußschen Systemen verwendet werden. Die Partikelmenge $\{\mathbf{x}_{0:k}^i, w_k^i\}$ mit $i \in \{1, \dots, M\}$ repräsentiert die a-posteriori PDF für die Zeit k :

$$p(\mathbf{x}_k | \mathbf{y}_{1:k}) \approx \sum_{i=1}^M w_k^i \delta(\mathbf{x}_{0:k} - \mathbf{x}_{0:k}^i), \quad (3.12)$$

wobei δ die Kronecker Delta-Funktion repräsentiert; die zugehörigen Gewichte w_k^i summieren sich zu eins. Die Gewichte w_k^i in Gl. 3.12 werden rekursiv basierend auf der Gewichtsaktualisierungs-

gleichung berechnet (Ristic et al., 2004):

$$w_{k+1}^i = w_k^i \frac{p(\mathbf{y}_{k+1}|\mathbf{x}_{k+1}^{(i)})p(\mathbf{x}_{k+1}^{(i)}|\mathbf{x}_k^{(i)})}{\pi(\mathbf{x}_{k+1}^{(i)}|\mathbf{x}_{1:k}^{(i)}, \mathbf{y}_{1:k})}, \quad (3.13)$$

wobei $\pi(\mathbf{x}_{1:k+1}|\mathbf{y}_{1:k+1})$ eine bekannte PDF ist, aus der möglichen Stichproben generiert werden können. Der SIS-Resampling-Algorithmus startet mit der Anzahl N der Initialisierungsstichproben des Zustandsvektors (symbolisiert mit $\mathbf{x}_0^i$ mit $i \in 1, \dots, N$), die zufällig aus der festgelegten initialen PDF $\pi(\mathbf{x}_0)$ generiert werden können. Diese Stichproben (genannt auch Partikel) werden zu jeder Epoche $k = 1, 2, \dots$ des Prädiktionsschritts durch Substitution in der dynamischen Gl. 3.1 propagiert. Die daraus resultierenden Partikel werden als $\mathbf{x}_{k,i}^-$ bezeichnet. Sobald die aktuellen Beobachtungen $\mathbf{y}_k$ verfügbar sind, wird die bedingte Wahrscheinlichkeit der einzelnen Partikel berechnet ($p(\mathbf{y}_k|\mathbf{x}_{k,i}^-)$). Die Auswertung der Wahrscheinlichkeit basiert auf der bekannten PDF des Messrauschens und auf der nicht-linearen Messgleichung (Gl. 3.2). Anhand von Gl. 3.13 werden die relativen Gewichte berechnet. Vor der Schätzung des aktuellen Zustandsvektors basierend auf dem MAP-Schätzer (Gl. 3.8) und vor dem Übergang zum nächsten Zeitschritt werden die Partikel neu generiert (Resampling), d. h., es werden zufällig neue Partikel $\mathbf{x}_{k,i}^+$ basierend auf den relativen Gewichten generiert. Partikel mit relativ geringem Gewicht werden eliminiert und Partikel mit großem Gewicht hingegen werden vervielfältigt. Dieser Resamplingsschritt wird verwendet, um das Problem der Degeneration gezogener Partikel zu vermeiden, siehe z. B. Doucet et al. (2001). Ein Defizit des PF-Algorithmus ist die durch die Erhöhung der generierten Partikel verursachte Rechenzeit. Die Generierung einer großen Anzahl von Partikeln ist erforderlich, um die Konvergenz des Algorithmus zu gewährleisten, damit die gesuchte Zustandsvektorschätzung und dessen VKM möglichst präzise approximiert werden können. In Alkhatib et al. (2012) **P#10** und in Alkhatib (2015) **P#11** wurden Algorithmen entwickelt, um die Leistung von PF zu verbessern und die erforderliche Anzahl der generierten Partikel zu verringern. Der entwickelte Algorithmus wurde als Extended Particle Filter (EKPF) bezeichnet und besteht aus einer Kombination von EKF und PF.

Extended Kalman Particle Filter (EKPF)

Der Hauptunterschied zwischen dem generischen PF-Algorithmus und dem EKPF-Algorithmus, ist die Aktualisierung jedes Partikels im Filterungsschritt $\mathbf{y}_k$ unter Verwendung des EKF. Dies bedeutet, dass für jedes Partikel i ein zusätzlicher EKF-Schritt durchgeführt wird. Die Berechnungsschritte des Filters werden durch den Algorithmus 7 zusammengefasst. Für die ausführliche Herleitung und Beschreibung wird auf Alkhatib et al. (2012) **P#10** verwiesen. Der entwickelte EKPF-Ansatz wurde auf ein simuliertes Experiment und auf eine reale Anwendung im Rahmen einer direkten Georeferenzierung eines MSS angewendet. Im simulierten Szenario wird ein Ziel verfolgt, welches sich entlang einer hochgradig nicht-linearen Trajektorie bewegt. Der Zustandsvektor beinhaltet neben der 2D-Position auch den 2D-Geschwindigkeitsvektor $\mathbf{x} = [x_k \ y_k \ \dot{x}_k \ \dot{y}_k]$ im kartesischen Koordinatensystem. Das Modell setzt sich aus einem linearen zeit-diskreten kinematischen Modell und einem nicht-linearen Messmodell zusammen. Das dynamische System wurde mit einem diskreten konstanten Wiener Geschwindigkeitsmodell modelliert (Bar-Shalom et al., 2001). In diesem Experiment wurden die horizontale Distanz und die Winkel von zwei Beobachtungsstationen an den bekannten Punkten S_1 und S_2 zu dem bewegten Fahrzeug gemessen. Eine ausführliche Beschreibung der durchgeführten MC-Simulation und deren Einstellungsparameter wurde in Alkhatib (2015) **P#11** gegeben. Die MC-Simulation wurde 500 mal durchgeführt. Die RMSE-Werte für die 500 Durchläufe sind in der Abbildung 3.7 dargestellt.

Es ist deutlich zu erkennen, dass in nahezu sämtlichen Durchläufen die RMSE-Werte für das EKPF geringer im Vergleich zum EKF und zum generischen PF sind.

Der Algorithmus 7 wurde außerdem zur Ableitung der Positions- und Orientierungsparameter für

Algorithmus 7: Das Erweiterte Kalman Partikel Filter (EKPF)

-
- 1 Systemmodell $\mathbf{x}_k = \mathbf{f}(\mathbf{x}_{k-1}, \mathbf{u}_{k-1}, \mathbf{w}_{k-1})$, $\mathbf{w}_{k-1} \sim N(\mathbf{0}, \Sigma_{ww})$
 - 2 Beobachtungsmodell $\mathbf{h}(\mathbf{l}_k + \mathbf{v}_k, \mathbf{x}_k) = \mathbf{0}$, $\mathbf{v}_k \sim N(\mathbf{0}, \Sigma_{vv})$
Input: Jakobi-Matrix $\mathbf{H}$; Transitionsmatrix $\mathbf{F}$; Beobachtungsvektor $\mathbf{y}_k$; VKM der Beobachtungen Σ_{vv} ; und VKM des Systemrauschens Σ_{ww}
Output: geschätzter Zustandsvektor $\hat{\mathbf{x}}_k^+$ und seine VKM $\Sigma_{\mathbf{x}_k^+, \mathbf{x}_k^+}$
 - 3 **Initialisierungsschritt:** Generiere $\mathbf{x}_{0,i}^+ \sim p(\mathbf{x}_0)$
 - 4 **for** $k = 1, 2, \dots$ **do**
 - 5
$$\Sigma_{k,i}^- = \mathbf{F}_{k-1,i} \Sigma_{k-1,i}^+ \mathbf{F}_{k-1,i}^T + \Sigma_{k-1,ww}$$

$$\mathbf{K}_{k,i} = \Sigma_{k,i}^- \mathbf{H}_{k,i}^T (\mathbf{H}_{k,i} \Sigma_{k,i}^- \mathbf{H}_{k,i}^T + \Sigma_{k,vv})^{-1}$$

$$\mathbf{x}_{k,i}^+ = \mathbf{x}_{0,i}^- + \mathbf{K}_{k,i} [\mathbf{y}_k - \mathbf{h}(\mathbf{x}_{k,i}^-)]$$

$$\Sigma_{k,i}^+ = (\mathbf{I} - \mathbf{K}_{k,i} \mathbf{H}_{k,i}) \Sigma_{k,i}^-$$

Partikel propagieren: $\mathbf{x}_{k,i} \sim p(\mathbf{x}_k^i | \mathbf{x}_{k-1}^i, \mathbf{y}_k) \sim \mathcal{N}(\mathbf{x}_{k,i}^+, \Sigma_{k,i}^+)$
 - 6 Gewichte aktualisieren: $w_{k+1}^i = w_k^i \frac{p(\mathbf{y}_{k+1} | \mathbf{x}_{k+1}^{(i)}) p(\mathbf{x}_{k+1}^{(i)} | \mathbf{x}_k^{(i)})}{\pi(\mathbf{x}_{k+1}^{(i)} | \mathbf{x}_{1:k}^{(i)}, \mathbf{y}_{1:k})}$
 - 7 Resamplingschritt: $\mathbf{x}_{k,i}^+ = \text{Resample}(\mathbf{x}_{k,i}^+)$
-

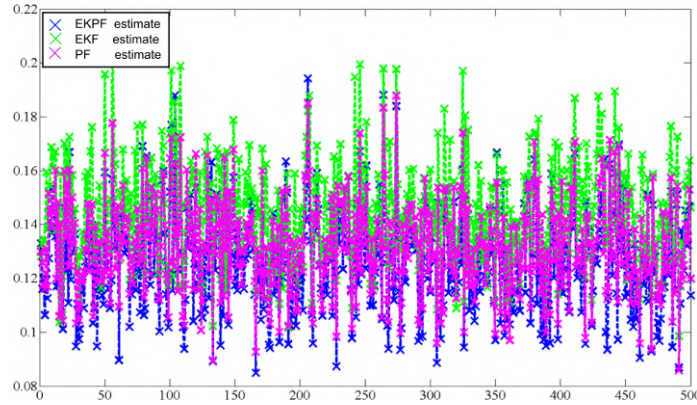


Abbildung 3.7: RMSE-Werte für die nicht-linearen Filterungsergebnisse mit dem neuen entwickelten EKPF-Algorithmus (blau) verglichen zu dem generischen PF (magenta) und zum EKF (grün). Insgesamt wurden 500 MC-Durchläufe wiederholt; Originalabbildung aus Alkhatib (2015).

die MSS angewendet, welche in Abschnitt 2.4.2 beschrieben und in Abbildung 2.4 (links) dargestellt sind. Abschließend wurden der klassische EKF-Algorithmus, welcher in Paffenholz (2012) entwickelt wurde, und der EKPF-Algorithmus (7) verwendet, um die genannten Transformationsparameter zu bestimmen.

Abbildung 3.8 stellt die geschätzten Zustandsparameter durch den klassischen EKF-Algorithmus (schwarz) und EKPF-Ansatz (grau) zusammen. Die obere Abbildung zeigt die Residuen, die innerhalb einer linearen Regression für die Orientierung erzielt wurden. Die Residuen sind in beiden Algorithmen vergleichbar und führen für die Azimutberechnung zu einer metrischen Unsicherheit von ca. 1,5 cm in einer Entfernung von bis zu 35 m. Der untere Teil von Abbildung 3.8 zeigt einen Vergleich zwischen den gefilterten Neigungen. Auch hier fällt der Filterungseffekt beim EKPF bei

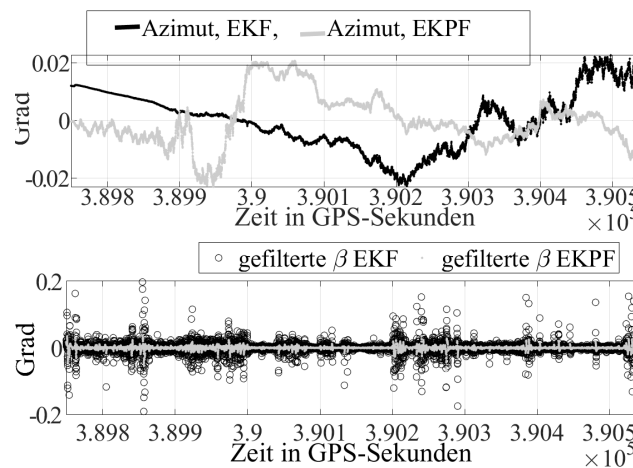


Abbildung 3.8: Ergebnisse von EKF- und EKPF-Algorithmen; oben: Residuen, die mittels einer linearen Regression zur Bestimmung der Orientierung gewonnen wurden; unten: die gefilterten Neigungen. Der EKPF-Ansatz zeigt eine signifikante Verbesserung der Filterwirkung für die gefilterte Neigung; Abbildung übersetzt aus Alkhatib et al. (2012).

den gefilterten Neigungen im Vergleich zum EKF auf, vor allem bei höherem Rauschniveau (z. B. am Ende der Zeitreihe).



Innovation kurz zusammengefasst | Die Überlegenheit der sequentiellen Monte-Carlo-Techniken in hochgradig nicht-linearen und nicht-Gaußschen Systemen zeigt sich in deren breitem Anwendungsspektrum in vielen Fachdisziplinen. In diesem Abschnitt wurde ein neues Filter (das Erweiterte Partikel-Kalmanfilter, EKPF) entwickelt, welches auf einer Kombination von Partikelfilter (PF)- und Erweiterten Kalmanfilter (EKF)-Techniken basiert. Als Ergebnis wird festgestellt, dass der entwickelte EKPF-Algorithmus die Standardfilteralgorithmen – PF und EKF – deutlich übertrifft. Der Hauptvorteil des entwickelten EKPF-Algorithmus liegt darin, bessere Schätzergebnisse bei hochgradig nicht-linearen Systemgleichungen zu erzielen. Ein zweiter wichtiger Vorteil ist die signifikante Verringerung der Anzahl der generierten Partikel im Vergleich zum reinen PF.

3.5 FILTERUNGSANSÄTZE III: Kombinierte stochastisch-deterministische Ansätze

Im Allgemeinen ist weder eine exakte Beschreibung der zugrunde liegenden Systemdynamik möglich noch sind Sensorbeobachtungen fehlerfrei. Daher ist die Ableitung spezifischer Schätzungen aus verrauschten Messungen wenig sinnvoll, wenn die damit verbundenen Unsicherheiten nicht adäquat berücksichtigt werden. Das bedeutet, dass unpräzise Informationen ungenaue Schätzergebnisse und Schlussfolgerungen verursachen. Eine zentrale Herausforderung in der Theorie der Zustandsschätzung besteht daher darin, geeignete Modelle zu definieren, um den Mangel an präzisen Informationen ausdrücken zu können. Der Einsatz von Unsicherheitsmodellen kann wesentlich dazu beitragen, Robustheit und Zuverlässigkeit in vielen geodätischen Anwendungen, z. B. zur Georeferenzierung eines MSS oder in Navigationsanwendungen, zu gewährleisten. Zu diesem Zweck ist es notwendig, Unsicherheitscharakterisierungen während des gesamten Zustandsschätzalgorithmus fortzupflanzen und zu aktualisieren. In einem dynamischen Zustandsschätzproblem wirkt sich beispielsweise die Unsicherheit, die der Anfangsposition eines MSS zugeordnet ist, auch auf die Positionsschätzung zu einem späteren Zeitpunkt aus und muss daher auch mit der relevanten Unsicherheit durch

das Bewegungsmodell propagiert werden. In der Theorie der Zustandsschätzung bezieht sich die Unsicherheit in den meisten Fällen auf eine probabilistische Beschreibung. Es ist jedoch zu berücksichtigen, dass ein probabilistischer Ansatz die Auswahl exakter Wahrscheinlichkeitsverteilungen erfordert und daher die Möglichkeiten, unvollständiges Wissen darzustellen, erheblich einschränken kann. Unsicherheiten in technischen Systemen werden im Allgemeinen als zufällig oder systematisch eingestuft. Leider weisen gängige Zustandsschätzverfahren Schwierigkeiten auf, mit beiden Unsicherheitsarten gleichzeitig umzugehen. Systematische Abweichungen sind mit Abweichungen in der Messausrüstung verbunden und können im Gegensatz zu zufälligen Fehlern nicht durch wiederholte Messungen ermittelt oder eliminiert werden, da jede Beobachtung möglicherweise durch die gleiche Abweichungen verfälscht wird (sogenannter Bias). In vielen ingenieurgeodätischen Anwendungen dienen beispielsweise Kalibrierverfahren seitens MSS dazu, systematischen Abweichungen entgegenzuwirken; sie können sie aber im Allgemeinen nur auf bestimmte Intervalle einschränken. Daher sind systematische Abweichungen prädestiniert, um durch ihre Zugehörigkeit zu *bounding sets* repräsentiert zu werden. Solche Anwendungen sind bekannt als UBB. Um eine gleichzeitige Berücksichtigung stochastischer und solcher mengenbasierter Unsicherheiten zu ermöglichen, wird durch Sun et al. (2019) **P#12** der EKF verallgemeinert. Dafür wird das zeit-diskrete Systemmodell aus Gl. 3.1 und Gl. 3.2 erweitert:

$$\mathbf{x}_{k+1} = f_k(\mathbf{x}_k, \mathbf{u}_k, \mathbf{w}_k, \mathbf{a}_k) \quad (3.14a)$$

$$\mathbf{y}_k = h_k(\mathbf{x}_k, \mathbf{v}_k, \mathbf{b}_k), \quad (3.14b)$$

wobei $\mathbf{w}_k \sim \mathcal{N}(\mathbf{0}, C_k^u)$ ein Gauß'sches Systemrauschen mit der VKM C_k^u , $\mathbf{a}_k \in \mathcal{E}(\mathbf{0}, S_k^u)$ ein Unknown But Bounded Uncertainty (UBB)-Systemstörvektor mit der Formmatrix S_k^u , $\mathbf{v}_k \sim \mathcal{N}(\mathbf{0}, C_k^z)$ ein Gaußsches Beobachtungsruschen mit der VKM C_k^z und $\mathbf{b}_k \in \mathcal{E}(\mathbf{0}, S_k^z)$ ein UBB-Beobachtungsstörvektor mit der Formmatrix S_k^z darstellt. Eine gleichzeitige Berücksichtigung von stochastischen und mengenbasierten Unsicherheiten erlaubt es, verschiedene Quellen von Schätzunsicherheiten flexibel modellieren zu können, von den individuellen Vorteilen zu profitieren und die Zuverlässigkeit oder aber zumindest realistische Beurteilung der Schätzergebnisse zu verbessern. Die in Sun et al. (2019) **P#12** entwickelte Filterungstechnik (sogenannte *Ellipsoidal and Gaussian Kalman Filter (EGKF)*) wird in Algorithmus 8 zusammengefasst. Auf eine ausführliche Herleitung des EGKF wird in dieser Arbeit verzichtet, auf Sun et al. (2019) wird verwiesen.

In Sun et al. (2019) wurde der entwickelte EGKF-Algorithmus 8 auf Basis simulierter Datensätze getestet und validiert und dem klassischen EKF-Algorithmus gegenübergestellt. Dabei wurden die RMSE-Werte für beide Algorithmen ermittelt und verglichen. Im ersten Simulationstest wurde das EGKF auf ein hochgradig nicht-lineares Benchmark-Problem in der nicht-linearen Schätztheorie angewendet (siehe z. B. Simon (2006)). Ein hoher Grad an Nichtlinearität sowohl in den Prozess- als auch in den Messgleichungen erschwert die Zustandsschätzung für dieses System. Zum Vergleich der Ergebnis wurden 100 Simulationen durchgeführt. In der Simulation wurde das Prozessrauschen mit $\mathbf{w}_k \sim \mathcal{N}(\mathbf{0}, 1)$, das Messrauschen mit $\mathbf{v}_k \sim \mathcal{N}(\mathbf{0}, 1)$, der UBB-Prozessstörvektor mit $\mathbf{w}_{i,k} \in \mathcal{E}(\mathbf{0}, 9)$ und $\mathbf{b}_k \in \mathcal{E}(\mathbf{0}, 4)$ gewählt. Der Gewichtungssparameter wurde mit $\eta = 0.5$ festgelegt. Für jedes der beiden Verfahren und für jede der 100 Iterationen wird zu jedem Zeitpunkt die Differenz zwischen dem aktuellen Zustand und der entsprechenden Schätzung berechnet. In den meisten Fällen führt der EGKF zu einem kleineren mittleren quadratischen Schätzfehler. Über alle 100 Simulationen liegt die durchschnittliche L_2 -Norm der EGKF-Schätzungen bei 148,70, während für den EKF die durchschnittliche L_2 -Norm viel höher bei 192,29 liegt. Gleiche Ergebnisse wurden auch bei der Verwendung der L_1 -Norm erzielt.

Im zweiten simulierten Testszenario wurde mit Hilfe von 1.000 MC Simulationen die Leistung vom EGKF und EKF in einem 2D-Trajektorien-schätzproblem aus Alkhatib (2015) (siehe auch Abschnitt 3.4) verglichen. Abbildung 3.9 zeigt den Vergleich des RMSE-Schätzfehlers des EGKF und

Algorithmus 8: Das Ellipsoidische und Gauß'sche Kalman Filter (EGKF)

Input: Nicht-lineares Systemmodell f_k , nicht-lineares Beobachtungsmodell h_k , Transitionsmatrix $F_{x,k}$, Designmatrix $H_{x,k}$, VKM des Prozessrauschens C_k^u , VKM des Messrauschens C_k^z , Formmatrix S_k^u , Formmatrix S_k^z , Gewichtungparameter: η , wobei $0 \leq \eta \leq 1$

Output: Gefilterter Zustandsvektor: $\hat{x}_k^{c+}$, seine VKM $C_k^+(\beta^*)$ und die a-posteriori Formmatrix $S_k^+(\beta^*)$

1 **Initialisierungsschritt:** c_0^+ , C_0^+ , S_0^+ .

2 **for** $k = 1, 2, \dots, K$ **do**

3 **Prädiktionsschritt:**

- Berechne VKM des prädierten Zustandsvektors (stochastischer Anteil):

$$C_k^- = F_{x,k-1} C_{k-1}^+ F_{x,k-1}^T + F_{w,k-1} C_{k-1}^u F_{w,k-1}^T.$$

- Berechne den Mittelpunkt des a-priori Ellipsoids: $\mathbf{c}_k^- = F_{x,k-1} \mathbf{c}_{k-1}^+ + \mathbf{\bar{u}}_{k-1}$
- Berechnen die Formmatrix des a-priori-Ellipsoids:

$$S_k^- = B \cdot \left(\frac{F_{x,k-1} S_{k-1}^+ F_{x,k-1}^T}{\sqrt{\text{tr}(F_{x,k-1} S_{k-1}^+ F_{x,k-1}^T)}} + \frac{F_{a,k-1} S_{k-1}^u F_{a,k-1}^T}{\sqrt{\text{tr}(F_{a,k-1} S_{k-1}^u F_{a,k-1}^T)}} \right).$$

$$\text{mit } B = \left(\sqrt{\text{tr}(F_{x,k-1} S_{k-1}^+ F_{x,k-1}^T)} + \sqrt{\text{tr}(F_{a,k-1} S_{k-1}^u F_{a,k-1}^T)} \right).$$

Filterungsschritt:

- Berechne Kalman-Gain-Matrix:

$$K_k = A^{-1} \cdot \left[(1 - \eta) C_k^- H_{x,k}^T + \eta \left(1 + \frac{1}{\beta} \right) S_k^- H_{x,k}^T \right].$$

mit

$$A = (1 - \eta) H_{x,k} C_k^- H_{x,k}^T + (1 - \eta) H_{v,k} C_k^z H_{v,k}^T + \eta \left(1 + \frac{1}{\beta} \right) H_{x,k} S_k^- H_{x,k}^T + \eta (1 + \beta) H_{b,k} S_k^z H_{b,k}^T.$$

- Berechne den Mittelpunkt der aktualisierten a-posterior-Schätzung $\hat{x}_k^+$:

$$\mathbf{c}_k^+ = (I - K_k H_{x,k}) \mathbf{c}_k^- + K_k (\mathbf{y}_k - \mathbf{\bar{z}}_k)$$

- Berechne die VKM des gefilterten Zustandsvektors:

$$C_k^+ = (I - K_k H_{x,k}) C_k^- (I - K_k H_{x,k})^T + F_{v,k} C_k^z F_{v,k}^T.$$

- Aktualisiere die Formmatrix:

$$S_k^+(\beta) = \left(1 + \frac{1}{\beta} \right) (I - K_k H_{x,k}) S_k^- (I - K_k H_{x,k})^T + (1 + \beta) K_k H_{b,k} S_k^z H_{b,k}^T K_k^T.$$

- Der optimale Parameter β^* kann gelöst werden durch $\beta^* = \arg \min_{\beta > 0} \{ (1 - \eta) \text{tr}[C_k^+(\beta)] + \eta \text{tr}[S_k^+(\beta)] \}$.

des EKF gemittelt über sämtliche 1.000 Monte-Carlo-Durchläufe sowie die geschätzte Trajektorie. Der Peak in der Abbildung 3.9 wird vermutlich durch die zufällige Unsicherheit verursacht, da sich der EKF in nicht-linearen Anwendungen gelegentlich entsprechend dieses Ergebnisses verhält. In fast allen Simulationen (nämlich in 99,7%) war das RMSE von EGKF deutlich geringer als beim EKF. Daraus lässt sich schließen, dass die neuen EGKF-Techniken eine genauere Schätzung für dieses nicht-lineare System liefern als der EKF.

Der EGKF-Algorithmus 8 wurde zudem in Sun et al. (2018) **P#13** auf einen Datensatz angewendet, der aus einem realen Experiment im Rahmen einer Georeferenzierung eines MSS gewonnen wurde. Das zugehörige Experiment wurde in Abschnitt 3.4 beschrieben. Der EGKF-Algorithmus 8 wurde verwendet, um die Positionen und Geschwindigkeiten dieser beiden Antennen zu schätzen. Die Ergebnisse wurden auch hier mit dem klassischen EKF-Algorithmus aus Paffenholz et al.

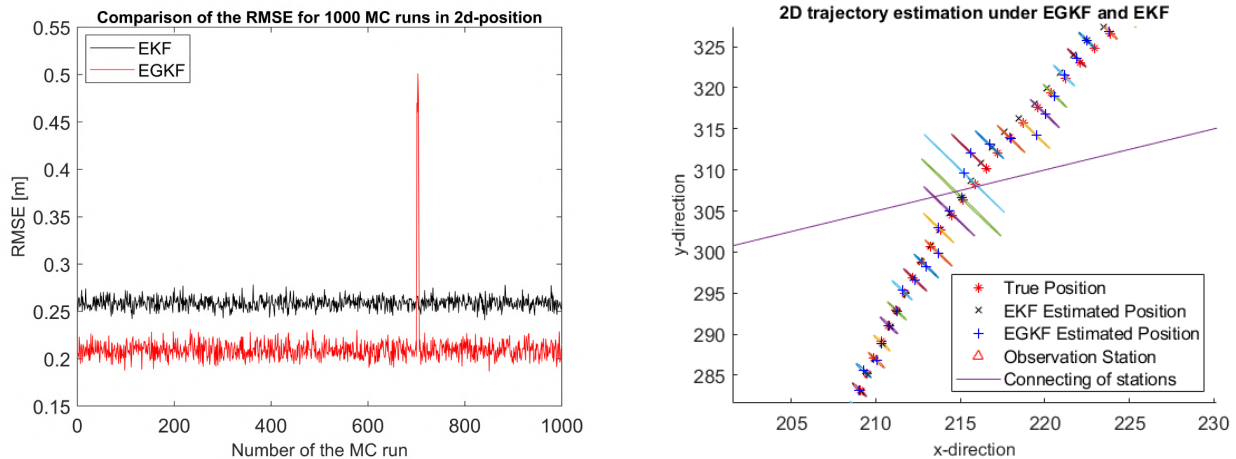


Abbildung 3.9: Links: RMSE-Vergleich von EGKF und EKF in 1000 Durchläufen; unten: mit EKF und EGKF geschätzte 2D Trajektorie, Originalabbildungen aus Sun et al. (2019).

(2010) verglichen. Die geschätzten Zustandsparameter für die 3D-Position unter der Verwendung des EGKF-Algorithmus sind in Abbildung 3.10 dargestellt.

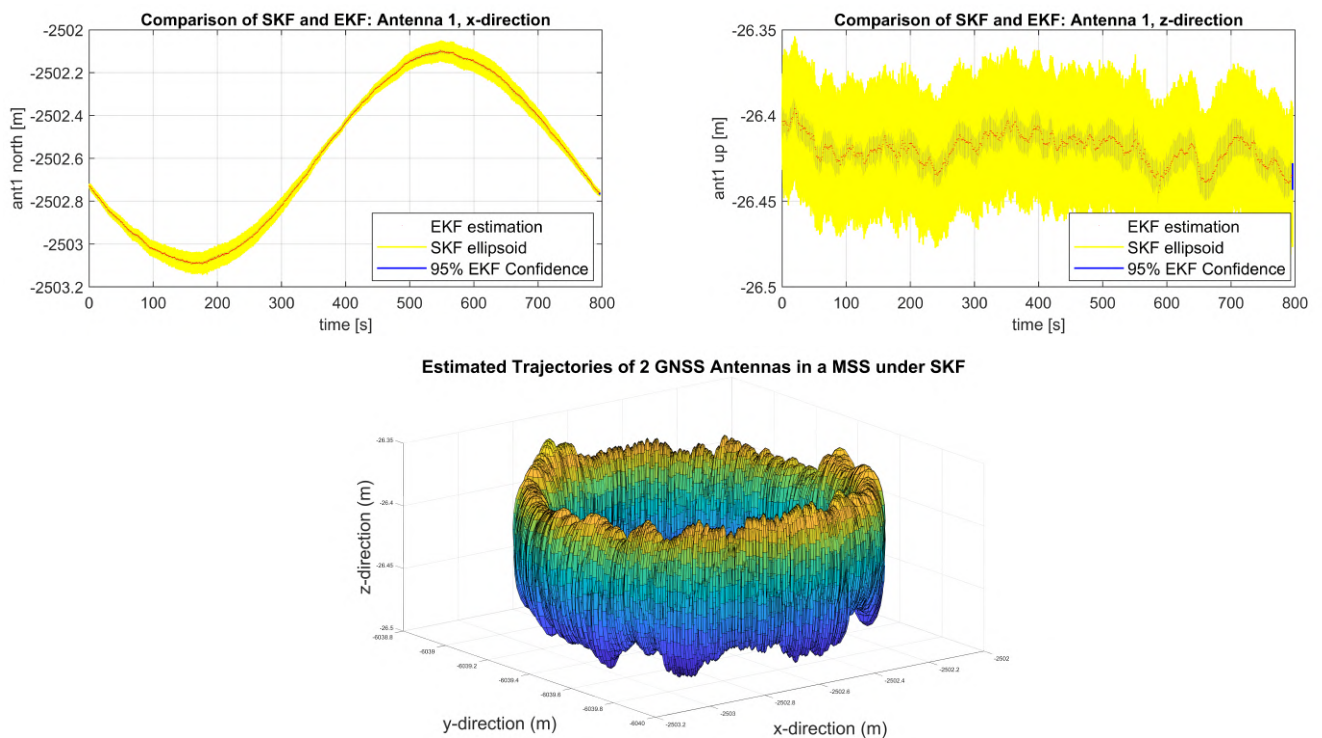


Abbildung 3.10: Gegenüberstellung der aus EKF- und EGKF geschätzten Positionen. Ca. 99,95% der EKF-Punkte befinden sich in den EGKF-Intervallen. Oben: Geschätzte Position in x- und z-Richtung. Unten: Geschätzte 3D-Trajektorien anhand des EGKF; Originalabbildungen aus Sun et al. (2018).

In jeder Sub-Abbildung nach 3.10 enthält der gelbe Teil sämtliche ausgegebenen eindimensionalen Ellipsoide (Intervalle). Die EKF-Schätzpunkte werden ebenfalls in der gleichen Abbildung dargestellt. Außerdem wurde der Prozentsatz der EKF-Punkte innerhalb der EGKF-Ellipsoide berechnet. Fast jeder EKF-Schätzpunkt liegt innerhalb eines geschätzten Ellipsoids nach der Verwendung des

EGKF-Algorithmus. Das blaue eindimensionale Ellipsoid (Intervall) kennzeichnet den 95% Konfidenzbereich für jeden EKF-Schätzpunkt, der im Vergleich zum geschätzten Set-Membership-Teil relativ klein ist. Das geschätzte Ellipsoid fällt in jeder Epoche größer aus als der 95% Konfidenzbereich des EKF. Die Geometrie der Ellipsoidreihe hängt stark von der Position der beiden Antennen ab. Darüber hinaus ist es offensichtlich, dass es gemäß EGKF größere Unsicherheiten in z -Richtung gibt, was mit dem Verlauf der 3D-Trajektorie (Abbildung 3.10 unten links) übereinstimmt. Der Hauptunterschied zum EKF besteht darin, dass die geschätzten Zustandsparameter Ellipsoide statt einzelne Punkte darstellen und jeder Punkt innerhalb eines Ellipsoids den gleichen Approximationsstatus hat. Aber man kann trotzdem eine Reihe von spezifischen Punkten in diesen Ellipsoiden wählen, wenn es notwendig ist.



Innovation kurz zusammengefasst | Das Kapitel befasste sich mit der Entwicklung von Filterverfahren für zuverlässige Lösungen für dynamische Systeme, in denen sowohl die Beobachtungen als auch das Systemwissen von zufälligen und unbekannten, aber begrenzten Unsicherheiten überlagert sind. Dafür wurde eine signifikante Weiterentwicklung des Ellipsoidisch-Gaußschen Kalmanfilter für zeit-diskrete, nicht-lineare Systeme durchgeführt. Hierbei werden die nicht-linearen Beobachtungs- und Systemgleichungen des Filters auf den unpräzisen Fall erweitert und mit maßgeschneiderten Methoden gelöst. Dies beinhaltet einerseits eine Erweiterung der verfügbaren Grundkonzepte um Elemente der garantierten Parameterschätzung und andererseits ein Filter zur Zusammenführung des Wissens aus der Beobachtung und über das System.

3.6 Zusammenfassung des Kapitels

Das Kapitel 3 gibt einen Überblick über entwickelte Filterungstechniken zur zeit-diskreten Zustandsschätzung von nicht-linearen Systemen. Die erste Gruppe, das IEKF, operiert im Sinne der Gaußschen Approximation für die a-posteriori Dichte $p(\mathbf{x}_k|\mathbf{y}_k)$. Dieser Filterungsansatz ist flexibel und kann verschiedene Arten von Beobachtungsgleichungen, sowohl explizite als auch implizite Formulierungen des funktionalen Zusammenhangs zwischen Zuständen und Beobachtungen, verwenden. Außerdem können Gleichungs- und Ungleichungsrestriktionen eingeführt werden. Diese Methode wurde zur Georeferenzierung eines Laserscanner-basierten kinematischen MSS in einer simulierten und realen Umgebung zur Steigerung der Integrität erfolgreich angewendet und evaluiert. Die zweite Gruppe, das EKPF, stellt eine Kombination von EKF und PF dar. Das PF führt hierbei sequentielle MC-Schätzungen basierend auf der Partikelrepräsentation von Wahrscheinlichkeitsdichten durch; die Kombination mit dem EKF verringert die Anzahl der Partikel signifikant. Sowohl der Implementierungs- als auch der Berechnungsaufwand dieses Filterungsansatzes sind gering. Umfangreiche Experimente mit der Georeferenzierung von MSS als Testanwendung zeigen, dass der EKPF-Ansatz eine erhebliche Verbesserung des klassischen EKF darstellt. Als dritte Gruppe wurde eine neue Filterungstechnik für Situationen vorgeschlagen (EGKF), in denen die Wahrscheinlichkeitsverteilungen für einige Teile der Unsicherheit bekannt sind, während bei anderen Teilen die Unsicherheit (UBB-Unsicherheit) nur über obere und untere Grenzen modelliert werden kann. Sämtliche entwickelten Algorithmen wurden anhand von simulierten und Realdatensätzen zur Georeferenzierung von MSS getestet und validiert. Hauptaugenmerk dieses Filters liegt auf der Fortpflanzung von Unsicherheiten ohne Überschätzung, welche bei der Anwendung intervallmathematischer Methoden oft vorkommt.

4 Simulationsbasierte Methoden für die Inferenzstatistik

4.1 Einführung in Monte-Carlo- und Bootstrapping-Methoden

Methoden der Inferenzstatistik (induktive Statistik oder schließende Statistik) werden verwendet, um Rückschlüsse auf eine Population zu ziehen und die Zuverlässigkeit der Rückschlüsse anhand von Informationen aus einer Stichprobe zu beurteilen. Inferenzstatistik beinhaltet Techniken wie Parameterschätzung mittels Punktschätzungen (siehe Kap. 2), die Schätzung von Konfidenzbereichen, Hypothesentests und Modellierungen (z. B. Regressionsmodelle). Um die Zuverlässigkeit der getroffenen Schlussfolgerungen zu bewerten, muss die Verteilung der in der Analyse verwendeten Statistiken bekannt sein. In Situationen, in denen eine leicht verständliche Statistik verwendet wird, wie z. B. der Mittelwert einer Stichprobe mit normalverteilten Messabweichungen, können die statistischen Analysen geschlossen durchgeführt werden. In vielen Anwendungen ist es jedoch nicht sinnvoll, sich auf die Verwendung einfacher Statistiken oder die Vereinfachung von Annahmen zu beschränken. In diesen Fällen können simulationsbasierte Techniken, wie MC-Simulation und Bootstrapping, herangezogen werden. In den letzten Kapiteln wurde gezeigt, dass eine MC-Simulation eine einfache und effiziente Möglichkeit ist, die Phänomene von Interesse nachzuvollziehen (Gentle, 2003; Kroese et al., 2011; Casella und Robert, 2002; Voss, 2014). Um eine MC-Simulation durchzuführen, wird ein Modell des betrachteten Phänomens und eine Methode zur Generierung von Zufallszahlen (entsprechend dem Modell) mittels Computer benötigt. Die simulierten Daten (Stichproben), die aus dem Modell generiert werden, können dann als Beobachtungen behandelt werden. Es können sodann verschiedene Statistiken und Momente auf der Grundlage der simulierten Daten (Mittelwert, Median, Modus, Varianz, Schiefe, etc.) verwendet werden, um Erkenntnisse über die Population zu gewinnen.

Ein weiterer bekannter, simulationsbasierter Ansatz in der rechnergestützten Statistik basiert auf Bootstrapping. Dieser rechenintensive, nicht formale Ansatz wurde erstmalig von Efron (1979) eingeführt und erlaubt es, eine Vielzahl von Inferenz-Schätzungen, einschließlich Punkt-, Unsicherheit- und Konfidenzregionschätzung, durchzuführen. Bootstrapping ermöglicht die Simulation einer Verteilung einer bestimmten Statistik zu simulieren (Givens und Hoeting, 2013). Die Idee besteht darin, die beobachteten Daten wiederholt neu zu generieren (*resampling*), wobei in jedem Schritt eine empirische Verteilungsfunktion aus den neu generierten Daten erzeugt wird. Für jeden neu generierten Datensatz - oder gleichwertig für jede empirische Verteilungsfunktion - kann ein neuer Wert der Statistik berechnet werden. Infolgedessen liefert die Zusammenstellung dieser Werte eine Schätzung der Stichprobenverteilung der Statistik. Weiterführende Informationen über die zugrunde liegende Theorie der Bootstrapping-Techniken findet man auch in Efron und Tibshirani (1993), Efron und Hastie (2016) sowie Givens und Hoeting (2013).

Die Stichprobenverteilung ist für viele Statistiken bekannt, wird jedoch typischerweise unter Verwendung von Annahmen über die zu untersuchende Grundgesamtheit oder für große Stichprobengrößen abgeleitet. In vielen Fällen sind die Stichprobenverteilungen für die Statistik nicht bekannt oder die Annahmen sind nicht eindeutig erfüllt. Diese Probleme lassen sich mit Hilfe von Bootstrap- und MC-Simulation lösen, was in diesem Kapitel behandelt wird. Einige der populärsten Anwendungen der MC-Simulation und Bootstrapping-Techniken sind (Martinez und Martinez, 2015):

- Durchführung von Hypothesentests, wenn die Verteilung der Teststatistik analytisch nicht bekannt ist,

- Bestimmung von Qualitätsmaßen bei Hypothesentests, wie z. B. der Testgüte
- Statistischer Vergleich der Qualität von verschiedenen Schätzern (siehe Kapitel 2) hinsichtlich Verzerrung und statistischen Momenten.

4.2 Vorstellung der simulationsbasierten Methoden für die Inferenzstatistik

Ziel dieses Abschnitts ist die Erläuterung, wie simulationsbasierte Techniken in Form von MC-Simulation und Bootstrapping verwendet werden können, um Inferenzen zu ziehen (Bestimmung von Konfidenzintervallen und Durchführung von Hypothesentests um daraus Schlussfolgerung zu ziehen), wenn die traditionellen oder analytischen Methoden versagen.

In Abschnitt 4.3 werden Situationen in der Inferenzstatistik behandelt, in denen die Verteilung der Grundgesamtheit bekannt ist oder Annahmen über die Verteilung der Grundgesamtheit getroffen werden können. Zu diesem Zweck werden zwei verschiedene Anwendungen demonstriert. Der Fokus liegt auf der Bestimmung numerischer Momente (Mittelwert, Standardabweichung, empirische Momente höhere Ordnung) und der Bereichsschätzung (Konfidenzintervalle).

- In der ersten Anwendung wird eine grundlegende Methode zur Bestimmung von Messunsicherheit nach dem Guide to the Expression of Uncertainty in Measurement (GUM) (ISO, 1995) von nicht-linearen Funktionen behandelt. Bei der Bestimmung der Messunsicherheit und ihrer Konfidenzintervalle werden MC-Simulationen verwendet. Diese werden gewonnen, indem wiederholt Stichproben aus derselben Population oder Grundverteilung gezogen werden, d. h., die Verteilung der Statistik wird durch die Generierung von zufälligen Stichproben aus festgelegten PDFs der Eingangsgrößen geschätzt. Die generierten Stichproben werden verwendet, um die PDF der Ausgangsgrößen abzuschätzen. Die simulationsbasierte Bestimmung von Messunsicherheit nach GUM wurde in Alkhatib et al. (2009) **P#14** und Alkhatib und Kutterer (2013) **P#15** vorgestellt.
- In der zweiten Anwendung wird demonstriert, wie anhand von Bootstrapping-Techniken die Kovarianzmatrix von geschätzten Parametern in einem GEM-Algorithmus (siehe Algorithmus 3 in Abschnitt 2.4) als Qualitätsparameter abgeleitet wird. Dieser Bootstrap-Algorithmus wurde in Alkhatib et al. (2018b) **P#16** vorgestellt.

In Abschnitt 4.4 wird der Fokus auf die Durchführung von statistischen Hypothesentests anhand simulationsbasierter Algorithmen gelegt. Exemplarisch wird der Ansatz anhand von zwei unterschiedlichen Testproblemen vorgestellt:

- Die Motivation des ersten Testproblems, welches in Zhao et al. (2018) **P#17** publiziert wurde, liegt in der Modellauswahl einer sparsamen, aber ausreichend genauen, parametrischen Beschreibung eines Objekts auf der Grundlage von TLS-Messungen anhand einer simulationsbasierten Version des Hypothesentests nach Cox (1961).
- Das zweite Testproblem nach Alkhatib et al. (2019) **P#18** besteht darin, zu überprüfen, ob die zufälligen Abweichungen einer beobachteten Regressionszeitreihe mit unbekannten Koeffizienten durch einen kovarianzstationären AR-Prozess beschrieben werden können oder ob ein AR-Prozess mit zeitvariablen Koeffizienten ausgewählt werden sollte.

4.3 Simulationsbasierte-Techniken zur Bestimmung von statistischen Momenten und Konfidenzintervallen

4.3.1 BEREICHSSCHÄTZUNG I: Monte Carlo-Methoden zur GUM Unsicherheitsmodellierung

Einführung zur GUM Unsicherheitsmodellierung

Das Problem der Bestimmung der Messunsicherheit von Messgrößen und Messprozessen tritt in zahlreichen Forschungsrichtungen auf, einschließlich, aber nicht beschränkt auf die Bereiche der Geodäsie und Geoinformation. Der GUM ist eine Standardreferenz für die Unsicherheitsmodellierung in den Ingenieur- und Mathematikwissenschaften sowie der Physik (vgl. ISO, 2008; JCGM, 2008). GUM gruppiert die auftretenden unsicheren Größen in “Typ A” und “Typ B”: Die Unsicherheiten vom “Typ A” werden mit den klassischen statistischen Methoden ermittelt, während Unsicherheiten von “Typ B” durch Ermittlungsverfahren gewonnen werden, die z. B. durch Erfahrung und Wissen über ein Messinstrument oder einen Messprozess entstehen. Während die Unsicherheiten der Messgrößen vom “Typ A” durch wiederholte Messung der interessierenden Messgröße geschätzt werden können, basieren die Unsicherheiten der Messgrößen vom “Typ B” damit i. d. R. auf Expertenwissen. Beide Typen können zufällige und systematische Unsicherheitskomponenten aufweisen:

- Zufällige Abweichungen ϵ ergeben sich aus nicht vorhersehbaren Abweichungen einzelner Einflussfaktoren unter scheinbar gleichen tatsächlichen Bedingungen (nicht reproduzierbare Effekte), siehe z. B. Bandemer (2006) und Grabe (2005).
- Systematische Abweichungen δ sind auf nicht kontrollierbare Effekte während des Messprozesses und der Vorverarbeitungsschritte der Messung zurückzuführen. Obwohl systematische Abweichungen unbekannt sind, verzerren sie das Messergebnis in eine bestimmte Richtung (reproduzierbare, aber unbekannte Effekte), siehe z. B. Bandemer (2006) und Grabe (2005).

GUM definiert die Ausgangsgrößen $\mathbf{Y}$ als Funktion der Eingangsgrößen $\mathbf{Z}$. Die Eingangsgrößen können als Einflussparameter betrachtet werden, die beispielsweise in Vorverarbeitungsschritten relevant sein können:

$$\mathbf{Y} = \mathbf{f}(Z_1, \dots, Z_n) = \mathbf{f}(\mathbf{Z}), \quad (4.1)$$

wobei $\mathbf{f}(\cdot)$ das lineare oder nicht-lineare Beobachtungsmodell und n die Anzahl der Eingangsgrößen ist, deren Werte und Unsicherheiten in der aktuellen Messung direkt bestimmt werden (Originalmessungen) oder von externen Quellen in die Messung eingebracht werden, z. B. durch die Kalibrierung eines Instruments (ISO, 2008). Die Größe Z_i kann sowohl zufällige als auch systematische Abweichungen enthalten. GUM schlägt vor, zufällige und systematische Abweichungen in einem stochastischen Rahmen zu behandeln. Die unsicheren Einflussfaktoren im Beobachtungsmodell 4.1 werden hier in drei Gruppen unterteilt: Zusatzinformationen, Sensorparameter und Modellkonstanten. Während die Unsicherheit der ursprünglichen Messung in der Regel von “Typ A” ist, kann die Unsicherheit der Einflussfaktoren von “Typ A” oder “Typ B” sein. Die Einführung von systematischen Abweichungen der Eingangsgrößen ist aus vielen Gründen sinnvoll: Beispielsweise sind Modellkonstanten für eine spezifische Situation nur teilweise repräsentativ (wie die Modellkonstanten für den Refraktionsindex bei Distanzmessungen) oder Messergebnisse werden durch Rundungsfehler beeinflusst.

Der GUM (ISO, 2008) beschreibt einen Ansatz zur Bestimmung der Standardunsicherheit ausgedrückt als VKM $\Sigma_{\mathbf{yy}}$ des Messergebnisses $\mathbf{y}$ ausgehend von der VKM $\Sigma_{\mathbf{zz}}$ der Eingangsgrößen unter Verwendung des Varianz-Fortpflanzungsgesetzes (vgl. Kutterer und Schön, 2004):

$$\Sigma_{\mathbf{yy}} = \mathbf{A}\Sigma_{\mathbf{zz}}\mathbf{A}^T, \quad (4.2)$$

wobei die Jakobi-Matrix $\mathbf{A}$ die partiellen Ableitungen von $\mathbf{Y} = \mathbf{f}(\mathbf{Z})$ in Bezug auf $Z_1, \dots, Z_n$ enthält, d. h.

$$\mathbf{A} = \begin{bmatrix} \frac{\partial f_1}{\partial Z_1} & \cdots & \frac{\partial f_1}{\partial Z_n} \\ \vdots & \cdots & \vdots \\ \frac{\partial f_m}{\partial Z_1} & \cdots & \frac{\partial f_m}{\partial Z_n} \end{bmatrix} \quad (4.3)$$

Sind die PDF der Messeingangsgrößen bekannt, können MC-Simulationen anstelle des oben erwähnten Varianz-Fortpflanzungsgesetzes zur Berechnung der kombinierten Unsicherheiten der Messungen verwendet werden (Hennes, 2007). Die *Erweiterung von GUM* (ISO, 2007) empfiehlt daher die Fortpflanzung von Unsicherheiten durch einen probabilistischen Ansatz. Innerhalb des genannten Ansatzes wird die Fortpflanzung von Unsicherheiten numerisch mit MC-Techniken durchgeführt. Die zufällig generierten Stichproben für die Eingangsgrößen führen zu den zufälligen Abweichungen der Messungen, mit denen die Varianzen und Kovarianzen der Messungen geschätzt werden (Koch, 2000). Die Methode wird von Siebert und Sommer (2004) als Erweiterung von GUM beschrieben und unter anderem zur Überprüfung der Unsicherheitsevaluierung durch Acko und Godina (2005) eingesetzt. Der Unterschied zwischen dem klassischen GUM (ISO, 2008) und der Erweiterung von GUM (ISO, 2007) unterscheidet sich im Falle von Nichtlinearität und/oder nicht-normalverteilten Eingangsgrößen nicht signifikant im ersten und zweiten zentralen Moment, sondern in der Schätzung des Konfidenzbereichs, der sich in der nicht-Gaußschen PDF der Ausgangsgrößen widerspiegelt.

Algorithmus zur Unsicherheitsschätzung mit MC Methoden

Die im Folgenden vorgestellte Unsicherheitsmodellierung zur Bestimmung von Qualitätsparametern nutzt die MC-Simulation zur flexiblen Modellierung und Fortpflanzung von Messunsicherheiten. Es wird davon ausgegangen, dass das Beobachtungsmodell in Gl. 4.1 vollständig formuliert ist, indem die Ausgangsgrößen mit den Eingangsgrößen in Zusammenhang gesetzt werden. Es wird außerdem angenommen, dass die PDFs der betrachteten Eingangsgrößen a-priori bekannt sind. Anschließend kann ein Stichprobenvektor der Eingangsgrößen mit Hilfe eines Zufallsgenerators mehrfach gezogen werden. Für jeden Vektor der Eingangsstichproben werden die entsprechenden Werte der Ausgangsgrößen unter Verwendung der funktionalen Beziehung berechnet. Der Satz der resultierenden Ausgangsstichproben ergibt eine empirische Verteilung, mit der die korrekte PDF der Ausgangsgrößen approximiert werden kann. Sämtliche erforderlichen Kennzahlen (Erwartungswerte, Varianzen und Kovarianzen) sowie durch zentrale Momente höherer Ordnung gebildete Größen, wie Schiefe und Kurtosis, können dann abgeleitet werden. Die Berechnungsschritte der Unsicherheitsevaluierung einer univariaten Ausgangsgröße unter Verwendung der MC-Simulation wird im Algorithmus 9 zusammengefasst.

Anwendung zu GUM-MC

Je nach Art der Unsicherheit wird ihre Übertragung von den Beobachtungen auf die Zielgrößen mit individuellen mathematischen Methoden gehandhabt. Diese beinhalten im Falle von Beobachtungsintervallen die Lösung einer Wertebereichsaufgabe, im Falle von unscharfen bzw. impräzisen Größen die Lösung einer Optimierungsaufgabe auf Basis des Erweiterungsprinzips nach Zadeh, und im Falle von Wahrscheinlichkeitsdichten in der Regel eine Monte-Carlo-Simulation (bei Vorliegen von Vorinformation auf Basis des Bayes-Theorems). Die letzten beiden Fälle wurden in Alkhatib et al. (2009) **P#14** kombiniert betrachtet. In diesem sogenannten Fuzzy-Random-Ansatz werden die zufälligen Komponenten in einem stochastischen Ansatz modelliert, während die deterministischen Unsicherheiten mit Hilfe eines Wertebereichs eines Suchproblems behandelt werden. In der vorliegenden Arbeit wird die Fuzzy-Komponente nicht näher erläutert und nur auf den reinen MC-Ansatz weiter eingegangen.

Algorithmus 9: GUM-Unsicherheitsevaluierung unter Verwendung der MC-Simulation für eine univariate Ausgangsgröße

Input : A-priori bekannte PDFs der Eingangsgrößen $Z_1, Z_2, \dots, Z_n$, das funktionale Modell $f(\mathbf{Z})$, Anzahl M der MC-Wiederholungen, Sicherheitsswahrscheinlichkeit γ

Output: Schätzwerte der statistischen Größen $\hat{E}(Y), \hat{u}(Y), \hat{s}k(Y), \hat{k}u(Y)$ und des Konfidenzintervalls $[\underline{y} = y_j, \bar{y} = y_k]$

1 **for** $j = 1 \dots M$ **do**

2 **Schritt 1:** Generiere einen Zufallszahlensatz aus den a-priori bekannten PDFs der Eingangsgrößen $z_1^{(i)}, z_2^{(i)}, \dots, z_n^{(i)}$

3 **Schritt 2:** Berechne die Ausgangsgrößen y mit Hilfe des funktionalen Modell:
 $y^{(i)} = f(z_1^{(i)}, z_2^{(i)}, \dots, z_n^{(i)}) = f(\mathbf{z}^{(i)}),$

4 **Schritt 3:** Ermittle alle relevanten Schätzwerte der statistischen Größen:

- Erwartungswert der Ausgangsgröße: $\hat{E}(f(\mathbf{z})) = \hat{E}(y) = \frac{1}{M} \sum_{i=1}^M f(\mathbf{z}^{(i)}),$
- Unsicherheit der Ausgangsgröße: $\hat{u}(Y) = \frac{1}{M} \sum_{i=1}^M \left(f(\mathbf{z}^{(i)}) - \hat{E}(f(\mathbf{z})) \right)^2$
- Schiefe der Ausgangsgröße $\hat{s}k(Y) = \frac{1}{(M-1)s^3} \sum_{i=1}^M \left(f(\mathbf{z}^{(i)}) - \hat{E}(f(\mathbf{z})) \right)^3$
- Kurtosis der Ausgangsgröße: $\hat{k}u(y) = \frac{1}{(M-1)s^4} \sum_{i=1}^M \left(f(\mathbf{z}^{(i)}) - \hat{E}(f(\mathbf{z})) \right)^4$

Schritt 4: Berechne das Konfidenzintervall. Ordne zuerst die generierten Ausgangsgrößen $y^{(i)}$ von der Kleinsten bis zur Größten an. Ein numerisch berechnetes $100(1 - 2\gamma)\%$

Konfidenzintervall für die Zufallsvariable Y ist:

$y_{\text{conf,MC}} = [\underline{y} = y_j, \bar{y} = y_k]$, wobei $j = (M + 1)\gamma$ und $k = (M + 1)(1 - \gamma)$ (j und k werden auf ganze Zahlen gerundet)

In Alkhatib und Kutterer (2013) **P#15** werden MC-Techniken betrachtet und eine Erweiterung der Eigenschaften der abgeleiteten Zeitreihen und deren Validierung mit realen k-TLS Beobachtungsdaten diskutiert. Exemplarisch werden die Unsicherheiten von 2D-TLS-Profilmessungen, dem typischen Messmodus beim Einsatz in k-TLS-basierten MSS, simuliert und auf die Zielgröße übertragen. Anschließend werden die aus der Simulation gewonnenen Erkenntnisse der Auswertung mit einem realen Datensatz kritisch gegenübergestellt. Die Modellierung der Unsicherheiten und deren Fortpflanzung auf die Messergebnisse sind für jede signifikante Einflussgröße des MSS vorzunehmen. Alkhatib und Kutterer (2013) **P#15** stellen Strategien für die Untersuchung von 2D-TLS-Profilmessungen (Vertikalprofile mit einer Wiederholfrequenz von 12,5 Profilen/Sekunde) eines TLS vor. Der reale Datensatz ist mit dem Ziel der Detektion von vertikalen Deformationen einer Autobahnbrücke im Rahmen von kontrollierten Belastungsszenarien mit einem terrestrischen Laserscanner vom Typ Z+F Imager 5006 erfasst worden. Hierfür wurden 2D-Profilmessungen mit einer Wiederholungsrate von 12,5 Profilen pro Sekunde durchgeführt, woraus 7.216 Punkte pro Epoche für einen definierten Abschnitt resultierten. 500 Epochen (2D-Profile) repräsentieren den unbelasteten Zustand der Autobahnbrücke. Die Zielgrößen Y_i sind die jeweiligen z-Koordinaten als Funktion der gemessenen Winkel und Schrägstrecken. In Alkhatib und Kutterer (2013) **P#15** wird das Simulationsverfahren allgemein eingeführt und an einem exemplarischen, realen Datensatz von 2D-Profilmessungen an einer Autobahnbrücke in Süddeutschland evaluiert. Der funktionale Zusammenhang ist für einen horizontalen Laserscanner durch eine klassische polare Punktbestimmung über die originären Messgrößen Horizontalrichtung, Zenitwinkel und Schrägstrecke gegeben. Zu Beginn werden für die MC-Simulation verschiedene Verteilungen, hier aus der Gruppe der Dreiecks-, Rechteck- und Normalverteilung, für die zu untersuchenden Einflussgrößen aus Erfahrungen, Untersuchungen und Annahmen festgelegt. Folgende Einflussgrößen mit ihren Verteilungsfunktionen

werden berücksichtigt:

$$Y_i = f(Z_1, Z_2, Z_3, Z_4), \quad (4.4)$$

wobei Z_1 der konstante Anteil der Unsicherheit der Distanz (Typ A, Normalverteilung), Z_2 der distanzabhängige Anteil (ppm-Wert) der Unsicherheit der Distanzmessung (Typ B, Normalverteilung), Z_3 die Unsicherheit des Zenitwinkels (Typ A, Normalverteilung) und Z_4 die vertikale Auflösung des Zenitwinkels, gegeben durch die Schrittweite des Motors (Typ B, Rechteckverteilung), darstellen. Es wird eine große Anzahl M (mindestens 100.000) von Stichproben der Einflussgrößen aus den PDFs $Z_1 \sim \mathcal{N}(0; 0,5 \text{ mm})$, $Z_2 \sim \mathcal{N}(0; 30 \text{ ppm})$, $Z_3 \sim \mathcal{N}(0; 10 \text{ mgon})$ und $Z_4 \sim \mathcal{U}(0; 20 \text{ mgon})$ generiert und mit dem funktionalen Zusammenhang der 2D-Profilmessung in die Zielgröße Y_i , transformiert. Die ausgewählten Parameter der genannten PDFs der Einflussgrößen werden weitestgehend dem Datenblatt des Laserscanners Z+F Imager 5006 entnommen. Das Ergebnis der Simulation sind PDFs der generierten, transformierten Stichproben, aus denen sich statistische Momente höherer Ordnungen numerisch berechnen lassen. Neben dem Mittelwert und der Standardabweichung werden die Schiefe und Kurtosis berechnet. Die Schiefe gibt Auskunft über die Symmetrie der Verteilung, wohingegen die Kurtosis angibt, ob die Verteilung im Gegensatz zur Normalverteilung spitz oder abgeflacht ist. Bei normalverteilten Daten nimmt die Schiefe den Wert 0 an und die Kurtosis den Wert 3. Je weiter die Schiefe der Datenverteilung von 0 und deren Kurtosis von 3 abweicht, umso größer ist die Abweichung der Daten von einer Normalverteilung. Anschließend können die simulierten Unsicherheiten der Zielgrößen mit den abgeleiteten statistischen Momenten aus realen Messungen verglichen werden. Für die empirische Schätzung der statistischen Momente der Zielgrößen in der Vorwärtsmodellierung wurden Simulationen mit den oben genannten Einflussgrößen Z_1 bis Z_4 berechnet. Unter dem Begriff Vorwärtsmodellierung wird die a-priori-Verfügbarkeit von Qualitätsparametern bezeichnet, die dann auf die entsprechenden Zielwerte propagiert werden (Paffenholz et al., 2017). Für weitere Simulationsvarianten, die sich durch eine Variation der Einflussgrößen ergeben, wird auf Alkhatib und Kutterer (2013) **P#15** verwiesen.

In Abbildung 4.1 sind die Ergebnisse der Simulationen für die Einflussgrößen Z_1 bis Z_4 gezeigt, welche sich durch die Berechnung der statistischen Momente aus dem realen Datensatz (Abbildung 4.2) validieren und bestätigen lassen. Im Hinblick auf die in Abbildung 4.1 und Abbildung 4.2 dargestellten Standardabweichungen wird der Effekt der Proportionalität der Distanz auf die Standardabweichungen der repräsentativen Profilpunkte deutlich sichtbar. Weiterhin ist das Wurzel-n-Gesetz für die Standardabweichung des Mittelwerts $s_{\bar{z}}$ in Bezug auf die Standardabweichung der einzelnen Werte ersichtlich. Abbildung 4.1 zeigt, dass sich die aus den Simulationen berechnete Schiefe der Datenverteilung bei sämtlichen berechneten Klassen (1, 5 und 10 Punkte pro Klasse) nicht signifikant vom Wert 0 unterscheiden, während die Abweichung der Kurtosis vom Wert 3 signifikant ist. Diese aus den Simulationen gewonnenen Ergebnisse werden durch die Auswertung des realen Datensatzes (Abbildung 4.2) bestätigt.

4.3.2 BEREICHSSCHÄTZUNG II: Bootstrapping zur Schätzung des Konfidenzbereichs für den EM-Algorithmus

Im Abschnitt 2.4 (Kargoll et al., 2018b; Alkhatib et al., 2017b, **P#3** und **P#4**) wurde ein GEM-Algorithmus (Algorithmus 3) vorgestellt, welcher die Schätzung der deterministischen Modellparameter, der AR-Koeffizienten, der Skalenparameter und der Freiheitsgrade der zugrunde liegenden t-Verteilungen ermöglicht. Dabei sind oft nicht nur die geschätzten Parameter von Interesse, sondern auch deren VKM. Die VKM, die im Folgenden mit $\Sigma\{\hat{\theta}\}$ bezeichnet wird, charakterisiert die Qualität der geschätzten Parameter. Da der Schätzer $\hat{\theta}$ eine komplexe nicht-lineare Funktion von zufälligen Abweichungen ist, kann die Fortpflanzung von Varianz-Kovarianz-Informationen nicht direkt im Schätzprozess durchgeführt werden. Bootstrapping, in Kombination mit MC-Techniken, kann in dieser Situation jedoch problemlos angewendet werden. In Alkhatib et al. (2018a) **P#16**

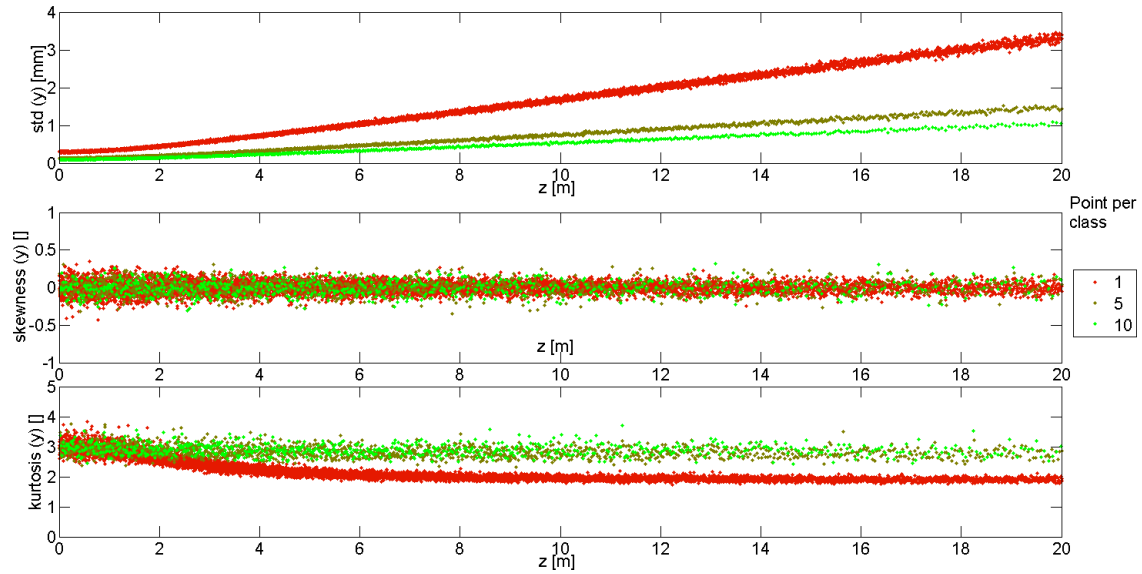


Abbildung 4.1: Statistische Momente zweiter bis vierter Ordnung der Zielgrößen (hier z -Koordinaten der 2D-Profilmessungen) aus Monte-Carlo-Simulationen für die Einflussgrößen Z_1 bis Z_4 aus dem realen Datensatz. Für die Standardabweichung (oben), die Schiefe (Mitte) und die Kurtosis (unten) werden jeweils die Ergebnisse für drei Klassen (Aggregation von 1, 5 und 10 Punkten pro Klasse) der 2D-Profilpunkte gezeigt (Alkhatib und Kutterer, 2013).

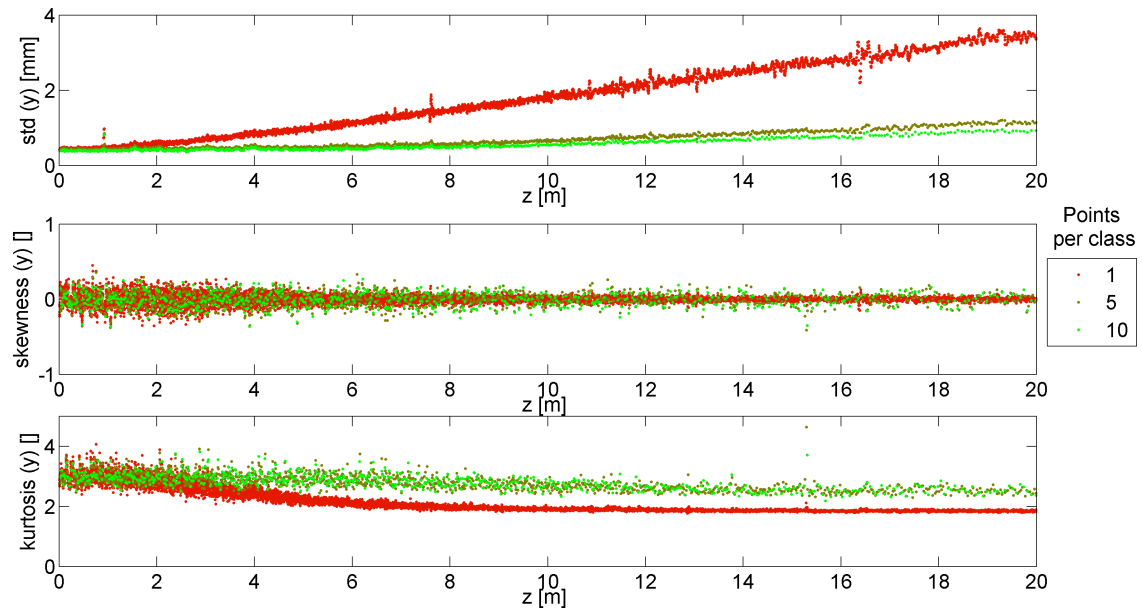


Abbildung 4.2: Statistische Momente der Zielgröße (hier z -Koordinaten der 2D-TLS-Profilmessungen) aus dem realen Datensatz für die Einflussgrößen Z_1 bis Z_4 . Für die Standardabweichung (oben), die Schiefe (Mitte) und die Kurtosis (unten) werden jeweils die Ergebnisse für drei Klassen (Aggregation von 1, 5 und 10 Punkten pro Klasse) der 2D-Profilpunkte gezeigt (Alkhatib und Kutterer, 2013).

wurde ein simulationsbasierter Ansatz zur numerischen Approximation der VKM und der entsprechenden Konfidenzintervalle demonstriert. Der Berechnungsprozess zur Ableitung der VKM der geschätzten Parameter $\Sigma\{\hat{\theta}\}$ ist in Algorithmus 10 zusammengefasst.

Algorithmus 10: Bootstrapping Algorithmus zur Schätzung der VKM der geschätzten Parameter und zur Berechnung der Konfidenzintervalle

Input : Funktionale Modellparameter $\hat{\xi}$, Skalenparameter $\hat{\sigma}_k^2$, Freiheitsgrad der t-PDF $\hat{\nu}_k$, geschätzte farbige Rauschkomponenten $\hat{\mathbf{e}}_k$, geschätzte weiße Rauschkomponenten $\hat{\mathbf{u}}_k$, Gewichtsmatrix $\hat{\mathbf{W}}_k$, AR Koeffizienten $\hat{\alpha}_k$, Sicherheitswahrscheinlichkeit γ

Output: $\Sigma\{\hat{\theta}\}$

1 Für jede Komponente $k = 1, \dots, N$, jede Zeitepoche $t = 1, \dots, n$ und jede Bootstrapping-Stichprobe $b = 1, \dots, B$,

- Generiere Samples des weißen Rauschens unter Verwendung der geschätzten t-Verteilungen:
 $u_{k,1}^{(b)}, \dots, u_{k,n}^{(b)} \stackrel{\text{ind.}}{\sim} t_{\hat{\nu}_k}(0, \hat{\sigma}_k^2).$
- Berechne die Stichproben des farbigen Rauschens unter Verwendung der geschätzten AR-Modelle:
 $e_{k,t}^{(b)} = \hat{\alpha}_{k,1}e_{k,t-1}^{(b)} + \dots + \hat{\alpha}_{k,p_k}e_{k,t-p_k}^{(b)} + u_{k,t}^{(b)}.$
- Berechne die Pseudo-Beobachtungsstichproben mit den geschätzten deterministischen Funktionen: $\ell_{k,t}^{(b)} = h_{k,t}(\hat{\xi}) + e_{k,t}^{(b)}.$
- Verwendung des Algorithmus 3, um die Bootstrap-Lösungen zu berechnen:
 $\hat{\theta}^{(1)}, \dots, \hat{\theta}^{(b)}.$

Berechnung der Mittelwerte der Bootstrap-Lösung:

$$\bar{\theta} = \frac{1}{B} \sum_{b=1}^B \hat{\theta}^{(b)}.$$

- Berechne die Bootstrap-VKM: $\Sigma\{\hat{\theta}\} \approx \frac{1}{B} \sum_{b=1}^B \left(\hat{\theta}^{(b)} - \bar{\theta} \right) \left(\hat{\theta}^{(b)} - \bar{\theta} \right)^T.$
- Berechne das Konfidenzintervall sämtlicher geschätzter Parameter $\hat{\theta}$. Ordne zuerst die generierten geschätzten Parametern aus den Bootstrap-Stichproben $\hat{\theta}^{(b)}$ von der Kleinsten bis zur Größten an. Finde die Quantile $\hat{t}^{(1-\gamma/2)}$ und $\hat{t}^{(\gamma/2)}$, wie es in Alkhatib et al. (2009) **P#14** beschrieben wurde. Schätze dann das Konfidenzintervall mit:

$$\left[\bar{\theta} - \sqrt{\text{diag}(\Sigma\{\hat{\theta}\}) \cdot \hat{t}^{(1-\gamma/2)}} \quad \bar{\theta} + \sqrt{\text{diag}(\Sigma\{\hat{\theta}\}) \cdot \hat{t}^{(\gamma/2)}} \right]$$

Anwendung des Algorithmus 10

Der Algorithmus 10 wurde verwendet, um das Bootstrap-Sampling mit dem GEM-Algorithmus (Algorithmus 3) zu veranschaulichen und die vollständige VKM sämtlicher geschätzter Parameter sowie deren Konfidenzintervalle abzuleiten. Dafür wurde der zweite reale Datensatz verwendet, welcher in Abschnitt 2.4.2 beschrieben wird. Zu diesem Zweck wurden $B = 500$ Samples generiert. In Tabelle 4.1 sind die 95% Konfidenzintervalle neben den Mittelwerten und Standardabweichungen in Bezug auf die sechs geschätzten Kreisparameter und die drei Skalierungsfaktoren dargestellt. Die metrischen Komponenten (Radius und Mittelpunkt) des Kreismodells werden mit Standardabweichungen auf Submillimeteerebene geschätzt. Für die azimutale Orientierung (Φ) wird der Modellparameter im centi-Grad-Bereich geschätzt, was in einer metrischen Unsicherheit von etwa 1,5 cm in einer Entfernung von 35 m resultiert. Im Vergleich zum Ansatz in Paffenholz (2012) stellen die Ergebnisse eine Verbesserung der geschätzten Parameter innerhalb der direkten Georeferenzierung von 3D-Punktwolken dar. Der geschätzte Winkel (θ) wird von der Höhenkomponente der GNSS-Beobachtungen dominiert. Da diese Höhenkomponente typischerweise durch ein höheres Messrauschen als die horizontalen Komponenten (Nord und Ost) gekennzeichnet ist, wird

Tabelle 4.1: Ergebnisse des Bootstrap-Algorithmus aus 500 Wiederholungen. Die Zeilen 1-6 zeigen die Mittelwerte der geschätzten Modellparameter, ihre Varianzen und die 95%-Konfidenzintervalle. Die Zeilen 7-9 geben die Ergebnisse für geschätzte Skalierungsfaktoren für die Nord- (X), Ost- (Y) und Höhen- (Z) Komponenten wieder.

	Mittelwert	Std	95% Konfidenzintervall
$r_{[m]}$	0.2971	0.00022	[0.2968 0.2974]
$\Phi_{[rad]}$	-0.00078	0.00061	[-0.00183 0.0003]
$\theta_{[rad]}$	-0.00665	0.00283	[-0.0112 - 0.00217]
$C_x[m]$	12.2340	0.00013	[12.2338 12.2342]
$C_y[m]$	-16.6317	0.00038	[-16.6321 - 16.6312]
$C_z[m]$	0.01628	0.00138	[0.01458 0.01776]
σ_X^2	$7.3 \cdot 10^{-7}$	$1.2 \cdot 10^{-8}$	$[7.1 \cdot 10^{-7} \ 7.6 \cdot 10^{-7}]$
σ_Y^2	$1.36 \cdot 10^{-6}$	$2.2 \cdot 10^{-8}$	$[1.3 \cdot 10^{-6} \ 1.4 \cdot 10^{-6}]$
σ_Z^2	$3.0 \cdot 10^{-6}$	$4.0 \cdot 10^{-8}$	$[2.9 \cdot 10^{-6} \ 3.1 \cdot 10^{-6}]$

eine höhere Standardabweichung für den Winkel erwartet. Das gleiche Verhalten zeigt sich für die geschätzten Skalierungsfaktoren und die entsprechenden Konfidenzintervalle, während die horizontalen Komponenten im Vergleich zur Höhenkomponente besser abschneiden.



Innovation kurz zusammengefasst | In Kapitel 4.3 wurden simulationsbasierte Techniken in der Inferenzstatistik zur Bestimmung von statistischen Momenten und Konfidenzintervallen vorgestellt. Einerseits wurde ein kombinierter probabilistischer und Fuzzy-Ansatz eingeführt, um zufällige und systematische Unsicherheiten zu modellieren und zu propagieren. Daraus lassen sich zwei wichtige Erkenntnisse ableiten. i) Ist die Abweichung von einer Gaußschen Dichtefunktion der Eingangsgrößen stark, so ist die Durchführung eines verfeinerten Fehlerfortpflanzungsprozesses mittels Monte-Carlo Techniken anstatt eines klassischen Fehlerfortpflanzungsprozesses vorteilhaft. ii) Monte-Carlo Techniken erlauben eine rigorose Berücksichtigung aller in einer Dichtefunktion enthaltenen Informationen innerhalb des Fortpflanzungsprozesses der zufälligen Unsicherheiten für die Ausgangsgröße. Andererseits wurde der GEM-Algorithmus um einen Monte-Carlo-basierten Bootstrap-Algorithmus erweitert, der die Berechnung der Kovarianzmatrix in Bezug auf alle geschätzten Parameter ermöglicht, um eine Qualitätsbewertung der resultierenden Schätzungen abzuleiten.

4.4 Statistische Beurteilung von Schätzergebnissen mit simulationsbasierten Hypothesentests

In der Praxis werden häufig Parametertests eingesetzt, um z. B. eine Deformationsanalyse oder eine Modellvalidierung durchzuführen. Unter den Standardannahmen linearer Beobachtungsgleichungen mit normalverteilten zufälligen Abweichungen und einer gegebenen VKM oder Gewichtsmatrix sind die Teststatistiken und zugehörigen Testverteilungen in solchen Situationen in der Regel bekannt. Für nicht-lineare oder ausreißerbehaftete Beobachtungsmodelle stehen Testverteilungen oftmals nicht oder nicht exakt zur Verfügung. In solchen komplexen Testfällen können Bootstrap-Tests basierend auf MC-Techniken eingesetzt werden (siehe z. B. MacKinnon, 2007; Martinez und Martinez, 2015). In Kombination mit der MC-Simulation (vgl. Koch, 2018) erfordern solche Tests keine Kenntnisse über die Verteilung der eingesetzten Teststatistik.

4.4.1 TESTPROBLEM I: Simulationsbasierter Ansatz für die Modellwahl zur Oberflächenmodellierung von 3D-Punktwolken

Um eine Deformationsdetektion aus Punktwolken zu gewährleisten, ist es entscheidend, die geometrischen Merkmale des Objekts durch ein geeignetes Kurven- oder Oberflächenregressionsmodell genau zu beschreiben. Zweck der Oberflächenanpassung ist es, eine kontinuierliche Modellfunktion aus den redundanten Daten, z. B. aus einer 3D-Punktwolke, zu schätzen. Es existieren zahlreiche Ansätze zur Approximation von Oberflächen, die auf einer impliziten, expliziten oder parametrischen Modellierung basieren. Parametrische Modelle werden in der Regel verwendet, um Daten aus der Punktwolke in Anwendungen, wie der Deformationsüberwachung, anzupassen. Verschiedene parametrische Modelle verhalten sich unterschiedlich in Bezug auf Genauigkeit und Anzahl der zu schätzenden Parameter. Unter den vielen Methoden, die in verschiedenen Anwendungen zur Approximation von Punktwolken verwendet werden, wird die Polynommodellanpassung aufgrund ihrer einfachen Bedienung in der Regel auf glatte und regelmäßige Objekte angewendet (siehe z. B. Yang et al., 2017). Zur Modellierung geometrisch komplizierterer Objekte wird oft die aufwändigere flächenhafte Approximation mittels B-Splines und NURBS bevorzugt. In diesem Zusammenhang wurden viele Forschungsarbeiten zur Optimierung der mathematischen und stochastischen Modelle durchgeführt (vgl. u. a. Harmening und Neuner, 2015b; Koch, 2010, 2009b; Ma und Kruth, 1995). Die Notwendigkeit der Modellauswahl und statistischen Validierung wurde von Wunderlich et al. (2016) hervorgehoben, der die Mängel der aktuellen flächenhaften Deformationsanalyse beschreibt und mögliche Strategien zur Verbesserung dieser Situation vorstellt. Typischerweise hängt die Auswahl des Oberflächenmodells von den Objekteigenschaften ab, z. B. ob die Oberfläche regelmäßig oder unregelmäßig ist. In den meisten Fällen ist jedoch unklar, ob das Objekt ausreichend glatt ist, um durch ein einfaches Modell, beispielsweise als niedergradige, globale Polynomoberfläche, beschrieben zu werden. Der Fokus in Zhao et al. (2018) **P#16** liegt auf der Auswahl der effizientesten parametrischen Beschreibung der Struktur aus TLS-Messungen mittels Hypothesentests. Die hierzu entwickelte Teststrategie ist allgemein zum Vergleich von zwei konkurrierenden parametrischen Modellen einsetzbar. Im Speziellen wurden Polynome höherer Grade und B-Splines zur Beschreibung einer deformierten Objektoberfläche gegeneinander getestet.

Definition des Testproblems

Ziel der Modellauswahl ist es, ein optimales Gleichgewicht zwischen der Parsimonie des Modells (Modell mit geringer Komplexität) und seiner Approximationsqualität zu finden (Harmening und Neuner, 2016). Im Gegensatz zu den Trial-and-Error-Verfahren und den Informationskriterien wird ein Likelihood-Quotienten (LQ)-Test zum Vergleich zwischen zwei konkurrierenden Modellen genutzt. LQ-Tests werden im Allgemeinen verwendet, um zwei verschachtelte (engl. nested) Modelle zu vergleichen. Im vorliegenden Fall sind die zwei konkurrierenden Modelle (Polynom- und B-Spline-Modell zur Approximation der Punktwolke) jedoch nicht verschachtelt (eng. non-nested), da keines der beiden Modelle durch die Einführung einer Reihe von parametrischen Restriktionen auf das andere reduziert werden kann. Hinsichtlich des Auswahlproblems nicht verschachtelter Modelle schlagen die Autoren in Cox (1961) und Vuong (1989) entsprechende Ansätze vor, um den LQ-Test auf nicht verschachtelte Fälle auszuweiten. In diesem Unterabschnitt wird ausschließlich die simulationsbasierte Version des Cox-Tests erklärt. Unter der Annahme von normalverteilten, unkorrelierten und homoskedastischen Zufallsabweichungen können die Beobachtungsmodelle in Bezug auf die generische Log-Likelihood-Funktion definiert werden als

$$\begin{aligned}
 L(\mathbf{x}, \sigma^2; \mathbf{l}) &= \ln \prod_{n=1}^N \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left\{ -\frac{1}{2} \left(\frac{l_n - \mathbf{A}_n \mathbf{x}}{\sigma} \right)^2 \right\} \\
 &= -\frac{N}{2} \ln(2\pi) - \frac{N}{2} \ln(\sigma^2) - \frac{1}{2} \sum_{n=1}^N \frac{(l_n - \mathbf{A}_n \mathbf{x})^2}{\sigma^2},
 \end{aligned} \tag{4.5}$$

wobei der Varianzfaktor σ^2 neben den funktionalen Parametern als unbekannter Parameter behandelt wird. Es werden $L_0(\mathbf{x}, \sigma^2; \mathbf{1})$ und $L_1(\mathbf{x}, \sigma^2; \mathbf{1})$ als spezifische Log-Likelihood-Funktionen in Bezug auf die Designmatrizen beider Modelle definiert. Die beiden Typen der Designmatrix definieren verschiedene funktionale Zusammenhänge, wobei keine der beiden ein Sonderfall der anderen darstellt. Somit sind die beiden Sätze von multivariaten Normalverteilungen definiert als nicht verschachtelt, so dass der LQ-Test nicht in seiner üblichen Form (siehe Koch, 1999) angewendet werden kann.

Gemäß Cox (1961) kann jedoch das logarithmischen LQ zum Testen der Korrektheit des Polynommodells gegenüber dem B-Spline-Modell verwendet werden:

$$\begin{aligned} L_{0,1} &= L_0(\hat{\mathbf{x}}, \hat{\sigma}^2; \mathbf{1}) - L_1(\tilde{\mathbf{x}}, \tilde{\sigma}^2; \mathbf{1}) \\ &= -\frac{N}{2} \ln(\hat{\sigma}^2) - \frac{1}{2\hat{\sigma}^2} \sum_{n=1}^N (l_n - \mathbf{A}_{0n}\hat{\mathbf{x}})^2 + \frac{N}{2} \ln(\tilde{\sigma}^2) + \frac{1}{2\tilde{\sigma}^2} \sum_{n=1}^N (l_n - \mathbf{A}_{1n}\tilde{\mathbf{x}})^2 \end{aligned} \quad (4.6)$$

Es ist zu beachten, dass die substituierten Lösungen der kleinsten Quadrate des Polynommodells sowie des B-Spline-Modells identisch mit den ML-Schätzungen sind. Zudem werden die beiden auftretenden ML-Schätzungen des Varianzfaktors σ^2 durch $\hat{\sigma}^2 = \frac{1}{N} \sum_{n=1}^N (l_n - \mathbf{A}_{0n}\hat{\mathbf{x}})^2$ und $\tilde{\sigma}^2 = \frac{1}{N} \sum_{n=1}^N (l_n - \mathbf{A}_{1n}\tilde{\mathbf{x}})^2$ (Koch, 1999) gegeben. Die Statistik $L_{0,1}$ in Gl. 4.6 folgt nur näherungsweise einer Normalverteilung

1. mit bestimmtem Erwartungswert μ_0 und Standardabweichung σ_0 , wenn das Polynommodell - gemäß Hypothesentest - wahr ist, sowie
2. mit bestimmtem Erwartungswert μ_1 und Standardabweichung σ_1 , wenn das B-Spline-Modell - gemäß Hypothesentest - wahr ist.

So lassen sich die annähernd standard-normalverteilten Testgrößen $T_0 = (L_{0,1} - \mu_0)/\sigma_0$ und $T_1 = (L_{0,1} - \mu_1)/\sigma_1$ für die Durchführung von zwei separaten Hypothesentests unter dem Signifikanzniveau α berechnen:

1. H_0 : Das Polynommodell ist wahr.
2. H_1 : Das B-Spline-Modell ist wahr.

Die Mittelwerte μ_0 und μ_1 sowie die Standardabweichungen σ_0 und σ_1 werden unter Berücksichtigung der beiden Parameterlösungen $(\hat{\mathbf{x}}, \hat{\sigma}^2)$ und $(\tilde{\mathbf{x}}, \tilde{\sigma}^2)$ durch eine MC-Simulation analog zu Williams (1970) bestimmt.

Nach diesem MC-Ansatz wird zunächst eine große Anzahl M von Beobachtungsvektoren $\mathbf{l}^{(1)}$ bis $\mathbf{l}^{(M)}$ zufällig aus der N -dimensionalen Gaußschen Verteilung $\mathcal{N}(\mathbf{A}\hat{\mathbf{x}}, \hat{\sigma}^2 \mathbf{I}_N)$ generiert. Basierend auf diesen generierten Stichproben werden die entsprechenden Lösungen $(\hat{\mathbf{x}}^{(1)}; \hat{\sigma}^{(1)}), \dots, (\hat{\mathbf{x}}^{(M)}; \hat{\sigma}^{(M)})$ in Bezug auf das Polynommodell und $(\tilde{\mathbf{x}}^{(1)}; \tilde{\sigma}^{(1)}), \dots, (\tilde{\mathbf{x}}^{(M)}; \tilde{\sigma}^{(M)})$ in Bezug auf das B-Spline-Modell berechnet. Mit der ersten Lösungsmenge werden die entsprechenden Log-Likelihood-Funktionen $L_0^{(1)}$ bis $L_0^{(M)}$ und mit dem zweiten Satz die entsprechenden Log-Likelihood-Funktionen $L_1^{(1)}$ bis $L_1^{(M)}$ berechnet, so dass die Realisierungen $L_{0,1}^{(1)}, \dots, L_{0,1}^{(M)}$ von (4.6) ermittelt werden können. Anschließend dienen das arithmetische Mittel und die empirische Standardabweichung dieser gesammelten logarithmischen LQ als Schätzungen von μ_0 und σ_0 , was zu einer standardisierten Gaußschen Teststatistik T_0 unter dem aktuell angenommenen Polynommodell führt. Die zweite Teststatistik T_1 (in Bezug auf den Test des B-Spline-Modells) wird in Analogie zum ersten durchgeführt, wobei nun M Pseudo-Beobachtungsvektoren von $\mathcal{N}(\mathbf{A}\tilde{\mathbf{x}}, \tilde{\sigma}^2 \mathbf{I}_N)$ gesammelt werden. Mit den beiden neu generierten Sätzen von Parameterlösungen (bezüglich des Polynom- und B-Spline-Modells) lassen sich die M -Realisierungen des Log-LQ sowie der daraus resultierenden Schätzungen von μ_1 und σ_1 berechnen.

Da Cox (1961) vorschlägt, die einseitigen Entscheidungsregeln

1. Lehne H_0 ab, falls $T_0 < k_\alpha^{N(0,1)}$,
2. Lehne H_1 ab, falls $T_1 > k_{1-\alpha}^{N(0,1)}$,

anzuwenden (wobei α die Irrtumswahrscheinlichkeit, $k_\alpha^{N(0,1)}$ das α -Quantil und $k_{1-\alpha}^{N(0,1)}$ das $1 - \alpha$ -Quantil der Standard-Normalverteilung sind), ist zur Durchführung der beiden Tests die Entscheidung aus vier sich gegenseitig ausschließenden Fällen auszuwählen:

1. Das Polynommodell wird abgelehnt und das B-Spline-Modell wird nicht abgelehnt im Falle von

$$T_0 < k_\alpha^{N(0,1)} \wedge T_1 \leq k_{1-\alpha}^{N(0,1)}. \quad (4.7)$$

2. Das B-Spline-Modell wird abgelehnt und das Polynommodell wird nicht abgelehnt im Falle von

$$T_0 \geq k_\alpha^{N(0,1)} \wedge T_1 > k_{1-\alpha}^{N(0,1)}. \quad (4.8)$$

3. Sowohl das Polynom als auch das B-Spline-Modell werden abgelehnt bei

$$T_0 < k_\alpha^{N(0,1)} \wedge T_1 > k_{1-\alpha}^{N(0,1)}. \quad (4.9)$$

4. Weder das Polynom- noch das B-Spline-Modell werden abgelehnt bei

$$T_0 \geq k_\alpha^{N(0,1)} \wedge T_1 \leq k_{1-\alpha}^{N(0,1)}. \quad (4.10)$$

Anwendung zum Testproblem I

In Zhao et al. (2018) werden Laserscanner-Punktwolken einer in mehreren Epochen durch Belastung deformierten Bogenstruktur (Yang et al., 2017) durch zwei konkurrierende Oberflächenmodelle approximiert: Polynome und B-Splines. Anschließend werden verschiedene angepasste Oberflächenmodelle durch den simulationsbasierten Cox-Test verglichen, um ein geeignetes parametrisches Modell auszuwählen, welches die geometrischen Strukturen (und eventuell die deformierten Flächen) in bestimmten ausgewählten Segmenten am Besten beschreibt. Abbildung 4.3 (a) zeigt die modellierten Polynomflächen vierten Grades für die erste (oben) und dreizehnte (unten) Epoche. In Abbildung 4.3 (b) sind die B-Spline-Approximationen für die gleichen Epochen dargestellt. Die B-Spline-Modellierung wurde mit 361 Parametern approximiert. Es ist deutlich zu erkennen, dass die B-Spline-Oberflächen in (b) detailliertere geometrische Merkmale beschreiben als das Polynommodell in (a). Die Deformationen im Bogenbereich des untersuchten Objekts werden anhand der Differenzbildung des jeweiligen Modells zwischen den untersuchten Epochen in den Z -Koordinaten ermittelt. Die in Abbildungen 4.3 (c) und (d) dargestellten Differenzen zeigen ebenfalls detailliertere Strukturen für das B-Spline-Modell im Vergleich zum Polynom-Modell.

4.4.2 TESTPROBLEM II: Bootstrap-Ansatz für das Testen auf Zeitvariabilität eines AR-Prozesses

Kargoll et al. (2018b) **P#3** und Alkhatib et al. (2018a) **P#4** (siehe Kapitel 2.4) befassen sich mit dem Fall, dass sowohl die Koeffizienten des AR-Modells der zufälligen Abweichungen in einem funktionalen Modell als auch die Formparameter des langschwänzigen Student t-verteilten Messabweichungen unbekannt sind. Mit einem *selbst-adaptiven* robusten Schätzer können die unbekannten Parameter des funktionalen und des AR-Modells sowie die Formparameter der t-Verteilung mit der IRLS-Methode berechnet werden. Dies ist nicht nur möglich, wenn der AR-Prozess kovarianzstationär ist, sondern auch, wenn die AR-Koeffizienten durch eine lineare Regression als zeitvariable

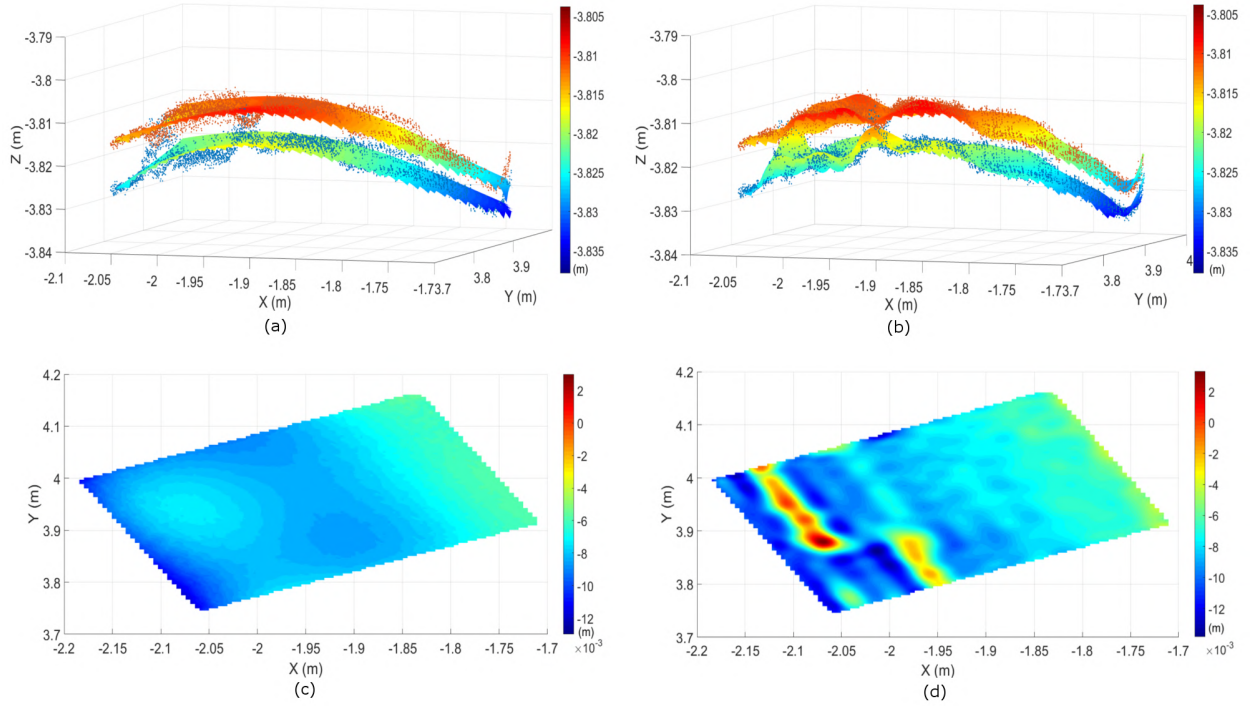


Abbildung 4.3: Vergleich von Polynom- (a,c) und B-Spline-Oberflächenmodellen (b,d) in Bezug auf die Unterschiede der 1. und 13. Epoche in Segment I (Zhao et al., 2018)

Parameter modelliert werden (Kargoll et al., 2018a). Solche Modelle haben sich als zweckmäßig erwiesen, um nicht-stationäre Effekte in Zeitreihenmessungen zu beschreiben, die nicht als Teil des deterministischen Modells auf der Ebene der Beobachtungsgleichungen beschrieben werden können (Kargoll et al., 2018a). Wie ein derartiges AR-Fehlermodell mit t-verteilten weißen Rauschkomponenten auf Zeitvariabilität getestet werden kann, ist eine Fragestellung, welche im Alkhatib et al. (2019) **P#18** behandelt wird. Während sich aus der Standard-Testtheorie eine entsprechend geeignete Teststatistik leicht ermitteln lässt, ist die zugehörige PDF der hier beschriebenen Teststatistik aufgrund der Komplexität des Modells und bisher fehlender analytischer Untersuchung nicht bekannt. Bei solchen schwer zu handhabenden Testproblemen kann das Bootstrapping-Verfahren eine adäquate Lösung bieten (Koch, 2018).

Definition des Testproblems

Es wird davon ausgegangen, dass Beobachtungen $\ell = [\ell_1, \dots, \ell_n]^\top$ durch eine nicht-lineare, vektorwertige Funktion $\mathbf{f}$ ('deterministisches Modell') und zufällige Abweichungen $\mathbf{e} = [e_1, \dots, e_n]^\top$ durch die sogenannten Beobachtungsgleichungen $\ell = \mathbf{f}(\mathbf{x}) + \mathbf{e}$ beschrieben werden können, wobei $\mathbf{x} = [x_1, \dots, x_q]^\top$ unbekannte Parameter sind. Es wird angenommen, dass die zufälligen Abweichungen einem AR(p)-Prozess

$$e_t = \sum_{j=1}^p \alpha_{t,j} e_{t-j} + u_t \quad (t = 1, \dots, n), \quad (4.11)$$

folgen, wobei die zeitvariablen Koeffizienten $\alpha_{t,j} = \mathbf{B}_t \mathbf{y}_j$ durch Linearkombinationen mit festen Vektoren $\mathbf{B}_t$ und unbekannten Parametern $\mathbf{y}_j = [y_{j,1}, \dots, y_{j,m}]^\top$ beschrieben werden. Die Zufallsvariablen $\mathbf{u} = [u_1, \dots, u_n]^\top$ in Gl. 4.11 werden als unabhängig und identisch $t_\nu(0, \sigma^2)$ -verteilt mit unbekanntem Skalenparameter σ^2 und unbekanntem Freiheitsgrad ν angenommen. Gängige Formen der Zeitvariabilitätsmodelle sind Polynome, die durch $\mathbf{B}_t = [1 \quad \tau_t^1 \quad \dots \quad \tau_t^{m-1}]$ definiert sind.

Falls nicht bekannt ist, ob der AR-Prozess zeitvariabel ist oder nicht, so ist es sinnvoll, die Zeitvariabilitätsparameter $\mathbf{Y} = [y_{1,2}, \dots, y_{1,m}, \dots, y_{p,2}, \dots, y_{p,m}]^\top$ im Modell zu behalten und auf ihre gemeinsame Signifikanz zu testen. Zu diesem Zweck wird die Null- und die Alternativhypothese definiert durch

$$H_0 : \mathbf{Y} = \mathbf{0} \quad \text{gegen} \quad H_1 : \mathbf{Y} \neq \mathbf{0}. \quad (4.12)$$

Eine sinnvolle Wahl für eine Teststatistik, welche die Abweichungen von H_0 (d. h. dem Grad, in dem $\mathbf{Y}$ ungleich Null ist) in kumulativer Weise ermittelt, ist durch die gewichtete Quadratsumme $T = \hat{\mathbf{Y}}^\top \hat{\Sigma}_{\hat{\mathbf{Y}}\hat{\mathbf{Y}}}^{-1} \hat{\mathbf{Y}}$ der geschätzten Parameter $\mathbf{Y}$ gegeben. Da die Verteilung der Teststatistik bzgl. der Zeitvariabilitätsparameter unbekannt ist, wird auf einen simulationsbasierten Bootstrap-Test zurückgegriffen. Dies ermöglicht eine Beurteilung, ob der sich aus den gegebenen Messungen ergebende Wert für die Teststatistik zu groß ist, um H_0 zu verwerfen. Dieser Fall tritt ein, falls der aus den Messungen berechnete Wert der Teststatistik zu einem hohen Prozentsatz (z. B. $> 95\%$) den Wert der aus den simulierten Beobachtungsreihen ermittelten Teststatistik überschreitet. Die einzelnen Schritte dieses Ansatzes werden im Folgenden für das allgemeine Problem des Testens gegen ein bestimmtes Zeitvariabilitätsmodell erläutert. Hierbei sind das deterministische Modell $\mathbf{f}(\mathbf{x})$, die AR-Modellordnung p , die Zeitvariabilitätsdesignmatrix $\mathbf{B}$ und das Signifikanzniveau α zu spezifizieren.

- **Schätzschritt:** Der in Kargoll et al. (2018a) beschriebene modifizierte EM-Algorithmus wird um das Linearisierungsverfahren (erklärt in Alkhatib et al. (2018a)) erweitert. Dieser Algorithmus liefert die Schätzwerte $\hat{\mathbf{x}}$ der funktionalen Parameter, die Schätzwerte $\hat{\mathbf{Y}}$ der zeitvariablen Parameter und ihre VKM $\hat{\Sigma}_{\hat{\mathbf{Y}}\hat{\mathbf{Y}}}$, den geschätzten Skalierungsfaktor $\hat{\sigma}^2$, die Schätzung $\hat{\nu}$ des Freiheitsgrades der zugrunde liegenden t-Verteilung sowie die geschätzten Residuen $\hat{\mathbf{u}}$ des weißen Rauschens.
- **Testschritt:** Der Wert $T = \hat{\mathbf{Y}}^\top \hat{\Sigma}_{\hat{\mathbf{Y}}\hat{\mathbf{Y}}}^{-1} \hat{\mathbf{Y}}$ der Teststatistik wird berechnet.
- **Generierungsschritt:** Das Generierungsschema beginnt mit der Generierung der Komponenten $u_t^{(k)}$ mit weißem Rauschen für die Zeitinstanzen $t = 1, \dots, n$ und die MC-Iterationen $k = 1, \dots, B$ (B ist die Gesamtanzahl der Bootstrap-Stichproben). Zu diesem Zweck werden die folgenden beiden Alternativen betrachtet (Efron und Hastie, 2016):
 - (1) *Parametrisches Bootstrapping:* Zufallszahlen in Bezug auf die $t_{\hat{\nu}}(0, \hat{\sigma}^2)$ -Verteilung werden unabhängig voneinander für die Komponente $u_t^{(k)}$ mit weißem Rauschen generiert.
 - (2) *Nichtparametrisches Bootstrapping:* Zufallszahlen $\lambda_t^{(k)}$ werden im Bezug auf die diskrete Gleichverteilung $\mathcal{U}(1, n)$ mit Zurücklegen gezogen, um $u_t^{(k)} = \hat{u}_{\lambda_t^{(k)}}$ zu definieren. Damit können die Residuen $\hat{\mathbf{u}}$ aus dem Schätzungsschritt erneut verwendet werden.

Anschließend werden die zuvor erzeugten Komponenten des weißen Rauschens durch Auswertung der zeitkonstanten AR-Modellgleichung $e_t^{(k)} = \sum_{j=1}^p \bar{\alpha}_j e_{t-j}^{(k)} + u_t^{(k)}$ mit den über die Zeit gemittelten einzelnen AR-Koeffizienten $\bar{\alpha}_j = \frac{1}{n} \sum_{t=1}^n \hat{\alpha}_{t,j}$ ($j = 1, \dots, p$) unter Verwendung der Startwerte $e_0^{(k)} = \dots = e_{t-p}^{(k)} = 0$ rekursiv korreliert. Das geschätzte deterministische Modell wird so dann zu diesen farbigen Rauschkomponenten addiert, woraus sich die generierten Beobachtungen $\ell_t^{(k)} = f_t(\hat{\mathbf{x}}) + e_t^{(k)}$ ergeben. Die resultierende Messreihe $\ell^{(k)}$ wird auf die gleiche Weise wie die tatsächliche Messreihe ℓ im vorherigen Schätzschritt ausgeglichen, was die Schätzungen aus B Stichproben für $\hat{\mathbf{x}}^{(k)}$, $\hat{\mathbf{y}}^{(k)}$, $\hat{\Sigma}_{\hat{\mathbf{y}}\hat{\mathbf{y}}}^{(k)}$, $(\hat{\sigma}^2)^{(k)}$ und $\hat{\nu}^{(k)}$ liefert. Die gesampelte Teststatistik wird durch $T^{(k)} = (\hat{\mathbf{Y}}^{(k)})^\top (\hat{\Sigma}_{\hat{\mathbf{Y}}\hat{\mathbf{Y}}}^{(k)})^{-1} \hat{\mathbf{Y}}^{(k)}$ erhalten.

- **Auswertungsschritt:** Um festzustellen, wie extrem der Testwert T im Vergleich zu den

Werten $T^{(1)}, \dots, T^{(B)}$ unter H_0 ist, wird der p-Wert $\widehat{pv}$ geschätzt durch

$$\widehat{pv} = \frac{1}{B} \sum_{k=1}^B I(T^{(k)} > T), \quad (4.13)$$

wobei I die Indikatorfunktion repräsentiert, die bei wahrem Argument den Wert 1 und ansonsten dem Wert 0 annimmt.

- **Entscheidungsschritt:** Lehne H_0 ab, wenn der geschätzte p-Wert kleiner als das vordefinierte Signifikanzniveau α ist.

Anwendung zum TESTPROBLEM II

Zur Demonstration dieses Testproblems wird als Beispiel die simulierte Fourierreihe in Gl. 2.16 aus Abschnitt 2.4.2 (siehe auch Kargoll et al., 2018b, **P#3**) herangezogen. Die Anzahl der Beobachtungen beträgt $n = 10.000$. Bezüglich des farbigen Rauschens e_t wird ein zeitvariabler AR(1)-Prozess mit dem globalen Polynom $\alpha_t = y_1 + y_2 \cdot \tau_t$ angegeben. Für $y_2 = 0$ ist der AR(1)-Prozess zeitkonstant. Die Simulation der p-Werte ($\widehat{pv}^{(i)}$) und der Gütefunktion des Testproblems besteht für jede Wiederholung $i \in \{1, \dots, 100\}$ aus den folgenden Schritten:

- Erzeuge das weiße Rauschen $u_t^{(i)}$ aus der wahren t-Verteilung $t_3(0, 10^{-6})$ für $t = 1, \dots, n$
- Korreliere das weiße Rauschen mit Hilfe des AR(1)-Prozesses $e_t^{(i)} = \alpha_{t,1}e_{t-1}^{(i)} + u_t^{(i)}$ mit $\alpha_{t,1} = y_1 + y_2 \cdot \tau_t$. Hierfür wird der wahre Achsenabschnittsparameter $y_1 = -0,5$ festgelegt, und die wahren Steigungsparameter y_2 werden in Schritten $[0 : 0.0001 : 0.003]$ variiert.
- Füge das farbige Rauschen dem wahren Schwingungsmodell (2.16) hinzu, um die simulierten Beobachtungen $\ell^{(i)}$ zu bestimmen.
- Führe den Schätzungsschritt aus, um $\hat{\mathbf{x}}^{(i)}, \hat{\mathbf{y}}^{(i)}, \hat{\Sigma}_{\hat{\mathbf{y}}\hat{\mathbf{y}}}^{(i)}, \hat{\sigma}^{(i)}, \hat{\nu}^{(i)}$ mit dem gleichen funktionalen und zeitvariablen AR-Modell zu erhalten, wie zuvor beschrieben.
- Führe den Testschritt zur Berechnung $T^{(i)} = (\hat{y}_2^{(i)} / \hat{\sigma}_{\hat{y}_2}^{(i)})^2$ durch.
- Führe den Generierungsschritt (parametrisch/nicht-parametrisch) zur Berechnung $T^{(i,k)}$ für $k = 1, \dots, B$ durch.
- Führe den Bewertungsschritt aus, um $\widehat{pv}^{(i)}$ zu berechnen.

Die oben beschriebene Simulation wurde mit unterschiedlichen Bootstrap-Stichproben durchgeführt. Die Variation von B ändert den ermittelten p-Wert innerhalb eines MC-Durchlaufs nicht stark. Insbesondere die Irrtumswahrscheinlichkeit $\alpha = 0,05$ wird im Durchschnitt sehr genau reproduziert, da die H_0 -Ablehnungsrate für sämtliche 100 Wiederholungen, die durch

$$\frac{1}{100} \sum_{i=1}^{100} I(\widehat{pv}^{(i)} < 0.05)$$

definiert sind, sich zu 0,04 für parametrisches und 0,05 für nicht-parametrisches Bootstrapping ergibt. Abbildung 4.4 zeigt die empirische Gütefunktion (Sensitivität der Hypothesentests).

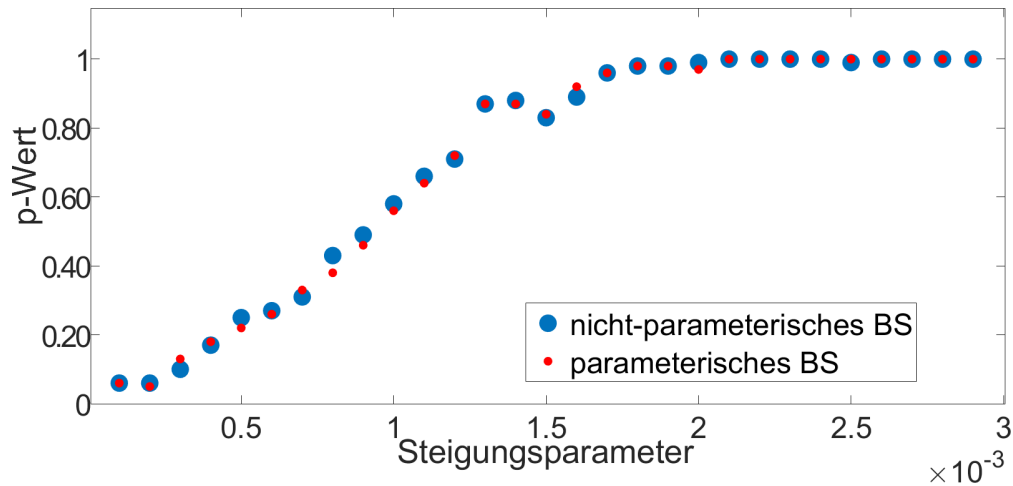


Abbildung 4.4: Vergleich des geschätzten $\widehat{pv}^{(i)}$ -Wertes (Ablehnung der Nullhypothese, wenn $\widehat{pv}^{(i)} < 0.05$) unter dem $AR(1)$ -Modell mit zeitvariablen Parametern im Wertebereich $y_2 = [0 : 0.0001 : 0.003]$ für parametrisches und nicht-parametrisches Bootstrapping; Übersetzung der Originalabbildung aus Alkhatib et al. (2019)



Innovation kurz zusammengefasst | In diesem Abschnitt wurden zwei simulationsbasierte Teststrategien demonstriert, mit denen anspruchsvolle Modellauswahlprobleme gelöst werden können, die mit bisherigen Standardtests nicht lösbar waren. Mit der ersten Teststrategie kann z. B. eine optimale VKM oder ein optimales B-Spline Modell identifiziert werden, wobei der Test (im Gegensatz zu den üblicherweise angewendeten Informationskriterien) eine Aussage hinsichtlich signifikanter Unterschiede in der Güte der einzelnen konkurrierenden Modelle macht. Die eingeführten Hypothesentests zur Modellauswahl für stochastische und funktionale Modelle konnten die Qualität des Deformationsmonitoring-Prozesses erheblich verbessern. Mit der zweiten Teststrategie kann überprüft werden, ob die zufälligen Abweichungen einer beobachteten Regressionszeitreihe mit unbekannten Regressionskoeffizienten kovarianz-stationär sind oder nicht. Dieses Verfahren wurde als neuartiger Bootstrap-Modellauswahltest auf Grundlage eines AR-Prozesses mit zeitvariablen Koeffizienten und t-verteilten Innovationen realisiert.

4.5 Zusammenfassung des Kapitels

Kapitel 4 befasst sich mit innovativen simulationsbasierten MC- und Bootstrapping-Methoden zur Ableitung von Unsicherheiten von Ausgangsgrößen (im Kontext von GUM), zur Berechnung von Konfidenzintervallen und zur Entscheidungsfindung mit Hilfe von Hypothesentests. Der erste Teil des Kapitels behandelt eine funktionale Messgleichung mit mehrdimensionalen Eingangs- und Ausgangsgrößen. Die Eingangsgrößen können entweder als zufällig oder systematisch betrachtet werden. Als Anwendungsszenario für die hier vorgestellte Methodik werden TLS-Profildaten zur Objektüberwachung und Deformationsanalyse kinematisch untersucht. Der Fokus liegt dabei auf einer differenzierteren Modellierung der Unsicherheiten der Beobachtungen und der abgeleiteten Positionen der Profilpunkte. Im Rahmen der zweiten Anwendung wird der bisherige GEM-Algorithmus um einen MC-Bootstrap-Algorithmus ergänzt, wodurch die Approximation der VKM der geschätzten Parameter ermöglicht wird. Zudem erlaubt der neue Bootstrap-Algorithmus die Berechnung von Konfidenzintervallen, ohne Kenntnis der zugrunde liegenden Verteilung. Der zweite Teil des

Kapitels behandelt zwei Testprobleme mit MC- und Bootstrapping-Techniken. Für beide Testprobleme standen bisher keine Standardtests zur Verfügung. Bei das erste Testproblem wurden Bootstrapping-Techniken zur optimalen Modellauswahl zwischen zwei Ansätze zur flächenhaften Approximation der 3D-Punktwolken durchgeführt. Die innovative Lösung des zweiten Testproblems erlaubt eine Entscheidung darüber, ob sich die Koeffizienten eines AR-Prozesses über die Zeit ändern oder nicht.

5 Zusammenfassung und Ausblick

Im Rahmen dieser Habilitationsschrift wurden computergestützte Methoden und Algorithmen dargestellt, die es ermöglichen, im gesamten Prozess der Unsicherheitsmodellierung, Schätzung und Hypothesentests die Charakteristik von Beobachtungen und funktionalen Modellen realistischer abzubilden. Die hier entwickelten Verfahren gehen deutlich über die klassischen geodätischen Auswertverfahren hinaus. Auf Beobachtungsebene sind beispielsweise sowohl nicht-normalverteilte zufällige und systematische Abweichungen im Unsicherheitsbudget berücksichtigt. Des Weiteren können Ausreißer in den Beobachtungsdaten effizient identifiziert und deren Einfluss auf die Bestimmung der Parameter bzw. Zustandsgrößen signifikant reduziert werden. Im Bereich der nicht-linearen funktionalen Modelle zwischen den Beobachtungen und Parametern bzw. Zustandsgrößen sowie der induktiven Statistik (insb. Hypothesentests) erlauben die simulationsbasierten Algorithmen eine signifikant bessere Bestimmung der Zielgrößen und eine realistischere Angabe von Bereichsschätzern.

5.1 Zusammenfassung

Die vorliegende kumulative Habilitation wurde in drei Hauptkapiteln (Kapitel 2 bis 4) zusammengefasst.

Kapitel 2 befasste sich mit Entwicklung von Optimierungsverfahren zur Parameterschätzung in linearen und nicht-linearen parametrischen Modellen. Zu diesem Zweck wurden Optimierungsaspekte anhand von drei parametrischen Modellen zur best-möglichen Anpassung an die Daten demonstriert. Im ersten Modell, der B-Spline Regression, wurden neue heuristische Verfahren zur Bestimmung des Knotenvektors in der Kurven-Approximation vorgestellt. Die EGA und MCM-Algorithmen ermöglichen es, Punktwolken von komplex geformten Objekten unter Vorliegen von Datenlücken zu behandeln. Beide entwickelten Ansätze konvergieren schneller zu einem globalen Optimum als die Standardverfahren, wie z. B. der CSA. Die Anwendung auf herausfordernde reale Datensätze hat gezeigt, dass der EGA- signifikant besser geeignet ist, um mit Datenlücken umzugehen und er erzielte zudem auch bei weiteren Datensätzen eindeutig die besten Ergebnisse. Der Grund für die Leistungsfähigkeit des EGA beruht auf zwei Faktoren: (i) Die angewandte Crossover-Technik erhöht die Wahrscheinlichkeit für multiple Knoten, die zur Approximation von Spitzen, Sprüngen oder Diskontinuitäten notwendig sind und (ii) die vorgeschlagene Initialisierung verhindert den ungünstigen Effekt von Singularitäten in der Kontrollpunktschätzung, die durch Datenlücken verursacht werden.

Im zweiten Modell wurden multivariate Regressionszeitreihen mit sowohl autoregressivem, farbigem Rauschen als auch ausreißerbehaftetem/langschwänzigem weißen Rauschen betrachtet. Letztere wurden durch skalierte t-Verteilungen mit geschätztem Freiheitsgrad modelliert. Dies ermöglicht eine datenadaptive robuste ML-Schätzung. Die Theorie und Implementierung eines GEM-Algorithmus wurde beschrieben. In diesem Algorithmus werden die deterministischen Modellparameter, die AR-Koeffizienten, die Skalenfaktoren und der Freiheitsgrad der t-Verteilung, für multiple Zeitreihen über den IRLS geschätzt. MC-Simulationen haben gezeigt, dass der entstehende Bias bei den geschätzten Parametern durch die Berücksichtigung einer ausreichend großen Anzahl von Beobachtungen deutlich verringert werden kann. Der vorgeschlagene Algorithmus wurde auch in realen Datenexperimenten mit Beschleunigungsmessungen und mit GNSS-Messungen erfolgreich getestet und lieferte sehr gute Ergebnisse. Eine Modellauswahl hinsichtlich der Ordnung des AR-Prozesses wurde mit Hilfe eines periodogramm-basierten Tests auf weißes Rauschen durchgeführt. Schließ-

lich wird die Analyse der geschätzten Freiheitsgrade in den verschiedenen multivariaten Zeitreihen gezeigt. Hier zeigte sich beispielsweise, dass die Residuen des weißen Rauschens der Hochwert-Komponente in der GNSS-Zeitreihe von einer Normalverteilung abweichen.

Der dritte Teil des Kapitels 2 beschäftigte sich mit Bayesschen linearen und nicht-linearen Regressionsmodellen unter der Annahme von t -verteilterm weißen Rauschen. Der Fokus lag auf der Entwicklung eines simulationsbasierten MCMC-Algorithmus zur Approximation der a-posteriori PDF. Der Vorteil dieses entwickelten Algorithmus gegenüber dem vorher entwickelten GEM-Algorithmus ist, dass ein nicht-lineares funktionales Modells nicht zwingend linearisiert werden muss. Zudem ist es beim Bayesschen Regressionsmodell möglich, a-priori Informationen hinsichtlich der unbekannten Parameter in das Modell zu integrieren. Die durchgeführte MC-Simulation zur Untersuchung des MCMC-Algorithmus hat gezeigt, dass der Einfluss von Ausreißern, durch die Annahme von Student-verteilten Abweichungen mit niedrigen Freiheitsgraden, im Vergleich zu der MkQ-Lösung signifikant reduziert wird. Der entwickelte Algorithmus wurde auf reale Datensätze zur Schätzung von GNSS-Trajektorien zur Georeferenzierung eines TLS-basierten MSS eingesetzt. Weiterhin wurde der Algorithmus zur Ableitung von Marktdaten für Wertermittlungszwecke in kaufpreisarmeren Lagen verwendet, um eine datengetriebene Prädiktion von Verkehrswerten zu ermöglichen. Der entwickelte MCMC-Algorithmus kann nur funktionale Modelle, die einem GMM entsprechen, lösen.

Im Kapitel 3 wurden neuartige Filterungstechniken zur zeit-diskreten Zustandsschätzung von nicht-linearen Systemen entworfen. Der erste Filterungsansatz basiert auf dem IEKF und kann erstmalig nichtlineare Gleichungs- und Ungleichungszustandsrestriktionen mit impliziten Messgleichungen berücksichtigen. Die Berücksichtigung solcher Restriktionen wird mittels eines flexiblen PDF-Truncation-Algorithmus realisiert. Diese Methode wurde für die Georeferenzierung von Indoor- und Outdoor-Laserscanner-basierten kinematischen MSS angewendet und evaluiert. Im Ergebnis kann festgestellt werden, dass die Berücksichtigung geeigneter Restriktionen zwischen den Zustandsparametern sinnvoll ist. Die Einführung von Ungleichungs- zusätzlich zu den Gleichungszustandsrestriktionen ermöglicht es, weitere Verbesserungen in Bezug auf die Genauigkeit der geschätzten Zustandsparameter zu erzielen. In dem vorgestellten IEKF werden die Ebenenparameter und Eckpunkte als Zustandsparameter eingeführt.

Im zweiten Teil des Kapitels 3 wurde der simulationsbasierte EKPF-Algorithmus entwickelt. Dieser Algorithmus kombiniert das klassische EKF und den PF, um die resultierende a-posteriori PDF durch generierte Partikel möglichst optimal zu approximieren, ohne die Anzahl der Partikel zu erhöhen. Somit wird die Rechenzeit bei mindestens gleichwertigen oder sogar besseren Zustandsschätzungen deutlich verringert. In der dritten Filterungstechnik wird ein neuer Algorithmus, der EGKF, zur Zustandsschätzung in Fällen entwickelt, in denen neben der probabilistischen Unsicherheit auch UBB-Unsicherheiten auftreten. Die Anwendung auf simulierte aber auch auf echte Daten hat gezeigt, dass die neue Methode zu wesentlich genaueren Schätzungen im Vergleich zu dem klassischen EKF führt, wenn UBB-Unsicherheiten vorhanden sind.

Kapitel 4 widmete sich neuartigen simulationsbasierten MC- und Bootstrapping-Methoden zur Ableitung von Unsicherheiten von Ausgangsgrößen (im Kontext von GUM), zur Berechnung von Konfidenzintervallen und zur Beurteilung von Entscheidungen in nicht-linearen Hypothesentests. Im ersten Teil dieses Kapitel wurde eine funktionale Messgleichung mit mehrdimensionalen Eingangs- und Ausgangsgrößen analysiert. Die Eingangsgrößen können sowohl zufällige als auch systematische Abweichungen beinhalten. Als Anwendungsfall für die entwickelte Methodik wurden kinematische TLS-Profildaten für die Objektüberwachung und Deformationsanalyse untersucht. Der Schwerpunkt wurde hier auf eine verfeinerte Modellierung der Unsicherheit der Beobachtungen und der abgeleiteten Positionen der Profilpunkte gelegt. Um die gesamte Datenverarbeitungskette zu berücksichtigen, wurde die Strategie zur Generierung und Analyse von Zeitreihen betrachtet. MC-Simulationstechniken wurden angewandt, um numerische Ergebnisse zur Diskussion und Validierung zu liefern. In der zweiten Anwendung wurde der bisherige GEM-Algorithmus um einen MC-Bootstrap-Algorithmus erweitert, um die Approximation der VKM von unbekannten Para-

metern zu ermöglichen. Der neuartige Bootstrap-Algorithmus erlaubt auch die Berechnung von Konfidenzintervallen ohne Angabe der Quantile einer bestimmten Verteilung.

Im zweiten Teil dieses Kapitel wurden zwei Testprobleme anhand MC- und Bootstrapping-Techniken behandelt. In beiden Testproblemen ist es nicht möglich, einen linearen Hypothesentest durchzuführen, da die Verteilung der Teststatistik analytisch nicht bekannt ist. Ziel des ersten Testproblems ist es, eine optimale Modellauswahl zwischen zwei funktionalen Modellen zur flächenhaften Approximation der 3D-Punktwolke durchzuführen: Polymome und B-Splines. Dieser Test wurde mit Hilfe der simulationsbasierten Version des Cox-Test durchgeführt. Im zweiten Testproblem geht es um die Signifikanzentscheidung in einer nichtlinearen Regression über zeitvariable AR-Prozesse. Dieses umfassende Beobachtungsmodell mit zeitvariablen Parameter, autoregressiven Zufallsabweichungen und t-verteilttem weißen Rauschen wurde bereits in Kapitel 2 entwickelt. Aufgrund der Komplexität des Modells hat die Teststatistik im Allgemeinen keine exakte, analytisch beschreibbare Verteilung. Gründe dafür sind die Nicht-linearität des Beobachtungsmodells und die Nicht-Normalverteiltheit der zufälligen Abweichungen. Um zu überprüfen, ob insbesondere die zeitliche Variabilität signifikant ist oder nicht, wurde ein Bootstrap-Test auf Basis der bekannten gewichteten quadratisch geschätzten Parameter vorgeschlagen und umgesetzt.

5.2 Ausblick

Für zukünftigen Arbeiten im Bereich der Bestimmung des Knotenvektors für die B-Spline Kurvenapproximation ist geplant, sämtliche verwendeten Parameter des EGA in einer umfassenden MC-Simulation zu analysieren. Zudem wird zukünftig auf die Umsetzung des hier entwickelten EGA in der Approximation von B-Spline-Oberflächen fokussiert. Ein weiterer wichtiger Punkt bei der B-Spline-Approximation ist die Berücksichtigung von Ausreißern bei der Kontrollpunkt-schätzung. Für die Knotenvektorbestimmung ist aktuell ein Machine-Learning-Algorithmus basierend auf einer Support Vector Machine in der Entwicklung. Der in dieser Arbeit durchgeführte Approximationsprozess kann auch für die B-Spline-Flächen angepasst werden. Dort ergeben sich jedoch andere Schwierigkeiten und Herausforderungen. Beispielsweise ist der Einsatz der B-Spline-Flächenapproximation in der Praxis bei großen Datensätzen mit hoher Variabilität und großen räumlichen Lücken eine Herausforderung. Deswegen werden in solchen Fällen alternative Methoden, wie Multilevel-B-Splines (MBS) oder Hierarchische B-Splines (HBS), vorgeschlagen. Gerade der Einsatz von MBS zur flächenhafte Approximationen von Punktwolken potenziert den Rechenaufwand.

Im Bereich der nicht-linearen Zeitreihenmodelle mit t-verteilttem Messrauschen wird beabsichtigt, den GEM-Algorithmus und den MCMC-Algorithmus zu erweitern, indem zusätzlich Kreuzkorrelationen zwischen den einzelnen Zeitreihen modelliert werden. Außerdem ist zu untersuchen, ob die GEM- und MCMC-Algorithmen derart erweiterbar sind, dass der Einfluss der Datenlücken minimiert werden kann. Des weiteren ist es geplant, beide Algorithmen so zu erweitern, dass auch funktionale Modelle im Kontext eines GHM gelöst werden können.

Die Entwicklung von Filterungstechniken zur effizienten und robusten Zustandsparemeter ist eine zentrale Aufgabe im DFG Graduiertenkolleg (2159) *Integrität und Kollaboration in dynamischen Sensornetzen (i.c.sens)*. Das Ziel im Graduiertenkolleg ist es, Konzepte zur Sicherstellung der Integrität von kollaborativ agierenden Systemen in dynamischen Sensornetzen zu erforschen. Der in Vogel et al. (2018, 2019) und Vogel (2020) entwickelte Filterungsalgorithmus wird im laufenden Arbeiten weiter optimiert. Mit zunehmender Anzahl von Ebenen und somit auch Eckpunkten wird die Zustandsparemeterschätzung aufgrund der Länge des Zustandsvektors ineffizient. Deswegen wird in zukünftigen Arbeiten der IEKF in ein duales Zustandsmodell umformuliert. Außerdem wird die Entwicklung eines robusten Partikelfilters zur kollaborativen Georeferenzierung unter Berücksichtigung der Integrität beabsichtigt. Die Positionierung und Navigation von Fahrzeugen basiert in der Regel auf der Kombination von globalen Navigationssatellitensystemen (GNSS) und zusätzlichen

IMU-Sensoren. Da bei langen Zeiträumen die Ergebnisse hauptsächlich auf den GNSS-Messungen beruhen, ist bei Unterbrechungen und/oder bei Ausreißern die Positionierung und Navigation nicht mehr integer. Dies gilt z.B. für schwierige Umgebungsbedingungen (Straßenschluchten) und für die Indoor-Anwendung. Um dieses Problem zu umgehen, wurde eine neue Methodik entwickelt, die den Fahrzeugdaten Zwangsinformationen hinzufügt. Die Bestimmung der Fassaden und ihrer geometrischen Charakteristik kann z.B. durch mobile, hochgenaue Kartierungssysteme oder durch eine kollaborative Kartierung aus einer großen Anzahl von autonomen Fahrzeugen erfolgen. Weiterhin stehen zur Beurteilung der Toleranzwerte der Bauwerke ISO-Konstruktionsstandards zur Verfügung. Im Rahmen dieses GRK-Projekts wird diese Methodik auf die kollaborative Positionierung mit Restriktionen für die Navigation in problematischen Umgebungen angewandt.

Es soll auch im Rahmen von zukünftigen Arbeiten eines robusten Partikelfilters für die Zustandsschätzung mit stochastischen und mengenbasierten Unsicherheiten in Sensornetzen weiterentwickelt werden. Obwohl sich das in Sun et al. (2019, 2018) entwickelte Filter als erfolgreich erwiesen hat, gibt es noch Raum für Verbesserungen. Erstens hängt die Effizienz dieser Methode von der initialen Auswahl der Parameter ab, die die System- und die Messunsicherheiten beschreiben. Bei sinnvoller Wahl dieser Vorinformationen konvergiert der vorgestellte Algorithmus schließlich zur korrekten Unsicherheitsschätzung, aber die Konvergenz hängt stark von der Auswahl der Startwerte der Unsicherheitsparameter ab, die eine schnellere Konvergenz garantieren. Aus diesem Grund soll zukünftig eine neue adaptive Komponente zur empirischen Schätzung der beiden Unsicherheitskomponenten integriert werden. Außerdem wurde in diesem Algorithmus angenommen, dass die entsprechende PDF des System- und Messrauschens normalverteilt ist. In der Praxis gilt diese Annahme aufgrund der im Rahmen der Arbeit ausführlich erläuterten Gründe und Beispiele oftmals nicht. Es ist daher geplant, diese Methode auf den allgemeinen Fall von nicht-normalverteilten Abweichungen zu erweitern – z. B. durch die Verwendung von anderen Filterungstechniken wie dem PF. Die in diesem Kapitel entwickelten Filterungsansätze können aufgrund ihrer Allgemeingültigkeit in verschiedenen Disziplinen adaptiert werden.

Die in Kapitel 4 entwickelten simulationsbasierte Teststrategien sind generell auf anderen Testprobleme übertragbar. Zukünftig sind neue Methoden für eine noch realistischere Fortpflanzung von zufälligen und systematischen Abweichungen zu entwickeln. Des Weiteren sind diese dann auf Ebene der Zielgrößen erhaltenen Unsicherheiten in gemeinsamen Hypothesentests zu analysieren. Der Bayessche Hypothesentest-Ansatz bietet eine leicht durchführbare und leicht verständliche Alternative zu klassischen statistischen Tests. Die a-posteriori PDF ermöglichen uns mathematische Konzepte, die im Vergleich zu Begriffen wie p-Werten, viel leichter zu erfassen, zu interpretieren und anzuwenden sind. Schließlich erlaubt uns das Konzept des Bayesschen Netzwerks viel komplexere Experimente zu entwickeln und jede Hypothese zu testen. Daher wird zukünftig die Formulierung des Testproblems in Bayesscher Inferenz angestrebt.

Teil II

TEIL 2: Eigene Veröffentlichungen

6 Liste der eigenen Publikation

Die für meine kumulative Habilitation an der Leibniz Universität Hannover eingereichte Zusammenstellung umfasst 18 Publikationen, die aus 52 begutachteten Manuskripten ausgewählt wurden, die ich im Zeitraum 2009-2019 veröffentlicht habe. Von den insgesamt 52 Publikationen sind 27 in internationalen Zeitschriften erschienen und 25 als Buchbeiträge oder als begutachtete Konferenzbeiträge. Im Folgenden findet sich die Zusammenfassung der für die Habilitation relevanten Publikationen.

6.1 Veröffentlichungen zur Kapitel 2

- P#1** Bureick, J.; **Alkhatib, H.**; Neumann, I. (2016): Robust Spatial Approximation of Laser Scanner Point Clouds by Means of Free-form Curve Approaches in Deformation Analysis, In: Journal of Applied Geodesy 10 (1), S. 27–35. DOI: 10.1515/jag-2015-0020
- P#2** Bureick, J.; **Alkhatib, H.**; Neumann, I. (2019): Fast converging elitist genetic algorithm for optimal knot adjustment in B-spline curve approximation, In: Journal of Applied Geodesy. DOI: 10.1515/jag-2018-0015.
- P#3** Kargoll, B.; Omidalizarandi, M.; Loth, I.; Paffenholz, J.-A.; **Alkhatib, H.** (2018): An iteratively reweighted least-squares approach to adaptive robust adjustment of parameters in linear regression models with autoregressive and t-distributed deviations, Journal of Geodesy, Jg. 92, Nr. 3, S. 271-297. DOI: 10.1007/s00190-017-1062-6
- P#4** **Alkhatib, H.**; Kargoll, B.; Paffenholz, J.-A. (2017): Robust multivariate time series analysis in nonlinear models with autoregressive and t-distributed errors, in: Valenzuela, O.; Rojas, F.; Pomares, H.; Rojas, I. (Hrsg.) Proceedings ITISE 2017, Granada, September, 18–20-International work-conference on Time Series, Vol. 1, S. 23-36. ISBN: 978-84-17293-01-7
- P#5** Dorndorf A., Kargoll B., Paffenholz JA., **Alkhatib H.** (2019): A Bayesian Nonlinear Regression Model Based on t-Distributed Errors. In: . International Association of Geodesy Symposia. Springer, Berlin, Heidelberg,
- P#6** Weitkamp, A.; **Alkhatib H.** (2014): Die Bewertung kaufpreisarmer Lagen mit multivariaten statistischen Verfahren - Möglichkeiten und Grenzen robuster Methoden bei der Auswertung weniger Kauffälle. In: allgemeine vermessungs-nachrichten (avn) (1), S. 3–12.

6.2 Veröffentlichungen zur Kapitel 3

- P#7** Vogel, S., **Alkhatib, H.**, Neumann, I. (2018): Iterated Extended Kalman Filter with Implicit Measurement Equation and Nonlinear Constraints for Information-Based Georeferencing, In: 21st International Conference on Information Fusion (FUSION) 2018, Cambridge, United Kingdom, July 10-13, 2018, p. 1209-1216. DOI: 10.23919/ICIF.2018.8455258
- P#8** Vogel, S.; **Alkhatib, H.**; Bureick, Johannes; Moftizadeh, Rozhin; Neumann, Ingo (2019): Georeferencing of Laser Scanner-Based Kinematic Multi-Sensor Systems in the Context of Iterated Extended Kalman Filters Using Geometrical Constraints. In: Sensors 19 (10). DOI: 10.3390/s19102280.

- P#9** Bureick, J.; Vogel, S.; Neumann, I.; Unger, J.; **Alkhatib, H.** (2019): Georeferencing of an Unmanned Aerial System by Means of an Iterated Extended Kalman Filter Using a 3D City Model. In: PFG 2 (3), S. 429. DOI: 10.1007/s41064-019-00084-x.
- P#10** **Alkhatib, H.**; Paffenholz, J.-A.; Kutterer, H. (2012) Sequential Monte Carlo filtering for nonlinear GNSS trajectories, In: N. Sneeuw, P. Novák, M. Crespi und F. Sansò (Hg.): VII Hotine-Marussi Symposium on Mathematical Geodesy. Proceedings of the Symposium in Rome. Rome, 06.-10.06.2009. International Association of Geodesy (IAG). 1st Edition. Berlin, Heidelberg: Springer Verlag (International Association of Geodesy Symposia, 137), S. 81–86.
- P#11** **Alkhatib, H.** (2015) Alternative Nonlinear Filtering Techniques In Geodesy For Dual State And Adaptive Parameter Estimation, In: Proceedings of the 1st International Workshop on the Quality of Geodetic Observation and Monitoring Systems (QuGOMS'11), H. Kutterer, F. Seitz, H. Alkhatib, M. Schmidt, (Ed.), IAG Symp. 140, pp. 107-113. ISBN: 978-3-319-10827-8
- P#12** Sun, L.; **Alkhatib, H.**; Kargoll, B.; Kreinovich, V.; Neumann, I. (2019): Ellipsoidal and Gaussian Kalman Filter Model for Discrete-Time Nonlinear Systems, In: Mathematics 7 (12), DOI: 10.3390/math7121168.
- P#13** Sun, L.; **Alkhatib, H.**; Paffenholz, J.-A.; Neumann, I. (2018): Geo-Referencing of a Multi-Sensor System Based on Set-Membership Kalman Filter. In: International Conference on Information Fusion (Hg.): 21st International Conference on Information Fusion (FUSION 2018). Cambridge. Piscataway, NJ: IEEE, S. 1–5, DOI: 10.23919/ICIF.2018.8455763.

6.3 Veröffentlichungen zur Kapitel 4

- P#14** **Alkhatib, H.**; Neumann, I.; Kutterer, H. (2009): Uncertainty modeling of random and systematic errors by means of Monte Carlo and fuzzy techniques, In: Journal of Applied Geodesy, Jg. 3, Nr. 2/2009, S. 67–79. DOI: 10.1515/JAG.2009.008
- P#15** **Alkhatib, H.**; Kutterer, H. (2013): Estimation of Measurement Uncertainty of kinematic TLS Observation Process by means of Monte-Carlo Methods, In: Journal of Applied Geodesy, Jg. 7, Nr. 2/2013, S. 125–134. DOI: 10.1515/jag-2013-0044
- P#16** **Alkhatib, H.**; Kargoll, B.; Paffenholz, J.A. (2018): Further results on robust multivariate time series analysis in nonlinear models with autoregressive and t-distributed errors, In: Rojas, I.; Pomares, H.; Valenzuela, O. (Hrsg.) Time Series Analysis and Forecasting. ITISE 2017. Contributions to Statistics, S. 25-38. Springer, Cham. DOI: 10.1007/978-3-319-96944-2_3, ISBN: 978-3-319-96943-5
- P#17** Zhao, X.; Kargoll, B.; Omidalizarandi, M.; Xu, X.; **Alkhatib, H.** (2018): Model selection for parametric surfaces approximating 3D point clouds for deformation analysis, Remote Sensing, Jg. 10, Heft 4, Nr. 634, Special Issue: "3D Modelling from Point Clouds: Algorithms and Methods. DOI: 10.3390/rs10040634
- P#18** **Alkhatib, H.**; Omidalizarandi, M.; Kargoll, B. (2019): A Bootstrap Approach to Testing for Time-Variability of AR Process Coefficients in Regression Time Series with t-Distributed White Noise Components. In: International Association of Geodesy Symposia. Springer, Berlin, Heidelberg. DOI: 10.1007/1345_2019_78

6.4 Beiträge der Autoren

Die Rolle des Verfassers dieser Habilitationsschrift wird in drei Klassen für die Artikel angegeben:

- **Erstautorschaft:** Hauptverantwortliche Koordinierung der Arbeiten und wesentlicher Anteil an der textlichen und graphischen Ausarbeitung des Artikels, korrespondierender Autor für die Einreichung und den Begutachtungsprozess.
- **Mitautorschaft aus wissenschaftlicher Zusammenarbeit:** Beteiligung an den im Artikel beschriebenen Forschungsergebnissen durch selbständige eigene Arbeitspakete oder durch koordinierende und beratende Tätigkeit, Beteiligung an der Ausarbeitung und Revision des Artikels.
- **Mitautorschaft als Themensteller und Betreuer:** Der Erstautor oder die Erstautorin wurde als Promovierende oder Promovierender vom Verfasser dieser Habilitationsschrift bei der Durchführung der Forschungsarbeiten und bei der Verfassung des Artikels hauptverantwortlich angeleitet, einschließlich der Themenstellung.

7 Originalpublikationen

7.1 Originalpublikationen zur Kapitel 2

- P#1** Bureick, J.; **Alkhatib, H.**; Neumann, I. (2016): Robust Spatial Approximation of Laser Scanner Point Clouds by Means of Free-form Curve Approaches in Deformation Analysis, In: Journal of Applied Geodesy 10 (1), S. 27–35. DOI: 10.1515/jag-2015-0020
- P#2** Bureick, J.; **Alkhatib, H.**; Neumann, I. (2019): Fast converging elitist genetic algorithm for optimal knot adjustment in B-spline curve approximation, In: Journal of Applied Geodesy. DOI: 10.1515/jag-2018-0015.
- P#3** Kargoll, B.; Omidalizarandi, M.; Loth, I.; Paffenholz, J.-A.; **Alkhatib, H.** (2018): An iteratively reweighted least-squares approach to adaptive robust adjustment of parameters in linear regression models with autoregressive and t-distributed deviations, Journal of Geodesy, Jg. 92, Nr. 3, S. 271-297. DOI: 10.1007/s00190-017-1062-6
- P#4** **Alkhatib, H.**; Kargoll, B.; Paffenholz, J.-A. (2017): Robust multivariate time series analysis in nonlinear models with autoregressive and t-distributed errors, in: Valenzuela, O.; Rojas, F.; Pomares, H.; Rojas, I. (Hrsg.) Proceedings ITISE 2017, Granada, September, 18–20-International work-conference on Time Series, Vol. 1, S. 23-36. ISBN: 978-84-17293-01-7
- P#5** Dorndorf A., Kargoll B., Paffenholz JA., **Alkhatib H.** (2019): A Bayesian Nonlinear Regression Model Based on t-Distributed Errors. In: . International Association of Geodesy Symposia. Springer, Berlin, Heidelberg,
- P#6** Weitkamp, A.; **Alkhatib H.** (2014): Die Bewertung kaufpreisarmer Lagen mit multivariaten statistischen Verfahren - Möglichkeiten und Grenzen robuster Methoden bei der Auswertung weniger Kauffälle. In: allgemeine vermessungs-nachrichten (avn) (1), S. 3–12.

Research Article

Johannes Bureick*, Hamza Alkhatib, Ingo Neumann

Robust Spatial Approximation of Laser Scanner Point Clouds by Means of Free-form Curve Approaches in Deformation Analysis

DOI 10.1515/jag-2015-0020

received November 19, 2015; accepted December 06, 2015.

Abstract: In many geodetic engineering applications it is necessary to solve the problem of describing a measured data point cloud, measured, e. g. by laser scanner, by means of free-form curves or surfaces, e. g., with B-Splines as basis functions. The state of the art approaches to determine B-Splines yields results which are seriously manipulated by the occurrence of data gaps and outliers.

Optimal and robust B-Spline fitting depend, however, on optimal selection of the knot vector. Hence we combine in our approach Monte-Carlo methods and the location and curvature of the measured data in order to determine the knot vector of the B-Spline in such a way that no oscillating effects at the edges of data gaps occur. We introduce an optimized approach based on computed weights by means of resampling techniques. In order to minimize the effect of outliers, we apply robust M-estimators for the estimation of control points.

The above mentioned approach will be applied to a multi-sensor system based on kinematic terrestrial laser-scanning in the field of rail track inspection.

Keywords: Deformation, Free-form Curve, B-Splines, Knot Adjustment, Robust Parameter Estimation, Monte-Carlo Resampling Techniques

1 Introduction

In several geodetic applications deformations and deflections to a target-state are derived from point clouds, captured,

e. g. by laser scanner. In order to determine deformations or deflections, the spatial object has to be modelled. Especially complex objects need to be approximated by free-form curves and surfaces, such as B-Splines, in a sophisticated manner.

Unfortunately, the measurements of the deformed object may contain data gaps and outliers. The state of the art approaches to determine B-Splines yields results which are seriously manipulated by the occurrence of data gaps and outliers. Missing data lead to oscillating effects at the edges of the data gap. Outliers could have an unlimited effect on the results, if the unknown parameters (the control points) are estimated by means of the least-squares methods. Furthermore, the outliers have to be distinguished from “real” deformations and wear marks.

B-Spline fitting usually consists of 3 main steps. First step is the parameterization of the measured data. The second step is the knot adjustment, which yields the knot vector $\mathbf{U}$. The third step is the determination of the control points by means of a linear Gauss-Markov-Model (GMM) with the previously determined parameterization and knot vector as input parameters.

The parameterization of the measured data can be achieved using the mentioned methods of Piegel and Tiller [11]: equally spaced, chord length and centripetal. Lai and Lu [8] introduced an approach to estimate location parameters of the measured points which leads to a non-linear least squares fit.

Knot adjustment for data fitting with B-Splines includes two main tasks. On the one hand the number of knots has to be determined. On the other hand the locations of the knots have to be adjusted.

The former task, a model selection problem, can be solved by applying an information criterion (Akaike or Bayesian, cf. Gálvez et al. [3]) or the usage of a significance test (cf. Liu and Wang [9]).

The latter task, an optimization problem, was tackled by plenty of researchers with a vast variety of approaches. Since the first works in the field of Splines in the 1960s and 1970s the optimal choice of the knot locations became important.

*Corresponding author: Johannes Bureick, Geodetic Institute, Leibniz Universität Hannover, Nienburger Str. 1, 30167 Hannover, Germany, e-mail: bureick@gih.uni-hannover.de

Hamza Alkhatib, Ingo Neumann, Geodetic Institute, Leibniz Universität Hannover, Nienburger Str. 1, 30167 Hannover, Germany, e-mail: alkhatib@gih.uni-hannover.de, neumann@gih.uni-hannover.de

However, two problems make the optimal choice of the knot locations difficult. First of all there is no analytic expression for the optimal knot locations and secondly there exist many local optima of the least-squares function (cf. Gálvez et al. [3]; Jupp [6]; Rice [12]).

Nevertheless, there are approaches to estimate the optimal knot locations. Schmitt and Neuner [14] try to estimate the knot locations and the position of the control points at the same time. In order to solve the resulting highly non-linear system, they introduce adequate initial values and constraints.

The approaches to align the knot vector to the measured points are well known in many research papers. Piegler and Tiller [11] align it to the location parameter of the measured points. Park and Lee [10] align it to the curvature of the measured points.

With the rising capability of information technology artificial intelligence techniques obtain good results in an adequate amount of time. Some approaches use neural and functional networks, respectively. Other approaches use metaheuristic techniques like genetic algorithms (cf. Sarfraz and Raza [13]; Yoshimoto et al. [18]), artificial immune systems (cf. Gálvez et al. [3]; Ülker and Arslan [16]) or estimation of distribution algorithms (cf. Zhao et al. [19]).

As far as we know, the artificial immune system of Gálvez et al. [3] is the approach which yields the best results in knot adjustment until now, especially for complex data with gaps, discontinuities and cusps. Nevertheless this approach is still time-consuming and CPU-intensive and it is not unusual that the final solution converges into a local optimum instead of the global optimum.

The third and final step in B-Spline fitting is the estimation of the position of the control points. Piegler and Tiller [11] and Koch [7] estimate the control points as parameters by means of a linear GMM. The observation vector is formed of the measured data. The design matrix consists of the basis functions. The parameters were determined by minimizing the residual sum of squares. As far as we know there is no work, which describes the usage of robust estimators, like Huber-, Hampel or L1-estimator, instead of the least-square-estimator to determine the position of control points of a B-Spline. Because of the characteristics of laser scan data, like the fast but uncontrolled acquisition of millions of data points, we have to consider the probability for a significant amount of outliers, which seriously affects the results. That is the reason for introducing robust estimators into the estimation of the control points.

The paper is organized as follows: In Section 2 the mathematical essentials for B-Spline curves and

the estimation of the position of the control points are briefly described. Section 3 points out the basic properties of robust parameter estimation. Section 4 contains a detailed description of our proposed approach of knot adjustment. In Section 5 the different results of the proposed approach and the robust estimation are presented and compared among each other and with an existing algorithm. This paper finishes with a short conclusion and an outlook in Section 6.

2 B-Spline Curves

The functional relation of a B-Spline curve is defined by (cf. Piegler and Tiller [11]):

$$\mathbf{C}(\bar{u}) = \sum_{i=0}^n N_{i,p}(\bar{u}) \mathbf{x}_i. \quad (1)$$

In Equation 1 the curve point $\mathbf{C}(\bar{u}) = |x(\bar{u}), y(\bar{u})|^T$ is calculated by a linear combination of the p -th-degree B-Spline basis functions $N_{i,p}(\bar{u})$ with index $i \in \{0, \dots, n\}$ and the control points $\mathbf{x}_i = |x_i, y_i|^T$, where $n+1$ is the number of control points. The p -th-degree basis functions can be calculated by a recursion formula (cf. Cox [1]; de Boor [2])

$$N_{i,0}(\bar{u}) = \begin{cases} 1 & \text{if } u_i \leq \bar{u} < u_{i+1} \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

$$N_{i,p}(\bar{u}) = \frac{\bar{u} - u_i}{u_{i+p} - u_i} N_{i,p-1}(\bar{u}) + \frac{u_{i+p+1} - \bar{u}}{u_{i+p+1} - u_{i+1}} N_{i+1,p-1}(\bar{u})$$

where

$$\mathbf{U} = |u_0, \dots, u_m| \text{ with } u_i \leq u_{i+1}, i \in \{0, \dots, m-1\} \quad (3)$$

The knot vector $\mathbf{U}$ is a nondecreasing sequence of real numbers. The real numbers u_i are called knots. The first $p+1$ knot of $\mathbf{U}$ usually consist of zeros. The last $p+1$ knots usually consist of ones. $m+1$ is the size of the knot vector and can be calculated by:

$$m = n + p + 1 \quad (4)$$

The parameterization $\bar{u}$ of the (measured) data points is also called location parameter. They are stored in the vector $\mathbf{U}_l = [\bar{u}_1, \dots, \bar{u}_r]$, with r the number of the data points. For example Piegler and Tiller [11] mention 3 methods (equally spaced, chord length, centripetal) to parameterize the observations. In the proposed approach we used the chord length, which sums up the Euclidean distance

between the sorted observations, to parameterize the measured data. Since the parameterization method is not in the focus of this paper, the widely used method chord length was chosen. Nevertheless our approach allows to use the other methods.

In order to fit a B-Spline to measured data $\mathbf{l}$ the knot vector $\mathbf{U}$ and the vector with parameterized data points $\mathbf{U}_i$ are determined previously. In a linear GMM the parameter $\mathbf{x}$, the positions of the control points, are estimated. The design matrix $\mathbf{A}$ is constructed by the basis functions:

$$\mathbf{A} = \begin{bmatrix} N_{0,p}(\bar{u}_1) & \cdots & N_{n,p}(\bar{u}_1) \\ \vdots & & \vdots \\ N_{0,p}(\bar{u}_r) & \cdots & N_{n,p}(\bar{u}_r) \end{bmatrix} \quad (5)$$

3 Robust Parameter Estimation

Robust estimators include the attribute, that their influence function Ψ is limited. That means that the influence of data with large residuals (possible outliers) on the parameter estimation is limited. The M-estimators, like Huber-, Hampel- or L1- estimator, can be distinguished in their influence function. For example, the influence function of the Huber-estimator weights the residuals of the measured data in the following way (see Equation 6 and Figure 1):

$$\psi(\varepsilon_i) = \begin{cases} \varepsilon_i & \text{for } |\varepsilon_i| < c \\ \text{sign}(\varepsilon_i) \cdot c & \text{for } |\varepsilon_i| \geq c \end{cases} \quad (6)$$

For the residuals ε_i , which are smaller than the so called tuning constant c , the influence function equals the influence function of the least-squares estimator. The influence

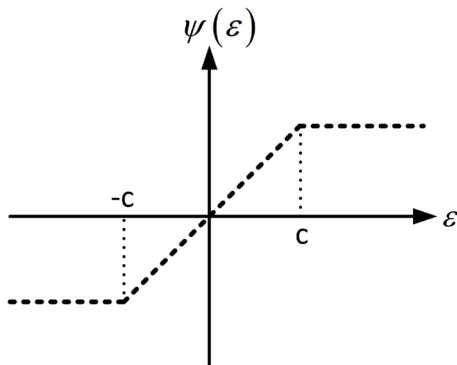


Figure 1: Influence function for Huber-estimator.

of residuals, which are larger than c , is limited to the value of c . For more information see, e. g. Hartung et al. [4] and Wicki [17].

By applying these estimators, a non-linear equation system has to be solved. This can be achieved by using the iterative reweighted least squares algorithm illustrated in Figure 2 (cf. Huber [5], p. 179 ff.). After an initial determination of the parameters $\mathbf{x}$, the residuals $\mathbf{v}$ and the variance factor σ (calculated by median absolute deviation (mad) of $\mathbf{v}$) with equal weights, the algorithm enters a while-loop which ends when the sum of absolute differences in $\mathbf{v}$ of 2 consecutive iterations is smaller than a certain threshold near 0. Until then in each iteration new weights p_i are calculated depending on $\mathbf{v}$ and σ of the previous iteration and depending on the influence function Ψ . The weights p_i are the main diagonal elements of the weight matrix $\mathbf{P}$. $\mathbf{x}$, $\mathbf{v}$ and σ are estimated iteratively by means of the updated weights.

Data : designmatrix $\mathbf{A}$;

measured data $\mathbf{l}$;

Result : parameter $\mathbf{x}$;

$$\mathbf{x}^{(0)} = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{l};$$

$$\mathbf{v}^{(0)} = \mathbf{A} \mathbf{x}^{(0)} - \mathbf{l};$$

$$\sigma^{(0)} = 1,4785 \cdot \text{mad}(\mathbf{v}^{(0)});$$

while $|\mathbf{v}^{(k+1)} - \mathbf{v}^{(k)}| \neq 0$ **do**

$$p_i^{(k+1)} = \frac{\Psi\left(\frac{v_i^{(k)}}{\sigma^{(k)}}\right)}{\frac{v_i^{(k)}}{\sigma^{(k)}}} \rightarrow \mathbf{P}^{(k+1)} = \text{diag}\left(\left[p_1^{(k+1)}, \dots, p_n^{(k+1)}\right]\right);$$

$$\mathbf{x}^{(k+1)} = (\mathbf{A}^T \mathbf{P}^{(k+1)} \mathbf{A})^{-1} \mathbf{A}^T \mathbf{P}^{(k+1)} \mathbf{l};$$

$$\mathbf{v}^{(k+1)} = \mathbf{A} \mathbf{x}^{(k+1)} - \mathbf{l};$$

$$\sigma^{(k+1)} = 1,4785 \cdot \text{mad}(\mathbf{v}^{(k)});$$

end

Figure 2: Algorithm iterative reweighted least squares.

4 Methodology of Knot Adjustment

Our methodology is illustrated in Figure 3. Before the algorithm starts, the number of control points n and the degree p of the basis function have to be chosen. As already mentioned in Section 1 this is a model selection problem which can be solved afterwards by applying an information criterion or a significance test to different solutions with a diverging number of control points or degree of basis function. This is not part of this work.

Also the maximal number of iterations $itermax$ has to be chosen. At the moment $itermax$ (in this case: 20,000) serves as stop-criterion of our algorithm.

Our algorithm offers 3 possible methods (“location”, “curvature”, “ranking”) to calculate the probability $\mathbf{R}$. All methods are described in Section 4.1. The probability calculating method pcm stores the selected method and has to be chosen previously. In case of selection of the method “ranking” the number of iterations with equally weighted $\mathbf{R}$ $iterchance$ has to be chosen (see Section 4.1). At the beginning the measured data $\mathbf{l}$ has to be parameterized (using chord length). The resulting location parameter is stored in $\mathbf{U}_l$.

In case of choosing “location” as pcm , $\mathbf{R}$ has to be calculated depending on $\mathbf{U}_l$. In case of choosing “curvature” as pcm , the curvature $\mathbf{cur}$ of the measured data has to be determined and subsequently $\mathbf{R}$ has to be obtained depending on $\mathbf{cur}$ and $\mathbf{U}_l$.

At the beginning of the following for-loop there is, in case of choosing “ranking” as pcm , an inquiry which checks if the actual number of iteration is lower or equal $iterchance$. In this case $\mathbf{R}$ is calculated equally weighted. If the actual number of iteration is larger than $iterchance$ $\mathbf{R}$ is calculated depending on $\mathbf{U}_{ranking}$.

In the following step, all $n-p$ internal knots $\mathbf{U}_{internal}$ are chosen randomly, but depending on $\mathbf{R}$.

$\mathbf{U}_{internal}$ together with the multiple start- and end-knots, has to be arranged to the complete knot vector $\mathbf{U}_{actual}$ in a non-decreasing way. For the choice of the knots, see Section 4.2.

In the next step the control points are estimated in a GMM by using $\mathbf{U}_l$ and $\mathbf{U}_{actual}$. It is possible to use a least squares estimator as well as a robust estimator. For this solution the residual sum of squares Ω_{actual} is calculated. Ω_{actual} has to be compared with the Ω stored in the ranking $\Omega_{ranking}$. When Ω_{actual} is smaller than one or more $\Omega_{ranking}$, $\mathbf{U}_{actual}$ and Ω_{actual} are stored in the ranking and the result with the highest Ω in the ranking will be deleted. These steps are repeated until $itermax$ is reached.

When $itermax$ is reached the knot vector $\mathbf{U}_{best}$ with the smallest Ω_{best} is chosen out of $\mathbf{U}_{ranking}$ and $\Omega_{ranking}$ and each internal knot of $\mathbf{U}_{best}$ is sequentially modified and stored as $\mathbf{U}_{best}$ and Ω_{best} when the resulting Ω_{actual} is smaller than Ω_{best} . After modifying each internal knot $\mathbf{U}_{best}$ and Ω_{best} are obtained and the algorithm ends.

4.1 Calculation of the Probability

As already mentioned in Section 4 and depicted in Figure 3 we introduce 3 methods to calculate the probability $\mathbf{R}$.

Data: measured data (sorted) $\mathbf{l}$;
 # control points n ;
 degree basis function p ;
 # iterations $itermax$;
 choice of probability calculating method pcm ;
 # iterations with equally weighted $\mathbf{R}$ $iterchance$
Result: optimal position of internal knots in the knot vector $\mathbf{U}_{best}$

```

initialization;
determine the location parameter  $\mathbf{U}_l$  of  $\mathbf{l}$ ;
if  $pcm = \text{"location"}$  then
  | calculate probability  $\mathbf{R}$  depending on  $\mathbf{U}_l$ 
else
  if  $pcm = \text{"curvature"}$  then
    | calculate the curvature  $\mathbf{cur}$  of  $\mathbf{l}$ ;
    | calculate  $\mathbf{R}$  depending on  $\mathbf{cur}$  and  $\mathbf{U}_l$ ;
  end
end
for  $iter = 1 : itermax$  do
  if  $pcm = \text{"ranking"}$  then
    if  $iter \leq iterchance$  then
      | calculate  $\mathbf{R}$  equally weighted
    else
      | calculate  $\mathbf{R}$  depending on  $\mathbf{U}_{ranking}$ 
    end
  end
  Choose the  $n-p$  internal knots  $\mathbf{U}_{internal}$  randomly depending on  $\mathbf{R}$ ;
  Arrange  $\mathbf{U}_{internal}$  to the knot vector  $\mathbf{U}_{actual}$ ;
  Estimate control points with a GMM depending on  $\mathbf{U}_l$  and  $\mathbf{U}_{actual}$ ;
  Calculate the residual sum of squares  $\Omega_{actual}$ ;
  if  $\Omega_{actual} < \Omega_{ranking}$  then;
    |  $\Omega_{ranking} = \Omega_{actual}$ ;
    |  $\mathbf{U}_{ranking} = \mathbf{U}_{actual}$ ;
  end
end
Choose  $\mathbf{U}_{ranking}$  with smallest  $\Omega_{ranking}$  as  $\mathbf{U}_{best}$ ,  $\Omega_{best}$ ;
for  $iter = p + 1 : n$  do
  for  $\mathbf{U}_{best}(1, iter) = 0 : 0.001$  do
    Arrange the complete knot vector  $\mathbf{U}_{actual}$ ;
    Estimate control points with a GMM depending on  $\mathbf{U}_l$  and  $\mathbf{U}_{actual}$ ;
    Calculate the residual sum of squares  $\Omega_{actual}$ ;
    if  $\Omega_{actual} < \Omega_{best}$  then
      |  $\Omega_{best} = \Omega_{actual}$ ;
      |  $\mathbf{U}_{best} = \mathbf{U}_{actual}$ 
    end
  end
end
end

```

Figure 3: Methodology of knot adjustment.

The first method “location” calculates the probability out of the parameterized location parameter $\mathbf{U}_l$ of the measured data $\mathbf{l}$.

First of all, the possible span of the internal knots (in this case the span ranged from 0 to 1, because of the parameterization of the measured data) is divided in many (in this case: 1000) parts. For each part where the mean distance to the next 2 location parameters exceeds a certain threshold (in this case: 0.02) the probability of this part is set to 0. In the other case the probability is set to 1. As a consequence the internal knots can only be chosen in areas where measured data is nearby. That means that internal knot spans are extreme unlikely to be located in data gaps, which has, due to possible singularities in the design matrix, negative effects on the appearance of the B-Spline.

The second method “curvature” calculates the probability depending on curvature values of the measured data points. The calculation is similar to the calculation of the method “location” with the difference that the probability of the parts lying under the threshold is calculated as the mean curvature of the 5 closest measured data points.

The third method “ranking” calculates the probability depending on a ranking list $\mathbf{U}_{ranking}$ of the knot vectors with the smallest sum of squares Ω . For the first iterations (in this case: $iterchance = 3000$) the whole knot span is weighted equally. Otherwise the unwished chance that the algorithm converges to a local optimum increases. The knot vectors with the smallest Ω are stored in $\mathbf{U}_{ranking}$ (in this case $\mathbf{U}_{ranking}$ consist of the top 20 knot vectors). For an iteration number larger $iterchance$ the probability is calculated depending on $\mathbf{U}_{ranking}$. For each part, where an internal knot of $\mathbf{U}_{ranking}$ is placed, the probability for that part is increased inversely proportional to Ω . As a consequence the probability for choosing parts increases, where a good solutions was achieved. In order to solve the problem that the algorithm converges to a local optimum, some knot vectors were still chosen with an equally weighted probability. Using the method “ranking”, our algorithm transforms into an evolutionary strategy.

4.2 Choice of the Knots

In order to determine knots out of the calculated probabilities $\mathbf{R}$ a resampling step, established for particle filter, is introduced (cf. Simon [15], pp. 466 f.). In step 1 $n \cdot p$ random numbers are generated uniformly distributed on $[0,1]$. In step 2 the probabilities of the 1000 parts are accumulated and stored for each part (see Figure 4).

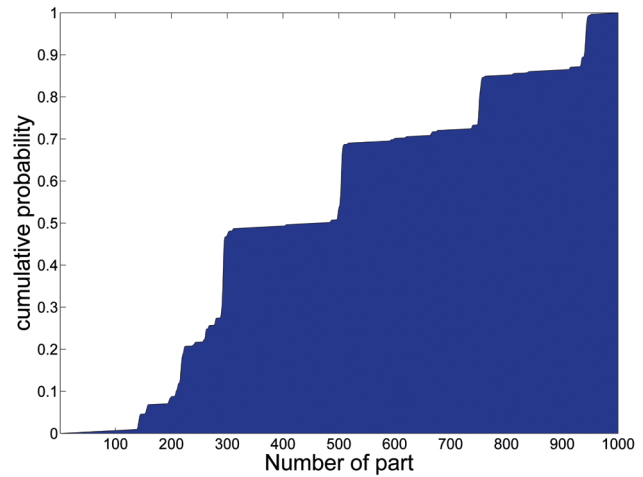


Figure 4: Cumulative probability in method “ranking” (after last iteration).

Finally, the part where the accumulated probability is greater than the randomly chosen number is chosen. That means that parts with low probabilities are unlikely to be chosen as the new internal knot.

5 Results

5.1 Knot Adjustment

In order to verify the capability of our algorithm it is applied to several test functions. Yoshimoto et al. [18] and Gálvez et al. [3] used the following functions (Equation 7–9) which represent complex data with discontinuities and cusps.

The first function represents a step function (see Figure 5):

$$\varphi_1(\omega) = \frac{90}{1 + e^{-100(\omega - 0.4)}} \quad (7)$$

The second function contains a discontinuity (comparable to a data gap, see Figure 6):

$$\varphi_2(\omega) = \begin{cases} \frac{1}{0.01 + (\omega - 0.3)^2} & \text{for } \omega < 0.6 \\ \frac{1}{0.015 + (\omega - 0.65)^2} & \text{for } \omega \geq 0.6 \end{cases} \quad (8)$$

The third function contains a cusp (see Figure 7):

$$\varphi_3(\omega) = \frac{100}{e^{|10\omega-5|}} + \frac{(10\omega-5)^5}{500} \quad (9)$$

For each test function 201 data points are generated using the Uniform distribution within the interval $U \sim [0,1]$. All data points are perturbed by an additive random noise that follows the normal distribution $N \sim [0,1]$ (cf. Gálvez et al. [3], p. 96 f.).

In the following Tables 1 to 3 Ω_{best} of the different probability calculating methods of the proposed approach are compared to Ω_{best} of the implemented clonal selection algorithm (csa) of Gálvez et al. [3]. κ is the number of internal knots and can be calculated according to Equation 10.

$$\kappa = n - p \quad (10)$$

For reason of comparability, Ω_{best} is calculated as average of 30 runs, without regarding the 5 best and worst runs (cf. Gálvez et al. [3], p. 98). The mentioned standard deviation σ is calculated out of all 30 runs.

Table 1: Ω_{best} ($\pm\sigma$) for function $\varphi_1(\omega)$ from $\kappa = 1$ to $\kappa = 7$.

κ	location	curvature	ranking	csa
1	2896.99 (± 0.00)	2896.99 (± 0.00)	2896.99 (± 0.00)	2896.99 (± 0.00)
2	651.46 (± 0.78)	651.19 (± 0.90)	606.02 (± 0.00)	606.02 (± 0.00)
3	272.25 (± 0.89)	272.39 (± 0.76)	271.53 (± 0.01)	271.09 (± 1.88)
4	229.81 (± 1.67)	229.71 (± 1.76)	228.19 (± 0.07)	225.84 (± 12.24)
5	217.61 (± 1.68)	216.45 (± 1.06)	215.74 (± 1.22)	219.24 (± 6.51)
6	172.60 (± 2.79)	171.29 (± 1.19)	170.08 (± 0.40)	169.69 (± 0.73)
7	166.86 (± 1.79)	165.60 (± 1.54)	163.51 (± 0.50)	166.03 (± 2.95)

Table 1 to 3 show that the method “ranking” always yields a smaller mean value for Ω_{best} than the methods “location” and “curvature”. In comparison to the csa the method “ranking” provides in the majority of the cases slightly smaller mean values for Ω_{best} . In the vast majority of the cases the standard deviation σ of the method “ranking” is significantly smaller than σ of the csa. Figure 5 to 7 show the solutions with the smallest Ω_{best} over 30 runs for the method “ranking”. The best results of the other methods weren’t displayed because the visual differences are too small.

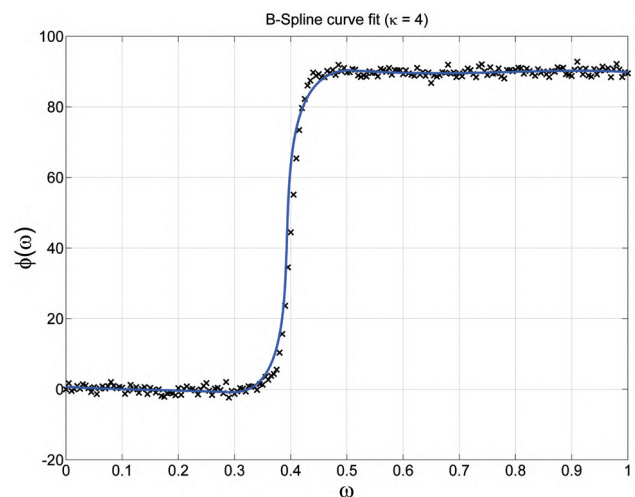


Figure 5: Best 3rd-order B-Spline fitting for function $\varphi_1(\omega)$ with method “ranking” ($\kappa = 4$).

Table 2: Ω_{best} ($\pm\sigma$) for function $\varphi_2(\omega)$ from $\kappa = 1$ to $\kappa = 17$.

κ	location	curvature	ranking	csa
1	43262.28 (± 0.00)	43262.28 (± 0.00)	43262.28 (± 0.00)	43262.28 (± 0.00)
2	23595.93 (± 1.12)	23595.72 (± 0.27)	23595.57 (± 0.00)	23595.57 (± 161.74)
3	1771.39 (± 20.18)	1747.55 (± 11.84)	1643.68 (± 0.00)	1643.69 (± 0.01)
4	1050.54 (± 23.47)	1038.23 (± 27.82)	998.13 (± 6.96)	995.18 (± 2.86)
5	834.02 (± 33.83)	818.52 (± 16.75)	772.96 (± 11.82)	778.14 (± 40.79)
6	466.34 (± 94.50)	438.82 (± 69.66)	354.27 (± 6.88)	424.73 (± 160.59)
7	207.52 (± 28.80)	220.28 (± 24.51)	157.59 (± 5.18)	153.52 (± 7.98)
8	162.83 (± 14.25)	166.93 (± 16.26)	139.54 (± 3.80)	141.07 (± 7.11)
9	137.11 (± 8.30)	141.92 (± 10.77)	124.49 (± 3.45)	132.51 (± 9.32)
10	125.14 (± 6.49)	131.93 (± 9.05)	112.65 (± 5.19)	124.10 (± 9.01)
11	115.70 (± 7.04)	122.42 (± 8.50)	103.17 (± 3.87)	118.86 (± 9.08)
12	108.01 (± 5.90)	113.30 (± 6.34)	99.64 (± 1.43)	112.75 (± 10.71)
13	103.60 (± 4.30)	106.93 (± 6.60)	97.03 (± 1.48)	104.39 (± 10.22)
14	97.23 (± 3.10)	101.35 (± 5.25)	93.22 (± 3.15)	96.55 (± 6.23)
15	94.39 (± 2.89)	95.35 (± 3.72)	90.39 (± 3.62)	93.26 (± 8.06)
16	91.49 (± 3.11)	93.56 (± 3.94)	85.04 (± 4.11)	90.94 (± 6.11)
17	89.47 (± 3.95)	89.45 (± 3.47)	82.21 (± 3.85)	87.44 (± 5.06)

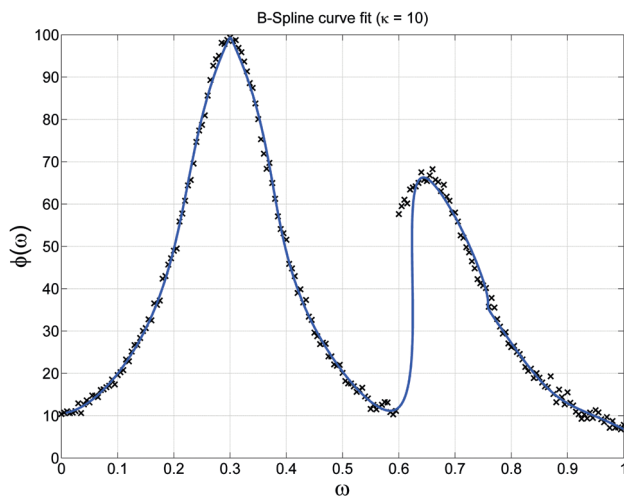


Figure 6: Best 3rd-order B-spline fitting for function $\varphi_2(\omega)$ with method “ranking” ($\kappa = 10$).

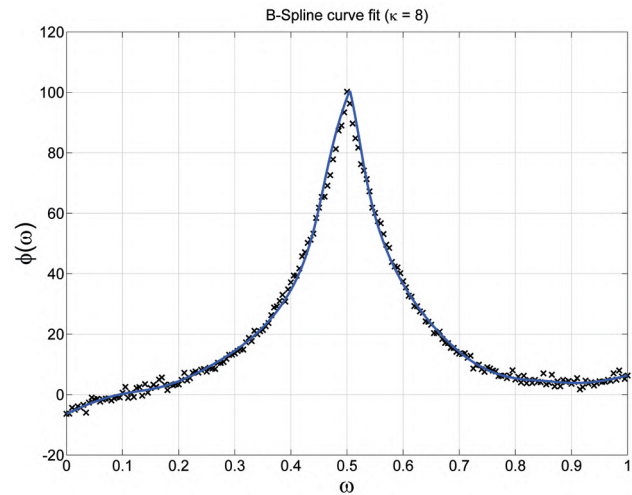


Figure 7: Best 3rd-order B-spline fitting for function $\varphi_3(\omega)$ with method “ranking” ($\kappa = 8$).

Table 3: $\Omega_{\text{best}} (\pm\sigma)$ for function $\varphi_3(\omega)$ from $\kappa = 1$ to $\kappa = 17$.

κ	location	curvature	ranking	csa
1	4912.60 (± 0.01)	4912.59 (± 0.00)	4912.55 (± 0.00)	4912.55 (± 0.00)
2	2672.91 (± 0.15)	2672.93 (± 0.25)	2672.87 (± 0.00)	2672.87 (± 0.00)
3	594.68 (± 1.03)	594.39 (± 0.87)	593.56 (± 0.05)	593.57 (± 1.73)
4	441.67 (± 3.69)	439.69 (± 2.59)	438.07 (± 8.34)	437.60 (± 10.45)
5	262.24 (± 10.10)	258.80 (± 6.57)	248.95 (± 0.26)	242.85 (± 1.49)
6	231.87 (± 8.23)	226.09 (± 4.85)	215.02 (± 0.72)	221.91 (± 7.19)
7	205.72 (± 6.19)	202.51 (± 5.33)	191.65 (± 0.82)	199.57 (± 10.99)
8	185.58 (± 6.78)	186.30 (± 5.07)	168.72 (± 1.41)	177.19 (± 10.87)
9	171.85 (± 3.84)	170.65 (± 4.45)	161.09 (± 0.57)	164.67 (± 10.25)
10	161.70 (± 4.24)	162.02 (± 3.77)	150.88 (± 1.18)	154.83 (± 6.82)
11	155.04 (± 3.36)	155.04 (± 4.40)	145.52 (± 2.20)	149.20 (± 5.15)
12	148.17 (± 2.92)	149.29 (± 3.95)	142.10 (± 1.01)	142.31 (± 5.28)
13	145.84 (± 2.07)	146.21 (± 2.55)	138.26 (± 1.93)	142.04 (± 4.89)
14	143.15 (± 4.09)	143.15 (± 3.87)	128.90 (± 3.48)	138.55 (± 5.05)
15	139.71 (± 3.96)	141.00 (± 3.76)	123.14 (± 1.83)	133.35 (± 6.70)
16	132.76 (± 4.41)	133.76 (± 4.67)	121.37 (± 1.18)	128.22 (± 7.61)
17	128.99 (± 4.00)	127.85 (± 4.29)	118.52 (± 2.98)	126.58 (± 6.85)

That means the results of the method “ranking” are more stable than the results of the csa, which converge, in a not negligible amount of runs, into a local optimum instead of the global optimum. Especially for large κ our algorithm yields better results than the csa.

5.2 Robust Parameter Estimation

The results shown in the previous Section 5.1 are obtained by a least squares estimation of the control points. Due to the fact that the test functions are perturbed by a normally distributed noise that is sufficient. In order to check the performance of robust estimation, we generated a point cloud of 911 points out of the desired values for a rail track. These desired values or true values, respectively, are stored in the vector $\bar{\mathbf{I}}$. Again $\bar{\mathbf{I}}$ is perturbed by a normally distributed noise $N \sim [0, 0.067]$, which conforms with the data sheet of a usual profile scanner for rail track inspection. Additionally, 3 outliers are inserted, arranged about 4 mm above the desired values. The resulting points are stored in $\mathbf{I}$. Figure 8 shows the comparison between an estimation with least squares and the estimation using a robust estimator, in this case the Huber-estimator (with the tuning constant $c = 1$). Both results are obtained using the method “ranking” for knot adjustment. In the area of the 3 outliers the least squares estimation is distorted in the direction of the outliers, whereas the fit using the Huber-estimator stays closer to the data without outliers.

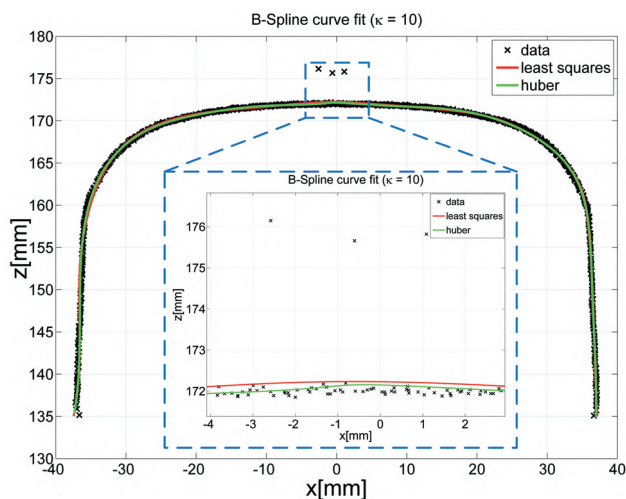


Figure 8: Comparison estimation with least squares and Huber.

In order to validate that visual result, we calculated the residual sum of squares of the estimated parameters to the “measured” values $\mathbf{l}$ (Ω) and to the true values $\tilde{\mathbf{l}}$ ($\tilde{\Omega}$).

$$\begin{aligned}\mathbf{v} &= \mathbf{A} \cdot \hat{\mathbf{x}} - \mathbf{l} \rightarrow \Omega = \mathbf{v}^T \mathbf{v} \\ \tilde{\mathbf{v}} &= \mathbf{A} \cdot \hat{\mathbf{x}} - \tilde{\mathbf{l}} \rightarrow \tilde{\Omega} = \tilde{\mathbf{v}}^T \tilde{\mathbf{v}}\end{aligned}\quad (11)$$

Table 4: Comparison of least squares and Huber-estimation.

	least squares	Huber
Ω	86.0781	86.9692
$\tilde{\Omega}$	43.6250	41.8616

Table 4 shows that the estimation using the Huber-estimator has a smaller $\tilde{\Omega}$ thus the effect of the outliers is lower than using the least squares estimation.

6 Conclusion and Outlook

In this paper we introduced an algorithm which determines the knot vector of a B-Spline with a mixture of Monte-Carlo methods and an evolutionary algorithm and simultaneously is robust against outliers. The results of knot adjustment are slightly better than the results of comparable algorithms. Especially for an increasing number of internal knots our algorithm yields better results. Furthermore the proposed algorithm produces more stable results, because the deviation of the results is significantly smaller. We also showed that robust parameter estimation for

B-Splines obtains good results and is essential in case of an outlier-contaminated point cloud.

Beside the extension of our approach on B-Spline surfaces, the input and tuning parameters, as the chosen assumptions at probability calculation (e.g. size of the ranking, thresholds and partitioning), are going to be implemented in a closed loop simulation. Especially, the extension with respect to a more sophisticated introduction of prior knowledge is planned.

Acknowledgement: A large part of the shown results were developed in the research project “Advanced Rail Track Inspection System (ARTIS)” supported by the Federal Ministry for Economic Affairs and Energy on the basis of a decision by the German Bundestag.

References

- [1] Cox, M. G., 1972. The Numerical Evaluation of B-Splines. *IMA J Appl Math* 10 (2), 134–149. doi:10.1093/imamat/10.2.134.
- [2] de Boor, C., 1972. On calculating with B-splines. *Journal of Approximation Theory* 6 (1), 50–62. doi:10.1016/0021-9045(72)90080-9.
- [3] Gálvez, A., Iglesias, A., Avila, A., Otero, C., Arias, R., Manchado, C., 2015. Elitist clonal selection algorithm for optimal choice of free knots in B-spline data fitting. *Applied Soft Computing* 26, 90–106. doi:10.1016/j.asoc.2014.09.030.
- [4] Hartung, J., Elpelt, B., Klösener, K.-H., 2009. *Statistik: Lehr- und Handbuch der angewandten Statistik*, 15th ed. Oldenbourg, München, XXXIII, 1145.
- [5] Huber, P. J., 1981. *Robust statistics*. Wiley, New York, ix, 308.
- [6] Jupp, D. L., 1978. Approximation to Data by Spline with Free Knots. *SIAM Journal on Numerical Analysis* 15 (2), 328–343.
- [7] Koch, K.-R., 2009. Fitting Free-Form Surfaces to Laserscan Data by NURBS. *Allgemeine Vermessungs-Nachrichten (AVN)* (4), 134–140.
- [8] Lai, J. -Y., Lu, C. -Y., 1996. Reverse Engineering of Composite Sculptured Surfaces. *Int J Advanced Manufacturing Technology* (12), 180–189.
- [9] Liu, A., Wang, Y., 2004. Hypothesis testing in smoothing spline models. *Journal of Statistical Computation and Simulation* 74 (8), 581–597. doi:10.1080/00949650310001623416.
- [10] Park, H., Lee, J. -H., 2007. B-spline curve fitting based on adaptive curve refinement using dominant points. *Computer-Aided Design* 39 (6), 439–451. doi:10.1016/j.cad.2006.12.006.
- [11] Piegl, L., Tiller, W., 1997. *The NURBS book*, 2nd ed. Springer, Berlin, New York, xiv, 646.
- [12] Rice, J. R., 1969. *The Approximation of Function*, 2nd ed. Addison-Wesley Publishing Company, Reading, MA.
- [13] Sarfraz, M., Raza, S. A., 2001. Capturing outline of fonts using genetic algorithm and splines, in: *Proceedings. IEEE Computer Soc., Los Alamitos, Calif. [et al.]*, pp. 738–743.

- [14] Schmitt, C., Neuner, H., 2015. Knot estimation on B-Spline curves. *Österreichische Zeitschrift für Vermessung & Geoinformation* 103 (2 + 3), 188–197.
- [15] Simon, D., 2006. Optimal state estimation: Kalman, H [infinity] and nonlinear approaches. Wiley-Interscience, Hoboken, N. J., xxvi, 526.
- [16] Ülker, E., Arslan, A., 2009. Automatic knot adjustment using an artificial immune system for B-spline curve approximation. *Information Sciences* 179 (10), 1483–1494. doi:10.1016/j.ins.2008.11.037.
- [17] Wicki, F., 1998. Robuste Schätzverfahren für die Parameterschätzung in geodätischen Netzen. Institut für Geodäsie und Photogrammetrie, Eidgenössische Technische Hochschule, Zürich, 185.
- [18] Yoshimoto, F., Moriyama, M., Harada, T., 1999. Automatic knot placement by a genetic algorithm for data fitting with a spline, in: *Proceedings of the international conference on shape modeling and applications*, pp. 162–169.
- [19] Zhao, X., Zhang, C., Yang, B., Li, P., 2011. Adaptive knot placement using a GMM-based continuous optimization algorithm in B-spline curve approximation. *Computer-Aided Design* 43 (6), 598–604. doi:10.1016/j.cad.2011.01.015.

Johannes Bureick*, Hamza Alkhatib, and Ingo Neumann

Fast converging elitist genetic algorithm for knot adjustment in B-spline curve approximation

<https://doi.org/10.1515/jag-2018-0015>

Received April 16, 2018; accepted August 7, 2019

Abstract: B-spline curve approximation is a crucial task in many applications and disciplines. The most challenging part of B-spline curve approximation is the determination of a suitable knot vector. The finding of a solution for this multimodal and multivariate continuous nonlinear optimization problem, known as knot adjustment problem, gets even more complicated when data gaps occur. We present a new approach in this paper called an elitist genetic algorithm, which solves the knot adjustment problem in a faster and more precise manner than existing approaches. We demonstrate the performance of our elitist genetic algorithm by applying it to two challenging test functions and a real data set. We demonstrate that our algorithm is more efficient and robust against data gaps than existing approaches.

Keywords: B-spline curve, approximation, knot adjustment, genetic algorithm, data gaps, Monte Carlo

1 Introduction

The approximation of point clouds by means of free-form curves (and surfaces) has been an important task in reverse engineering for the past few decades. The applications in automotive, computer vision or medical imaging are well-known and extensively described.

Many of the point clouds are captured by a laser scanner, which has proved to be a useful observation sensor for a variety of emerging applications in geodesy and other disciplines, such as constructional engineering. The data has been used, for instance, for three-dimensional (3D) surface modelling by means of NURBS and B-splines [1, 2, 3]. Laser scanning has also been used for the monitoring of different structures, such as rails, arches or tunnels [4, 5, 6, 7].

***Corresponding author: Johannes Bureick**, Geodetic Institute, Leibniz University Hannover, Nienburger Str. 1, 30167 Hannover, Germany, e-mail: bureick@gih.uni-hannover.de

Hamza Alkhatib, Ingo Neumann, Geodetic Institute, Leibniz University Hannover, Nienburger Str. 1, 30167 Hannover, Germany, e-mails: alkhatib@gih.uni-hannover.de, neumann@gih.uni-hannover.de

However, laser scanners provide point clouds, which are affected by unavoidable noise and which may contain data gaps. Furthermore, the object measured may exhibit a complex shape by means of cusps or leaps.

Free-form curves, especially B-spline curves, are basically well-suited to deal with such challenges in the approximation process. In addition to a smoothing of the data, B-spline curves may adapt easily to complex objects without the need to change the general functional relation. The resulting B-spline curve needs considerably less memory space than the raw point cloud and can be uploaded easily into computer-aided design software for further analysis steps.

However, some challenges have to be tackled to provide a good approximation by a B-spline curve. The placement of the knots in B-spline approximation, known as knot adjustment problem, which is a multimodal and multivariate continuous nonlinear optimisation problem [8, 9], could have a very considerable effect on the final estimation of the curve and surface.

Due to the lack of an analytic expression for optimal knot locations, different methodologies in the specialized literature have been demonstrated for the selection and optimisation of knot vectors. Some fast deterministic methods employ, for instance, the position [10] or the curvature [11] of the point cloud to determine the knot vector. In the case of complex point clouds, the results are far away from the optimum. Alternatively, metaheuristic methods, such as artificial immune systems (AIS) [9], genetic algorithms (GA) [12] or estimation of distribution algorithms (EDA) [13], yield knot vectors which are very close to the optimum, but only converge slowly and are, therefore, time- and computing power-consuming. Furthermore, the performance of these algorithms is seriously affected by the occurrence of data gaps.

In this paper, we present a new method, the elitist genetic algorithm (EGA), which is designed to solve the knot adjustment problem for B-spline curves in a fast and more precise manner, despite the possible occurrence of data gaps. Although not shown within this paper, the idea of the EGA can easily be transferred to B-spline surfaces.

The content of this paper is structured as follows. In Chapter 2, we depict the mathematical basics for B-spline curves and give an overview of related work in B-spline curve approximation. In Chapter 3, we describe explicitly

an existing method and our proposed EGA for knot vector determination. Both methods are applied to three data sets and discussed in Chapter 4. The paper closes with the conclusion and an outlook in Chapter 5.

2 B-spline curve approximation

A 3D curve point $\mathbf{C}(\bar{u})$ lying on a B-spline curve is calculated by the sum of the linear combination of the basis functions $N_{i,p}(\bar{u})$ and the 3D control points $\mathbf{x}_i$:

$$\mathbf{C}(\bar{u}) = [x(\bar{u}), y(\bar{u}), z(\bar{u})]^T = \sum_{i=0}^n N_{i,p}(\bar{u}) \mathbf{x}_i \text{ with } \mathbf{x}_i = [x_i, y_i, z_i]^T. \quad (1)$$

Formally $n+1$ linear combinations and, therefore, the same amount of control points and basis functions contribute to $\mathbf{C}(\bar{u})$. p denotes the degree of the basis function and $p+1$ denotes the order of the B-spline curve. Practically only $p+1$ control points and basis functions contribute to a certain curve point $\mathbf{C}(\bar{u})$. This feature is called local support (see [10]). The location parameter $\bar{u}$ specifies the position of the curve point on the B-spline curve. The first location parameter is usually defined by 0 and the last location parameter is defined by 1. We adopt this regulation schema in this paper.

Cox [14] and de Boor [15] introduced a recursive function to calculate the basis functions $N_{i,p}(\bar{u})$. In addition to the degree of the basis function p , the number of control points $n+1$ and the location parameter $\bar{u}$, the knots u_i with $i \in \{0, \dots, m\}$ affect the calculation of the basis functions. The $m+1$ knots are stored in the knot vector $\mathbf{U}$ in a non-decreasing sequence, whereby the knots can be separated into external and internal knots. The first $p+1$ knots contain the value 0, the last $p+1$ knot contain the value 1. These knots are called external knots. All other knots are called internal knots.

When B-spline curves are applied to approximate a 3D (dimension $d=3$) point cloud $\mathbf{I}$:

$$\mathbf{I}_{[r,d=3]} = \begin{bmatrix} l_{x,1} & l_{y,1} & l_{z,1} \\ \vdots & \vdots & \vdots \\ l_{x,r} & l_{y,r} & l_{z,r} \end{bmatrix} \quad (2)$$

consisting of r points, four steps usually have to be realised: *model selection*, *parameterisation*, *knot vector determination* and *control point estimation*. Because of the focus of this paper, only the different methodologies for *knot vector determination* will be discussed in detail in the following.

Model selection comprises the selection of proper values for the degree of basis function p and number of control points $n+1$. For more information, please refer to e. g. [9, 12, 16, 17].

Parameterisation deals with the determination of a location parameter $\bar{u}$ for every point in the point cloud. Proper methods for this step can be found in e. g. [10, 18, 19, 20].

Knot vector determination has an important and considerable effect on the estimating results of the B-spline curve. Simultaneously, it is also the most critical step in B-spline approximation.

Some works in the 1960s and 1970s showed that B-spline curves can significantly be improved if the knots were optimised during the approximation process (e. g. [21, 22]). However, the optimisation of the knots leads to a multimodal, multivariate, continuous, non-linear optimisation problem, known as knot adjustment problem. The multimodality of the knot adjustment problem enables that totally different knot vectors may lead to almost identical results, whereby small differences in the knot vectors may cause totally different results.

The occurrence of data gaps makes the knot adjustment problem even more difficult, because the placement of $p+1$ knots in the data gap causes singularities in the *control point estimation*. A lot of researchers tackled the knot adjustment problem, and selected works are briefly depicted in the following.

In many cases, deterministic approaches are used to determine the knot vector. These approaches do not really try to solve the knot adjustment problem, but they try to yield a good knot vector, which leads to well-conditioned matrices for *control point estimation* afterwards and to a smooth B-spline curve. These approaches are mostly sufficient for smooth point clouds and work fast because of their deterministic condition. Here it is mentionable that the deterministic approaches given in [10, 18, 23] ensure that at least one location parameter lies in every knot span, the area between two knots. A different deterministic approach presented by Park and Lee [11] employs the location parameters of previously identified dominant points as knots.

However, the deterministic approaches mentioned are not well-suited to approximate point clouds with cusps, leaps or discontinuities ([9]). The reason is the lack to generate multiple knots. Multiple knots are identical knots which cause a reduction of continuity at that position. The number of knots with the same value is also called multiplicity k . A B-spline curve is C^{p-k} -continuous, which means, for example, that a cubic B-spline curve with three identical knots is C^0 -continuous at that position (see [10],

p. 88f.). Consequently, the approximation of cusps or data gaps requires multiple knots. Otherwise, $k > p$ multiple knots cause singularities in the *control point estimation*.

Artificial intelligence techniques have emerged with the rising capability of information technology and have obtained good results. Metaheuristic approaches have especially resumed to solve the knot adjustment problem. There are different kinds of approaches which were divided roughly into the EDA, GA and AIS, whereby the last two are closely related. All methods apply a kind of evolutionary strategy, often adapted from real natural processes, to obtain the optimal knot vector.

EDA uses the best knot vectors of the previous iteration to estimate a distribution for the knot locations. Two EDA approaches can be found in Zhao et al. [13] and Bureick et al. [4].

GA imitate the evolution of creatures. To our knowledge the first GA dealing with knot vector determination, was presented by Yoshimoto et al. [16]. In [12] the authors presented an adapted version of the GA. A further application can be found in Sarfraz and Raza [24]. Nevertheless, the GA does not provide optimal knot vectors compared to other metaheuristic approaches. The reason for that may be the mutation operator, which completely deletes and inserts random knots in the knot vector. Applying B-spline approximation on various data sets, we noticed that it is better to mutate the knots by small shifts instead of replacing it by a uniformly distributed new knot.

AIS approaches simulate the immune system of human beings. Two approaches can be found in Ülker and Arslan [25] and Gálvez et al. [9]. The latter approach is called elitist clonal selection algorithm (CSA) and performs very well. We think that the main reason for the good performance of the CSA is located in the mutation scheme applied. In the mutation scheme, only one randomly chosen knot value is shifted by a small random value, whose magnitude is controlled by the distance to the external knots. The CSA has the drawback that it sometimes converges into a local optimum and the resulting knot vector is far away from the optimum. Furthermore, this approach frequently fails when the point cloud contains larger data gaps, because the algorithm may consistently place more than p knots inside the data gap and therefore cause singularities in control point estimation. The CSA is extensively described in Chapter 3.1.

In the scope of this paper, we pick up the drawbacks of the CSA and present the EGA, a new method to determine the knot vector in a fast and precise manner, which is robust against the occurrence of data gaps. The EGA is an enhancement of the CSA and extensively described in Chapter 3.2.

In *control point estimation*, as the final step of B-spline curve approximation, the control points are optimally estimated, as the unknown parameters $\mathbf{x}$ in a linear Gauss Markov model (GMM):

$$\mathbf{l} = \mathbf{A} \cdot \mathbf{x} + \boldsymbol{\epsilon}. \quad (3)$$

The design matrix $\mathbf{A}$ of this GMM is constructed by the basis functions:

$$\mathbf{A}_{[r,n+1]} = \begin{bmatrix} N_{0,p}(\bar{u}_1) & \cdots & N_{n,p}(\bar{u}_1) \\ \vdots & & \vdots \\ N_{0,p}(\bar{u}_r) & \cdots & N_{n,p}(\bar{u}_r) \end{bmatrix}. \quad (4)$$

The estimates for the control points $\hat{\mathbf{x}}_{[n+1,d]}$ are then approximated in a least square sense by minimizing the weighted squared errors $\boldsymbol{\epsilon}$ (e. g. [10]), representing the distances between point cloud and estimated points on the B-spline curve, which yields to:

$$\hat{\mathbf{x}}(:, i) = (\mathbf{A}^T \mathbf{P}(:, :, i) \mathbf{A})^{-1} \mathbf{A}^T \mathbf{P}(:, :, i) \mathbf{l}(:, i) \text{ with } i \in \{1, \dots, d\}. \quad (5)$$

The dimension of the control point matrix equals the dimension d of the point cloud (see Eq. 2). $\mathbf{P}(:, :, i)$ represents a weight matrix for the i -th coordinate component of the point cloud. In the case of identical and independent normally distributed errors $\boldsymbol{\epsilon}$, $\mathbf{P}(:, :, i)$ is given by the identity matrix $\mathbf{I}$.

A requirement for the application of the linear GMM is that the steps *model selection*, *parameterisation* and *knot vector determination* are realised a priori.

3 Approaches for knot vector determination

In the following, we will introduce and extensively describe our new algorithm, the EGA (see Chapter 3.2). Since the EGA is an enhancement of CSA, regarding efficiency and precise estimation, we will first recapitulate the CSA, demonstrated in Gálvez et al. [9] (see Chapter 3.1).

In these chapters we will frequently use the term “individual”. One individual represents all internal knots of a single knot vector.

3.1 Elitist clonal selection algorithm

Gálvez et al. [9] solve the knot adjustment problem, which needs a powerful global optimization method, by means of evolutionary computation techniques. Key details in their

Data: Number of iterations $iter_{max}$;

Number of individuals N ;

Number of individuals to be selected N_b ;

Number of elitist individuals N_e ;

Number of random individuals N_r ;

Result: Fittest individual I_{best} ;

Generate randomly N individuals and allocate them to $\mathcal{P}_N$;

Calculate the fitness value Ω_i for each individual in $\mathcal{P}_N$ (B.1);

Set run variable $iter = 1$;

while $iter \leq iter_{max}$ **do**

 Select the N_b individuals of $\mathcal{P}_N$ with smallest Ω_i ;

 Clone the N_b individuals according their Ω_i and allocate them to $\mathcal{P}_M$

 (B.2);

 Perform mutation in $\mathcal{P}_M$ (B.3);

 Calculate Ω_i for each individual in $\mathcal{P}_M$;

 Select the N_e individuals of $\mathcal{P}_N$ with smallest Ω_i and allocate them to

$\mathcal{P}_e$ (B.4);

 Clear $\mathcal{P}_N$;

 Select the $N - (N_e + N_r)$ individuals of $\mathcal{P}_M$ with smallest Ω_i and allocate them to $\mathcal{P}_N$;

 Allocate $\mathcal{P}_e$ to $\mathcal{P}_N$ (B.4);

 Randomly generate N_r individuals, calculate their Ω_i and allocate

 them to $\mathcal{P}_N$ (B.5);

$iter = iter + 1$

end

Select the individual I_{best} with smallest Ω_i of $\mathcal{P}_N$;

Algorithm 1: Pseudocode of the CSA.

research will be summarized in this chapter. The CSA proposed in Gálvez et al. [9] is depicted in Algorithm 1. The key steps are tagged with a circle around a letter-number combination and described in detail.

The CSA starts with a random generation of N individuals according uniform distribution. The authors in Gálvez et al. [9] apply several fitness functions to evaluate the individuals. In addition to the sum of squared residuals, the root mean square error and two information criteria (AIC and BIC) are applied for model selection. Since we do not focus on model selection, we apply the sum of squared residuals in step B.1.

The matrix of residuals $\hat{\mathbf{e}}$ is calculated by Eq. 6.

$$\hat{\mathbf{e}} = \mathbf{I} - \mathbf{A} \cdot \hat{\mathbf{x}} \quad (6)$$

The sum of squared residuals Ω is calculated by Eq. 7.

$$\Omega = \sum_{i=1}^d \hat{\mathbf{e}}(:, i)^T \cdot \mathbf{P}(:, :, i) \cdot \hat{\mathbf{e}}(:, i) \quad (7)$$

Gálvez et al. [9] presented a fitness proportionate technique for selection and cloning in step B.2. Therefore, a certain number of individuals N_b with the best fitness value Ω_i are selected and increasingly ordered by their Ω_i for the cloning process. These individuals are replicated

based on their fitness value Ω_i . The number of clones for each individual $N_{ind,i}$ is calculated by (see [9], p. 95, Eq. 9):

$$N_{ind,i} = \text{round}[(v_i + 1) \cdot 0.1 \cdot N]. \quad (8)$$

Gálvez et al. [9] specified the vector $\mathbf{v}$ by:

$$\mathbf{v}_{[N_b, 1]} = [5, 3, 1, 0, \dots, 0]^T \quad (9)$$

in order to achieve higher cloning rates for the individuals with a good fitness value. The total number of clones N_c is calculated by:

$$N_c = \sum_{i=1}^{N_b} N_{ind,i}. \quad (10)$$

The mutation technique proposed for the CSA in step B.3 differs from traditional mutation operators, because it does not discard an old knot u_i and generate a new random knot u_i^* from uniform distribution. It shifts the old knot u_i by a small random value by the mechanism presented in Gálvez et al. ([9], p. 95):

$$u_i^* = u_i - \left(\frac{\Delta}{2}\right)(\sigma - 0.5) \text{ with } \Delta = \min\left\{\frac{u_i}{2}, 1 - u_i\right\} \quad (11)$$

where σ is a random value chosen from a uniform distribution $\mathcal{U}[0, 1]$. Although the perturbation is limited to a small value, it is possible and not unfavourable that the order of knots might change. Therefore, the knot vector with the new knot u_i^* has to be assorted. Additionally, the applied mutation technique ensures that the new knots u_i^* are located inside the interval $[0, 1]$, by taking into account the closest distance Δ to one of the interval boundaries.

In the CSA (step B.4), it is proposed to transfer the fittest N_e individuals of each iteration unchanged into the next iteration. This feature, called elitism, ensures that a good solution will not be randomly discarded during the iterations. Furthermore, a certain number of individuals N_r per iteration is chosen from uniform distribution $\mathcal{U}[0, 1]$ in step B.5. That feature should avoid the convergence of the algorithm to a local optimum by enhancing the exploration ability.

Finally, the individual with the best fitness value I_{best} is selected from the population $\mathcal{P}_N$.

3.2 Elitist genetic algorithm

The idea for the proposed EGA arose by comparing the CSA of Gálvez et al. [9] and the GA of Yoshimoto et al. [12]. It is worth mentioning that both algorithms focused on the determination of knot location and knot vector size, which

Table 1: Key features of CSA and GA.

	CSA	GA
selection	fitness proportionate	tournament
crossover	–	two-point
mutation	small random shift	replacement

means that they try to realise model selection and knot vector determination simultaneously. By contrast our proposed EGA focuses solely on the knot vector determination. Nevertheless, an extension to model selection is possible. The CSA and the GA are closely related, whereas the CSA provides better results. If we consider the CSA as a genetic algorithm, its key features would be those presented in Table 1. The key features of the GA are also depicted for comparison.

The CSA is obviously not designated with crossover, but, especially for complex objects, crossover is a further suitable tool to obtain good results in an adequate amount of time. The reason therefore is that crossover allows the combination of locally well approximating knot vector parts and hence increases the diversity of results.

Therefore, our proposed EGA is similar to the CSA, but allows the occurrence of crossover. Furthermore, the mutation is slightly changed and the initialisation is different to apply the EGA to point clouds with larger data gaps.

The processing chain of the EGA is summarised in Algorithm 2.

The fitness function (see **B.1** in Chapter 3.1), selection and cloning (see **B.2** in Chapter 3.1), elitism and exploration (see **B.4** and **B.5** in Chapter 3.1) in the EGA are equivalent to the CSA. We tagged these steps (**B.1**, **B.2**, **B.4** and **B.5**) and the additional or modified steps of the EGA (**C.1**–**C.3**) with a circle around a letter-number combination. In the following, only the additional or modified steps of the EGA are described in detail.

The initialisation in step **C.1** of the EGA has been slightly changed over the CSA to deal with the possible occurrence of data gaps. Whereas all N individuals in the CSA are generated randomly in the initialisation, only $N-1$ individuals are generated randomly in the EGA. The last individual is calculated by the deterministic method from Piegł and Tiller [23]. This deterministic calculation ensures that at least one individual avoids singularities and has a Ω_i which is a real number.

For crossover in step **C.2** there are several possible techniques which could be applied (for more information, see for instance [26]). We investigated single point, two point and an adapted version of uniform crossover, sub-

Data: Number of iterations $iter_{max}$;

Number of individuals N ;

Number of individuals to be selected N_b ;

Number of elitist individuals N_e ;

Number of random individuals N_r ;

Rate of random pointwise crossover r_{rc} ;

Result: Fittest individual I_{best} ;

Randomly generate $N-1$ individuals and allocate them to $\mathcal{P}_N$;

Generate 1 individual with deterministic method and allocate it to $\mathcal{P}_N$

(C.1);

Calculate the fitness value Ω_i for each individual in $\mathcal{P}_N$ **(B.1)**;

Set run variable $iter = 1$;

while $iter \leq iter_{max}$ **do**

 Select the N_b individuals of $\mathcal{P}_N$ with smallest Ω_i ;

 Clone the N_b individuals according their Ω_i and allocate them to $\mathcal{P}_c$

(B.2);

 Randomly relocate $r_{rc} \cdot N_c$ of $\mathcal{P}_c$ to $\mathcal{P}_{rc}$;

 Perform random pointwise crossover in $\mathcal{P}_{rc}$ **(C.2)**;

 Allocate $\mathcal{P}_{rc}$ and the remaining individuals of $\mathcal{P}_c$ to $\mathcal{P}_M$;

 Perform mutation in $\mathcal{P}_M$ **(C.3)**;

 Calculate Ω_i for each individual in $\mathcal{P}_M$;

 Select the N_e individuals of $\mathcal{P}_N$ with smallest Ω_i and allocate them to

$\mathcal{P}_e$ **(B.4)**;

 Clear $\mathcal{P}_N$;

 Select the $N - (N_e + N_r)$ individuals of $\mathcal{P}_M$ with smallest Ω_i and allocate them to $\mathcal{P}_N$;

 Randomly generate N_r individuals, calculate their Ω_i and allocate them to $\mathcal{P}_N$ **(B.5)**;

 Allocate $\mathcal{P}_e$ to $\mathcal{P}_N$ **(B.4)**;

$iter = iter + 1$

end

Select the individual I_{best} with smallest Ω_i of $\mathcal{P}_N$;

Algorithm 2: Pseudocode of the EGA.

sequently called random pointwise crossover, by applying the EGA on several data sets, varying the rates of the different crossover techniques (single point r_{sc} , two point r_{tc} and random pointwise r_{rc}) between 0 (0 %) and 1 (100 %) with a step width of 0.1 (10 %). The analysed data sets are shown in Figure 1. They cover a huge variety of possible data sets and differ in terms of complexity and the occurrence of data gaps and cusps. Please note that some of this data has been already used as a benchmark in numerous publications (see for instance [16, 9, 4]) and is also used in Section 4 to evaluate the performance of the algorithms. The data sets are approximated by the EGA until Ω_i reaches a certain threshold thr_e (for more information see Section 4 and Eq. 15). Threshold thr_e is calculated from the noise added to the data sets. Each data set is approximated by the EGA for different p and n with all possible combinations of crossover rates, whereby the sum of crossover rates can not exceed 1:

$$r_{sc} + r_{tc} + r_{rc} \leq 1. \quad (12)$$

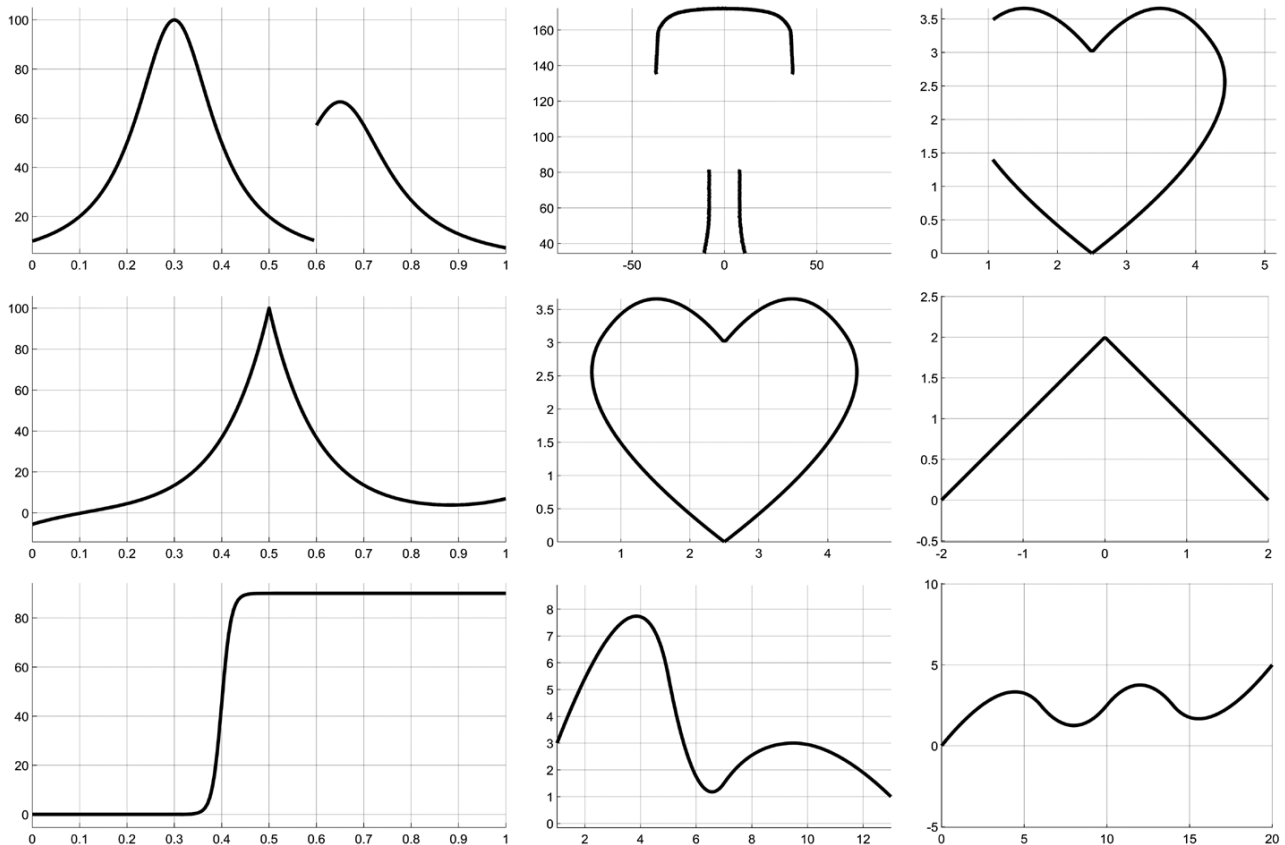


Figure 1: Analysed data sets for crossover: The left column shows three benchmark test functions introduced in [16]. In the middle and right column data sets with data gaps (first row), data sets with cusps (second row) and data sets without data gaps and cusps (third row) are shown.

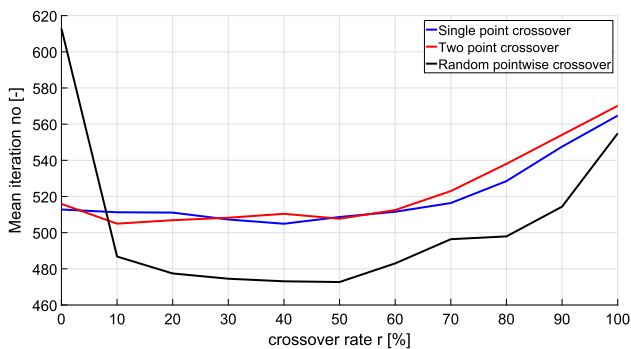


Figure 2: Results for different crossover rates: The y-axis displays the mean (total) iteration number needed to approximate all data sets. The x-axis displays the crossover rates. Single point crossover is depicted in blue, two point crossover in red and random pointwise crossover in black.

With the chosen step width of 0.1 for r_{sc} , r_{tc} and r_{rc} we analysed in total 859 different combinations, which fulfill Eq. 12. This process is repeated 50 times in order to obtain reliable and accurate results. In Figure 2 we depict the mean iteration number, summarised for all data

sets, where Ω_i reached thr_e for different crossover rates. As we can see in Figure 2, the mean iteration number significantly decreases in case of the rate for random pointwise crossover r_{rc} increases from 0 to 0.1 or more. This significant decrease is more distinct for a data set which comprises data gaps and cusps (see the upper two rows in Figure 1) than for rather simple data sets (see the left and right data set in the bottom row). Furthermore, we can see that the variation of the rates for single point r_{sc} and two point crossover r_{tc} does not have a significant impact on the mean iteration number. The mean iteration number rises for all three crossover techniques when the rate for crossover increases to more than 0.5. All described effects occur for curves which differ in the number of required control points $n + 1$, basis function degree p , and number of data points.

Based on the investigations, we recommend to use random point crossover with a rate r_{rc} between 0.2 and 0.5 and to neglect the application of single point or two point crossover. For random pointwise crossover two random individuals are chosen. In each individual one knot is



Figure 3: Example for random pointwise crossover: Both individuals consist of $n - p = 5$ internal knots. In individual 1 the second internal knot and in individual 2 the fourth internal knot is chosen randomly. These knots are exchanged between the individuals and subsequently each individual has to be assorted.

chosen randomly. The chosen knots were exchanged (see Figure 3). Subsequently, the knot vectors of the new individuals have to be assorted. This crossover technique enhances the probability to generate individuals with multiple knots. The reason for that is the fact that different knot positions could be exchanged between the individuals, while single and two point crossover exchange the knots at the same knot positions. Especially, when we keep in mind that a huge number of the individuals in $\mathcal{P}_c$ and subsequently in $\mathcal{P}_{rc}$ conform to the best individual of the former epoch and are therefore identical clones, this feature is beneficial.

The mutation mechanism in C.3 is slightly changed in comparison to Eq. 11. As already mentioned in Section 3.1 the mutation mechanism applied in Eq. 11 reduces the magnitude of perturbation by taking the closest distance Δ to one of the interval boundaries $[0, 1]$ into account. Nevertheless, it is inexplicable, why Δ is restricted to the half magnitude for knots closer to the interval boundary 0 in comparison to knots closer to the interval boundary 1. There is no known and profound reason to restrict the Δ to this value, so we suspect that in [9] a spelling error occurred. So, for the EGA, the mutation is calculated by:

$$u_i^* = u_i - \left(\frac{\Delta}{2}\right)(\sigma - 0.5) \text{ with } \Delta = \min\{u_i, 1 - u_i\}. \quad (13)$$

4 Results and discussion

In order to evaluate the efficiency of the developed algorithm EGA, both algorithms, CSA and EGA described in Chapter 3, are applied to two challenging test functions (see Chapter 4.2 and 4.3) as well as to a real data set from the field of rail track inspection (see Chapter 4.4). The test functions were introduced in Yoshimoto et al. [16] and also used in Gálvez et al. [9] and Bureick et al. [4] to evaluate the capability of the algorithms. The chosen parameters for both algorithms to obtain the subsequent results are gathered mainly from the literature. More details are described in Chapter 4.1.

4.1 Chosen parameters

Gálvez et al. [9] analysed the parameters of their algorithm for the CSA in a study. The optimal parameters determined in that study are displayed in Table 2.

For the EGA, we generally utilized the parameters suggested for the CSA. The parameter r_{rc} , comprising the rate for random pointwise crossover, is chosen from experimental investigations (see Section 3.2). All chosen parameters are depicted in Table 2.

Table 2: Parameters chosen for the CSA and EGA.

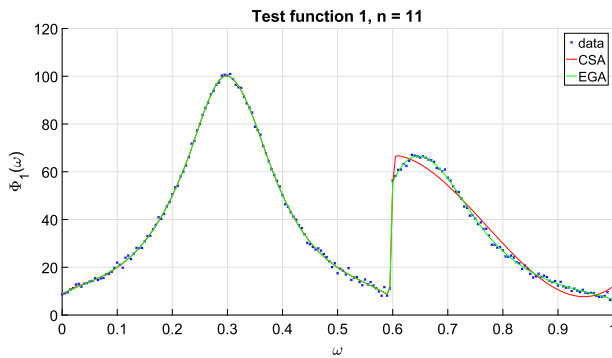
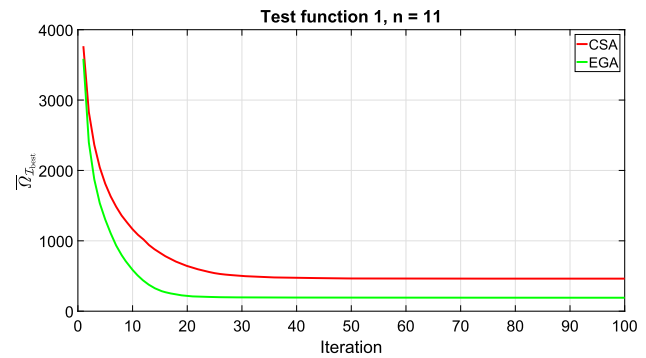
Parameter	CSA	EGA
$iter_{max}$	100	100
N	100	100
N_b	10	10
N_r	20	20
N_c	190	190
N_e	10	10
r_{rc}	–	0.3

4.2 Test function 1

The first test function is a discontinuous function and is given by:

$$\phi_1(\omega) = \begin{cases} \frac{1}{0.01 + (\omega - 0.3)^2} & \text{for } \omega < 0.6 \\ \frac{1}{0.015 + (\omega - 0.65)^2} & \text{for } \omega \geq 0.6. \end{cases} \quad (14)$$

The function $\phi_1(\omega)$ is evaluated at 201 uniformly distributed values of ω on the interval $[0, 1]$. Those one-dimensional data points represent the true point cloud. Random noise, which is generated independently from the normal distribution $N(0, 1)$, is added to the true point cloud by means of the random number generator of MATLAB[®]. Subsequently, the noise-vector is denoted by ϵ . We repeated the generation process 100 times to obtain a reliable and accurate result. Each resulting noisy point

Figure 4: Approximation result test function 1 with $p = 3$ and $n = 11$.Figure 5: Convergence rates for test function 1 with $n = 11$.

cloud is approximated by CSA and EGA (see Figure 4). The degree of the basis function is set to $p = 3$ and the parameterisation of the data points is realised by directly assigning ω to $\bar{u}$. Both, point cloud and control points are one-dimensional.

Figure 4 shows a single simulation run where the CSA converged to a local optimum. The approximation by the CSA differs significantly, especially with ω larger than 0.6.

Table 3 shows the mean sum of squared residuals $\bar{Q}_{T_{\text{best}}}$ of the best individual T_{best} at the end of each method and the percentage the method provided the best result for different numbers of control points $n + 1$. When the algorithms provided the same result, no algorithm is classified to have the best result. This occurs frequently, especially when n is small. The mean sum of squared residuals $\bar{Q}_{T_{\text{best}}}$ of the results produced with the EGA is smallest in

nearly all the cases. The percentage where the EGA produced the best results is also the highest. The CSA often converges to a local optimum, especially for $n = 9$ to $n = 13$, whereas EGA always tends to converge to the global optimum.

The convergence rates of both algorithms are depicted in Figure 5. As we can see, the EGA converges very fast. After 20 iterations, the best result does not improve significantly. The convergence rate of the CSA is similar but slower. Additionally, the CSA does not converge to the global optimum. The convergence rates for other n look very similar. For $n > 13$, the CSA also tends to converge to the global optimum.

In order to evaluate the convergence of the algorithms, we performed a slightly different analysis. Each algorithm is executed until the residual sum of squares of the best individual reaches a certain threshold thr_ϵ . The threshold thr_ϵ depends on the noise added to the curve:

$$thr_\epsilon = \sum_{i=1}^{201} \epsilon_i^2. \quad (15)$$

Table 3: $\bar{Q}_{T_{\text{best}}}$ and rate of best $Q_{T_{\text{best}}}$ in test function 1.

n	$\bar{Q}_{T_{\text{best}}}$		best $Q_{T_{\text{best}}}$ [%]	
	CSA	EGA	CSA	EGA
4	68494.55	68494.55	2	2
5	37724.65	37724.65	44	13
6	10750.44	10750.49	64	34
7	6757.83	6756.43	54	45
8	4360.35	4315.55	59	41
9	3791.96	3541.63	33	67
10	1159.55	895.18	55	44
11	462.66	191.26	42	57
12	322.76	183.54	26	74
13	190.58	180.35	23	77
14	195.62	178.75	21	79
15	178.21	174.02	28	72
16	176.80	173.31	22	78
17	170.93	168.14	34	66
18	172.15	168.09	22	78
19	171.94	167.93	27	73
20	164.71	160.47	20	80

To avoid an infinite runtime the execution is stopped after 1000 iterations. This analysis is only done for n , where it is possible to reach thr_ϵ . Again we repeated this experiment 100 times. Table 4 shows the mean time until thr_ϵ is reached and the percentage thr_ϵ was reached within 1000 iterations. In case of an iteration, where thr_ϵ was not reached, the time is stopped after iteration 1000. The calculations were executed on an Intel® Core™ i5-4590 CPU with 3.30 GHz and 8.00 GB RAM using the MATLAB®-software. As we can see, the EGA reaches thr_ϵ most rapidly and in nearly every iteration. The CSA generally requires more time to reach thr_ϵ and, for small n , fails to reach thr_ϵ more often. With an increasing value for n the differences between the algorithms decrease, because the approximation becomes more precise in general.

Table 4: Mean time and rate to reach thr_ϵ in test function 1.

n	Mean time [s]		Rate [%]	
	CSA	EGA	CSA	EGA
11	15.31	2.83	59	95
12	6.96	1.78	83	99
13	4.09	0.53	91	100
14	1.56	0.40	98	100
15	0.57	0.34	100	100
16	0.47	0.31	100	100
17	0.45	0.28	100	100
18	0.39	0.25	100	100
19	0.37	0.25	100	100
20	0.35	0.24	100	100

4.3 Test function 2

Test function 2 contains a cusp (see Figure 6) and is given by:

$$\phi_2(\omega) = \frac{100}{e^{|10\omega-5|}} + \frac{(10\omega-5)^5}{500}. \quad (16)$$

The further steps are performed analogously to the ones depicted in Chapter 4.2. Figure 6 shows a simulations run where both methods converged towards the global optimum. Additionally, the approximation of the cusp ($\omega = 0.5$) does not differ significantly.

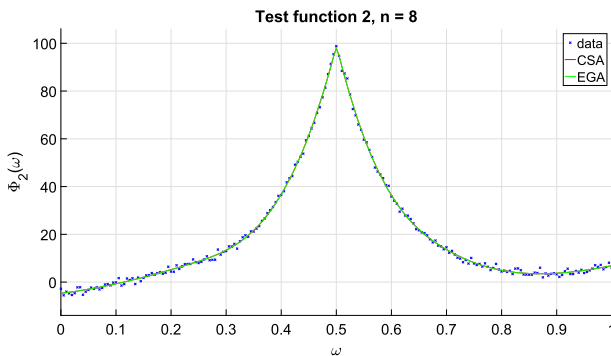
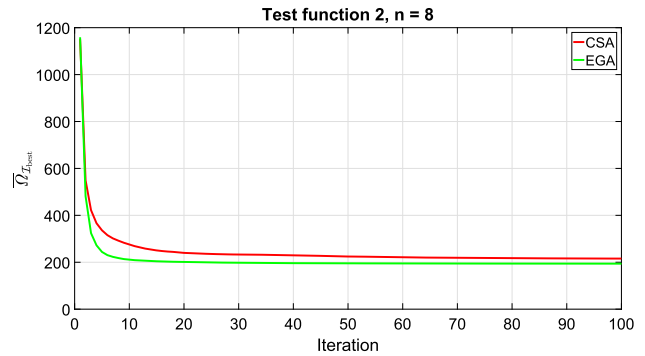
**Figure 6:** Approximation result test function 2 with $p = 3$ and $n = 8$.

Table 5 shows the mean residual sum of squares $\bar{Q}_{\mathcal{I}_{\text{best}}}$ of the best individual $\mathcal{I}_{\text{best}}$ at the end of each method and the percentage where the method provided the best result for different n . The $\bar{Q}_{\mathcal{I}_{\text{best}}}$ of the results produced with the EGA is smallest for this test function in nearly all the cases. In most cases the EGA provided the best result. Except for $n = 7$ and $n = 8$, both algorithms tend to converge to the global optimum. For $n = 7$ and $n = 8$, the CSA occasionally converged to a local optimum. The convergence rates of both algorithms for $n = 8$ are depicted in Figure 7. Again,

Table 5: $\bar{Q}_{\mathcal{I}_{\text{best}}}$ and rate of best $\mathcal{Q}_{\mathcal{I}_{\text{best}}}$ in test function 2.

n	$\bar{Q}_{\mathcal{I}_{\text{best}}}$		best $\mathcal{Q}_{\mathcal{I}_{\text{best}}}$ [%]	
	CSA	EGA	CSA	EGA
4	24188.69	24188.69	6	0
5	17528.69	17528.69	22	7
6	592.12	592.20	65	32
7	401.78	385.65	36	64
8	215.68	194.02	48	52
9	189.30	187.67	41	58
10	186.01	183.84	31	69
11	183.62	181.95	34	65
12	181.04	178.96	29	71
13	181.82	179.19	21	79
14	176.48	173.66	24	76
15	172.09	168.65	25	75
16	175.27	170.77	20	79
17	170.10	166.37	26	74
18	166.57	162.53	25	75
19	162.53	157.69	22	78
20	165.07	160.53	20	80

**Figure 7:** Convergence rates for test function 2 with $n = 8$.

EGA and CSA show a similar convergence rate, whereas the CSA sometimes converges to a local optimum.

Analogously to the analysis in Chapter 4.2 and Table 4 we analysed the performance of the algorithms on this test function (see Table 6).

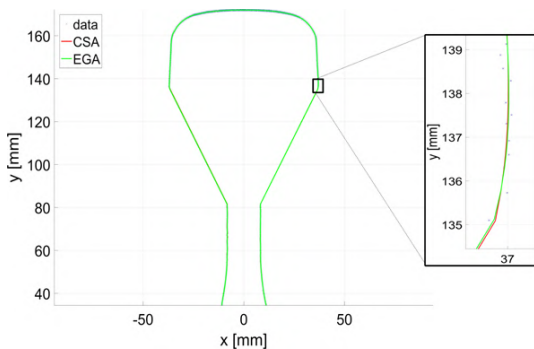
As one can see, the performance of the two algorithms varies significantly less than for test function 1. But again the EGA reaches thr_ϵ most rapidly. Both algorithms reached thr_ϵ in a comparable amount of cases. As already mentioned and explained for test function 1 the difference between the algorithms decreases when n increases.

4.4 Real data set

Both methods are applied to a real data set, which was captured by the advanced rail track inspection system (ARTIS)

Table 6: Mean time and rate to reach thr_ϵ in test function 2.

n	Mean time [s]		Rate [%]	
	CSA	EGA	CSA	EGA
8	7.81	6.62	83	83
9	1.31	0.70	98	99
10	1.08	0.21	98	100
11	0.22	0.13	100	100
12	0.20	0.12	100	100
13	0.15	0.10	100	100
14	0.12	0.09	100	100
15	0.09	0.08	100	100
16	0.08	0.07	100	100
17	0.07	0.07	100	100
18	0.07	0.07	100	100
19	0.07	0.06	100	100
20	0.07	0.06	100	100

**Figure 8:** Approximation result real data with $p = 3$ and $n = 60$.

(for more information, see [27]). Among others, ARTIS is equipped with two profile laser scanner (PLS), which measure the rail geometry. Because of the measurement configuration, the point cloud contains two large data gaps within the concealed range below the rail head. The manufacturer quantifies the measurement uncertainty of the PLS at $\sigma = 0.16$ mm. The point cloud, consisting of 1144 two-dimensional points, and an approximation result of the two methods is depicted in Figure 8. The point cloud is parameterised using the chord length (see Piegel and Tiller [10]). As we can see, the visual result of both approximation methods is similar. Only by zooming in slight differences become visible. We repeated the approximation with both methods 50 times. The mean Ω_i of the best individual $\mathcal{I}_{\text{best}}$ at the end of each method and the percentage where the method provided the best result is depicted for different n in Table 7. The result was NaN for the CSA in many repetitions, especially for $n = 75$ and $n = 80$. These repetitions are excluded from the mean value. The EGA provided the best results in all cases. The results of the CSA are numerically unstable when data gaps occur.

Table 7: $\bar{\Omega}_{\mathcal{I}_{\text{best}}}$ and rate of best $\Omega_{\mathcal{I}_{\text{best}}}$ in real data set.

n	$\bar{\Omega}_{\mathcal{I}_{\text{best}}}$		best $\Omega_{\mathcal{I}_{\text{best}}}$ [%]	
	CSA	EGA	CSA	EGA
60	9.67	8.39	0	100
65	9.45	8.17	0	100
70	9.19	8.00	0	100
75	10.59	7.84	0	100
80	8.68	7.72	0	100

Table 8: Mean time and rate to reach thr_ϵ in real data set.

n	Mean time [s]		Rate [%]	
	CSA	EGA	CSA	EGA
60	86.76	1.55	93	100
65	160.54	0.93	99	100
70	404.40	1.04	70	100
75	635.17	1.11	34	100
80	791.09	1.22	9	100

Analogously to the analysis in Chapter 4.2 (Table 4) and 4.3 (Table 6) we also analysed the performance of the algorithms for the real data set (see Table 8). In case of the real data set we do not know the “true” noise for Eq. 15, so we derived thr_ϵ utilizing the measurement uncertainty σ specified by the manufacturer. On the one hand, the EGA converges to thr_ϵ faster than CSA. On the other hand, the EGA reaches thr_ϵ in each repetition, while the CSA fails between 1% and 91%. For $n \geq 65$ the Ω_i obtained by the deterministic method were smaller than thr_ϵ and the EGA already converged in the first iteration. Furthermore, this analysis emphasizes the numerical instability of the CSA for data sets containing data gaps in combination with an increasing n . The number of repetitions, where the CSA reaches thr_ϵ , decreases, when n increases, because many repetitions produce numerically instable results due to singularity of the normal equations. In these repetitions the CSA was stopped after 1000 iterations.

Obviously the CSA was not developed to deal with data gaps. Due to the initialisation (see C.1 in Chapter 3.2) our developed EGA is significantly better to deal with these data gaps. A CSA with a comparable initialisation would produce significantly better and more stable results than it does now for this data set.

5 Conclusion and outlook

In this paper, we presented a new approach to determine the knot vector in B-spline curve approximation. The EGA

is capable of dealing with point clouds containing objects with complex shapes as well as with the occurrence of data gaps. Additionally, the EGA converges quickly towards the global optimum. The application to two challenging test functions showed that the EGA outperforms existing methods in most cases. The EGA obtained the best results for the first test function in about 70 % of the simulation runs. The EGA obtained the best mean result for nearly all the numbers of control points. For the second test function the EGA also obtained the best results in nearly 70 % of the simulation runs and the best mean result for nearly all numbers of control points. Additionally the EGA reached a previously defined threshold most rapidly and more often than the CSA. The application on a challenging real data set showed that the EGA is much better suited to deal with data gaps. Here, the EGA clearly obtained the best results, whereas the CSA occasionally yielded singular results with NaN.

The reason for the good performance of the EGA is based on two features:

1. The applied crossover technique enhances the probability for multiple knots, which are a necessity for approximating cusps, leaps or discontinuities.
2. The presented initialisation eliminates the unfavorable effect of singularities in the control point estimation caused by the occurrence of data gaps.

In future work, we will analyse all used parameters of the EGA in a comprehensive investigation. Hereafter, we attend to implement the invented EGA in B-spline surface approximation.

Another important issue for B-spline approximation is the handling of outliers within the knot vector determination and control point estimation. Since we did not focus on that topic within this paper, we plan to further analyse the dealing with outliers in future.

In a further study we plan to analyse the applicability of B-splines, especially using the EGA, for the modeling of non linear, high frequent trajectories of Unmanned Aerial Vehicles.

Funding: This work was partly funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – NE 1453/5-1.

References

- [1] K. R. Koch, Fitting free-form surfaces to laserscan data by NURBS, *Allgemeine Vermessungs-Nachrichten (AVN)* 116 (4) (2009) 134–140.
- [2] C. Harmening, H. B. Neuner, A constraint-based parameterization technique for B-spline surfaces, *Journal of Applied Geodesy* 9 (3) (2015) 143–161. doi:10.1515/jag-2015-0003.
- [3] J. Bureick, H. B. Neuner, C. Harmening, I. Neumann, Curve and surface approximation of 3D point clouds, *Allgemeine Vermessungs-Nachrichten (AVN)* 123 (11–12) (2016) 315–327.
- [4] J. Bureick, H. Alkhatib, I. Neumann, Robust spatial approximation of laser scanner point clouds by means of free-form curve approaches in deformation analysis, *Journal of Applied Geodesy* 10 (1) (2016) 27–35. doi:10.1515/jag-2015-0020.
- [5] X. Xu, X. Zhao, H. Yang, I. Neumann, TLS-based feature extraction and 3D modeling for arch structures, *Journal of Sensors* 2017 (2017) 1–8. doi:10.1155/2017/9124254.
- [6] X. Xu, J. Bureick, H. Yang, I. Neumann, TLS-based composite structure deformation analysis validated with laser tracker, *Composite Structures*. doi:10.1016/j.compstruct.2017.10.015.
- [7] X. Xu, B. Kargoll, J. Bureick, H. Yang, H. Alkhatib, I. Neumann, TLS-based profile model analysis of major composite structures with robust B-spline method, *Composite Structures* 184 (2018) 814–820. doi:10.1016/j.compstruct.2017.10.057.
- [8] P. Dierckx, *Curve and Surface Fitting with Splines*, repr Edition, Oxford science publications, Clarendon Press, Oxford [i. a.], 1993.
- [9] A. Gálvez, A. Iglesias, A. Avila, C. Otero, R. Arias, C. Manchado, Elitist clonal selection algorithm for optimal choice of free knots in B-spline data fitting, *Applied Soft Computing* 26 (2015) 90–106. doi:10.1016/j.asoc.2014.09.030.
- [10] L. A. Piegl, W. Tiller, *The NURBS Book*, 2nd Edition, Monographs in visual communications, Springer, Berlin and New York, 1997.
- [11] H. Park, J. H. Lee, B-spline curve fitting based on adaptive curve refinement using dominant points, *Computer-Aided Design* 39 (6) (2007) 439–451. doi:10.1016/j.cad.2006.12.006.
- [12] F. Yoshimoto, T. Harada, Y. Yoshimoto, Data fitting with a spline using a real-coded genetic algorithm, *Computer-Aided Design* 35 (8) (2003) 751–760. doi:10.1016/S0010-4485(03)00006-X.
- [13] X. Zhao, C. Zhang, B. Yang, P. Li, Adaptive knot placement using a GMM-based continuous optimization algorithm in B-spline curve approximation, *Computer-Aided Design* 43 (6) (2011) 598–604. doi:10.1016/j.cad.2011.01.015.
- [14] M. G. Cox, The numerical evaluation of B-Splines, *IMA Journal of Applied Mathematics* 10 (2) (1972) 134–149. doi:10.1093/imat/10.2.134.
- [15] C. de Boor, On calculating with B-splines, *Journal of Approximation Theory* 6 (1) (1972) 50–62. doi:10.1016/0021-9045(72)90080-9.
- [16] F. Yoshimoto, M. Moriyama, T. Harada, Automatic knot placement by a genetic algorithm for data fitting with a spline, in: *Proceedings of the International Conference on Shape Modeling and Applications*, Aizu-Wakamatsu, IEEE Computer Society Press, 1999, pp. 162–169.
- [17] C. Harmening, H. B. Neuner, Choosing the optimal number of B-spline control points (Part 1: Methodology and approximation of curves), *Journal of Applied Geodesy* 10 (3) (2016) 139–157. doi:10.1515/jag-2016-0003.
- [18] W. Ma, J. P. Kruth, Parameterization of randomly measured points for least squares fitting of B-spline curves and

- surfaces, *Computer-Aided Design* 27 (9) (1995) 663–675. doi:10.1016/0010-4485(94)00018-9.
- [19] J. Hoschek, Intrinsic parametrization for approximation, *Computer Aided Geometric Design* 5 (1) (1988) 27–31. doi:10.1016/0167-8396(88)90017-9.
- [20] B. Sarkar, C. H. Menq, Parameter optimization in approximating curves and surfaces to measurement data, *Computer Aided Geometric Design* 8 (4) (1991) 267–290. doi:10.1016/0167-8396(91)90016-5.
- [21] J. R. Rice, *The Approximation of Function*, 2nd Edition, Addison-Wesley Publishing Company, Reading, MA, 1969.
- [22] D. L. Jupp, Approximation to data by spline with free knots, *SIAM Journal on Numerical Analysis* 15 (2) (1978) 328–343.
- [23] L. A. Piegl, W. Tiller, Surface approximation to scanned data, *The Visual Computer* 16 (7) (2000) 386–395. doi:10.1007/PL00013393.
- [24] M. Sarfraz, S. A. Raza, Capturing outline of fonts using genetic algorithm and splines, in: E. Banissi (Ed.), *Proceedings Fifth International Conference on Information Visualisation*, IEEE Computer Soc., Los Alamitos, Calif. [i. a.], 2001, pp. 738–743.
- [25] E. Ülker, A. Arslan, Automatic knot adjustment using an artificial immune system for B-spline curve approximation, *Information Sciences* 179 (10) (2009) 1483–1494. doi:10.1016/j.ins.2008.11.037.
- [26] A. A. Adewuya, *New methods in genetic search with real-valued chromosomes*, Ph. D. thesis, Massachusetts Institute of Technology, (1996).
- [27] D. Dennig, J. Bureick, J. Link, D. Diener, C. Hesse, I. Neumann, Comprehensive and highly accurate measurements of crane runways, profiles and fastenings, *Sensors* 17 (5) (2017) 1118. doi:10.3390/s17051118.

An iteratively reweighted least-squares approach to adaptive robust adjustment of parameters in linear regression models with autoregressive and t -distributed deviations

Boris Kargoll¹  · Mohammad Omidalizarandi¹ · Ina Loth² · Jens-André Paffenholtz¹ · Hamza Alkhatib¹

Received: 10 May 2017 / Accepted: 26 August 2017 / Published online: 9 September 2017
© Springer-Verlag GmbH Germany 2017

Abstract In this paper, we investigate a linear regression time series model of possibly outlier-afflicted observations and autocorrelated random deviations. This colored noise is represented by a covariance-stationary autoregressive (AR) process, in which the independent error components follow a scaled (Student's) t -distribution. This error model allows for the stochastic modeling of multiple outliers and for an adaptive robust maximum likelihood (ML) estimation of the unknown regression and AR coefficients, the scale parameter, and the degree of freedom of the t -distribution. This approach is meant to be an extension of known estimators, which tend to focus only on the regression model, or on the AR error model, or on normally distributed errors. For the purpose of ML estimation, we derive an expectation conditional maximization either algorithm, which leads to an easy-to-implement version of iteratively reweighted least squares. The estimation performance of the algorithm is evaluated via Monte Carlo simulations for a Fourier as well as a spline model in connection with AR colored noise models of different orders and with three different sampling distributions generating the white noise components. We apply the algorithm to a vibration dataset recorded by a high-accuracy, single-axis accelerometer, focusing on the evaluation of the estimated AR colored noise model.

Keywords Linear regression model · Autoregressive process · Scaled t -distribution · Adaptive robust estimation ·

Expectation maximization (EM) algorithm · Iteratively reweighted least squares

1 Introduction

The normal distribution has traditionally played a prominent role in geodetic adjustment theory for both practical and theoretical reasons (cf. Koch 1999). As Helmert (1907, §2) noted, it is known from experience that the Gaussian error law usually approximates the occurrence of random measurement errors very well; however, the shape of the error law can only be determined by means of observations. Helmert (1907) continued by considering two other error laws besides the Gaussian one. Nowadays, there is growing evidence that random deviations have indeed, in many cases, a non-Gaussian (cf. Box and Andersen 1955; Orlov 1991), thick-tailed probability distribution (cf. Tukey 1960; Hampel 2001; Resnick 2007; Wiśniewski 2014). This means that large deviations have, in practice, often been found to occur with a larger probability than for the Gaussian distribution.

For instance, Lehmann (2015) considers in his recent study of observation error laws, besides the Laplace distribution and the distribution based on the Huber function (Huber 1964) (which are the probabilistic backbones of the L_1 -norm and Huber's M-estimator), also the scaled (Student's) t -distribution. The t -distribution, in particular, which is used as an error law also in the present contribution, was in a number of studies found to have more adequate tails (cf. Nadarajah 2009; Geweke 1993; Fraser 1976) and to have a number of useful mathematical properties which arise from common-sense requirements (cf. Alkhatib et al. 2017; Schön et al. 2017).

The focus on the tails of a distribution is important to this examination since they are naturally related to outliers

Boris Kargoll
kargoll@gih.uni-hannover.de

¹ Geodetic Institute, Leibniz Universität Hannover, Nienburger Str. 1, 30167 Hannover, Germany

² Institute of Geodesy and Geoinformation, Rheinische Friedrich-Wilhelms Universität Bonn, Nussallee 17, 53115 Bonn, Germany

in case the latter are defined through a threshold, say, by means of the popular 3σ -rule (cf. [Lehmann 2013](#)). Outliers are a common and unavoidable phenomenon in time series measured by modern geodetic sensors, and they easily distort the estimation of (spatial) parametric models in various significant ways. In the context of linear models, one classical way of dealing with outlying observations is to apply Baarda's or Pope's outlier test ([Baarda 1968](#); [Pope 1976](#)) and to eliminate the detected observations via data snooping (cf. [Knight et al. 2010](#); [Kok 1984](#)). The deletion of observations may however be undesirable, for instance, if time series models such as autoregressive (AR) processes are employed, as these usually require an equidistant sampling without gaps. This common yet problematic scenario can be avoided by the use of robust parameter estimators, which aim at limiting the effect of outlying observations on model estimates.

The robust estimators considered most frequently for a geodetic data analysis are M -estimators having a bounded loss function, here in particular, the L_1 - and L_p -norm estimators (see [Nowel and Kamiński 2014](#); [Marx 2013](#); [Götzelmann et al. 2006](#); [Junhuan 2005](#); [Marshall 2002](#)), Huber's M -estimator (see [Huber 1964](#); [Bureick et al. 2016](#); [Koch 2013b](#); [Guo et al. 2010](#); [Götzelmann et al. 2006](#); [Chang and Guo 2005](#); [Hekimoglu and Berber 2003](#); [Yang et al. 2002](#)), and robust maximum likelihood (ML) estimators based on a thick-tailed probability density function (e.g., [Wiśniewski 2014](#)). Other robust estimators have also been suggested and applied, e.g., the Danish Method ([Wiśniewski 2009](#); [Götzelmann et al. 2006](#); [Zhong 1997](#); [Kubik 1982](#); [Krarup et al. 1980](#)) and R -estimators (see [Duchnowski 2009, 2013](#); [Kargoll 2005](#)). We should note that the L_1 -norm estimator and Huber's M -estimator were proven to be non-robust under a certain condition of weightings ([Xu 2005](#)).

Another disadvantage of the aforementioned robust estimators is their lack of precision when the true error distribution deviates from the assumed probabilistic model (e.g., in case of applying the L_1 -norm estimator to roughly normally distributed observations). Robust estimators that involve free distributional shape parameters or tuning constants (e.g., the constant " k " for Huber's estimator) appear to be more flexible in this regard. However, such tuning constants need to be fixed *before* the actual parameter estimation and thus usually in absence of sufficient knowledge about the outlier characteristics. To overcome this problem, [Xu \(1993\)](#) proposed to assign individual tuning constants, e.g., regarding Huber's estimator and the Danish Method, to the residuals within a probabilistic framework essentially as functions of the design matrix, the *a priori* weight matrix, and the redundancy of the adjustment problem.

In contrast, there exist robust estimators that involve tuning parameters fixed *during* the process of adjusting the data, based on the stochastic properties of estimated residuals. For instance, [Hogg \(1974\)](#) investigated α -trimmed least-squares

and the L_p -norm estimator with adaptable trimming ratio α and norm parameter p , respectively, which he called *adaptive estimators*. [Xu \(2005\)](#) introduced an (unsymmetrically) adaptive trimmed least-squares estimator, for which robustness is achieved by means of constraints about the signs of the residuals and the weights of the observations. Once a choice is made concerning a threshold constant (depending, for instance, on whether the contamination by outliers is less or greater than 50%), this so-called *sign-constrained robust least squares* method is data-adaptive in the sense that the correct percentage of outliers is automatically accommodated by removing their effect on the parameter estimates. Another type of adaptive robust estimator was constructed by means of the convex combination of the L_1 -norm and, respectively, the L_2 -norm or Huber's estimator, where the optimal weight of each norm is adapted to the data (see [Brockmann and Kargoll 2012](#); [Dodge and Jureckova 1987, 1988](#)).

Still another type of adaptive robust estimator was based on the aforementioned scaled t -distribution, which involves—besides a scaling parameter—a data-adaptable degree of freedom, to control the thickness of the tails of the defining density function. [Lange et al. \(1989\)](#) showed that ML estimation for the scaled t -distribution can be conveniently carried out by using an expectation maximization (EM) algorithm (cf. [Peng 2009](#); [Gupta and Chen 2011](#); [Luxen and Brunn 2003](#); [Dempster et al. 1977](#)) in the form of iteratively reweighted least squares (IRLS, in the sense of [Dempster et al. 1980](#)). Various modifications of EM such as expectation conditional maximization (ECM) and ECM *either* (ECME) were proposed to speed up the convergence of the algorithm (see [McLachlan and Krishnan 2008](#); [Little and Rubin 2002](#); [Liu 1997](#); [Liu and Rubin 1995, 1994](#); [Meng and Rubin 1993](#)). [Koch and Kargoll \(2013\)](#) used an ECM algorithm for an adaptive robust estimation of a B-spline surface to approximate a measured 3D point cloud with one stochastic coordinate component. This approach was subsequently combined with an outlier test, which successfully identified additive outliers among those residuals with small weights ([Koch 2013a](#)). [Koch \(2013b\)](#) later adapted the underlying EM algorithm to the Gauss–Helmert model to approximate a 3D point cloud measured by a terrestrial laser scanner (TLS), and effectively extended this study to the case of unknown variance components ([Koch 2014](#)). [Koch and Kargoll \(2015\)](#) further used the EM algorithm based on the scaled t -distribution to compare the success of outlier detection in a Gauss–Helmert model for a polar TLS coordinate system and rectangular output coordinate system.

The issue of robust estimation of linear regression models involving autoregressive (AR) or autoregressive moving average (ARMA) deviations has not received much attention in geodetic data analysis. Recent work on related combined models can be found in [Schuh et al. \(2014\)](#), [Kargoll \(2012\)](#), [Schuh \(2003\)](#), [Klees et al. \(2003\)](#), and [Koch and](#)

Schmidt (1994). In practice, the AR(MA) model component was usually used in a non-robust setting to model Gaussian colored noise for different types of measuring sensors, for instance, inertial sensors (see Wang et al. 2012; Nassar and El-Sheimy 2005; Nassar et al. 2004), Global Navigation Satellite System (GNSS) (e.g., Luo et al. 2012; Li 2011; Amiri-Simkooei et al. 2007), microelectromechanical systems (MEMS) based inertial sensors in combination with a low-cost Global Positioning System (GPS) receiver and a digital compass (Park and Gao 2008), and the Gravity Field and Steady-State Ocean Circulation Explorer (GOCE) satellite gravity gradiometer (e.g., Brockmann 2015; Krasbutter et al. 2011, 2014, 2015; Schall et al. 2014; Cai and Sneeuw 2014; Siemes 2013; Pail et al. 2011; Schuh et al. 2010; Brockmann et al. 2010; Alkhatib and Schuh 2007; Klees et al. 2004; Schuh 2003). AR and, in particular, ARMA processes constitute attractive models due to their capability of capturing quite intricate autocorrelation patterns through a number of coefficients which is relatively small in comparison with the length of data-based autocovariance functions (ACF). In addition, there are a number of easy-to-implement mathematical relationships between AR(MA) coefficients, the ACF and the power spectral density (PSD), which further add to the usability of AR(MA) processes. For instance, if the colored noise of a sensor is specified through the PSD, then its representation in terms of (AR)MA coefficients is readily available (see, for instance, Krasbutter et al. 2015) and thus integrable into the observation model. It should be further noted that AR(MA) processes are also useful for modeling the measurements themselves instead of their colored noise. For example, an ARMA process was applied by Xu (1988) to predict deformations of a large dam and by Niedzielski and Kosek (2010) to model sea level fluctuations. Lindenberger (1993) also used an integrated AR process to model surface profiles (see also Förstner and Wrobel 2016), and Niedzielski and Kosek (2012) employed a vector ARMA process for predictions of Universal Time (UT1-UTC) based on geodetic and geophysical data.

Our approach with the current paper is to extend the ECM algorithm for adaptive robust estimation of a Gauss–Markov model to include an additional AR model (with t -distributed independent errors), which is adjusted jointly with the functional model. This more comprehensive adjustment model is inspired by Christmas and Everson (2011), who adaptively estimated the coefficients of a pure AR process (with t -distributed independent errors), though without consideration of a spatial regression model. For the same kind of process (which was established theoretically by Heyde and Leonenko 2005), Sanubari and Tokuda (1992) and Sanubari (1999) carried out a non-adaptive ML estimation in the sense that the value of the degree of freedom of the underlying t -distribution was assumed and fixed *a priori*.

We will not attempt to include an ARMA model into the Gauss–Markov model since we found it to be incompatible with EM-type algorithms. If such an attempt was to be made, then a state-space representation of the ARMA process would likely need first to be established (as indicated by Metaxoglou and Smith 2007), which is beyond the scope of the present paper.

We describe the overall observation model in greater detail in Sect. 2 and outline the adaptive robust estimation procedure in Sect. 3. In particular, we show how the E-step yields adapted observation weights and how the M-steps are realized for the individual parameter groups via accordingly reweighted least squares. In Sect. 3.2, we investigate the stochastic properties of the proposed estimator in terms of unbiasedness and variability. In Sect. 3.3, we present the analysis of a long time series of measured accelerometer data, recorded as part of a vibration monitoring experiment. The Fourier coefficients with respect to the occurring oscillations and the AR coefficients of the accelerometer’s colored noise model are jointly estimated with the variance factor and the degree of freedom of the underlying t -distribution. Based on these findings, the hypothesis that “accelerometer measurements are normally distributed” is addressed. Section 6 contains the conclusions and briefly outlines a plan for further investigations.

2 The basic observation model

We assume that observables $\mathcal{L} = (\mathcal{L}_1, \dots, \mathcal{L}_n)$ give rise to observation equations

$$\mathcal{L}_t = \mathbf{A}_t \boldsymbol{\xi} + \mathcal{E}_t \quad (t = 1, \dots, n) \quad (1)$$

in form of a time series.¹ Here, $\mathbf{A}_t \boldsymbol{\xi}$ stands for a time-dependent, purely deterministic functional model involving $m < n$ unknown parameters $\boldsymbol{\xi} = [\xi_1, \dots, \xi_m]^T$. The rows $\mathbf{A}_1, \dots, \mathbf{A}_n$ form the $(n \times m)$ -design matrix $\mathbf{A}$, which we assume to have full rank. Furthermore, the components of $\mathcal{E} = [\mathcal{E}_1, \dots, \mathcal{E}_n]^T$ shall represent random deviations, for which we obtain from (1)

$$\mathcal{E}_t = \mathcal{L}_t - \mathbf{A}_t \boldsymbol{\xi} \quad (t = 1, \dots, n) \quad (2)$$

Concerning the stochastic model, we assume the random deviations to be correlated through autoregressive (AR) model equations

$$\mathcal{E}_t = \alpha_1 \mathcal{E}_{t-1} + \dots + \alpha_p \mathcal{E}_{t-p} + \mathcal{U}_t \quad (t = 1, \dots, n), \quad (3)$$

¹ We denote unknown parameters by Greek letters, random variables by calligraphic letters, and constants by Roman letters. Thus, we distinguish between a random variable (e.g., $\mathcal{E}_t$) and its realization (e_t). Furthermore, matrices and vectors are represented by bold letters.

where $\alpha = [\alpha_1, \dots, \alpha_p]^T$ is a constant vector of unknown parameters and $\mathcal{U} = [\mathcal{U}_1, \dots, \mathcal{U}_n]^T$ a vector of independently and identically distributed random variables with mean 0 and variance σ_0^2 (i.e., *white noise*). Due to the recursive nature of the *colored noise* process (3), random deviations $\mathcal{E}_0, \dots, \mathcal{E}_{1-p}$ without associated observables occur; we assume these deviations to take the constant value 0, via assumption of the initial conditions $\mathcal{L}_0 = \dots = \mathcal{L}_{1-p} = 0$ and $\mathbf{A}_0 = \dots = \mathbf{A}_{1-p} = \mathbf{0}_{[1 \times m]}$. In view of (3), we obtain for the white noise variables

$$\mathcal{U}_t = \mathcal{E}_t - \alpha_1 \mathcal{E}_{t-1} - \dots - \alpha_p \mathcal{E}_{t-p} \quad (t = 1, \dots, n). \quad (4)$$

Applying now the abbreviating notations $L^j \mathcal{E}_t := \mathcal{E}_{t-j}$, $\mathbf{L} := [L, L^2, \dots, L^p]$ and $\alpha(\mathbf{L}) := 1 - \alpha^T \mathbf{L}$ (known as the *lag polynomial*) we can rewrite (4) as

$$\mathcal{U}_t = \alpha(\mathbf{L}) \mathcal{E}_t \quad (t = 1, \dots, n). \quad (5)$$

Here, we may interpret $\alpha(\mathbf{L})$ as a *decorrelation filter*, transforming the colored noise $\mathcal{E}$ into white noise $\mathcal{U}$, and we assume that all (possibly complex-valued) roots of $\alpha(\mathbf{z}) = 0$ are located inside the unit circle. Since the assumed constancy of the random variables $\mathcal{E}_0, \dots, \mathcal{E}_{1-p}$ distorts the initial stage of the AR process—this effect is known as *warm-up* (cf. Section 4.1, Siemes 2013)—the AR process is then *asymptotically stationary up to order 2* (see Section 3.5, Priestley 1981). Combining (5) with (2) yields also the equations

$$\mathcal{U}_t = \alpha(\mathbf{L})(\mathcal{L}_t - \mathbf{A}_t \xi) \quad (t = 1, \dots, n). \quad (6)$$

The case $\alpha(\mathbf{L}) = 1$ (implied by $p = 0$) “switches off” the AR process, resulting in purely uncorrelated random deviations $\mathcal{E} = \mathcal{U}$.

The observation model defined up to this point was already considered by Durbin (1960, see Eq. (3) and subsequent assumptions), who additionally required that all moments of $\mathcal{U}$ are finite. As a particular instance of Durbin’s model specifications, Schuh et al. (2014) assumed $\mathcal{U}$ to be Gaussian white noise. The extension of this AR component of that model to an ARMA process was widely used for representing a spherical harmonics model for the Earth’s gravity field in connection jointly with the colored noise of the measuring GOCE satellite gravity gradiometer. In both cases, normal least-squares theory was applied to estimate the model parameters. However, least-squares approaches are non-robust by themselves, so that a data preprocessing in terms of outlier detection and removal is necessary.

Instead, we follow the general idea of Lange et al. (1989) and employ a robust parameter estimation based on Student’s *t*-distribution. A robust procedure avoids data deletion, so that the recursive structure of the AR process is not hampered

by data gaps. More specifically, we assume for every $t \in \{1, \dots, n\}$ that the white noise variable $\mathcal{U}_t$ follows a scaled *t*-distribution with degree of freedom ν , mean 0 and scale parameter σ^2 , symbolically

$$\mathcal{U}_t \sim t_\nu(0, \sigma^2) \quad (t = 1, \dots, n). \quad (7)$$

This distribution, which approximates the special case of $N(0, \sigma^2)$ -distributed white noise for $\nu \rightarrow \infty$ (cf. Lange et al. 1989), is defined by the probability density function (pdf)

$$f(u_t) = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi} \sigma \Gamma(\frac{\nu}{2})} \left[1 + \left(\frac{u_t}{\sigma} \right)^2 / \nu \right]^{-\frac{\nu+1}{2}}, \quad (8)$$

where Γ denotes the gamma function. Since we have assumed the variance σ_0^2 of the white noise to be defined, ν must be assumed to be greater than 2. Recalling that the random variables $\mathcal{U}_1, \dots, \mathcal{U}_n$ were assumed to be stochastically independent and writing (6) in terms of realizations of the random variables $\mathcal{U}_t$ and $\mathcal{E}_t$, the joint pdf of $\mathcal{U}$ factorizes into

$$\begin{aligned} f(\mathbf{u}) &= f(u_1, \dots, u_n) = \prod_{t=1}^n f(u_t) \\ &= \prod_{t=1}^n \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi} \sigma \Gamma(\frac{\nu}{2})} \left[1 + \left(\frac{\alpha(\mathbf{L})(\ell_t - \mathbf{A}_t \xi)}{\sigma} \right)^2 / \nu \right]^{-\frac{\nu+1}{2}}. \end{aligned} \quad (9)$$

In contrast to the model of Lange et al. (1989), in which the independently *t*-distributed random variables $\mathcal{U}_1, \dots, \mathcal{U}_n$ directly serve as the random deviations within the observation equations, our additional model component (3) first transforms these variables into autocorrelated random deviations. The AR(*p*) model Eq. (3) in connection with the *t*-distribution specified in (7) yields a so-called *Student AR* (or *t-AR*) process (cf. Heyde and Leonenko 2005). We consider (10) as the basic parametric observation model, expressed as a function of the white noise input $\mathbf{u}$.

3 Adjustment of the observation model

3.1 The problem of maximum likelihood estimation

Maximum likelihood (ML) estimation of the model parameters $(\xi, \sigma^2, \alpha, \nu)$ from given observations $\ell = [\ell_1, \dots, \ell_n]^T$ can be carried out by maximizing the pdf (10) with respect to the parameters, using the likelihood function

$$L(\xi, \sigma^2, \alpha, \nu; \ell) = f(\mathbf{u}). \quad (11)$$

Due to the aforementioned warm-up effect of the initial conditions on the subsequent random variables of the AR pro-

cess, (11) is not identical to the likelihood function obtained by treating $\mathcal{E}_0, \dots, \mathcal{E}_{1-p}$ as genuine random variables within the process (see for a general discussion of this issue Priestley 1981). As our observation model should be applied only to rather large datasets so that the degree of freedom of the t -distribution is reliably estimated, we currently do not attempt to derive the exact (unconditional) likelihood function which may even be intractable, but employ instead the approximating (conditional) likelihood function (11). The method of conditional likelihoods is well known in time series analysis (cf. Hamilton 1994) and has been widely applied, for instance, by Tiku et al. (2000) to AR models whose input noise has a distribution defined by a certain parametric family of symmetric pdfs, by McDonald (1989) in the context of partially adaptive ML estimation of ARMA models with generalized t -distributed white input noise, and by Bera and Jarque (1982, p. 66) regarding a regression model with AR deviations and white input noise following an unspecified distribution from the Pearson Family.

The ML estimates $(\hat{\xi}, \hat{\sigma}^2, \hat{\alpha}, \hat{\nu})$ of $(\xi, \sigma^2, \alpha, \nu)$ can then be obtained by maximizing (11) itself or its natural logarithm, that is,

$$(\hat{\xi}, \hat{\sigma}^2, \hat{\alpha}, \hat{\nu}) = \arg \max_{\xi, \sigma^2, \alpha, \nu} L(\xi, \sigma^2, \alpha, \nu; \ell) \quad (12)$$

$$= \arg \max_{\xi, \sigma^2, \alpha, \nu} \log L(\xi, \sigma^2, \alpha, \nu; \ell). \quad (13)$$

Neither of these maximization problems is appealing in terms of algebraic simplicity and computational convenience. Note that the first problem (12) is non-quadratic in (ξ, α) due to the exponent $-\frac{\nu+1}{2}$ in (10). While the natural logarithm occurring in the second problem (13) turns this exponent into a harmless factor in view of

$$\begin{aligned} \log L(\xi, \sigma^2, \alpha, \nu; \ell) &= n \log \left[\frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi\sigma^2}\Gamma(\frac{\nu}{2})} \right] - \frac{\nu+1}{2} \\ &\quad \times \sum_{t=1}^n \log \left[1 + \frac{1}{\nu} \left(\frac{\alpha(\mathbf{L})(\ell_t - \mathbf{A}_t \xi)}{\sigma} \right)^2 \right], \end{aligned} \quad (14)$$

it then involves the logarithmic functions of the unknown parameters. In the following section, we show how such difficulties may be resolved to a great extent.

3.2 Alternative formulation of the observation model

Following the suggestion of Dempster et al. (1977), we introduce *latent variables* to bring the adjustment problem into an equivalent but easier manageable form. Such variables may be considered as data which are not observable but *imputable* as conditional expectations. To begin with,

we expand the observables $\mathcal{L} = (\mathcal{L}_1, \dots, \mathcal{L}_n)$ by additional latent variables $\mathcal{W} = (\mathcal{W}_1, \dots, \mathcal{W}_n)$, which will later be recognized as observation weights within an *iteratively reweighted least-squares* (IRLS) scheme. ML estimation from given observations ℓ and given “weights” $\mathbf{w}$ would then be based on the likelihood function $L(\xi, \sigma^2, \alpha, \nu; \ell, \mathbf{w})$, which requires the specification of a corresponding joint pdf. As with the definition of the likelihood function (11), we express the probabilistic model through the white noise $\mathbf{u}$ by establishing the pdf $f(\mathbf{u}, \mathbf{w}|\theta)$, which we now write explicitly as being conditional on given parameter values $\theta = (\xi, \sigma^2, \alpha, \nu)$, in the sense of Koch (2007, Section 2.2.2). This pdf will be defined in such a way that the marginal distribution of each $\mathcal{U}_t$ is identical with the scaled t -distribution (7). Thus, using $f(\mathbf{u}, \mathbf{w}|\theta)$ leads to an ML estimation problem which is equivalent to, but easier to handle than, the original problems illustrated in (12) and (13). The following description of this method was given in a similar form in Koch and Kargoll (2013) but without consideration of an AR model component. Here we add further explanations which are helpful in the present context.

Applying a known technique for representing the scaled t -distribution as a marginal distribution (cf. Section 4.6, Dempster et al. 1977), we firstly assume that the random variables $\mathcal{W}_1, \dots, \mathcal{W}_n$ are, conditional on the value of ν , independently and identically gamma-distributed according to

$$\mathcal{W}_t | \xi, \sigma^2, \alpha, \nu \sim \frac{\chi_\nu^2}{\nu} = G\left(\frac{\nu}{2}, \frac{\nu}{2}\right) \quad (t = 1, \dots, n), \quad (15)$$

which distribution is defined by the pdf

$$f(w_t|\theta) = \begin{cases} \left(\frac{\nu}{2}\right)^{\frac{\nu}{2}} \cdot w_t^{\frac{\nu}{2}-1} \cdot \exp\left(-\frac{\nu}{2}w_t\right) & \text{for } w_t > 0 \\ 0 & \text{for } w_t \leq 0 \end{cases} \quad (16)$$

(see also “Appendix A.2”) and where χ_ν^2 is the chi-squared distribution with ν degrees of freedom. The conditional independence of the latent variables $\mathcal{W}_1, \dots, \mathcal{W}_n$ implies that the product

$$f(\mathbf{w}|\theta) = \prod_{t=1}^n f(w_t|\theta) \quad (17)$$

yields their joint, multivariate pdf. Secondly, we assume that each random variable $\mathcal{U}_t$ follows a normal distribution conditional on the occurrence of the value w_t and the parameter values θ , defined by the conditional pdf

$$f(u_t|w_t, \theta) = \frac{1}{\sqrt{2\pi(\sigma/\sqrt{w_t})^2}} \exp\left(-\frac{u_t^2}{2(\sigma/\sqrt{w_t})^2}\right). \quad (18)$$

Moreover, we assume that $\mathcal{U}_t$ is conditionally independent from $\mathcal{U}_1, \mathcal{W}_1, \dots, \mathcal{U}_{t-1}, \mathcal{W}_{t-1}, \mathcal{U}_{t+1}, \mathcal{W}_{t+1}, \dots, \mathcal{U}_n$ and $\mathcal{W}_n$ with respect to given values w_t and θ , so that any of these random variables can be removed from the list of givens in $f(u_t|u_1, w_1, \dots, u_{t-1}, w_{t-1}, u_{t+1}, w_{t+1}, \dots, u_n, w_n, w_t, \theta)$, say

$$f(u_t|u_1, w_1, \dots, u_{t-1}, w_{t-1}, u_{t+1}, w_{t+1}, \dots, u_n, w_n, w_t, \theta) = f(u_t|w_t, \theta) \quad (19)$$

(cf. Section 2.2.7, Koch 2007). This independence assumption, which represents a kind of *hidden Markov property* with real-valued *hidden variables* (see Section 4, Bilmes 1998), ensures in our particular model that the variance of each $\mathcal{U}_t$ is rescaled by means of the time-wise associated weight $\mathcal{W}_t$, without being affected by other weights and other noise variables. The characterization of (18) as a conditional pdf is justified since it is known that the product

$$\begin{aligned} f(u_t, w_t|\theta) &= f(w_t|\theta) f(u_t|w_t, \theta) \\ &= \left(\frac{\nu}{2}\right)^{\frac{\nu}{2}} w_t^{\frac{\nu}{2}-1} \exp\left(-\frac{\nu}{2} w_t\right) \\ &\quad \times \frac{1}{\sqrt{2\pi(\sigma/\sqrt{w_t})^2}} \exp\left(-\frac{u_t^2}{2(\sigma/\sqrt{w_t})^2}\right) \end{aligned} \quad (20)$$

of (16) and (18) indeed constitutes a pdf (cf. p. 58 McLachlan and Krishnan 2008). It is also known that the determination of the marginal density of $\mathcal{U}_t$ from the preceding joint pdf gives (cf. p. 58 McLachlan and Krishnan 2008)

$$f(u_t|\theta) = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi} \sigma \Gamma(\frac{\nu}{2})} \left[1 + \left(\frac{u_t}{\sigma}\right)^2 / \nu\right]^{-\frac{\nu+1}{2}}, \quad (21)$$

which is precisely the pdf (8). Thus, the original probabilistic observation model in terms of the scaled t -distribution can be retrieved from the model that involves the additional latent variables. The joint pdf (20) also defines

$$\begin{aligned} f(w_t, u_t|\theta) &= f(u_t|\theta) f(w_t|u_t, \theta) \\ &= f(w_t|\theta) f(u_t|w_t, \theta) [= f(u_t, w_t|\theta)] \end{aligned} \quad (22)$$

applying the multiplication rule (cf. Section 10.5, Kusolitsch 2014) twice. Similarly, the joint pdf of $\mathcal{U}$ and $\mathcal{W}$ underlying the conditional pdf $f(u_t|u_1, \dots, u_{t-1}, u_{t+1}, \dots, u_n, \mathbf{w}, \theta)$

in (19) can be used to establish a relationship with the joint pdf of $\mathcal{W}$ and $\mathcal{U}$ via

$$f(\mathbf{u}, \mathbf{w}|\theta) = f(\mathbf{w}|\theta) f(\mathbf{u}|\mathbf{w}, \theta) \quad (23)$$

$$= f(\mathbf{u}|\theta) f(\mathbf{w}|\mathbf{u}, \theta) = f(\mathbf{w}, \mathbf{u}|\theta). \quad (24)$$

As a consequence of the conditional independence properties (17) and (19), we can establish the factorization

$$f(\mathbf{w}, \mathbf{u}|\theta) = \prod_{t=1}^n f(w_t, u_t|\theta) \quad (25)$$

by applying the multiplication rule multiple times [see the proof of equation (1.7) in replacing the total data $\mathbf{x}$ by the latent weights $\mathbf{w}$ and the observed values $\mathbf{y}$ by the noise vector $\mathbf{u}$ Gupta and Chen 2011]. In view of (22)–(24), we then find also the factorization

$$f(\mathbf{u}, \mathbf{w}|\theta) = \prod_{t=1}^n f(u_t, w_t|\theta), \quad (26)$$

so that all pairs $(\mathcal{U}_t, \mathcal{W}_t)$ are conditionally independent. We are now in a position to define the approximate likelihood function with respect to the alternative observation model in terms of the pdf (20), in which we substitute the decorrelated observation equations (6) to obtain

$$\begin{aligned} L(\xi, \sigma^2, \alpha, \nu; \ell, \mathbf{w}) &:= f(\mathbf{u}, \mathbf{w}|\theta) = \prod_{t=1}^n f(u_t, w_t|\theta) \\ &= \prod_{t=1}^n \left(\frac{\nu}{2}\right)^{\frac{\nu}{2}} w_t^{\frac{\nu}{2}-1} \exp\left(-\frac{\nu}{2} w_t\right) \frac{1}{\sqrt{2\pi(\sigma/\sqrt{w_t})^2}} \\ &\quad \times \exp\left(-\frac{[\alpha(\mathbf{L})(\ell_t - \mathbf{A}_t \xi)]^2}{2(\sigma/\sqrt{w_t})^2}\right) \end{aligned} \quad (27)$$

$$\begin{aligned} &= (2\pi)^{-\frac{n}{2}} (\sigma^2)^{-\frac{n}{2}} \left(\frac{\nu}{2}\right)^{\frac{n\nu}{2}} \left[\Gamma\left(\frac{\nu}{2}\right)\right]^{-n} \\ &\quad \times \prod_{t=1}^n w_t^{-\frac{1}{2}} \exp\left(-\frac{1}{2\sigma^2} \sum_{t=1}^n w_t [\alpha(\mathbf{L})(\ell_t - \mathbf{A}_t \xi)]^2\right. \\ &\quad \left.+ \frac{\nu}{2} \sum_{t=1}^n (\log w_t - w_t)\right). \end{aligned} \quad (28)$$

In comparison with the likelihood function defined by equation (24) in Koch and Kargoll (2013), the random deviations $\ell_t - \mathbf{A}_t \xi$ are now considered to be autocorrelated, involving the additional (inverted) AR model $\alpha(\mathbf{L})$. The ML estimates within the extended model framework now solve the maximization problem

$$(\hat{\xi}, \hat{\sigma}^2, \hat{\alpha}, \hat{\nu}) = \arg \max_{\xi, \sigma^2, \alpha, \nu} L(\xi, \sigma^2, \alpha, \nu; \ell, \mathbf{w}). \quad (29)$$

To establish that this ML estimation problem and the original one (12) are equivalent (in the sense of Lange et al. 1989), it is required that integrating out the latent variables $\mathbf{w} = (w_1, \dots, w_n)$ from $f(\mathbf{u}, \mathbf{w}|\boldsymbol{\theta}) = L(\boldsymbol{\xi}, \sigma^2, \boldsymbol{\alpha}, \nu; \boldsymbol{\ell}, \mathbf{w})$ results in $f(\mathbf{u}|\boldsymbol{\theta}) = L(\boldsymbol{\xi}, \sigma^2, \boldsymbol{\alpha}, \nu; \boldsymbol{\ell})$. Indeed, forming the marginal density of $\mathcal{U}_t$ from $f(u_t, w_t|\boldsymbol{\theta})$ for every $t \in \{1, \dots, n\}$ in (26) yields on the left-hand side $f(\mathbf{u}|\boldsymbol{\theta})$ as $\mathbf{w}$ is eliminated through integrating out $w_1, \dots, w_n$; on the right-hand side, we have the product of the pdfs (8) as noted right after (20), which product we previously identified with the likelihood function $L(\boldsymbol{\xi}, \sigma^2, \boldsymbol{\alpha}, \nu; \boldsymbol{\ell})$ by virtue of (9) and (11).

Although the original and the alternative model descriptions are equivalent, the underlying theories differ in that they allow for two qualitatively different ways of looking at the outlier modeling scheme. In the first case (i.e., without the latent variables), outliers are taken into account through the thicker tails of the t -distribution in comparison with a corresponding normal distribution. In the second case, the errors are modeled to follow a (conditional) normal distribution (18) where the variance of each error is rescaled by individual factors w_t , which mechanism accommodates for outlying errors by assigning larger variances to them.

As the likelihood function now takes the form of an exponential function, it seems natural to try to obtain the ML estimates from the logarithm of the likelihood function in (28), which can be written as

$$\begin{aligned} \log L(\boldsymbol{\xi}, \sigma^2, \boldsymbol{\alpha}, \nu; \boldsymbol{\ell}, \mathbf{w}) &= -\frac{n}{2} \log(2\pi) - \frac{n}{2} \log(\sigma^2) \\ &+ \frac{n\nu}{2} \log\left(\frac{\nu}{2}\right) - n \log \Gamma\left(\frac{\nu}{2}\right) - \frac{1}{2} \sum_{t=1}^n \log w_t \\ &- \frac{1}{2\sigma^2} \sum_{t=1}^n w_t [\alpha(L)(\ell_t - \mathbf{A}_t \boldsymbol{\xi})]^2 + \frac{\nu}{2} \sum_{t=1}^n (\log w_t - w_t). \end{aligned} \quad (30)$$

As the quantities $w_1, \dots, w_n$ on the right-hand side are neither observed nor parameters, such a direct approach is, however, impossible. In the following section, we will solve this problem by determining the expectation of the log-likelihood function over the assumed distribution of $\mathbf{w}$, i.e., by replacing these unobserved data by their conditional expectations.

Before proceeding with that task, it is instructive to rewrite the likelihood function (28) in terms of sufficient statistics, which often allow for simplifications of statistical procedures in the context of parameter estimation and testing. It is known that the likelihood function admits such a representation in case of the linear model with independently t -distributed random deviations (cf. Liu and Rubin 1995; Little 1988), which constitutes the particular case of $p = 0$ or $\alpha(L) = 1$ in the present model; see also Kargoll (2012) for the concept of suf-

ficient statistics in the more restrictive case of independently and normally distributed random deviations (represented by $p = 0$ and a “large” value for ν). It is also known that sufficient statistics for AR processes with normally distributed white input noise exist (Arato 1961). Concerning the present, more general model, we can write the likelihood function (28) as

$$\begin{aligned} L(\boldsymbol{\theta}; \mathbf{y}) &= (2\pi)^{-\frac{n}{2}} (\sigma^2)^{-\frac{n}{2}} \left(\frac{\nu}{2}\right)^{\frac{n\nu}{2}} \left[\Gamma\left(\frac{\nu}{2}\right)\right]^{-n} \prod_{T=1}^n w_T^{-\frac{1}{2}} \\ &\times \exp\left(-\frac{1}{2\sigma^2} \sum_{t=1}^n w_t (\ell_t - \mathbf{A}_t \boldsymbol{\xi} - \alpha_1 \ell_{t-1} + \alpha_1 \mathbf{A}_{t-1} \boldsymbol{\xi} \right. \\ &\quad \left. - \dots - \alpha_p \ell_{t-p} + \alpha_p \mathbf{A}_{t-p} \boldsymbol{\xi})^2 + \frac{\nu}{2} \sum_{t=1}^n (\log w_t - w_t)\right). \end{aligned} \quad (31)$$

Defining

$$\mathbf{W} = \begin{bmatrix} w_1 & 0 \\ & \ddots \\ 0 & w_n \end{bmatrix}, \quad \mathbf{1}_n = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}, \quad \bar{\mathbf{w}} = \begin{bmatrix} \log w_1 - w_1 \\ \vdots \\ \log w_n - w_n \end{bmatrix}$$

and for every $j \in \{0, \dots, p\}$

$$L^j \boldsymbol{\ell} = \begin{bmatrix} \ell_{1-j} \\ \vdots \\ \ell_{n-j} \end{bmatrix}, \quad L^j \mathbf{A} = \begin{bmatrix} \mathbf{A}_{1-j} \\ \vdots \\ \mathbf{A}_{n-j} \end{bmatrix}$$

with $\mathbf{A}_{i-j} = \mathbf{0}_{[1 \times m]}$ for any $i \leq j$, we can write the likelihood function in the form

$$L(\boldsymbol{\theta}; \mathbf{y}) = g(\mathbf{T}_y, \boldsymbol{\theta}) \cdot h(\mathbf{y}) \quad (32)$$

with $h(\mathbf{y}) = \prod_{T=1}^n w_T^{-\frac{1}{2}}$, $\mathbf{T}_y = [T_1(\mathbf{y}), \dots, T_{3N+1}(\mathbf{y})]^T$ and

$$T_1(\mathbf{y}) = \boldsymbol{\ell}^T \mathbf{W} \boldsymbol{\ell}, \dots, \quad T_N(\mathbf{y}) = (L^p \boldsymbol{\ell})^T \mathbf{W} (L^p \boldsymbol{\ell}), \quad (33)$$

$$T_{N+1}(\mathbf{y}) = \mathbf{A}^T \mathbf{W} \mathbf{A}, \dots, \quad T_{2N}(\mathbf{y}) = (L^p \mathbf{A})^T \mathbf{W} (L^p \mathbf{A}), \quad (34)$$

$$T_{2N+1}(\mathbf{y}) = \mathbf{A}^T \mathbf{W} \boldsymbol{\ell}, \dots, \quad T_{3N}(\mathbf{y}) = (L^p \mathbf{A})^T \mathbf{W} (L^p \boldsymbol{\ell}) \quad (35)$$

$$T_{3N+1}(\mathbf{y}) = \mathbf{1}_n^T \bar{\mathbf{w}} \quad (36)$$

(see “Appendix B”). Consequently, the data-dependent vector $\mathbf{T}_y$ of statistics is sufficient for $\boldsymbol{\theta}$ according to Neyman’s factorization theorem (cf. Section 2.5.2, Kargoll 2012). Here it should be noted that the four sufficient statistics $T_1(\mathbf{y})$, $T_{N+1}(\mathbf{y})$, $T_{2N+1}(\mathbf{y})$ and $T_{3N+1}(\mathbf{y})$ correspond to the sufficient statistics for $(\boldsymbol{\xi}, \sigma^2, \nu)$ in the linear model with

t -distributed random deviations without AR process [see Eqs. (14) and (35) Liu and Rubin 1995]. The additional sufficient statistics in (33)–(36) are necessary because of the unknown autoregressive coefficients. The statistics

$$\begin{aligned} T_2(\mathbf{y}) &= \boldsymbol{\ell}^T \mathbf{W}(L^1 \boldsymbol{\ell}) = \sum_{t=1}^n w_t \ell_t \ell_{t-1} =: (n-1) \cdot \hat{\gamma}_1 \\ &\vdots \\ T_{p+1}(\mathbf{y}) &= \boldsymbol{\ell}^T \mathbf{W}(L^p \boldsymbol{\ell}) = \sum_{t=1}^n w_t \ell_t \ell_{t-p} =: (n-p) \cdot \hat{\gamma}_p \end{aligned}$$

can be interpreted essentially as values of a weighted empirical autocovariance function $\hat{\gamma}$ of the raw data $\boldsymbol{\ell}$. The remaining sufficient statistics arise because the AR coefficients are connected also with the functional model $\mathbf{A}\boldsymbol{\xi}$.

3.3 EM algorithm for the alternative observation model

According to Dempster et al. (1977), the ML estimation of the parameters $\boldsymbol{\theta}$ of a random log-likelihood function $\log L(\boldsymbol{\theta}; \mathcal{Y})$ involving a random data vector $\mathcal{Y} = (\mathcal{Y}_{\text{obs}}, \mathcal{Y}_{\text{lat}})$ of observables $\mathcal{Y}_{\text{obs}}$ and latent variables (or missing data) $\mathcal{Y}_{\text{lat}}$ can be carried out iteratively by computing the estimate $\boldsymbol{\theta}^{(k+1)}$ for any iteration step $k+1$ ($k = 0, 1, \dots$) in such a way that the conditional expectation of the log-likelihood function under the occurred measurement results $\mathcal{Y}_{\text{obs}}$ and parameter estimates $\boldsymbol{\theta}^{(k)}$ from the preceding iteration step is maximized, in the sense of

$$\boldsymbol{\theta}^{(k+1)} = \arg \max_{\boldsymbol{\theta}} E_{\mathcal{Y}_{\text{lat}}|\mathcal{Y}_{\text{obs}}, \boldsymbol{\theta}^{(k)}} \{ \log L(\boldsymbol{\theta}; \mathcal{Y}_{\text{obs}}, \mathcal{Y}_{\text{lat}}) \} \quad (37)$$

(see also Gupta and Chen 2011, Section 1.1). To emphasize that the preceding conditional expectation constitutes a function of $\boldsymbol{\theta}$ involving given parameter values $\boldsymbol{\theta}^{(k)}$, this expectation is usually denoted by the Q -function $Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(k)})$, as suggested by Dempster et al. (1977). As shown for instance in (Section 1.4.1, Gupta and Chen 2011), the conditional expectation can be equivalently formed with respect to the latent variables $\mathcal{Y}_{\text{lat}}$, so that

$$Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(k)}) = E_{\mathcal{Y}_{\text{lat}}|\mathcal{Y}_{\text{obs}}, \boldsymbol{\theta}^{(k)}} \{ \log L(\boldsymbol{\theta}; \mathcal{Y}_{\text{obs}}, \mathcal{Y}_{\text{lat}}) \}. \quad (38)$$

The E-Step consists of the determination of the conditional expectation and the M-Step of the subsequent maximization of the conditional expectation with respect to the variables $\boldsymbol{\theta}$.

3.3.1 The E-step

To apply the previous principles to our specific observation model, we first rewrite (30) as

$$\begin{aligned} \log L(\boldsymbol{\xi}, \sigma^2, \boldsymbol{\alpha}, \nu; \boldsymbol{\ell}, \mathbf{w}) &= -\frac{n}{2} \log(2\pi) - \frac{n}{2} \log(\sigma^2) \\ &+ \frac{n\nu}{2} \log\left(\frac{\nu}{2}\right) - n \log \Gamma\left(\frac{\nu}{2}\right) \\ &- \sum_{t=1}^n \frac{1}{2} \left[\nu + \left(\frac{\boldsymbol{\alpha}(\mathbf{L})(\ell_t - \mathbf{A}_t \boldsymbol{\xi})}{\sigma} \right)^2 \right] w_t \\ &+ \sum_{t=1}^n \frac{1}{2} (\nu - 1) \log w_t. \end{aligned}$$

Next, we identify (i) $\boldsymbol{\theta}$ with the vector of parameters formed by $\boldsymbol{\xi}, \sigma^2, \boldsymbol{\alpha}$ and ν , (ii) $\mathcal{Y}_{\text{obs}}$ with the observables $\mathcal{L}$, and (iii) $\mathcal{Y}_{\text{lat}}$ with the latent weights $\mathcal{W}$, so that the total random data vector $\mathcal{Y}$ consists in our model of $\mathcal{L}$ and $\mathcal{W}$. The Q -function (38) takes then the form

$$Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(k)}) = E_{\mathcal{W}|\mathcal{L}, \boldsymbol{\theta}^{(k)}} \{ \log L(\boldsymbol{\theta}; \mathcal{L}, \mathcal{W}) \}.$$

Since we defined the probabilistic model initially in terms of the white noise $\mathbf{U}$, we condition primarily on the outcome $\mathbf{u}$ (which in turn determines the values $\boldsymbol{\ell}$). We then obtain from the previous equations

$$\begin{aligned} Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(k)}) &= E_{\mathcal{W}|\mathbf{u}, \boldsymbol{\theta}^{(k)}} \{ \log L(\boldsymbol{\theta}; \boldsymbol{\ell}, \mathcal{W}) \} \\ &= -\frac{n}{2} \log(2\pi) - \frac{n}{2} \log(\sigma^2) \\ &+ \frac{n\nu}{2} \log\left(\frac{\nu}{2}\right) - n \log \Gamma\left(\frac{\nu}{2}\right) \\ &- \sum_{t=1}^n \frac{1}{2} \left[\nu + \left(\frac{\boldsymbol{\alpha}(\mathbf{L})(\ell_t - \mathbf{A}_t \boldsymbol{\xi})}{\sigma} \right)^2 \right] E_{\mathcal{W}|\mathbf{u}, \boldsymbol{\theta}^{(k)}} \{ \mathcal{W}_t \} \\ &+ \sum_{t=1}^n \frac{1}{2} (\nu - 1) E_{\mathcal{W}|\mathbf{u}, \boldsymbol{\theta}^{(k)}} \{ \log \mathcal{W}_t \}. \end{aligned} \quad (39)$$

Here, the two conditional expectations can be simplified to

$$\begin{aligned} E_{\mathcal{W}|\mathbf{u}, \boldsymbol{\theta}^{(k)}} \{ \mathcal{W}_t \} &= E_{\mathcal{W}_t|\mathbf{u}, \boldsymbol{\theta}^{(k)}} \{ \mathcal{W}_t \} \\ &= \int_0^\infty w_t f(w_t|\mathbf{u}; \boldsymbol{\theta}^{(k)}) dw_t \\ E_{\mathcal{W}|\mathbf{u}, \boldsymbol{\theta}^{(k)}} \{ \log \mathcal{W}_t \} &= E_{\mathcal{W}_t|\mathbf{u}, \boldsymbol{\theta}^{(k)}} \{ \log \mathcal{W}_t \} \\ &= \int_0^\infty \log w_t f(w_t|\mathbf{u}; \boldsymbol{\theta}^{(k)}) dw_t \end{aligned}$$

by applying basic properties of conditional expectations of functions of real-valued random variables (cf. Chapter 2, Parzen 1999). In light of the multiplication rule, the method of marginalization and the factorization (25), we observe that

the conditional pdf occurring in the conditional expectation $E_{\mathcal{W}_t|u;\theta^{(k)}}\{\cdot\}$ satisfies

$$f(w_t|u, \theta^{(k)}) = f(w_t|u_t, \theta^{(k)}) \quad (t = 1, \dots, n)$$

(see Gupta and Chen 2011, equation (1.8) and the subsequent proof with weights $\mathbf{w}$ and noise values $\mathbf{u}$ in place of the total data x and the observation y , respectively). Therefore, we obtain

$$\begin{aligned} E_{\mathcal{W}_t|u;\theta^{(k)}}\{\mathcal{W}_t\} &= \int_0^\infty w_t f(w_t|u_t; \theta^{(k)}) dw_t \\ &= E_{\mathcal{W}_t|u_t;\theta^{(k)}}\{\mathcal{W}_t\}, \\ E_{\mathcal{W}_t|u;\theta^{(k)}}\{\log \mathcal{W}_t\} &= \int_0^\infty \log w_t f(w_t|u_t; \theta^{(k)}) dw_t \\ &= E_{\mathcal{W}_t|u_t;\theta^{(k)}}\{\log \mathcal{W}_t\}. \end{aligned}$$

Here, the conditional pdf $f(w_t|u_t; \theta)$ can be obtained from the joint pdf (20) in exactly the same way as described in Koch and Kargoll (2013, see Equation (27)). This pdf takes the form of (68) with $x = w_t$, $a = \frac{\nu+1}{2}$, and $b = \frac{1}{2} \left[\nu + \left(\frac{u_t}{\sigma} \right)^2 \right]$, so that each $\mathcal{W}_t$ follows a gamma distribution $G(a, b)$ with parameters a and b , conditional on the value $u_t = \alpha(\mathbf{L})(\ell_t - \mathbf{A}_t \boldsymbol{\xi})$. Since the expected value of a $G(a, b)$ -distributed random variable is given by a/b , we obtain with the current parameter estimates $\theta^{(k)}$ for the desired conditional expectation of the latent weight

$$w_t^{(k)} := E_{\mathcal{W}_t|u_t;\theta^{(k)}}\{\mathcal{W}_t\} = \frac{\nu^{(k)} + 1}{\nu^{(k)} + \left(\frac{\alpha^{(k)}(\mathbf{L})(\ell_t - \mathbf{A}_t \boldsymbol{\xi}^{(k)})}{\sigma^{(k)}} \right)^2}. \quad (40)$$

It can also be shown that the expectation of the natural logarithm of a $G(a, b)$ -distributed random variable is $\psi(a) - \log(b)$ where $\psi(\cdot)$ is the digamma function (see “Appendix A.2”). As shown by Koch and Kargoll (2013, see Equation (34)), the second desired conditional expectation then follows to be

$$\begin{aligned} E_{\mathcal{W}_t|u;\theta^{(k)}}\{\log \mathcal{W}_t\} &= \log w_t^{(k)} + \psi \left(\frac{\nu^{(k)} + 1}{2} \right) \\ &\quad - \log \left(\frac{\nu^{(k)} + 1}{2} \right). \end{aligned} \quad (41)$$

Substituting (40) and (41) for the conditional expectations in (39) and combining furthermore all of the constant (const.) terms which do not depend on any of the unknowns ($\boldsymbol{\xi}$, σ^2 , α , ν) (which may therefore be neglected in subsequent maximization of the Q -function with respect to these parameters), we arrive after further rearrangement and simplification at

$$\begin{aligned} Q(\boldsymbol{\theta}|\theta^{(k)}) &= E_{\mathcal{W}_t|u;\theta^{(k)}}\{\log L(\boldsymbol{\theta}; \boldsymbol{\ell}, \mathcal{W})\} \\ &= \text{const.} - \frac{n}{2} \log \sigma^2 - \frac{1}{2\sigma^2} \sum_{t=1}^n w_t^{(k)} [\alpha(\mathbf{L})(\ell_t - \mathbf{A}_t \boldsymbol{\xi})]^2 \\ &\quad + \frac{n\nu}{2} \log \nu - n \log \Gamma \left(\frac{\nu}{2} \right) + \frac{n\nu}{2} \left[\psi \left(\frac{\nu^{(k)} + 1}{2} \right) \right. \\ &\quad \left. - \log \left(\frac{\nu^{(k)} + 1}{2} \right) + \frac{1}{n} \sum_{t=1}^n (\log w_t^{(k)} - w_t^{(k)}) \right]. \end{aligned} \quad (42)$$

Before deriving the M-Step based on this representation of the Q -function, we note in light of (30) that taking the natural logarithm of (31) creates a linear dependence of the log-likelihood function on the weights w_t and their natural logarithms $\log w_t$. Consequently, the conditional expectations of the corresponding random variables affect only the components of the weight matrix $\mathbf{W}$ and the vector $\bar{\mathbf{w}}$ within the sufficient statistics (33)–(36). The structure of the resulting particular sufficient statistics $T_{N+1}^{(k)} = \mathbf{A}^T \mathbf{W}^{(k)} \mathbf{A}$ and $T_{2N+1}^{(k)} = \mathbf{A}^T \mathbf{W}^{(k)} \boldsymbol{\ell}$, being based on an iteratively updated weight matrix, suggests the combination $(\mathbf{A}^T \mathbf{W}^{(k)} \mathbf{A})^{-1} \mathbf{A}^T \mathbf{W}^{(k)} \boldsymbol{\ell}$ for the purpose of estimating the functional parameters $\boldsymbol{\xi}$. Thus, before actually deriving the first-order conditions of the M-Step, we can already anticipate an IRLS scheme, at least for that parameter group.

3.3.2 The M-step

The M-Step consists of the computation of the zeros of the first partial derivatives of the Q -function (42) with respect to the four different groups of parameters $\boldsymbol{\xi}$, α , σ^2 , and ν . As shown in “Appendix C.1”, we find for the parameter group $\boldsymbol{\xi}$

$$\mathbf{0} = \begin{bmatrix} \alpha(\mathbf{L})A_{1,1} & \cdots & \alpha(\mathbf{L})A_{n,1} \\ \vdots & & \vdots \\ \alpha(\mathbf{L})A_{1,m} & \cdots & \alpha(\mathbf{L})A_{n,m} \end{bmatrix} \mathbf{W}^{(k)} \begin{bmatrix} \alpha(\mathbf{L})(\ell_1 - \mathbf{A}_1 \boldsymbol{\xi}) \\ \vdots \\ \alpha(\mathbf{L})(\ell_n - \mathbf{A}_n \boldsymbol{\xi}) \end{bmatrix}. \quad (43)$$

Note that $\boldsymbol{\xi}$ and α obviously cannot be determined simultaneously and independently. We then consider $\boldsymbol{\xi}$ to be the parameter group estimated first and determine its new estimate $\boldsymbol{\xi}^{(k+1)}$ by fixing the values $\alpha = \alpha^{(k)}$ based on the preceding iteration step k . This calculation of $\boldsymbol{\xi}^{(k+1)}$ thus corresponds to a *conditional maximization* (CM) step within the current M-Step (in the sense of Meng and Rubin 1993). Denoting the $\alpha^{(k)}(\mathbf{L})$ -filtered components of the observation vector and design matrix as

$$\bar{\ell}_t^{(k)} := \alpha^{(k)}(\mathbf{L})\ell_t, \quad (44)$$

$$\bar{A}_{t,j}^{(k)} := \alpha^{(k)}(\mathbf{L})A_{t,j}, \quad (45)$$

$$\bar{\mathbf{A}}_t^{(k)} := \alpha^{(k)}(\mathbf{L})\mathbf{A}_t, \quad (46)$$

we obtain the normal equations

$$\mathbf{0} = \begin{bmatrix} \bar{\mathbf{A}}_{1,1}^{(k)} & \cdots & \bar{\mathbf{A}}_{n,1}^{(k)} \\ \vdots & & \vdots \\ \bar{\mathbf{A}}_{1,m}^{(k)} & \cdots & \bar{\mathbf{A}}_{n,m}^{(k)} \end{bmatrix} \mathbf{W}^{(k)} \begin{bmatrix} \bar{\ell}_1^{(k)} - \bar{\mathbf{A}}_1^{(k)} \boldsymbol{\xi}^{(k+1)} \\ \vdots \\ \bar{\ell}_n^{(k)} - \bar{\mathbf{A}}_n^{(k)} \boldsymbol{\xi}^{(k+1)} \end{bmatrix},$$

$$=: (\bar{\mathbf{A}}^{(k)})^T \mathbf{W}^{(k)} (\bar{\boldsymbol{\ell}}^{(k)} - \bar{\mathbf{A}}^{(k)} \boldsymbol{\xi}^{(k+1)}),$$

consequently

$$\boldsymbol{\xi}^{(k+1)} = \left((\bar{\mathbf{A}}^{(k)})^T \mathbf{W}^{(k)} \bar{\mathbf{A}}^{(k)} \right)^{-1} (\bar{\mathbf{A}}^{(k)})^T \mathbf{W}^{(k)} \bar{\boldsymbol{\ell}}^{(k)}. \quad (47)$$

We can now compute the residuals at the level of the auto-correlated random deviations (i.e., the colored noise)

$$e_t^{(k+1)} := \ell_t - \mathbf{A}_t \boldsymbol{\xi}^{(k+1)}. \quad (48)$$

Setting up the matrix

$$\mathbf{E}^{(k+1)} := \begin{bmatrix} e_0^{(k+1)} & \cdots & e_{1-p}^{(k+1)} \\ \vdots & & \vdots \\ e_{n-1}^{(k+1)} & \cdots & e_{n-p}^{(k+1)} \end{bmatrix} \quad (49)$$

and using the initial conditions $e_0^{(k+1)} = \cdots = e_{1-p}^{(k+1)} = 0$, we can derive a second IRLS scheme to estimate $\boldsymbol{\alpha}$, that is,

$$\boldsymbol{\alpha}^{(k+1)} = \left((\mathbf{E}^{(k+1)})^T \mathbf{W}^{(k)} \mathbf{E}^{(k+1)} \right)^{-1} (\mathbf{E}^{(k+1)})^T \mathbf{W}^{(k)} \mathbf{e}^{(k+1)} \quad (50)$$

(see “Appendix C.2”). Having based the estimation of $\boldsymbol{\alpha}^{(k+1)}$ on the current estimates $\boldsymbol{\xi}^{(k+1)}$, we thus carry out the second CM-Step within the current M-Step. Knowledge of $\boldsymbol{\alpha}^{(k+1)}$ allows us to compute the estimated residuals now at the level of the white noise input to the AR process, by applying the currently designed decorrelation filter $\alpha^{(k+1)}(\mathbf{L})$ to the colored noise residuals, that is,

$$u_t^{(k+1)} = \alpha^{(k+1)}(\mathbf{L}) e_t^{(k+1)}. \quad (51)$$

The third CM-Step then estimates the scaling parameter σ^2 by means of the average weighted sum of squared residuals

$$(\sigma^2)^{(k+1)} = \frac{1}{n} \sum_{t=1}^n w_t^{(k)} \left(u_t^{(k+1)} \right)^2$$

$$= \frac{(\mathbf{u}^{(k+1)})^T \mathbf{W}^{(k)} \mathbf{u}^{(k+1)}}{n} \quad (52)$$

(see “Appendix C.3”).

Regarding the final parameter ν , it is shown in “Appendix C.4” that a solution $\nu^{(k+1)}$ satisfies

$$0 = \log \nu^{(k+1)} - \psi \left(\frac{\nu^{(k+1)}}{2} \right) + \psi \left(\frac{\nu^{(k)} + 1}{2} \right)$$

$$- \log \left(\nu^{(k)} + 1 \right) + \frac{1}{n} \sum_{t=1}^n \left(\log w_t^{(k)} - w_t^{(k)} \right) + 1. \quad (53)$$

According to [Liu and Rubin \(1995\)](#) and [McLachlan and Krishnan \(2008, Section 5.8.4\)](#), we can hope to speed up the convergence of the estimate of ν by maximizing the log-likelihood function (14) with respect to ν , instead of maximizing its conditional expectation $Q(\boldsymbol{\theta} | \boldsymbol{\theta}^{(k)})$. Employing the notation

$$w_t^{(k+1)} = \frac{\nu^{(k+1)} + 1}{\nu^{(k+1)} + \left(\frac{\alpha^{(k+1)}(\mathbf{L})(\ell_t - \mathbf{A}_t \boldsymbol{\xi}^{(k+1)})}{\sigma^{(k+1)}} \right)^2} \quad (54)$$

in analogy to (40) gives us the alternative equation

$$0 = \log \nu^{(k+1)} + 1 - \psi \left(\frac{\nu^{(k+1)}}{2} \right) + \psi \left(\frac{\nu^{(k+1)} + 1}{2} \right)$$

$$- \log \left(\nu^{(k+1)} + 1 \right) + \frac{1}{n} \sum_{t=1}^n \left(\log w_t^{(k+1)} - w_t^{(k+1)} \right) \quad (55)$$

to be solved for $\nu^{(k+1)}$ (see “Appendix C.4”). This step turns the ECM algorithm into a so-called ECM *either* (ECME) algorithm, which is also known to increase the log-likelihood (14) in every iteration step ([Liu and Rubin 1994](#)).

3.4 Implementation of the EM algorithm

The implemented EM algorithm consists of the following steps.

Initial EM-Step: The algorithm starts with the determination of an initial weight matrix $\mathbf{W}^{(0)}$ via (40), setting $k = 0$; if the required initial values $\boldsymbol{\xi}^{(0)}$, $\boldsymbol{\alpha}^{(0)}$, $\sigma^{(0)}$ and $\nu^{(0)}$ are not or only partially available, we assume tentatively that all of the observations contribute with equal weights

$$w_i^{(0)} = 1, \quad (i = 1, \dots, n) \quad (56)$$

to the subsequent M-Step.

CM-Step for $\boldsymbol{\xi}$: The first stage of the M-Step consists of the computation of the functional parameter estimates $\boldsymbol{\xi}^{(1)}$, which requires

knowledge of the AR-filtered quantities (44)–(46) for $k = 0$, according to (47). If these initial values are not given, then we tentatively set

$$\alpha^{(0)} = \mathbf{0}, \quad (57)$$

which corresponds to the initial assumption of uncorrelated random deviations; in this case we can compute the ordinary least squares solution

$$\xi^{(1)} = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \ell. \quad (58)$$

CM-Step for α : The computation of $\alpha^{(1)}$ is then possible by virtue of (50) with $k = 1$. Since (weighted) least-squares estimation of an AR models does not necessarily give a covariance-stationary and invertible process Porat (1994 see p. 173), we check whether all roots of the equation $\alpha^{(1)}(z) = 0$ are within the unit circle, for instance by checking in MATLAB whether the call

$$\text{all}(\text{abs}(\text{roots}([1; -\alpha]))) < 1$$

gives the corresponding value 1. If not, then we mirror the exterior roots (having magnitudes greater than 1) into the unit circle by stabilizing the polynomial (see p. 197, Porat 1994), using the MATLAB routine `polystab.m`. **CM-Step for σ^2 :** After determining the decorrelated residuals by means of the inverted AR model, we estimate $(\sigma^2)^{(1)}$ by using (52). **CME-Step for ν :** The final task of the current first step is the computation of $\nu^{(1)}$. We apply the standard MATLAB routine `fzero.m` to find the zero of (55). Since $\nu \rightarrow \infty$ holds for a normal distribution, the zero $\nu^{(1)}$ does not necessarily exist as a finite, real number. In such a case, the signs of the values that the function in (55) takes for both small and very large degrees of freedom are identical. In our implementation, we compare the signs at $\nu^{(1)} = 10^{-8}$ and $\nu^{(1)} = 10^8$; in case of identical signs, we set $\nu^{(1)} = 10000$ and bypass the call of `fzero.m`. **E-Step:** The E-Step (40) is now carried out to estimate the new weight matrix $\mathbf{W}^{(1)}$, given the current estimates $\theta^{(1)} = (\xi^{(1)}, (\sigma^2)^{(1)}, \alpha^{(1)}, \nu^{(1)})$.

M-Step: The M-Step, which consists of the CM(E)-Steps with respect to the four parameter groups, determines $\theta^{(2)}$ via (47), (50), (52), and (55). Note that the estimates $\alpha^{(1)}$ of the preceding step are employed for the computation of $\xi^{(2)}$.

Termination step: Finally, the newly estimated parameters $\theta^{(2)}$ are tested for convergence by comparing them to

$\theta^{(1)}$. For this purpose, we determine the greatest absolute value of the differences between the estimates of iteration 2 and iteration 1 with respect to the functional parameters ξ , the AR parameters α , and the scaling parameter σ^2 . If that maximum is less than 10^{-8} and if, in addition $|\nu^{(2)} - \nu^{(1)}| < 10^{-4}$, then the algorithm stops. Otherwise, we repeat the E-Step and subsequently the M-Step, until the stop-criterion is eventually satisfied by the estimates (or the maximum number $itermax = 500$ of iterations is reached). Evaluation of the original log-likelihood function (14) for the current parameter estimates yields another possibility for assessing the overall convergence of the EM algorithm.

4 Monte Carlo results

In this section we consider two different linear regression models. The first model

$$\mathcal{L}_t = \frac{a_0}{2} + \sum_{j=1}^{12} a_j \cos(2\pi f_j x_t) + b_j \sin(2\pi f_j x_t) + \mathcal{E}_t \quad (59)$$

($t = 1, \dots, n$) consists of sine and cosine basis functions with unknown (“Fourier”) coefficients $a_0, a_1, \dots, a_{12}$ and $b_1, \dots, b_{12}$, collected in the parameter vector ξ . The Fourier frequencies

$$f_j = 8 \cdot j \text{ [Hz]} \quad (j = 1, \dots, 12) \quad (60)$$

are treated as error-free quantities. Furthermore, $x_1, \dots, x_n$ are given time instances sampled with constant sampling rate $\Delta x = 0.00512$ s, beginning at time $x_1 = 67.68128$ s. The number of observations is either $n = 100$, $n = 1000$, $n = 10,000$ or $n = 100,000$. This functional model gives rise to design matrices $\mathbf{A}_{[n \times 25]}$ differing essentially in the number of rows.

The second model

$$\mathcal{L}_t = \sum_{j=1}^{10} a_j b_j(x_t) + \mathcal{E}_t \quad (t = 1, \dots, n) \quad (61)$$

is based on shifted cubic splines $b_j(x) = b(x - j - 1)$ with

$$b(x) = \begin{cases} 0 & \text{if } x < -2 \\ \frac{1}{6}(x+2)^3 & \text{if } -2 \leq x < -1 \\ \frac{1}{6}(x+2)^3 - \frac{2}{3}(x+1)^3 & \text{if } -1 \leq x < 0 \\ \frac{1}{6}(2-x)^3 - \frac{2}{3}(1-x)^3 & \text{if } 0 \leq x < 1 \\ \frac{1}{6}(2-x)^3 & \text{if } 1 \leq x < 2 \\ 0 & \text{if } x \geq 2 \end{cases} \quad (62)$$

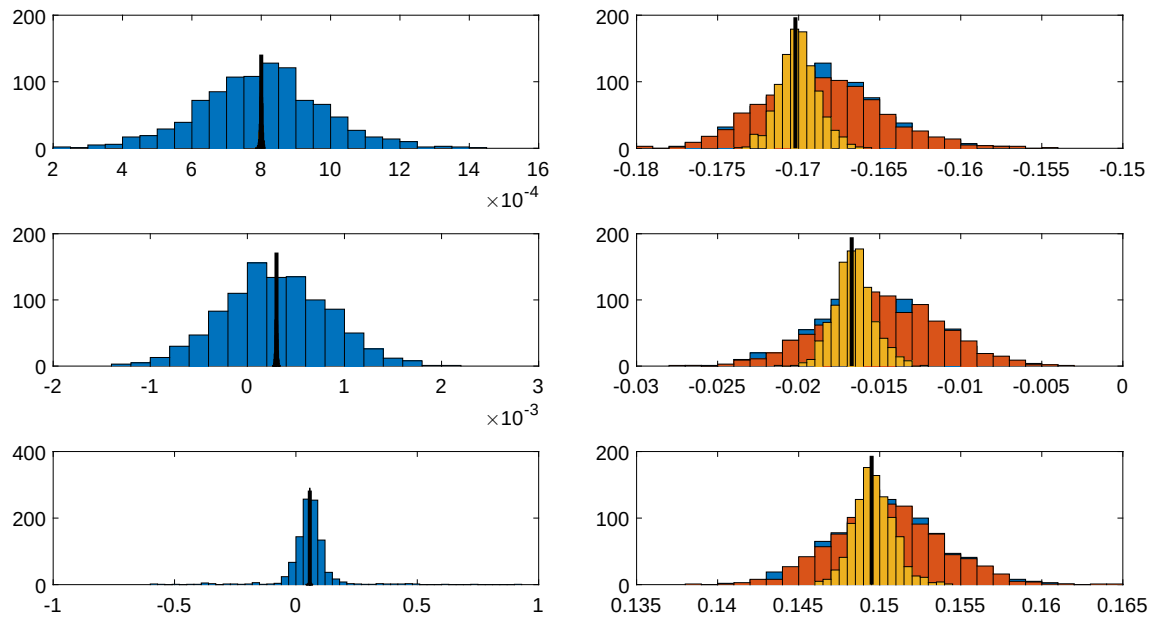


Fig. 1 Histogram of the Monte Carlo estimates of the coefficients a_0 (upper left), a_1 (middle left), a_2 (lower left) and the AR coefficients α_1 (upper right), α_2 (middle right), α_3 (lower right) for $n = 100,000$ observations of the Fourier model with AR(100) deviations and white

noise following the $t_2(0, 0.001^2)$ distribution (yellow), the $N(0, 0.001^2)$ distribution (red), and the $0.6 \cdot N(0, 0.001^2) + 0.4 \cdot N(0, 0.008^2)$ distribution (blue). Heavy black lines indicate the location of the true value

The time instances are sampled equidistantly between $x_1 = 0.1$ s and $x_n = 9.9$ s with rate $\Delta x = 9.8/(n-1)$, where again $n \in \{100, 1000, 10,000, 100,000\}$. We thus obtain four differently sized design matrices $\mathbf{A}_{[n \times 10]}$ for the spline model.

The random deviations in (59) and (61) satisfy the AR model equation

$$\mathcal{E}_t = \sum_{j=1}^p \alpha_j \mathcal{E}_{t-j} + \mathcal{U}_t, \quad (t = 1, \dots, n), \quad (63)$$

with $p \in \{0, 1, 10, 100\}$; thus, the random deviations either constitute white noise (in case of $p = 0$) or colored noise. The white noise components $\mathcal{U}_1, \dots, \mathcal{U}_n$ are generated independently either by: (i) the central t -distribution with degree of freedom $\nu = 2$ and scale parameter $\sigma = 0.001$, i.e.,

$$\mathcal{U}_t \stackrel{\text{ind}}{\sim} t_2(0, 0.001^2), \quad (64)$$

or (ii) the central normal distribution with standard deviation $\sigma = 0.001$, i.e.,

$$\mathcal{U}_t \stackrel{\text{ind}}{\sim} N(0, 0.001^2), \quad (65)$$

or (iii) the central contaminated normal distribution

$$\mathcal{U}_t \stackrel{\text{ind}}{\sim} 0.6 \cdot N(0, 0.001^2) + 0.4 \cdot N(0, 0.008^2). \quad (66)$$

whereas the first two distributions (64) and (65) constitute special cases of the family of scaled t -distributions, the con-

taminated normal distribution is not a member of that family. The random number generator of MATLAB is used to generate 1000 samples for each of the distributions and each of the sample sizes. These white noise samples are turned into autocorrelated colored noise by means of the AR(1), AR(10) and AR(100) model. To avoid the distorting warm-up effect resulting from the constant initial values used in that recursive filter operation, the generated white noise samples are initially much longer than the desired sample length n , and only the tail of length n of the filtered colored noise sequence is kept. The true coefficients of the AR models are listed in Table 7 (“Appendix D”). The noise samples are then added to the true observations $\mathbf{A}\xi$; the true parameter values for the Fourier and the spline model are given in Table 8 (“Appendix D”).

The ECME algorithm described in Sect. 3.4 is used to adjust each generated Monte Carlo sample for the observations, based on the design matrix and autoregressive model order employed to generate that particular observation sample. Thus, from the final ECME iteration steps, we obtain the parameter estimates $\hat{\xi}$, $\hat{\alpha}$, $\hat{\sigma}$ and $\hat{\nu}$ for each regression model, each AR model, each sampling distribution, and each Monte Carlo run.

The implemented ECME algorithm typically needs at least ten iteration steps in order for the solution to satisfy the convergence criterion; in some cases, a few hundred iteration steps may be necessary. The estimation results are summarized in Figs. 1, 2, 3 and 4 and Tables 1, 2, 3, 4, 5 and 6.

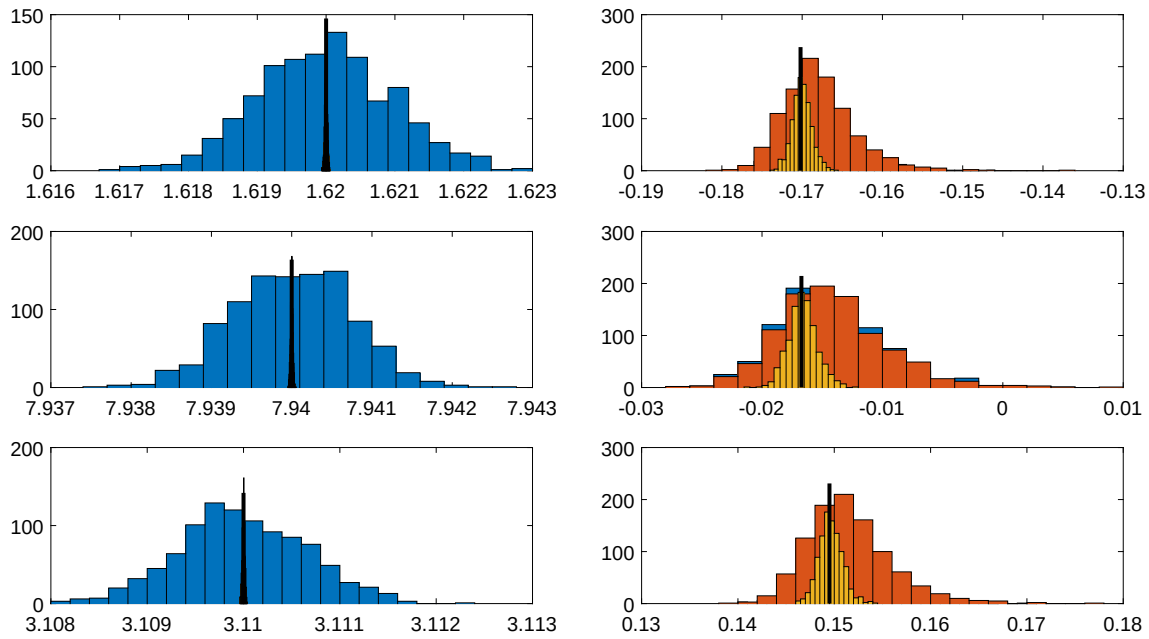


Fig. 2 Histogram of the Monte Carlo estimates of the coefficients a_0 (upper left), a_1 (middle left), a_2 (lower left) and the AR coefficients α_1 (upper right), α_2 (middle right), α_3 (lower right) for $n = 100,000$ observations of the spline model with AR(100) deviations and white

noise following the $t_2(0, 0.001^2)$ distribution (yellow), the $N(0, 0.001^2)$ distribution (red), and the $0.6 \cdot N(0, 0.001^2) + 0.4 \cdot N(0, 0.008^2)$ distribution (blue). Heavy black lines indicate the location of the true value

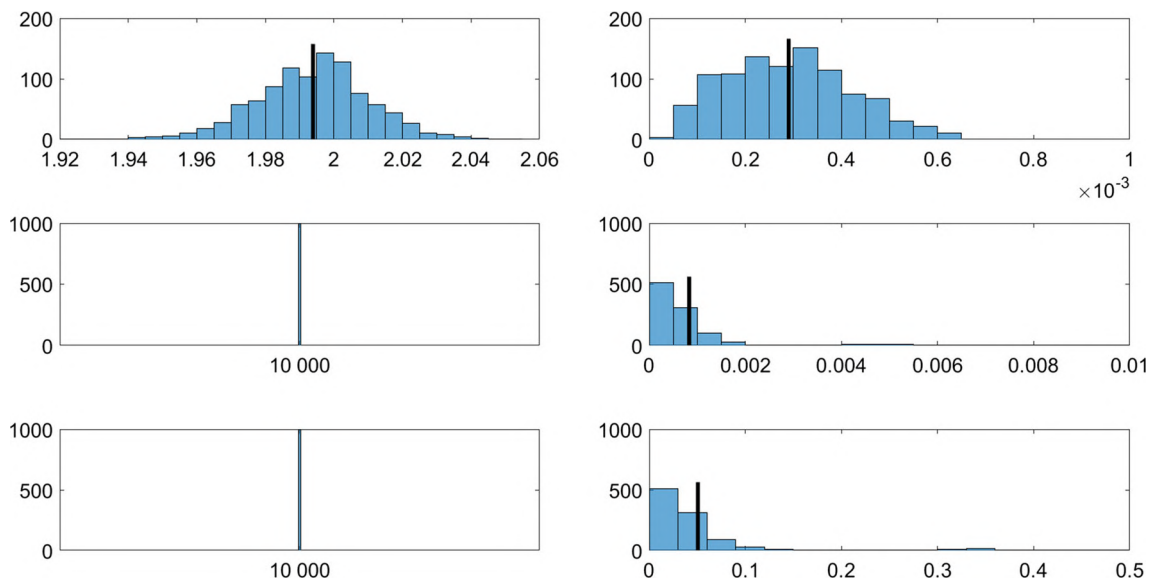


Fig. 3 Histogram of the Monte Carlo estimates of v (left column) and the RMSE (right column) for $n = 100,000$ observations of the Fourier model with AR(100) deviations and white noise following the

$t_2(0, 0.001^2)$ distribution (first row), the $N(0, 0.001^2)$ distribution (second row), and the $0.6 \cdot N(0, 0.001^2) + 0.4 \cdot N(0, 0.008^2)$ distribution (third row). Heavy black lines indicate the location of the mean value

Concerning the functional model parameters ξ , the tables give in particular the means of the estimates of the first Fourier coefficient ($\hat{a}_0/2$) and of the first spline coefficient ($\hat{a}_1$), computed from the 1000 Monte Carlo runs. The approximation of the true parameter values by these means clearly

improves with increasing sample size n , and the bias appears to be insignificant by $n = 100,000$. To validate this finding, the histograms of different functional model parameters were computed for the case of $n = 100,000$ observations and an AR model order of $p = 100$. The left columns in Figs. 1 and

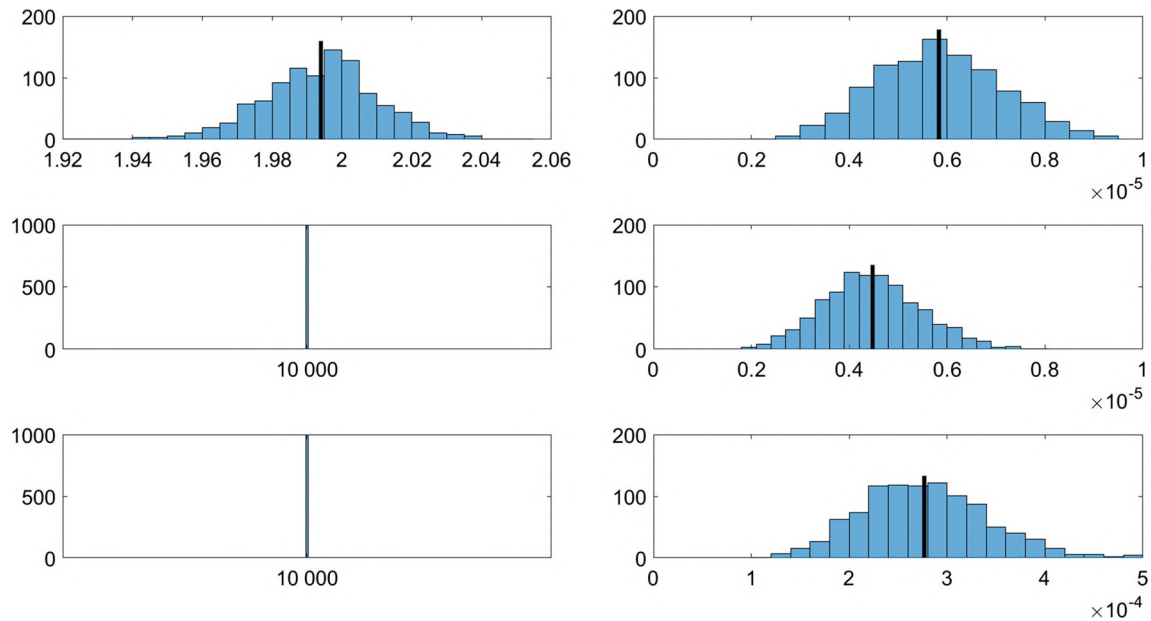


Fig. 4 Histogram of the Monte Carlo estimates of v (left column) and the RMSE (right column) for $n = 100,000$ observations of the spline model with AR(100) deviations and white noise following the

$t_2(0, 0.001^2)$ distribution (first row), the $N(0, 0.001^2)$ distribution (second row), and the $0.6 \cdot N(0, 0.001^2) + 0.4 \cdot N(0, 0.008^2)$ distribution (third row). Heavy black lines indicate the location of the mean value

2 show the histograms for the first three estimated parameters of the two regression models. Here we see that the estimates of the functional parameters are indeed practically unbiased across all three sampling distributions for both regression models. Furthermore, the spread of the histograms is generally largest in case of the contaminated normal distribution, and smallest in case of the t -distribution. The histograms regarding the t - and the normal distribution are very similar, and they almost coincide with the heavy black lines that indicate the true parameter values. Thus, the precision of the estimated parameters is much higher for these two sampling distributions than for the contaminated normal distribution.

To assess the goodness of fit of the AR models, a white noise test (WNT) is applied within each Monte Carlo run to the decorrelation-filtered residuals $\hat{u}_1, \dots, \hat{u}_n$, which are expected to approximately constitute white noise, if the estimation of the AR coefficients were successful. The test is based on the empirical periodogram of the estimated noise $\hat{u}_1, \dots, \hat{u}_n$. For this purpose, the MATLAB routine `periodogram` is used to compute the raw, one-sided periodogram $I_1, \dots, I_M$ with the sampling rate (in Hz) determined by $Fs = 1/\Delta x$. The periodogram values for all Fourier frequencies $1/n, \dots, M/n$ (where M is the lower integer of $n/2$) are used to compute the normalized cumulated periodogram values

$$S_0 = 0, \quad S_i = \frac{\sum_{k=1}^i I_k}{\sum_{k=1}^M I_k} \quad (i = 1, \dots, M).$$

The test statistic determines the maximum cumulated periodogram excess over a cumulated white noise periodogram, that is,

$$T = \max_i |S_i - i/M|.$$

The white noise hypothesis is rejected if this maximum excess is too large compared to a critical value based on a specified significance level $1 - \alpha$. Since the standard distribution of this test (cf. Section 7.3.3, Schlittgen and Streitberg 2001) cannot be applied reliably to our adjustment problem in view of the estimation step and the non-normality of the data, the critical value is simulated via Monte Carlo. More specifically, the critical value of the test is determined individually for each sample size n and each probability distribution in such a way that the acceptance rate, throughout all Monte Carlo runs with generated random deviations $e_1, \dots, e_n$ following the AR(0)-white noise model, is identical with the desired significance level 0.95. This critical value is then employed for the current sample size and probability distribution to determine the acceptance rates with respect to the estimated AR(1), AR(10) and AR(100) models (see the WNT values in Tables 1, 2, 3, 4, 5 and 6).

Generally, the WNT acceptance rates increase with sample size and reach at least 90% for $n = 100,000$. The AR models up to order $p = 10$ are already estimated reasonably well for the sample size $n = 1000$, whereas a reliable estimation of an AR(100) model in terms of the WNT acceptance rate requires more than 10,000 observations. As with

Table 1 Model: student's $t_2(0, 0.001^2) +$ Fourier series with mean $a_0/2 = 0.00080$ for $n = 100/1000/10,000/100,000$

AR(0)	AR(1)	AR(10)	AR(100)
Mean ($\hat{a}_0/2$)			
0.00082	0.00078	0.00078	NaN
0.00080	0.00080	0.00080	NaN
0.00080	0.00079	0.00080	0.00080
0.00080	0.00080	0.00080	0.00080
WNT			
0.950	0.978	0.779	NaN
0.950	0.963	0.948	NaN
0.950	0.955	0.950	0.920
0.950	0.953	0.957	0.902
Mean ($\hat{\sigma}$)			
0.0016	0.0002	0.0025	NaN
0.0013	0.0010	0.0010	NaN
0.0010	0.0010	0.0010	0.0010
0.0010	0.0010	0.0010	0.0010
Mode ($\hat{\nu}$)			
9500.01	500.10	500.10	NaN
501.48	1.91	501.41	NaN
501.75	1.97	1.97	2.00
2.01	2.00	2.00	2.00
Mean (RMSE)			
0.00107	0.00121	0.00647	NaN
0.00025	0.00027	0.00552	NaN
0.00006	0.00008	0.00064	0.00421
0.00002	0.00003	0.00016	0.00029

The shown quantities are computed from the 1000 Monte Carlo runs: (1) Mean($\hat{a}_0/2$) is the (arithmetic) mean of the estimated first coefficient of the Fourier series, (2) WNT is the acceptance rate of the periodogram-based white noise test, (3) Mean($\hat{\sigma}$) is the mean of the estimated scale factor, (4) Mode($\hat{\nu}$) is the mode of the estimated degrees of freedom, and (5) Mean(RMSE) is the mean of the root-mean-square error with respect to the true error-free observations and the adjusted observations. "NaN" indicates models that were not computed

the functional model parameters, we computed histograms of different estimated AR coefficients in order to evaluate the biases and variances. The right columns of Figs. 1 and 2 show the histograms of the first three coefficients. On the one hand, the AR coefficients for the normal and the contaminated normal distribution (whose histograms are very similar and therefore strongly overlapping) are slightly biased, whereas the estimates resulting from the t -distribution appear to be unbiased. On the other hand, the spread of the histograms is largest now for both the normal and for the contaminated normal sampling distribution, whereas the t -distribution yields relatively precise estimates.

The performance of the estimation of the scale factor σ with respect to the $t_\nu(0, \sigma^2)$ distribution underlying the ECME algorithm can be assessed only in the two cases where

Table 2 Model: student's $t_2(0, 0.001^2) +$ spline with first coefficient $a_1 = 1.62$ for $n = 100/1000/10,000/100,000$

AR(0)	AR(1)	AR(10)	AR(100)
Mean ($\hat{a}_1$)			
1.61996	1.61993	1.62024	NaN
1.61998	1.62000	1.61999	NaN
1.62000	1.62000	1.62000	1.61999
1.62000	1.62000	1.62000	1.62000
WNT			
0.950	0.978	0.624	NaN
0.950	0.958	0.878	NaN
0.950	0.955	0.950	0.711
0.950	0.953	0.957	0.902
Mean ($\hat{\sigma}$)			
0.0018	0.0009	0.0023	NaN
0.0013	0.0010	0.0010	NaN
0.0010	0.0010	0.0010	0.0010
0.0010	0.0010	0.0010	0.0010
Mode ($\hat{\nu}$)			
9500.04	500.81	500.10	NaN
501.50	1.90	1.95	NaN
501.75	2.01	1.97	1.93
2.00	2.00	2.00	2.00
Mean (RMSE)			
0.00063	0.00132	0.00086	NaN
0.00015	0.00041	0.00007	NaN
0.00004	0.00013	0.00002	0.00002
0.00001	0.00004	0.00001	0.00001

The shown quantities are computed from the 1000 Monte Carlo runs: Mean($\hat{a}_1$) is the arithmetic mean of the estimated first coefficient of the spline model, the other quantities are as described in Table 1

the white noise sampling distribution is given by (64) or (65), because the latter distributions are special cases of the family of scaled t -distributions. In these two cases, the mean value of the Monte Carlo estimates $\hat{\sigma}$ coincides with the true value 0.001 already for $n = 10,000$ (see Tables 1, 2, 3, 4), regardless of the particular form of the regression and AR model. It thus becomes evident that when sampling is carried out by means of the contaminated normal distribution (66), the estimated scale factor is not able to capture the effect of the two different variance components in the data.

The evaluation of the algorithm's performance in estimating the degree of freedom of the underlying t -distribution is based on the mode of the Monte Carlo estimates $\hat{\nu}$ since the shape of their histograms is slightly non-symmetric (see Figs. 3, 4, left columns). As for the scale factor, the two sampling distributions (64) and (65) allow for direct comparisons of the mode of $\hat{\nu}$ with the corresponding true values $\nu = 2$ and $\nu \rightarrow \infty$. As mentioned in Sect. 3.4, the maximum value of an estimated $\hat{\nu}$ is 10,000 for numerical reasons, which

Table 3 Model: Gaussian $N(0, 0.001^2)$ + Fourier series with mean $a_0/2 = 0.00080$ for $n = 100/1000/10,000/100,000$

AR(0)	AR(1)	AR(10)	AR(100)
Mean ($\hat{a}_0/2$)			
0.00081	0.00082	0.00080	NaN
0.00080	0.00079	0.00080	NaN
0.00080	0.00080	0.00080	0.00080
0.00080	0.00080	0.00080	0.00080
WNT			
0.950	0.981	0.828	NaN
0.950	0.981	0.998	NaN
0.950	0.979	1.000	1.000
0.950	0.984	1.000	1.000
Mean ($\hat{\sigma}$)			
0.0009	0.0007	0.0012	NaN
0.0010	0.0010	0.0010	NaN
0.0010	0.0010	0.0010	0.0010
0.0010	0.0010	0.0010	0.0010
Mode ($\hat{\nu}$)			
10,000.00	9500.01	9500.01	NaN
10,000.00	9500.48	9500.46	NaN
10,000.00	9501.72	526.98	530.93
10,000.00	10,000.00	10,000.00	10,000.00
Mean (RMSE)			
0.00049	0.00071	0.00245	NaN
0.00016	0.00021	0.00229	NaN
0.00005	0.00007	0.00104	0.00279
0.00002	0.00002	0.00016	0.00083

The shown quantities are as described in Table 1

we therefore take as a sufficient approximation of $\nu \rightarrow \infty$. Tables 1, 2, 3 and 4 show that the modes of the estimated $\hat{\nu}$ converge to the given true values for both regression models and for all AR processes. The results for the contaminated normal sampling distribution are quite different for the various sample sizes. For instance, the most frequent estimates for ν are values around 3 for $n = 10,000$, and values around 10,000 for $n = 100,000$.

The root-mean-square error

$$\text{RMSE} = \sqrt{\sum_{t=1}^n (\hat{e}_t - e_t)^2} = \sqrt{(\mathbf{A}\hat{\xi} - \mathbf{A}\xi)^T (\mathbf{A}\hat{\xi} - \mathbf{A}\xi)}$$

measures the estimator's ability to predict the true observations $\mathbf{A} = \mathbf{A}\xi$. Here, $\hat{e}_t$ denotes the estimate of the residual (2) at the level of the observation equations. Since the predicted or adjusted observations $\mathbf{A}\hat{\xi}$ are a consequence of the estimation of all four parameter groups ξ , α , σ^2 and ν , the RMSE expresses the overall performance of the proposed ECME algorithm. This error measure includes both

Table 4 Model: Gaussian $N(0, 0.001^2)$ + spline with first coefficient $a_1 = 1.62$ for $n = 100/1000/10,000/100,000$

AR(0)	AR(1)	AR(10)	AR(100)
Mean ($\hat{a}_1$)			
1.62002	1.62007	1.61993	NaN
1.61999	1.61994	1.62000	NaN
1.62000	1.61999	1.62000	1.62000
1.62000	1.62000	1.62000	1.62000
WNT			
0.950	0.991	0.728	NaN
0.950	0.989	0.989	NaN
0.950	0.985	1.000	0.948
0.950	0.984	1.000	1.000
Mean ($\hat{\sigma}$)			
0.0009	0.0009	0.0019	NaN
0.0010	0.0010	0.0010	NaN
0.0010	0.0010	0.0010	0.0010
0.0010	0.0010	0.0010	0.0010
Mode ($\hat{\nu}$)			
10,000.00	9500.10	500.48	NaN
10,000.00	9500.59	506.40	NaN
10,000.00	9501.75	520.91	509.59
9505.42	10,000.00	10,000.00	10,000.00
Mean (RMSE)			
0.00031	0.00088	0.00051	NaN
0.00010	0.00030	0.00005	NaN
0.00003	0.00010	0.00002	0.00001
0.00001	0.00003	0.00001	0.00000

The shown quantities are as described in Table 2

the variance and the bias of the estimator, and should therefore approach 0 for increasing sample size. The RMSE is computed for each Monte Carlo run, and the resulting mean values are given in Tables 1, 2, 3, 4, 5 and 6. It can be seen that the mean RMSE is substantially reduced with each increase in sample size n for both regression models and all AR model orders. Moreover, the right columns of Figs. 3 and 4 show that the range of values of the RMSE is about two magnitudes larger for the Fourier model in comparison with the spline model, independently of the employed sampling distribution. In addition, it can be seen that the distribution of the RMSE is approximately normal for the spline model, contrary to the Fourier model. One reason for these large discrepancies could be that the Fourier model is more unstable due to its interaction with the AR model. Further investigations concerning this issue should be carried out in the future.

We conclude that the implemented ECME algorithm can be expected to give approximately unbiased estimates for the various model parameters if a sufficiently large number of observations is available. This number is clearly influenced

Table 5 Model: mixture normal $0.6 \cdot N(0, 0.001^2) + 0.4 \cdot N(0, 0.008^2) + \text{Fourier series with mean } a_0/2 = 0.00080$ for $n = 100/1000/10,000/100,000$

AR(0)	AR(1)	AR(10)	AR(100)
Mean ($\hat{a}_0/2$)			
0.00176	0.00408	0.00093	NaN
0.00071	0.00061	0.00079	NaN
0.00074	0.00066	0.00078	0.00078
0.00080	0.00080	0.00080	0.00080
WNT			
0.950	0.985	0.808	NaN
0.950	0.979	0.998	NaN
0.950	0.970	0.993	1.000
0.950	0.976	1.000	1.000
Mean ($\hat{\sigma}$)			
0.0530	0.0148	0.0658	NaN
0.0607	0.0399	0.0406	NaN
0.0616	0.0407	0.0407	0.0407
0.0616	0.0616	0.0616	0.0617
Mode ($\hat{\nu}$)			
10,000.00	500.10	500.10	NaN
10,000.00	3.01	502.08	NaN
10,000.00	3.04	3.03	3.04
10,000.00	10,000.00	10,000.00	10,000.00
Mean (RMSE)			
0.03064	0.04409	0.14870	NaN
0.00965	0.01038	0.13762	NaN
0.00303	0.00318	0.02563	0.15250
0.00096	0.00127	0.00941	0.05067

The shown quantities are as described in Table 1

Table 6 Model: mixture normal $0.6 \cdot N(0, 0.001^2) + 0.4 \cdot N(0, 0.008^2) + \text{Spline with first coefficient } a_1 = 1.62$ for $n = 100/1000/10,000/100,000$

AR(0)	AR(1)	AR(10)	AR(100)
Mean ($\hat{a}_1$)			
1.61874	1.61935	1.62191	NaN
1.61972	1.61763	1.61994	NaN
1.62001	1.62014	1.62001	1.61997
1.62000	1.61991	1.61999	1.61998
WNT			
0.950	0.985	0.683	NaN
0.950	0.976	0.953	NaN
0.950	0.969	0.993	0.739
0.950	0.975	1.000	1.000
Mean ($\hat{\sigma}$)			
0.0582	0.0372	0.0938	NaN
0.0612	0.0405	0.0403	NaN
0.0616	0.0407	0.0407	0.0406
0.0407	0.0616	0.0616	0.0617
Mode ($\hat{\nu}$)			
10,000.00	501.05	500.10	NaN
10,000.00	3.09	2.82	NaN
10,000.00	3.05	3.04	2.99
3.05	10,000.00	10,000.00	10,000.00
Mean (RMSE)			
0.01900	0.04762	0.02787	NaN
0.00598	0.01494	0.00261	NaN
0.00189	0.00469	0.00080	0.00069
0.00048	0.00191	0.00032	0.00028

The shown quantities are as described in Table 2

by the complexity of the functional and AR model, as well as by the underlying probability distribution of the white noise components.

5 Example based on real data

In this example, the applicability of the proposed approach is examined for modeling the colored noise behavior of a single-axis accelerometer within a vibration analysis experiment. This accelerometer is a highly accurate analog sensor which converts acceleration data to the unit m/s^2 by means of certain calibration parameters. These parameters are available from a calibration certificate issued by the American Association for Laboratory Accreditation (A2LA). Calibration data relevant to this experiment are: sensitivity (1029 mV/g), output bias (11.0 VDC), transverse sensitivity (0.5%), and resonant frequency (11.3 KHz) (Christman 2015).

The vibration monitoring experiment was performed at the Institute of Concrete Construction (Leibniz Universität Hannover) and involved a PCB Piezotronics accelerometer (used as the reference sensor) as well as other low-cost sensors.² These sensors were mounted on a plexiglass plate, which was fixed between two wooden supports (see Fig. 5). The oscillation frequencies were adjusted throughout the experiment by two imbalance motors located in the center of the plexiglass plate. Such a *shaker table* allows for measuring vibrations in a controlled manner. The acceleration data was acquired over approximately 45 min with an oscillation frequency of 16 Hz. The sampling rate was 0.00512 s, which corresponds to a sampling frequency of approximately 195 Hz. The first 5 s of the data were discarded as transient oscillation.

² One ADXL (ADXL345 chip from Analog Devices company with Arduino UNO Board), one NAMS (BNO055 chip with Arduino UNO board and Nine-Axis-Motion-Shield), and one IMU (Bosch BNO055-Chip Tinkerforge-Board).

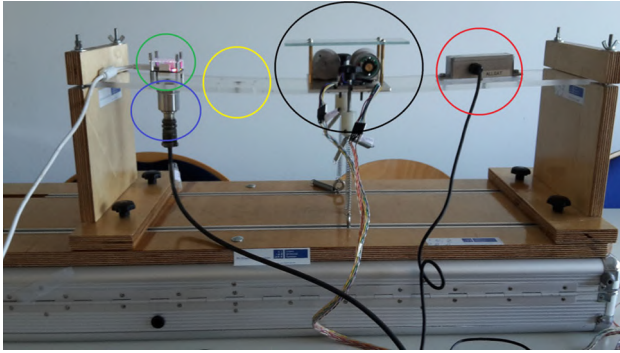


Fig. 5 Experiment setup of the shaker table and the mounted sensors on the plexiglass plate: PCB Piezotronics (blue), IMU (green), ADXL (red), NAMS (yellow), and two imbalance motors (black) (see Kemkes 2016)

Whereas the experiment aimed at comparing the amplitudes and frequencies detected by all of the sensors, we analyze in the present paper only the data stemming from the PCB Piezotronics reference sensor, focusing especially on the determination of its colored noise characteristics. A Fourier series (59) is used as a linear regression model to capture the oscillations, and an AR process (3) is employed for the modeling of the colored noise. Since it is unknown whether the dataset contains outliers and whether the observables follow a normal distribution, the scaled t -distribution (7) is applied as the stochastic model for the white noise.

To apply the proposed ECME algorithm for the estimation of these model components, the Fourier frequencies and the model order of the AR process must be specified beforehand. The 16 Hz frequency induced by the shaker is clearly visible in the discrete Fourier transform (DFT) of the acceleration data (see Fig. 6). Aside from this main frequency, multiples of 8 Hz are associated with small amplitudes. Sampling and windowing of the observed, originally continuous-time phenomenon create artificial oscillations in the DFT. In addition, the magnetic field inside the imbalance motors, the weights of the imbalance motors, and the condition of the plexiglass plate can impose further unexpected vibrations. To take the visible amplitudes into account, sine and cosine basis functions are assigned to the frequencies of 8, 16, 32, 40, 48, 56, 64, 72, 80, 88 and 96 Hz.

Unfortunately, there is no comparable direct approach to finding the most appropriate model order of the AR(p) process. Since the colored noise characteristics of the employed accelerometer are unknown, a wide range of potential model orders should be considered. To evaluate the performance of the proposed ECME algorithm, a large number of adjustments was computed. To enable this, the dataset was divided into 536 segments of $n = 1000$ consecutive measurements (each spanning approximately 5 s). Each segment was then adjusted individually for all model orders between $p = 1$

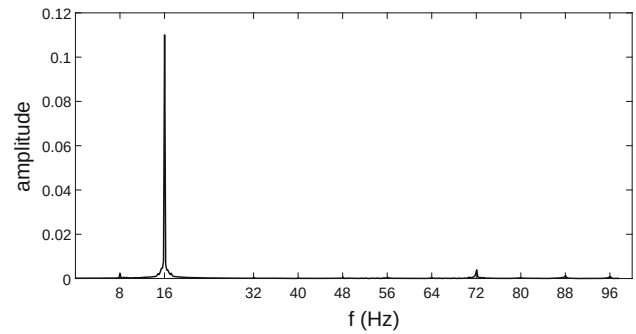


Fig. 6 Typical discrete Fourier transform of one segment of the acceleration dataset, showing the main amplitude at 16 Hz and the minor amplitudes at multiples of 8 Hz

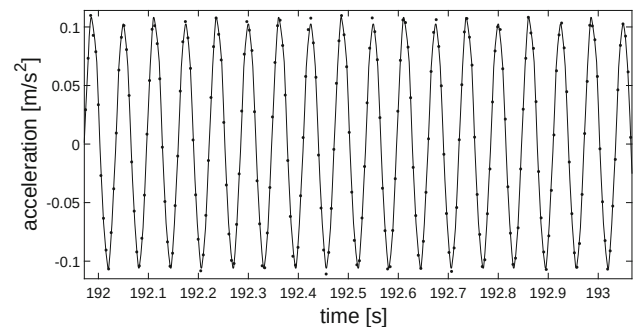


Fig. 7 Typical section of the Fourier model (solid line) fitted to the given measurements (dots)

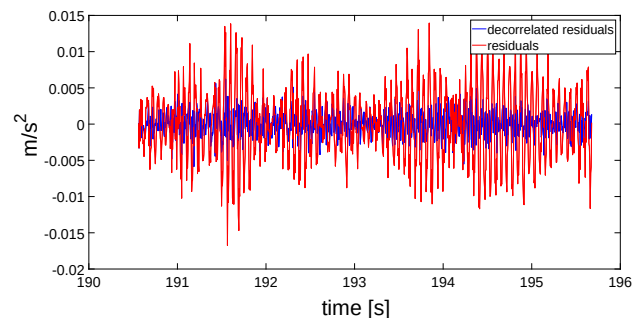


Fig. 8 Typical segment showing the estimated colored noise residuals and the decorrelated residuals

and $p = 60$. A few model orders greater than $p = 100$ were also considered.

The given measurements are generally approximated closely by the estimated trigonometric basis function models (see Fig. 7 for a typical time section). Figure 8 shows for one typical data segment the colored noise residuals (generated by an estimated AR process) and the corresponding decorrelated residuals resulting from the filtering of the former residuals by means of the inverted AR model.

The adequacy of the estimated AR colored noise models was assessed by means of the periodogram-based white noise test described in Sect. 4. For this purpose, for each model order p , it was determined in how many cases the decorre-

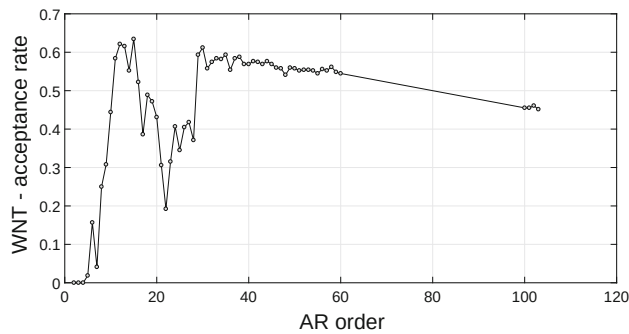


Fig. 9 Acceptance rates of the periodogram-based *white* noise test determined from all 536 data segments for each considered AR model order (no values were computed between the orders 61 and 99)

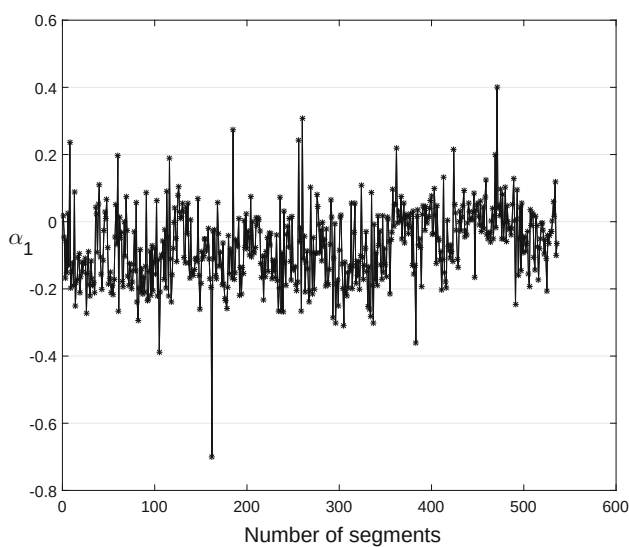


Fig. 10 Plot of the first coefficients of the AR(15) processes estimated for all 536 data segments

lated noise vector $[\hat{u}_1, \dots, \hat{u}_{1000}]$ (resulting from each of the 536 estimated AR(p) models) actually constitute white noise. This yields the WNT acceptance rate as a function of the AR order, shown in Fig. 9. The highest acceptance rate (63%) is obtained for $p = 15$, corresponding to 340 acceptances for the 536 data segments. In light of Fig. 10, the estimated first AR coefficient appears to oscillate throughout all segments, a phenomenon which is also noticeable for some other coefficients of both the Fourier and the AR model. This effect could be an indication of the interaction of both models, which was already suspected in the previous Monte Carlo simulation.

To assess how the variations of the individual coefficients of the estimated AR(15) models affect the overall colored noise characteristics, the PSDs can be computed directly from these models (cf. p. 180 Schlittgen and Streitberg 2001). Figure 11 indicates that these PSDs are relatively stable throughout the segments. We therefore conclude that the

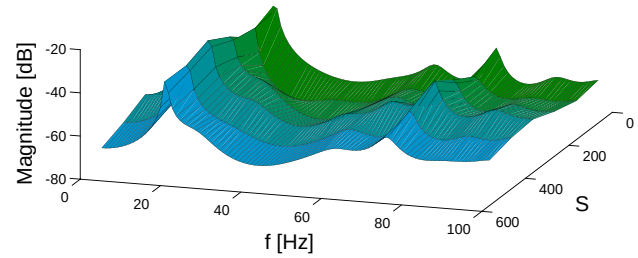


Fig. 11 Plot of the PSD based on the estimated AR(15) models for all 536 data segments

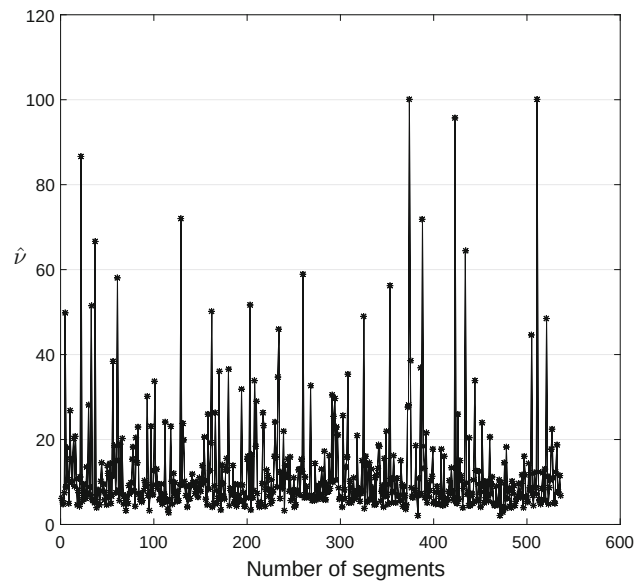


Fig. 12 Plot of the degree of freedom of the modeled $t_\nu(0, \sigma^2)$ -distribution, estimated for all 536 data segments

overall colored noise pattern is captured by the ensemble of estimated AR models to an acceptable degree.

Figure 12 reveals that the estimated degree of freedom of the underlying t -distribution mainly takes values less than 20. Despite the many spikes, the average degree of freedom throughout all segments is approximately 12. We may therefore conclude that the random deviations of the accelerometer measurements does not follow a normal distribution under the current experimental and modeling setup.

Figure 13 shows a typical histogram of the weights for the single data segment 300. It can be seen that the weights increase rather smoothly from values around 0.25, and that a few very low weights occur. The latter indicates that the corresponding white noise components are located in the tails of the t -distribution, which could therefore be labeled as “outliers”. However, outliers are in general rather difficult to detect by means of the EM algorithm based on the t -distribution due to the aforementioned effect of smoothly changing weights (Koch and Kargoll 2013). To obtain a clear separation between the smallest and the largest weights, the weights could be used as prior information for a more rig-

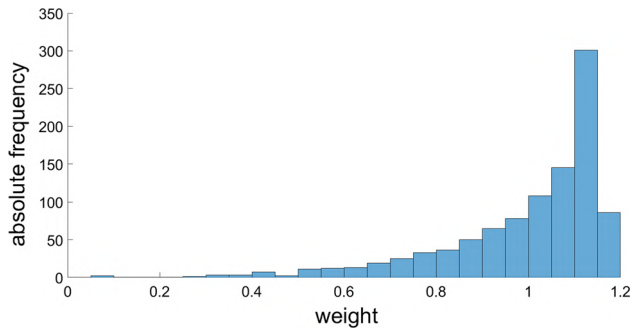


Fig. 13 Histogram plot of the weights for segment number 300

orous outlier test based on the EM algorithm developed in Koch (2012). Koch (2017) recently determined the minimal detectable outliers for such a test procedure.

6 Conclusions and outlook

To achieve an adaptive robust adjustment of possibly outlier-afflicted or non-Gaussian regression time series with autocorrelated errors, we outlined the theory and implementation of a modification of the expectation maximization algorithm, which takes the form of IRLS. Monte Carlo simulations showed that the bias of the parameter estimates can be reduced considerably by adjusting a sufficiently large number of observations. Sample sizes of at least 1000 observations are recommended for parsimonious regression and AR models, whereas more complex models (involving for instance an AR(100) process) should generally be based on at least 10,000 measurements. The accuracy of the parameter estimates was shown to diminish significantly in case the random deviations are generated by a contaminated normal distribution in comparison with a sampling normal or t -distribution.

The proposed algorithm was also tested in a real-data experiment using high-accuracy accelerometer measurements. A recorded vibration time series with autocorrelated colored noise was divided into a large number of segments. A model selection with respect to the order of the AR process was performed by employing a periodogram white noise test. The majority of segments could be decorrelated by means of estimated AR(15) models. As some of the estimated Fourier and AR coefficients themselves exhibit visible oscillations throughout the segments, a perfect separation of the two models can apparently not be achieved. In order to shed more light on these phenomena, we intend to extend the proposed EM algorithm in the future to allow for time-variable functional and AR models.

Since it is also possible that the variability of the estimated coefficients is caused by different local maxima attained by the ECME algorithm, it would also be important to investigate whether the incorporation of a global optimization

procedure (such as methods elaborated in Xu 2002, 2003) into the M-Step could be accomplished. This task could be also useful in situations where the frequencies are unknown, as the Fourier series is then a highly nonlinear function.

Finally, the analysis of the estimated degree of freedom of the underlying scaled t -distribution throughout the segments revealed an average value of 12, showing that the accelerometer measurements appear to deviate from a normal distribution.

Acknowledgements We thank the editors and the reviewers for their constructive comments and valuable suggestions, which helped to improve this paper. The presented application of the PCB Piezotronics accelerometer within the vibration analysis experiment was performed as a part of the collaborative project “Spatio-temporal monitoring of bridge structures using low cost sensors” with ALLSAT GmbH, which is funded by the German Federal Ministry for Economic Affairs and Energy (BMWi) and the Central Innovation Programme for SMEs (ZIM Kooperationsprojekt, ZF4081803DB6). In addition, the authors would like to acknowledge the Institute of Concrete Construction (Leibniz Universität Hannover) for providing the shaker table and the reference accelerometer used within this experiment.

Appendix

A Used distributions and useful properties

This section reviews the main distributions as well as corresponding densities and expected values used throughout the paper. This compilation is partly based on Koch (1999), but also offers other results that are less well known. Throughout this section, $\mathcal{X}$ denotes any real-valued random variable (r.v.).

A. 1 The univariate, scaled (Student’s) t -distribution

If $\mathcal{X}$ follows a univariate, scaled Student’s t -distribution with $\nu > 0$ degrees of freedom, location parameter μ and scale parameter $\sigma > 0$, symbolically $\mathcal{X} \sim t_\nu(\mu, \sigma)$, then the density of $\mathcal{X}$ is given by

$$f_{\mathcal{X}}(x) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\nu\pi}\sigma^2\Gamma\left(\frac{\nu}{2}\right)} \left[1 + \left(\frac{x-\mu}{\sigma}\right)^2/\nu\right]^{-\frac{\nu+1}{2}}. \quad (67)$$

Here $\Gamma(\cdot)$ denotes the gamma function (cf. Abramowitz and Stegun 1972, p. 255). If $\nu > 1$, then the expectation of $\mathcal{X}$ is defined and equals μ ; if $\nu > 2$, then the variance of $\mathcal{X}$ is finite and equals $\frac{\nu}{\nu-2} \cdot \sigma^2$ (cf. Lange et al. 1989). Setting $\sigma = 1$ leads to the (unscaled) univariate Student’s t -distribution introduced by Student (Gosset WS) (1908).

A.2 The univariate gamma distribution

If $\mathcal{X}$ follows a univariate gamma distribution with parameters $a > 0$ and $b > 0$, symbolically $\mathcal{X} \sim G(a, b)$, then the density of $\mathcal{X}$ is given by

$$f_{\mathcal{X}}(x) = \begin{cases} \frac{b^a}{\Gamma(a)} x^{a-1} e^{-bx} & \text{for } x > 0 \\ 0 & \text{for } x \leq 0 \end{cases} \quad (68)$$

The chi-squared distribution (say, with ν degrees of freedom) is a special case of the gamma distribution in the sense that $\chi_{\nu}^2 = G\left(\frac{1}{2}, \frac{\nu}{2}\right)$.

The moment generating function of $G(a, b)$ is given by

$$M_{G(a,b)}(t) = \left(1 - \frac{t}{b}\right)^{-a}, \quad (69)$$

which allows for the convenient determination of the expectation of $\mathcal{X}$ via the relationship $E_{\mathcal{X}} = M'_{G(a,b)}(0)$, where $M'(\cdot)$ denotes the derivative of the moment generating function. It is easily verified that differentiation of (69) with respect to t and evaluation of that derivative for $t = 0$ yields the value $\frac{a}{b}$ for the expectation of $\mathcal{X}$.

Another property we exploit in this paper is the following scaling property. If $s \neq 0$, then

$$s \cdot G(a, b) = G\left(\frac{b}{s}, a\right). \quad (70)$$

Using this property and the above-mentioned relationship between the gamma and the chi-squared distribution, we see that

$$G\left(\frac{\nu}{2}, \frac{\nu}{2}\right) = \frac{1}{\nu} \cdot G\left(\frac{1}{2}, \frac{\nu}{2}\right) = \frac{\chi_{\nu}^2}{\nu}.$$

Next, we show how the expected value of the natural logarithm of a gamma-distributed random variable can be obtained, that is,

$$E_{\mathcal{X}}\{\log \mathcal{X}\} = \int_0^{\infty} \log x f_{\mathcal{X}}(x) dx. \quad (71)$$

Using (68) and the substitution $z = bx$ (with $x > 0$, thus $z > 0$), we obtain

$$\begin{aligned} E_{\mathcal{X}}\{\log \mathcal{X}\} &= \int_0^{\infty} \log\left(\frac{z}{b}\right) \frac{b^a}{\Gamma(a)} \left(\frac{z}{b}\right)^{a-1} e^{-z} \frac{1}{b} dz \\ &= \int_0^{\infty} (\log z - \log b) \frac{1}{\Gamma(a)} z^{a-1} e^{-z} dz \end{aligned}$$

$$\begin{aligned} &= \frac{1}{\Gamma(a)} \int_0^{\infty} \log(z) z^{a-1} e^{-z} dz \\ &\quad - \frac{\log b}{\Gamma(a)} \int_0^{\infty} z^{a-1} e^{-z} dz. \end{aligned} \quad (72)$$

The first integral takes the value $\Gamma'(a)$ (cf. p. 331, for the derivatives of the gamma function [Bronstein and Semendjajew 1991](#)), and the second integral takes the value $\Gamma(a)$ by definition of the gamma function. Thus, we obtain

$$E_{\mathcal{X}}\{\log \mathcal{X}\} = \frac{\Gamma'(a)}{\Gamma(a)} - \log b = \psi(a) - \log b, \quad (73)$$

where $\psi(\cdot)$ denotes the digamma (or psi) function (cf. [Abramowitz and Stegun 1972](#), p. 258f.). To evaluate the digamma function we used the standard MATLAB function call `psi(0, x)`; an algorithm for this purpose is given in [Bernardo \(1976\)](#).

B Factorization of the likelihood function

We obtain for the argument of the exponential function in (31) the $N = (p+1)(p+2)/2$ terms

$$\begin{aligned} &-\frac{1}{2\sigma^2} \ell^T \mathbf{W} \ell + \frac{\alpha_1}{\sigma^2} \ell^T \mathbf{W} (L^1 \ell) + \dots + \frac{\alpha_p}{\sigma^2} \ell^T \mathbf{W} (L^1 \ell) \\ &\quad - \frac{\alpha_1^2}{2\sigma^2} (L^1 \ell)^T \mathbf{W} (L^1 \ell) - \dots - \frac{\alpha_1 \alpha_p}{\sigma^2} (L^1 \ell)^T \mathbf{W} (L^p \ell) \\ &\quad \vdots \\ &\quad - \frac{\alpha_p^2}{2\sigma^2} (L^p \ell)^T \mathbf{W} (L^p \ell), \end{aligned}$$

the further N terms

$$\begin{aligned} &-\frac{1}{2\sigma^2} \xi^T \mathbf{A}^T \mathbf{W} \mathbf{A} \xi + \frac{\alpha_1}{\sigma^2} \xi^T \mathbf{A}^T \mathbf{W} (L^1 \mathbf{A}) \xi + \dots \\ &\quad + \frac{\alpha_p}{\sigma^2} \xi^T \mathbf{A}^T \mathbf{W} (L^1 \mathbf{A}) \xi \\ &\quad - \frac{\alpha_1^2}{2\sigma^2} \xi^T (L^1 \mathbf{A})^T \mathbf{W} (L^1 \mathbf{A}) \xi - \dots \\ &\quad - \frac{\alpha_1 \alpha_p}{\sigma^2} \xi^T (L^1 \mathbf{A})^T \mathbf{W} (L^p \mathbf{A}) \xi \\ &\quad \vdots \\ &\quad - \frac{\alpha_p^2}{2\sigma^2} \xi^T (L^p \mathbf{A})^T \mathbf{W} (L^p \mathbf{A}) \xi, \end{aligned}$$

as well as the N terms

$$\begin{aligned} & -\frac{1}{2\sigma^2} \xi^T \mathbf{A}^T \mathbf{W} \ell + \frac{\alpha_1}{\sigma^2} \xi^T \mathbf{A}^T \mathbf{W} (L^1 \ell) + \dots \\ & + \frac{\alpha_p}{\sigma^2} \xi^T \mathbf{A}^T \mathbf{W} (L^p \ell) \\ & - \frac{\alpha_1^2}{2\sigma^2} \xi^T (L^1 \mathbf{A})^T \mathbf{W} (L^1 \ell) - \dots \\ & - \frac{\alpha_1 \alpha_p}{\sigma^2} \xi^T (L^1 \mathbf{A})^T \mathbf{W} (L^p \ell) \\ & \vdots \\ & - \frac{\alpha_p^2}{2\sigma^2} \xi^T (L^p \mathbf{A})^T \mathbf{W} (L^p \ell), \end{aligned}$$

and finally, the single term $\frac{\nu}{2} \mathbf{1}_n^T \bar{\mathbf{w}}$. All of the $3N+1$ occurring scalar products are functions of the augmented data $\mathbf{y} = (\mathbf{u}, \mathbf{w})$ and do not involve any parameters. Therefore, they may be viewed as data-dependent functions, as defined by (33)–(36).

C Derivation of the M-step

C.1 Regarding the CM-step for ξ

Concerning the j th functional parameter ξ_j ($j = 1, \dots, m$), we seek the solution of

$$\begin{aligned} 0 &= \frac{\partial}{\partial \xi_j} Q(\theta | \theta^{(k)}) \\ &= -\frac{1}{2\sigma^2} \sum_{t=1}^n w_t^{(k)} \frac{\partial}{\partial \xi_j} [\alpha(L)(\ell_t - \mathbf{A}_t \xi)]^2 \\ &= -\frac{1}{2\sigma^2} \sum_{t=1}^n w_t^{(k)} \cdot 2 [\alpha(L)(\ell_t - \mathbf{A}_t \xi)] \\ &\quad \times \frac{\partial}{\partial \xi_j} [-\alpha(L)(\mathbf{A}_t \xi)], \end{aligned}$$

or equivalently of

$$\begin{aligned} 0 &= \sum_{t=1}^n w_t^{(k)} [\alpha(L) \mathbf{A}_{t,j}] [\alpha(L)(\ell_t - \mathbf{A}_t \xi)] \\ &= [\alpha(L) \mathbf{A}_{1,j} \dots \alpha(L) \mathbf{A}_{n,j}] \mathbf{W}^{(k)} \begin{bmatrix} \alpha(L)(\ell_1 - \mathbf{A}_1 \xi) \\ \vdots \\ \alpha(L)(\ell_n - \mathbf{A}_n \xi) \end{bmatrix}. \end{aligned}$$

Writing all m of these equations jointly, we find (43).

C.2 CM-step for α

Recalling (42) and (2), with respect to the h -th autoregressive coefficient α_h ($h \in \{1, \dots, p\}$), we obtain for the first-order condition

$$\begin{aligned} 0 &= \frac{\partial}{\partial \alpha_h} Q(\theta | \theta^{(k)}) \\ &= -\frac{1}{2\sigma^2} \sum_{t=1}^n w_t^{(k)} \cdot \frac{\partial}{\partial \alpha_h} [\alpha(L)(\ell_t - \mathbf{A}_t \xi)]^2 \\ &= -\frac{1}{2\sigma^2} \sum_{t=1}^n w_t^{(k)} \cdot 2 [\alpha(L)(\ell_t - \mathbf{A}_t \xi)] \\ &\quad \times \frac{\partial}{\partial \alpha_h} [\alpha(L)(\ell_t - \mathbf{A}_t \xi)] \\ &= -\frac{1}{\sigma^2} \sum_{t=1}^n w_t^{(k)} [e_t - \alpha_1 e_{t-1} - \dots - \alpha_p e_{t-p}] \\ &\quad \times \frac{\partial}{\partial \alpha_h} [e_t - \alpha_1 e_{t-1} - \dots - \alpha_p e_{t-p}] \\ &= -\frac{1}{\sigma^2} \sum_{t=1}^n e_{t-h} w_t^{(k)} (e_t - \alpha_1 e_{t-1} - \dots - \alpha_p e_{t-p}) \\ &= -\frac{1}{\sigma^2} [e_{1-h} \dots e_{n-h}] \mathbf{W}^{(k)} \\ &\quad \times \begin{bmatrix} e_1 - \alpha_1 e_0 - \dots - \alpha_p e_{1-p} \\ \vdots \\ e_n - \alpha_1 e_{n-1} - \dots - \alpha_p e_{n-p} \end{bmatrix}. \end{aligned}$$

Writing all p equations in matrix notation, we thus have

$$\begin{aligned} \mathbf{0} &= \begin{bmatrix} e_0 & \dots & e_{n-1} \\ \vdots & & \vdots \\ e_{1-p} & \dots & e_{n-p} \end{bmatrix} \mathbf{W}^{(k)} \\ &\quad \times \begin{bmatrix} e_1 - \alpha_1 e_0 - \dots - \alpha_p e_{1-p} \\ \vdots \\ e_n - \alpha_1 e_{n-1} - \dots - \alpha_p e_{n-p} \end{bmatrix}. \end{aligned}$$

Denoting the solution of this equation system by $\xi = \xi^{(k+1)}$ and $\alpha = \alpha^{(k+1)}$, and using furthermore the notations (48)–(49), we find the normal equations

$$\begin{aligned} \mathbf{0} &= (\mathbf{E}^{(k+1)})^T \mathbf{W}^{(k)} \\ &\quad \times \begin{bmatrix} e_1^{(k+1)} - \alpha_1^{(k+1)} e_0^{(k+1)} - \dots - \alpha_p^{(k+1)} e_{1-p}^{(k+1)} \\ \vdots \\ e_n^{(k+1)} - \alpha_1^{(k+1)} e_{n-1}^{(k+1)} - \dots - \alpha_p^{(k+1)} e_{n-p}^{(k+1)} \end{bmatrix} \\ &= (\mathbf{E}^{(k+1)})^T \mathbf{W}^{(k)} (\mathbf{e}^{(k+1)} - \mathbf{E}^{(k+1)} \alpha^{(k+1)}), \end{aligned}$$

leading to (50).

C.3 CM-step for σ^2

Solving the first-order condition

$$\begin{aligned} 0 &= \frac{\partial}{\partial \sigma^2} Q(\boldsymbol{\theta} | \boldsymbol{\theta}^{(k)}) \\ &= -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{t=1}^n w_t^{(k)} [\alpha(\mathbf{L})(\ell_t - \mathbf{A}_t \boldsymbol{\xi})]^2 \end{aligned}$$

for σ^2 by using the already available estimates $\boldsymbol{\xi}^{(k+1)}$, $\boldsymbol{\alpha}^{(k+1)}$, $\mathbf{e}^{(k+1)}$ and (51), we obtain (52).

C.4 CM-step for ν

The first-order condition with respect to ν reads

$$\begin{aligned} 0 &= \frac{\partial}{\partial \nu} Q(\boldsymbol{\theta} | \boldsymbol{\theta}^{(k)}) \\ &= \frac{n}{2} \log \nu + \frac{n}{2} - \frac{n}{2} \cdot \frac{\Gamma'(\frac{\nu}{2})}{\Gamma(\frac{\nu}{2})} + \frac{n}{2} \left[\psi\left(\frac{\nu^{(k)} + 1}{2}\right) \right. \\ &\quad \left. - \log(\nu^{(k)} + 1) + \frac{1}{n} \sum_{t=1}^n (\log w_t^{(k)} - w_t^{(k)}) \right]. \end{aligned}$$

Multiplying with $2/n$ and using the definition of the digamma function, a zero $\nu^{(k+1)}$ of this equation satisfies (53).

Next, we derive the CME-Step (55). Rewriting the log-likelihood function (14) in the form

$$\begin{aligned} \log L(\boldsymbol{\theta}; \boldsymbol{\ell}) &= -\frac{n}{2} \log \pi - \frac{n}{2} \log(\sigma^2) \\ &\quad + n \log \Gamma\left(\frac{\nu+1}{2}\right) - n \log \Gamma\left(\frac{\nu}{2}\right) + \frac{n}{2} \nu \log \nu \\ &\quad - \frac{1}{2}(\nu+1) \sum_{t=1}^n \log \left[\nu + \left(\frac{\alpha(\mathbf{L})(\ell_t - \mathbf{A}_t \boldsymbol{\xi})}{\sigma} \right)^2 \right], \end{aligned}$$

we obtain for the first-order condition

$$\begin{aligned} 0 &= \frac{\partial}{\partial \nu} \log L(\boldsymbol{\xi}, \sigma^2, \boldsymbol{\alpha}, \nu; \boldsymbol{\ell}) \\ &= \frac{n}{2} \psi\left(\frac{\nu+1}{2}\right) - \frac{n}{2} \psi\left(\frac{\nu}{2}\right) + \frac{n}{2} (\log \nu + 1) \\ &\quad - \frac{1}{2} \sum_{t=1}^n \log \left[\nu + \left(\frac{\alpha(\mathbf{L})(\ell_t - \mathbf{A}_t \boldsymbol{\xi})}{\sigma} \right)^2 \right] \\ &\quad - \frac{1}{2}(\nu+1) \sum_{t=1}^n \left[\nu + \left(\frac{\alpha(\mathbf{L})(\ell_t - \mathbf{A}_t \boldsymbol{\xi})}{\sigma} \right)^2 \right]^{-1}. \end{aligned}$$

Substituting here $\boldsymbol{\xi}^{(k+1)}$, $\boldsymbol{\alpha}^{(k+1)}$, $(\sigma^2)^{(k+1)}$, $\mathbf{e}^{(k+1)}$ from (48), and finally $\mathbf{u}^{(k+1)}$ from (51), it follows from the preceding

equation that

$$\begin{aligned} 0 &= \frac{n}{2} \log \nu + \frac{n}{2} - \frac{n}{2} \psi\left(\frac{\nu}{2}\right) + \frac{n}{2} \psi\left(\frac{\nu+1}{2}\right) - \frac{n}{2} \log(\nu+1) \\ &\quad + \frac{n}{2} \cdot \frac{1}{n} \sum_{t=1}^n \left(\log \left[\frac{\nu+1}{\nu + \left(\frac{\alpha^{(k+1)}(\mathbf{L})(\ell_t - \mathbf{A}_t \boldsymbol{\xi}^{(k+1)})}{\sigma^{(k+1)}} \right)^2} \right] \right. \\ &\quad \left. - \frac{\nu+1}{\nu + \left(\frac{\alpha^{(k+1)}(\mathbf{L})(\ell_t - \mathbf{A}_t \boldsymbol{\xi}^{(k+1)})}{\sigma^{(k+1)}} \right)^2} \right). \end{aligned}$$

Using notation (54) and multiplying the preceding equation with $2/n$ then yields finally (55).

D Data for the Monte Carlo simulation

See Tables 7 and 8.

Table 7 Values of the true AR(10) and AR(100) coefficients (the true coefficient of the AR(1) model is $\alpha_1 = 0.6828$)

$\alpha_1 - \alpha_{10}$	$\alpha_1 - \alpha_{25}$	$\alpha_{26} - \alpha_{50}$	$\alpha_{51} - \alpha_{75}$	$\alpha_{76} - \alpha_{100}$
-0.0030	-0.1702	-0.0414	-0.3229	0.2017
0.2851	-0.0167	0.2126	-0.1451	0.1376
0.3023	0.1495	0.0650	-0.1991	0.0280
-0.3529	-0.3306	0.3787	-0.0763	0.0688
-0.5831	-0.5497	0.1526	-0.0528	0.1017
-0.3516	-0.5205	0.3607	-0.0091	0.0827
-0.0573	-0.4985	0.2288	-0.0021	0.0205
0.2241	-0.2508	0.2887	-0.0225	0.0310
-0.0651	-0.3785	0.2459	-0.1005	0.0793
-0.2377	-0.3944	0.1473	-0.0338	0.0582
-	-0.2039	0.0459	-0.0157	-0.0245
-	-0.3326	0.2886	0.1211	0.0283
-	-0.4594	0.0660	0.1164	0.0250
-	-0.2461	0.1879	0.0749	0.0417
-	-0.3301	0.0739	0.1334	-0.0651
-	-0.2583	0.1925	0.1102	-0.0259
-	-0.3912	0.0744	0.1341	-0.0092
-	-0.1610	0.0557	0.1391	0.0178
-	-0.3047	-0.0739	0.1532	0.0148
-	-0.0866	-0.0291	0.1424	0.0100
-	-0.2139	-0.0982	0.2017	-0.0510
-	0.1241	-0.1093	0.2219	-0.0161
-	-0.2184	0.0152	0.0558	0.0083
-	0.3288	-0.0579	0.0871	-0.0383
-	0.1770	-0.0170	0.2721	-0.0233

Table 8 Values of the true Fourier coefficients (columns 2 and 3) and of the true spline coefficients (column 4)

j	a_j	b_j	a_j
0	0.0016	—	—
1	0.0003	0.0021	1.62
2	0.0572	−0.0950	7.94
3	−0.0002	0.0003	3.11
4	−0.0005	−0.0002	5.29
5	0.0003	−0.0002	1.66
6	0.0006	−0.0002	6.02
7	−0.0004	0.0006	2.63
8	0.0005	0.0001	6.54
9	−0.0038	0.0003	6.89
10	0.0006	0.0002	7.48
11	0.0016	0.0008	—
12	0.0009	−0.0010	—

References

- Abramowitz M, Stegun IA (1972) Handbook of mathematical functions. Dover Publications, New York
- Alkhatib H, Schuh WD (2007) Integration of the Monte Carlo covariance estimation strategy into tailored solution procedures for large-scale least squares problems. *J Geod* 81(1):53–66. doi:10.1007/s00190-006-0034-z
- Alkhatib H, Kargoll B, Neumann I, Kreinovich V (2017) Normalization-invariant fuzzy logic operations explain empirical success of student distributions in describing measurement uncertainty. Technical Report UTEP-CS-17-55, Department of Computer Science, The University of Texas at El Paso. To appear in: proceedings of the 2017 annual conference of the North American fuzzy information processing society NAFIPS'2017, Cancun, Mexico, 16–18 Oct 2017
- Amiri-Simkooei AR, Tiberius CCJM, Teunissen PJG (2007) Assessment of noise in GPS coordinate time series: methodology and results. *J Geophys Res*. doi:10.1029/2006JB004913
- Arato M (1961) On the sufficient statistics for stationary Gaussian random processes. *Theor Prob Appl* 6(2):199–201
- Baarda W (1968) A testing procedure for use in geodetic networks. Publications on Geodesy (New Series), vol 2, 5th edn. Netherlands Geodetic Commission, Delft
- Bera AK, Jarque CM (1982) Model specification tests: a simultaneous approach. *J Econom* 20(1):59–82. doi:10.1016/0304-4076(82)90103-8
- Bernardo JM (1976) Algorithm AS 103 psi (digamma function) computation. *J R Stat Soc C* 25:315–317
- Bilmes JA (1998) A gentle tutorial of the EM algorithm and its application to parameter estimation for Gaussian mixture and hidden Markov models. In: International computer science institute, Berkeley. <http://www.icsi.berkeley.edu/ftp/global/pub/techreports/1997/tr-97-021.pdf>. Accessed 17 Jul 2017
- Box GEP, Andersen SL (1955) Permutation theory in the derivation of robust criteria and the study of departures from assumption. *J R Stat Soc B* 17(1):1–34
- Brockmann JM (2015) On high performance computing in geodesy: applications in global gravity field determination. Deutsche Geodätische Kommission, Series C (Dissertations), no 761, Munich
- Brockmann JM, Kargoll B (2012) Uncertainty assessment of some data-adaptive M-estimators. In: Sneeuw N, Novák P, Crespi M, Sansò F (eds) Proceedings of the VII Hotine–Marussi symposium on mathematical geodesy. International association of geodesy symposia, vol 137. Springer, Berlin, pp 87–92. doi:10.1007/978-3-642-22078-4_13
- Brockmann JM, Kargoll B, Krasbutter I, Schuh W-D, Wermuth M (2010) GOCE data analysis: from calibrated measurements to the global earth gravity field. In: Flechtner F, Manda M, Gruber T, Rothacher M, Wickert J, Güntner A (eds) System earth via geodetic-geophysical space techniques. Advanced technologies in earth sciences. Springer, Berlin, pp 213–229. doi:10.1007/978-3-642-10228-8_17
- Bronstein IN, Semendjajew KA (1991) Taschenbuch der Mathematik, 25th edn. Teubner, Stuttgart
- Bureick J, Alkhatib H, Neumann I (2016) Robust spatial approximation of laser scanner point clouds by means of free-form curve approaches in deformation analysis. *J Appl Geod* 10(1):27–35. doi:10.1515/jag-2015-0020
- Cai J, Sneeuw N (2014) Stochastic modeling of GOCE gravitational tensor invariants. In: Observation of the system earth from space—CHAMP, GRACE, GOCE and future missions, GEOTECHNOLOGIEN science report No. 20. Springer, Berlin, pp 115–121. doi:10.1007/978-3-642-32135-1_15
- Chang XW, Guo Y (2005) Huber's M-estimation in relative GPS positioning: computational aspects. *J Geod* 79:351–362. doi:10.1007/s00190-005-0473-y
- Christman M (2015) Calibration of industrial accelerometers. IMI division of PCB Piezotronics, Depew, New York. http://www.pcb.com/contentstore/mktgcontent/IMI_Downloads/calibration-of-industrial-accelerometers.pdf. Accessed 17 Jul 2017
- Christmas J, Everson R (2011) Robust autoregression: student-t innovations using variational Bayes. *IEEE Trans Signal Process* 59:48–57. doi:10.1109/TSP.2010.2080271
- Dempster AP, Laird NM, Rubin DB (1977) Maximum likelihood from incomplete data via the EM algorithm. *J R Stat Soc B* 39:1–38
- Dempster AP, Laird NM, Rubin DB (1980) Iteratively reweighted least squares for linear regression when errors are normal/independent distributed. In: Krishnaiah PR (ed) Multivariate analysis-V. North-Holland, Amsterdam, pp 35–57
- Dodge Y, Jureckova J (1987) Adaptive combination of least squares and least absolute derivations estimators. In: Dodge Y (ed) Statistical data analysis based on the L1-norm and related methods. North Holland, Amsterdam, pp 275–284
- Dodge Y, Jureckova J (1988) Adaptive combination of M-estimator and L1-estimator. In: Dodge Y, Fedorov VV, Wynn HP (eds) Optimal design and analysis of experiments. North Holland, Amsterdam, pp 167–176
- Duchnowski R (2009) Geodetic application of R-estimation: levelling network examples. *Tech Sci* 12:135–144
- Duchnowski R (2013) Hodges–Lehmann estimates in deformation analyses. *J Geod* 87(10–12):873–884. doi:10.1007/s00190-013-0651-2
- Durbin J (1960) The fitting of time series models. *Rev Inst Int Stat* 28(3):233–244
- Förstner W, Wrobel BP (2016) Photogrammetric computer vision. Springer, New York
- Fraser DAS (1976) Necessary analysis and adaptive inference. *J Am Stat Assoc* 71(353):99–110
- Geweke J (1993) Bayesian treatment of the independent Student-t linear model. *J Appl Econom* 8(S1):19–40. doi:10.1002/jae.3950080504
- Götzelmann M, Keller W, Reubelt T (2006) Gross error compensation for gravity field analysis based on kinematic orbit data. *J Geod* 80:184–198. doi:10.1007/s00190-006-0061-9

- Guo J, Ou J, Wang H (2010) Robust estimation for correlated observations: two local sensitivity-based downweighting strategies. *J Geod* 84:243–250. doi:[10.1007/s00190-009-0361-y](https://doi.org/10.1007/s00190-009-0361-y)
- Gupta MR, Chen Y (2011) Theory and use of the EM algorithm. *Found Trends Signal Process* 4(3):223–296. doi:[10.1561/20000000034](https://doi.org/10.1561/20000000034)
- Hamilton JD (1994) *Time series analysis*. Princeton University Press, New Jersey
- Hampel F (2001) Robust statistics: a brief introduction and overview. In: Carosio A, Kutterer H (eds) *First international symposium on robust statistics and fuzzy techniques in geodesy and GIS*. Institute of Geodesy and Photogrammetry, ETH Zurich, Zurich
- Hekimoglu S, Berber M (2003) Effectiveness of robust methods in heterogeneous linear models. *J Geod* 76:706–713. doi:[10.1007/s00190-002-0289-y](https://doi.org/10.1007/s00190-002-0289-y)
- Helmert FR (1907) *Die Ausgleichungsrechnung nach der Methode der kleinsten Quadrate*, 2nd edn. Teubner, Leipzig
- Heyde CC, Leonenko NN (2005) Student processes. *Adv Appl Probab* 37:342–365. doi:[10.1239/aap/1118858629](https://doi.org/10.1239/aap/1118858629)
- Hogg RV (1974) Adaptive robust procedures: a partial review and some suggestions for future applications and theory. *J Am Stat Assoc* 69:909–923
- Huber PJ (1964) Robust estimation of a location parameter. *Ann Math Stat* 35(1):73–101
- Junhuan P (2005) The asymptotic variance-covariance matrix, Baarda test and the reliability of L_1 -norm estimates. *J Geod* 78:668–682. doi:[10.1007/s00190-004-0433-y](https://doi.org/10.1007/s00190-004-0433-y)
- Kargoll B (2005) Comparison of some robust parameter estimation techniques for outlier analysis applied to simulated GOCE mission data. In: Jekeli C, Bastos L, Fernandes J (eds) *Gravity, geoid and space missions*. International association of geodesy symposia, vol 129. Springer, Berlin, pp 77–82. doi:[10.1007/3-540-26932-0_14](https://doi.org/10.1007/3-540-26932-0_14)
- Kargoll B (2012) *On the theory and application of model misspecification tests in geodesy*. Deutsche Geodätische Kommission, Series C (Dissertations), no 674, Munich
- Kemkes E (2016) *Suitability testing of micro-electro-mechanical-systems (MEMS) for integration into a geosensor network for the automated monitoring of the swing structures*. Master thesis. Leibniz Universität Hannover, Geodetic Institute
- Klees R, Ditmar P, Broersen P (2003) How to handle colored observation noise in large least-squares problems. *J Geod* 76(11):629–640. doi:[10.1007/s00190-002-0291-4](https://doi.org/10.1007/s00190-002-0291-4)
- Klees R, Ditmar P, Kusche J (2004) Numerical techniques for large least-squares problems with applications to GOCE. In: Sansó F (ed) *Proceedings of the V Hotine-Marussi symposium on mathematical geodesy, international association of geodesy symposia*, vol 127. Springer, Berlin, pp 12–21. doi:[10.1007/978-3-662-10735-5_3](https://doi.org/10.1007/978-3-662-10735-5_3)
- Knight NL, Wang J, Rizos C (2010) Generalised measures of reliability for multiple outliers. *J Geod* 84(10):625–635. doi:[10.1007/s00190-010-0392-4](https://doi.org/10.1007/s00190-010-0392-4)
- Koch KR (1999) *Parameter estimation and hypothesis testing in linear models*. Springer, Berlin
- Koch KR (2007) *Introduction to Bayesian statistics*, 2nd edn. Springer, Berlin
- Koch KR (2013a) Robust estimation by expectation maximization algorithm. *J Geod* 87:107–116. doi:[10.1007/s00190-012-0582-3](https://doi.org/10.1007/s00190-012-0582-3)
- Koch KR (2013) Comparison of two robust estimations by expectation maximization algorithms with Huber's method and outlier tests. *J Appl Geod* 7:115–123. doi:[10.1515/jag-2013-0050](https://doi.org/10.1515/jag-2013-0050)
- Koch KR (2012) Robust estimation by expectation maximization algorithm. *J Geod* 87(2):107–116. doi:[10.1007/s00190-012-0582-3](https://doi.org/10.1007/s00190-012-0582-3)
- Koch KR (2014) Outlier detection for the nonlinear Gauss Helmert model with variance components by the expectation maximization algorithm. *J Appl Geod* 8(3):185–194. doi:[10.1515/jag-2014-0004](https://doi.org/10.1515/jag-2014-0004)
- Koch KR (2017) Expectation maximization algorithm and its minimal detectable outliers. *Stud Geophys Geod* 61:1–18. doi:[10.1007/s11200-016-0617-y](https://doi.org/10.1007/s11200-016-0617-y)
- Koch KR, Kargoll B (2013) Expectation-maximization algorithm for the variance-inflation model by applying the t distribution. *J Appl Geod* 7:217–225. doi:[10.1515/jag-2013-0007](https://doi.org/10.1515/jag-2013-0007)
- Koch KR, Kargoll B (2015) Outlier detection by the EM algorithm for laser scanning in rectangular and polar systems. *J Appl Geod* 9(3):162–173. doi:[10.1515/jag-2015-0004](https://doi.org/10.1515/jag-2015-0004)
- Koch KR, Schmidt M (1994) *Deterministische und stochastische Signale*. Dümmler
- Kok JJ (1984) *On data snooping and multiple outlier testing*. NOAA Technical Report NOS NGS 30. US Department of Commerce, National Geodetic Survey, Rockville/Maryland
- Krarup T, Juhl J, Kubik K (1980) Götterdämmerung over least squares adjustment. In: *Proceedings of the 14th congress ISP Hamburg, international archives of photogrammetry, XXIII, B3*. Commission III, pp 369–378
- Krasbutter I, Brockmann JM, Kargoll B, Schuh W-D, Goiginger H, Pail R (2011) Refinement of the stochastic model of GOCE scientific data in a long time series. In: Ouwehand L (ed) *Proceedings of the 4th international GOCE user workshop*. ESA Publication SP-696
- Krasbutter I, Brockmann JM, Kargoll B, Schuh WD (2014) Adjustment of digital filters for decorrelation of GOCE SGG data. In: Flechtner F, Sneeuw N, Schuh WD (eds) *Observation of the system earth from space: CHAMP, GRACE, GOCE and future missions*. GEOTECHNOLOGIEN Science Report no 20. Springer, Berlin, pp 109–114. doi:[10.1007/978-3-642-32135-1_14](https://doi.org/10.1007/978-3-642-32135-1_14)
- Krasbutter I, Kargoll B, Schuh W-D (2015) Magic square of real spectral and time series analysis with an application to moving average processes. In: Kutterer H, Seitz F, Alkhatib H, Schmidt M (eds) *The 1st international workshop on the quality of geodetic observation and monitoring systems (QuGOMS'11)*. International association of geodesy symposia, vol 140. Springer, Berlin, pp 9–14. doi:[10.1007/978-3-319-10828-5_2](https://doi.org/10.1007/978-3-319-10828-5_2)
- Kubik K (1982) An error theory for the Danish method. In: *Proceedings of the symposium of the International Society of Photogrammetry, Commission III, Helsinki*
- Kusolitsch N (2014) *Mass- und Wahrscheinlichkeitstheorie*. Springer, Berlin
- Lange KL, Little RJA, Taylor JMG (1989) Robust statistical modeling using the t distribution. *J Am Stat Assoc* 84:881–896. doi:[10.2307/2290063](https://doi.org/10.2307/2290063)
- Lehmann R (2013) 3σ -rule for outlier detection from the viewpoint of geodetic adjustment. *J Surv Eng* 139(4):157–165. doi:[10.1061/\(ASCE\)SU.1943-5428.0000112](https://doi.org/10.1061/(ASCE)SU.1943-5428.0000112)
- Lehmann R (2015) Observation error model selection by information criteria vs. normality testing. *Stud Geophys Geod* 59(4):489–504. doi:[10.1007/s11200-015-0725-0](https://doi.org/10.1007/s11200-015-0725-0)
- Li L (2011) *Separability of deformations and measurement noises of GPS time series with modified Kalman filter for landslide monitoring in real-time*. Ph.D. thesis, Institute of Geodesy and Geoinformation, Universität Bonn
- Lindenberger J (1993) *Laser-Profilmessungen zur topographischen Geländeaufnahme*. Deutsche Geodätische Kommission, Series C (Dissertations), no 400, Munich
- Little RJA (1988) Robust estimation of the mean and covariance matrix from data with missing values. *Appl Stat* 37(1):23–38
- Little RJA, Rubin DB (2002) *Statistical analysis with missing data*, 2nd edn. Wiley, Hoboken
- Liu CH, Rubin DB (1994) The ECME algorithm: a simple extension of EM and ECM with faster monotone convergence. *Biometrika* 81:633–648. doi:[10.1093/biomet/81.4.633](https://doi.org/10.1093/biomet/81.4.633)
- Liu CH, Rubin DB (1995) ML estimation of the t distribution using EM and its extensions, ECM and ECME. *Stat Sin* 5:19–39

- Liu CH (1997) ML estimation of the multivariate t distribution and the EM algorithm. *J Multivar Anal* 63:296–312. doi:[10.1006/jmva.1997.1703](#)
- Luo X, Mayer M, Heck B (2012) Analysing time series of GNSS residuals by means of AR(I)MA processes. In: Sneeuw N, Novák P, Crespi M, Sansó F (eds) Proceedings of the VII Hotine–Marussi symposium on mathematical geodesy International association of geodesy symposia, vol 137. Springer, Berlin, pp 129–134. doi:[10.1007/978-3-642-22078-4_19](#)
- Luxen M, Brunn A (2003) Parameterschätzung aus unvollständigen Beobachtungsdaten mittels des EM-Algorithmus. *Z Geod Geoinform Landmanag* 128:71–78
- Marshall J (2002) L1-norm pre-analysis measures for geodetic networks. *J Geod* 76:334–344. doi:[10.1007/s00190-002-0254-9](#)
- Marx C (2013) On resistant Lp-norm estimation by means of iteratively reweighted least squares. *J Appl Geod* 7(1):1–10. doi:[10.1515/jag-2012-0042](#)
- McDonald JB (1989) Partially adaptive estimation of ARMA time series models. *Int J Forecast* 5:217–230. doi:[10.1016/0169-2070\(89\)90089-7](#)
- McLachlan GJ, Krishnan T (2008) The EM algorithm and extensions, 2nd edn. Wiley, Hoboken
- Meng X-L, Rubin DB (1993) Maximum likelihood estimation via the ECM algorithm: a general framework. *Biometrika* 80:267–278
- Metaxoglou K, Smith A (2007) Maximum likelihood estimation of VARMA models using a state-space EM algorithm. *J Time Ser Anal* 28(5):666–685. doi:[10.1111/j.1467-9892.2007.00529.x](#)
- Nadarajah S (2009) The product t density distribution arising from the product of two Student's t PDFs. *Stat Pap* 50:605–615. doi:[10.1007/s00362-007-0088-x](#)
- Nassar S, El-Sheimy N (2005) Accuracy improvement of stochastic modeling of inertial sensor errors. *Z Geod Geoinform Landmanag* 130:146–155
- Nassar S, Schwarz KP, El-Sheimy N, Noureldin A (2004) Modeling inertial sensor errors using autoregressive (AR) models. *Navigation* 51:259–268. doi:[10.1002/j.2161-4296.2004.tb00357.x](#)
- Niedzielski T, Kosek W (2010) An application of low-order ARMA and GARCH models for sea level fluctuations. *Artif Satell* 45(1):27–39. doi:[10.2478/v10018-010-0003-x](#)
- Niedzielski T, Kosek W (2012) Prediction analysis of UT1-UTC time series by combination of the least-squares and multivariate autoregressive method. In: Sneeuw N, Novák P, Crespi M, Sansó F (eds) Proceedings of the VII Hotine–Marussi symposium on mathematical geodesy. International association of geodesy symposia, vol 137. Springer, Berlin, pp 153–157. doi:[10.1007/978-3-642-22078-4_23](#)
- Nowel K, Kamiński W (2014) Robust estimation of deformation from observation differences for free control networks. *J Geod* 88(8):749–764. doi:[10.1007/s00190-014-0719-7](#)
- Orlov A (1991) How often are the observations normal? *Ind Lab* 57(7):770–772
- Pail R, Bruinsma S, Migliaccio F, Förste C, Goiginger H, Schuh WD, Höck E, Reguzzoni M, Brockmann JM, Abrikosov O, Veicherts M, Fecher T, Mayrhofer R, Krasbutter I, Sansó F, Tscherning CG (2011) First GOCE gravity field models derived by three different approaches. *J Geod* 85:819–843. doi:[10.1007/s00190-011-0467-x](#)
- Park M, Gao Y (2008) Error and performance analysis of MEMS-based inertial sensors with a low-cost GPS receiver. *Sensors* 8(4):2240–2261. doi:[10.3390/s8042240](#)
- Parzen E (1999) Stochastic processes. Society for industrial and applied mathematics (SIAM). Holden Day, Oakland
- Peng J (2009) Jointly robust estimation of unknown parameters and variance components based on expectation-maximization algorithm. *J Surv Eng* 135:1–9. doi:[10.1061/\(ASCE\)0733-9453\(2009\)](#)
- Pope AJ (1976) The statistics of residuals and the detection of outliers. NOAA Technical Report NOS65 NGS1, US Department of Commerce, National Geodetic Survey, Rockville, Maryland
- Porat B (1994) Digital processing of random signals. Dover Publications, Mineola
- Priestley MB (1981) Spectral analysis and time series. Academic Press, Cambridge
- Resnick SI (2007) Heavy-tail phenomena: probabilistic and statistical modeling. Springer, New York
- Sanubari J (1999) Robust spectral estimation based on ARMA model excited by a t -distribution process. In: Proceedings of the IEEE-EURASIP workshop on nonlinear signal and image processing (NSIP'99), vol 2, pp 607–611
- Sanubari J, Tokuda K, Onoda M (1992) Spectral estimation based on AR-model, excited by t -distribution process. *Proc IEEE Int Conf Acoust Speech Signal Process* 5:521–524. doi:[10.1109/ICASSP.1992.226568](#)
- Schall J, Eicker A, Kusche J (2014) The ITG-Goce02 gravity field model from GOCE orbit and gradiometer data based on the short arc approach. *J Geod* 88(4):403–409. doi:[10.1007/s00190-014-0691-2](#)
- Schlittgen R, Streitberg BHJ (2001) Zeitreihenanalyse, 9th edn. R. Oldenbourg Verlag, Munich
- Schön S, Kermarrec G, Kargoll B, Neumann I, Kosheleva O, Kreinovich V (2017) Why student distributions? Why Matern's covariance model? A symmetry-based explanation. Technical Report UTEP-CS-17-65, Department of Computer Science, The University of Texas at El Paso
- Schuh WD (2003) The processing of band-limited measurements; filtering techniques in the least squares context and in the presence of data gaps. *Space Sci Rev* 108(1):67–78. doi:[10.1023/A:1026121814042](#)
- Schuh WD, Brockmann JM, Kargoll B, Krasbutter I, Pail R (2010) Refinement of the stochastic model of GOCE scientific data and its effect on the in-situ gravity field solution. In: Lacoste-Francis H (ed) Proceedings of the ESA Living planet symposium. ESA publication SP-686, ESA/ESTEC. ISBN (Online) 978-9292212506, ISSN 1609-042X
- Schuh WD, Krasbutter I, Kargoll B (2014) Korrelierte Messung: was nun? In: Neuner H (ed) Zeitabhängige Messgrößen - Ihre Daten haben (Mehr-)Wert
- Siemes C (2013) Digital Filtering Algorithms for Decorrelation within Large Least Squares Problems. Deutsche Geodätische Kommission, Series C (Dissertations), no 707, Munich
- Student (Gosset WS), (1908) The probable error of a mean. *Biometrika* 6:1–25. doi:[10.1093/biomet/6.1.1](#)
- Tiku ML, Wong W-K, Vaughan DC, Bian G (2000) Time series models in non-normal situations: symmetric innovations. *J Time Ser Anal* 21:571–596. doi:[10.1111/1467-9892.00199](#)
- Tukey JW (1960) A survey of sampling from contaminated distributions. In: Olkin I (ed) Contributions to probability and statistics: essays in honor of Harold Hotelling. Stanford University Press, Stanford, pp 448–485
- Wang K, Xiong S, Li Y (2012) Modeling with noises for inertial sensors. In: Position location and navigation symposium (PLANS) 2012. IEEE/ION, pp 625–632. doi:[10.1109/PLANS.2012.6236937](#)
- Wiśniewski Z (2009) Estimation of parameters in a split functional model of geodetic observations (M_{split} estimation). *J Geod* 83(2):105–120. doi:[10.1007/s00190-008-0241-x](#)
- Wiśniewski Z (2014) M -estimation with probabilistic models of geodetic observations. *J Geod* 88(10):941–957. doi:[10.1007/s00190-014-0735-7](#)
- Xu PL (1988) Application of time series methods in prediction of displacements on large dams. *J Wuhan Tech Univ Surv Mapp* 13(3):23–31

- Xu PL (1993) Consequences of constant parameters and confidence intervals of robust estimation. *Boll Geod Sci Affini* 52:231–249
- Xu PL (2002) A hybrid global optimization method: the one-dimensional case. *J Comput Appl Math* 147(2):301–314. doi:[10.1016/S0377-0427\(02\)00438-7](https://doi.org/10.1016/S0377-0427(02)00438-7)
- Xu PL (2003) A hybrid global optimization method: the multi-dimensional case. *J Comput Appl Math* 155(2):423–446. doi:[10.1016/S0377-0427\(02\)00878-6](https://doi.org/10.1016/S0377-0427(02)00878-6)
- Xu PL (2005) Sign-constrained robust least squares, subjective breakdown point and the effect of weights of observations on robustness. *J Geod* 79:146–159. doi:[10.1007/s00190-005-0454-1](https://doi.org/10.1007/s00190-005-0454-1)
- Yang Y, Song L, Xu T (2002) Robust estimator for correlated observations based on bifactor equivalent weights. *J Geod* 76:353–358. doi:[10.1007/s00190-002-0256-7](https://doi.org/10.1007/s00190-002-0256-7)
- Zhong D (1997) Robust estimation and optimal selection of polynomial parameters for the interpolation of GPS geoid heights. *J Geod* 71(9):552–561. doi:[10.1007/s001900050123](https://doi.org/10.1007/s001900050123)

Robust Multivariate Time Series Analysis in Nonlinear Models with Autoregressive and t-Distributed Errors

Hamza Alkhatib, Boris Kargoll, and Jens-André Paffenholtz

Leibniz Universität Hannover, Geodätisches Institut
Nienburger Str. 1, 30167 Hannover, Germany
{alkhatib,kargoll,paffenholtz}@gih.uni-hannover.de
<https://www.gih.uni-hannover.de>

Abstract. We study a time series model which can generally be described as the additive combination of a multivariate, nonlinear deterministic model with multiple univariate, covariance-stationary autoregressive (AR) processes whose white noise components follow independent scaled t-distributions. These distributions allow for the stochastic modeling of heavy tails or multiple outliers and provide the framework for a partially adaptive, robust maximum likelihood (ML) estimation of the deterministic model parameters, of the AR coefficients, of the scale parameters, and of the degrees of freedom of the underlying t-distributions. To obtain the ML estimator, we derive a generalized expectation maximization (GEM) algorithm, which takes the form of linearized, iteratively reweighted least squares. The performance of this estimator is evaluated by means of a Monte Carlo simulation for the observations of a circle in three dimensions, involving different noise models encountered typically in the analysis of global navigation satellite system (GNSS) time series.

Keywords: multivariate time series, nonlinear regression model, AR process, scaled t-distribution, partially adaptive estimation, robust parameter estimation, GEM algorithm, GNSS time series

1 Introduction

Robust estimation is important in many fields of application where the probability density function (pdf) of the random deviations is expected to be heavy-tailed (e.g., as a consequence of multiple outliers). [5] was an early exposition demonstrating the use and usefulness of the scaled (Student) t-distribution in robust maximum likelihood (ML) estimation for regression models. As already indicated by [1], this kind of ML estimation can be expressed in a computationally convenient form as iteratively reweighted least squares, where the weights are used to rescale the variances of the random deviations according to their locations under the pdf. It is possible with this approach to estimate the degree of freedom of the underlying t-distribution, alongside the regression parameters and the scale parameter, turning it into a so-called (partially) adaptive estimator.

In a multivariate regression model, each observable is modeled as a random vector which is explained by a vector-valued (possibly non-linear) deterministic regression function and a vector of random deviations. [6] assumed a multivariate t -distribution with unknown scale factor and unknown degree of freedom for each vector of random deviations and investigated different forms of the expectation maximization (EM) algorithm for the purpose of estimating the unknown model parameters. It was shown earlier in [7] and [10] that the expectation conditional maximization (ECM) and the expectation conditional maximization either (ECME) variants can speed up the convergence of the EM algorithm considerably. To handle models that do not allow for closed form solutions by EM, the optimization principle of generalized expectation maximization (GEM) was proposed by [1]. The idea is to approach the maximum expectation within each EM step rather than trying to reach it fully. GEM algorithms employing Newton-Raphson steps have been applied frequently [9]. A GEM algorithm can in particular be used to handle non-linear regression models. In this situation, an iteratively reweighted least squares algorithm with Gauss-Newton steps was found to be a suitable form of GEM [14, 5].

Besides heavy tails, multivariateness and non-linearity, a further aspect that complicates (partially adaptive) parameter estimation consists in the frequently encountered autocorrelatedness of the random deviations. For instance, many types of sensor data such as inertial sensor data, satellite gravity gradiometry data and GNSS data give measurement results where the random deviations exhibit pronounced colored noise characteristics (see, e.g., [17, 13, 16, 8]). Typically, such datasets contain numerous outliers, so that robust estimation approach is generally desirable. To deal with situation, the aforementioned partially adaptive estimator for regression models based on the scaled t -distribution was extended in [4] to include autoregressive (AR) random deviations, where the white noise components of the AR process are independently and identically t -distributed. A limitation of that method is however that the observables describe only a univariate time series involving a linear regression model.

The purpose of this contribution is to extend the existing univariate, linear model to a multivariate and nonlinear (differentiable) regression model. Concerning the setup of the AR model, we currently limit ourselves to the case where each time series component is associated with a univariate AR process of individual order, independently of the AR processes of the other components. We thus exclude the modeling of cross-correlations, a task which would require the use of vector AR (VAR) processes and which is beyond the scope of the present contribution. The paper is organized as follows.

First, the time series model is described in detail in Sect. 2, and the derivation of a corresponding GEM algorithm is outlined in Sect. 3. Here, it is shown on the one hand how the scaled t -distributions are taken into account within the E step. On the other hand, the linearization of the nonlinear deterministic model is demonstrated in connection with the M step, which is broken up into conditional maximization steps with respect to the different groups of estimated model parameters. In Sect. 4, a time series model for GNSS observations of a

circle in 3D is proposed, and the results of a Monte Carlo simulation as well as real world application based on this observation model are discussed. These findings are used to evaluate the performance of the implemented GEM algorithm in this scenario.

2 The Observation Model

We consider q -dimensional observables $\mathbf{Y}_t = [Y_{1,t} \cdots Y_{N,t}]^T$ measured at equidistant time instances $t = 1, \dots, n$. The task is to approximate the corresponding measurement results $\mathbf{y}_1, \dots, \mathbf{y}_n$ by a (vector-valued) nonlinear function $\mathbf{h}_t(\boldsymbol{\xi}) = [h_{1,t}(\boldsymbol{\xi}) \cdots h_{N,t}(\boldsymbol{\xi})]^T$ of unknown parameters $\boldsymbol{\xi} = [\xi_1, \dots, \xi_m]^T$. We model the uncertainties of the measurement process by means of random deviations $\mathbf{E}_t = [E_{1,t} \cdots E_{N,t}]^T$ between the observables and the functional model, so that the observation equations take the form

$$\mathbf{Y}_t = \mathbf{h}_t(\boldsymbol{\xi}) + \mathbf{E}_t \quad (t = 1, \dots, n). \quad (1)$$

Here, we assume that each of the N components of the random deviations is subject to autocorrelations in the form of a covariance-stationary autoregressive (AR) model

$$E_{k,t} = \alpha_{k,1}E_{k,t-1} + \dots + \alpha_{k,p_k}E_{k,t-p_k} + U_{k,t} \quad (k = 1, \dots, N; t = 1, \dots, n), \quad (2)$$

in which the random variables $U_{k,1}, \dots, U_{k,n}$ are, for every $k = 1, \dots, N$, independently and identically t-distributed according to

$$U_{k,t} \sim t_{\nu_k}(0, \sigma_k^2) \quad (k = 1, \dots, N; t = 1, \dots, n). \quad (3)$$

Thus, we allow each white noise series $U_{k,1}, \dots, U_{k,n}$ to have an individual fluctuation and tail behavior, as determined by the component-dependent scale parameter σ_k^2 and degree of freedom ν_k . These quantities, alongside the AR coefficients, are considered as additional unknowns to be estimated jointly with the functional parameters $\boldsymbol{\xi}$. The probability density function (pdf) of the scaled t-distributed white noise components $U_{k,t}$ is thus defined by

$$f(u_{k,t}) = \frac{\Gamma\left(\frac{\nu_k+1}{2}\right)}{\sqrt{\nu_k\pi\sigma_k^2}\Gamma\left(\frac{\nu_k}{2}\right)} \left[1 + \left(\frac{u_{k,t}}{\sigma_k}\right)^2 / \nu_k\right]^{-\frac{\nu_k+1}{2}} \quad (4)$$

(where Γ represents the gamma function). The preceding assumption of stochastic independence of the white noise components $\mathbf{u}_k = [u_{k,1} \cdots u_{k,n}]^T$ for each $k = 1, \dots, N$ implies that their joint pdf is given by

$$f(\mathbf{u}_k) = \prod_{t=1}^n f(u_{k,t}) = \prod_{t=1}^n \frac{\Gamma\left(\frac{\nu_k+1}{2}\right)}{\sqrt{\nu_k\pi\sigma_k^2}\Gamma\left(\frac{\nu_k}{2}\right)} \left[1 + \left(\frac{u_{k,t}}{\sigma_k}\right)^2 / \nu_k\right]^{-\frac{\nu_k+1}{2}}. \quad (5)$$

We assume that no stochastic dependencies between the N white noise series exist, so that the joint pdf of the white noise components throughout all series

can be written in the factorized form $f(\mathbf{u}) = f(\mathbf{u}_1) \cdots f(\mathbf{u}_N)$. This implies also that the N colored noise processes (2) can be treated separately. Note that we generally allow these AR processes to have different orders $p_1, \dots, p_N$. Since we intend to apply the preceding model to rather large time series (with n being at least 100), we deal with the initialization problem of the AR processes in a practical manner, by setting all quantities occurring at time instances $t = 0, -1, \dots$ equal to 0. Moreover, we assume all AR processes to be invertible, so that we can rewrite them in the form

$$U_{k,t} = E_{k,t} - \alpha_{k,1}E_{k,t-1} - \dots - \alpha_{k,p_k}E_{k,t-p_k} = \alpha_k(L)E_{k,t}, \quad (6)$$

using the lag operator $L^j E_t := E_{t-j}$ and the lag polynomial $\alpha_k(L) := 1 - \alpha_{k,1}L - \dots - \alpha_{k,p_k}L^{p_k}$. We can interpret the latter as a digital filter, which decorrelates the colored noise series $e_{k,1}, \dots, e_{k,n}$ (into the white noise series $u_{k,1}, \dots, u_{k,n}$).

A maximum likelihood estimation of the unknown model parameters $\xi, \alpha_1, \dots, \alpha_N, \sigma_1^2, \dots, \sigma_N^2$ and $\nu_1, \dots, \nu_N$ based on the pdf $f(\mathbf{u})$ or its natural logarithm

$$\begin{aligned} \log \mathcal{L}(\theta; \mathbf{y}) = \log f(\mathbf{u}) = \log [f(\mathbf{u}_1) \cdots f(\mathbf{u}_N)] &= \sum_{k=1}^N \left(n \log \left[\frac{\Gamma(\frac{\nu_k+1}{2})}{\sqrt{\nu_k \pi \sigma_k^2} \Gamma(\frac{\nu_k}{2})} \right] \right. \\ &\quad \left. - \frac{\nu_k + 1}{2} \sum_{t=1}^n \log \left[1 + \left(\frac{\alpha_k(L)(y_{k,t} - h_{k,t}(\xi))}{\sigma_k} \right)^2 / \nu_k \right] \right) \end{aligned} \quad (7)$$

and given measurement results $\mathbf{y}$ requires numerical optimization since a closed-form expression of the estimator is unavailable. Following the ideas of [1] and [5], we transform the preceding t-distribution observation model into an easier-to-manage form by introducing latent variables

$$W_{k,t} | \xi, \sigma_k^2, \alpha_k, \nu_k \sim \frac{\chi_{\nu_k}^2}{\nu_k} = G\left(\frac{\nu_k}{2}, \frac{\nu_k}{2}\right) \quad (k = 1, \dots, N; t = 1, \dots, n), \quad (8)$$

where the gamma distribution is defined by the pdf

$$f(w_{k,t} | \theta) = \begin{cases} \frac{(\frac{\nu_k}{2})^{\nu_k/2}}{\Gamma(\frac{\nu_k}{2})} \cdot (w_{k,t})^{\nu_k/2-1} \cdot e^{-\nu_k/2 \cdot w_{k,t}} & \text{if } w_{k,t} > 0, \\ 0 & \text{if } w_{k,t} \leq 0 \end{cases} \quad (9)$$

(using θ for convenience as the vector consisting of all the unknown model parameters). These variables are assumed to be stochastically independent within each series, resulting in the factorization $f(\mathbf{w}_k | \theta) = \prod_{t=1}^n f(w_{k,t} | \theta)$. The idea is now to define further stochastic properties of the white noise $\mathbf{U}_k = [U_{k,1} \cdots U_{k,n}]^T$ and the latent variables $\mathbf{W}_k = [W_{k,1} \cdots W_{k,n}]^T$ in such a way that the Student pdf (5) is obtained as the marginal distribution from the joint pdf $f(\mathbf{u}_k, \mathbf{w}_k | \theta)$ (cf. [9]). This is achieved on the one hand by employing the conditional Gaussian

$$f(u_{k,t} | w_{k,t}, \theta) = \frac{1}{\sqrt{2\pi(\sigma_k/\sqrt{w_{k,t}})^2}} \exp \left\{ -\frac{u_{k,t}^2}{2(\sigma_k/\sqrt{w_{k,t}})^2} \right\}. \quad (10)$$

On the other hand, $\mathbf{U}_{k,t}$ is assumed to be independent of the white noise components and latent variables occurring within the series k at the other time instants $1, \dots, t-1, t+1, \dots, n$ and within the other series $1, \dots, k-1, k+1, \dots, N$ at all time instances, conditional on the values $w_{k,t}$ and $\boldsymbol{\theta}$. This conditional independence assumption allows us to apply for instance the simplification

$$\begin{aligned} f(u_{k,t} | u_{k,1}, w_{k,1}, \dots, u_{k,t-1}, w_{k,t-1}, u_{k,t+1}, w_{k,t+1}, \dots, u_{k,n}, w_{k,n}, w_{k,t}, \boldsymbol{\theta}) \\ = f(u_{k,t} | w_{k,t}, \boldsymbol{\theta}) \end{aligned} \quad (11)$$

in the derivation of the desired joint pdf (similarly to the proof in [2])

$$\begin{aligned} f(\mathbf{u}, \mathbf{w} | \boldsymbol{\theta}) &= \prod_{t=1}^n f(u_{1,t}, w_{1,t} | \boldsymbol{\theta}) \cdots \prod_{t=1}^n f(u_{N,t}, w_{N,t} | \boldsymbol{\theta}) \\ &= \prod_{t=1}^n f(w_{1,t} | \boldsymbol{\theta}) f(u_{1,t} | w_{1,t}, \boldsymbol{\theta}) \cdots \prod_{t=1}^n f(w_{N,t} | \boldsymbol{\theta}) f(u_{N,t} | w_{N,t}, \boldsymbol{\theta}). \end{aligned} \quad (12)$$

We define this be the likelihood function $\mathcal{L}(\boldsymbol{\theta}; \mathbf{y}, \mathbf{w})$ of the extended observation model. Before proceeding with the corresponding ML estimation, we note that the second factor in $f(w_{k,t}, u_{k,t} | \boldsymbol{\theta}) = f(u_{k,t} | \boldsymbol{\theta}) f(w_{k,t} | u_{k,t}, \boldsymbol{\theta})$ defines the conditional gamma distribution $G(a, b)$ with parameters $a = (\nu_k + 1)/2$ and $b = (\nu_k + u_{k,t}^2/\sigma_k^2)/2$, given the value $u_{k,t}$ (applying a proof in analogy to [3]).

3 The Generalized EM Algorithm

In view of (12), (9) and (10), the log-likelihood function takes the form

$$\begin{aligned} \log \mathcal{L}(\boldsymbol{\theta}; \mathbf{y}, \mathbf{w}) &= \text{const.} - \frac{n}{2} \sum_{k=1}^N \log(\sigma_k^2) + \frac{n}{2} \sum_{k=1}^N \nu_k \log\left(\frac{\nu_k}{2}\right) - n \sum_{k=1}^N \log \Gamma\left(\frac{\nu_k}{2}\right) \\ &\quad - \sum_{k=1}^N \sum_{t=1}^n \frac{1}{2} \left[\nu_k + \left(\frac{\boldsymbol{\alpha}_k(L)(y_{k,t} - h_k(\boldsymbol{\xi}))}{\sigma_k} \right)^2 \right] w_{k,t} + \sum_{k=1}^N \sum_{t=1}^n \frac{1}{2} (\nu_k - 1) \log w_{k,t}. \end{aligned} \quad (13)$$

To set up the generalized EM (GEM) algorithm, we define the Q -function as the conditional expectation of the preceding log-likelihood function (treated now as a random function), given measurement results $\mathbf{y}$ and trial parameter values $\boldsymbol{\theta}^{(i)}$, in the sense of

$$Q(\boldsymbol{\theta} | \boldsymbol{\theta}^{(i)}) = E_{\mathbf{W} | \mathbf{y}; \boldsymbol{\theta}^{(i)}} \{ \log \mathcal{L}(\boldsymbol{\theta}; \mathbf{y}, \mathbf{W}) \}. \quad (14)$$

3.1 The E Step

Recalling that the likelihood function was defined by (12), we condition directly on the white noise outcome $\mathbf{u}$ and on $\boldsymbol{\theta}^{(i)}$ (which values give $\mathbf{y}$ through the

equations (1) and (2)). Then, (13) yields

$$\begin{aligned}
Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(i)}) &= \text{const.} - \frac{n}{2} \sum_{k=1}^N \log(\sigma_k^2) + \frac{n}{2} \sum_{k=1}^N \nu_k \log\left(\frac{\nu_k}{2}\right) - n \sum_{k=1}^N \log \Gamma\left(\frac{\nu_k}{2}\right) \\
&\quad - \sum_{k=1}^N \sum_{t=1}^n \frac{1}{2} \left[\nu_k + \left(\frac{\boldsymbol{\alpha}_k(L)(y_{k,t} - h_{k,t}(\boldsymbol{\xi}))}{\sigma_k} \right)^2 \right] E_{\mathbf{W}|\mathbf{u};\boldsymbol{\theta}^{(i)}} \{W_{k,t}\} \\
&\quad + \sum_{k=1}^N \sum_{t=1}^n \frac{1}{2} (\nu_k - 1) E_{\mathbf{W}|\mathbf{u};\boldsymbol{\theta}^{(i)}} \{\log W_{k,t}\}. \tag{15}
\end{aligned}$$

Here, we observe in light of [2] that the two conditional expectations simplify to

$$\begin{aligned}
E_{\mathbf{W}|\mathbf{u};\boldsymbol{\theta}^{(i)}} \{W_{k,t}\} &= E_{W_{k,t}|u_{k,t};\boldsymbol{\theta}^{(i)}} \{W_{k,t}\}, \\
E_{\mathbf{W}|\mathbf{u};\boldsymbol{\theta}^{(i)}} \{\log W_{k,t}\} &= E_{W_{k,t}|u_{k,t};\boldsymbol{\theta}^{(i)}} \{\log W_{k,t}\}.
\end{aligned}$$

Since the latent variable $W_{k,t}$ given the value $u_{k,t}$ follows the gamma distribution $G(a, b)$, the previous two expectations are, respectively, a/b and $\psi(a) - \log(b)$ (where ψ is the digamma function), so that we obtain (cf. [3] for details)

$$w_{k,t}^{(i)} := E_{W_{k,t}|u_{k,t};\boldsymbol{\theta}^{(i)}} \{W_{k,t}\} = \frac{\nu_k^{(i)} + 1}{\nu_k^{(i)} + \left(\frac{\boldsymbol{\alpha}_k^{(i)}(L)(y_{k,t} - h_{k,t}(\boldsymbol{\xi}^{(i)}))}{\sigma_k^{(i)}} \right)^2}, \tag{16}$$

$$E_{W_{k,t}|u_{k,t};\boldsymbol{\theta}^{(i)}} \{\log W_{k,t}\} = \log w_{k,t}^{(i)} + \psi\left(\frac{\nu_k^{(i)} + 1}{2}\right) - \log\left(\frac{\nu_k^{(i)} + 1}{2}\right). \tag{17}$$

Consequently, we may rewrite (15) as

$$\begin{aligned}
Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(i)}) &= \text{const.} - \frac{n}{2} \sum_{k=1}^N \log(\sigma_k^2) - \sum_{k=1}^N \frac{1}{2\sigma_k^2} \sum_{t=1}^n w_{k,t}^{(i)} [\boldsymbol{\alpha}_k(L)(y_{k,t} - h_{k,t}(\boldsymbol{\xi}))]^2 \\
&\quad + \frac{n}{2} \sum_{k=1}^N \nu_k \log \nu_k - n \sum_{k=1}^N \log \Gamma\left(\frac{\nu_k}{2}\right) \\
&\quad + \frac{n}{2} \sum_{k=1}^N \nu_k \left[\psi\left(\frac{\nu_k^{(i)} + 1}{2}\right) - \log\left(\nu_k^{(i)} + 1\right) + \frac{1}{n} \sum_{t=1}^n (\log w_{k,t}^{(i)} - w_{k,t}^{(i)}) \right]. \tag{18}
\end{aligned}$$

We see in light of (16) that the computation of initial weights requires initial parameter values. In cases where these are not given, we choose unit weights $w_{k,t}^{(0)} = 1$ for all $k = 1, \dots, N$ and all $t = 1, \dots, n$ for the subsequent M step.

3.2 The M Step

We break up the M step into four conditional maximization (CM) steps (see [10]), one for each of the parameter groups, and substituting the most recent

available estimates whenever needed. Since the regression function $h_{k,t}$ were assumed to be nonlinear functions of $\boldsymbol{\xi}$, it is linearized within the first CM-Step with respect to that parameter group. Choosing for the Taylor point the estimate $\boldsymbol{\xi}^{(i)}$ of the preceding iteration step, we obtain for the partial derivative of the Q -function with respect to ξ_j

$$\begin{aligned} 0 &= \frac{\partial}{\partial \xi_j} Q(\boldsymbol{\theta} | \boldsymbol{\theta}^{(i)}) = - \sum_{k=1}^N \frac{1}{2\sigma_k^2} \sum_{t=1}^n w_{k,t}^{(i)} \frac{\partial}{\partial \xi_j} [\boldsymbol{\alpha}_k(L)(y_{k,t} - h_{k,t}(\boldsymbol{\xi}))]^2 \\ &= - \sum_{k=1}^N \frac{1}{2\sigma_k^2} \sum_{t=1}^n w_{k,t}^{(i)} \frac{\partial}{\partial \xi_j} \left[\boldsymbol{\alpha}_k(L) \left(y_{k,t} - \left[h_{k,t}(\boldsymbol{\xi}^{(i)}) + \frac{\partial h_{k,t}(\boldsymbol{\xi}^{(i)})}{\partial \boldsymbol{\xi}} (\boldsymbol{\xi} - \boldsymbol{\xi}^{(i)}) \right] \right) \right]^2 \\ &= - \sum_{k=1}^N \frac{1}{2\sigma_k^2} \sum_{t=1}^n w_{k,t}^{(i)} \frac{\partial}{\partial \xi_j} \left[\boldsymbol{\alpha}_k(L) \left(\Delta y_{k,t} - \mathbf{A}_{k,t}^{(i)} \Delta \boldsymbol{\xi} \right) \right]^2, \end{aligned}$$

where $\Delta y_{k,t} = y_{k,t} - h_{k,t}(\boldsymbol{\xi}^{(i)})$, $\Delta \boldsymbol{\xi} = \boldsymbol{\xi} - \boldsymbol{\xi}^{(i)}$, and $\mathbf{A}_{k,t}^{(i)} = \frac{\partial h_{k,t}(\boldsymbol{\xi}^{(i)})}{\partial \boldsymbol{\xi}}$. Denoting in addition $\mathbf{A}_{k,t,j}^{(i)} = \frac{\partial h_{k,t}(\boldsymbol{\xi}^{(i)})}{\partial \xi_j}$ and forming also the diagonal matrix $\mathbf{W}^{(i)}$ from the values $w_1^{(i)}, \dots, w_n^{(i)}$, we can derive the system of m equations

$$\mathbf{0} = \sum_{k=1}^N \frac{1}{\sigma_k^2} \begin{bmatrix} \boldsymbol{\alpha}_k(L) A_{k,1,1}^{(i)} & \cdots & \boldsymbol{\alpha}_k(L) A_{k,n,1}^{(i)} \\ \vdots & & \vdots \\ \boldsymbol{\alpha}_k(L) A_{k,1,m}^{(i)} & \cdots & \boldsymbol{\alpha}_k(L) A_{k,n,m}^{(i)} \end{bmatrix} \mathbf{W}_k^{(i)} \begin{bmatrix} \boldsymbol{\alpha}_k(L) (\Delta y_{k,1} - \mathbf{A}_{k,1}^{(i)} \Delta \boldsymbol{\xi}) \\ \vdots \\ \boldsymbol{\alpha}_k(L) (\Delta y_{k,n} - \mathbf{A}_{k,n}^{(i)} \Delta \boldsymbol{\xi}) \end{bmatrix}.$$

Fixing now the values of the unknown scale parameters and AR coefficients by taking the estimates from the preceding M step i , we can filter the reduced observations and the Jacobi matrices (for every $k = 1, \dots, N$ and every $t = 1, \dots, n$) according to

$$\overline{\Delta y}_{k,t}^{(i)} := \boldsymbol{\alpha}_k^{(i)}(L) \Delta y_{k,t}, \quad \overline{A}_{k,t,j}^{(i)} := \boldsymbol{\alpha}_k^{(i)}(L) A_{k,t,j}, \quad \overline{\mathbf{A}}_{k,t}^{(i)} := \boldsymbol{\alpha}_k^{(i)}(L) \mathbf{A}_{k,t} \quad (19)$$

and restate the preceding normal equation system as

$$\begin{aligned} \mathbf{0} &= \sum_{k=1}^N \frac{1}{(\sigma_k^2)^{(i)}} \begin{bmatrix} \overline{A}_{k,1,1}^{(i)} & \cdots & \overline{A}_{k,n,1}^{(i)} \\ \vdots & & \vdots \\ \overline{A}_{k,1,m}^{(i)} & \cdots & \overline{A}_{k,n,m}^{(i)} \end{bmatrix} \mathbf{W}_k^{(i)} \begin{bmatrix} \overline{\Delta y}_{k,1} - \overline{\mathbf{A}}_{k,1}^{(i)} \Delta \boldsymbol{\xi} \\ \vdots \\ \overline{\Delta y}_{k,n} - \overline{\mathbf{A}}_{k,n}^{(i)} \Delta \boldsymbol{\xi} \end{bmatrix} \\ &= \sum_{k=1}^N \frac{1}{(\sigma_k^2)^{(i)}} \overline{\mathbf{A}}_k^{(i)} \mathbf{W}_k^{(i)} \left(\overline{\Delta \mathbf{y}}_k - \overline{\mathbf{A}}_k^{(i)} \Delta \boldsymbol{\xi} \right). \end{aligned}$$

Consequently, the estimate of the update $\Delta \boldsymbol{\xi}$ is given by

$$\Delta \boldsymbol{\xi}^{(i+1)} = \left(\sum_{k=1}^N \frac{1}{(\sigma_k^2)^{(i)}} (\overline{\mathbf{A}}_k^{(i)})^T \mathbf{W}_k^{(i)} \overline{\mathbf{A}}_k^{(i)} \right)^{-1} \sum_{k=1}^N \frac{1}{(\sigma_k^2)^{(i)}} (\overline{\mathbf{A}}_k^{(i)})^T \mathbf{W}_k^{(i)} \overline{\Delta \mathbf{y}}_k^{(i)}. \quad (20)$$

This update is added entirely or partially to the trial solution (in the sense of a Gauss-Newton step with step size $\gamma \in (0, 1]$), resulting in

$$\boldsymbol{\xi}^{(i+1)} = \boldsymbol{\xi}^{(i)} + \gamma \boldsymbol{\Delta} \boldsymbol{\xi}^{(i+1)}. \quad (21)$$

In the first iteration step, we would typically use unit weight matrices $\mathbf{W}_k^{(0)} = \mathbf{I}_n$, neutral filters $\boldsymbol{\alpha}_k^{(0)}(L) = 1$ and identity scale factors $(\sigma_k^2)^{(i)} = 1$, corresponding to the initial assumption of normally distributed, uncorrelated and homoskedastic white noise components throughout all time series. For the subsequent CM step with respect to the autoregressive coefficients, the colored noise residuals

$$e_{k,t}^{(i+1)} := y_{k,t} - h_{k,t}(\boldsymbol{\xi}^{(i+1)}) \quad (k = 1, \dots, N; t = 1, \dots, n). \quad (22)$$

will play a central role. We assemble for this purpose the matrices

$$\mathbf{E}_k^{(i+1)} := \begin{bmatrix} e_{k,0}^{(i+1)} & \dots & e_{k,1-p_k}^{(i+1)} \\ \vdots & & \vdots \\ e_{k,n-1}^{(i+1)} & \dots & e_{k,n-p_k}^{(i+1)} \end{bmatrix} \quad (k = 1, \dots, N), \quad (23)$$

in which we substitute the initial values $e_{k,0}^{(i+1)} = \dots = e_{k,1-p_k}^{(i+1)} = 0$. Setting now the first partial derivative of (18) with respect to the j th AR coefficient within the K th time series equal to zero, we have

$$\begin{aligned} 0 &= \frac{\partial}{\partial \alpha_{K,j}} Q(\boldsymbol{\theta} | \boldsymbol{\theta}^{(i)}) = - \frac{\partial}{\partial \alpha_{K,j}} \sum_{k=1}^N \frac{1}{2\sigma_k^2} \sum_{t=1}^n w_{k,t}^{(i)} [\boldsymbol{\alpha}_k(L)(y_{k,t} - h_{k,t}(\boldsymbol{\xi}))]^2 \\ &= - \frac{1}{2\sigma_K^2} \sum_{t=1}^n w_{K,t}^{(i)} \frac{\partial}{\partial \alpha_{K,j}} [\boldsymbol{\alpha}_K(L)e_{K,t}]^2. \end{aligned}$$

Substituting for the unknowns $\boldsymbol{\xi}$ within the residual $e_{K,t}$ the already available estimates $\boldsymbol{\xi}^{(i+1)}$ (according to the principle of conditional maximization) and collecting all j partial derivative with respect to the K th time series in a single equation system, we obtain then for every $K = 1, \dots, N$ the iteratively reweighted least squares scheme for the estimation of the AR coefficients $\boldsymbol{\alpha}_K$

$$\boldsymbol{\alpha}_K^{(i+1)} = \left((\mathbf{E}_K^{(i+1)})^T \mathbf{W}_K^{(i)} \mathbf{E}_K^{(i+1)} \right)^{-1} (\mathbf{E}_K^{(i+1)})^T \mathbf{W}_K^{(i)} \mathbf{e}_K^{(i+1)}. \quad (24)$$

Since we aim for covariance-stationary and invertible AR processes, it is necessary to determine whether all roots of $\boldsymbol{\alpha}_K^{(i+1)}(z) = 0$ are located within the unit circle. In case this is not true, we stabilize the preceding polynomial by mirroring all roots with magnitude exceeding 1 into the unit circle (cf. [15]), using MATLAB's `polystab` routine. We see from (24) that the individual AR processes can be determined independently, and we use them to filter the colored noise residuals according to (6) through

$$u_{k,t}^{(i+1)} = \boldsymbol{\alpha}_k^{(i+1)}(L) e_{k,t}^{(i+1)} \quad (k = 1, \dots, N; t = 1, \dots, n) \quad (25)$$

in order to obtain the estimated white noise residuals. We are now in a position to estimate within the third CM step each scale factor σ_K^2 via the N independent conditions

$$\begin{aligned} 0 &= \frac{\partial}{\partial \sigma_K^2} Q(\boldsymbol{\theta} | \boldsymbol{\theta}^{(i)}) \\ &= -\frac{n}{2} \frac{\partial}{\partial \sigma_K^2} \log(\sigma_K^2) - \frac{\partial}{\partial \sigma_K^2} \frac{1}{2\sigma_K^2} \sum_{t=1}^n w_{K,t}^{(i)} [\boldsymbol{\alpha}_K(L)(y_{K,t} - h_{K,t}(\boldsymbol{\xi}))]^2, \end{aligned}$$

in which we substitute the current estimates $\boldsymbol{\xi}^{(i+1)}$ and $\boldsymbol{\alpha}_K^{(i+1)}$. Making use of (25), we therefore arrive at the solutions

$$(\sigma_K^2)^{(i+1)} = \frac{1}{n} \sum_{t=1}^n w_{K,t}^{(i)} \left(u_{K,t}^{(i+1)} \right)^2 = \frac{(\mathbf{u}_K^{(i+1)})^T \mathbf{W}_K^{(i)} \mathbf{u}_K^{(i+1)}}{n}. \quad (26)$$

It remains for us to compute the solutions for the degrees of freedom of the t-distributions underlying the N time series. Instead of using the Q -function for this purpose, we follow the recommendation of [7] and maximize the log-likelihood function (7) with respect to these parameters (which turns the current ECM algorithm into an ECME algorithm). Using the digamma function ψ , it can be shown that

$$\begin{aligned} 0 &= \frac{\partial}{\partial \nu_K} \log \mathcal{L}(\boldsymbol{\theta}; \mathbf{y}) = \frac{n}{2} \psi \left(\frac{\nu_K + 1}{2} \right) - \frac{n}{2} \psi \left(\frac{\nu_K}{2} \right) + \frac{n}{2} (\log \nu_K + 1) \\ &\quad - \frac{1}{2} \sum_{t=1}^n \log \left[\nu_K + \left(\frac{\boldsymbol{\alpha}_K(L)(y_{K,t} - h_{K,t}(\boldsymbol{\xi}))}{\sigma_K} \right)^2 \right] \\ &\quad - \frac{1}{2} (\nu_K + 1) \sum_{t=1}^n \left[\nu_K + \left(\frac{\boldsymbol{\alpha}_K(L)(y_{K,t} - h_{K,t}(\boldsymbol{\xi}))}{\sigma_K} \right)^2 \right]^{-1} \end{aligned}$$

As with the previous three CM steps, we utilize the most up-to-date parameter estimates, now for $\boldsymbol{\xi}$, $\boldsymbol{\alpha}_1, \dots, \boldsymbol{\alpha}_N$, $\sigma_1^2, \dots, \sigma_N^2$. Denoting furthermore the desired solution by $\nu_K^{(i+1)}$ for every $K = 1, \dots, N$, where we define $w_{K,t}^{(i+1)}$ in analogy to (16), we can derive the N equations

$$\begin{aligned} 0 &= \log \nu_K^{(i+1)} + 1 - \psi \left(\frac{\nu_K^{(i+1)}}{2} \right) + \psi \left(\frac{\nu_K^{(i+1)} + 1}{2} \right) - \log \left(\nu_K^{(i+1)} + 1 \right) \\ &\quad + \frac{1}{n} \sum_{t=1}^n \left(\log w_{K,t}^{(i+1)} - w_{K,t}^{(i+1)} \right). \end{aligned} \quad (27)$$

Thus, the estimates $\nu_1^{(i+1)}, \dots, \nu_N^{(i+1)}$ constitute the zeros of these equations, which are to be found numerically (using for instance MATLAB's `fzero` routine). Note for normally distributed white noise components that these degrees of freedom tend to infinity, in which case the function on the right-hand side

of (27) does not change its sign. This numerical problem with the zero search is circumvented by testing for the existence sign change over a sufficiently large interval, say, over $[10^{-8}, 10^8]$; if this does not happen, the estimated degree of freedom should be set to a large value (say, to 10000). We stopped the reiteration in case the maximum number of iteration steps (500) was reached or in case the parameter values of the preceding step i did not change significantly within step $i + 1$. We specified two thresholds with respect to largest maximum parameter changes: 10^{-4} for the degrees of freedom, and 10^{-8} for all other parameters.

4 Monte Carlo (MC) Results and Real World Application

4.1 The Framework of the Simulation

We consider in this section a multivariate, non-linear regression model in terms of a circle in $N = 3$ dimensions, having the following six parameters: two for the orientation (azimuth angle $\Phi \in [-\pi, \pi]$ and zenith angle $\theta \in [0, \pi]$) of its unit normal vector, one for the radius (r), and three for the circle center (C_x, C_y, C_z) (see pp. 24-27 in [11]). The observable 3D circle points are described by

$$\begin{pmatrix} X_t \\ Y_t \\ Z_t \end{pmatrix} = \begin{pmatrix} -r \cos(T_t) \sin(\Phi) + r \sin(T_t) \cos(\theta) \cos(\Phi) + C_x \\ r \cos(T_t) \cos(\Phi) + r \sin(T_t) \cos(\theta) \sin(\Phi) + C_y \\ -r \sin(T_t) \sin(\theta) + C_z \end{pmatrix} + \begin{pmatrix} E_{1,t} \\ E_{2,t} \\ E_{3,t} \end{pmatrix} \quad (28)$$

with $t = 1, \dots, n$. In our current simulation study, $n = 100,000$ time instances in (28) are sampled equidistantly between $T_1 = 0$ and $T_n = 2\pi$ (corresponding to the time interval $[1, 10000]$ sec), and the circle parameters ξ were assumed to take the true values: $r = 0.487$ m, $\Phi = 0$ rad, $\theta = -\pi$ rad, $C_x = -2487.211$ m, $C_y = -6053.041$ m and $C_z = -26.293$ m. according to a realistic scenario within the aforementioned application. Concerning the random deviations $\mathbf{E}_t$, we generated three different kinds of time series: (1) a pure white noise process, which may be viewed as an AR(0) process, (2) the AR(1) process

$$E_{k,t} = -0.9E_{k,t-1} + U_{k,t} \quad (k = 1, \dots, 3; t = 1, \dots, n), \quad (29)$$

and (3) the ARMA(3,2) process (used for all $k = 1, \dots, 3$)

$$E_{k,t} = -0.73E_{k,t-1} - 0.38E_{k,t-2} + 0.14E_{k,t-3} + U_{k,t} - 0.33U_{k,t-1} - 0.35U_{k,t-2}. \quad (30)$$

These models were investigated in the extensive study [8] (see pp. 230) on the stochastic modeling of GNSS data, where the white noise processes $U_{k,1}, \dots, U_{k,n}$ were assumed to be Gaussian. Besides generating the white noise components $U_{k,1}, \dots, U_{k,n}$ with the Gaussian sampling distributions

$$U_{1,t}, U_{2,t} \stackrel{\text{ind}}{\sim} N(0, 0.001^2), \quad U_{3,t} \stackrel{\text{ind}}{\sim} N(0, 0.002^2), \quad (31)$$

we sampled also from the scaled t-distributions

$$U_{1,t}, U_{2,t} \stackrel{\text{ind}}{\sim} t_{2.5}(0, 0.001^2), \quad U_{3,t} \stackrel{\text{ind}}{\sim} t_2(0, 0.002^2) \quad (32)$$

and from the contaminated normal distributions

$$U_{1,t}, U_{2,t} \stackrel{\text{ind}}{\sim} 0.6 \cdot N(0, 0.001^2) + 0.4 \cdot N(0, 0.008^2) \quad (33)$$

$$U_{3,t} \stackrel{\text{ind}}{\sim} 0.6 \cdot N(0, 0.002^2) + 0.4 \cdot N(0, 0.008^2) \quad (34)$$

to induce heavy tails or outliers. As the Z coordinates measured by GNSS are known to have much larger random fluctuations than the other coordinate components, the true variances in (31), true scale factors in (32) and true variances of the first Gaussian mixture component of (33) – (34) were chosen differently for the X/Y components (corresponding to $k = 1/k = 2$) and the Z component (associated with $k = 3$). Fluctuations due to systematic effects can also be expected to be largest for the Z components, so that the degree of freedom with respect to the variables $U_{3,t}$ in (32) is assumed to be less than for the components $U_{1,t}$ and $U_{2,t}$; thus, we assume the Student white noise in the Z coordinates (vertical coordinates) to be more heavy-tailed than the noise in the other components.

We generated 1000 random samples for the multivariate white noise series $\mathbf{U}_1, \dots, \mathbf{U}_n$ from each of the distributions, from which we subsequently computed the corresponding noise series $\mathbf{E}_1, \dots, \mathbf{E}_n$ and then via (28) the simulated observation time series $\mathbf{Y}_1, \dots, \mathbf{Y}_n$. The proposed GEM algorithm was applied to each of these observation samples in order to estimate the six circle parameters, the coefficients of AR processes (having a suitable, identical order for each coordinate component), and the scale factors as well as degrees of freedom of the three underlying t-distributions. Note that neither the ARMA(3,2) model (30) nor the contaminated normal distributions (33) – (34) constitute special cases of the stochastic model (2) and (4) underlying the applied GEM algorithm.

4.2 Results of the Simulation and Real Data Application

Concerning the functional parameter ξ , the Table 1 gives the means of the estimates of the first and third parameter ($\hat{r}$ and $\hat{\theta}$), computed from the 1000 MC runs. The approximation of the true parameter values by these means leads to bias free estimates for the AR(0) and AR(1) model. Only an insignificant bias in $\hat{\theta}$ in case of the ARMA(3,2) model can be detected.

To assess the goodness-of-fit of the AR models, a periodogram-based white noise test (WNT) is applied within each MC run to each of the three decorrelation-filtered residual series $\hat{u}_{k,1}$, $\hat{u}_{k,2}$ and $\hat{u}_{k,3}$. The test statistic determines the maximum cumulated periodogram excess over a cumulated, theoretical white noise periodogram (see [4] for detailed information concerning the computation of the test value). The white noise hypothesis is rejected if this maximum excess is larger than the critical value at a 95% significance level. More specifically, the critical value of the test is determined individually for each sample size n and each probability distribution in such a way that the acceptance rate, throughout all MC runs with generated random deviations e_1 , e_2 and e_3 following the AR(0)-white noise model, is identical with the desired significance level 0.95. This critical value is then employed for the current sample size and probability distribution to determine the acceptance rates with respect to the estimated

AR(1) and ARMA(3,2) models. To approximate the ARMA(3,2) model in the applied GEM algorithm we increased the order an AR-processes gradually until the white noise test has been accepted. This results in appropriate model order 30. The WNT results are given in Table 1. Generally, the WNT acceptance rates increase with the AR model and reach 95,0% for AR(0). Apparently, the ARMA(3,2) models (approximated by an AR(30)) are estimated already reasonably well for this large sample size.

The performance of the estimation of the scale factor σ with respect to the $t_\nu(0, \sigma^2)$ -distribution underlying the algorithm in Sect. 3 can be assessed only in the two cases that the white noise sampling distribution is (32) or (31), because the latter distributions are special cases of the family of scaled t -distributions. For the AR(0) and AR(1) models the mean value of the MC estimates $\hat{\sigma}$ coincides with the true value 0.001 for X/Y and 0.002 for Z (see Table 1). In contrast, for the ARMA(3,2) model the estimated scale factor is underestimated. In case of sampling by means of the contaminated normal (CN), the estimated scale factor can evidently not capture the effect of the two different variances in the data.

The evaluation of the algorithm's performance in estimating the degree of freedom of the underlying t -distribution is based on the mode of the MC estimates $\hat{\nu}$. As for the scale factor, the sampling distributions (31 and 32) allow for direct comparisons of the mode of the $\hat{\nu}$ with the corresponding true values $\nu = 2/2/2.5$ (with respect to $X/Y/Z$) and $\nu \rightarrow \infty$. The maximum value of an estimated $\hat{\nu}$ is 10000 for numerical reasons, which we therefore take as a sufficient approximation of $\nu \rightarrow \infty$. Table 1 shows that the degree of freedom tends to be overestimated for the ARMA(3,2) model.

Finally, the root mean square error (RMSE) measures the estimator's ability to predict the true observations. Since the predicted or adjusted observations are a consequence of the estimation of all four groups parameter groups ξ , α , σ^2 and ν , the RMSE expresses the overall performance of the proposed GEM algorithm. This error measure includes both the variance and the bias of the estimator, and should therefore approach 0 for different AR models. The RMSE is computed for each MC run, and the resulting mean value is given in the Table 1. It can be seen that the mean of RMSE is substantially reduced with each increase in the AR model orders and for all error models. Only in case of t -distributed errors for the ARMA(3,2) model, one sample from the tail of the distribution occurred, which lead to an extreme estimation result and therefore to an unusually high RMSE value. To accommodate for this sampling effect we computed also the median of the RMSE values, as a robust measure of goodness of fit. As could be expected, the model reproductions based on the t - and the normal sampling distributions are much superior to the contaminated normal.

We also applied the GEM algorithm to approximate a measured and preprocessed 3D GNSS time series (see [12]) by the circle given in (28). One application of this model serves the geo-referencing of terrestrial laser scanner data where the 3D circle describes the circular, horizontal motion of two global navigation satellite system (GNSS) antenna reference points. Dual frequency receivers with individually and absolutely calibrated GNSS antennas were used. The origin of

the coordinates lies in the nearby reference station with a baseline length of approximately 14 m. For further information on the measurement setup (see [12], p. 69). A full rotation consists of 7609 points (acquired with a data rate of 1 Hz) with respect to one antenna. We employed an AR model of order 12 for each time series component. Figure 1 shows the adjusted circle and the observed 3D points. Having obtained an estimated degree of freedom of 10,000 for each component we conclude that given GNSS series are normally distributed.

Table 1. Estimation results based on 1000 MC runs from the generated Student (t), normal (N) and contaminated normal (CN) error models according to (31) - (34). For WNT acceptance rates, $\text{Mean}(\hat{\sigma})$ and $\text{Mode}(\hat{\nu})$ results are listed one below the other for the three time series components (X/Y/Z).

	AR(0)			AR(1)			ARMA(3,2)		
Error model	t	N	CN	t	N	CN	t	N	CN
$\text{Mean}(\hat{r})$	0.4874	0.4874	0.4874	0.4874	0.4874	0.4874	0.4874	0.4874	0.4874
$\text{Mean}(\hat{\theta})$	-3.141593	-3.141592	-3.141585	-3.141592	-3.141592	-3.141589	-3.141593	-3.141592	-3.141573
WNT	0.95	0.95	0.95	0.962	0.972	0.955	0.999	1	1
	0.95	0.95	0.95	0.962	0.972	0.955	0.999	1	1
	0.95	0.95	0.95	0.962	0.972	0.955	0.999	1	1
$\text{Mean}(\hat{\sigma})$	0.0010	0.0010	0.0616	0.0010	0.0010	0.0616	0.0006	0.0005	0.0296
	0.0010	0.0010	0.0616	0.0010	0.0010	0.0616	0.0006	0.0005	0.0296
	0.0020	0.0020	0.0663	0.0020	0.0020	0.0663	0.0012	0.0010	0.0318
$\text{Mode}(\hat{\nu})$	2.50	10000	10000	2.50	10000	10000	3.00	10000	10000
	2.50	10000	10000	2.50	10000	10000	3.01	10000	10000
	2.00	10000	10000	2.00	10000	10000	2.40	10000	10000
$\text{Mean}(\text{RMSE}) \times 10^{-6}$	8	6	271	4	3	142	7961	4	191
$\text{Median}(\text{RMSE}) \times 10^{-6}$	8	6	267	4	3	140	6	4	187

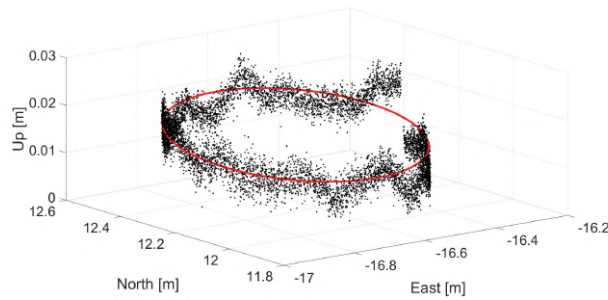


Fig. 1. 3D view of observed (black points) and adjusted circle (red line) for $n = 7827$ real three-dimensional GNSS measurements taken from [12], displayed in a North East Up (NEU) coordinate system.

5 Conclusions

To achieve an adaptive robust adjustment of a multivariate regression time series with outlier-afflicted/heavy-tailed and autocorrelated errors, we described the theory and implementation of a generalized expectation maximization algorithm. Monte Carlo simulations based on different error sampling distributions showed that the bias of the parameter estimates is insignificant when a sufficiently large number of observations (here 100,000) is adjusted. The presented algorithm was also tested in a real-data experiment using GNSS measurements.

References

1. Dempster, A.P., Laird, N.M., Rubin, D.B.: Maximum Likelihood from Incomplete Data via the EM Algorithm. *J R Stat Soc (Series B)* 39, 1-38 (1977)
2. Gupta, M.R., Chen, Y.: Theory and Use of the EM Algorithm. *Foundations and Trends in Signal Processing* 4, 223-296 (2011)
3. Koch, K.R., Kargoll, B.: Expectation Maximization Algorithm for the Variance-Inflation Model by Applying the t-Distribution. *J Appl Geod* 7, 217-225 (2013)
4. Kargoll, B., Omidalizarandi, M., Loth, I., Paffenholz, J.A., Alkhatib, H.: An Iteratively Reweighted Least-Squares Approach to Adaptive Robust Adjustment of Parameters in Linear Regression Models with Autoregressive and t-Distributed Deviations. *Journal of Geodesy*, <https://doi.org/10.1007/s00190-017-1062-6> (2017)
5. Lange, K.L., Little, R.J.A., Taylor, J.M.G.: Robust Statistical Modeling Using the t-Distribution. *J Am Stat Ass* 84, 881-896 (1989)
6. Liu, C.H.: ML Estimation of the Multivariate t Distribution and the EM Algorithm. *J Multivar Anal* 63, 296-312 (1997)
7. Liu, C.H., Rubin, D.B.: The ECME Algorithm: A Simple Extension of EM and ECM with Faster Monotone Convergence. *Biometrika* 81, 633-648 (1994)
8. Luo, X.: GPS Stochastic Modelling: Signal Quality Measures and ARMA Processes. Springer, Berlin, Heidelberg (2013)
9. McLachlan, G.J., Krishnan, T.: The EM Algorithm and Extensions. John Wiley & Sons, Hoboken, New Jersey (2008)
10. Meng, X., Rubin, D.B.: Maximum Likelihood Estimation via the ECM Algorithm: A General Framework. *Biometrika* 80, 267-278 (1993).
11. Moon, P., Spencer, D.E.: Field Theory Handbook – Including Coordinate Systems, Differential Equations, and Their Solutions. 3rd edition. Springer, New York (1988)
12. Paffenholz, J.A.: Direct Geo-Referencing of 3D Point Clouds with 3D Positioning Sensors. PhD thesis. Deutsche Geodätische Kommission (Series C, 689), Munich, Online available under <http://www.dgk.badw.de/fileadmin/docs/c-689.pdf> (2013)
13. Park, M., Gao, Y.: Error and Performance Analysis of MEMS-based Inertial Sensors with a Low-cost GPS Receiver. *Sensors* 8, 2240-2261 (2008)
14. Phillips, R.F.: Least Absolute Deviations Estimation via the EM Algorithm. *Statistics and Computing* 12:281-285 (2002)
15. Porat, B.: Digital Processing of Random Signals. Dover, Mineola/New York (1994)
16. Schuh, W.D.: The Processing of Band-Limited Measurements; Filtering Techniques in the Least Squares Context and in the Presence of Data Gaps. *Space Science Reviews* 108, 67-78 (2003)
17. Wang, K., Xiong, S., Li, Y.: Modeling with Noises for Inertial Sensors. In: Position Location and Navigation Symposium (PLANS) 2012, IEEE/ION, 625-632 (2012)



A Bayesian Nonlinear Regression Model Based on t-Distributed Errors

Alexander Dorndorf, Boris Kargoll, Jens-André Paffenholz, and Hamza Alkhatib

Abstract

In this contribution, a robust Bayesian approach to adjusting a nonlinear regression model with t-distributed errors is presented. In this approach the calculation of the posterior model parameters is feasible without linearisation of the functional model. Furthermore, the integration of prior model parameters in the form of any family of prior distributions is demonstrated. Since the posterior density is then generally non-conjugated, Monte Carlo methods are used to solve for the posterior numerically. The desired parameters are approximated by means of Markov chain Monte Carlo using Gibbs samplers and Metropolis-Hastings algorithms. The result of the presented approach is analysed by means of a closed-loop simulation and a real world application involving GNSS observations with synthetic outliers.

Keywords

Bayesian nonlinear regression model · Gibbs sampler · Markov Chain Monte Carlo · Metropolis-Hastings algorithm · Scaled t-distribution

1 Introduction

The estimation of model parameters is a fundamental task in geodetic applications. One possibility for accomplishing this task is provided by Bayesian inference, which is based on Bayes' theorem and utilizes probability density functions of observations and parameters. Bayesian inference also enables hypothesis testing and the determination of confidence regions. In comparison to classical non-Bayesian statistics, Bayesian inference is more intuitive and “methods become apparent which in traditional statistics give the impression of arbitrary computational rules” (according to Koch 2007, p. 1). The fields of application of Bayesian statistics are diverse and include disciplines such as biological, social and economic sciences (see Gelman et al. 2014, for the fundamental basics of Bayesian inference and some appli-

cation examples from these disciplines). Bayesian statistics has also been used for different geodetic applications for decades (Bossler 1972; Koch 1988, 2007, 2018; Riesmeier 1984; Schaffrin 1987; Yang 1991; Zhu et al. 2005). However, in all of these studies simple linear functions are used, simple conjugate prior density functions are assumed or outlier-affected observations are not considered. A particular problem that arises with nonlinear models or non-conjugate priors in connection with popular classes of algorithms, such as Markov chain Monte Carlo (MCMC) methods (cf. Gamerman and Lopes 2006; Gelman et al. 2014), one generally cannot sample directly from the posterior density function. The class of MCMC methods includes the well known Metropolis-Hastings algorithm and Gibbs sampler. The latter has been originally developed for the Bayesian restoration of digital images and later used for a variety of problems of Bayesian inference. Such problems include nonlinear inverse problems (see, for instance, Haario et al. 2006; Johnathan et al. 2014). In a geodetic context Gibbs sampler methods have been used for the purpose of error propagation and inversion of large matrices (cf. Koch 2017; Alkhatib and Schuh 2006; Gundlich et al. 2003).

A. Dorndorf (✉) · B. Kargoll · J.-A. Paffenholz · H. Alkhatib
Geodetic Institute, Leibniz University Hannover, Hannover, Germany
e-mail: dorndorf@gih.uni-hannover.de; kargoll@gih.uni-hannover.de;
paffenholz@gih.uni-hannover.de; alkhatib@gih.uni-hannover.de

The likelihood function and prior distribution are generally assumed to be Gaussian in this context, so that these approaches are not robust against outliers. To obtain a robust approach, the family of normal distributions is replaced by a family of heavy-tailed distributions. Koch and Kargoll (2013) introduced the scaled t-distribution as a heavy-tailed error law in a geodetic application involving a linear model. An extension of this approach to nonlinear models was developed in Alkhatib et al. (2017). In both works, the solution was obtained by means of an expectation maximization algorithm, which is only based on the likelihood function and thereby does not allow for the integration of prior knowledge about the parameters. In a Bayesian context involving linear models with t-distributed errors, a variety of solution approaches utilizing MCMC methods, in particular a Gibbs sampler, exist (e.g., Geweke 1993; Gelman et al. 2014). In our current contribution, we extend the previous approaches to a Bayesian approach to solving a nonlinear model based on the t-distribution error law. The suggested approach allows for the integration of prior model parameters by assuming any family of prior distributions.

2 Bayesian Inference

2.1 Fundamentals of Bayesian Inference

In many geodetic applications one typically works with models which depend upon parameters to be estimated. We limit ourselves in this paper to the linear or nonlinear regression models. Let I be a vector of data and Θ be a vector which contains the parameters for a model which seeks to explain I . The relationship between observations and unknown parameters is described by means of a functional model $f(\Theta)$:

$$I + \epsilon = f(\Theta). \quad (1)$$

The residuals (“errors”) ϵ arise in an overdetermined system and describe the precision of the observations. The usual assumption is that the residuals are normally distributed with zero mean and variance-covariance matrix Σ . The vector-valued function $f(\Theta)$ can be linear (e.g., distance estimation from repeated measurements) or nonlinear (e.g., estimation of geometric shape parameters from 2D or 3D points), but the focus is on the nonlinear case in this paper. The previous model defines a Gauss-Markov model (GMM), which is usually adjusted by means of the method of least squares (cf. Koch 1999).

As an extension of that approach, Bayesian inference uses probability distributions to determine the unknown parameters of a model and is based on Bayes’ theorem:

$$p(\bar{\Theta}|I) \propto p(\underline{\Theta}) \cdot p(I|\Theta). \quad (2)$$

Here, $p(\bar{\Theta}|I)$ is the so-called posterior density, from which the unknown model parameters can be derived for given observations I . The prior density $p(\underline{\Theta})$ expresses all additional information about the unknown parameters and may be obtained from, e.g., results of a previous adjustment or a manufacturer’s data sheet. $p(I|\Theta)$ is the likelihood function, which represents the information of the observations conditional on the unknown parameters. In case of Bayesian inference the solution approach is based on marginal and conditional densities, where one distinguishes between conjugated and non-conjugated prior distributions. To evaluate the mean of the posterior density as a point estimate, one can solve the integral $E(\Theta|I) = \int \Theta p(\bar{\Theta}|I) d\Theta$. However, except for special cases involving linear functional models and normally distributed residuals, it is impossible to evaluate this integral analytically. To overcome this limitation, Monte Carlo (MC) techniques can be employed to approximate desired statistical measures such as expectations, variances, covariances, skewness, and kurtosis (cf. Koch 2017; Gelman et al. 2014).

2.2 A Robust Bayesian Model

The general assumption of normally distributed residuals in Eq. (1) does not account for outliers, which may therefore deteriorate the inference about the model parameters. To deal with outliers, hypothesis tests for outlier detection or a robust adjustment should be carried out. The focus of this contribution is on the latter. For this purpose, the normal distribution may be replaced by a longer-tailed family of distributions (cf. Gelman et al. 2014), for instance, by the family of scaled (Student’s) t-distributions, frequently used in Bayesian and likelihood inference. Accordingly, the stochastic model for each residual is assumed to be

$$\epsilon_i \sim t_\nu(0, s_i^2). \quad (3)$$

The degree of freedom ν controls the thickness of the tails. Outliers, being located in the tails, are more abundant for small ν , whereas the t-distribution approaches a normal distribution with increasing ν . Inference about ν is possible, so that estimators based on the t-distribution model have been called *adaptive* or *self-tuning* robust estimators (cf. Koch and Kargoll 2013; Parzen 1979). The residuals are assumed to have expectation 0, and the scale factor s_i is related to their variances through

$$\sigma_\epsilon^2 = \frac{\nu}{\nu - 2} s_i^2, \quad (4)$$

defined for $\nu > 2$. The t-distribution model Eq. (3) can be re-formulated conveniently and equivalently as the rescaled

normal distribution (cf. Gelman et al. 2014)

$$\begin{aligned}\epsilon_i &\sim N(0, \alpha^2 W_i), \\ W_i &\sim \text{Inv-}\chi^2(\nu, \tau^2), \\ \text{with } s_i^2 &= \alpha^2 \tau^2.\end{aligned}\quad (5)$$

This equivalence is enabled by the introduction of additional, scaled inverse-chi-square distributed weights W_i . Whereas ν itself is also a degree of freedom of that $\text{Inv-}\chi^2$ -distribution, the scale factor s_i is factorized into the scale factor α with respect to the normal distribution and the parameter τ of the $\text{Inv-}\chi^2$ -distribution.

As this t-distribution results in a non-conjugated prior and as the functional model is nonlinear, MCMC methods are required to calculate the posterior density. The fundamentals and further discussion of the solution of non-conjugated prior densities by means of MCMC can be found, e.g., in Kroese et al. (2011) and Gelman et al. (2014). In this contribution, the Gibbs sampler is used for the generation of Markov chains, by sequentially drawing the unknown posterior parameters from their conditional densities. The purpose of the next section is to provide the required fundamentals of MCMC as well as a calculation procedure based on the Gibbs sampler.

3 A Bayesian Approach to Parameter Estimation

3.1 Specification of the Bayesian Model

Without loss of generality, the calculation procedure is developed for the typical task of adjusting n 3D points. More specifically, each observation

$$I_i = [x_i, y_i, z_i] \quad (i = 1, \dots, n), \quad (6)$$

is defined by an x -, a y - and a z -coordinate. To demonstrate the flexibility of the Bayesian model, outliers are assumed to occur only in the z -coordinate, so that a t-distribution is associated with the corresponding residuals, whereas the residuals of the other two coordinate components are assumed to be normally distributed at the outset. The residuals for all three coordinate components may have different levels of variance, and stochastic dependencies between them are currently neglected. Thus, the stochastic model reads

$$\begin{aligned}\epsilon_{x_i} &\sim N(0, \sigma_x^2), \\ \epsilon_{y_i} &\sim N(0, \sigma_y^2), \\ \epsilon_{z_i} &\sim t_\nu(0, s_z^2).\end{aligned}\quad (7)$$

The functional model f generally involves a parameter vector $\Theta = [\theta_1, \theta_2, \dots, \theta_u]^T$ consisting of u unknowns. As part of a Bayesian model, prior distributions are assumed for all unknown parameters, including the three scale factors and the degree of freedom ν of the t-distribution. An informative prior is assigned to the functional model parameters Θ and non-informative priors to all other parameters. As non-informative priors may be defined by constant probability, only the prior of Θ remains to be specified, which choice depends on the inferential procedure about the parameters Θ . If the redundancy of the prior adjustment problem is large, it is frequently adequate to assume a multivariate normal distribution

$$p(\Theta) = \frac{1}{\sqrt{(2\pi)^u \det \Sigma_\Theta}} \exp \left(-\frac{(\Theta - \underline{\Theta})^T \Sigma_\Theta^{-1} (\Theta - \underline{\Theta})}{2} \right). \quad (8)$$

The prior knowledge of the model parameters $\underline{\Theta}$ is the expected value of the multivariate normal distribution, and Σ_Θ controls the spread of the distribution. Θ is an arbitrary realization of the vector of functional model parameters. If the redundancy of the prior adjustment is small, a multivariate t-distribution would be an adequate choice for the prior. The likelihood function is determined by the observation equations $I_i + \epsilon_i = f_i(\Theta)$ and the stochastic model (7) for the residuals. Expressing the t-distribution for the residuals of the z -component as the rescaled normal distribution (5), the likelihood function is defined by the (factorized) multivariate normal distribution

$$\begin{aligned}p(I|\Theta, \sigma_x, \sigma_y, \alpha, W, \nu, \tau) &= \prod_{i=1}^n \frac{1}{\sqrt{(2\pi)^k \det \Sigma_{I_i}}} \\ &\times \exp \left(-\frac{(I_i - f_i(\Theta))^T \Sigma_{I_i}^{-1} (I_i - f_i(\Theta))}{2} \right), \\ \text{with } \Sigma_{I_i} &= \begin{bmatrix} \sigma_x^2 & 0 & 0 \\ 0 & \sigma_y^2 & 0 \\ 0 & 0 & \alpha^2 W_i \end{bmatrix} \text{ and } k = 3,\end{aligned}\quad (9)$$

where the variance-covariance matrix Σ_{I_i} of one observed point fulfills the assumption of uncorrelated coordinate components. The following section demonstrates the calculation of the posterior density.

3.2 Calculation of Posterior Parameters

The posterior density $p(\overline{\Theta}, \overline{\sigma_x}, \overline{\sigma_y}, \overline{\alpha}, \overline{W}, \overline{\nu}, \overline{\tau}|I)$ summarizes the information of the prior (8) and the likelihood function (9) via Bayes's theorem (2). As indicated in Sect. 2,

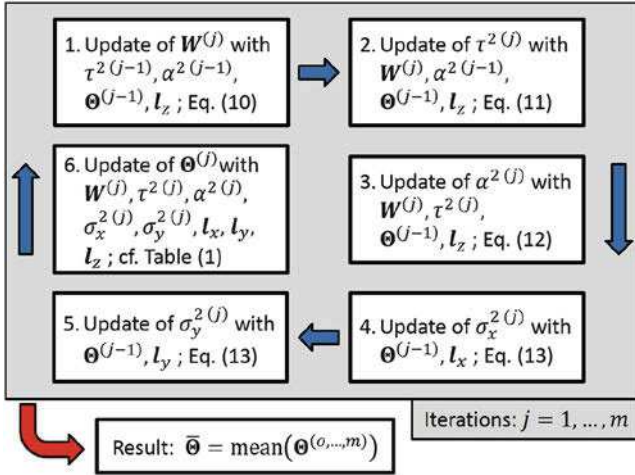


Fig. 1 Computational steps of the Gibbs sampler for adjusting 3D points based on the Bayesian model described in Sect. 3.1 with fixed degree of freedom ν

the prior is non-conjugated for the current model, so that the posterior is solved for numerically. The calculation of the degree of freedom is an intricate step, for which various approaches have been proposed in the case of a linear regression model (see Geweke 1993; Gelman et al. 2014). Due to limited space, the current model is simplified by assuming that the degree of freedom is known with $\nu = 4$, which value has been recommended for the purpose of a robust estimation (cf. Gelman et al. 2014). For the calculation of the posterior unknowns a Markov chain is generated by means of a Gibbs sampler (see Fig. 1). The convergence of Gibbs sampler depends strongly on the choice of the initial parameter values $\Theta^{(0)}$. Less critically, initial values for the parameters τ and α must also be specified, e.g., setting $\tau^{2(0)} = 1$ and $\alpha^{2(0)}$ equal to the variance of the z-coordinate (thus assuming that the z-coordinate is initially normally distributed).

For the generation of Markov chains using the Gibbs sampler, the conditional distributions of the unknowns are required, which are generally well known from Bayesian literature in the context of linear models (e.g. Gelman et al. 2014). These are now adapted to the present nonlinear model. The Gibbs sampler starts with step 1. in any iteration step j , where the weights for the z-coordinate are drawn depending on τ , α , Θ from the preceding iteration $j - 1$. According to

$$W_i | \tau^2, \alpha^2, \Theta, l_{z_i} \sim \text{Inv-}\chi^2 \left(\nu + 1, \frac{\nu \tau^2 + (l_{z_i} - f_{z_i}(\Theta))^2 / \alpha^2}{\nu + 1} \right), \quad (10)$$

the weights are Inv- χ^2 -distributed, where the parameter α scales the residual square of a measured z_i ; ν and τ control the robustness. In step 2. of the Gibbs sampler, the

distribution of τ is updated depending on the new weights W as well as parameter values Θ and α of the iteration before:

$$\tau^2 | W, \alpha^2, \Theta, l_z \sim \text{Gamma} \left(\frac{n\nu}{2}, \frac{\nu}{2} \sum_{i=1}^n \frac{1}{W_i} \right). \quad (11)$$

For the generation of the gamma-distributed random variable τ , the parameter α and the observations l_z are not used directly. These values are included indirectly in the weights W and the number of observations n . In step 3., the scale factor α is generated from a Inv- χ^2 -distribution as follows:

$$\alpha^2 | W, \tau^2, \Theta, l_z \sim \text{Inv-}\chi^2 \left(n, \frac{1}{n} \sum_{i=1}^n \frac{(l_{z_i} - f_{z_i}(\Theta))^2}{W_i} \right). \quad (12)$$

Now, the generation of the variance σ_x^2 within step 4. differs from (12) only by the replacement of the z_i by the x_i -components while omitting the weight W_i , that is,

$$\sigma_x^2 | \Theta, l_x \sim \text{Inv-}\chi^2 \left(n, \frac{1}{n} \sum_{i=1}^n (l_{x_i} - f_{x_i}(\Theta))^2 \right). \quad (13)$$

The generation of the variance σ_y^2 in step 5. is then carried out in the same way as σ_x^2 , using y_i - instead of x_i -components.

In the remaining step 6., the generation of $\Theta^{(j)}$ is not possible by means of a conditional distribution as the latter is unknown. This, however, is not a problem as random numbers can be directly generated from the combination of the prior density and the likelihood function by means of a Metropolis-Hastings (MH) algorithm (Hastings 1970), as shown in Table 1. Firstly, a random number θ_i^{new} is generated using the values $\theta_i^{(j-1)}$ from preceding iteration step and using a scale factor λ_{θ_i} . The selection of the distribution family for this random number generation influences the efficiency of the MH algorithm. In view of the application given

Table 1 Generation of posterior $\Theta^{(j)}$ by means of the MH algorithm

1. Generate	$\theta_i^{\text{new}} \sim N(\theta_i^{(j-1)}, \lambda_{\theta_i})$
2. Set	$\Theta^{\text{new}} = [\theta_1^{(j)}, \theta_2^{(j)}, \dots, \theta_i^{\text{new}}, \dots, \theta_u^{(j-1)}]^T$ $\Theta^{\text{old}} = [\theta_1^{(j)}, \theta_2^{(j)}, \dots, \theta_i^{(j-1)}, \dots, \theta_u^{(j-1)}]^T$
3. Calculate	$\Psi = \min \left[1, \frac{p(\Theta^{\text{new}}) * p(\Theta^{\text{new}} l, \sigma_x, \sigma_y, \alpha, W, \nu, \tau)}{p(\Theta^{\text{old}}) * p(\Theta^{\text{old}} l, \sigma_x, \sigma_y, \alpha, W, \nu, \tau)} \right]$
4. Accept or reject	
Generate:	$\rho \sim U(0, 1)$
If $\rho \leq \Psi$:	$\theta_i^{(j)} = \theta_i^{\text{new}}$
If $\rho > \Psi$:	$\theta_i^{(j)} = \theta_i^{(j-1)}$

in Sect. 4, a normal distribution is employed for this purpose since the posterior of Θ will approximately be gaussian. The efficiency of the MH algorithm also depends on the scale factor λ_{θ_i} . An approach to choosing adequate value for λ_{θ_i} in application is presented in Sect. 4.3. In step 2. the parameter vectors Θ are set up with θ_i^{new} and $\theta_i^{(j-1)}$. The parameters $\theta_1^{(j)}, \dots, \theta_{i-1}^{(j)}$ are from the current iteration j , which means that these values have been updated before θ_i by the MH algorithm. The model parameter values $\theta_{i+1}^{(j-1)}, \dots, \theta_u^{(j-1)}$ stem from the iteration before. These values are gradually regenerated randomly by means of the MH algorithm after updating θ_i . In step 3. the probability ratio of Θ^{new} and Θ^{old} is calculated according to Eqs. (8) and (9). The value Ψ is used for the decision in step 4., whether θ_i^{new} or $\theta_i^{(j-1)}$ is the new generated $\theta_i^{(j)}$. A convenient feature of this procedure is the direct usage of the functional model $f(\Theta)$, without the need for derivatives.

With the conclusion of the MH algorithm, the Gibbs sampler complete one iteration. This procedure is carried out in total m times and thereby yields the Markov chain results for the unknown posterior parameters. By means of resulting Markov chains, the posterior results for the key parameters (functional model parameters, variances σ_x^2 and σ_y^2 , the unknown weights and scale parameters of t-distribution) can be approximated. The first half of the drawn chains are considered as *burn-in replications* and discarded. The remaining samples may serve for the estimation of the parameters from $\Theta^{(0, \dots, m)}$, as their mean value (see Fig. 1). By choosing m sufficiently large, the approximation error implicit in the estimate can be reduced (see, e.g. Kroese et al. 2011; Gelman et al. 2014).

4 Application and Results

4.1 Application

In our real-world application we use a multi-sensor-system (MSS) composed by a laser scanner and two firmly attached GNSS equipment proposed by Paffenholtz (2012). The aim of the MSS is to efficiently geo-reference 3D point clouds by means of 3D coordinates in a superior coordinate frame. For ease of understanding, we consider only the obtained 3D coordinates by one GNSS equipment according Eq. (6) for n observed 3D points. These points describe a circle in 3D due to the rotation of the laser scanner around its vertical z -axis. For the geo-referencing of the laser scanner the unknown circle parameters must be estimated. The parameterisation of a circle in 3D is given by: $\Theta = [c_x, c_y, c_z, r, \omega, \varphi]^T$.

The parameter c is the centre point of the circle for the x -, y - and z -coordinate and r is the radius. The angles ω and φ describe the orientation of the circle in 3D by means of the rotations around the x - and y -axis. By the parameter Θ the

functional model can be set up for the different coordinate components as follows:

$$f_{x_i}(\Theta) = r \cos(t_i) \cos(\varphi) + c_x, \quad (14)$$

$$f_{y_i}(\Theta) = r \cos(t_i) \sin(\varphi) \sin(\omega) + r \sin(t_i) \cos(\omega) + c_y,$$

$$f_{z_i}(\Theta) = -r \cos(t_i) \sin(\varphi) \cos(\omega) + r \sin(t_i) \sin(\omega) + c_z,$$

$$\text{with } f_i(\Theta) = [f_{x_i}(\Theta), f_{y_i}(\Theta), f_{z_i}(\Theta)].$$

The splitting of the functional model in the three coordinate components corresponds to a nonlinear regression model, which allows for the estimation of the unknown parameters by means of a GMM. The equations arise from the combination of the polar coordinate equation of the circle and a 3D rotation matrix. The variable t is the rotation angle of the laser scanner around its z -axis. To simplify the model we assume that t is known and error free. This assumption is possible, because the horizontal angle measurement of the laser scanner is significantly preciser than the 3D points obtained by the GNSS equipment. In addition to the GNSS observations, prior knowledge is available from calibration measurements by means of a laser tracker. Therefore, the value r and the corresponding variance σ_r^2 are known. We assume a non-informative prior for the other model parameters c , ω and φ . Hence, we use a normal distribution as prior instead of the general presented multiple prior distribution in Eq. (8). For the stochastic model the assumption of Eq. (7) is used. For further information about the MSS such as the specific sensors, the calibration and the geo-referencing approach employed, see Paffenholtz (2012).

4.2 Closed Loop Simulation

The investigation of the presented Bayesian estimation approach in Sect. 3 is based on a closed loop (CL) simulation and on a real data example. In this section we describe the generation of the CL simulation and the results. Firstly, we define the true parameter values $\Theta_{\text{true}} = [1716.00 \text{ cm}, 3012.00 \text{ cm}, 1054.00 \text{ cm}, 30.00 \text{ cm}, 0.40^\circ, 0.08^\circ]$ for the 3D circle. With the functional model in Eq. (14) and Θ_{true} we calculated 50 uniformly distributed observations on the 3D circle. After that, random normal distributed values are generated and added to these observations. For the noise generation we use the following parameters derived from the real data: $\sigma_x = 0.2 \text{ cm}$, $\sigma_y = 0.1 \text{ cm}$ and $\sigma_z = 0.4 \text{ cm}$. In addition to the normally distributed errors, we create 15% of the z -observations as outliers by means of the uniform distribution $U(0.93, 1.71)$ with minimum value 0.93 and maximum value 1.71. The outliers are ill-conditioned distributed and are spread out over the true circle. In each iteration j_{CL} of the CL simulation the noise and outliers are generated randomly. The simulation is repeated 10,000

times, where three different approaches are used for the estimation of $\hat{\Theta}^{(j_{CL})}$. j_{CL} denotes the counter of the CL simulations. The approaches are based on:

- (1) a linearised non-robust GMM (cf. Koch 1999),
- (2) the presented robust Bayesian approach with non-informative prior, and
- (3) the presented robust Bayesian approach with prior information. The prior knowledge about the radius is defined here to be the true value $\underline{r} = r_{\text{true}} = 30$ cm as well as the standard deviation $\underline{\sigma}_r = 0.05$ cm. Basics and further information about CL simulation are presented, e.g., in Saltelli et al. (2008).

In each iteration of the CL simulation initial values for the unknown parameters Θ are required. For $\mathbf{c}^{(0)}$ we use the mean of the observations $\mathbf{I}$. $\mathbf{r}^{(0)}$ is derived by means of the euclidean distance between $\mathbf{c}^{(0)}$ and an arbitrarily observed point. The initial angles $\omega^{(0)}$ and $\varphi^{(0)}$ can be set to zero since the MMS is levelled. These initial values are used for all three estimation approaches. Furthermore the Gibbs sampler requires starting values for α^2 , τ^2 (cf. Sect. 3.2), λ_{Θ} and m . The last two parameter values are essential for the convergence of the Gibbs sampler, and they depend on the quality of the initial values of the unknown parameters. For example, if the initial values $\Theta^{(0)}$ deviate significantly from the true parameter values Θ_{true} then the number of generated samples in each Gibbs sampler chain m should be increased. In each iteration of the CL simulation a Markov chain with $m = 7,000$ and a warm-up period of $o = 3,500$ is generated. The convergence of the Gibbs sampler and the definition of λ_{Θ} are discussed in Sect. 4.3.

The results of the CL simulation are the estimates $\hat{\Theta}$ for the three approaches. The means of the 10,000 vectors $[\hat{c}_x, \hat{c}_y, \hat{r}, \hat{\omega}, \hat{\varphi}]$ differ from Θ_{true} by less than 10^{-4} [cm] resp. [deg]. Only the mean of $\hat{c}_z$ deviates significantly from $c_{z\text{true}}$ (see Table 2) due to the generated outliers in the z-coordinate.

Table 2 Estimated results of the CL simulation

	Bayes		
	GMM	non-informative	Bayes informative
Mean (RMSE) [cm]	0.25	0.19	0.19
σ_{RMSE} [cm]	0.05	0.06	0.06
Mean ($\hat{c}_z$) - $c_{z\text{true}}$ [cm]	0.21	0.14	0.14
$\hat{\sigma}_{c_z}$ [cm]	0.053	0.063	0.063
Mean ($\hat{r}$) [cm]	30.00	30.00	30.00
$\hat{\sigma}_r$ [cm]	0.020	0.020	0.018

The outliers have a smaller influence on the estimate $\hat{c}_z$ by the robust Bayesian approach than on this estimate by the non robust GMM. For a comparison between the three approaches in terms of all estimates $\hat{\Theta}$ simultaneously, the root mean square error (RMSE) is calculated. The RMSE represents the average euclidean distance between predicted and true 3D points.

$$\text{RMSE}^{(j_{CL})} = \sqrt{\frac{1}{n} \sum_{i=1}^n (d_{xi}^2 + d_{yi}^2 + d_{zi}^2)},$$

$$\text{with } d_{xi}^2 = [f_{xi}(\Theta_{\text{true}}) - f_{xi}(\hat{\Theta}^{(j_{CL})})]^2, \quad (15)$$

$$d_{yi}^2 = [f_{yi}(\Theta_{\text{true}}) - f_{yi}(\hat{\Theta}^{(j_{CL})})]^2,$$

$$d_{zi}^2 = [f_{zi}(\Theta_{\text{true}}) - f_{zi}(\hat{\Theta}^{(j_{CL})})]^2.$$

The corresponding distributions of the 10,000 RMSE results are shown in Fig. 2. The average RMSE for the GMM (blue) is larger than the values for the Bayesian non-informative (green) and informative (red) approaches, which are identical (see Table 2). The only difference between the Bayesian results can be observed for the estimated radius (see Table 2): In the informative case, the prior knowledge about the radius

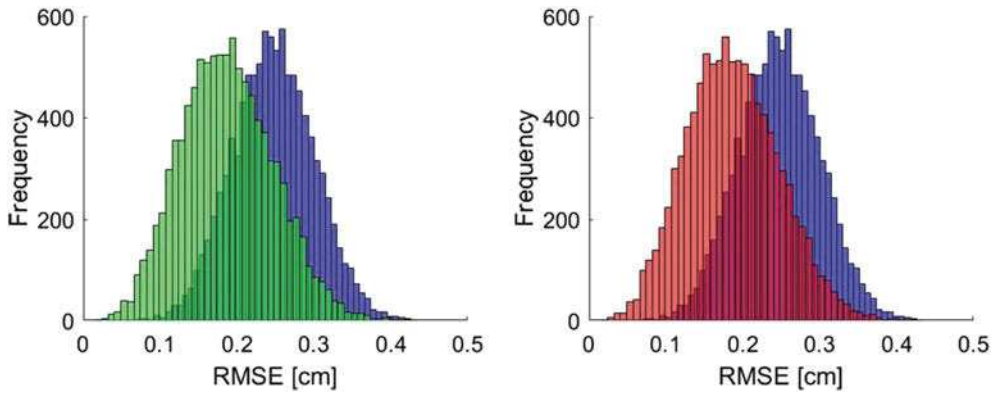


Fig. 2 Distribution of the 10,000 RMSE results for the three approaches: GMM (blue), Bayesian non-informative (green), and Bayesian informative (red)

reduces the standard deviation of the estimated radius but the difference is very small.

4.3 Convergence Analysis of the Markov Chains

Any posterior simulation approach such as the Gibbs sampler presented in Sect. 3.2 provides us with an $\hat{\Theta}$ which is an estimate of $E(f(\Theta)|I)$. By choosing m sufficiently large and the starting values for the parameters close enough to the true values, the convergence of the Markov chain is very probable. To analyse the convergence behaviour of the Gibbs sampler, the CL simulation is repeated. Contrary to the calculated values for $\Theta^{(0)}$ in Sect. 4.2 we generate these values randomly. The random generated values are now not close to the values Θ_{true} as the calculated initial values given in Sect. 4.2. We will demonstrated here that the Markov chain for the random values $\Theta^{(0)}$ will independently converge to the true values Θ_{true} . The random initial values for $\Theta^{(0)}$ are generated by means of uniform distributions with the intervals:

$$\begin{aligned} c^{(0)} &\in [c_{\text{true}} - 50, c_{\text{true}} + 50] [\text{cm}], \\ r^{(0)} &\in [r_{\text{true}} - 10, r_{\text{true}} + 10] [\text{cm}], \\ \omega^{(0)} &\in [\omega_{\text{true}} - 10, \omega_{\text{true}} + 10] [\text{deg}], \\ \varphi^{(0)} &\in [\varphi_{\text{true}} - 10, \varphi_{\text{true}} + 10] [\text{deg}]. \end{aligned}$$

In case of random initial values $\Theta^{(0)}$ the choice of an adequate value for λ_{θ_i} is a challenging task. The use of a fixed λ_{θ_i} has the following difficult problem. If the value is too large, then only a few new random numbers for θ_i will be accepted in the MH algorithm. Conversely, a very small λ_{θ_i} results in a high acceptance rate for θ_i . In both cases a convergence of the Gibbs sampler cannot be guaranteed.

Alternatively, an iterative adaptation can be used to determine λ_{θ_i} . The used approach is based on the assumption that an adequate λ_{θ_i} is chosen if the acceptance rate is about 50% (cf. Gelman et al. 2014). After every 250 iterations of the Gibbs sampler the acceptance rate of the currently created chain θ_i is calculated. If the acceptance ratio is less than 40%, $\lambda_{\theta_i}^{(j)}$ will be reduced by the factor γ . In contrast, an acceptance ratio greater than 60% will be increased the value $\lambda_{\theta_i}^{(j)}$ by the factor γ . For the presented application the values $\gamma = 2$ and $\gamma = 5$ have proven to be suitable. Both values are used in alternating order (cf. Fig. 3). In a future work we will investigate the dependency of the value of the alternating factor γ on the number of iterations m .

In the CL simulation the initial values $\lambda_{\Theta}^{(0)} = [5, 5, 5, 1, 1, 1]$ are used. The results of the adaptively estimated λ_{Θ} for one CL simulation are shown in Fig. 3. It can be seen that these estimates become constant around the warm up period. Figure 4 shows the results of the Markov chain for Θ for the same simulation. The chains spread constantly around their mean values close to Θ_{true} . These results are similar to the results of the other 9,999 iterations of the CL simulation. Consequently, it can be assumed that the Markov chain for Θ converges for the used start values m , o , $\Theta^{(0)}$ and $\lambda_{\Theta}^{(0)}$. For a final statement of convergence a hypotheses test would be required. Due to the limited space in this contribution a general examination of the determination of λ_{θ_i} and convergence analysis is not feasible and will be discussed in future studies.

4.4 Real Data Example

For the real data example we use a data set with 1,580 GNSS observations. The noise of the x - and y -coordinates are approximately equal to the variances in the CL simulation. Contrary to the simulated data, the residuals of the measured z -coordinates have additional error effects,

Fig. 3 Results of λ_{Θ} for one CL simulation: iterative adaptation of λ_{θ_i} (blue), and limit of warm-up period (red)

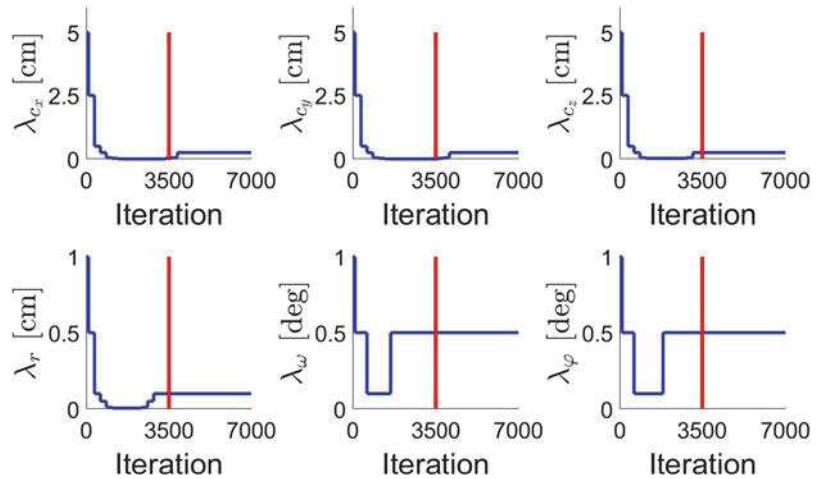


Fig. 4 Results of estimating parameters Θ for one CL simulation: Markov chains of the Gibbs sampler after warm-up period (blue), mean values of these chains (red), and Θ_{true} (green)

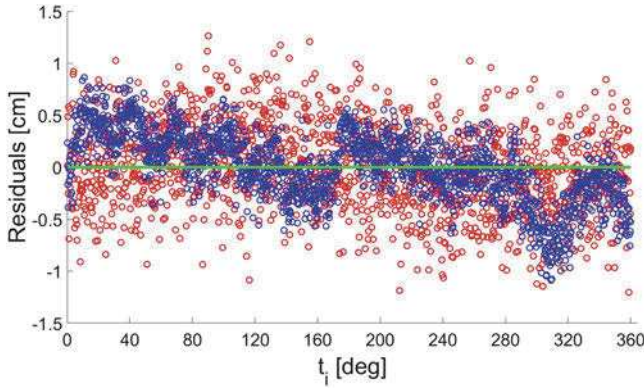
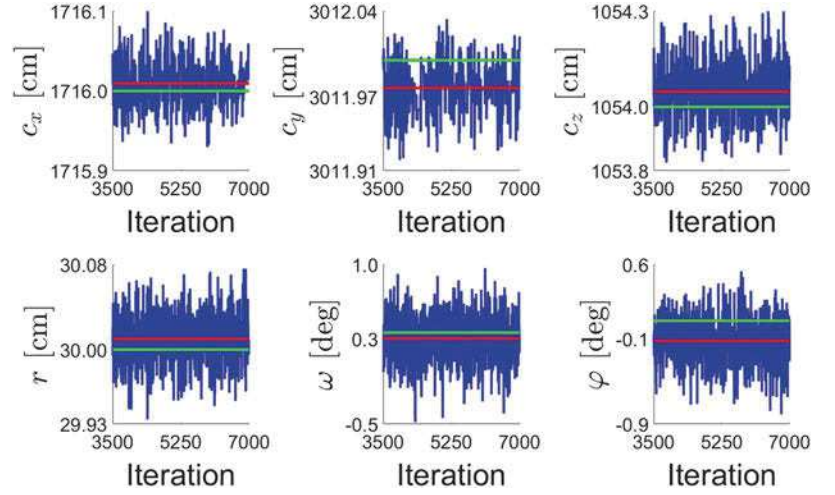


Fig. 5 Residuals of z -coordinate. $[z_i - \text{Mean}(z)]$ of the real data (blue), and noise generated from the normal distribution $N(0, 0.4 \text{ [cm]})$ as in the CL simulation (red)

which are not represented by a normal distributed noise (see Fig. 5). However, these are no significant outliers but probably time-dependent effects which are not considered further in this paper. Hence, we generate randomly outliers for 15% of the z -coordinate by means of the uniform distribution $U(0.5, 1.0)$. The informative Bayesian approach uses the same prior values for the radius as in the CL simulation ($\underline{r} = 30 \text{ cm}$, $\underline{\sigma}_r = 0.05 \text{ cm}$), which were estimated from a calibration measurement by means of a laser tracker. The initial values for Θ are calculated as described in Sect. 4.2, and the initial values for α^2 , τ^2 , λ_Θ , o and m are the same as in the CL simulation. As the true circle parameters are unknown, the RMSE cannot be calculated. Therefore, we estimate the circle parameters without any outliers in the data set by means of a GMM. These parameters are used to evaluate the results of the contaminated data set. In Table 3 the estimated results of $\hat{c}_z$ and $\hat{r}$ are presented. We compare only these two parameters, because the outliers influence primarily the estimated circle center point of the z -coordinate and the radius should be influenced by the prior knowledge.

Table 3 Estimation results for the real data example

Estimation approach	$\hat{c}_z$ [cm]	$\hat{\sigma}_{c_z}$ [cm]	$\hat{r}$ [cm]	$\hat{\sigma}_r$ [cm]
GMM without outliers	1.70	0.008	29.75	0.004
GMM	1.81	0.009	29.75	0.004
Bayes non-informative	1.73	0.010	29.73	0.004
Bayes informative	1.73	0.010	29.73	0.004

The centre points of the x - and y -coordinate and the angles do not differ significantly between the different estimations. The results of $\hat{c}_z$ show that the robust Bayesian results are closer to the result of the GMM without outliers as the result of the GMM. In case of $\hat{r}$ the results of all estimations are close together without significant differences.

5 Conclusions and Outlook

In this contribution, a robust Bayesian approach to adjusting a nonlinear functional model based on normally or t -distributed residuals was presented. In this approach one observation was introduced as a 3D point, to demonstrate that different observation groups may have different stochastic models. The selection of a prior distribution and a likelihood function was described with regard to geodetic applications. t -distributions, having longer tails than normal distributions, were used within the likelihood function for individual components to account for expected outliers. A numerical approach to calculating the unknown posterior parameters based on a Gibbs sampler was suggested. As the non-linearity of the functional model excludes the usage of a conditional distribution for the functional model parameters, a Metropolis-Hastings algorithm was outlined.

Finally, a geodetic application was presented where the parameters of a 3D circle have been of interest. The results of this application show that the introduced t -distribution

in the Bayesian model reduces the influence of outliers on the estimated parameters. The introduced prior information for the radius affects only the precision $\hat{\sigma}_r$ in the range of 0.02 mm in the closed loop simulation. This is to be expected since r_{true} and $\underline{r}$ are identical. In case of the real world data no significant differences between non-informative and informative Bayesian approach is detectable for $\hat{r}$ and $\hat{\sigma}_r$. The reason for this is the larger number of observations in the real world data. The likelihood function thus dominates the posterior density, which can be interpreted as a down weighting of the prior density in the estimation.

In future studies the robust Bayesian approach should be improved to deal with fully correlated observations. Furthermore, the approach should be extended by the estimation of the degree of freedom of the t-distribution.

References

- Alkhatib H, Schuh W-D (2006) Integration of the Monte Carlo covariance estimation strategy into tailored solution procedures for large-scale least squares problems. *J Geod* 81:53–66
- Alkhatib H, Kargoll B, Paffenholz J-A (2017) Further results on robust multivariate time series analysis in nonlinear models with autoregressive and t-distributed errors. In: Rojas I, Pomares H, Valenzuela O (eds) *Time series analysis and forecasting. ITISE 2017. Contributions to statistics*. Springer, Cham, pp 25–38
- Bossler JD (1972) *Bayesian inference in geodesy*. Dissertation, The Ohio State University, Columbus, Ohio, USA
- Gamerman D, Lopes HF (2006) *Markov chain Monte Carlo – stochastic simulation for bayesian inference*, 2nd edn. Chapman and Hall/CRC, Boca Raton
- Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A, Rubin DB (2014) *Bayesian data analysis*, 3rd edn. CRC Press, Taylor & Francis Group, Boca Raton
- Geweke J (1993) Bayesian treatment of the independent student-t linear model. *J Appl Economet* 8:19–40
- Gundlich B, Koch KR, Kusche J (2003) Gibbs sampler for computing and propagating large covariance matrices. *J Geod* 77:514–528
- Haario H, Laine M, Mira A, Saksman E (2006) DRAM: efficient adaptive MCMC. *Stat Comput* 16:339–354
- Hastings WK (1970) Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* 57:97–109
- Johnathan MB, Solonen A, Haario H, Laine M (2014) Randomize-then-optimize: a method for sampling from posterior distributions in nonlinear inverse problems. *SIAM J Sci Comput* 36:A1895–A1910
- Koch KR (1988) Bayesian statistics for variance components with informative and noninformative priors. *Manuscr Geodaet* 13:370–373
- Koch KR (1999) *Parameter estimation and hypothesis testing in linear models*, 2nd edn. Springer, Berlin
- Koch KR (2007) *Introduction to Bayesian statistics*, 2nd edn. Springer, Berlin
- Koch KR (2017) Monte Carlo methods. *Int J Geomath* 9:117–143
- Koch KR (2018) Bayesian statistics and Monte Carlo methods. *J Geod Sci* 8:18–29
- Koch KR, Kargoll B (2013) Expectation-maximization algorithm for the variance-inflation model by applying the t-distribution. *J Appl Geod* 7:217–225
- Kroese DP, Taimre T, Botev ZI (2011) *Handbook of Monte Carlo methods*. Wiley, Hoboken
- Paffenholz J-A (2012) *Direct geo-referencing of 3D point clouds with 3D positioning sensors*. Deutsche Geodätische Kommission, Series C (Dissertation), No. 689, Munich
- Parzen E (1979) A density-quantile function perspective on robust estimation. In: Launer L, Wilkinson GN (eds) *Robustness in statistics*. Academic Press, New York, pp 237–258
- Riesmeier K (1984) *Test von Ungleichungshypothesen in linearen Modellen mit Bayes-Verfahren*. Deutsche Geodätische Kommission, Series C (Dissertation), No. 292, Munich
- Saltelli A, Ratto M, Andres T, Campolongo F, Cariboni J, Gatelli D, Saisana M, Tarantola S (2008) *Global sensitivity analysis: the primer*. Wiley, Chichester
- Schaffrin B (1987) Approximating the Bayesian estimate of the standard deviation in a linear model. *Bull Geod* 61(3):276–280
- Yang Y (1991) Robust Bayesian estimation. *Bull Geod* 65(3):145–150
- Zhu J, Santerre R, Chang X-W (2005) A Bayesian method for linear, inequality-constrained adjustment and its application to GPS positioning. *J Geod* 78:528–534

Die Bewertung kaufpreisarmer Lagen mit multivariaten statistischen Verfahren – Möglichkeiten und Grenzen robuster Methoden bei der Auswertung weniger Kauffälle

Valuation in Areas with less Transactions Using Multivariate Statistical Methods – Possibilities and Limitations of Robust Methods in the Analysis of less Purchases

Alexandra Weitkamp, Hamza Alkhatib

Das Vergleichswertverfahren hat sich als marktnächstes Verfahren als sehr praktikabel in der Immobilienbewertung etabliert. Allerdings bedarf es, wie alle statistischen Methoden, einer geeigneten Stichprobengröße: normalerweise werden 15 Kauffälle pro unabhängige Variable in einer Regressionsanalyse benötigt. In Gebieten mit wenigen Kauffällen stehen den Sachverständigen oft nur sehr wenige Kauffälle zur Verfügung (z.B. 10 bis 30 Kauffälle/Teilmarkt). Er oder sie schätzt den Wert durch seine oder ihre Erfahrung unter Berücksichtigung dieser wenigen Information ab. In diesem Fall wird die klassische statistische Auswertung nur unzuverlässige Ergebnisse liefern oder nicht möglich sein.

Das Ziel dieser Untersuchung ist es, einen Ansatz zu präsentieren, der eine zuverlässige Auswertung auch in Lagen mit wenigen Kauffällen ermöglicht. Hierzu wird ein robuster Bayesischer Ansatz eingeführt. Damit ist es möglich, Expertenwissen in datengestützte Modelle – wie die multiple lineare Regressionsanalyse – zu integrieren, die auf einer kleinen Stichprobe (kaufpreisarme Lage) gründen. Der Fokus liegt dabei auf Ein- und Zweifamilienhäusern. Um kaufpreisarme Lagen zu simulieren, werden die Daten systematisch reduziert, wie z.B. in Teilstichproben in den Rändern oder in der Mitte der Daten. Basierend auf dem Bayesischen Ansatz werden Daten und Expertenwissen in einem umfassenden Modell verarbeitet. Dieses Modell verwendet die unabhängige Student-t-Verteilung im linearen Modell, sodass es auch in Fällen funktioniert, in denen signifikante Abweichungen von den optimalen Annahmen auftreten. Die Lösung des funktionalen Zusammenhangs erfolgt mithilfe der Markov-Chain-Monte-Carlo-Methode.

Schlüsselwörter: Wertermittlung, Vergleichswertverfahren, kaufpreisarme Lage, Bayesische Statistik, multiple lineare Regression, robuste Schätzer, Markov-Chain-Monte-Carlo, Gibbs-Sampler.

The Comparison Approach is established as market next method in property valuation. Like all statistical methods, it needs an appropriate sample size: normally 15 purchases per independent variable in a regression analysis. In areas with few purchases, the expert does not receive an adequate sample size, maybe there are only a few cases like 10 to 30 purchases by example. He or she has to estimate the

value by his or her experience with a view to this few purchases. In this case, the classical statistical analysis will suffer, or even fail.

The purpose of this study is to present an approach in order to solve the problem of small sample sizes. Here, a robust Bayesian approach is introduced to integrate experts' knowledge in data-driven models with a small sample size, such as the multiple linear regression analysis. The focus lies on one and two family houses. In order to simulate the market with few purchases, the data is methodically reduced, such as for cases in the edges or in the center of the data. Based on the Bayesian approach, data and knowledge were used in a comprehensive model. This model is established by using independent Student-t linear model, which is working properly in cases where significant deviations from the optimal assumptions occur. The computations are performed by means of the Markov Chain Monte Carlo method.

Keywords: Valuation, comparison approach, areas with few purchases, Bayesian statistics, multiple linear regression, robust estimators, Markov Chain Monte Carlo, Gibbs Sampler.

1 MOTIVATION

Die zuverlässige Immobilienbewertung ist seit der Finanz- und Immobilienkrise immer mehr in den Fokus geraten. Dies gelingt für viele Teilmärkte sehr gut. Liegen Informationen und insbesondere Kauffälle vor, so kann der Verkehrswert mit den standardisierten Wertermittlungsmethoden nach den Immobilienwertermittlungsverordnung entsprechend den Vorgaben der Rechtsprechung mit $\pm 20\%$ – 30% des wahren Wertes (BVerfG, Urt. v. 07.11.2006 1 BvL 10/02) ermittelt werden.¹

Die standardisierten Bewertungsverfahren erreichen ihre Grenzen allerdings sehr schnell, sowie nur wenige oder keine Informationen (speziell Kauffälle) über den Teilmarkt vorliegen. Diese sogenannten kaufpreisarmen Lagen sind sehr unterschiedlich: Es kann sich ebenso um den Teilmarkt der unbebauten Grundstücke (Ableitung von Bodenrichtwerten) in hochpreisigen Innenstadtlagen wie auch um Immobilienwerte in peripheren ländlichen Räumen handeln. Je geringer die Informationen sind, desto anspruchsvoller wird die Bewertung. Die Sachverständigen müssen die Werte aus anderen Teilmärkten ableiten oder nach gutachterlichem Sachverstand bestimmen. Hier wäre eine Methode von Vorteil, die es erlaubt, die wenigen Kauffälle und den objektiv erhobenen gutachterlichen Sachverstand zusammen auszuwerten.

In den vorangegangenen Veröffentlichungen /Alkhatib et al. 2012/, /Weitkamp et al. 2012a/ und /Weitkamp et al. 2012b/ konnte nachgewiesen werden, dass die Bayesische Regression² in simulierten kaufpreisarmen Lagen im Vergleich zur klassischen Regression sehr gute Ergebnisse liefert. Werden Informationen in Form von Prioriwissen eingeführt, so kann der funktionale Zusammenhang in simulierten Lagen besser ermittelt werden als mit der klassischen Regression. Wird nur ein Teil der ursprünglichen Stichprobe aufgrund der Reduktion betrachtet, bedarf es der Extrapolation der Regressionsfunktion: dies scheitert in der Regel und führt zu unzuverlässigen und verzerrten Ergebnissen. Hier hat die Bayesische Regression den

Vorteil, dass das Prioriwissen in der Lage ist, die fehlenden Daten zu kompensieren. Damit wird eine angemessene Lösung erzielt, die den Anforderungen der Rechtsprechung entspricht.

In den vorausgegangenen Publikationen, wie in /Weitkamp et al. 2012a/, wurden zwar die ursprünglichen Stichproben reduziert, allerdings mussten aufgrund numerischer Probleme mehr Kauffälle in die Auswertung einfließen als in kaufpreisarmen Lagen zu erwarten ist (150 statt 30 Kauffälle). Hier bedurfte es der Entwicklung eines Verfahrens, welches trotz weniger Kauffälle eine numerische Lösung des funktionalen Zusammenhangs ermöglicht.

Hinzu kommt, dass in einem klassischen Bayesischen Ansatz ähnlich wie in der klassischen Regressionsanalyse Ausreißer eliminiert werden müssen, da beide mittelwertbasierte Verfahren sind. Wird darauf verzichtet, Ausreißer zu beseitigen, so verzerrt dies aufgrund der Hebelwirkung das Ergebnis. In kaufpreisarmen Lagen bedeutet dies allerdings, dass auf Informationen verzichtet werden muss, die für eine zuverlässige Schätzung benötigt werden. Es wird ein Verfahren angestrebt, das trotz Ausreißer eine robuste Lösung erreichen kann.

In /Alkhatib et al. 2013/ wird daher erstmals ein robustes Bayesisches Verfahren eingeführt, dessen Ergebnisse im Folgenden im Vergleich zu den Ergebnissen aus anderen robusten, nicht robusten sowie informativen und nicht-informativen Bayesischen Verfahren gezeigt werden.

2 ANWENDUNG ROBUSTER VERFAHREN IN DER WERTERMITTLUNG

2.1 Das Vergleichswertverfahren

Das klassische Vergleichswertverfahren (nach Immobilienwertermittlungsverordnung) stellt in der Bewertungspraxis das marktnächste Verfahren dar, sofern genügend vergleichbare Kauffälle zur Verfügung stehen. Sein Einsatzgebiet umfasst unbebaute Grundstücke oder andere gut vergleichbare Objekte wie beispielsweise Eigentumswohnungen mit hinreichend übereinstimmenden Grundstücksmerkmalen. Vorteile des Vergleichswertverfahren liegen vor allem in seiner Marktnähe, da direkt reelle Kauffälle zur Ermittlung des Ver-

¹ Allerdings ist ein Nachweis dieser Kauffälle derzeit nur im Vergleichswertverfahren anhand von Residuenuntersuchungen nach der Regressionsanalyse erfolgt bzw. könnte erfolgen. Sensibilitätsanalysen der modellhaften Verfahren erfolgen in der Bewertungspraxis in der Regel nicht.

² Im deutschen Sprachgebrauch werden auch die Begriffe Bayessche oder Bayes'sche Regression verwendet und sind entsprechend gleichbedeutend. Hier wird der Version mit Fugenvokal Vorrang eingeräumt.

kehrswertes verwendet werden. Dieser wird somit nicht wie in den anderen normierten Verfahren modellhaft ermittelt und es muss der Bezug zum Markt nicht über z.B. Sachwertfaktoren oder Liegenschaftszins hergestellt werden. Über diverse Einflussgrößen wird der funktionale Zusammenhang für eine Zielgröße wie der Bodenpreis pro Quadratmeter oder der Wohnflächenpreis geschätzt. Dafür werden in der Regel 10 – 15 Kauffälle pro abzuleitender Einflussgröße benötigt /Ziegenbein 2009/, /Kleiber et al. 2010/.

Der funktionale Zusammenhang zwischen der abhängigen Variablen y (Zielgröße) und den k unabhängigen Variablen $X_1, \dots, X_k$ zusammengefasst in der Matrix X (Einflussgrößen, mit vollem Rang), ergibt sich durch:

$$y = X\beta + \varepsilon, \varepsilon \sim N(0, \sigma^2 Q_y). \quad (1)$$

Die Schätzung der Regressionskoeffizienten β erfolgt nach der Methode der kleinsten Quadrate mit den Restfehlern bzw. Residuen ε (z.B. /Gao et al. 2009/ oder /Koch 2007/). Aufgrund der Modellannahmen müssen für die korrekte Lösung die Residuen normalverteilt sein und im Mittelwert um 0 streuen; in der diagonalen Kofaktormatrix mit $Q_{yy} = \text{diag}(\omega_i)$ stellen ω_i die gewichteten Größen der Kauffälle (Beobachtungen) dar, σ^2 ist die Varianz der Beobachtungen (Kauffälle). Weitergehende Beschreibungen und Diskussionen finden sich in Standardwerken der Statistik (vgl. z.B. /Fahrmeir et al. 2009/, /Urban et al. 2011/) bzw. in entsprechenden Werken zum Vergleichswertverfahren in der Wertermittlungsliteratur (hier z.B. /Kleiber et al. 2010/ einführend, /Ziegenbein 1977/ und /Brückner 1976/ weitergehend).

2.2 Kaufpreisarme Lagen

Kaufpreisarme Lagen zeichnen sich dadurch aus, dass nur vereinzelt reale Kaufpreise verfügbar sind. Dies kann sowohl Geschäfts- und Wohnlagen in Städten als auch Ortslagen in ländlichen Räumen betreffen. Daneben benennt /Reuter 2006/ besondere Grundstückseigenschaften wie Mängel in der Beschaffenheit oder dingliche Belastungen, unklare Nutzungsaussichten bei gestörten Marktverhältnissen, insbesondere wegen erheblichen Wohnungs- und Gewerbeleerstands, sowie besondere gesetzliche Gegebenheiten, wie das Bodenrecht in förmlich festgelegten Sanierungsgebieten, als weiteres Problem der Wertermittlung. Damit sind diese Kauffälle eher als Ausreißer zu betrachten; im Ergebnis stehen somit zu wenige vergleichbare Kauffälle für eine angemessene Bewertung zur Verfügung. Kaufpreise werden vergleichsweise häufig ausgeschlossen; aus einer Umfrage von /Reuter 2006/ ergibt sich, dass 20 – 35 %, zumeist aufgrund von persönlichen Verhältnissen, nicht weiter berücksichtigt werden.

Bezogen auf die Ableitung von Bodenwerten wird anstelle des Vergleichswertverfahrens die periodische Ableitung und Fortschreibung von Bodenrichtwerten als häufigste verwendete Methode benannt; daneben erfolgt die Bewertung durch Heranziehung des Bodenanteils von Kaufpreisen bebauter Grundstücke sowie der Einsatz von Lagewertverfahren (zumeist in Sanierungsgebieten verwendet) und der Gebrauch von bodenpreisrelevanten Faktorleistungen (deduktiver Preisvergleich: über Mieten für den Ausgleich von Lageunterschieden oder über Entwicklungs- und Vorhaltekosten für die Er-

fassung unterschiedlicher Entwicklungszustände des Grund und Bodens) genannt /Reuter 2006/. /Dransfeld 2007/ unterscheidet diese Verfahren in rechnende Verfahren, in denen Fakten aus der Analyse des Marktes abgeleitet werden, und nutzwertanalytische Verfahren einschließlich der freien Schätzung /Dransfeld 2007/.

Für kaufpreisarme Lagen kann somit subsummierend festgehalten werden, dass das klassische Vergleichswertverfahren versagt. Es wird eine Methode benötigt, die es ermöglicht, die wenigen Kauffälle zu verarbeiten (numerisch wie auch zuverlässig), zusätzlich zu den wenigen Informationen den Sachverstand des Gutachters einzubringen und auf Ausreißerelimination³ zu verzichten.

2.3 Klassische robuste Methoden

Sollen die Ausreißer in der Stichprobe verbleiben und dennoch die Hebelwirkung vermieden werden, bieten sich Verfahren der robusten Parameterschätzung aus der klassischen Statistik an. Diese Methoden sind gegenüber Abweichungen von den Modellannahmen unempfindlich, wie beispielsweise hinsichtlich der Abweichung von der Normalverteilung durch Ausreißer (die z.B. nur in der Nähe der Verteilung liegen).

Robuste Verfahren führen zu funktionalen Zusammenhängen, die nicht durch Ausreißer verzerrt werden. Während die klassische Regressionsanalyse zu den mittelwertbasierten Verfahren zählt, sind die robusten Methoden mit dem Median vergleichbar. Sie sind ähnlich resistent gegenüber Ausreißern, haben aber den Vorteil, dass im Gegensatz zum Mittelwert keine Verzerrung durch Hebelwirkung erfolgt. Die Robustheit des Medians geht allerdings im Vergleich zum arithmetischen Mittel mit einer reduzierten Güte einher /Hartung et al. 2009/.

Es haben sich zahlreiche robuste Verfahren in der klassischen Inferenz etabliert, die hinsichtlich ihrer Schätzer unterschieden werden können: M-Schätzer, L-Schätzer und R-Schätzer. Sie unterscheiden sich u.a. in ihrem Umgang mit Ausreißern /Hartung et al. 2009/.

Speziell der Huber-Schätzer (zur Klasse der M-Schätzer gehörend) hat sich als sehr praktikabel erwiesen (Abb. 1). Hier werden Ausreißer durch eine Einflussfunktion begrenzt. In einem definierten Abstand der Ausreißer vom Großteil der Stichprobe wird das Gewicht der Ausreißer durch die Funktion $\psi(x)$ begrenzt. Der Abstand zwischen Ausreißer und dem Großteil der Stichprobe wird durch eine sogenannte *Tuning-Konstante* c reguliert: alle Stichprobenwerte (hier Kauffälle), die vom Median einen Abstand größer c haben, werden als Ausreißer abgewichtet. Nach /Hartung et al. 2009/, ist die Einflussfunktion der Huber-Methode wie folgt definiert:

$$\psi(x) = \begin{cases} x & \text{für } |x| \leq c \\ \text{sgn}(x) \cdot c & \text{für } |x| > c \end{cases} \text{ mit } \text{sgn}(x) = \begin{cases} 1 & x > 0 \\ 0 & \text{falls } x = 0 \\ -1 & x < 0 \end{cases} \quad (2)$$

³ Ausreißer im Sinne der Wertermittlung sind Kauffälle, in denen in Bezug auf die wertbeeinflussenden Einflussgrößen ein zu hoher oder zu niedriger Kaufpreis gezahlt wurde. Oftmals sind dies nicht erkannte Kauffälle, die einem ungewöhnlichen oder persönlichen Geschäftsverkehr entspringen /Kleiber et al. 2010/. /Ziegenbein 2009/ empfiehlt, Ausreißer über den teilmarktüblichen Variationskoeffizienten in Verbindung mit der Annahme einer zweifachen Standardabweichung zu ermitteln.

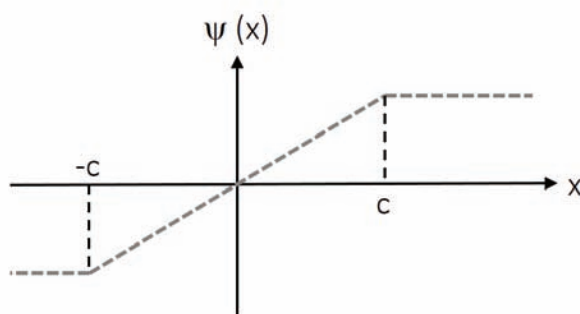


Abb. 1 | Robuste Parameterschätzung nach Huber

Für weiterführende Informationen und andere robuste Methoden wird auf z.B. /Hartung et al. 2009/, /Hampel 1980/, /Huber et al. 2009/, /Wicki 1998/, /Lindley 1971/ und /Kargoll 2004/ verwiesen.

Von besonderer Wichtigkeit ist für die Parameterschätzung nach Huber somit die Festlegung der Tuning-Konstanten. Ihr Wert ist abhängig vom Anteil der Ausreißer in den Daten. Beträgt ihr Anteil etwa 4 %, ist $c = 1,5$ zu wählen. Bei etwas unter 1 % wäre $c = 2,0$ /Koch 1997/. Für den Einsatz in kaufpreisarmen Lagen gilt es abzuschätzen, wie viele Kauffälle Ausreißer darstellen und welche Kauffälle im „normalen“ Geschäftsverkehr der kaufpreisarmen Lage vollzogen wurden. Um optimale Ergebnisse zu erzielen, wurde die Tuning-Konstante mit $c = 1,15$ gewählt. Die robuste Regression wurde – anders als in /Alkhatib et al. 2013/ – in einem iterativen Ansatz anstelle einer einfachen Lösung mit Annahme für s streng gelöst. Der Algorithmus basiert auf /Dutter et al. 1981/.

2.4 Einbindung gutachterlichen Sachverständs

Neben dem Verzicht, Aufreißer zu eliminieren, soll im Rahmen der Regression Vorwissen in Form von gutachterlichem Sachverstand eingebunden werden (vgl. hierzu /Alkhatib et al. 2012/, /Weitkamp et al. 2012a/ und /Weitkamp et al. 2012b/). Daher werden hohe Anforderungen an das Vorwissen gestellt: Während in der klassischen Regressionsanalyse der Sachverstand lediglich in die Modellwahl einfließt (Auswahl der Einflussgrößen und des grundsätzlichen funktionalen Zusammenhangs, z.B. linearer, polynomialer Zusammenhang), wird im Bayesischen Modell das Vorwissen den Kauffällen gegenübergestellt. Die Experten schätzen den Verkehrswert von Testobjekten aus den Teilmärkten bzw. es werden Gutachten der entsprechenden Teilmärkte ausgewertet. Diese Informationen bilden das Vorwissen – die Priori-Dichte in Form von Wahrscheinlichkeitsdichtefunktionen. Während eine korrekte Einschätzung bzw. korrekte Gutachten das Ergebnis (hier: der funktionale Zusammenhang) verbessern bzw. unterstützen (zuverlässigeres Ergebnis), würde eine schlechte oder falsche Einschätzung das Ergebnis verschlechtern. Speziell in kaufpreisarmen Lagen mit wenigen Daten hat das Vorwissen einen großen Einfluss auf das Ergebnis: hier gleicht es die fehlenden Daten aus, während nicht-informative Methoden aufgrund von Extrapolationen ggf. fehlschlagen (vgl. hierzu im Detail /Weitkamp et al. 2012b/).

Für informative Bayesische Verfahren ist es daher notwendig, dass nur Experten oder Gutachten zur Generierung des Vorwissens hin-

zugezogen werden, die in der Lage sind, den untersuchten Teilmarkt korrekt einzuschätzen. Dies kann durch Zertifizierung sichergestellt werden, z.B. nach /ISO/IEC 17024 2003/ – wobei diese in Deutschland fakultativ ist /Gondring 2004/. Nach /Ziegenbein 1999/ ist daneben von besonderer Bedeutung, dass Sachverständige fachkundig handeln und auf Fakten basierend und objektiv bewerten. Experten sollten unabhängig und neutral agieren und nicht finanziell oder persönlich von dem Auftraggeber abhängig sein. Diese hohen Anforderungen müssen speziell in Teilmärkten mit geringen Informationen wie kaufpreisarmen Lagen erfüllt werden: Je weniger Informationen vorhanden sind, desto mehr Erfahrung erfordert die Bewertung. Hier versagen die normierten Verfahren vielfach, sodass mit anderen Methoden der Wert ermittelt werden muss (vgl. 2.2 Kaufpreisarme Lagen).

Statt zertifizierter Sachverständiger wurde für die Expertenbefragung auf die Gutachterausschüsse für Grundstückswerte zurückgegriffen. Diese bewerten in einem unabhängigen Gremium und sind in den untersuchten Teilmärkten erfahren. Befragt wurden verschiedene unabhängige Mitglieder des Gutachterausschusses. Sie bewerteten Objekte nach ihrer üblichen Vorgehensweise /Weitkamp et al. 2012a/, /Weitkamp et al. 2012b/.

2.5 Die Bayesische Interferenz als robuster Ansatz

Im Folgenden wird der Bayesische Ansatz als robustes Modell eingeführt. Wie auch in der klassischen Regression bzw. Bayesischen Regression ist das Ziel, die Lösung des funktionalen Zusammenhangs und damit die Ermittlung der Regressionskoeffizienten β gegeben durch die Zielgröße y . Dies basiert entsprechend auf dem Bayes-Theorem:

$$p(\beta | y) \propto p(\beta)p(y | \beta). \quad (3)$$

Hierbei stellt $p(\beta | y)$ die Posteriori-Dichte dar. Diese enthält das Wissen über die Regressionskoeffizienten (Parameter) gegeben durch die Zielgröße y . Die Likelihood-Funktion $p(y | \beta)$ beinhaltet die Einflussgrößen (wertbeeinflussende Parameter). $p(\beta)$ enthält das Vorwissen. Zur Vertiefung der klassischen Bayesischen Interferenz sei auf die Vielzahl von Standardwerken verwiesen, z.B. /Koch 2007/ und /Kacker et al. 2003/.

Das Vorwissen wurde aus der Expertenbefragung oder Gutachten gewonnen und als sogenannte Pseudokauffälle ausgewertet. Im Interview sollen die Sachverständigen den Verkehrswert und Auswirkungen auf diesen durch Änderungen in den abhängigen Variablen schätzen. Die Ergebnisse dienen zur Erzeugung einer „neuen Stichprobe“ von Daten (Pseudokauffälle). Aus dieser Stichprobe, den Pseudokauffällen, wurden die normalverteilte Priori-Dichte $p(\beta)$, der Priori-Einheitsfaktor σ^2 und die Gewichtskoeffizienten der Kofaktormatrix der Kauffälle Q_{yy} abgeleitet. Die *Priori-Dichte* als Ergebnis stellt sich wie folgt dar:

$$p(\beta, \sigma, \omega) \propto p(\beta)p(\sigma)p(\omega). \quad (4)$$

(4) liegt die Annahme zugrunde, dass der Varianzfaktor, die Gewichtskoeffizienten und die Regressionsparameter als stochastisch unabhängig betrachtet werden. Die Varianz ist wie in der Bayesischen Regression unbekannt, somit wird die aus /Alkhatib et al. 2012/ übernommen. Hinzu kommt allerdings die Gewichtsfunktion,

da die Daten nicht ausreißerbereinigt wurden und dementsprechend heterogen sind. Durch die Dichtefunktion werden die Ausreißer entsprechend über die unbekannten Gewichtsparameter w_i abgewichtet. Die Priori-Verteilung für die Gewichtung wird entsprechend /Lindley 1971/ als unabhängige Student-t-Verteilung angenommen. Daraus ergibt sich die Notwendigkeit, die Freiheitsgrade angeben zu müssen (als Bedingung aus der t-Verteilung), die allerdings nicht bekannt sind. Die Varianz der Residuen ergibt sich zu:

$$\text{var}(\boldsymbol{\varepsilon}) = \sigma^2 \mathbf{Q}_{yy} \text{ mit } \mathbf{Q}_{yy} = \begin{bmatrix} \omega_1 & & 0 \\ & \omega_2 & \\ 0 & & \ddots \\ & & & \omega_n \end{bmatrix}^{-1}. \quad (5)$$

Die *Likelihood-Funktion* kann basierend auf Annahmen über die Residuen $\boldsymbol{\varepsilon}$ und die Einflussgrößen $\mathbf{X}$ berechnet werden; hier wurde der Ansatz von /Geweke 1993/ gewählt:

$$p(\mathbf{y} | \mathbf{X}\boldsymbol{\beta}, \text{var}(\boldsymbol{\varepsilon})) \sim N(\mathbf{X}\boldsymbol{\beta}, \text{var}(\boldsymbol{\varepsilon})). \quad (6)$$

Vorteil dieses Ansatzes ist es, dass heteroskedastische Daten (mit Ausreißern behaftet) verwendet werden können. Die Varianz-Kovarianz-Matrix ist diagonal besetzt und gewichtet die Daten. Anstelle der Normalverteilung wird die t-Verteilung verwendet und die Freiheitsgrade ν sind unbekannt.

Durch Faltung der Priori-Dichte (Formel (4)) mit der Likelihood-Funktion (Formel (6)) kann die *Posteriori-Dichte* ermittelt werden. Die Posteriori-Dichte hat in diesem Fall keine geschlossene Form und ist somit im robusten Ansatz analytisch nicht lösbar – im Gegensatz zu der klassischen Bayesischen Regression (unter Verteilungsannahme Normal-Gamma) nach /Koch 2007/, umgesetzt in /Alkhatib et al. 2012/. Damit ist eine Lösung nur numerisch möglich; dies erfolgt als Markov-Chain-Monte-Carlo-Algorithmus, in diesem Fall als Gibbs-Sampler, in dem eine stochastische Kette generiert wird, welche zu der gesuchten Verteilung der zu schätzenden Parameter konvergiert. Aus dieser Kette können die verschiedenen Momente der Verteilungsfunktion ermittelt werden: die Regressionskoeffizienten $\boldsymbol{\beta}$, die Varianz σ^2 , die Gewichtung $\boldsymbol{\omega}$ und die Freiheitsgrade ν . Im dem speziellen Verfahren, dem Gibbs-Sampler, wird die Kette so generiert, dass die bedingte Verteilungsfunktion jeder gesuchten Größe (Regressionskoeffizienten $\boldsymbol{\beta}$, Varianz σ , Gewichtskoeffizienten $\boldsymbol{\omega}$ und Freiheitsgrade ν) ermittelt wird. Dies erfolgt iterativ mit bestimmten Annahmen als Startwerte. Dieser Prozess muss genügend oft erfolgen, sodass eine Konvergenz der Kette zu den gesuchten Verteilungsfunktionen sichergestellt werden kann (hier $m = 5.000$ mal). Um den Einfluss der Startwerte auf das Ergebnis zu vermindern, werden die ersten Iterationen als *Burn-In-Phase* nicht zur Berechnung der Momente verwendet (hier 500 mal) /Kroese et al. 2013/.

/Koch 2007/. Für weitere Ausführungen sei insbesondere auf /Geweke 1993/ verwiesen.

2.6 Verwendete Methoden im Vergleich

Im Folgenden werden die oben eingeführten Methoden gegenübergestellt (Tab. 1). Einerseits werden die klassische und die klassische Bayesische Regressionsanalyse verwendet. Der Unterschied zwischen beiden Methoden besteht darin, dass in der klassischen Regression kein Sachverstand als informatives Priori-Wissen eingeht (hier wird sogenanntes nicht-informatives Priori-Wissen verwendet), während in der klassischen Bayesischen Regression Sachverstand, z.B. in Form von Pseudokauffällen, gewonnen aus Expertenbefragung oder Gutachten, einfließt. Aufgrund der Invertierung der Normalgleichungen ist eine Lösung vielfach nicht zuverlässig, da die Matrix schlecht konditioniert ist, bzw. gar nicht möglich. Damit versagen die Verfahren häufig in kaufpreisarmen Lagen.

Als robuster Schätzer kommt die Robuste Parameterschätzung nach Huber zum Einsatz; diese ist gegenüber Ausreißern unempfindlich, benötigt allerdings rein numerisch mehr Daten als gemeinhin in kaufpreisarmen Lagen zu finden ist. Als Weiterentwicklung wird daher die in /Alkhatib et al. 2013/ erstmals in die Wertermittlung eingeführte Methode der robusten Bayesischen Regression angewandt. Diese ist in der Lage, Daten und Sachverstand zu kombinieren, wenige Daten verarbeiten zu können und robust gegenüber Ausreißern zu sein.

3 ANWENDUNG ROBUSTER VERFAHREN IN SIMULIERTEN KAUFPREISARMEN LAGEN – EINE VERGLEICHSTUDIE

Eine erste Untersuchung erfolgt im räumlichen Teilmarkt von Osnabrück (Stadt und Landkreis). Hier wurde als sachlicher Teilmarkt der für freistehende Ein- und Zweifamilienhäuser, Doppelhaushälften und Reihenhäuser gewählt. Die Kauffälle sind im gewöhnlichen Geschäftsverkehr entstanden und frei von ungewöhnlichen oder persönlichen Verhältnissen. Der funktionale Zusammenhang ergibt sich zwischen der Zielgröße Wohnflächenpreis (€/m²) und den Einflussgrößen Grundstücksfläche (m²), Bodenrichtwert (€/m²), Baujahr (a), Wohnfläche (m²) und Ausstattungsstandard (einheitenlos). Dieser Teilmarkt war auch Gegenstand vorheriger Untersuchungen /Alkhatib et al. 2012/, /Weitkamp et al. 2012b/, /Weitkamp et al. 2012a/ und /Alkhatib et al. 2013/, was einen Vergleich der verschiedenen Methoden ermöglicht. Das Vorwissen wurde aus Experteninterviews bzw. aus der Auswertung von Gutachten gewonnen. Dadurch konn-

Verfahren	Informative Methode (Kombination Daten und Sachverstand)	Verarbeitung weniger Daten möglich	Robust gegenüber Ausreißer
Klassische Regression	Nein	Nein	Nein
Klassische Bayesische Regression	Ja	Bedingt	Nein
Huber	Nein	Nein	Ja
Robuste Bayesische Regression	Ja	Ja	Ja

Tab. 1 | Gegenüberstellung der Methoden

ten 270 bzw. 75 Pseudokauffälle erzeugt werden. Während in /Alkhatib et al. 2013/ nur Vorwissen aus der Expertenbefragung eingebracht wurde, sollen nun beide Arten des Vorwissens analysiert werden.

3.1 Reduzierung der Kauffälle und Validierungsstrategie

Anders als in /Alkhatib et al. 2013/ wird nun die Reduzierung der Gesamtstichprobe nicht nur auf das Maximum beschränkt. Die Stichprobe wird zunächst zufällig auf $n = 30$ Kauffälle reduziert. Die weitere Reduktion erfolgt systematisch und ist in *Abb. 2* anhand des Histogramms der Zielgröße Wohnflächenpreis dargestellt: die $n = 30$ Kauffälle werden aus der Mitte, aus beiden Rändern bzw. aus dem Minimum oder dem Maximum gezogen. Die Wohnflächenpreise liegen im Median bei 1.157 €/m² Wohnfläche.

Für jede Art der Reduzierung wird das Experiment 1.000 mal wiederholt (Monte-Carlo-Run). Die Ergebnisse werden durch eine vorab ausgeschiedene Stichprobe validiert. Hierzu werden 250 Kauffälle zur Überprüfung der Ergebnisse zurückgelegt und im Rahmen einer Kreuzvalidierung verwendet. Zur Überprüfung der Ergebnisse wird die Wurzel des mittleren quadratischen Fehlers (root mean squared error/RMSE) verwendet /Willmott et al. 2005/, /Hartung et al. 2009/:

$$RMSE = \sqrt{\frac{1}{n} \times \sum_{i=1}^n (\hat{y}_i - y_i)^2} \quad (7)$$

Dadurch ist eine generelle Aussage bzgl. der Verhaltensweise der Verfahren möglich. Neben der Beurteilung der Güte der Ergebnisse anhand der RMSE soll das Ergebnis hinsichtlich der numerischen Lösbarkeit beurteilt werden.

3.2 Gegenüberstellung der Ergebnisse

Nach systematischer Reduzierung der Gesamtstichprobe wurden die oben benannten Verfahren der klassischen Bayesischen Regres-

sion, der robusten Regression nach Huber (nicht-informativ) sowie die robuste Bayesischen Regression informativ wie auch nicht-informativ angewandt. Für die informativen Verfahren wurden als Priori-Informationen sowohl das Expertenwissen aus der Befragung als auch die ausgewerteten Gutachten verwendet.

Die Gegenüberstellung der Ergebnisse erfolgt zunächst anhand der RMSE (*Tab. 2* und *Tab. 3*) und der Abweichung der jeweiligen RSME vom Median der Zielgröße Wohnflächenpreis. Schwankungen in den RMSE der Gesamtstichprobe ergeben sich daraus, dass zunächst eine Validierungsstichprobe entnommen wird, die für jedes Verfahren erneut gezogen wird. Die RMSE der Gesamtstichprobe liegt bei durchschnittlich 226 €/m². Dies entspricht einer prozentualen Abweichung von durchschnittlich 19,5 % vom Median des Wohnflächenpreises von 1.157 €/m².

Werden zunächst die informativen Verfahren, die Vorwissen nutzen, und nicht-informative Verfahren gegenüber gestellt, so kann festgestellt werden, dass die informativen Verfahren immer dann den nicht-informativen Verfahren überlegen sind, wenn Daten fehlen, die die Steigung der Regressionsgeraden beeinflussen. Dies betrifft die Reduktion auf die Mitte sowie die einzelnen Extrema (Minimum bzw. Maximum). Während die nicht-informativen Verfahren 40 – 60 % höhere RMSE im Vergleich zu den RMSE der Originalstichprobe aufweisen, liegen die RMSE der informativen Verfahren nur 10 – 20 % über den RMSE der Originalstichprobe. Hier versagen die nicht-informativen Verfahren, da eine Extrapolation der Regressionsfunktion erfolgen muss, die eigentlich nicht zulässig ist (der funktionale Zusammenhang ist nur innerhalb der Stichprobe gültig). Dies entspricht den Erkenntnissen aus /Weitkamp et al. 2012b/: die robusten Verfahren verhalten sich diesbezüglich ähnlich wie die klassischen.

Am besten passen sich die Funktionen im Bereich an, in dem Daten vorliegen. *Abb. 3* veranschaulicht das Verhalten für die Reduktion auf die Mitte, *Abb. 4* für die Reduktion auf das Minimum und *Abb. 5* für die Reduktion auf das Maximum jeweils der Originalstichprobe (*Abb. 3* verwenden Expertenwissen als Priori-Informationen, *Abb. 4* und *Abb. 5* Gutachtendaten). Sehr gut deutlich wird, dass die beiden nicht-informativen keine Extrapolation außerhalb der Daten

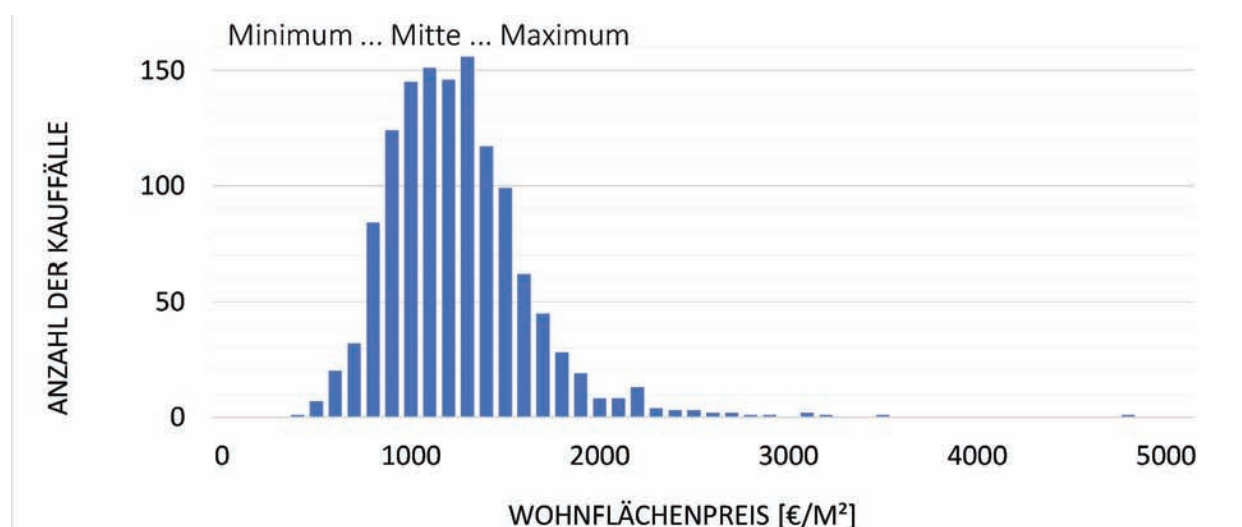


Abb. 2 | Histogramm der abhängigen Variable Wohnflächenpreis

	Zufällig		Mitte		Ränder		Maximum		Minimum	
Klassische (gesamt)	225	19,4 %	227	19,6 %	225	19,4 %	226	19,5 %	227	19,6 %
Klassische Bayesische, informativ	242	20,9 %	247	21,3 %	240	20,7 %	276	23,8 %	254	21,9 %
Huber, nicht-informativ	255	22,0 %	419	36,2 %	277	23,9 %	482	41,6 %	558	48,2 %
Robuste Bayesische, nicht-informativ	544	47,0 %	672	58,1 %	234	20,2 %	493	42,6 %	694	60,0 %
Robuste Bayesische, informativ	242	20,9 %	250	21,6 %	241	20,8 %	252	21,8 %	254	21,9 %

Tab. 2 | Vergleich der verschiedenen Verfahren anhand der RMSE [€/m²] und der prozentualen Abweichung der RMSE vom Median der Zielgröße Wohnflächenpreis (€/m²) für die Expertenbefragung

	Zufällig		Mitte		Ränder		Maximum		Minimum	
Klassische (gesamt)	226	19,5 %	225	19,4 %	226	19,5 %	225	19,4 %	226	19,5 %
Klassische Bayesische, informativ	260	22,5 %	261	22,5 %	267	23,1 %	338	29,2 %	277	23,9 %
Huber, nicht-informativ	254	21,9 %	475	41,0 %	278	24,0 %	489	42,2 %	564	48,7 %
Robuste Bayesische, nicht-informativ	622	53,7 %	543	46,9 %	313	27,0 %	497	42,9 %	651	56,2 %
Robuste Bayesische, informativ	267	23,1 %	263	22,7 %	293	25,3 %	355	30,7 %	288	24,9 %

Tab. 3 | Vergleich der verschiedenen Verfahren anhand der RMSE [€/m²] und der prozentualen Abweichung der RMSE vom Median der Zielgröße Wohnflächenpreis (€/m²) für die Gutachten

erlauben. Die beiden Bayesischen informativen Verfahren lösen auch außerhalb der Daten den funktionalen Zusammenhang vergleichsweise gut. Die klassische Bayesische Regression löst die Regressionsfunktion unwesentlich besser als das robuste Verfahren, versagt aber häufig numerisch aufgrund der wenigen Daten.

Während in Weitkamp et al. 2012b/ mit ausreißerbefreiten Stichproben gearbeitet wurde, wurde nunmehr die Originalstichprobe aus der Automatisierten Kaufpreissammlung verwendet. Infolge dessen sind die RMSE leicht erhöht (180 €/m² auf 225 €/m²). Ein signifikanter Unterschied zwischen den informativen Bayesischen Verfahren ist allerdings nicht zu verzeichnen – allerdings ist die Stichprobe sehr homogen (vgl. Histogramm in Abb. 1). Von allen Verfahren ist die robuste Regression von Huber am instabilsten. Bis zu 20 % der Monte-Carlo-Durchläufe können aufgrund numerischer Probleme nicht gelöst werden.

Die RMSE der informativen Bayesischen Verfahren weichen durchschnittlich 23 % vom Median ab, während die nicht-informativen Verfahren durchschnittlich 40 % Abweichung aufweisen. Wird die Forderung der Rechtsprechung nach einer Genauigkeit von $\pm 20 - 30$ % zugrunde gelegt, so können die informativen Verfahren dieser Forderung gerecht werden – auch in Lagen, in denen nur wenige Daten (Kauffälle) vorliegen. Die nicht-informativen Verfahren, sofern eine Lösung numerisch herbeigeführt werden kann, können dem nicht entsprechen.

Im Vergleich der unterschiedlichen Priori-Daten (Tab. 4) ist das Wissen aus der Expertenbefragung etwas geeigneter; hier werden leicht bessere Ergebnisse erzielt.

Dies liegt aber zum Teil an der unterschiedlichen Anzahl der Daten (270 Pseudokauffälle durch Expertenbefragung versus 70 Pseudokauffälle aus Gutachten) und auch an der Qualität des Vorwissens. Während in der Expertenbefragung der untersuchte Teilmarkt sehr genau abgebildet wurde, mussten für die Generierung des Priori-Wissens aus Gutachten alle vorliegenden Gutachten verwendet wer-

	Expertenbefragung		Auswertung Gutachten	
Klassische Bayesische Regression, informativ	252	21,8 %	281	24,2 %
Robuste Bayesische Regression, informativ	248	21,4 %	293	25,3 %

Tab. 4 | Vergleich der unterschiedlichen Priori-Informationen bezogen auf die durchschnittlichen RMSE bzw. die durchschnittliche Abweichung der RMSE vom Median

den. Hierunter befanden sich durchaus auch Gutachten sehr schlechter bzw. sehr guter Objekte aus den Randbereichen.

4 FAZIT UND AUSBLICK

Die informativen Bayesischen Verfahren eignen sich sehr gut, wenn nur wenige Daten vorliegen wie in kaufpreisarmen Lagen, in denen ggf. in den Extrema Daten fehlen – beispielsweise, wenn sich nur gute Objekte verkaufen lassen oder aber ggf. nur schlechte/günstige Objekte vermarktbare sind. Hier gleicht das Vorwissen ein Fehlen von Daten aus. Speziell das robuste Bayesische Verfahren ist dazu geeignet, die numerische Instabilität zu regularisieren und zudem den Einfluss von Ausreißern herab zu gewichten. Zukünftiger Forschungsbedarf ist hinsichtlich der Verarbeitung einer größeren Anzahl von Ausreißern und in heterogeneren Teilmärkten zu verzeichnen. Eine Weiterentwicklung der klassischen robusten Verfahren wie z.B. nach Huber scheint für kaufpreisarme Lagen allerdings nicht zielführend. Robuste Bayesische Verfahren jedoch erscheinen diesbezüglich vielversprechend. Hier sollten Weiterentwicklungen

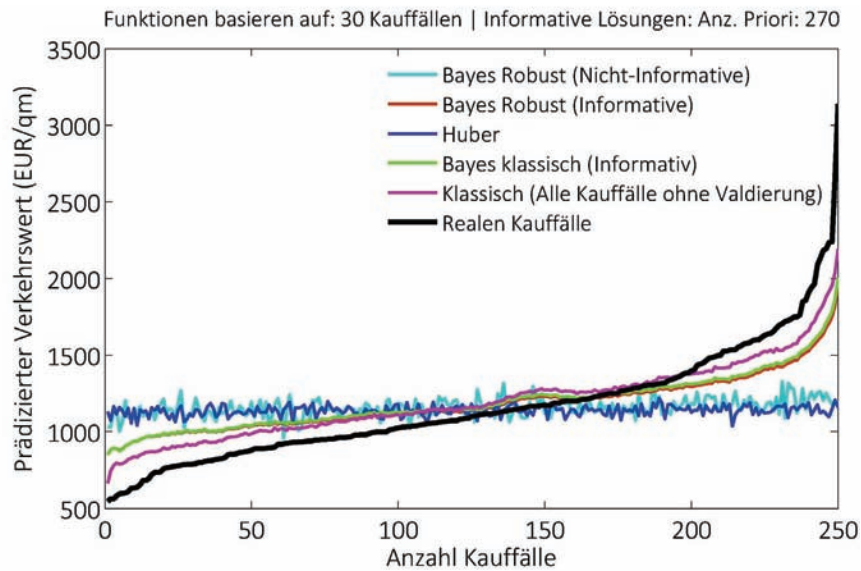


Abb. 3 | RMSE für die Reduktion auf die Mitte der Originalstichprobe und die Nutzung von Expertenwissen als Priori-Information

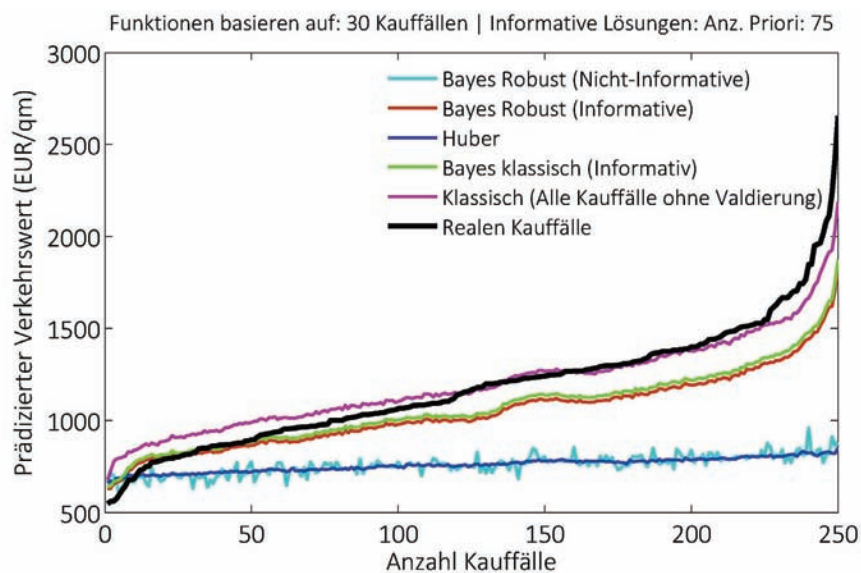


Abb. 4 | RMSE für die Reduktion auf das Minimum der Originalstichprobe und die Nutzung von Gutachten als Priori-Information

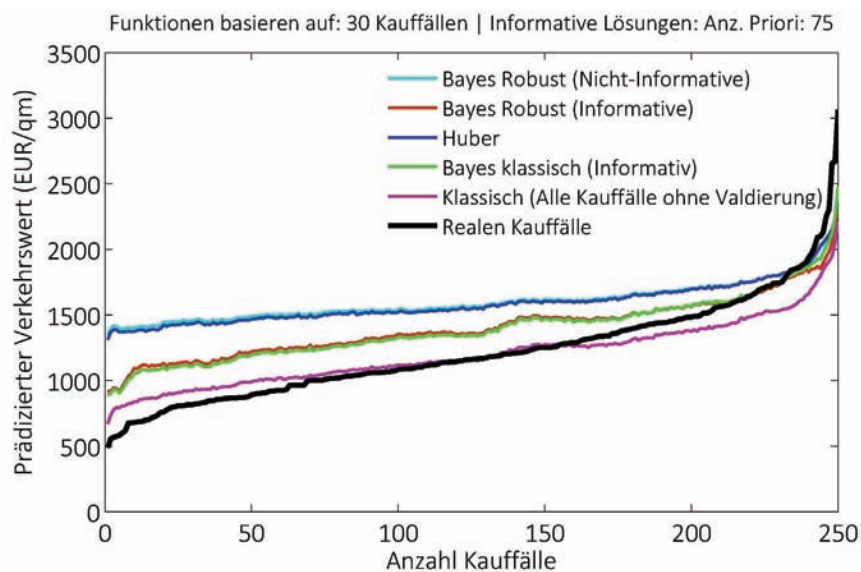


Abb. 5 | RMSE für die Reduktion auf das Maximum der Originalstichprobe und die Nutzung von Gutachten als Priori-Information

bzgl. der Verteilungsannahmen, der Gewichtung zwischen Vorwissen und Daten sowie der Monte-Carlo-Methoden erfolgen.

Daneben bedarf es der Erprobung der Verfahren in realen kaufpreisarmen Lagen: dazu muss eine Validierungsstrategie entwickelt werden. Eine Validierung mit Daten ist höchstens über die Delete-1-Kreuzvalidierung möglich. Hier sollte eine Strategie über Expertenbefragung erarbeitet werden.

Zudem besteht Forschungsbedarf hinsichtlich der Integration von Vorwissen: Sowohl die Expertenbefragung als auch Gutachten haben sich als geeignet erwiesen – Letzteres wäre mit deutlich geringerem Aufwand verbunden, wenn auch die Qualität des Vorwissens nicht ganz an die der Expertenbefragung heranreicht und in kaufpreisarmen Lagen auch oftmals Gutachten fehlen. Hier sollte geprüft werden, welche Art von Daten ebenfalls mit der entsprechenden Gewichtung zielführend eingesetzt werden könnte: Daten von Banken scheinen vielversprechend, aber auch Angebotsdaten könnten sich mit entsprechender Gewichtung als hilfreich in kaufpreisarmen Lagen erweisen – nutzen doch Sachverständige schon heute in Lagen, in denen keine Daten vorliegen, diese Informationsquelle.

Robuste Bayesische Verfahren könnten in kaufpreisarmen Lagen Sachverständigen zukünftig eine weitere Hilfestellung bei der Bewertung bieten. Eine statistische Untersuchung, die die Bewertung stützt, ist speziell für den Streitfall vor Gericht eine gute Unterstützung und Begründung; diese ist objektiver als die Bewertung aufgrund der reinen Erfahrung des Sachverständigen.

LITERATUR

Alkhatib, H.; Weitkamp, A. (2012): Bayesischer Ansatz zur Integration von Expertenwissen in die Immobilienbewertung, Teil 1. In: Zeitschrift für Vermessungswesen (zfv) 137(2012)2, 93-102.

Alkhatib, H.; Weitkamp, A. (2013): Robust Bayesian Regression Approach For Areas With Small Numbers Of Purchases. In: Royal Institution of Chartered Surveyors (Ed.): Proceedings of COBRA Conference, New Delhi 2013.

Brückner, R. (Hrsg.) (1976): Mathematische Statistik bei der Ermittlung von Grundstückswerten – Lehrbriefe und Vorlesungen zum Kontaktstudium des Geodatischen Instituts 1976. Eigenverlag, Hannover (Wissenschaftliche Arbeiten der Lehrstühle Geodäsie, Photogrammetrie und Kartographie an der Technischen Hochschule Hannover, 65).

Dransfeld, E. (2007): Grundstückswertermittlung im Stadtumbau. Verkehrswertermittlung bei Schrumpfung und Leerstand. Selbstverl. des Bundesamtes für Bauwesen und Raumordnung, Bonn (Forschungen / Bundesministerium für Verkehr, Bau und Stadtentwicklung, 127).

Dutter, R.; Huber, P. (1981): Numerical methods for the nonlinear robust regression problem. In: Journal of Statistical Computation and Simulation, 13(1981)2, 79-113.

Fahrmeir, L.; Kneib, T.; Lang, S. (2009): Regression. Modelle, Methoden und Anwendungen. 2. Aufl. Springer, Heidelberg.

Gao, Y.; Sun, D. (2009): Calibrating least squares covariance matrix problems with equality and inequality constraints. In: SIAM Journal on Matrix Analysis and Applications, 31(2009)3, 1432-1457.

Geweke, J. (1993): Bayesian Treatment of the Independent Student-t Linear Model. In: Journal of Applied Econometrics, (1993)8, 19-40.

Gondring, H. (2004): Immobilienwirtschaft. Handbuch für Studium und Praxis. Vahlen, München.

Hampel, F. (1980): Robuste Schätzungen: ein anwendungsorientierter Überblick. In: Biom. J., 22(1980)1, 3-21.

Hartung, J.; Elpelt, B.; Klösener, K.-H. (2009): Statistik. Lehr- und Handbuch der angewandten Statistik [mit zahlreichen durchgerechneten Beispielen]. 15. Aufl. Oldenbourg, München.

Huber, P.; Ronchetti, E. (2009): Robust statistics. 2. Aufl. Wiley, Hoboken, N.J. (Wiley series in probability and statistics).

ISO/IEC 17024 (2003): Conformity assessment – General requirements for bodies operating certification of persons, Geneva.

Kacker, R.; Jones, A. (2003): On use of Bayesian statistics to make the guide to the expression of uncertainty in measurement consistent. In: Metrologia, 40(2003), 235-248.

Kargoll, B. (2004): Comparison of some robust parameter estimation techniques for outlier analysis applied to simulated GOCE mission data. In: Jekeli, C.; Bastos, L.; Fernandes, J.; Sansò, F. (Eds.): Gravity, Geoid and Space Missions.: GGSM 2004. IAG International Symposium Porto, Portugal. Springer, Berlin/Heidelberg/New York.

Kleiber, W.; Fischer, R.; Simon, J. (2010): Verkehrswertermittlung von Grundstücken. Kommentar und Handbuch zur Ermittlung von Marktwerten (Verkehrswerten), Versicherungs- und Beileihungswerten unter Berücksichtigung der ImmoWertV. 6. Aufl. Bundesanzeiger Verlag mbH, Köln.

Koch, K.-R. (1997): Parameterschätzung und Hypothesentests in linearen Modellen. 3. Aufl. Dümmler, Bonn.

Koch, K.-R. (2007): Introduction to Bayesian Statistics. Springer, Berlin/Heidelberg.

Kroese, D.; Taimre, T.; Botev, Z. (2013): Handbook of Monte Carlo Methods. 1. Aufl. Wiley, s.l.

Lindley, D. (1971): The Estimation of Many Parameters. In: Godambe, V.; Sprott, D. (Eds.): Foundations of statistical inference: Proceedings of the Symposium on the Foundations of Statistical Inference, Toronto 1971, 435-455.

Reuter, F. (2006): Zur Ermittlung von Bodenwerten in kaufpreisarmen Lagen. In: Flächenmanagement und Bodenordnung, 68(2006)3, 97-107.

Urban, D.; Mayerl, J. (2011): Regressionsanalyse: Theorie, Technik und Anwendung. 4. Aufl. VS Verlag für Sozialwissenschaften, Wiesbaden (Studienkripten zur Soziologie).

Weitkamp, A.; Alkhatib, H. (2012): Bayesischer Ansatz zur Integration von Expertenwissen in die Immobilienbewertung, Teil 2. In: Zeitschrift für Vermessungswesen (zfv), 137(2012)2, 103-114.

Weitkamp, A.; Alkhatib, H. (2012b): The Bayesian approach in the valuation – a strategy to handle markets with low purchasing prices? In: FIG (Eds.): Proceedings of FIG Working Week 2012, Rome, Italy 2012.

Wicki, F. (1999): Robuste Schätzverfahren für die Parameterschätzung in geodätischen Netzen. Mitteilung Nr. 67, Institut für Geodäsie und Photogrammetrie an der ETH Zürich.

Willmott, C.; Matsuura, K. (2005): Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. In: Climate Research, 30(2005), 79-82.

Ziegenbein, W. (1977): Zur Anwendung multivarianter Verfahren in der mathematischen Statistik in der Grundstückswertermittlung. Dissertation, Technische Universität Hannover.

Ziegenbein, W. (1999): Wertinformationssystem Niedersachsen. In: Nachrichten der Niedersächsischen Vermessungs- und Katasterverwaltungen, 4(1999)3, 321-335.

Ziegenbein, W. (2009): Immobilienwertermittlung. In: Kummer, K.; Frankenberg, J. (Hrsg.): Das deutsche Vermessungs- und Geoinformationswesen 2009, 423-471. Wichmann, Heidelberg.

Dr.-Ing. Alexandra Weitkamp

LEIBNIZ UNIVERSITÄT HANNOVER
GEODÄTISCHES INSTITUT – FLÄCHEN-
UND IMMOBILIENMANAGEMENT

Nienburger Str. 1 | 30167 Hannover
weitkamp@gih.uni-hannover.de



Dr.-Ing. Hamza Alkhatib

LEIBNIZ UNIVERSITÄT HANNOVER
GEODÄTISCHES INSTITUT – INGENIEURGEODÄSIE
UND AUSWERTEMETHODEN

Nienburger Str. 1 | 30167 Hannover
alkhatib@gih.uni-hannover.de



Manuskript eingereicht: 18.10.2013 | Im Peer-Review-Verfahren begutachtet

7.2 Originalpublikationen zur Kapitel 3

- P#7** Vogel, S., **Alkhatib, H.**, Neumann, I. (2018) ¹: Iterated Extended Kalman Filter with Implicit Measurement Equation and Nonlinear Constraints for Information-Based Georeferencing, In: 21st International Conference on Information Fusion (FUSION) 2018, Cambridge, United Kingdom, July 10-13, 2018, p. 1209-1216. DOI: 10.23919/ICIF.2018.8455258
- P#8** Vogel, S.; **Alkhatib, H.**; Bureick, Johannes; Moftizadeh, Rozhin; Neumann, Ingo (2019): Georeferencing of Laser Scanner-Based Kinematic Multi-Sensor Systems in the Context of Iterated Extended Kalman Filters Using Geometrical Constraints. In: Sensors 19 (10). DOI: 10.3390/s19102280.
- P#9** Bureick, J.; Vogel, S.; Neumann, I.; Unger, J.; **Alkhatib, H.** (2019): Georeferencing of an Unmanned Aerial System by Means of an Iterated Extended Kalman Filter Using a 3D City Model. In: PFG 2 (3), S. 429. DOI: 10.1007/s41064-019-00084-x.
- P#10** **Alkhatib, H.**; Paffenholz, J.-A.; Kutterer, H. (2012) Sequential Monte Carlo filtering for nonlinear GNSS trajectories, In: N. Sneeuw, P. Novák, M. Crespi und F. Sansò (Hg.): VII Hotine-Marussi Symposium on Mathematical Geodesy. Proceedings of the Symposium in Rome. Rome, 06.-10.06.2009. International Association of Geodesy (IAG). 1st Edition. Berlin, Heidelberg: Springer Verlag (International Association of Geodesy Symposia, 137), S. 81–86.
- P#11** **Alkhatib, H.** (2015) Alternative Nonlinear Filtering Techniques In Geodesy For Dual State And Adaptive Parameter Estimation, In: Proceedings of the 1st International Workshop on the Quality of Geodetic Observation and Monitoring Systems (QuGOMS'11), H. Kutterer, F. Seitz, H. Alkhatib, M. Schmidt, (Ed.), IAG Symp. 140, pp. 107-113. ISBN: 978-3-319-10827-8
- P#12** Sun, L.; **Alkhatib, H.**; Kargoll, B.; Kreinovich, V.; Neumann, I. (2019): Ellipsoidal and Gaussian Kalman Filter Model for Discrete-Time Nonlinear Systems, In: Mathematics 7 (12), DOI: 10.3390/math7121168.
- P#13** Sun, L.; **Alkhatib, H.**; Paffenholz, J.-A.; Neumann, I. (2018)²: Geo-Referencing of a Multi-Sensor System Based on Set-Membership Kalman Filter. In: International Conference on Information Fusion (Hg.): 21st International Conference on Information Fusion (FUSION 2018). Cambridge. Piscataway, NJ: IEEE, S. 1–5, DOI: 10.23919/ICIF.2018.8455763.

¹© 2020 IEEE. Reprinted with permission from Sören Vogel, Hamza Alkhatib and Ingo Neumann

²© 2020 IEEE. Reprinted with permission from Ligang Sun, Hamza Alkhatib, Jens-André Paffenholz and Ingo Neumann

Iterated Extended Kalman Filter with Implicit Measurement Equation and Nonlinear Constraints for Information-Based Georeferencing

1st Sören Vogel
Geodetic Institute
Leibniz Universität Hannover
Hanover, Germany
vogel@gih.uni-hannover.de

2nd Hamza Alkhatib
Geodetic Institute
Leibniz Universität Hannover
Hanover, Germany
alkhatib@gih.uni-hannover.de

3rd Ingo Neumann
Geodetic Institute
Leibniz Universität Hannover
Hanover, Germany
neumann@gih.uni-hannover.de

Abstract—Accurate, reliable and complete georeferencing with kinematic multi-sensor systems (MSS) is very demanding if common types of observations (e.g. usually GNSS) are imprecise or completely absence. The main reasons for this are challenging areas of indoor applications or inner-city areas with shadowing and multipath effects. However, those complex and tough environments are rather the rule than the exception. Consequently, we are developing an information-based georeferencing approach which can still estimate precise and accurate pose parameters when other current methods may fail. We modified an iterated extended Kalman filter (IEKF) approach which can deal with implicit measurement equations and introduced nonlinear equality constraints for the state parameters to integrate additional information. Hence, we can make use of geometric circumstances in the direct environment of the MSS and provide a more precise and reliable georeferencing.

Index Terms—IEKF, georeferencing, nonlinear constraints, implicit model, multi-sensor systems, 6-DOF

I. INTRODUCTION

There are many different sensors and methods used for georeferencing with an arbitrary kinematic multi-sensor system (MSS) in its environment [1]. In this paper, *georeferencing* describes the reference to a superordinate system and is generally used. It will be assumed that the realizations of such superordinate systems are possible by means of the control points available.

Accurate georeferencing of kinematic MSS within complex indoor environments and in inner-city outdoor environments is very challenging and needs a lot of effort due mainly to imprecise and unreliable GNSS observations (e.g. noisy, drifted or missing data) and significant weaknesses of other sensors. Thus, real-time processing to estimate the pose parameters (3 translations and 3 rotations, in total 6 degrees of freedom (6-DOF)) with high accuracy is difficult to achieve. However, there is a great demand for suitable solutions. The mapping component for building information modeling (BIM) can be mentioned as an example of the importance of accurate and reliable approaches for indoor use. Over the last few

years, planning, building, development and management of buildings required reliable, accurate and dense point clouds within indoor environments [2]. Accurate georeferencing with reliability aspects has also become increasingly important for outdoor applications, especially in case of autonomous driving. A reliable, continuously available, accurate and precise georeferencing solution of an autonomous vehicle is indispensable in this field.

Instead of using a wide variety of heterogeneous sensors, we aim to use data from only one high accurate laser scanner in profile mode and one moderate inertial measurement unit (IMU). Both sensors are rigidly interconnected with a vertical alignment of the 2D scanning plane. We developed a recursive estimation approach for georeferencing by including different information (e.g. derived directly from the environment) to obtain accurate solutions. Therefore, we focus on an optimal integration process of object space information from the laser scanner and a mathematical formulation of prior information by means of geometrical constraints. This available information will be combined within an adjustment approach to enhance the accuracy and reliability of the pose obtained. We use a modified version of an iterated extended Kalman filter (IEKF) with an implicit measurement equation and nonlinear equality constraints for the online compatibility of such a combination.

Finally, in contrast to neglecting the additional information, it is possible to achieve a more accurate, reliable and complete georeferencing solution. The mandatory requirements for our approach are consistently available laser scanner profiles (LSP) and IMU observations as well as basic structural conditions, such as accessible elements of walls, ceilings, floors, facades or road surfaces. Therefore, it will be assumed that all prior information (in accordance with the respective environment) is already segmented in advance and can be used directly for the estimation process. All such information will also be strongly linked with corresponding quality parameters (e.g. full variance-covariance matrix (VCM) of the pose estimated).

The paper is structured as follows. In the next Section, related work of current georeferencing methods and its suit-

This work was funded by the German Research Foundation (DFG) as part of the Research Training Group i.c.sens [RTG 2159] and NE 1453/5-1.

ability for the challenging environments mentioned is given. The IEKF used and extended is introduced mathematically in Section III. Section IV highlights a specific application of the theoretical approach by means of a simulated data set. The paper ends in Section V with conclusions and an outlook.

II. RELATED WORK

An overview of the basic types of georeferencing with kinematic MSS within indoor environments is given in [1]. Mainly GNSS observation data must be added to that compilation only in case of outdoor applications. Other sensors for outdoor georeferencing (e.g. magnetometer, radar or ultrasound) are not applicable due to demands on highly accurate pose parameters. Nevertheless, position information based on GNSS observations (even in combination with additional precise correction data from a reference station) is unreliable, especially in the challenging inner-city environments mentioned. Therefore, they are unsuitable for accurate georeferencing. Consequently, object space information should be used to obtain an optimal gain in accuracy and reliability.

Object space information is frequently used to enhance the georeferencing of a kinematic MSS. All simultaneous localization and mapping (SLAM) approaches, for example, are based on such ideas [3]. A particular subcategory focused on line or surface features as constrained SLAM also exists. Moreover, existing approaches distinguish between line features within a 2D laser scan or whole surfaces in 3D point clouds. It is obvious that additional information from the environment must be used especially for more reliable indoor mapping and localization applications. These include primarily available maps, previous measurements or common assumptions such as parallel or perpendicular main structures [4].

Geometric constraints (e.g. walls are detected as line segments and should be straight, perpendicular or parallel to each other) for a SLAM approach are used in [5]. [6] also used such constraints for an orthogonal SLAM approach by only allowing parallel or perpendicular lines to be mapped. In addition, weights with respect to the particular lengths of the line segment are used in this approach and define reference alignment. A more efficient approach was carried out by [7]. They used a Rao-Blackwellized particle filter SLAM approach in combination with geometric constraints to provide an efficient solution by ignoring non-perpendicular line segments. Soft constraints within the framework of a SLAM approach are used in [8] within a camera-based system. Here, they used point and line features and their relationship to each other.

A broad overview of possible 6-DOF SLAM approaches is given by [9]. Here, they just mentioned the 6-DOF pose representation by a 2D laser scanner as '*slice-wise 6D SLAM*'. The use of this method is depicted in [10] to build a 3D point cloud on the basis of profile laser scanners. However, they neither estimate the pose of the robot by means of the laser scanner observations, nor do they use information from object space by geometrical constraints. A spring-mounted combination of a low-cost 2D laser scanner with an IMU

is introduced in [11]. Within their online SLAM approach, they estimated 6-DOF pose parameters by a repetitive iterative closest point algorithm for time-windowed segments of the trajectory. However, they did not consider geometrical constraints for improving the pose parameters. [12] estimated 6-DOF pose information, but used three rigidly aligned 2D laser scanners. Therefore, the range measurements regarding planes of walls, ceilings and floors are used in a simulation environment. However, he performed a post-processing adjustment and only considered geometrical constraints for initialization. Nevertheless, he worked independent of IMU information.

As far as we know, the quality parameters have not been implied during pose parameter estimation in all SLAM approaches mentioned. This drawback will be eliminated in this paper. Strict consideration of uncertainties regarding input and output variables in the course of estimation is very important to ensure the integrity of the MSS. Additionally, the risk exists that individual pose parameters are not or (only) are inaccurately assignable (e.g. roll angle within a consistent straight long corridor) in case of adverse environments. Additionally, these approaches used either 3D laser scanners or the scanning plane of the profile laser scanner used is aligned horizontally. Both are in contrast to our approach by improving full 6-DOF pose parameters by means of vertically aligned laser scanner profiles. The idea of using such line segments within the object space observed by a high accurate laser scanner in a vertical profile mode is described by [13]. However, this additional information is only used to improve the roll angle of a MSS.

It is mentioned in [14], that they can improve performance for reliable, more accurate and robust localization by combination with a SLAM approach with an available digital road map, as some kind of prior approximate knowledge.

Reference [15] used laser scanner observations for improving the position and orientation information of a vehicle in the field of driver assistance systems. Therefore, he estimated the proper motion of the vehicle (based on environment sensors like an odometer) and supported this solution by means of highly accurate and very detailed digital maps of the environment as well as different kinds of SLAM approaches. He gives an overview of different scan matching approaches and developed a hybrid SLAM approach which uses grid- and feature-based concepts depending on the environment.

In a similar way, [16] developed a self-localization approach with 6-DOF. They used laser scanner data in combination with a feature-based approach using straight line or plane features based on an extended Kalman filter (EKF). The algorithms enable the possibility of localization by using sidewalls, fences or the road surface as additional features especially in challenging situations without structures in the environment (e.g. in case of a straight tunnel). However, they used point clouds from a 3D scanner for their approach. Geometrical constraints such as parallelism or perpendicularity are also not used. Nevertheless, they considered measurement and related position uncertainty and compared their approach against odometry-only solutions. The feature-based only solution has to deal with a lack of features in the direction of motion (inside a tunnel), but they

showed the benefit if they combine odometry and their method.

Using directly measured laser scanner point clouds to improve the trajectory of a mobile mapping system is given in [17]. By doing this, he compensated for the imprecise initial solution of a GNSS/IMU combination by the alignment of scanned points by deforming trajectories. However, repeated acquisition of the same environment is needed.

In [18], they used linear vertical, horizontal and right-angled structures by means of man-made objects (e.g. building facades) as soft constraints. With those geometrical restrictions, they could reduce the number of ground control points needed for the bundle adjustment of airborne images significantly. Due to usage of this total bundle adjustment, there is no option for online compatibility.

Closing all these gaps (e.g. neglect quality parameters, necessity for 3D laser scanners, need for a map in advance) is the intention of this paper.

III. METHODOLOGY

The approach for information-based georeferencing with an MSS introduced here is based on an IEKF which was initially published by [19]. Using this IEKF, it is possible to handle nonlinear measurement equations by performing Taylor series expansion (first- or second-order) for linearization iteratively. This well-known IEKF was generally developed for explicit measurement equations of type $\mathbf{l} = \mathbf{h}(\mathbf{x})$, also known as Gauss Markov models (GMM), where the observations $\mathbf{l}$ (with $\dim(\mathbf{l}) = i \times 1$ and i represents the number of observations) are separated from a nonlinear function $\mathbf{h}$ of the desired parameters $\mathbf{x}$ (with $\dim(\mathbf{x}) = j \times 1$ and j represents the number of parameters). However, as we will see in Section IV, we are going to deal with implicit measurement equations of type $\mathbf{h}(\mathbf{l}, \mathbf{x}) = 0$ (Gauss Helmert model (GHM)), where it is not possible to separate observations from parameters. Thus, we make use of an IEKF which deals with implicit measurement equations introduced by [20] and [21]. We will call this filter *IEKF-GHM* in the following due to the underlying model.

A. Introduction of the IEKF-GHM

Similar to the usual Kalman filters (KF), apart from the nonlinear measurement equation $\mathbf{h}(\cdot)$ in terms of the observation model, we also have a system model which represent the physical behaviour of the system by means of a (non)linear functional model $\mathbf{f}(\cdot)$ for epochs $k = 1, \dots, K$:

$$\mathbf{x}_k = \mathbf{f}(\mathbf{x}_{k-1}, \mathbf{u}_{k-1}, \mathbf{w}_{k-1}), \quad \mathbf{w}_{k-1} \sim N(\mathbf{0}, \Sigma_{ww}) \quad (1)$$

$$\hat{\mathbf{l}}_k = \mathbf{l}_k + \mathbf{v}_k, \quad \mathbf{h}(\mathbf{l}_k, \mathbf{x}_k) = \mathbf{0}, \quad \mathbf{v}_k \sim N(\mathbf{0}, \Sigma_{vv}) \quad (2)$$

where $\mathbf{u}$ is the deterministic control by means of external controls, $\mathbf{w}$ is the normally distributed system noise according to VCM of the system noise Σ_{ww} and $\mathbf{v}$ is the normally distributed measurement noise according to their VCM Σ_{vv} at epoch k . The estimation of the IEKF-GHM is divided into a *prediction* step, by only using the system model, and into an *update* step, by making use of the observation model.

Within the *prediction*, the state transition from the previous epoch $k - 1$ to the current epoch k is estimated. Current

or previous observations are not taken into account within this step. Underlying model is the carefully selected system model by means of the (non)linear functional model $\mathbf{f}(\cdot)$. The Jacobian matrices $\mathbf{F}_{x,k}$ and $\mathbf{F}_{w,k}$ are the results of the first-order Taylor series expansion of the functional model $\mathbf{f}(\cdot)$ regarding the previous filtered state vector $\hat{\mathbf{x}}_{k-1}^+$, control $\mathbf{u}_{k-1}$ and system noise vector $\mathbf{w}_{k-1}$ from epoch $k - 1$. The predicted state vector $\hat{\mathbf{x}}_k^-$ and, by weighting with their respective VCM, its related VCM $\Sigma_{xx,k}^-$ can be estimated:

$$\mathbf{F}_{x,k} = \left. \partial \mathbf{f} / \partial \mathbf{x} \right|_{\hat{\mathbf{x}}_{k-1}^+, \mathbf{u}_{k-1}, \mathbf{w}_{k-1}} \quad (3)$$

$$\mathbf{F}_{w,k} = \left. \partial \mathbf{f} / \partial \mathbf{w} \right|_{\hat{\mathbf{x}}_{k-1}^+, \mathbf{u}_{k-1}, \mathbf{w}_{k-1}} \quad (4)$$

$$\hat{\mathbf{x}}_k^- = \mathbf{f}(\hat{\mathbf{x}}_{k-1}^+, \mathbf{u}_{k-1}, \mathbf{w}_{k-1}) \quad (5)$$

$$\Sigma_{xx,k}^- = \mathbf{F}_{x,k} \Sigma_{xx,k-1}^+ \mathbf{F}_{x,k}^T + \mathbf{F}_{w,k} \Sigma_{ww} \mathbf{F}_{w,k}^T \quad (6)$$

Due to nonlinearity within the observation model, we have to estimate the Jacobian matrix $\mathbf{H}_{x,k}$ within the *update* step regarding the state vector and $\mathbf{H}_{l,k}$ regarding the observations. Both are computable by means of the measurement equation $\mathbf{h}(\cdot)$ concerning their related derivations. We have to perform an iterated optimization problem because of needed initial values and highly nonlinear functions needed. Thus, we have to consider a loop $m = 0, \dots, M - 1$ inside the *update* step until a predefined threshold is exceeded or maximum iteration number is reached. The parameters iterated and observations are marked with a caron symbol, like $\check{\mathbf{l}}_{k,m}$ and $\check{\mathbf{x}}_{k,m}$:

$$\mathbf{H}_{x,k,m} = \left. \partial \mathbf{h} / \partial \mathbf{x} \right|_{\check{\mathbf{l}}_{k,m}, \check{\mathbf{x}}_{k,m}} \quad (7)$$

$$\mathbf{H}_{l,k,m} = \left. \partial \mathbf{h} / \partial \mathbf{l} \right|_{\check{\mathbf{l}}_{k,m}, \check{\mathbf{x}}_{k,m}} \quad (8)$$

Because of the complexity in the later measurement equation (cf. Section IV-B), we are performing numerical differentiation by using INTerval LABoratory (INTLAB) Toolbox [22] for MATLAB [23]. The iterated state vector $\check{\mathbf{x}}_{k,m+1}$ depends on the Kalman matrix $\mathbf{K}_{k,m}$ and the innovation vector $\mathbf{r}_{k,m}$:

$$\mathbf{O}_{k,m} = \mathbf{H}_{x,k,m} \Sigma_{xx,k}^- \mathbf{H}_{x,k,m}^T \quad (9)$$

$$\mathbf{S}_{k,m} = \mathbf{H}_{l,k,m} \Sigma_{vv} \mathbf{H}_{l,k,m}^T \quad (10)$$

$$\mathbf{K}_{k,m} = \Sigma_{xx,k}^- \mathbf{H}_{x,k,m}^T [\mathbf{O}_{k,m} + \mathbf{S}_{k,m}]^{-1} \quad (11)$$

$$\mathbf{r}_{k,m} = \mathbf{H}_{l,k,m} \cdot (\mathbf{l}_k - \check{\mathbf{l}}_{k,m}) + \mathbf{H}_{x,k,m} \cdot (\hat{\mathbf{x}}_k^- - \check{\mathbf{x}}_{k,m}) \quad (12)$$

$$\check{\mathbf{x}}_{k,m+1} = \hat{\mathbf{x}}_k^- - \mathbf{K}_{k,m} \cdot (\mathbf{h}(\check{\mathbf{l}}_{k,m}, \check{\mathbf{x}}_{k,m}) + \mathbf{r}_{k,m}) \quad (13)$$

Afterwards, the iterated observation vector $\check{\mathbf{l}}_{k,m+1}$ can be estimated:

$$\mathbf{G}_{k,m} = \Sigma_{vv} \mathbf{H}_{l,k,m}^T (\mathbf{O}_{k,m} + \mathbf{S}_{k,m})^{-1} \quad (14)$$

$$\check{\mathbf{l}}_{k,m+1} = \mathbf{l}_k - \mathbf{G}_{k,m} \cdot (\mathbf{h}(\check{\mathbf{l}}_{k,m}, \check{\mathbf{x}}_{k,m}) + \mathbf{r}_{k,m}) \quad (15)$$

Together with the iterated state vector $\check{\mathbf{x}}_{k,m+1}$ and iterated observation vector $\check{\mathbf{l}}_{k,m+1}$, the iteration starts again until a predefined threshold is exceeded or maximum iteration number ($m = M - 1$) is reached.

Afterwards, we finally obtained the filtered observation vector $\hat{l}_k^+$ and the filtered state vector $\hat{x}_k^+$ as well as their corresponding VCM $\Sigma_{\hat{x}\hat{x},k}^+$:

$$\hat{l}_k^+ = \check{l}_{k,M} \quad (16)$$

$$\hat{x}_k^+ = \check{x}_{k,M} \quad (17)$$

$$L_k = \mathbf{I}_{j \times j} - K_{k,M-1} H_{x,k,M-1} \quad (18)$$

$$\Sigma_{\hat{x}\hat{x},k}^+ = L_k \Sigma_{xx,k}^- L_k^T + K_{k,M-1} \cdot S_{k,M-1} \cdot K_{k,M-1}^T \quad (19)$$

In addition to the IEKF approach from [20] and [21], we are performing error propagation to get the VCM of the filtered observations $\Sigma_{\hat{v}\hat{v},k}^+$ within every epoch k :

$$\hat{l}_{k,m}^* = -H_{l,k,m} l_k - h(\check{l}_{k,m}, \check{x}_{k,m}) \quad (20)$$

$$\check{l}_{k,m+1} = l_k + G_{k,m} \cdot (\hat{l}_{k,m}^* - H_{x,k,m} \hat{x}_{k,m}^-) \quad (21)$$

$$U_k = G_{k,M-1} \cdot H_{x,k,M-1} \quad (22)$$

$$\Sigma_{\hat{v}\hat{v},k}^+ = \Sigma_{vv} + G_{k,M-1} S_{k,M-1} G_{k,M-1}^T - U_k \Sigma_{xx,k}^- U_k^T \quad (23)$$

B. Extending the IEKF-GHM by nonlinear equality constraints

In order to improve the estimates of the IEKF-GHM (non)linear equality constraints of the state parameters can be included. The projection algorithm, given in [24] and [25], is used for inserting linear equality constraints, which can be formulated as:

$$Dx_k \stackrel{!}{=} d \quad (24)$$

where D is a known and constant matrix with $\dim(D) = c \times j$, d is a known vector with $\dim(d) = c \times 1$ and c is the number of constraints. In the case of nonlinear equality constraints of the state parameters:

$$g(x_k) \stackrel{!}{=} b \quad (25)$$

we perform Taylor series expansion for linearization:

$$g'(\hat{x}_k^-) \cdot x_k \stackrel{!}{=} b - g(\hat{x}_k^-) + g'(\hat{x}_k^-) \cdot \hat{x}_k^- \quad (26)$$

Here, we neglected the second-order term, as suggested in [24] and [25]:

$$D = g'(\hat{x}_k^-) \quad (27)$$

$$d = b - g(\hat{x}_k^-) + g'(\hat{x}_k^-) \cdot \hat{x}_k^- \quad (28)$$

We have the linearized form according to (24) and can estimate the constrained state parameter vector $\hat{x}_k^+$:

$$\hat{x}_k^+ = \hat{x}_k^+ - W^{-1} D^T (D W^{-1} D^T)^{-1} (D \hat{x}_k^+ - d) \quad (29)$$

$$\Sigma_{\hat{x}\hat{x},k}^+ = \Sigma_{\hat{x}\hat{x},k}^+ - \Sigma_{\hat{x}\hat{x},k}^+ D^T (D \Sigma_{\hat{x}\hat{x},k}^+ D^T)^{-1} D \Sigma_{\hat{x}\hat{x},k}^+ \quad (30)$$

where the weight matrix W can be chosen by $W = \mathbf{I}_{j \times j}$ which will result in a constrained solution closer to the true state than an unconstrained estimation [25].

C. Final IEKF-GHM with nonlinear equality constraints

The final algorithm of the IEKF-GHM is summarized within algorithm 1. It gives an overview of the initialization, the predicted state parameters, the updated state parameters and the observations, as well as the nonlinear equality constraints finally applied.

Algorithm 1: IEKF-GHM for observation models with nonlinear implicit measurement equation and nonlinear equality constraints

```

1 System model  $x_k = f(x_{k-1}, u_{k-1}, w_{k-1})$ ,  $w_{k-1} \sim N(0, \Sigma_{ww})$ 
2 Observation model  $l_k = l_k + v_k$ ,  $h(l_k, x_k) = 0$ ,  $v_k \sim N(0, \Sigma_{vv})$ 
3 Initial parameter vector and its VCM:
    $\hat{x}_0^+ = x_0$ ,  $\Sigma_{\hat{x}\hat{x},0}^+ = \Sigma_{xx,0}$ ,  $k = 1$ 
4 while  $k < K$  do
5   Prediction
6    $F_{x,k} = \partial f / \partial x|_{\hat{x}_{k-1}^+, u_{k-1}, w_{k-1}}$ 
7    $F_{w,k} = \partial f / \partial w|_{\hat{x}_{k-1}^+, u_{k-1}, w_{k-1}}$ 
8    $\hat{x}_k^- = f(\hat{x}_{k-1}^+, u_{k-1}, w_{k-1})$ 
9    $\Sigma_{xx,k}^- = F_{x,k} \Sigma_{xx,k-1}^+ F_{x,k}^T + F_{w,k} \Sigma_{ww} F_{w,k}^T$ 
10  Update
11   $\check{l}_{k,0} = l_k$ ,  $\check{x}_{k,0} = \hat{x}_k^-$ 
12  for  $m = 0 \dots M-1$  do
13     $H_{l,k,m} = \partial h / \partial l|_{\check{l}_{k,m}, \check{x}_{k,m}}$ ,  $H_{l,k,m} = \partial h / \partial l|_{\check{l}_{k,m}, \check{x}_{k,m}}$ 
14     $O_{k,m} = H_{x,k,m} \Sigma_{xx,k}^- H_{x,k,m}^T$ 
15     $S_{k,m} = H_{l,k,m} \Sigma_{vv} H_{l,k,m}^T$ 
16     $K_{k,m} = \Sigma_{xx,k}^- H_{x,k,m}^T (O_{k,m} + S_{k,m})^{-1}$ 
17     $r_{k,m} = H_{l,k,m} \cdot (\check{l}_k - \check{l}_{k,m}) + H_{x,k,m} \cdot (\hat{x}_k^- - \check{x}_{k,m})$ 
18     $\check{x}_{k,m+1} = \check{x}_k^- - K_{k,m} \cdot (h(\check{l}_{k,m}, \check{x}_{k,m}) + r_{k,m})$ 
19     $G_{k,m} = \Sigma_{vv} H_{l,k,m}^T (O_{k,m} + S_{k,m})^{-1}$ 
20     $\check{l}_{k,m+1} = l_k - G_{k,m} \cdot (h(\check{l}_{k,m}, \check{x}_{k,m}) + r_{k,m})$ 
21   $\hat{x}_k^+ = \check{x}_{k,M}$ 
22   $\check{l}_k^+ = \check{l}_{k,M}$ 
23   $L_k = \mathbf{I}_{j \times j} - K_{k,M-1} H_{x,k,M-1}$ 
24   $\Sigma_{\hat{x}\hat{x},k}^+ = L_k \Sigma_{xx,k}^- L_k^T + K_{k,M-1} \cdot S_{k,M-1} \cdot K_{k,M-1}^T$ 
25   $U_k = G_{k,M-1} \cdot H_{x,k,M-1}$ 
26   $\Sigma_{\hat{v}\hat{v},k}^+ = \Sigma_{vv} + G_{k,M-1} S_{k,M-1} G_{k,M-1}^T - U_k \Sigma_{xx,k}^- U_k^T$ 
27  Applying nonlinear equality constraints
28   $D = g'(\hat{x}_k^-)$ 
29   $d = b - g(\hat{x}_k^-) + g'(\hat{x}_k^-) \cdot \hat{x}_k^-$ 
30  Set  $W = \mathbf{I}_{j \times j}$ 
31   $\hat{x}_k^+ = \hat{x}_k^+ - W^{-1} D^T (D W^{-1} D^T)^{-1} (D \hat{x}_k^+ - d)$ 
32   $\Sigma_{\hat{x}\hat{x},k}^+ = \Sigma_{\hat{x}\hat{x},k}^+ - \Sigma_{\hat{x}\hat{x},k}^+ D^T (D \Sigma_{\hat{x}\hat{x},k}^+ D^T)^{-1} D \Sigma_{\hat{x}\hat{x},k}^+$ 
33  Set  $\hat{x}_k^+ = \hat{x}_k^+$  and  $\Sigma_{\hat{x}\hat{x},k}^+ = \Sigma_{\hat{x}\hat{x},k}^+$ 

```

IV. APPLICATION

We simulate a realistic LSP within a simplified 3D environment of a corridor by various linear and nonlinear trajectories to verify that the algorithm developed in Section III works properly. Pose observations of an IMU are generated based on ground truth information from the simulation. Therefore, synthetic additional values for drift and noise effects are added by means of typical specifications of a low-cost IMU. Therefore, we are assuming to get the position and orientation information

directly from the IMU instead of original acceleration and angular rates.

For reasons of simplification, by performing the IEKF-GHM from Section III, we will set the transition matrix $F_{x,k}$ as identity matrix $\mathbf{I}_{j \times j}$, i.e. we assume a static physical model for the system model given in (2) from epoch $k-1$ to epoch k . Furthermore, the VCM of the system noise Σ_{ww} will be equal to zero matrix and the control u will also be zero. Thus, within each *prediction* step, the predicted state vector $\hat{\mathbf{x}}_k^-$ and its VCM $\Sigma_{xx,k}^-$ will be equal to filtered state vector $\hat{\mathbf{x}}_{k-1}^+$ or rather filtered VCM $\Sigma_{xx,k-1}^+$ from epoch $k-1$. Due to consideration of an implicit measurement equation and neglecting the *prediction*, we are performing a recursive GHM.

A. Overview of the simulated data set

It will be assumed that necessary geometrical calibration parameters between the reference point of the IMU and laser scanner origin are already attached and, hence, all the information is referred to one common reference point. Therefore, we choose the origin of the laser scanner itself (cf. Fig 1). Under the simplifying assumption that all measurements are synchronized and available with an identical sampling rate, we get at each epoch k the 3D position $\mathbf{t}_k$ and 3D rotation matrix $\mathbf{R}_k$ of the IMU information as well as N (number of points within current LSP) single 3D scan points within each LSP $\mathbf{P}_k^{\text{local}}$ in their local sensor coordinate system:

$$\mathbf{t}_k = [X_k, Y_k, Z_k]^T, \dim(\mathbf{R}_k) = 3 \times 3 \quad (31)$$

$$\mathbf{P}_k^{\text{local}} = [x_{1,k}, y_{1,k}, z_{1,k}, \dots, x_{N,k}, y_{N,k}, z_{N,k}]^T \quad (32)$$

Note that the 3D rotation matrix $\mathbf{R}_k$ will be transformed into Euler angles $\mathbf{R}_k(\Omega_k, \Phi_k, K_k)$. On this basis, the observation vector $\mathbf{l}_k$ can be defined:

$$\mathbf{l}_k = [\mathbf{P}_k^{\text{local}}, X_k, Y_k, Z_k, \Omega_k, \Phi_k, K_k]^T \quad (33)$$

Because the laser scanner is capturing in 2D mode ($x_{1:N,k}, y_{1:N,k}$), we add the z-component of an expected value of zero and a typical noise of this type of laser scanner.

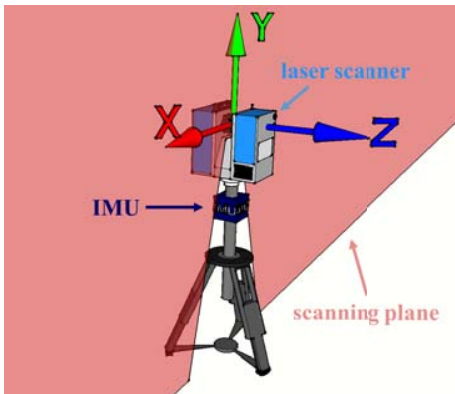


Fig. 1. Theoretical MSS with a highly accurate laser scanner in profile mode (top) and IMU (middle) as well as related superordinated coordinate system in the origin of the laser scanner. The scanning plane is depicted in red.

Secondly, and strongly related to the observation data, are the corresponding VCM of the observations Σ_{vv} . These consists of the quality information of the IMU composed in $\Sigma_{vv\text{IMU}}$ as well as the quality information of LSP $\Sigma_{vv\text{LSP}}$ (cf. (36)). Note that variance information for the 3D rotation matrix $\mathbf{R}_k$ is also given regarding transformed Euler angles Ω_k, Φ_k and K_k :

$$\Sigma_{vv\text{IMU}} = \text{diag}(\sigma_X^2, \sigma_Y^2, \sigma_Z^2, \sigma_{\Omega}^2, \sigma_{\Phi}^2, \sigma_K^2) \quad (34)$$

$$\Sigma_{vv\text{LSP}} = \text{diag}(\sigma_{x_1}^2, \sigma_{y_1}^2, \sigma_{z_1}^2, \dots, \sigma_{x_N}^2, \sigma_{y_N}^2, \sigma_{z_N}^2) \quad (35)$$

$$\Sigma_{vv} = \begin{bmatrix} \Sigma_{vv\text{LSP}} & \mathbf{0} \\ \mathbf{0} & \Sigma_{vv\text{IMU}} \end{bmatrix} \quad (36)$$

Regarding the simulated trajectory, it should be noted that we assume almost error-free observations within the direction of movement. This is certainly not true in reality, but we assume this to show our theoretical approach.

B. Nonlinear implicit measurement equation

Together with the available IMU pose information $\mathbf{t}_k$ and $\mathbf{R}_k$, basic referencing of every LSP is possible by the well-known Helmert transformation (excluding a scaling factor):

$$\mathbf{P}_k^{\text{IMU}} = \mathbf{t}_k + \mathbf{R}_k \cdot \mathbf{P}_k^{\text{local}} \quad (37)$$

Due to the fact of commonly noisy and drifted IMU observations, from a global perspective, the transformed and so-called referenced point cloud $\mathbf{P}_k^{\text{IMU}}$ will be strongly inaccurate in comparison to reality. Only the inner geometry of neighbored LSP are precise to each other. This temporal correlation depends strongly on the quality of the IMU itself. However, in principle, the translation $\mathbf{t}_k$ and rotation $\mathbf{R}_k$ need to be corrected by some additional terms $\Delta\mathbf{t}_k$ and $\Delta\mathbf{R}_k$ in the sense that the global point cloud $\mathbf{P}_k^{\text{global}}$ corresponds to the real environment. These are generally also the desired parameters, named as position $\mathbf{t}_k^{\text{MSS}}$ and orientation by means of rotation matrix $\mathbf{R}_k^{\text{MSS}}$ of the MSS:

$$\mathbf{t}_k^{\text{MSS}} = \mathbf{t}_k + \Delta\mathbf{t}_k \quad (38)$$

$$\mathbf{R}_k^{\text{MSS}} = \Delta\mathbf{R}_k \cdot \mathbf{R}_k \quad (39)$$

After this correction, the corrected LSP $\mathbf{P}_k^{\text{global}}$ can be estimated by updated Helmert transformation:

$$\mathbf{P}_k^{\text{global}} = \mathbf{t}_k^{\text{MSS}} + \mathbf{R}_k^{\text{MSS}} \cdot \mathbf{P}_k^{\text{local}} \quad (40)$$

However, in our scenario, these pose parameters of the MSS $\mathbf{t}_k^{\text{MSS}}$ and $\mathbf{R}_k^{\text{MSS}}$ are unknown and not directly observable. But instead, we can make use of the additional terms $\Delta\mathbf{t}_k$ and $\Delta\mathbf{R}_k$ and choose them as our requested parameters:

$$\mathbf{P}_k^{\text{global}} = (\mathbf{t}_k + \Delta\mathbf{t}_k) + (\Delta\mathbf{R}_k \cdot \mathbf{R}_k) \cdot \mathbf{P}_k^{\text{local}} \quad (41)$$

Finally, we can set up the desired state parameter vector $\mathbf{x}'_k$ with $\dim(\mathbf{x}'_k) = 6 \times 1$ according to 3 translations $\Delta\mathbf{t}_k = [\Delta X_k, \Delta Y_k, \Delta Z_k]^T$ and 3 Euler angles $\Delta\mathbf{R}_k(\Delta\Omega_k, \Delta\Phi_k, \Delta K_k)$ regarding their 3D rotation matrix:

$$\mathbf{x}'_k = [\Delta X_k, \Delta Y_k, \Delta Z_k, \Delta\Omega_k, \Delta\Phi_k, \Delta K_k]^T \quad (42)$$

We use all the available information from object space which was already observed by the laser scanner to estimate the values requested in (42). In combination with external prior information (e.g. documented by standards for building industry), constraints can be formulated and integrated within our approach. Therefore, we use already segmented LSP information $\mathbf{C}_k^{\text{local}}$ regarding left and right wall, ceiling and floor of the environment. Thus, in total, $\mathbf{C}_k^{\text{local}}$ is equal to $\mathbf{P}_k^{\text{local}}$ but contains additional segmentation information for every measured scan point. This can be done within a real data set by known segmentation algorithms from the literature, see, for example, [26] and [27].

If we use this additional information, we can formulate the nonlinear implicit measurement equation based on well-known Hesse normal form of a plane:

$$\mathbf{n} \cdot [(\mathbf{t}_k + \Delta \mathbf{t}_k) + (\Delta \mathbf{R}_k \cdot \mathbf{R}_k) \cdot \mathbf{C}_k^{\text{local}}] - d = 0 \quad (43)$$

where $\mathbf{n}$ is the 3×1 normal vector of the left wall (or rather right wall, ceiling, floor) and d the related distance to the origin. Hence, we have to extend our state parameter vector $\mathbf{x}'_k$ by additional state parameters $\mathbf{x}''_k$. This vector consists of four sets of plane parameters with each four parameters $n_{e_x}, n_{e_y}, n_{e_z}, d_e$. Here, e can stand for the left wall (or rather right wall, ceiling or floor):

$$\mathbf{x}''_k = \left[\underbrace{n_{\zeta_x}, n_{\zeta_y}, n_{\zeta_z}, d_{\zeta}}_{\text{left wall}}, \dots, \underbrace{n_{\xi_x}, n_{\xi_y}, n_{\xi_z}, d_{\xi}}_{\text{floor}} \right]^T \quad (44)$$

Finally, the extended state parameter vector $\mathbf{x}_k$ is given by:

$$\mathbf{x}_k = [\mathbf{x}'_k, \mathbf{x}''_k]^T \quad (45)$$

C. Nonlinear equality constraints for the state parameters

We introduce nonlinear equality constraints of the state parameters by means of geometrical restrictions according to Section III-B to improve the filtered estimates from IEKF-GHM. Regarding our case of application, we chose concurrency between left and right wall as well as between ceiling and floor. The third and fourth constraint represents perpendicularity between the left wall and ceiling as well as between the right wall and floor. These geometrical constraints have to be applied within every epoch and affect the plane parameters of those four planes. These nonlinear constraints are based on the estimation of the angle between to planes:

$$\cos(\alpha) = \frac{|\mathbf{n}_q \cdot \mathbf{n}_s|}{|\mathbf{n}_q| \cdot |\mathbf{n}_s|} \quad (46)$$

where $\mathbf{n}_q$ and $\mathbf{n}_s$ are the parameters of two not identical planes within the Hesse normal form. Applying this will lead to the nonlinear function $g(\mathbf{x}_k)$ with constrained values $\mathbf{b}$:

$$g(\mathbf{x}_k) = \frac{|n_{q_x} n_{s_x} + n_{q_y} n_{s_y} + n_{q_z} n_{s_z}|}{\sqrt{n_{q_x}^2 + n_{q_y}^2 + n_{q_z}^2} \cdot \sqrt{n_{s_x}^2 + n_{s_y}^2 + n_{s_z}^2}} \stackrel{!}{=} \mathbf{b} \quad (47)$$

$$\mathbf{b} = [1, 1, 0, 0]^T \quad (48)$$

where the vector $\mathbf{b}$ is chosen according to the four constraints of concurrency and perpendicularity mentioned. After performing the IEKF-GHM with nonlinear equality constraints according to algorithm (1), we also have to apply (38) and (39) to get the desired pose parameters $\mathbf{t}_k^{\text{MSS}}$ and $\mathbf{R}_k^{\text{MSS}}$.

D. Performing error propagation for getting $\Sigma_{\hat{\mathbf{x}},k}^{+, \text{MSS}}$

Because we intend to estimate the pose parameters $\mathbf{t}_k^{\text{MSS}}$ and $\mathbf{R}_k^{\text{MSS}}$, we are also interested in estimating their VCM. Instead, we estimate the additional terms $\Delta \mathbf{t}_k$ and $\Delta \mathbf{R}_k$ — in combination with their respective VCM $\Sigma_{\hat{\mathbf{x}},k}^{+, \text{MSS}}$ — which have to be applied to noisy and drifted IMU solution $\mathbf{t}_k$ and $\mathbf{R}_k$ according to (38) and (39).

We have to perform error propagation to get the VCM of the requested pose parameters $\Sigma_{\hat{\mathbf{x}},k}^{+, \text{MSS}}$:

$$\Sigma_{\hat{\mathbf{x}},k}^{+, \text{MSS}} = \mathbf{J} \begin{bmatrix} \Sigma_{\hat{\mathbf{x}},k}^{+, \text{MSS}} & \mathbf{0} \\ \mathbf{0} & \Sigma_{\hat{\mathbf{v}},k}^{+, \text{MSS}} \end{bmatrix} \mathbf{J}^T \quad (49)$$

where $\mathbf{J}$ is the Jacobian matrix of (38) and (39) regarding filtered state parameters $\hat{\mathbf{x}}_k^+$ and filtered observations $\hat{\mathbf{l}}_k^+$. It should be noted that the values of $\Sigma_{\hat{\mathbf{x}},k}^{+, \text{MSS}}$ depend highly on the initial VCM of the IMU observation $\Sigma_{\hat{\mathbf{v}},k}^{+, \text{MSS}}$ which are passed through the IEKF-GHM into $\Sigma_{\hat{\mathbf{v}},k}^{+, \text{MSS}}$.

E. Results

We generated 500 epochs of the simulated data from Section IV-A to test our approach. The total distance in the direction of motion (Z component) is about 5 m. As corresponding VCM of the observations $\Sigma_{\hat{\mathbf{v}}}$ we chose realistic values by $\sigma_x = \sigma_y = \sigma_z = 3 \text{ mm}$ standard deviation for the scan points of the LSP measured and $\sigma_x = \sigma_y = 100 \text{ mm}$ for first two components of the IMU observations. As has already been mentioned, the Z component will be accurate by $\sigma_z = 0,01 \text{ mm}$ and the Euler angles will be less accurate with $\sigma_\Omega = \sigma_\Phi = \sigma_K = 0,1^\circ$.

The nonlinear change in the pose is depicted in Fig. 2. It can be seen that there is a linear trend in the x component and a quadratic trend in the y component of the position. Both are overlapped with random noise. The biggest amount of linear

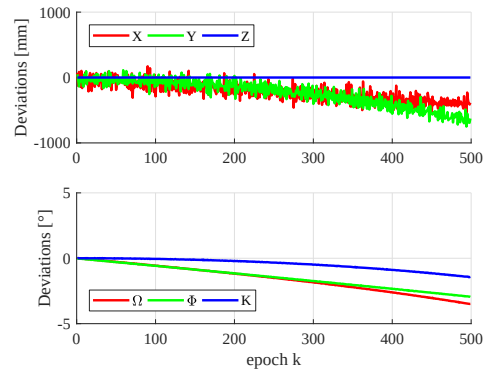


Fig. 2. Differences between noisy and drifted position (X (red), Y (green) and Z (blue), top) and orientation (Ω (red), Φ (green) and K (blue), bottom) of the IMU regarding ground truth of the MSS.

change in rotation is given by the heading component Φ with $\sim 30^\circ$ within these 500 epochs. The deviations of all Euler angles between IMU and ground truth is depicted in the lower graph of Fig. 2.

The estimated pose parameters by means of performing the IEKF-GHM considering the nonlinear inequality constraints mentioned from Section IV-C are presented in Fig. 3. Within position, the deviations regarding ground truth increase linearly until 5 mm (x component) and 7 mm (y component). Deviations in orientations are higher in the initial phase and then around 0° (Euler angle Ω) or rather decreasing quadratically towards 0.5° (Euler angle Φ). Due to simplification mentioned regarding drift, noise and change in movement within the z component, the deviations in position are a constant zero. However, the Euler angle K — which represents the roll angle — is determinable precisely.

By contrast, the estimated pose parameters by neglect of the nonlinear inequality constraints mentioned from Section IV-C are depicted in Fig. 4. It is obvious that the approach will fail if those additional constraints are not considered. However, the estimates are still better than pose information only based on the IMU observations.

In the following, the respective standard deviations of the pose parameters are depicted in Fig. 5 by means of the VCM $\Sigma_{\hat{x}\hat{x},k}^+$ according to directly estimated parameters of the IEKF-GHM and VCM $\Sigma_{\hat{x}\hat{x},k}^{+,MSS}$ after performing error propagation (cf. Section IV-D). After performing the IEKF-GHM, the standard deviations σ_x and σ_y are equal up to a hundredth of a millimetre and decrease quite fast up to 8 mm. The standard deviations σ_Ω , σ_Φ and σ_K are also quite similar in comparison to each other and decrease up to 0.3° after the first epochs. However, after performing error propagation, the standard deviations σ_x and σ_y are still equal up to a hundredth of a millimetre but increase quite fast up to 140 mm. The same holds for the standard deviations σ_Ω , σ_Φ and σ_K , which also increase up to 8.1° (σ_Φ and σ_K) respectively 8.3° for σ_Ω quite fast after the first epochs. This significant

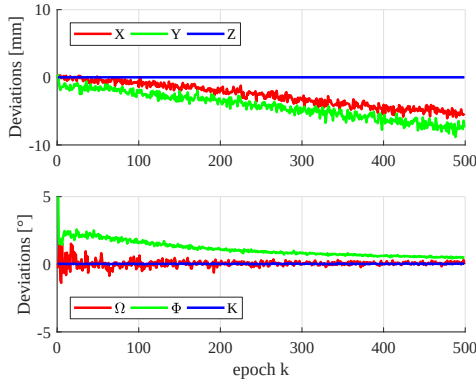


Fig. 3. Differences between estimated position (X (red), Y (green) and Z (blue), top) and orientation (Ω (red), Φ (green) and K (blue), bottom) of the IEKF-GHM with nonlinear equality constraints regarding ground truth of the MSS.

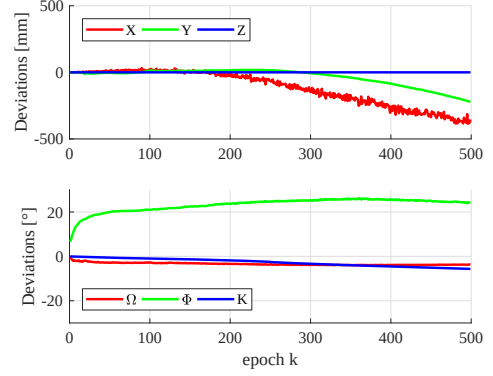


Fig. 4. Differences between estimated position (X (red), Y (green) and Z (blue), top) and orientation (Ω (red), Φ (green) and K (blue), bottom) of the IEKF-GHM without nonlinear equality constraints regarding ground truth of the MSS.

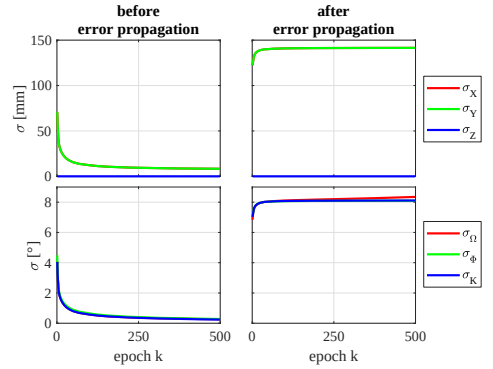


Fig. 5. Standard deviations according to VCM $\Sigma_{\hat{x}\hat{x},k}^+$ before error propagation (left) and VCM $\Sigma_{\hat{x}\hat{x},k}^{+,MSS}$ after error propagation (right) of the estimated position (X (red), Y (green) and Z (blue), top) and orientation (Ω (red), Φ (green) and K (blue), bottom) of the IEKF-GHM.

difference in standard deviations between VCM $\Sigma_{\hat{x}\hat{x},k}^+$ before error propagation and VCM $\Sigma_{\hat{x}\hat{x},k}^{+,MSS}$ after performing error propagation appears because of (38) and (39) mentioned. The originally inaccurate observations of the IMU are still inaccurate after estimating the filtered observations $\hat{l}_k^+$ and influence the propagated standard deviations significantly. This is because we only consider the inner geometry and, thus, only improve precision.

The root mean square error (RMSE) — divided into position and orientation — is depicted in Fig. 6. It is shown in particular that the RMSE of the position (black line) increases from 1 to 6 mm, whereas the RMSE of the orientation (magenta line) decreases from 115 up to 22 m° .

V. CONCLUSIONS AND OUTLOOK

The method presented enables new possibilities to ensure precise, accurate, reliable and complete georeferencing with kinematic MSS in challenging environments. This improvement has been achieved by using available information directly

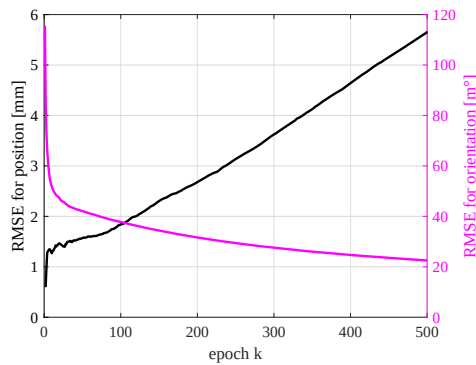


Fig. 6. Root mean square error (RMSE) for position (black) and orientation (magenta).

from the environment and integrating it by means of geometrical equality constraints into the estimation. Currently, we only considered simple environments to verify our approach. We have to extend the method in the future to deal with more complex scenarios.

We have to introduce absolute information (e.g. by means of horizontal floor and vertical walls) to get rid of just improving the inner geometry. Furthermore, we will consider drift and noise in the direction of movement. Therefore, we intend to include bias and drift parameters within the states of the IEKF-GHM used. It is also planned to take an suitable system model for the prediction of the IEKF-GHM into account to describe the motion of the system.

We intend to implement other filtering approaches in future research for solving implicit measurement equations by means of other nonlinear filters (e.g. unscented KF [28]) to avoid building Taylor expansion and for reducing the error resulting from the linearization of such equations. In addition, we are intending to expand our simplified system to include more applications regarding real environment in indoor and outdoor scenarios. In doing so, more complex environments with turnings around corners, protruding elements (e.g. doors, furniture) and changes of walls have to be considered. Under such circumstances, robustification of the estimation and soft constraints or nonlinear inequality constraints will also be of particular importance.

REFERENCES

- [1] S. Vogel, H. Alkhatib, and I. Neumann, "Accurate indoor georeferencing with kinematic multi sensor systems," in *2016 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, 2016, pp. 1–8.
- [2] C. Thomson, G. Apostolopoulos, D. Backes, and J. Böhm, "Mobile laser scanning for indoor modelling," *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. II-5/W2, pp. 289–293, 2013.
- [3] R. C. Smith and P. Cheeseman, "On the representation and estimation of spatial uncertainty," *The International Journal of Robotics Research*, vol. 5, no. 4, pp. 56–68, 1986.
- [4] J. Jung, S. Yoon, S. Ju, and J. Heo, "Development of kinematic 3d laser scanning system for indoor mapping and as-built bim using constrained slam," *Sensors*, vol. 15, no. 10, pp. 26 430–26 456, 2015.
- [5] G. Zunino, *Simultaneous localization and mapping for navigation in realistic environments: Dissertation*, Stockholm, Kungl Tekniska Högskolan (KTH), 2002, vol. 0203.
- [6] V. Nguyen, A. Harati, A. Martinelli, R. Siegwart, and N. Tomatis, "Orthogonal slam: A step toward lightweight indoor autonomous navigation," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2006, pp. 5007–5012.
- [7] B.-W. Kuo, H.-H. Chang, Y.-C. Chen, and S.-Y. Huang, "A light-and-fast slam algorithm for robots in indoor environments using line segment map," *Journal of Robotics*, vol. 2011, no. 3, pp. 1–12, 2011.
- [8] H. Choi, R. Kim, and E. Kim, "An efficient ceiling-view slam using relational constraints between landmarks," *International Journal of Advanced Robotic Systems*, vol. 11, no. 1, pp. 1–11, 2014.
- [9] A. Nüchter, K. Lingemann, J. Hertzberg, and H. Surmann, "6d slam—3d mapping outdoor environments," *Journal of Field Robotics*, vol. 24, no. 8–9, pp. 699–722, 2007.
- [10] S. Thrun, M. Montemerlo, and A. Aron, "Probabilistic terrain analysis for high-speed desert driving," in *Robotics: Science and Systems II. Robotics: Science and Systems Foundation*, 2006, pp. 1–7.
- [11] M. Bosse, R. Zlot, and P. Flier, "Zebedee: Design of a spring-mounted 3-d range sensor with application to mobile mapping," *IEEE Transactions on Robotics*, vol. 28, no. 5, pp. 1104–1119, 2012.
- [12] G. Vosselman, "Design of an indoor mapping system using three 2d laser scanners and 6 dof slam," *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. II-3, pp. 173–179, 2014.
- [13] H. Vennegeerts, *Objektraumgestützte kinematische Georeferenzierung für Mobile-Mapping-Systeme: Dissertation*, ser. Reihe C. München: DGK, 2011, vol. 657.
- [14] K. W. Lee, S. Wijesoma, and J. I. Guzmán, "A constrained slam approach to robust and accurate localisation of autonomous ground vehicles," *Robotics and Autonomous Systems*, vol. 55, no. 7, pp. 527–540, 2007.
- [15] T.-T. Weiss, *Hochgenaue Positionierung und Kartographie mit Laserscannern für Fahrerassistenzsysteme: Dissertation*, ser. Schriftenreihe des Instituts für Mess-, Regel- und Mikrotechnik. Ulm: Univ., [Fak. für Ingenieurwiss. und Informatik], 2011.
- [16] M. R. Schmid, M. Mählich, J. Dickmann, and H.-J. Wünsche, "Straight-feature-based self-localization for urban scenarios," in *Proceedings of 8th International Workshop on Intelligent, Hamburg, Germany*, 2011.
- [17] C. Brenner, "Scalable estimation of precision maps in a mapreduce framework," in *Proceedings of the 24th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems - GIS '16*, New York, NY, USA, 2016, pp. 1–10.
- [18] M. Gerke, "Using horizontal and vertical building structure to constrain indirect sensor orientation," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 66, no. 3, pp. 307–316, 2011.
- [19] W. F. DENHAM and S. PINES, "Sequential estimation when measurement function nonlinearity is comparable to measurement error," *AIAA Journal*, vol. 4, no. 6, pp. 1071–1076, 1966.
- [20] T. Dang, *Kontinuierliche Selbstkalibrierung von Stereokameras: Dissertation*, ser. Schriftenreihe / Institut für Mess- und Regelungstechnik, KIT. Karlsruhe: Univ.-Verl. Karlsruhe, 2007, vol. 8.
- [21] T. Dang, "An iterative parameter estimation method for observation models with nonlinear constraints," *Metrology and Measurement Systems*, vol. 15, no. 4, pp. 421–432, 2008.
- [22] S. R. Rump, "Intlab - interval laboratory 10.2," in *Developments in Reliable Computing*, Kluwer Academic Publishers, Ed. Dordrecht: Tibor Csendes, 1999, pp. 77–104. [Online]. Available: <http://www.ti3.tuhh.de/rump/>
- [23] The MathWorks Inc., "Matlab r2017b," Natick, MA, USA, 2017.
- [24] D. Simon and T. L. Chia, "Kalman filtering with state equality constraints," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 38, no. 1, pp. 128–136, 2002.
- [25] D. Simon, "Kalman filtering with state constraints: A survey of linear and nonlinear algorithms," *IET Control Theory & Applications*, vol. 4, no. 8, pp. 1303–1318, 2010.
- [26] V. Nguyen, S. Gächter, A. Martinelli, N. Tomatis, and R. Siegwart, "A comparison of line extraction algorithms using 2d range data for indoor mobile robotics," *Autonomous Robots*, vol. 23, no. 2, pp. 97–111, 2007.
- [27] H. L. Nguyen, P. Helmholz, D. Belton, and G. West, "New methods for segmentation of sparse mobile laser scanning point clouds: Proceedings," in *9th International Symposium on Mobile Mapping Technology*, 2015.
- [28] Julier, Simon, J. and Uhlmann, Jeffrey, K., "A new extension of the kalman filter to nonlinear systems," in *SPIE Proceedings of AeroSense. The 11th International Symposium on Aerospace/Defense Sensing, Simulation and Controls*, 1997, vol. 3068, pp. 182–193.

Article

Georeferencing of Laser Scanner-Based Kinematic Multi-Sensor Systems in the Context of Iterated Extended Kalman Filters Using Geometrical Constraints

Sören Vogel , Hamza Alkhatib , Johannes Bureick , Rozhin Moftizadeh and Ingo Neumann

Geodetic Institute, Leibniz Universität Hannover, Nienburger Str. 1, 30167 Hannover, Germany; alkhatib@gih.uni-hannover.de (H.A.); bureick@gih.uni-hannover.de (J.B.); rozhin.moftizadeh@stud.uni-hannover.de (R.M.); neumann@gih.uni-hannover.de (I.N.)

* Correspondence: vogel@gih.uni-hannover.de; Tel.: +49-511-762-4394

Received: 1 March 2019; Accepted: 15 May 2019; Published: 17 May 2019



Abstract: Georeferencing is an indispensable necessity regarding operating with kinematic multi-sensor systems (MSS) in various indoor and outdoor areas. Information from object space combined with various types of prior information (e.g., geometrical constraints) are beneficial especially in challenging environments where common solutions for pose estimation (e.g., global navigation satellite system or external tracking by a total station) are inapplicable, unreliable or inaccurate. Consequently, an iterated extended Kalman filter is used and a general georeferencing approach by means of recursive state estimation is introduced. This approach is open to several types of observation inputs and can deal with (non)linear systems and measurement models. The capability of using both explicit and implicit formulations of the relation between states and observations, and the consideration of (non)linear equality and inequality state constraints is a special feature. The framework presented is evaluated by an indoor kinematic MSS based on a terrestrial laser scanner. The focus here is on the impact of several different combinations of applied state constraints and the dependencies of two classes of inertial measurement units (IMU). The results presented are based on real measurement data combined with simulated IMU measurements.

Keywords: georeferencing; kinematic multi-sensor system; implicit model; iterated extended Kalman filter; inequality state constraints; probability density function truncation

1. Introduction

Multi-sensor systems (MSS) are greatly used nowadays in geodesy to capture an environment for various applications. Georeferencing is required in most cases for these data to be applicable. In simple and straightforward words, georeferencing is to derive the position and orientation of a platform with respect to a superordinate coordinate system. Therefore, in a static case, georeferencing would mean to derive six pose parameters (three translations and three rotations), whereas for kinematic platforms, the six degrees of freedom (DOF) should be calculated separately for each time epoch [1]. In general, there is no need to consider a scale factor as additional DOF, as long as sensors (e.g., laser scanner) of the MSS are consistent with each other during data acquisition. If there is a necessity, extension to a three-dimensional (3D) similarity transformation can be applied [2].

This indispensable necessity of precise and accurate pose parameters is a frequent challenge for outdoor and indoor mapping applications. Depending on respective complex or challenging environments common methods for georeferencing might fail, are unreliable or are at least inaccurate. The main reasons for this are missing or inaccurate observations from a global navigation satellite

system (GNSS) within indoor spaces or inner-city environments caused by shadowing or multipath effects. In addition, further methods (e.g., visual odometry) and sensors (e.g., inertial measurement unit (IMU)) have to deal with significant drifts in pose estimation [1,3–5]. In order to improve georeferencing for such challenging circumstances, a Kalman filter-based approach is extended and validated by means of a laser scanner-based kinematic MSS within this paper. As a novelty, arbitrary explicit and implicit measurement equations as well as nonlinear equality and inequality state constraints can be applied.

1.1. Georeferencing of Kinematic Multi-Sensor Systems

Georeferencing is generally realized through three different approaches, each of them deals with various methods that are based on the sensors available and environmental conditions. These approaches are called direct, indirect and data-driven georeferencing [2]. In direct georeferencing, position and orientation of a measuring platform are derived directly from the sensors available on board, such as a GNSS antenna [2], an IMU or an external sensor, such as a laser tracker [6] or a total station [4]. However, this approach depends highly on the environmental circumstances (e.g., visibility in complex indoor interior or absence of GNSS observations). In indirect georeferencing, observations of other sensors available on the platform, such as laser scanners or cameras are taken into account. In this approach, common environmental information (e.g., known control points for laser scanners by means of artificial targets) which are captured both in the local sensors' coordinate system and in a superordinate coordinate system are linked together [7]. Approaches for data-driven georeferencing require point cloud information which has already been georeferenced. This can be given by means of 3D city models, floor plans or other maps of the environment requested. Several arbitrary matching algorithms can be applied to get the position and orientation of an MSS which is acquiring point cloud data regarding models or maps mentioned. Uncertainty of the prior information affects the final georeferencing solution significantly. Known approaches for this method generally rely on iterative closest point (ICP) algorithms or rather on simultaneous localization and mapping (SLAM) methods [5,8–11].

1.2. Kalman Filter Techniques for Georeferencing

Combinations of aforementioned approaches are also possible and advisable to increase the accuracy and precision of the georeferencing of a kinematic MSS. This data fusion is commonly covered within the system state of a filtering approach. Such recursive approaches enable possibilities to handle big data, which come along with present and future multi-sensor technologies. Furthermore, they are suitable for online applications and usually require less memory and computational effort than batch algorithms that have been adapted for online georeferencing applications [12].

A Kalman filter (KF) is a well-known two-step procedure for this in which the next system state is estimated based on the previous state information and recent observations subsequently. Therefore, an iterative process is required by utilizing nonlinear measurement equations, which seem to be the most logical choice in case of trajectories. Consequently, such a procedure is called iterated extended Kalman filter (IEKF). A standard extended Kalman filter (EKF) can also handle nonlinear equations. However, an IEKF with further iterations is more suitable in the case of high nonlinear functions and will provide more accurate results using only small additional computational effort. Handling with nonlinear equations can also be done by means of unscented transformation (UT) as part of the unscented Kalman filter (UKF).

So far, in almost every research only explicit measurement equations are considered for such filter approaches. This means that the observations are taken into account as a function of the state parameters. Such a model is generally referred to a Gauss–Markov model (GMM). However, the use of a Gauss–Helmert model (GHM), which gives the possibility to implicitly link the observations to the state parameters, has also been studied by a few researchers [13–17]. Such a methodology provides the opportunity to include all kinds of measurement equations into the filtering approach,

regardless whether they are of an implicit or explicit nature. A basic algorithm for nonlinear implicit measurement equations within an IEKF for extrinsic auto-calibration of a stereo rig is proposed in [13,15]. The algorithm is used for the extrinsic auto-calibration of a stereo rig which has led to satisfying results. Furthermore, a linear KF with respect to GHM is developed in [17] and applied for orientation determination with smartphone sensors. However, both contributions do not consider state constraints. In addition, the latter is based on a linear KF approach. In [14], implicit measurement equations within a recursive estimation approach for Kalman filtering are referred to as implicit constraints. Usage of implicit measurement equations in terms of an UKF does not exist at all.

1.3. Contribution

Except for [16], the approaches mentioned within Section 1.2 have neglecting additional state constraints in common. Although, it is very useful to consider suitable environmental scene information by means of equality or inequality state constraints. This possibility is frequently used and evidenced in terms of well-known filter approaches with explicit formulations [2,18–20]. Horizontal and vertical lines, parallel or perpendicular lines and different planes in a scene are examples of such information which could be used as assigned geometric constraints during data analysis. In the recent work by [16], an IEKF by means of implicit measurement equations and nonlinear equality constraints is used for georeferencing of a simulated kinematic MSS. As a novelty, this approach is extended by nonlinear inequality state constraints within this paper. This increases the possibilities to apply any suitable geometrical prior information and to improve the georeferencing solution even in such challenging environments mentioned. Fundamental applicability is shown by means of a real kinematic MSS within an indoor environment and validated by highly accurate reference information. Additionally, a more general overview of the filter approach is given within this paper to make the approach independent from specific MSS, environments and prior information used.

1.4. Outline

The dedicated sections of this paper are as follows. An overview of the general georeferencing approach by means of a recursive state estimation is introduced in Section 2. This algorithm proposed is confirmed by being applied to a real data-set of an indoor environment that is captured by a kinematic MSS equipped with a TLS and tracked by a laser tracker in Section 3. The paper ends with a discussion of the results presented in Section 4 whereas Section 5 concludes this contribution.

2. General Georeferencing Approach by Means of Recursive State Estimation

A standardized estimation approach is indispensable to ensure an accurate, precise, reliable and complete georeferencing solution of different arbitrary kinematic MSS. The drawbacks mentioned in Section 1 could be eliminated only by providing a generally valid framework, which is applicable to as many use cases and systems as possible. For this reason, a recursive state estimation approach, which is compatible with various types of input data (e.g., requested states, available sensor observations and additional prior information), is formulated in this paper. However, the basic structure and equations used are with respect to [16]. The carefully selected information depends on each individual application and its respective circumstances. However, they are combined and fused in a unified way within the general valid framework to deliver optimal results. Necessary demands on the input data can be divided into four interconnected questions:

- Which types of sensor observations (e.g., laser scanner, GNSS, IMU, total station) are available and what are their accuracies?
- Which suitable and reliable prior information (e.g., geometrical circumstances, landmarks, maps) are available?
- What is the mathematical relationship between all input data?
- What information about the physical model of the system is known?

In theory, all possible input data should be considered. These input possibilities are restricted to the most common ones for the sake of simplification and according to the current paper perspectives. However, it should be noted that other types of input data are also possible and should be considered based on the application. A schematic overview of the universal recursive filter approach for georeferencing of a kinematic MSS together with corresponding relations between states, observations, prior information and respective parts of an IEKF is illustrated in Figure 1. In addition to the input data (states and observations) and prior information, sets of fundamental functions have to be formulated and integrated into the process. An arbitrary system model will describe the physical behaviour of the MSS between neighbouring epochs. Any model from the current state of the art can be selected for this. Total neglect of the system model is also possible and will result in a sequential adjustment approach. Formulation of a measurement model can happen in an implicit and/or explicit manner. (Non)linear functions regarding the states can also be added by means of equality and/or inequality formulations. Such state constraints can be integrated by means of several different methods (e.g., pseudo observations, projection method, probability density function (PDF) truncation or soft constraints). However, equality and inequality constraints have the crucial advantage of including specific further information into the filter approach by means of clear values (in the case of equalities) or thresholds (in the case of inequalities).

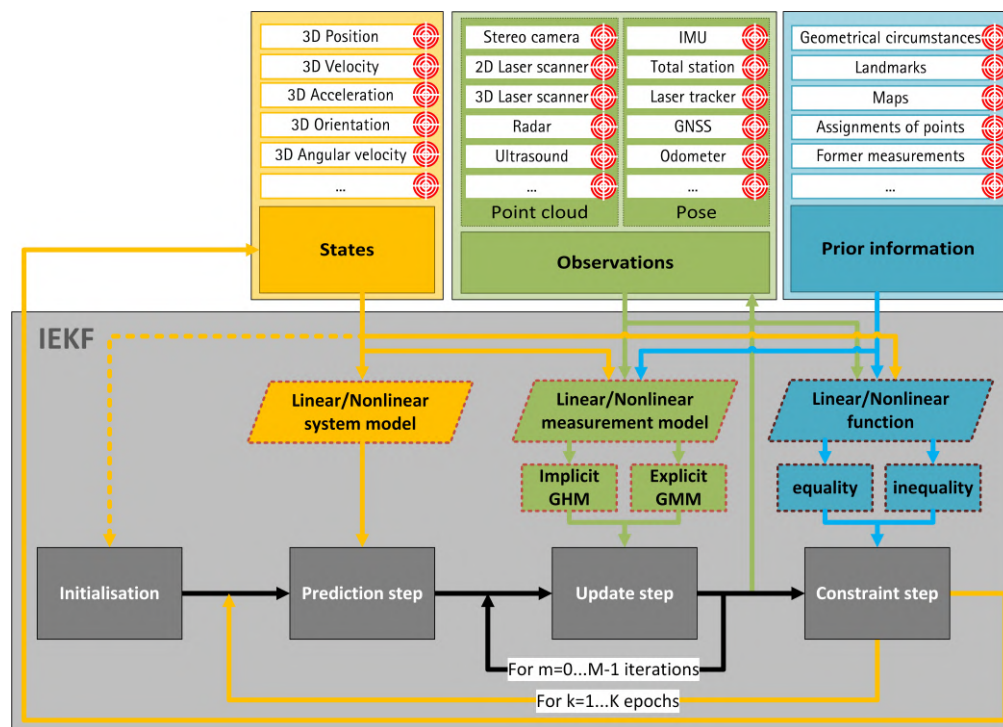


Figure 1. Schematic procedure of the universal recursive filter approach for georeferencing of a kinematic multi-sensor systems (MSS). Steps of the iterated extended Kalman filter (IEKF) (grey) are depicted with possible requested states (yellow), available observations (green) and known prior information (blue). Respective uncertainty information are depicted by red target circles.

2.1. Iterated Extended Kalman Filter with Nonlinear Implicit Measurement Equation

The basic structure of the georeferencing approach applied is based on the IEKF, which was published by [13] and enables the possibility of integrating implicit measurement equations (of type $h(l, x) = 0$) within the recursive estimation process. This gives the possibility to consider inextricable relationships between states x and observations l within the observation model. Only explicit relationships (of type $l = h(x)$) are allowed within normal KF, which results in major restrictions. As shown in Section 3, there is an important demand for implicit equations within the IEKF process in

order to consider more challenging relationships. However, explicit measurement equations are still possible to use but will be converted into $l - h(x) = 0$ to fulfil the implicit statement.

There are nonlinear measurement equations $h(\cdot)$ within the IEKF which provide a connection between the observations measured and states requested. The physical behaviour of the MSS over time is formulated within the nonlinear system model $f(\cdot)$ over all epochs $k = 1, \dots, K$ theoretically:

$$h(l_k + v_k, x_k) = 0, \quad v_k \sim N(0, \Sigma_{vv}) \quad (1)$$

$$x_k = f(x_{k-1}, u_{k-1}, w_{k-1}), \quad w_{k-1} \sim N(0, \Sigma_{ww}). \quad (2)$$

Here, u is the deterministic control by means of external controls. The variance-covariance matrix (VCM) Σ_{ww} of the system noise w and Σ_{vv} of the measurement noise v are normally distributed with zero mean. Regarding our universal recursive filter approach, the IEKF is divided into a prediction step using the system model, an update step making use of the measurement model and into a constrained step for applying known prior information by means of linear Dx_k and/or nonlinear $g(x_k)$ state constraints. All three steps are described within [16] in detail and are summarized in Algorithm 1.

Algorithm 1: Iterated extended Kalman filter (IEKF) with nonlinear implicit measurement equation and nonlinear equality state constraints.

```

1 System model  $x_k = f(x_{k-1}, u_{k-1}, w_{k-1}), \quad w_{k-1} \sim N(0, \Sigma_{ww})$ 
2 Observation model  $h(l_k + v_k, x_k) = 0, \quad v_k \sim N(0, \Sigma_{vv})$ 
3 Initial parameter vector and its VCM:  $\hat{x}_0^+ = x_0, \quad \Sigma_{\hat{x}\hat{x},0}^+ = \Sigma_{xx,0}, \quad k = 1$ 
4 while  $k < K$  do
5   Prediction step
6    $F_{x,k} = \partial f / \partial x|_{\hat{x}_{k-1}^+, u_{k-1}, w_{k-1}}$ 
7    $F_{w,k} = \partial f / \partial w|_{\hat{x}_{k-1}^+, u_{k-1}, w_{k-1}}$ 
8    $\hat{x}_k^- = f(\hat{x}_{k-1}^+, u_{k-1}, w_{k-1})$ 
9    $\Sigma_{xx,k}^- = F_{x,k} \Sigma_{xx,k-1}^+ F_{x,k}^T + F_{w,k} \Sigma_{ww} F_{w,k}^T$ 
10  Update step
11   $\check{l}_{k,0} = l_k, \quad \check{x}_{k,0} = \hat{x}_k^-$ 
12  for  $m = 0 \dots M-1$  do
13     $H_{x,k,m} = \partial h / \partial x|_{\check{l}_{k,m}, \check{x}_{k,m}}, \quad H_{l,k,m} = \partial h / \partial l|_{\check{l}_{k,m}, \check{x}_{k,m}}$ 
14     $K_{k,m} = \Sigma_{xx,k}^- H_{x,k,m}^T (H_{x,k,m} \Sigma_{xx,k}^- H_{x,k,m}^T + H_{l,k,m} \Sigma_{vv} H_{l,k,m}^T)^{-1}$ 
15     $\check{x}_{k,m+1} = \hat{x}_k^- - K_{k,m} \cdot (h(\check{l}_{k,m}, \check{x}_{k,m}) + H_{l,k,m} \cdot (l_k - \check{l}_{k,m}) + H_{x,k,m} \cdot (\hat{x}_k^- - \check{x}_{k,m}))$ 
16     $G_{k,m} = \Sigma_{vv} H_{l,k,m}^T (H_{x,k,m} \Sigma_{xx,k}^- H_{x,k,m}^T + H_{l,k,m} \Sigma_{vv} H_{l,k,m}^T)^{-1}$ 
17     $\check{l}_{k,m+1} = l_k - G_{k,m} \cdot (h(\check{l}_{k,m}, \check{x}_{k,m}) + H_{l,k,m} \cdot (l_k - \check{l}_{k,m}) + H_{x,k,m} \cdot (\hat{x}_k^- - \check{x}_{k,m}))$ 
18   $\hat{x}_k^+ = \check{x}_{k,M}$ 
19   $\hat{l}_k^+ = \check{l}_{k,M}$ 
20   $L_k = \mathbf{I}_{j \times j} - K_{k,M-1} H_{x,k,M-1}$ 
21   $\Sigma_{\hat{x}\hat{x},k}^+ = L_k \Sigma_{xx,k}^- L_k^T + K_{k,M-1} H_{l,k,M-1} \Sigma_{vv} H_{l,k,M-1}^T K_{k,M-1}^T$ 
22   $U_k = G_{k,M-1} \cdot H_{x,k,M-1}$ 
23   $\Sigma_{\hat{v}\hat{v},k}^+ = \Sigma_{vv} + G_{k,M-1} H_{l,k,M-1} \Sigma_{vv} H_{l,k,M-1}^T G_{k,M-1}^T - U_k \Sigma_{xx,k}^- U_k^T$ 
24  Constraint step
25   $D = g'(\hat{x}_k^-)$ 
26   $d = b - g(\hat{x}_k^-) + g'(\hat{x}_k^-) \cdot \hat{x}_k^-$ 
27  Set  $W = \mathbf{I}_{j \times j}$ 
28   $\tilde{x}_k^+ = \hat{x}_k^+ - W^{-1} D^T (D W^{-1} D^T)^{-1} (D \hat{x}_k^+ - d)$ 
29   $\Sigma_{\tilde{x}\tilde{x},k}^+ = \Sigma_{\hat{x}\hat{x},k}^+ - \Sigma_{\hat{x}\hat{x},k}^+ D^T (D \Sigma_{\hat{x}\hat{x},k}^+ D^T)^{-1} D \Sigma_{\hat{x}\hat{x},k}^+$ 
30  Set  $\hat{x}_k^+ = \tilde{x}_k^+$  and  $\Sigma_{\hat{x}\hat{x},k}^+ = \Sigma_{\tilde{x}\tilde{x},k}^+$ 

```

It is worth mentioning that up to now, to the best of our knowledge, no research which deals with state constraints in connection with implicit measurement equations are being investigated. All published methods are regarding explicit measurement equations where it is possible to separate states and observations from each other. However, in terms of implicit measurement equations of type $h(l, x) = 0$, modified state parameters during the constraint step will violate this equation because of the unaffected observations during the constraint step. This fact is currently being dealt with by the authors and is under investigation. However, our results in Section 3 show the fundamental validity.

2.2. Inequality State Constraints by Means of Probability Density Function Truncation

The constrained step within this paper is extended to allow the possibility of considering inequality state constraints. Instead of using the projection method (cf. Algorithm 1, line 28–30), the flexible PDF truncation method is used, and both given in [18] and [20]. In theory, other methods also mentioned above (e.g., pseudo observations) can be applied in order to consider state constraints. However, usage of state constraints in combination with implicit measurement equations is so far not considered for any Kalman filtering technique. By using the PDF truncation, equality and inequality constraints can be included simultaneously by the same method and there is no need to perform inefficient quadratic programming techniques. Furthermore, numerical instabilities resulting from e.g., singular measurement noise covariance in the context of perfect measurements can be avoided [19]. Depending on the respective conditions, the thresholds could be set by means of lower $lb_{i,k}$ and upper $ub_{i,k}$ boundaries for s scalar two-sided state constraints for any arbitrary nonlinear functions g_i of the states.

$$lb_{i,k} \leq g_i(x_k) \leq ub_{i,k} \quad i = 1, \dots, s. \quad (3)$$

Within this PDF truncation method, the estimated PDF of the IEKF (assumed in this paper as Gaussian) is truncated by means of the defined lower and upper boundaries and, subsequently, recomputed to the constrained estimate at the mean of the truncated PDF. Realization of the PDF truncation method is carried out for every single constraint $i = 1 \dots s$ successively. Furthermore, $\tilde{x}_{i,k}$ will be the state estimate after applying the i -th constraint and $\tilde{\Sigma}_{i,k}$ will be its respective VCM. Their initialization for $i = 0$ is achieved by updated KF estimations $\tilde{x}_{0,k} = \hat{x}_k^+$ and $\tilde{\Sigma}_{0,k} = \Sigma_{\hat{x}\hat{x},k}^+$. Afterwards, a transformation from $x_{i,k}$ to $z_{i,k}$ is performed for the decoupling of the s constraints:

$$z_{i,k} = S_i \cdot W_i^{-\frac{1}{2}} \cdot T_i^T (x_k - \tilde{x}_{i,k}). \quad (4)$$

The diagonal matrices W_i and orthogonal matrices T_i are obtained by performing Jordan canonical decomposition of the VCM $\tilde{\Sigma}_{i,k}$:

$$T_i \cdot W_i \cdot T_i^T = \tilde{\Sigma}_{i,k}. \quad (5)$$

The orthogonal matrix S_i is determined by using Gram–Schmidt orthogonalization (cf. [18]) and satisfies:

$$S_i \cdot W_i^{-\frac{1}{2}} \cdot T_i^T \cdot g_i(x_k) = \left[\left(g_i(x_k)^T \cdot \tilde{\Sigma}_{i,k} \cdot g_i(x_k) \right)^{\frac{1}{2}} \quad 0 \quad \dots \quad 0 \right]^T. \quad (6)$$

The normalized scalar constraint could be derived using this transformation, where $z_{i,k}$ has a zero mean and a VCM of identity.

$$a_{i,k} \leq \begin{bmatrix} 1 & 0 & \dots & 0 \end{bmatrix} \cdot z_{i,k} \leq b_{i,k}. \quad (7)$$

The transformed boundaries $a_{i,k}$ and $b_{i,k}$ are:

$$a_{i,k} = \frac{lb_{i,k} - g_i(\mathbf{x}_k) \cdot \tilde{\mathbf{x}}_{i,k}}{\left(g_i(\mathbf{x}_k)^T \cdot \tilde{\Sigma}_{i,k} \cdot g_i(\mathbf{x}_k)\right)^{\frac{1}{2}}}, \quad b_{i,k} = \frac{ub_{i,k} - g_i(\mathbf{x}_k) \cdot \tilde{\mathbf{x}}_{i,k}}{\left(g_i(\mathbf{x}_k)^T \cdot \tilde{\Sigma}_{i,k} \cdot g_i(\mathbf{x}_k)\right)^{\frac{1}{2}}}. \quad (8)$$

Truncation of the Gaussian PDF by means of the lower and upper bound and the integration variables ζ and γ is implemented by:

$$\int_{a_{i,k}}^{b_{i,k}} \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{\zeta^2}{2}} d\zeta = \frac{1}{2} \left[\operatorname{erf}\left(\frac{b_{i,k}}{\sqrt{2}}\right) - \operatorname{erf}\left(\frac{a_{i,k}}{\sqrt{2}}\right) \right], \quad \operatorname{erf}(u) = \frac{2}{\sqrt{\pi}} \int_0^u e^{-\gamma^2} d\gamma. \quad (9)$$

The normalized truncated PDF within the boundaries $a_{i,k}$ and $b_{i,k}$ is given by:

$$\operatorname{pdf}(\zeta) = \beta_{i,k} \cdot e^{-\frac{\zeta^2}{2}}, \quad \beta_{i,k} = \frac{\sqrt{2}}{\sqrt{\pi} \cdot \left[\operatorname{erf}\left(\frac{b_{i,k}}{\sqrt{2}}\right) - \operatorname{erf}\left(\frac{a_{i,k}}{\sqrt{2}}\right) \right]}. \quad (10)$$

The mean $\mu_{i,k}$ and variance $\sigma_{i,k}^2$ of the i -th element of $\mathbf{z}_{i,k}$ is computed by:

$$\mu_{i,k} = \beta_{i,k} \cdot \left[e^{-\frac{a_{i,k}^2}{2}} - e^{-\frac{b_{i,k}^2}{2}} \right] \quad (11)$$

$$\sigma_{i,k}^2 = \beta_{i,k} \cdot \left[e^{-\frac{a_{i,k}^2}{2}} \cdot (a_{i,k} - 2 \cdot \mu_{i,k}) - e^{-\frac{b_{i,k}^2}{2}} \cdot (b_{i,k} - 2 \cdot \mu_{i,k}) \right] + \mu_{i,k}^2 + 1. \quad (12)$$

With this, the mean $\tilde{\mathbf{z}}_{i+1,k}$ and VCM $\tilde{\mathbf{C}}_{i+1,k}$ of the transformed state could be estimated:

$$\tilde{\mathbf{z}}_{i+1,k} = \begin{bmatrix} \mu_{i,k} & 0 & \dots & 0 \end{bmatrix}^T \quad (13)$$

$$\tilde{\mathbf{C}}_{i+1,k} = \operatorname{diag}(\sigma_{i,k}^2, 1, \dots, 1). \quad (14)$$

The $\tilde{\mathbf{x}}_{i+1,k}$ and its corresponding VCM $\tilde{\Sigma}_{i+1,k}$ of the state are estimated by means of inversion of transformation in (4):

$$\tilde{\mathbf{x}}_{i+1,k} = \mathbf{T}_i \cdot \mathbf{W}_i^{-\frac{1}{2}} \cdot \mathbf{S}_i^T \cdot \tilde{\mathbf{z}}_{i+1,k} + \tilde{\mathbf{x}}_{i,k} \quad (15)$$

$$\tilde{\Sigma}_{i+1,k} = \mathbf{T}_i \cdot \mathbf{W}_i^{-\frac{1}{2}} \cdot \mathbf{S}_i^T \cdot \tilde{\mathbf{C}}_{i+1,k} \cdot \mathbf{S}_i \cdot \mathbf{W}_i^{-\frac{1}{2}} \cdot \mathbf{T}_i^T. \quad (16)$$

Finally, after performing this for all s constraints in series, the constrained states $\tilde{\mathbf{x}}_k$ and their VCM $\tilde{\Sigma}_k$ could be derived as:

$$\tilde{\mathbf{x}}_k = \tilde{\mathbf{x}}_{s,k} \quad (17)$$

$$\tilde{\Sigma}_k = \tilde{\Sigma}_{s,k}. \quad (18)$$

The whole procedure of PDF truncation for involving inequality constraints is depicted in Algorithm 2. In order to handle one-sided inequality constraints, $lb_{i,k} = -\infty$ or $ub_{i,k} = \infty$ could

be used. In the case of equality constraints, Equations (7), (8), (11), (12) are required to be changed to perform PDF truncation:

$$c_{i,k} = \begin{bmatrix} 1 & 0 & \dots & 0 \end{bmatrix} \cdot \mathbf{z}_{i,k} \quad (19)$$

$$c_{i,k} = \frac{d_{i,k} - \mathbf{g}_i(\mathbf{x}_k) \cdot \tilde{\mathbf{x}}_{i,k}}{\left(\mathbf{g}_i(\mathbf{x}_k)^T \cdot \tilde{\mathbf{\Sigma}}_{i,k} \cdot \mathbf{g}_i(\mathbf{x}_k) \right)^{\frac{1}{2}}} \quad (20)$$

$$\mu_{i,k} = c_{i,k} \quad (21)$$

$$\sigma_{i,k}^2 = 0. \quad (22)$$

Algorithm 2: Probability density function (PDF) truncation for inequality state constraints.

```

1 Initial parameter vector  $\tilde{\mathbf{x}}_{0,k} = \hat{\mathbf{x}}_k^+$  and its VCM  $\tilde{\mathbf{\Sigma}}_{0,k} = \mathbf{\Sigma}_{\hat{\mathbf{x}}\hat{\mathbf{x}},k}^+$  for  $i = 0$ 
2 while  $i < s$  do
3   Transformation:  $\mathbf{z}_{i,k} = \mathbf{S}_i \cdot \mathbf{W}_i^{-\frac{1}{2}} \cdot \mathbf{T}_i^T (\mathbf{x}_k - \tilde{\mathbf{x}}_{i,k})$ 
4   Jordan canonical decomposition of  $\tilde{\mathbf{\Sigma}}_{i,k}$ :  $\mathbf{T}_i \cdot \mathbf{W}_i \cdot \mathbf{T}_i^T = \tilde{\mathbf{\Sigma}}_{i,k}$ 
5   Gram–Schmidt orthogonalization for  $\mathbf{S}_i$ :  $\mathbf{S}_i \cdot \mathbf{W}_i^{-\frac{1}{2}} \cdot \mathbf{T}_i^T \cdot \mathbf{g}_i(\mathbf{x}_k) = \begin{bmatrix} \left( \mathbf{g}_i(\mathbf{x}_k)^T \cdot \tilde{\mathbf{\Sigma}}_{i,k} \cdot \mathbf{g}_i(\mathbf{x}_k) \right)^{\frac{1}{2}} & 0 & \dots & 0 \end{bmatrix}^T$ 
6    $a_{i,k} = \frac{lb_{i,k} - \mathbf{g}_i(\mathbf{x}_k) \cdot \tilde{\mathbf{x}}_{i,k}}{\left( \mathbf{g}_i(\mathbf{x}_k)^T \cdot \tilde{\mathbf{\Sigma}}_{i,k} \cdot \mathbf{g}_i(\mathbf{x}_k) \right)^{\frac{1}{2}}}$ ,  $b_{i,k} = \frac{ub_{i,k} - \mathbf{g}_i(\mathbf{x}_k) \cdot \tilde{\mathbf{x}}_{i,k}}{\left( \mathbf{g}_i(\mathbf{x}_k)^T \cdot \tilde{\mathbf{\Sigma}}_{i,k} \cdot \mathbf{g}_i(\mathbf{x}_k) \right)^{\frac{1}{2}}}$ 
7    $\beta_{i,k} = \frac{\sqrt{2}}{\sqrt{\pi} \cdot \left[ \operatorname{erf}\left(\frac{b_{i,k}}{\sqrt{2}}\right) - \operatorname{erf}\left(\frac{a_{i,k}}{\sqrt{2}}\right) \right]}$ 
8    $\mu_{i,k} = \beta_{i,k} \cdot \left[ e^{-\frac{a_{i,k}^2}{2}} - e^{-\frac{b_{i,k}^2}{2}} \right]$ ,  $\sigma_{i,k}^2 = \beta_{i,k} \cdot \left[ e^{-\frac{a_{i,k}^2}{2}} \cdot (a_{i,k} - 2 \cdot \mu_{i,k}) - e^{-\frac{b_{i,k}^2}{2}} \cdot (b_{i,k} - 2 \cdot \mu_{i,k}) \right] + \mu_{i,k}^2 + 1$ 
9    $\tilde{\mathbf{z}}_{i+1,k} = \begin{bmatrix} \mu_{i,k} & 0 & \dots & 0 \end{bmatrix}^T$ ,  $\tilde{\mathbf{C}}_{i+1,k} = \operatorname{diag}(\sigma_{i,k}^2, 1, \dots, 1)$ 
10   $\tilde{\mathbf{x}}_{i+1,k} = \mathbf{T}_i \cdot \mathbf{W}_i^{-\frac{1}{2}} \cdot \mathbf{S}_i^T \cdot \tilde{\mathbf{z}}_{i+1,k} + \tilde{\mathbf{x}}_{i,k}$ ,  $\tilde{\mathbf{\Sigma}}_{i+1,k} = \mathbf{T}_i \cdot \mathbf{W}_i^{-\frac{1}{2}} \cdot \mathbf{S}_i^T \cdot \tilde{\mathbf{C}}_{i+1,k} \cdot \mathbf{S}_i \cdot \mathbf{W}_i^{-\frac{1}{2}} \cdot \mathbf{T}_i^T$ 
11   $\tilde{\mathbf{x}}_k = \tilde{\mathbf{x}}_{s,k}$ 
12   $\tilde{\mathbf{\Sigma}}_k = \tilde{\mathbf{\Sigma}}_{s,k}$ 

```

3. Application in Terms of Accurate Indoor Georeferencing of a k-TLS

Various MSS in terms of calibration, acquisition and georeferencing are developed and used in practice at the Geodetic Institute Hannover (GIH) of the Leibniz University Hannover. Within this case study, the proposed general georeferencing approach from Section 2 is applied to a kinematic MSS extensively described in [6,21]. Utilizing such a proved MSS allows us to focus on the application of the proposed theoretical approach and to rely on an already calibrated and synchronized system. Furthermore, highly accurate validation by means of a laser tracker is possible. Such a comparison based on the trajectory is much more accurate than based on the 3D point cloud (e.g., TLS targets).

3.1. Overview

The kinematic MSS consists of a 3D TLS, a laser tracker, and a special probe (a normally hand-held combination of a reflector and ten LEDs for pose estimation regarding the laser tracker). The TLS is a Zoller + Fröhlich Imager 5016, which is used in a 2D profile mode for this application. The measuring rate of this sensor is 55 profiles/second and its range noise is 0.3 mm for a distance of 10 m [22]. The TLS is mounted on a rollable platform. On the other hand, the laser tracker used is a Leica AT960 LR with its Leica T-Probe. The T-Probe is rigidly mounted on top of the TLS and, therefore, moves with the TLS along the trajectory (cf. Figure 2a). Combination of the laser tracker with the T-probe gives the position with an accuracy of $\pm 15 \mu\text{m} + 6 \mu\text{m}/\text{m}$ (as a maximum permissible error (MPE) for the 3D position) and the orientation with an MPE of $0.01^\circ = 18 \mu\text{m}/100 \text{ mm}$ for the accuracy of each orientation-direction, respectively [23]. Due to the integration of such a highly accurate laser tracker into this MSS, the reference pose information with superordinated accuracy could be derived directly.

Data acquisition for this case study was carried out within the basement of the GIH inside a selected section of a corridor (cf. Figure 2b). Characteristics of this section are ideal with extensive walls and obstacles such as pipes below the ceiling and a door in one wall. The kinematic MSS was moved through this environment on an almost linear six meter long trajectory for about 25 s at a slow walking speed. This corresponds to 1311 epochs in total. We were aware that the used trajectory was limited in time and space. However, we intended here only to focus on the applicability of our developed georeferencing approach. The laser tracker was referenced in advance by given control points to ensure transformation to a superordinated coordinate system. The TLS targets regarding this coordinate system are also provided inside the environment measured. They will support further validation of our approach based on 3D point cloud information in the future. However, this issue was not in the focus of this paper. Instead, we will use the highly accurate reference pose by means of the laser tracker for validation. The TLS captured 3D points in a profile mode regarding its local sensor coordinate system. Every full laser scanner profile (LSP) was linked to 6D pose information by means of the laser tracker and T-probe. The right geometrical relation of this 6D pose to the reference point of the TLS was done by means of given calibration parameters (cf. [6]). The kinematic MSS utilized together with all the coordinate systems mentioned are depicted in Figure 2a. Thus, the direction of movement of the MSS is in x-direction. Consequently, the LSPs captured were in the “y–z” plane of the local laser scanner coordinate system. A highly accurate static full 3D laser scan of the captured section of the environment by means of the same laser scanner in 3D mode is also performed for further investigations. An overview of the true trajectory is pictured in Figure 3 as a top view by means of the laser tracker measurements in two different scales for the y-axis.

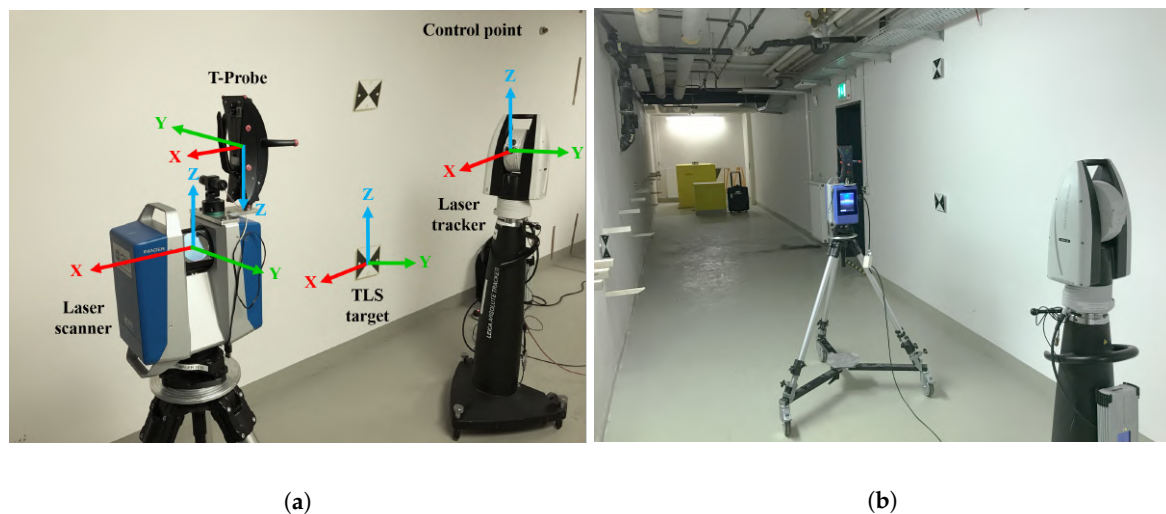


Figure 2. A general view of the kinematic MSS with its coordinate systems (a) used in the basement of the Geodetic Institute Hannover (GIH) (b).

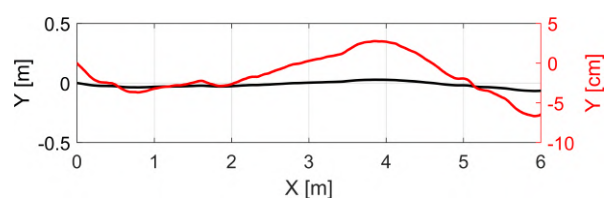


Figure 3. Top view of the measured trajectory obtained by the laser tracker. Two visualizations of the same trajectory in order to highlight the almost linear course. The black curve is regarding the left y-axis (meter) and the red curve regarding the right y-axis (centimetre).

3.2. Methodology

3.2.1. Observation Vector

The local 2D LSP and 6D pose information by means of the laser tracker in combination with the T-Probe were already at hand by means of the sensor data available from the MSS mentioned in Section 3.1. The 6D pose observations of an IMU are simulated on the basis of such highly accurate reference pose information. For this purpose, noise and a linear drift were added to the reference given to model realistic observations of a moderate and an accurate IMU. This was obviously just a rough approximation and did not reproduce observations of an IMU in reality. However, additional influencing parameters were neglected for the sake of simplicity. Single or rather double integration over a time of 25 s (regarding the duration of measurements mentioned in Section 3.1) for angular velocity and acceleration stability was used. Based on this, it was assumed that a drift in the position of ~ 16 m (moderate IMU) or rather ~ 2.5 m (accurate IMU) and a drift in the orientation of $\sim 5^\circ$ (moderate IMU) or rather $\sim 0.2^\circ$ (accurate IMU) was acquired. Due to the lack of information perpendicular to the scanning plane of the laser scanner (in the direction of movement), the position information in the x-direction was not affected by these changes and was consistently equal to a respected reference. The sampling rate of the simulated IMU observations was identical to that of the reference data.

Afterwards, the observation vector $\mathbf{l}_k$, consisting of one local LSP $\mathbf{P}_k^{\text{local}}$ (which consists, in turn, of N single 3D scan points), the 3D position $\mathbf{t}_k$ and 3D rotation matrix $\mathbf{R}_k$ (which is set up based on the three Euler angles Ω_k , Φ_k and K_k) of the IMU are derived for each epoch $k = 1 \dots K$. Apart from that, the 6D reference pose of the laser tracker (position $\mathbf{t}_k^*$ and rotation matrix $\mathbf{R}_k^*$) could be relied directly on for the purpose of validation.

$$\mathbf{l}_k = \left[\underbrace{x_{1,k}, y_{1,k}, z_{1,k}, \dots, x_{N,k}, y_{N,k}, z_{N,k}}_{\mathbf{P}_k^{\text{local}}}, \underbrace{X_k, Y_k, Z_k}_{\mathbf{t}_k}, \underbrace{\Omega_k, \Phi_k, K_k}_{\mathbf{R}_k} \right]^T. \quad (23)$$

Additionally, the corresponding VCM Σ_{vv} of the observations $\mathbf{l}_k$ can be set up, which consists of variances of the IMU $\Sigma_{vv\text{IMU}}$ and the quality information of LSP in the form of variances $\Sigma_{vv\text{LSP}}$. Related standard deviations for the VCM Σ_{vv} are given in Table 1. As has already been mentioned, IMU observations in the direction of movement (x) were assumed to be considerably more accurate than the ones in the perpendicular direction (Y and Z). Furthermore, the VCM $\Sigma_{vv\text{LSP}}$ applied for LSP was not based directly on the range noise of the laser scanner given by the manufacture. It was concluded in the context of former investigations that such specifications were overoptimistic within the scope of the current approach proposed. This was due to the fact that the observations had to fulfil additional equations (e.g., geometrical constraints) and needed to be more variable. Consequently, standard deviations of the VCM $\Sigma_{vv\text{LSP}}$ for LSP were larger than the manufacturer's specifications and selected generously.

$$\Sigma_{vv\text{LSP}} = \text{diag} \left(\sigma_{x_1}^2, \sigma_{y_1}^2, \sigma_{z_1}^2, \dots, \sigma_{x_N}^2, \sigma_{y_N}^2, \sigma_{z_N}^2 \right) \quad (24)$$

$$\Sigma_{vv\text{IMU}} = \text{diag} \left(\sigma_X^2, \sigma_Y^2, \sigma_Z^2, \sigma_\Omega^2, \sigma_\Phi^2, \sigma_K^2 \right) \quad (25)$$

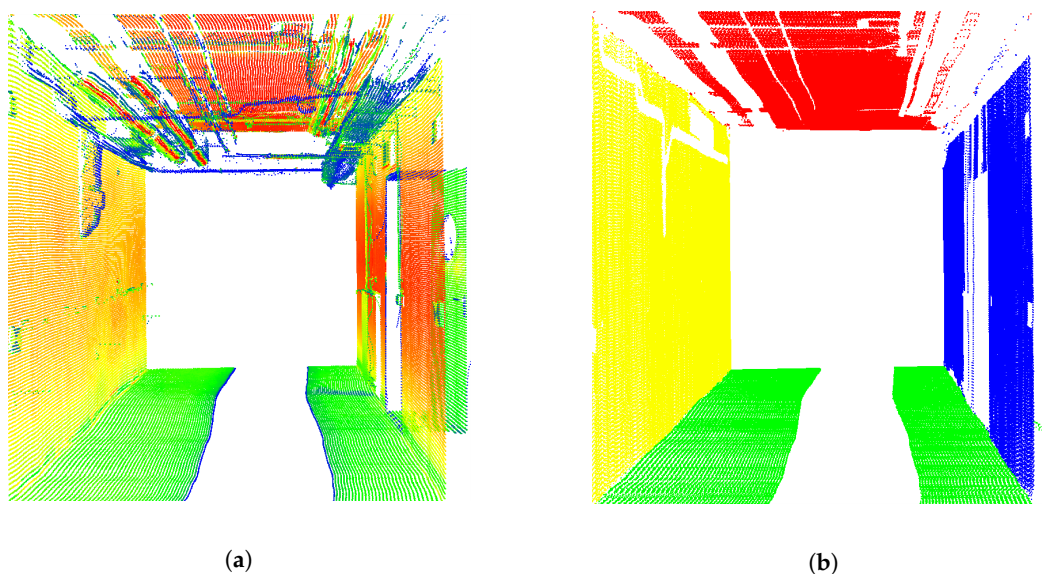
$$\Sigma_{vv} = \begin{bmatrix} \Sigma_{vv\text{LSP}} & \mathbf{0} \\ \mathbf{0} & \Sigma_{vv\text{IMU}} \end{bmatrix} \quad (26)$$

Table 1. Scheduled standard deviations σ for the variance-covariance matrix (VCM) Σ_{vv} of the observation vector I_k .

Sensor Type	Observation	Assumed σ	
		Moderate IMU	Accurate IMU
Laser scanner	x, y, z	3 mm	3 mm
IMU	X	0.01 mm	0.01 mm
	Y, Z	80 mm	20 mm
	Ω, Φ, K	0.2°	0.07°

3.2.2. Assignment Algorithm for Distinctive Planes

Further indispensable information were the assignments of every single 3D scan point to distinctive planes (left wall, right wall, ceiling and floor) of the environment. For this purpose, every captured LSP P_k^{local} was segmented individually to identify the walls, ceiling and floor properly. This is done by a RANSAC algorithm in order to find suitable line segments within each single LSP. Applied distance threshold for the consensus set was 5 mm in combination with a maximum of 30 iterations. Suitable candidates have a minimum percentage (2%) of points in comparison to the total number of points within the respective LSP. Additionally, at least 20 points needed to be assigned to a line segment. In order to only identify lines, which represented left or right walls or rather ceilings or floors, only those candidates were selected which are almost parallel or perpendicular regarding the standing axis of the laser scanner (which is known by means of the local coordinate system). In order to avoid doors, leads or other obstacles, line candidates were compared regarding averaged assignments of several past LSPs (named as memory subsequently). The criteria used for this are changes in distance between the respective line and origin of the laser scanner and the variation of the averaged intensity of respective scan points. Both criteria are analyzed regarding the memory mentioned. Rough outliers in the assignment could be identified by applying such a restriction. Finally, every N single 3D scan point of the LSP P_k^{local} within each epoch $k = 1 \dots K$ is assigned to left wall, right wall, ceiling, floor or remains as unused. These extended LSP are denoted C_k^{local} subsequently. Thus, in total, C_k^{local} is equal to P_k^{local} but contains mentioned additional segmentation information for every measured scan point. The results of the assignment algorithm introduced in relation to the case study are depicted in Figure 4a,b. Interfering objects (e.g., pipes and cables) are erased.

**Figure 4.** Georeferenced 3D point cloud of the environment measured based on the reference pose by means of laser tracker and T-Probe. Original scan points with colors by means of intensity (a). Assigned scan points regarding the left wall (yellow), right wall (blue), ceiling (red) and floor (green) (b).

3.2.3. Measurement Equation and State Parameter Vector

The state parameters desired were inter alia, relative changes in 3D position Δt_k , 3D orientation ΔR_k ($\Delta \Omega_k, \Delta \Phi_k, \Delta K_k$) and 3D velocity Δv_k . In combination with the (simulated) noisy and drifted IMU pose (t_k, R_k) they ended up in the almost true position t_k^{MSS} and orientation R_k^{MSS} ($\Omega_k^{\text{MSS}}, \Phi_k^{\text{MSS}}, K_k^{\text{MSS}}$) of the MSS at each epoch k :

$$t_k^{\text{MSS}} = t_k + \Delta t_k, \quad R_k^{\text{MSS}} = \Delta R_k \cdot R_k. \quad (27)$$

This formulation of relative changes as states might evoke a relation towards an error-state KF (ErKF, or indirect KF) [24,25]. However, the underlying concept of an ErKF is different. Instead of direct relative measurements (e.g., from an IMU), we included laser scanner observations by means of an implicit formulation. Georeferencing of every local LSP P_k^{local} regarding the superordinated coordinate system P_k^{global} can be applied by transformation using the estimated pose of the MSS:

$$P_k^{\text{global}} = (t_k + \Delta t_k) + (\Delta R_k \cdot R_k) \cdot P_k^{\text{local}}. \quad (28)$$

At this point, prior information is integrated into our approach. Several geometrical circumstances could be taken into consideration during the movement through the corridor (cf. Figure 2b). In the current case, it is assumed that certain parts of all individual LSP's captured some random parts of the left wall, right wall, ceiling and floor of the environment. Within a certain region, it could also be presumed that respective detected points on the left wall, right wall, ceiling and floor each refer to the same geometrical planes, respectively.

By using such information, the measurement equation could be formulated by means of the well-known Hesse normal form of a plane:

$$n_e \cdot \underbrace{\left[(t_k + \Delta t_k) + (\Delta R_k \cdot R_k) \cdot C_k^{\text{local}} \right]}_{C_k^{\text{global}}} - d_e = 0, \quad (29)$$

where n_e is the 3×1 normal vector of the left wall (or rather right wall, ceiling, floor) and d_e the related distance to the origin. Additionally, the segmented LSP information mentioned C_k^{local} regarding the left and right wall, ceiling and floor of the environment (cf. Section 3.2.2) could also be taken into consideration. In relation to Section 2.1 the given overall measurement Equation in (29) has an implicit formulation of type $h(l, x) = 0$.

Hence, the 25-dimensional state parameter vector x_k could be set up by means of the relative changes requested in position Δt_k , orientation ΔR_k ($\Delta \Omega_k, \Delta \Phi_k, \Delta K_k$) and velocity Δv_k and four sets of plane parameters with each four parameters $n_{e_x}, n_{e_y}, n_{e_z}, d_e$. Here, e can stand for the left wall (or rather right wall, ceiling or floor):

$$x_k = \left[\underbrace{\Delta X_k, \Delta Y_k, \Delta Z_k}_{\Delta t_k}, \underbrace{\Delta \Omega_k, \Delta \Phi_k, \Delta K_k}_{\Delta R_k}, \underbrace{\Delta v_{x_k}, \Delta v_{y_k}, \Delta v_{z_k}}_{\Delta v_k}, \underbrace{n_{\zeta_x}, n_{\zeta_y}, n_{\zeta_z}, d_{\zeta}}_{\text{left wall}}, \dots, \underbrace{n_{\zeta_x}, n_{\zeta_y}, n_{\zeta_z}, d_{\zeta}}_{\text{floor}} \right]^T. \quad (30)$$

It is worth mentioning that the increase of epochs in trajectories is associated with the increase of geometric details (e.g., walls) of buildings in the environment within real world application. This leads to an unlimited expansion of the state vector. The usage of a dual state Kalman filter (DKF) in such a case might be suitable. This would enable strict separation of time changing states (e.g., position, orientation, velocity) and other over time static parameters (e.g., normal vector and distances to origin of a plane) [26]. However, interaction of DKF and implicit measurement equations was not treated in this paper.

3.2.4. System Equation

A simple physical model was used to predict the constraint states from previous epoch $k - 1$ to the current k . This state transition was based on a constant velocity model, which only affected the six pose parameters and three velocities of the state parameter vector [27]. All plane parameters were unaffected by this prediction step and were equal to the constraint state $\hat{\mathbf{x}}_{k-1}^+$:

$$\Delta \hat{\mathbf{t}}_k^- = \Delta \hat{\mathbf{t}}_{k-1}^+ + \Delta \hat{\mathbf{v}}_{k-1}^+ \cdot \Delta \tau + \mathbf{w}_{\Delta t, k-1} \quad (31)$$

$$\Delta \hat{\mathbf{R}}_k^- = \Delta \hat{\mathbf{R}}_{k-1}^+ + \mathbf{w}_{\Delta R, k-1} \quad (32)$$

$$\Delta \hat{\mathbf{v}}_k^- = \Delta \hat{\mathbf{v}}_{k-1}^+ + \mathbf{w}_{\Delta v, k-1} \quad (33)$$

$$\hat{\mathbf{n}}_{e, k}^- = \hat{\mathbf{n}}_{e, k-1}^+ \quad (34)$$

$$\hat{\mathbf{d}}_{e, k}^- = \hat{\mathbf{d}}_{e, k-1}^+ \quad (35)$$

where $\mathbf{w}_{\Delta t, k-1}$, $\mathbf{w}_{\Delta R, k-1}$ and $\mathbf{w}_{\Delta v, k-1}$ are the process noise vectors and $\Delta \tau$ is the time interval between two consecutive epochs. The VCM of the process noise Σ_{ww} represents related system noise during the prediction step. Due to simplicity, all variances and covariances were zero, except for the process noise of the velocity. Within this case study a definition of $\sigma_{v, w} = 5 \cdot \Delta \tau$ is selected.

$$\Sigma_{ww} = \begin{bmatrix} \text{diag} \left(\mathbf{0}_{[1 \times 6]}, \sigma_{v, w}^2, \sigma_{v, w}^2, \sigma_{v, w}^2 \right) & \mathbf{0} \\ \mathbf{0} & \text{diag} \left(\mathbf{0}_{[1 \times 16]} \right) \end{bmatrix}. \quad (36)$$

3.2.5. Nonlinear Equality and Inequality Constraint for the State Parameters

In addition to the measurement Equation (29), the geometric prior information by means of equality and inequality constraints is also used to improve the georeferencing of the MSS. Due to the fact that the plane parameters $\mathbf{n}_e$ within the state parameter vector $\mathbf{x}_k$ were used, the unity of normal vectors had to be ensured. In order to do so, nonlinear equality constraints can be used:

$$g(\mathbf{x}_k) = \|\mathbf{n}_e\| = \sqrt{n_{e_x}^2 + n_{e_y}^2 + n_{e_z}^2} = b = 1. \quad (37)$$

Furthermore, inequality constraints regarding intersection angles of related planes are also implemented. In this context, obvious conditions for concurrency and perpendicularity between distinctive walls are relied on. It would also be possible to formulate these constraints by means of equality constraints. However, instead of using such hard constraints, the use of inequality constraints together with lower $lb_{i, k}$ and upper $ub_{i, k}$ boundaries (cf. Section 2.2) are preferred. Applying such inequality constraints is more consistent with reality, where such perfect conditions are rather rare or can be rarely fulfilled. Selected thresholds for this are derived based on documented standards for the building industry [28] and should be stated around 0° (for concurrency) and 90° (for perpendicularity). As a further basis, the information based on the highly accurate static 3D laser scanner point cloud mentioned are used to set up the boundaries. By means of this reference, true intersection angles between walls can be determined and applied. Consequently, the boundaries are selected by considering 0.5° for the intersection angles mentioned:

$$g_i(\mathbf{x}_k) = \cos^{-1} \left(\frac{|\mathbf{n}_{\zeta} \cdot \mathbf{n}_{\bar{\zeta}}|}{|\mathbf{n}_{\zeta}| \cdot |\mathbf{n}_{\bar{\zeta}}|} \right) = \cos^{-1} \left(\frac{|n_{\zeta_x} n_{\bar{\zeta}_x} + n_{\zeta_y} n_{\bar{\zeta}_y} + n_{\zeta_z} n_{\bar{\zeta}_z}|}{\sqrt{n_{\zeta_x}^2 + n_{\zeta_y}^2 + n_{\zeta_z}^2} \cdot \sqrt{n_{\bar{\zeta}_x}^2 + n_{\bar{\zeta}_y}^2 + n_{\bar{\zeta}_z}^2}} \right) \quad (38)$$

$$lb_{i, k} \leq g_i(\mathbf{x}_k) \leq ub_{i, k} \quad \text{with: } lb_{i, k} = g_i(\mathbf{x}_k) - 0.5^\circ, \quad ub_{i, k} = g_i(\mathbf{x}_k) + 0.5^\circ. \quad (39)$$

Due to the geometrical behaviour of the environment, there are several possibilities to apply Equation (38) in terms of concurrency or rather perpendicularity. For this reason, there are also different options for the number of geometrical inequality constraints selected. Within Section 3.3, the respective impacts and benefits of combining several constraints in contrast to individual use cases are shown. Regardless of the respective combination, which constraints are active within each epoch should be checked. This means that constraints can only be applied if at least five points of the related walls, ceiling or floor are segmented within this epoch. If there is a lack of one or several walls, all respective constraints to this wall will be inactive for this epoch.

3.2.6. Initialization

Initialization of approximate values for the state vector $\mathbf{x}_0$ and the related VCM $\Sigma_{xx,0}$ are needed to perform the IEKF. Initial relative changes in position $\Delta \mathbf{t}_0$ and orientation $\Delta \mathbf{R}_0$ are selected by means of the difference between the reference pose and IMU regarding first epoch $k = 0$. Relative velocities $\Delta \mathbf{v}_0$ are initialized as zero. Initial values for the normal vectors of the planes $\mathbf{n}_{e,0}$, are estimated by means of the first LSP and its respective points for left wall, right wall, ceiling and floor. Related standard deviations for the VCM $\Sigma_{xx,0}$ are given in Table 2.

$$\mathbf{x}_0 = [\Delta X_0, \Delta Y_0, \Delta Z_0, \Delta \Omega_0, \Delta \Phi_0, \Delta K_0, \Delta v_{x_0}, \Delta v_{y_0}, \Delta v_{z_0}, n_{\zeta_x,0}, n_{\zeta_y,0}, n_{\zeta_z,0}, d_{\zeta,0}, \dots, n_{\zeta_x,0}, n_{\zeta_y,0}, n_{\zeta_z,0}, d_{\zeta,0}]^T \quad (40)$$

$$\Sigma_{xx,0} = \text{diag} \left(\sigma_{\Delta X}^2, \sigma_{\Delta Y}^2, \sigma_{\Delta Z}^2, \sigma_{\Delta \Omega}^2, \sigma_{\Delta \Phi}^2, \sigma_{\Delta K}^2, \sigma_{\Delta v_x}^2, \sigma_{\Delta v_y}^2, \sigma_{\Delta v_z}^2, \sigma_{n_{\zeta_x}}^2, \sigma_{n_{\zeta_y}}^2, \sigma_{n_{\zeta_z}}^2, \sigma_{d_{\zeta}}^2, \dots, \sigma_{n_{\zeta_x}}^2, \sigma_{n_{\zeta_y}}^2, \sigma_{n_{\zeta_z}}^2, \sigma_{d_{\zeta}}^2 \right). \quad (41)$$

Table 2. Scheduled standard deviations σ for the initial VCM $\Sigma_{xx,0}$ of the initial state vector $\mathbf{x}_0$.

State Parameter	σ
$\Delta X, \Delta Y, \Delta Z$	0.1 m
$\Delta \Omega, \Delta \Phi, \Delta K$	5.7°
$\Delta v_x, \Delta v_y, \Delta v_z$	0.1 m/s
$n_{e_x}, n_{e_y}, n_{e_z}$	0.1
d_e	0.1 m

3.3. Results

In order to ensure independence from simulated IMU pose information, the results within this Section 3.3 are, with respect to the mean of 500 replications, of slightly different realizations of the IMU pose information. Additionally, to investigate the differences with respect to a moderate and an accurate IMU, results of two sets of simulations are presented within this Section 3.3. A schematic overview of this procedure is depicted in Figure 5. Evaluation is done by means of the estimated pose parameters of the kinematic MSS $\mathbf{t}_k^{\text{MSS}}$ and $\mathbf{R}_k^{\text{MSS}}$ and the ground truth by means of the laser tracker $\mathbf{t}_k^{\text{GT}}$ and $\mathbf{R}_k^{\text{GT}}$. Based on these pose information, the root mean square error (RMSE) for the combined position in the x-, y-, z-direction can be calculated (cf. (42)). In order to give a quality parameter for combined orientation, transformation from the rotation matrix $\mathbf{R}_k$ to the axis-angle representation by a normalized vector $\mathbf{r}_k = [r_1, r_2, r_3]$ and rotation angle Θ_k is performed. Afterwards, the mean error (ME) of the representative angle between estimation Θ_k^{MSS} and ground truth Θ_k^{GT} is calculated and used (cf. (43)). Presentation of the results by means of combined position and orientation instead of a single axis is intended. In such a manner, we can identify the most suitable combination of state constraints for this approach while keeping the results clear.

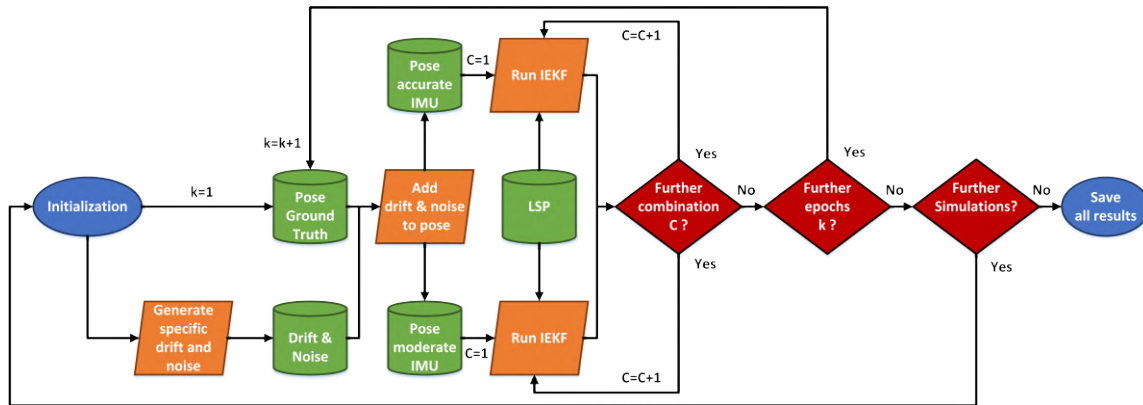


Figure 5. Schematic overview of the 500 replications performed for two types of IMUs as required input data for the iterated extended Kalman filter (IEKF) from Section 2.1 and its related combination $C = 1 \dots X$ of applied state constraints. The Roman numerals refer to respective state constraints applied regarding Table 3.

$$RMSE = \sqrt{\frac{1}{k} \sum_{i=1}^k \left((X_k^{GT} - X_k^{MSS})^2 + (Y_k^{GT} - Y_k^{MSS})^2 + (Z_k^{GT} - Z_k^{MSS})^2 \right)} \quad (42)$$

$$ME = \frac{1}{k} \sum_{i=1}^k \left(|\Theta_k^{GT} - \Theta_k^{MSS}| \right). \quad (43)$$

The difference between both classes of IMUs, as well as ground truth, is shown in Figure 6 by means of their averaged change in position and orientation over all corresponding 500 replications. As it has already been mentioned in Section 3.2.1, position in the x-direction for both IMUs was identical to the ground truth by means of the laser tracker. However, a major linear drift is visible (~ 15 m for moderate IMU or rather ~ 2.5 m for accurate IMU) for position in both other directions. Due to different assumed uncertainties for both IMUs, the drift in orientation for the accurate IMU was rather negligible, whereas the drift for the moderate IMU was about $\sim 5^\circ$ for all axes.

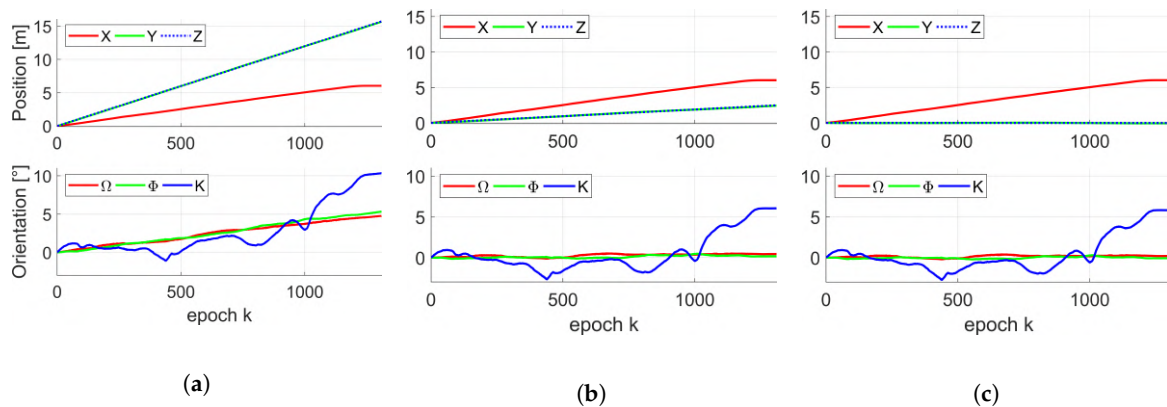


Figure 6. Mean change in position (top) and orientation (bottom) of the kinematic MSS by means of 500 simulated moderate IMU poses (a) and accurate IMU poses (b) over K epochs. True change in position (top) and orientation (bottom) of the kinematic MSS by means of laser tracker pose (c) over K epochs.

Based on this IMU pose information, methods from Section 3.2 are applied. In addition to both classes of IMUs, combinations of respective equality and inequality state constraints also affects the pose parameters requested. In the framework of this paper, estimates of ten different combinations (listed in Table 3) are presented. In order to compare various constraints in the developed IEKF algorithm, the combination I was designed without using any constraints and will be considered as a reference solution. All other combinations (II–X) relied on different equality and inequality constraints which include concurrency and/or perpendicularity between assigned left/right wall or rather ceiling and floor. In all these combinations the constraints regarding normal plane vectors (cf. (37)) were formulated in order to ensure numerical and geometrical stability. The inequality constraints in combinations III–VI were applied independently from each other whereas in combination VII–X a collaboration between concurrency and perpendicularity was enabled in order to evaluate respective impact on the state estimates for each collaboration. However, it is worth mentioning that the impact of the individual combinations might vary depending on respective application and related environmental circumstances.

Table 3. Investigated combinations of respective equality (red) and inequality (green) state constraints. Applied constraints within each combination are depicted with a ✓ symbol.

	Combinations of Respective Equality and Inequality State Constraints									
	I	II	III	IV	V	VI	VII	VIII	IX	X
unit vector for left wall		✓	✓	✓	✓	✓	✓	✓	✓	✓
unit vector for right wall		✓	✓	✓	✓	✓	✓	✓	✓	✓
unit vector for ceiling		✓	✓	✓	✓	✓	✓	✓	✓	✓
unit vector for floor		✓	✓	✓	✓	✓	✓	✓	✓	✓
left/right wall are parallel			✓				✓	✓		✓
ceiling/floor are parallel				✓			✓		✓	✓
left wall/ceiling are perpendicular					✓			✓	✓	✓
right wall/floor are perpendicular						✓				✓

The results achieved over all 500 replications for both moderate and accurate IMU observations in relation to the ten different combinations of applied state constraints are summarized in Table 4 for combined position by means of RMSE and Table 5 for combined orientation by means of ME. Comparison between the results was determined by means of minimum (min), maximum (max), mean, median and standard deviation (SD), as well as lower bound ($\downarrow$) and upper bound ($\uparrow$) of the 95% confidence interval (CI), calculated numerically from the 500 samples, as selected characteristic values.

It is notable that, independently from the IMU used, pose estimation fails if no constraints (combination I) are applied. It further stands out that there was an impact of the RMSE and ME depending on the constraints applied. In terms of position, combination III delivers the lowest estimates for both IMUs. Whereas for the moderate IMU, combination X was the lowest in terms of orientation, and for the accurate IMU, combination III is also the lowest (both judged by median). However, without taking into account combination I, all solutions by means of the applied state constraints were smaller than the ME for orientation of the moderate IMU. For the accurate IMU only combinations III, V, VI and VIII were smaller than the noisy and drifted IMU solution. However, the gain in accuracy is much higher for the position compared to the orientation.

Due to the conclusions provided by means of Tables 4 and 5, the temporal progress in position and orientation of the RMSE or rather ME for different combinations of state constraints are depicted in Figure 7 for the moderate IMU. The same results regarding the accurate IMU are depicted in Figure 8. For presentation purposes, inaccurate solutions (e.g., IMU in terms of position; combination I) are omitted. The basic behaviour of the temporal progress of the RMSE for position of both IMUs was very similar. All combinations increased drastically within the very first epochs. After this running-in effect of the filter they decrease quickly and continue differently over time. Over all epochs, combination II leads to a significant larger RMSE and has the largest increase. This is of interest, except normalized plane normal vectors, no further geometrical constraints like concurrency or perpendicularity were considered within this configuration. Combination III, V and VIII are very similar and lead to the best results around 1.5 cm. Remaining combinations have a slightly larger increase and will end up between 2–5 cm. Temporal behaviour of the ME for orientation is slightly different for both IMUs. They also increased drastically within the very first epochs. Afterwards, the gradient was related to the initially drifted IMU solution. However, all presented solutions for the moderate IMU were lower than respective initial IMU solution. In addition, gradient and progress are almost identical for this type of IMU. In case of the accurate IMU, there was a slight variation between all combinations. But from epoch $k = 800$ the increase was for all combinations lower than the IMU solution. Combination III behaves most similar to the IMU solution in the beginning and undercut the IMU curve at epoch $k = 400$.

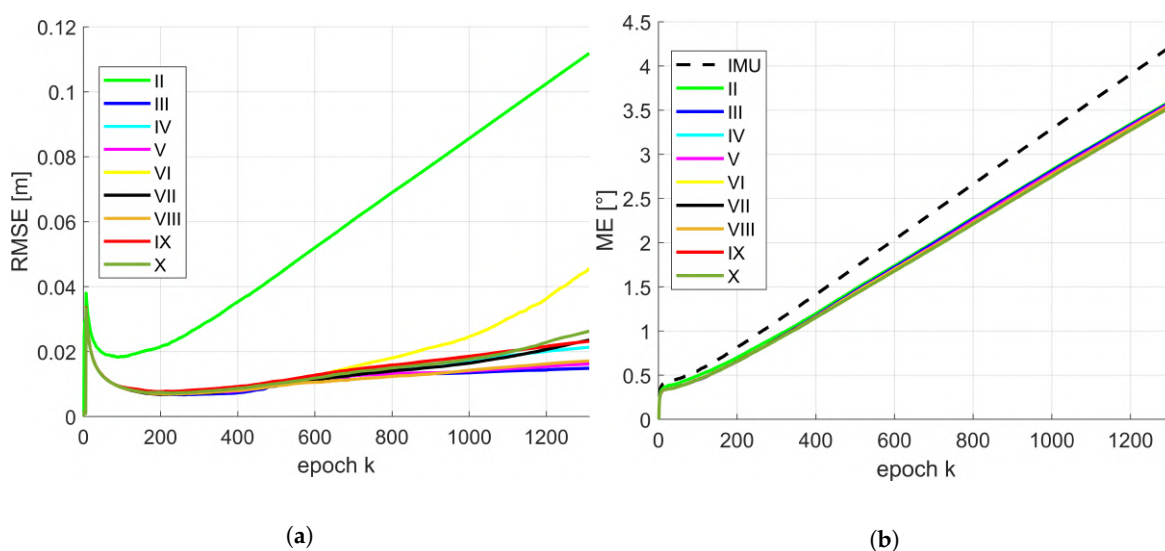


Figure 7. Moderate IMU: temporal progress of the median of the root mean square error (RMSE) for position (a) and mean error (ME) for orientation (b) by means of 500 replications for respective combinations of the state constraints applied. The Roman numerals refer to respective state constraints applied regarding Table 3.

In order to investigate the individual best results for position and orientation, respective histograms regarding the related 500 replications are depicted in Figure 9 for the moderate and accurate IMU with respect to each other. Based on these representations, further conclusions can be drawn. All histograms show distributions which are right-skew symmetric. This indicates that, independent from the IMU observation applied, there are a few configurations within the respective 500 replications which lead to a slightly larger RMSE or rather ME or even outliers. However, this skew is much more pronounced in case of the RMSE for position. The histograms for the ME of the orientation are similar to a Gaussian distribution. This different behaviour is not directly explainable and further investigations are needed. For this reason, a more detailed arrangement regarding the

single coordinate axis in contrast to the combined presentation appears appropriate and will be realized in the future.

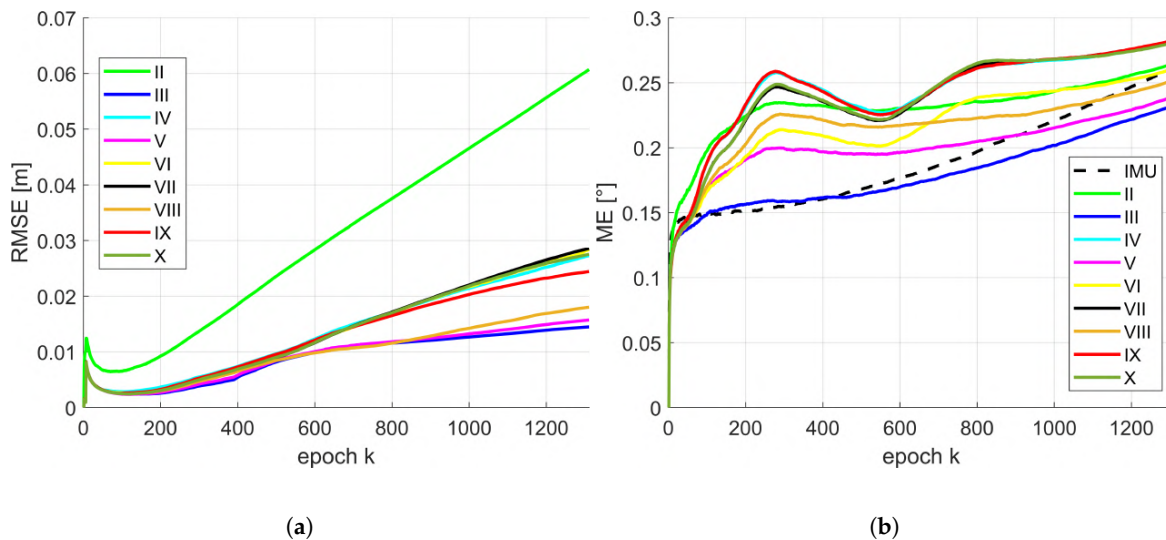


Figure 8. Accurate IMU: temporal progress of the median of the RMSE for position (a) and ME for orientation (b) by means of 500 replications for respective combinations of the state constraints applied. The Roman numerals refer to respective state constraints applied regarding Table 3.

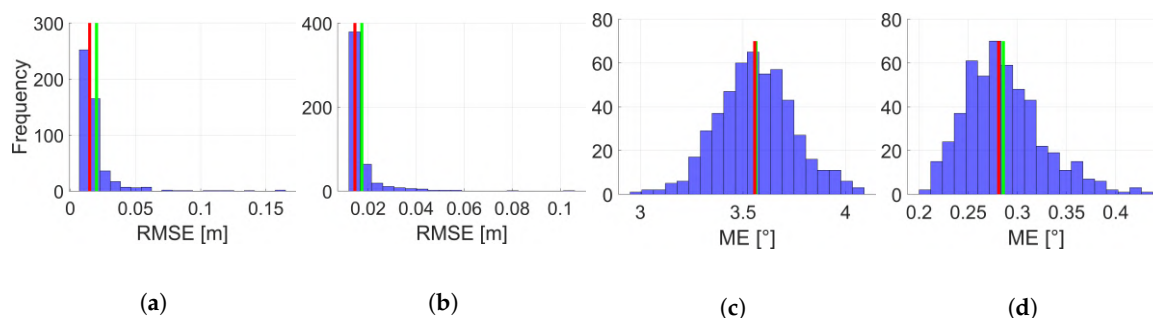


Figure 9. Histograms of the RMSE for position by means of 500 replications for combination III (moderate IMU (a)) and combination III (accurate IMU (b)) of state constraints applied. Related histograms of the ME for orientation by means of 500 replications for combination X (moderate IMU (c)) and combination III (accurate IMU (d)) of state constraints applied. Respective mean is given by a red bar and respective median by a green bar. The Roman numerals refer to respective state constraints applied regarding Table 3.

In general, it can be summarized that the consideration of state constraints improved state estimation significantly. However, differences between individual combinations were quite small. For this reason, geometrical restrictions regarding perpendicularity and concurrency depend strongly on respective environments.

4. Discussion

The results presented in Section 3.3 indicate significant dependencies of the estimated pose parameters on the respective equality and inequality state constraints applied. Moreover, no prominent combination of constraints exists which fits to all requirements in terms of position and orientation. The two different types of IMU observations demonstrate additional dependencies. Depending on the respective accuracy class, the use of certain constraints can significantly improve pose estimation. This applies particularly to the orientation estimation, whereas position estimation benefits from almost

every constraint applied, although to different levels. Overall, it could be seen that the usage of state constraints results in an important added value. However, there is an important need to define and apply a suitable model selection procedure into the current filter approach. Various different constraint combinations are applied and such a procedure should determine which combination is most suitable. Our priors will be obtained by considering which constraints are representative of features we expect to see in the data, and which would produce biased or inaccurate estimates. In addition, the effects of different individual constraints, in contrast to combined constraints, will be investigated more extensively in the future. Possible linear dependencies between individual constraints need to be analyzed and, if necessary, neglected.

As it has already been mentioned in Section 2.1, the compatibility of implicit measurement equations and state constraints as part of the IEKF is an important issue which needs further consideration. Due to implicit formulation, both states and observations are corrected within the update step to fulfill the measurement equation. During the constraint step, only the states are affected, while observations are unaffected. Consequently, this leads to a violated measurement equation. For this reason, the constraint step is going to be directly integrated into the update step of the IEKF. In terms of equality constraints, extension of the objective function should be sufficient. However, in terms of inequality constraints, this is not possible straightforwardly. Combination of inequality state constraints in the scope of GHM is treated in [29] and shows the complexity relating thereto. Another more promising approach for this task will be to use soft constraints instead of inequality constraints, which will be investigated in the future. Also, consideration of perfect measurements is possible, however, this would only allow equality constraints. In theory, state constraints can also be applied in combination with other classes of Kalman filters (e.g., UKF) to avoid the factual linearization issues of the EKF and IEKF. However, compatibility of implicit measurement equations besides linear KF, EKF and IEKF need to be solved first.

Table 4. Root mean square error (RMSE) for position by means of 500 replications. The Roman numerals refer to respective state constraints applied regarding Table 3. Each of the seven characteristic values (minimum, maximum, mean, median, standard deviation (SD) as well as lower bound ($\downarrow$) and upper bound ($\uparrow$) of the 95% confidence interval (CI)) are divided into two additional rows regarding moderate (above) and accurate (below) inertial measurement unit (IMU). The largest (red) and lowest (green) estimates are marked for first five rows.

Combinations of Respective Equality and Inequality State Constraints											
	IMU	I	II	III	IV	V	VI	VII	VIII	IX	X
Min [m]	12.646	1.4684	0.0118	0.0128	0.0147	0.0130	0.0135	0.0161	0.0130	0.0168	0.0164
	1.9833	0.8992	0.0114	0.0140	0.0168	0.0142	0.0143	0.0209	0.0142	0.0159	0.0209
Max [m]	13.030	3469.2	5.7065	0.1649	0.2269	0.1360	0.2460	0.1234	0.2278	0.4138	141.08
	2.0864	$2.4 \cdot 10^5$	4.3490	0.1049	0.1425	0.0561	0.1122	0.0966	0.0781	0.2269	0.0699
Mean [m]	12.835	79.778	0.3188	0.0201	0.0280	0.0218	0.0470	0.0269	0.0226	0.0335	0.3139
	2.0336	1974.7	0.1828	0.0174	0.0312	0.0182	0.0327	0.0304	0.0206	0.0282	0.0291
Median [m]	12.832	9.9179	0.1118	0.0149	0.0214	0.0162	0.0455	0.0236	0.0172	0.0232	0.0263
	2.0337	8.8968	0.0607	0.0145	0.0273	0.0157	0.0284	0.0286	0.0180	0.0244	0.0275
SD [m]	0.0678	320.29	0.6136	0.0160	0.0192	0.0151	0.0272	0.0116	0.0163	0.0325	6.3077
	0.0176	16083	0.3384	0.0080	0.0130	0.0064	0.0167	0.0081	0.0075	0.0147	0.0063
$\downarrow$ 95% CI [m]	12.714	2.7926	0.0189	0.0134	0.0163	0.0133	0.0140	0.0171	0.0135	0.0179	0.0178
	1.9983	2.0866	0.0137	0.0141	0.0194	0.0143	0.0146	0.0236	0.0144	0.0181	0.0227
$\uparrow$ 95% CI [m]	12.960	559.22	1.9255	0.0596	0.0742	0.0729	0.1042	0.0563	0.0654	0.1155	0.0947
	2.0707	7126.2	0.9893	0.0405	0.0647	0.0409	0.0771	0.0525	0.0414	0.0570	0.0498

The dependency mentioned regarding the IMU observations applied is an argument to consider initial biases and drift parameters of the IMU as additional state parameters within further developments. By doing so, direct estimation and consideration within the IEKF are possible and should further enhance the pose estimation. As a consequence of such an extension, this would be accompanied by further development of a more suitable system model within the prediction step.

Table 5. Mean error (ME) for orientation by means of 500 replications. The Roman numerals refer to respective state constraints applied regarding Table 3. Each of the seven characteristic values (minimum, maximum, mean, median, SD as well as lower bound ($\downarrow$) and upper bound ($\uparrow$) of the 95% CI) are divided into two additional rows regarding moderate (above) and accurate (below) IMU. The largest (red) and lowest (green) estimates are marked for first five rows.

Combinations of Respective Equality and Inequality State Constraints											
	IMU	I	II	III	IV	V	VI	VII	VIII	IX	X
Min [°]	3.6742	4.6565	2.9283	2.9843	2.9558	3.0133	2.9283	2.9544	2.9793	2.9560	2.9503
	0.1548	2.1602	0.1522	0.1541	0.1986	0.1557	0.1498	0.2151	0.1535	0.2084	0.2100
Max [°]	4.8145	42.297	4.3253	4.1203	4.0770	4.0935	4.0896	4.0788	4.0887	4.0779	4.0766
	0.4367	32.819	0.4542	0.4114	0.4455	0.4197	0.4218	0.4434	0.4279	0.4460	0.4369
Mean [°]	4.2414	11.222	3.6246	3.6140	3.5657	3.5990	3.5691	3.5653	3.5873	3.5643	3.5625
	0.2654	9.4379	0.2727	0.2371	0.2864	0.2447	0.2628	0.2860	0.2562	0.2857	0.2856
Median [°]	4.2471	10.355	3.6242	3.6117	3.5619	3.5992	3.5689	3.5619	3.5874	3.5615	3.5568
	0.2611	8.8777	0.2649	0.2322	0.2821	0.2397	0.2600	0.2822	0.2520	0.2825	0.2809
SD [°]	0.1890	4.2427	0.2302	0.1807	0.1863	0.1814	0.1893	0.1871	0.1838	0.1866	0.1874
	0.0529	4.1022	0.0598	0.0463	0.0400	0.0463	0.0482	0.0405	0.0453	0.0400	0.0400
$\downarrow$ 95% CI [m]	3.8684	5.7887	3.1598	3.2611	3.1979	3.2463	3.1921	3.1992	3.2241	3.1982	3.1970
	0.1793	3.9401	0.1762	0.1667	0.2212	0.1668	0.1706	0.2216	0.1836	0.2226	0.2206
$\uparrow$ 95% CI [m]	4.6200	22.2408	4.0898	3.9940	3.9578	3.9841	3.9604	3.9565	3.9702	3.9562	3.9542
	0.3758	19.2219	0.4079	0.3426	0.3809	0.3443	0.3699	0.3787	0.3612	0.3822	0.3804

5. Conclusions

We presented a novel method to consider nonlinear equality and inequality state constraints within the framework of an IEKF with implicit measurement equations. Consideration of such restrictions is realized by means of flexible PDF truncation. This method was applied and evaluated for georeferencing of an indoor laser scanner-based kinematic MSS. Therefor, different combinations of geometrical constraints were applied for real measurement data.

In conclusion, the consideration of appropriate restrictions between the state parameters is desirable. The use of inequality constraints in addition to equality constraints offers further possibilities in terms of accuracy. This justifies general consideration of inequality state constraints for georeferencing of a kinematic MSS.

Furthermore, adaptation and application of the general georeferencing approach by means of an IEKF with respect to other kinematic MSS is planned to verify its general validity. The focus there is to apply the approach on an UAV and an outdoor mobile mapping system. Both applications require special demands concerning 3D point cloud assignments in terms of facades, building models and further external influences which may occur within outdoor environments. In addition, the approach presented in this article has to be applied for longer data sets (with respect to spatial and temporal expansion). For this, it is assumed that the RMSE and ME will increase slightly over time unless absolute landmarks are integrated at certain points in time or assumed geometrical constraints are not applicable. Also nonlinear trajectories with turning manoeuvres need to be considered. This might make it necessary to introduce new planes and respective parameters into the model. However, applicability will be ensured as long as sufficient additional information from object space are available, assignable and applicable.

In addition, a more simple application example might be suitable to evaluate a comparison of different constraint combinations and methods to apply them to the fundamental algorithm of the IEKF with implicit measurement equations. As mentioned in Section 2, also other methods can be applied to consider state constraints. While PDF truncation provides great flexibility in simultaneously applying equality and inequality constraints, other methods mentioned above may be more appropriate (e.g., with respect to the uncertainty of the estimated state parameters and the computing time of the algorithm). However, methods for inequality constraints are limited as long as quadratic programming problems should be avoided.

Author Contributions: S.V., H.A. and I.N.: initial idea for this paper and its structure. S.V. and R.M.: performed the measurement with the MSS. S.V., J.B. and H.A.: design, conduction and analysis of experiment. S.V., J.B., R.M., H.A. and I.N.: writing of original draft and review of the manuscript.

Funding: This work was funded by the German Research Foundation (DFG) as part of the Research Training Group i.c.sens (RTG 2159) and NE 1453/5-1. The computations were performed by the compute cluster, which is funded by the Leibniz Universität Hannover, the Lower Saxony Ministry of Science and Culture (MWK) and DFG.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Vogel, S.; Alkhatib, H.; Neumann, I. Accurate Indoor Georeferencing with Kinematic Multi Sensor Systems. In Proceedings of the International Conference on Indoor Positioning and Indoor Navigation (IPIN), Alcalá de Henares (Madrid), Spain, 4–7 October 2016; pp. 1–8. [\[CrossRef\]](#)
2. Paffenholz, J.A. Direct Geo-Referencing of 3D Point Clouds with 3D Positioning Sensors. Ph.D. Thesis, Deutsche Geodätische Kommission (DGK), München, Germany, 2012; Volume 689.
3. Mautz, R. The challenges of indoor environments and specification on some alternative positioning systems. In Proceedings of the 6th Workshop on Positioning, Navigation and Communication (WPNC), Hannover, Germany, 19 March 2009; pp. 29–36. [\[CrossRef\]](#)
4. Keller, F.; Sternberg, H. Multi-Sensor Platform for Indoor Mobile Mapping: System Calibration and Using a Total Station for Indoor Applications. *Remote Sens.* **2013**, *5*, 5805–5824. [\[CrossRef\]](#)
5. Jung, J.; Yoon, S.; Ju, S.; Heo, J. Development of Kinematic 3D Laser Scanning System for Indoor Mapping and As-Built BIM Using Constrained SLAM. *Sensors* **2015**, *15*, 26430–26456. [\[CrossRef\]](#) [\[PubMed\]](#)
6. Hartmann, J.; Paffenholz, J.A.; Strübing, T.; Neumann, I. Determination of Position and Orientation of LiDAR Sensors on Multisensor Platforms. *J. Surv. Eng.* **2017**, *143*, 1–11. [\[CrossRef\]](#)
7. Abmayr, T.; Härtl, F.; Hirzinger, G.; Burschka, D.; Fröhlich, C. A correlation based target finder for terrestrial laser scanning. *J. Appl. Geod.* **2008**, *2*, 131–137. [\[CrossRef\]](#)
8. Nguyen, V.; Harati, A.; Martinelli, A.; Siegwart, R.; Tomatis, N. Orthogonal SLAM: A Step toward Lightweight Indoor Autonomous Navigation. In Proceedings of the International Conference on Intelligent Robots and Systems, Beijing, China, 9–15 October 2006; pp. 5007–5012. [\[CrossRef\]](#)
9. Lee, K.W.; Wijesoma, S.; Guzmán, J.I. A constrained SLAM approach to robust and accurate localisation of autonomous ground vehicles. *Robot. Auton. Syst.* **2007**, *55*, 527–540. [\[CrossRef\]](#)
10. Jutzi, B.; Weinmann, M.; Meidow, J. Improved UAV-Borne 3D Mapping by Fusing Optical and Laserscanner Data. *ISPRS Int. Arch. Photogr. Remote Sens. Spat. Inf. Sci.* **2013**, *XL-1/W2*, 223–228. [\[CrossRef\]](#)
11. Kaul, L.; Zlot, R.; Bosse, M. Continuous-Time Three-Dimensional Mapping for Micro Aerial Vehicles with a Passively Actuated Rotating Laser Scanner. *J. Field Robot.* **2016**, *33*, 103–132. [\[CrossRef\]](#)
12. Schön, S.; Brenner, C.; Alkhatib, H.; Coenen, M.; Dbouk, H.; Garcia-Fernandez, N.; Fischer, C.; Heipke, C.; Lohmann, K.; Neumann, I.; et al. Integrity and Collaboration in Dynamic Sensor Networks. *Sensors* **2018**, *18*, 21. [\[CrossRef\]](#) [\[PubMed\]](#)
13. Dang, T. *Kontinuierliche Selbstkalibrierung von Stereokameras: Dissertation*; Schriftenreihe/Institut für Messung und Regelungstechnik, KIT, Univ.-Verl. Karlsruhe: Karlsruhe, Germany, 2007; Volume 8.
14. Steffen, R.; Beder, C. Recursive Estimation with Implicit Constraints. In *Pattern recognition*; Lecture Notes in Computer Science; Hamprecht, F.A., Schnörr, C., Jähne, B., Eds.; Springer: Berlin, Germany, 2007; Volume 4713, pp. 194–203.

15. Dang, T. An Iterative Parameter Estimation Method for Observation Models with Nonlinear Constraints. *Metrol. Meas. Syst.* **2008**, *15*, 421–432.
16. Vogel, S.; Alkhatib, H.; Neumann, I. Iterated Extended Kalman Filter with Implicit Measurement Equation and Nonlinear Constraints for Information-Based Georeferencing. In Proceedings of the 21st International Conference on Information Fusion (FUSION), IEEE, Cambridge, UK, 10–13 July 2018; pp. 1209–1216. [[CrossRef](#)]
17. Ettlinger, A.; Neuner, H.; Burgess, T. Development of a Kalman Filter in the Gauss-Helmert Model for Reliability Analysis in Orientation Determination with Smartphone Sensors. *Sensors* **2018**, *18*, 414. [[CrossRef](#)] [[PubMed](#)]
18. Simon, D. *Optimal State Estimation*; John Wiley & Sons: Hoboken, NJ, USA, 2006.
19. Simon, D. Kalman filtering with state constraints: A survey of linear and nonlinear algorithms. *IET Control Theory Appl.* **2010**, *4*, 1303–1318. [[CrossRef](#)]
20. Simon, D.; Simon, D.L. Constrained Kalman filtering via density function truncation for turbofan engine health estimation. *Int. J. Syst. Sci.* **2010**, *41*, 159–171. [[CrossRef](#)]
21. Hartmann, J.; Trusheim, P.; Alkhatib, H.; Paffenholz, J.A.; Diener, D.; Neumann, I. High Accurate Pointwise (Geo-)Referencing of a k-TLS based Multi-Sensor-System. *ISPRS Ann. Photogr. Remote Sens. Spat. Inf. Sci.* **2018**, *IV-4*, 81–88. [[CrossRef](#)]
22. Zoller + Fröhlich GmbH. Z + F Imager 5016, 3D Laser Scanner–Data Sheet. Available online: <https://www.zf-laser.com/Z-F-IMAGER-R-5016.184.0.html> (accessed on 12 October 2018).
23. Hexagone Metrology. Leica Absolute Tracker AT960 (Product Brochure): Absolute Portability. Absolute Speed. Absolute Accuracy. Available online: https://w3.leica-geosystems.com/downloads123/m1/metrology/general/brochures/leica%20at960%20brochure_en.pdf (accessed on 12 October 2018).
24. Roumeliotis, S.I.; Sukhatme, G.S.; Bekey, G.A. Circumventing dynamic modeling: evaluation of the error-state Kalman filter applied to mobile robot localization. In Proceedings of the IEEE International Conference on Robotics and Automation, Piscataway, NJ, USA, 1999; pp. 1656–1663. [[CrossRef](#)]
25. Madyastha, V.; Ravindra, V.; Mallikarjunan, S.; Goyal, A. Extended Kalman Filter vs. Error State Kalman Filter for Aircraft Attitude Estimation. In *AIAA Guidance, Navigation, and Control Conference*; American Institute of Aeronautics and Astronautics: Reston, VA, USA, 2011.
26. Boada, B.L.; Garcia-Pozuelo, D.; Boada, M.J.L.; Diaz, V. A Constrained Dual Kalman Filter Based on pdf Truncation for Estimation of Vehicle Parameters and Road Bank Angle: Analysis and Experimental Validation. *IEEE Trans. Intell. Trans. Syst.* **2017**, *18*, 1006–1016. [[CrossRef](#)]
27. Bar-Shalom, Y.; Li, X.R.; Kirubarajan, T. *Estimation with Applications to Tracking and Navigation: Theory, Algorithms and Software*; Wiley: New York, NY, USA, 2001.
28. Deutsches Institut für Normung e.V. Toleranzen im Hochbau - Bauwerke (DIN 18202), 05.2015. Available online: https://bvn.de/Mitgliederservice/Infoline/Downloads/Technik/Toleranzen_im_Hochbau_2015.pdf (accessed on 12 October 2018).
29. Roese-Koerner, L.R. Convex Optimization for Inequality Constrained Adjustment Problems. Ph.D. Thesis, Deutsche Geodätische Kommission (DGK), München, Germany, 2015; Volume 759.



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).



Georeferencing of an Unmanned Aerial System by Means of an Iterated Extended Kalman Filter Using a 3D City Model

Johannes Bureick¹ · Sören Vogel¹ · Ingo Neumann¹ · Jakob Unger² · Hamza Alkhatib¹

Received: 15 May 2019 / Accepted: 30 September 2019
© The Author(s) 2019

Abstract

In engineering geodesy, the technical progress leads to various kinds of multi-sensor systems (MSS) capturing the environment. Multi-sensor systems, especially those mounted on unmanned aerial vehicles, subsequently called unmanned aerial system (UAS), have emerged in the past decade. Georeferencing for MSS and UAS is an indispensable task to obtain further products of the data captured. Georeferencing comprises at least the determination of three translations and three rotations. The availability and accuracy of Global Navigation Satellite System (GNSS) receivers, inertial measurement units, or other sensors for georeferencing is not or not constantly given in urban scenarios. Therefore, we utilize UAS-based laser scanner measurements on building facades. The building latter are modeled as planes in a three-dimensional city model. We determine the trajectory of the UAS by combining the laser scanner measurements with the plane parameters. The resulting implicit measurement equations and nonlinear equality constraints are covered within an iterated extended Kalman filter (IEKF). We developed a software simulation for testing the IEKF using different scenarios to evaluate the functionality, performance, strengths, and remaining challenges of the IEKF implemented.

Keywords Iterated extended Kalman filter · 3D city model · Unmanned aerial system · Laser scanner measurements · Implicit measurement equation · Equality constraint

Zusammenfassung

Georeferenzierung von Unmanned Aerial Systems mit Hilfe eines iterativen erweiterten Kalman Filters und eines 3D Gebäudemodells. In der Ingenieurgeodäsie führt der technische Fortschritt zu verschiedenen Arten von Multisensorsystemen (MSS), die der Erfassung der Umgebung dienen. In der vergangenen Dekade sind sehr viele MSS hinzugekommen, die auf einem Unmanned Aerial Vehicle montiert wurden. Diese MSS werden nachfolgend als Unmanned Aerial Systems (UAS) bezeichnet. Die Georeferenzierung von MSS und UAS ist ein notwendiger Schritt zur weiteren Datenverarbeitung. Die Georeferenzierung beinhaltet mindestens die Bestimmung von drei Translationen und drei Rotationen. Die erforderlichen Daten aus GNSS-Empfängern, inertialen Messsystemen oder anderen Sensoren zur Georeferenzierung sind in urbanem Umfeld nicht immer lückenlos und mit der erforderlichen Genauigkeit verfügbar. Deshalb werden in diesem Ansatz die Messungen UAS-basierter Laserscanner auf Gebäudefassaden verwendet. Letztere sind als Ebenen in einem 3D-Gebäudemodell modelliert. Die Trajektorie des UAS wird durch Kombination der Laserscanner-Messungen mit den Ebenenparametern ermittelt. Die daraus resultierenden impliziten Beobachtungsgleichungen und die nichtlinearen Restriktionsgleichungen werden innerhalb eines iterativen erweiterten Kalman-Filters (IEKF) modelliert. Außerdem wurde eine Softwaresimulation für den Test des IEKF entwickelt, um mit verschiedenen Szenarien die Funktionalität, Leistungsfähigkeit und verbleibende Herausforderungen zu bewerten.

✉ Johannes Bureick
bureick@gih.uni-hannover.de

¹ Geodetic Institute, Leibniz University Hannover, Nienburger Str. 1, 30167 Hannover, Germany

² Institute of Photogrammetry and GeoInformation, Leibniz University Hannover, Nienburger Str. 1, 30167 Hannover, Germany

1 Introduction

It is possible nowadays using the increasing capability and ongoing miniaturization of hardware and sensors to construct various kinds of multi-sensor systems (MSS). The MSS are frequently used in engineering geodesy to capture the environment. It is crucial to know the position and orientation of the MSS with respect to a superordinate coordinate system¹ for a further processing of the data captured. The task to determine position and orientation in a superordinate coordinate system is called georeferencing. Georeferencing usually comprises the determination of three translations and three rotations, also known as six Degrees of Freedom (DoF) or pose.

Unmanned aerial vehicles (UAV) have become a widespread, useful and affordable platform for the MSS in the past decade, optimally suited to capture common scenarios in engineering geodesy. The MSS mounted on a UAV are subsequently called unmanned aerial systems (UAS).

Precise georeferencing of a UAS is a challenging task, especially in urban areas. Global Navigation Satellite System (GNSS) data are affected by shadowing and multipath effects and are, therefore, inaccurate or even unavailable. Basically, low-cost GNSS receivers obtain positional accuracies of a few meters. These accuracies may be improved to decimeter or even centimeter level under good GNSS conditions and using satellite-based augmentation systems, differential GNSS, or real-time kinematics. Lightweight and low-cost inertial measurement units (IMU)² obtain orientation accuracies of 0.1° for roll and pitch and slightly worse accuracy, 0.8°, for heading. Combined GNSS-IMU systems improve the heading accuracy to 0.2°–0.5°. However, a general problem with IMU data is that it is often seriously affected by drifts.

A further possibility for georeferencing of an MSS is tracking using an external sensor. This is a cumbersome and inefficient task in the case of georeferencing a UAS due to limited measuring ranges and occultations.

In this paper, we present the georeferencing of a UAS, which is equipped with a three-dimensional (3D) laser scanner, a low-cost GNSS receiver, and an IMU, among other sensors. We set ourselves the requirement to determine the pose of the UAS with a precision (as a measure of the accuracy) better than 10 cm in position and 0.1° for orientation. The accuracy of the pose is decisive for the further results derived, such as digital terrain models or detailed 3D city models. Therefore, we fused the UAS-based laser scanner measurements towards building facades, which are modeled in a 3D city model, and the other sensor data within an iterated extended Kalman filter (IEKF). We developed

a software simulation for testing the IEKF developed using different scenarios to evaluate the functionality and performance of the IEKF implemented. All results presented are part of a research project, subsequently called the *UAS project*.

1.1 Georeferencing of MSS/UAS

The strategies to accomplish the georeferencing vary depending on the measurement configuration, the sensors and hardware available, the environment, and the accuracies to be fulfilled. The georeferencing strategies can be generally classified into *direct*, *indirect*, and *data-driven* georeferencing (see, e.g., Schuhmacher and Böhm 2005; Paffenholz 2012; Holst et al. 2015). An alternative classification regarding indoor applications is given in Vogel et al. (2016).

For *direct*³ georeferencing, sensors which measure the pose of an MSS directly are integrated into the MSS. These sensors could be, for example, a GNSS receiver (Paffenholz 2012; Talaya et al. 2004), an IMU (Talaya et al. 2004), or an external sensor, such as a total station or a laser tracker (Dennig et al. 2017; Hartmann et al. 2018), which determine the pose by angle and distance measurements to a reflector on the MSS.

*Indirect*⁴ georeferencing comprises methods where the pose of the MSS is determined by measurements towards known targets. These targets may be flat markers with a specific pattern (Abmayr et al. 2008) or simple 3D geometries, such as cylinders or spheres (Elkhrachy and Niemeier 2006). The position of the targets within the superordinate coordinate system is determined using an external sensor, for example, a total station. Indirect georeferencing is commonly used in bundle adjustment or for the georeferencing of static terrestrial laser scanners (TLS).

Data-driven georeferencing conforms basically to the indirect georeferencing. Instead of known targets, the data sets are matched to reference data sets. These reference data sets may be point clouds georeferenced already (Soloviev et al. 2007; Glira et al. 2015), digital surface models or 3D city models (Hebel et al. 2009; Li-Chee-Ming and Armenakis 2013; Unger et al. 2016, 2017). The matching is accomplished, for instance, via point-to-point assignment, for example, an iterative closest point algorithm (Besl and McKay 1992) or point-to-object assignment (see, e.g., Schuhmacher and Böhm 2005). Please note that *data-driven* georeferencing approaches vary widely. A huge part can be found in the commonly used Simultaneous Localization and Mapping (SLAM) approaches (see, e.g., Nguyen et al. 2006; Lee et al. 2007; Jutzi et al. 2013; Kaul et al. 2015; Nüchter et al. 2015).

¹ Also called the world coordinate system.

² E.g. the GNSS-IMU system in SBG Systems (2019).

³ Also called *sensor-driven*.

⁴ Also called *target-based*.

In principle, all the georeferencing strategies mentioned are suitable for georeferencing static and kinematic MSS. For a static MSS, the six DoF have to be determined only once, whereas for a kinematic MSS, the six DoF for each measuring epoch should be determined. It may be indispensable to model the pose almost continuously depending on the measuring frequency of the sensors used.

1.2 Filtering Techniques for Georeferencing

Georeferencing of a kinematic MSS is commonly covered as part of the state parameter vector within a filtering approach. Filtering is a two-step procedure in which the current state parameter vector is estimated based on the previous state parameter vector and the current observations. In the first step, called prediction, the current state parameter vector is predicted using the previous state parameter vector and a specific system model. The system model takes into account all the controlling variables and environmental noises that affect the current state. In the second step, called the measurement update step, the predictions are modified based on the observation equations which compare the predicted observations with the current original ones.

In recent years, many pose estimation algorithms based on the extended Kalman filter (EKF) have been successfully applied to solve the pose estimation problem of a UAS. Tailanian et al. (2014) focus on the sensor fusion of the local sensors of a UAS in which the GNSS and the IMU have been combined by means of an EKF. In Hol et al. (2007), pose estimation on a six DOF robot using an EKF to fuse this information has been shown. Forster et al. (2013) use the collaboration of multiple UAS for pose estimation to combine multiple SLAM algorithms and create an accurate pose estimation. In their approach, real-time camera pose estimation is accomplished by combining the inertial and vision measurements using nonlinear state estimation approaches.

In the case of using nonlinear observation equations, a linearization (realized by means of Taylor series expansion) has to be applied to overcome the nonlinearity issue (see Denham and Pines 1966). In the case where the effects of the linearization errors tend to affect the efficiency of the filter or its convergence, the re-linearization of the measurement equation around the updated state may reduce these difficulties. Therefore, such a procedure is called IEKF. Researchers commonly use explicit observation equations in the IEKF. This means that the observations are considered as a function of the state parameters. Such an observation model is generally called a Gauss Markov model (GMM).

If the equations relating the observations to the state parameters are condition equations, for example, some 3D points should fulfill the plane equations, then we do not exhibit the typical formulation of the GMM. In other words, the observations and the state parameters are not separable

(implicit measurement equations, see Dang 2007). In such a case, we are dealing with a Gauss Helmert model (GHM). In Dang (2007, 2008), Steffen and Beder (2007), Steffen (2013) and Ettlinger et al. (2018), an IEKF is used which deals with implicit measurement equations. In Vogel et al. (2018), an IEKF is used for georeferencing and extended with additional nonlinear equality constraints. In Vogel et al. (2019), the approach is further extended by integrating nonlinear inequality constraints. With the ability to handle implicit measurement equations and nonlinear inequality constraints, it is possible to depict almost any mathematical relationship within an IEKF.

1.3 Contribution

In this paper, we present an adaption of the IEKF in Vogel et al. (2018) for *data-driven* georeferencing of a UAS with an accuracy better than 10 cm in position and 0.1° in orientation. The highlight of the IEKF presented is the fusion of laser scanner measurements towards building facades, modeled as planes in a 3D city model, with the corresponding plane parameters of the 3D city model. This fusion leads to nonlinear implicit measurement equations and additional nonlinear equality constraints, which are covered within the IEKF. To evaluate the functionality and performance of the IEKF, we developed a simulation tool with a simple dynamic system model defining the state of the UAS and other states. On the basis of the states defined, we simulated sensor measurements regarding the sensor specifications and further characteristics. We tested the IEKF for different scenarios and analyzed its sensitivity towards different data characteristics.

1.4 Outline

The dedicated sections are as follows:

In Sect. 2 we give a detailed description of the UAS, the IEKF implemented and the entire workflow for georeferencing. Section 3 displays the simulation tool developed and gives an overview of the specifically chosen values for the IEKF. We present and discuss the results of the IEKF applied to the simulation in Sect. 4. Finally, conclusions and an outlook are drawn in Sect. 5.

2 IEKF for Georeferencing of a UAS

The Geodetic Institute (GIH) and the Institute of Photogrammetry and GeoInformation (IPI) of Leibniz University Hannover (LUH) are currently working on the *UAS project*. The *UAS project* deals with the precise determination of the trajectory of a UAS by integrating camera and laser scanner data in combination with generalized object information. Within this paper, we will focus on the usage of laser scanner

data in combination with a non-generalized 3D city model. We will exclusively use simulated data to evaluate the functionality and performance of the algorithms developed. The simulation scenarios were chosen in accordance with common measurement scenarios and data characteristics.

2.1 General Idea

The UAS moves through an urban area where the building facades are modeled as planes within a 3D city model. Three-dimensional city models with a level of detail 1 (block model) and a level of detail 2 (model with differentiated roof structures and boundary surfaces) are freely available for many cities. A detailed classification of the different levels of detail is given in Gröger et al. (2012). The pose of the UAS is roughly known, for example, from measurements of a GNSS receiver and an IMU. A 3D laser scanner captures the environment (red dots in Fig. 1) continuously. These captured 3D points may represent the ground, vegetation, building facades, or other objects. The laser scanner measurements are given in the local sensor coordinate system. The laser scanner measurements are transformed in a superordinate coordinate system according to the roughly known translation $t = [t_x, t_y, t_z]$ and orientation $\theta = [\omega, \phi, \kappa]$ (see Fig. 2). The transformed laser scanner measurements are assigned to planes of the 3D city model based on the distance between the scanner points and the planes of the building model (see Sect. 2.3). Only the points that are close enough to a plane of the 3D city model (green points in Fig. 2) are used as observations afterwards. The final pose parameters are estimated within a GHM, resp. filtering approach, by minimizing the distance between the assigned laser scanner measurements and the planes of the 3D city model to which they are assigned.

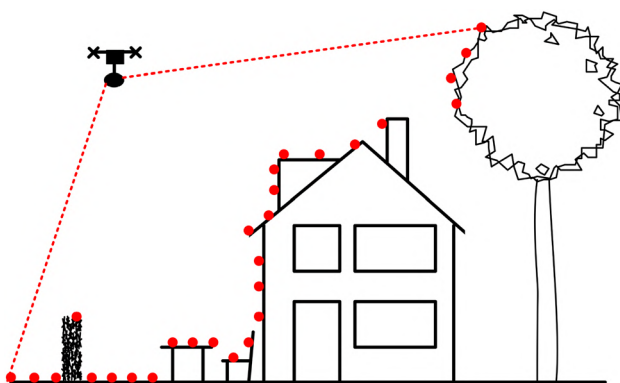


Fig. 1 Schematic depiction of the UAS capturing the environment. The points measured (red dots) may represent the ground, vegetation, building facades, or other objects

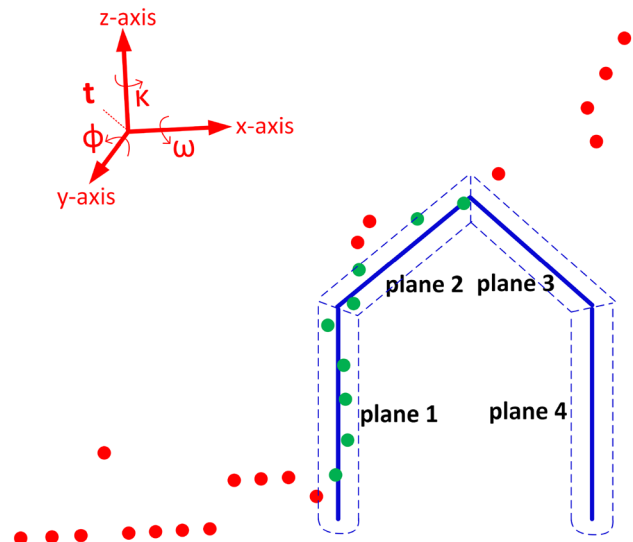


Fig. 2 Schematic depiction of the assignment of the transformed measured points. Some building facades are represented as planes in a generalized 3D city model (solid blue lines). The laser scanner measurements are transformed in the superordinate coordinate system according to roughly known translation $t = [t_x, t_y, t_z]$ and orientation $\theta = [\omega, \phi, \kappa]$. The origin of the laser scanner is depicted in the upper left corner. Note the slight difference to the original measurement configuration due to the roughly known pose. The transformed points which are close to one of the planes are assigned to that plane (green dots). The closeness is determined by a distance threshold (dashed blue lines). In this example, 5 points are assigned to plane 1 and 4 points are assigned to plane 2, whereas the other points (red dots) are not assigned. Subsequently, only the green dots are used as observations in the GHM, resp. filtering approach

2.2 UAS

A UAV is equipped with a 3D laser scanner, cameras, a GNSS receiver and an IMU in the *UAS project*. Figure 3 depicts a simplified sketch of the platform setup (without cameras). The laser scanner⁵ scans 16 scan lines which are nearly perpendicular to the sensors' vertical axis (that corresponds to the z -axis in Fig. 3). The divergence between each scan line is 2° . Thus, the laser scanner has a field of view of $30^\circ \times 360^\circ$. It is possible to set the resolution of the points in the scan lines between 0.1° and 0.4° . The rotation rate depends on that setting and is between 5 Hz and 20 Hz. We set the resolution to 0.4° to obtain a higher rotation rate of 20 Hz. Furthermore, it is possible to exclude certain angle areas from the measurement, because they cannot provide any data or the data generated are not needed for further processing. In our case, it would make sense to exclude the angle areas where the laser scanner measures towards the UAV.

⁵ Conforms to the Velodyne LiDAR PUCK VLP-16.

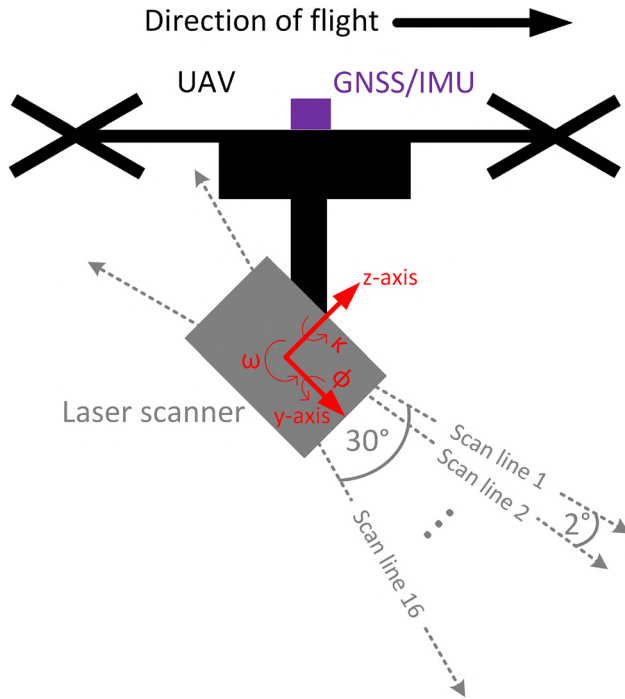


Fig. 3 Sketch of the platform setup. The x -axis of the local laser scanner coordinate system (red) points in readers line of sight

2.3 Assignment Algorithm

A crucial task of our georeferencing process is the assignment of measured 3D points C^{loc} , given in a local coordinate system of the laser scanner, to a plane of the city model, given in the global coordinate system. Therefore, we first have to transform C^{loc} to C^{glo} :

$$C^{\text{glo}} = t + R \cdot C^{\text{loc}}. \quad (1)$$

The parameters of the translation t describe the position of the origin of the local coordinate system. The rotation matrix R is obtained based on the orientation $\phi = [\omega, \phi, \kappa]$ of the local coordinate system according to Luhmann (2013):

$$R_{\omega} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos \omega & -\sin \omega \\ 0 & \sin \omega & \cos \omega \end{bmatrix} \quad (2)$$

$$R_{\phi} = \begin{bmatrix} \cos \phi & 0 & \sin \phi \\ 0 & 1 & 0 \\ -\sin \phi & 0 & \cos \phi \end{bmatrix} \quad (3)$$

$$R_{\kappa} = \begin{bmatrix} \cos \kappa & -\sin \kappa & 0 \\ \sin \kappa & \cos \kappa & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (4)$$

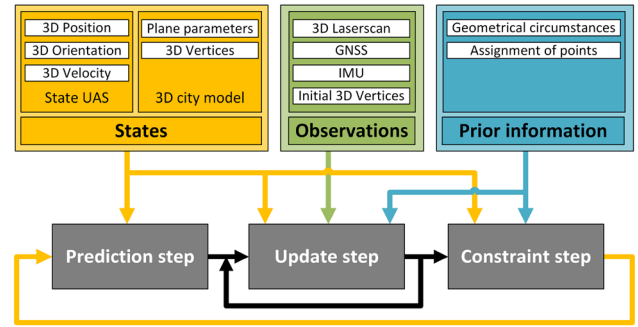


Fig. 4 Simplified basic workflow of the IEKF (grey) for georeferencing of a UAS by means of unknown states (yellow), available observations (green), and known prior information (blue)

$$R = R_{\omega} \cdot R_{\phi} \cdot R_{\kappa}. \quad (5)$$

The assignment is realized by a simple distance criterion, as described in Unger et al. (2016, 2017). The Euclidean distance to each plane of the city model is calculated for each 3D point in C^{glo} . Whether the projection of the points into the plane lies within the bounding polygon of the planes is checked for points where the distance to the nearest plane is less than a threshold value d_{assign} . If it is outside, the distance from the point to the boundary polygon of the plane is calculated and replaces the orthogonal distance. Points that project inside the boundary or are closer than d_{assign} are assigned to their closest plane. d_{assign} has to be selected regarding the accuracy of the points as well as the accuracy of the city model. The accuracy of the points depends mainly on the accuracy of the translation and orientation parameters t and ϕ and the accuracy of the measured points C^{loc} . The accuracy of the city model depends on the geometrical accuracy of its vertices and the extent of generalization effects by which the model deviates from the reality captured. The larger the threshold is fixed, the more points are assigned to planes, but the higher the probability is that incorrect assignments will be set that lead to incorrect results.

2.4 IEKF (Iterated Extended Kalman Filter)

The IEKF which we adapted for georeferencing of the UAS is given in Vogel et al. (2018). This approach can be applied for many different use cases. All available observations, types of prior information, and requested states need to be linked with the respective uncertainty information. Based on this, all information can be introduced into the IEKF.

According to the respective application, several relationships have to be established by means of linear or rather non-linear functional models. A system model $f(\cdot)$ is needed for the requested states x_k to describe the physical behavior of the system from epoch $k - 1$ to the current epoch k . Realization of this

transfer is carried out within the prediction step of the IEKF. Suitable functional relationships for each observation $\mathbf{l}_k$ have to be formulated within the measurement model $\mathbf{h}(\cdot)$. Based on implicit ($\mathbf{h}(E(\mathbf{l}), \mathbf{x}) = \mathbf{0}$) or explicit ($E(\mathbf{l}) - \mathbf{h}(\mathbf{x}) = \mathbf{0}$) formulations, relations between the observations available and requested states are considered. $E(\cdot)$ is the expected operator of a random vector; here, the expected value of the observation vector $\mathbf{l}$, which can be replaced by $E(\mathbf{l}) = \mathbf{l} + \mathbf{v}$. This consideration of current observations is carried out within the measurement update step. Available prior information can also be integrated within the measurement model during the measurement update step. In addition, further suitable prior information can also be applied in terms of state constraints by means of additional linear ($\mathbf{D}\mathbf{x}_k = \mathbf{d}$) or nonlinear ($g(\mathbf{x}_k) = \mathbf{b}$) functions. Here, $\mathbf{D}$ is a known matrix and $\mathbf{d}$ and $\mathbf{b}$ are known vectors with respect to related state constraints. Formulation of such restrictions can be implemented by means of equality constraints within the constraint step.

The basic workflow of this IEKF is depicted in Fig. 4 in a simplified way. Detailed equations for initialization, prediction, measurement update, and constraints are given in the following Sects. 2.4.1–2.4.6. Algorithm 1 given in Sect. 2.5 summarizes the required input, different computations steps and the output.

2.4.1 State Parameter Vector

The state parameter vector $\mathbf{x}_k$ consists of two parts, as already depicted in Fig. 4. The first part $\mathbf{x}_{\text{State},k} \in \mathbb{R}^9$ consists of the current position $\mathbf{t}_k$, orientation $\mathbf{o}_k$ and velocity $\mathbf{v}_k$ of the UAS (see Eq. 6). These state parameters describe the state of the UAS.

The second part consists, on one hand, of the vector $\mathbf{x}_{\text{Plane},k} \in \mathbb{R}^{4 \cdot E}$, which contains the parameters of all the E city model planes in Hesse normal form (see Eq. 7). The Hesse normal form is defined by the 3×1 normal vector $\mathbf{n}_e = [n_{x,e}, n_{y,e}, n_{z,e}]^T$ and the distance to the origin d_e with $e = 1 \dots E$. On the other hand, it consists of the vector $\mathbf{x}_{V,k} \in \mathbb{R}^{3 \cdot M}$, which contains all the M vertices of the building model (see Eq. 8):

$$\mathbf{x}_{\text{State},k} = \begin{bmatrix} \underbrace{t_{x,k}, t_{y,k}, t_{z,k}}_{\mathbf{t}_k}, \underbrace{\omega_k, \phi_k, \kappa_k}_{\mathbf{o}_k}, \underbrace{v_{x,k}, v_{y,k}, v_{z,k}}_{\mathbf{v}_k} \end{bmatrix}^T \quad (6)$$

$$\mathbf{x}_{\text{Plane},k} = \begin{bmatrix} \underbrace{\mathbf{n}_{1,k}, d_{1,k}}_{\text{plane 1}}, \dots, \underbrace{\mathbf{n}_{E,k}, d_{E,k}}_{\text{plane E}} \end{bmatrix} \quad (7)$$

$$\mathbf{x}_{V,k} = [\mathbf{v}_{1,k}^{\text{glo}}, \dots, \mathbf{v}_{m,k}^{\text{glo}}, \dots, \mathbf{v}_{M,k}^{\text{glo}}] \quad (8)$$

with

$$\mathbf{v}_{m,k}^{\text{glo}} = [\mathbf{v}_{x,m,k}^{\text{glo}}, \mathbf{v}_{y,m,k}^{\text{glo}}, \mathbf{v}_{z,m,k}^{\text{glo}}]^T. \quad (9)$$

The integration of plane parameters and vertices into the state parameter vector is used to identify and correct planes that are not accurately represented in the 3D city model. Although this purpose is not part of this paper, we are already introducing the mathematical relationship.

The complete state parameter vector $\mathbf{x}_k$ is arranged according to:

$$\mathbf{x}_k = [\mathbf{x}_{\text{State},k}; \mathbf{x}_{\text{Plane},k}; \mathbf{x}_{V,k}]. \quad (10)$$

2.4.2 Observation Vector

The MSS mentioned in Sect. 2.2 provides discrete 3D laser scanner point clouds (LSC) and 6D pose information by means of a GNSS receiver and an IMU. An LSC consists of a full scan rotation ($30^\circ \times 360^\circ$). For the sake of simplification, we will not consider the movement of the UAV during the period of a full scan rotation in the following. This time period is a maximum of 0.05 s for the configuration chosen. It is possible to exclude certain angle areas of the laser scanner, as has already been described in Sect. 2.2. The time period for the non-excluded areas is further reduced by excluding the angular areas in which the laser scanner measures in the direction of the UAV. This is acceptable due to the planned velocity of the UAV of about 1 m s^{-1} and an angular velocity of 2° s^{-1} . In further development steps, the LSC should consist of fewer and temporally closer 3D points, for example, just a half or a quarter scan rotation. In addition, the initial vertices of the 3D city model are introduced as observations into our approach. The observation vector $\mathbf{l}_k$ consists of three parts for each epoch $k = 1 \dots K$. The first part $\mathbf{l}_{\text{Scan},k}^{\text{loc}}$ consists of the measured LSC in the local sensor coordinate system with a total of N 3D points. $\mathbf{l}_{\text{Scan},k}^{\text{loc}}$ only contains the 3D points, which were assigned to a plane of the city model⁶ (for the assignment algorithm, see Sect. 2.3). These points are assorted on the basis of the planes they are assigned to:

$$\mathbf{l}_{\text{Scan},k}^{\text{loc}} = [\mathbf{l}_{\text{Scan},1,k}^{\text{loc}}; \dots; \mathbf{l}_{\text{Scan},e,k}^{\text{loc}}; \dots; \mathbf{l}_{\text{Scan},E,k}^{\text{loc}}] \quad (11)$$

with

$$\mathbf{l}_{\text{Scan},e,k}^{\text{loc}} = [\mathbf{p}_{e,1,k}^{\text{loc}}; \dots; \mathbf{p}_{e,i,k}^{\text{loc}}; \dots; \mathbf{p}_{e,N_e,k}^{\text{loc}}] \quad (12)$$

and

$$\mathbf{p}_{e,i,k}^{\text{loc}} = [X_{e,i,k}^{\text{loc}}, Y_{e,i,k}^{\text{loc}}, Z_{e,i,k}^{\text{loc}}]^T. \quad (13)$$

⁶ In photogrammetry, tie points are observed in more than one image and, therefore, connect poses. By contrast, points measured by the laser scanner are only observed from one pose and do not, therefore, contribute to pose estimation in this scenario as long as they are not related to a model plane.

$\mathbf{l}_{\text{Scan},e,k}^{\text{loc}}$ is representing the vector with all N_e points assigned to plane e . The total number N of 3D points of the LSC stored in the observation vector is calculated by:

$$N = \sum_{e=1}^E N_e. \quad (14)$$

The second part $\mathbf{l}_{\text{Pose},k}^{\text{glo}}$ consists of the direct GNSS and IMU observations for the position and orientation:

$$\mathbf{l}_{\text{Pose},k}^{\text{glo}} = \begin{bmatrix} \underbrace{\mathbf{l}_{x,k}^{\text{GNSS}}, \mathbf{l}_{y,k}^{\text{GNSS}}, \mathbf{l}_{z,k}^{\text{GNSS}}}_{\mathbf{l}_k^{\text{GNSS}}}, \underbrace{\omega_k^{\text{IMU}}, \phi_k^{\text{IMU}}, \kappa_k^{\text{IMU}}}_{\mathbf{o}_k^{\text{IMU}}} \end{bmatrix}^T. \quad (15)$$

Finally, the third part $\mathbf{l}_{V,0}^{\text{glo}}$ consists of the, in total M , initial vertices of the E model planes⁷:

$$\mathbf{l}_{V,0}^{\text{glo}} = [\mathbf{V}_{1,0}^{\text{glo}}, \dots, \mathbf{V}_{m,0}^{\text{glo}}, \dots, \mathbf{V}_{M,0}^{\text{glo}}] \quad (16)$$

with

$$\mathbf{V}_{m,0}^{\text{glo}} = [\mathbf{V}_{x,m,0}^{\text{glo}}, \mathbf{V}_{y,m,0}^{\text{glo}}, \mathbf{V}_{z,m,0}^{\text{glo}}]^T \quad (17)$$

representing a 3D point which is the vertex of at least one plane.

The observation vector is arranged as follows:

$$\mathbf{l}_k = [\mathbf{l}_{\text{Scan},k}^{\text{loc}}; \mathbf{l}_{\text{Pose},k}^{\text{glo}}; \mathbf{l}_{V,0}^{\text{glo}}]. \quad (18)$$

We apply Eq. (19) for building the stochastic model of the observation vector. We neglect correlations in all cases for a better discussion of the mainly important issues:

$$\Sigma_{ll,k} = \text{diag} \left(\underbrace{\sigma_{\text{LS}}^2, \dots, \sigma_{\text{LS}}^2}_{3 \cdot N}, \underbrace{\sigma_t^2, \sigma_o^2, \sigma_o^2}_{6}, \underbrace{\sigma_V^2, \dots, \sigma_V^2}_{3 \cdot M} \right). \quad (19)$$

The standard deviation of a laser scanner coordinate component is denoted by σ_{LS} . $\sigma_t = [\sigma_t, \sigma_t, \sigma_t]$ and $\sigma_o = [\sigma_o, \sigma_o, \sigma_o]$ denote the standard deviation of the GNSS and IMU observations, and σ_V denotes the standard deviation of the initial vertices.

2.4.3 System Equation

Formulating the system equation of the kind:

$$\mathbf{x}_k = \mathbf{f}(\mathbf{x}_{k-1}, \mathbf{u}_{k-1}) + \mathbf{w}_{k-1}, \quad (20)$$

we neglected the control vector $\mathbf{u}$ by setting it to zero. The system noise is normally distributed with $\mathbf{w}_{k-1} \sim N(\mathbf{0}, \Sigma_{ww})$. Equation (20) can be formulated as:

$$\mathbf{x}_k = \mathbf{F}_{x,k} \cdot \mathbf{x}_{k-1} + \mathbf{w}_{k-1}, \quad (21)$$

where $\mathbf{F}_{x,k}$ denotes the transition matrix:

$$\mathbf{F}_{x,k} = \partial \mathbf{f} / \partial \mathbf{x} |_{\hat{\mathbf{x}}_{k-1}, \mathbf{u}_{k-1}}. \quad (22)$$

Due to the fact that we do not intend to develop an optimal system model within this paper, we applied rather simple linear system equations. Nevertheless, a subsequent adaption of the system model with more complex system equations is easily possible if the future flight characteristics of the UAS require it. For the system equations chosen, the transition matrix $\mathbf{F}_{x,k}$ is given by:

$$\mathbf{F}_{x,k} = \begin{bmatrix} \mathbf{F}_{x,\text{State},k}^{[9 \times 9]} & \mathbf{0}_{[9 \times 4 \cdot E]} & \mathbf{0}_{[9 \times 3 \cdot M]} \\ \mathbf{0}_{[4 \cdot E \times 9]} & \mathbf{I}_{[4 \cdot E \times 4 \cdot E]} & \mathbf{0}_{[4 \cdot E \times 3 \cdot M]} \\ \mathbf{0}_{[3 \cdot M \times 9]} & \mathbf{0}_{[3 \cdot M \times 4 \cdot E]} & \mathbf{I}_{[3 \cdot M \times 3 \cdot M]} \end{bmatrix}, \quad (23)$$

with

$$\mathbf{F}_{x,\text{State},k} = \begin{bmatrix} \mathbf{I}_{[3 \times 3]} & \mathbf{0}_{[3 \times 3]} \text{diag}([\Delta \tau, \Delta \tau, \Delta \tau])_{[3 \times 3]} \\ \mathbf{0}_{[6 \times 3]} & \mathbf{I}_{[6 \times 6]} \end{bmatrix}. \quad (24)$$

$\Delta \tau$ is the time period between two consecutive epochs.

We apply the following equations for building the variance-covariance matrix (VCM) of the system noise Σ_{ww} :

$$\Sigma_{ww,\text{State}} = \text{diag}(\sigma_{t,w}^2, \sigma_{o,w}^2, \sigma_{v,w}^2) \quad (25)$$

$$\Sigma_{ww,\text{Plane}} = \text{diag}(\sigma_{n,w}^2, \sigma_{d,w}^2, \dots, \sigma_{n,w}^2, \sigma_{d,w}^2) \quad (26)$$

$$\Sigma_{ww,V} = \text{diag}(\sigma_{V,w}^2, \dots, \sigma_{V,w}^2) \quad (27)$$

$$\Sigma_{ww} = \begin{bmatrix} \Sigma_{ww,\text{State}} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \Sigma_{ww,\text{Plane}} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \Sigma_{ww,V} \end{bmatrix}. \quad (28)$$

The standard deviations of the system noise for translation, orientation, and velocity are denoted by $\sigma_{t,w} = [\sigma_{t,w}, \sigma_{t,w}, \sigma_{t,w}]$, $\sigma_{o,w} = [\sigma_{o,w}, \sigma_{o,w}, \sigma_{o,w}]$ and $\sigma_{v,w} = [\sigma_{v,w}, \sigma_{v,w}, \sigma_{v,w}]$. The system noise for the plane parameters and vertices is given by the standard deviations $\sigma_{n,w} = [\sigma_{n,w}, \sigma_{n,w}, \sigma_{n,w}]$, $\sigma_{d,w}$ and $\sigma_{V,w}$.

⁷ Please note the missing k in the index, which implies that this observation remains unchanged over the epochs.

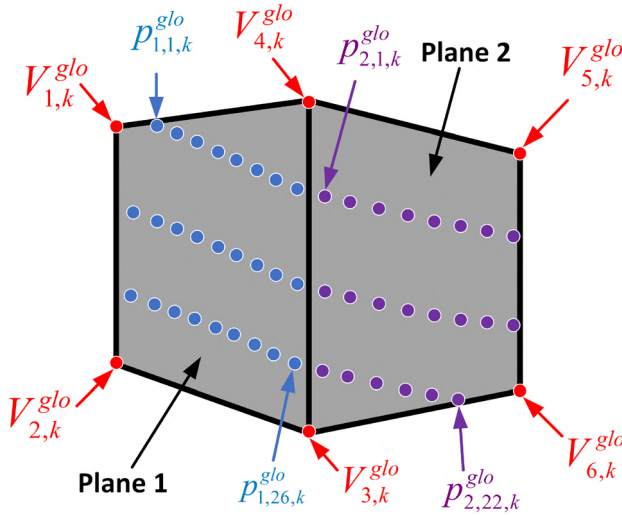


Fig. 5 Example of the relationship between state parameters, observations, and measurement equations in an arbitrary epoch k : The $N_1 = 26$ blue points $p_{1,1,k}^{glo}$ to $p_{1,26,k}^{glo}$ are assigned to plane 1. Subsequently, $\mathcal{I}_{Scan,1,k}^{loc}$ consists of these blue points in local sensor coordinate system. The $N_2 = 22$ purple points $p_{2,1,k}^{glo}$ to $p_{2,22,k}^{glo}$ are assigned to plane 2. Subsequently, $\mathcal{I}_{Scan,2,k}^{loc}$ consists of these purple points in local sensor coordinate system. If we assume that there are only those $E = 2$ planes in the 3D city model, the observation vector $\mathcal{I}_{Scan,k}^{loc}$ has the dimension $[144 \times 1]$. According to Eq. (14), N is equal to 48. The state parameter vector $\mathbf{x}_{V,k}$ consists of the six vertices $\mathbf{V}_{1,k}^{glo}$ to $\mathbf{V}_{6,k}^{glo}$ (in global coordinate system) and has the dimension $[18 \times 1]$. The vector $\mathcal{I}_{V,0}^{glo}$ has the same dimension. With solely two planes in the 3D city model, $\mathbf{x}_{Plane,k}$ has the dimension $[8 \times 1]$

2.4.4 Measurement Equation

With the state parameter and observation vector given in Sects. 2.4.1 and 2.4.2, we can formulate three types of nonlinear implicit measurement equations. The first type of measurement equation $\mathbf{h}_I$ causes that each transformed LSC $\mathcal{I}_{Scan,e,k}^{glo}$ has to be located in the assigned plane of the city model represented by $\mathbf{n}_e$ and d_e . If the point $\mathbf{p}_{e,i,k}^{loc} = [X_{e,i,k}^{loc}, Y_{e,i,k}^{loc}, Z_{e,i,k}^{loc}]^T$ is in $\mathcal{I}_{Scan,e,k}^{loc}$, which consists of all points assigned to plane e of the 3D city model, the following measurement equation is applied:

$$h_{I,i}(E(\mathbf{l}_k), \mathbf{x}_k) : \mathbf{n}_e^T \cdot \underbrace{[\mathbf{t}_k + \mathbf{R}_k \cdot E(\mathbf{p}_{e,i,k}^{loc})]}_{\mathbf{p}_{e,i,k}^{glo}} - d_e = 0, \quad (29)$$

with $i \in \{1 \dots N\}$.

The rotation matrix $\mathbf{R}_k$ is calculated based on the current orientation $\mathbf{o}_k$ according to Eqs. (2)–(5). Subsequently, we obtain N measurement equations of type $\mathbf{h}_I$:

$$\mathbf{h}_I(E(\mathbf{l}_k), \mathbf{x}_k) = [h_{I,1}(E(\mathbf{l}_k), \mathbf{x}_k), \dots, h_{I,N}(E(\mathbf{l}_k), \mathbf{x}_k)]^T. \quad (30)$$

Figure 5 shows an illustration of the relationship between state parameters, observations, and measurement equations.

In the second type of measurement equation $\mathbf{h}_{II}$, we calculate the difference between the estimated pose $[\mathbf{t}_k; \mathbf{o}_k]$ and observed pose $\mathcal{I}_{Pose,k}^{glo}$ of GNSS and IMU:

$$\mathbf{h}_{II}(E(\mathbf{l}_k), \mathbf{x}_k) : [\mathbf{t}_k; \mathbf{o}_k] - E(\mathcal{I}_{Pose,k}^{glo}) = \mathbf{0}. \quad (31)$$

If the GNSS and IMU data are available, we obtain six measurement equations of type $\mathbf{h}_{II}$. When the GNSS signal is missing, the measurement equation changes to:

$$\mathbf{h}_{II}(E(\mathbf{l}_k), \mathbf{x}_k) : \mathbf{o}_k - E(\mathbf{o}_k^{IMU}) = \mathbf{0}, \quad (32)$$

and we only obtain three measurement equations of this type.

The third type of measurement equation $\mathbf{h}_{III}$ is introduced to avoid a datum defect. It computes the difference between the estimated vertices in $\mathbf{x}_{V,k}$ and the initial vertices in $\mathcal{I}_{V,0}^{glo}$. If the vertex $\mathbf{V}_{m,k}^{glo}$ is stored in $\mathbf{x}_{V,k}$ and $\mathbf{V}_{m,0}^{glo}$ is its initial observation stored in $\mathcal{I}_{V,0}^{glo}$, the following measurement equations are applied:

$$\mathbf{h}_{III,m}(E(\mathbf{l}_k), \mathbf{x}_k) : \mathbf{V}_{m,k}^{glo} - E(\mathbf{V}_{m,0}^{glo}) = \mathbf{0}, \quad (33)$$

with $m \in \{1 \dots M\}$.

For a three-dimensional vertex Eq. (33) consists of three measurement equations. Subsequently, we obtain $3 \cdot M$ measurement equations of type $\mathbf{h}_{III}$:

$$\mathbf{h}_{III}(E(\mathbf{l}_k), \mathbf{x}_k) = \begin{bmatrix} \mathbf{h}_{III,1}(E(\mathbf{l}_k), \mathbf{x}_k) \\ \vdots \\ \mathbf{h}_{III,M}(E(\mathbf{l}_k), \mathbf{x}_k) \end{bmatrix}. \quad (34)$$

Altogether, the measurement equations are given by:

$$\mathbf{h}(E(\mathbf{l}_k), \mathbf{x}_k) = \begin{bmatrix} \mathbf{h}_I(E(\mathbf{l}_k), \mathbf{x}_k) \\ \mathbf{h}_{II}(E(\mathbf{l}_k), \mathbf{x}_k) \\ \mathbf{h}_{III}(E(\mathbf{l}_k), \mathbf{x}_k) \end{bmatrix}. \quad (35)$$

2.4.5 Nonlinear Equality Constraint for the State Parameters

We apply two types of nonlinear equality constraints for the state parameters in our IEKF. The prior information which we want to model describes hard constraints which have to be fulfilled.

The first type of nonlinear equality constraint $\mathbf{g}_I$ arises due to the fact that we are using plane parameters within the state parameter vector $\mathbf{x}_k$. Here, we must ensure unit normal vectors by means of a length of one. For this, we can make use of the nonlinear equality constraints:

$$g_{I,e}(\mathbf{x}_k) = \|\mathbf{n}_e\| = \sqrt{n_{e_x}^2 + n_{e_y}^2 + n_{e_z}^2} = b_{I,e} = 1, \quad (36)$$

with $e \in \{1 \dots E\}$.

The nonlinear equality constraint of type g_I applies to each of the E planes normal vectors:

$$g_I(\mathbf{x}_k) = [g_{I,1}(\mathbf{x}_k), \dots, g_{I,E}(\mathbf{x}_k)]^T. \quad (37)$$

The right side of the equal sign is stored in vector $\mathbf{b}_I$:

$$\mathbf{b}_I = [b_{I,1}, \dots, b_{I,E}]^T. \quad (38)$$

With the second type of nonlinear equality constraint g_{II} , we ensure that each vertex of a plane is located in the plane. By this, we preserve the topology of the 3D city model. If we assume that $\mathbf{V}_{m,k}^{\text{glo}}$ is a vertex of plane e , represented by $\mathbf{n}_e$ and d_e , the following equality constraint is applied:

$$g_{II,i}(\mathbf{x}_k) = \mathbf{n}_e^T \cdot \mathbf{V}_{m,k}^{\text{glo}} - d_e = \mathbf{b}_{II,i} = 0 \quad (39)$$

with $i \in \{1 \dots 4 \cdot E\}$.

This type of nonlinear equality constraint must be fulfilled for each vertex of all planes. In general, 3D city model planes have 4 vertices, but it is also possible for them to have 3 or more than 4 vertices. Thus, the number $4 \cdot E$ is just a rough estimate for the total amount of equality constraints of type g_{II} :

$$g_{II}(\mathbf{x}_k) = [g_{II,1}(\mathbf{x}_k), \dots, g_{II,4 \cdot E}(\mathbf{x}_k)]^T. \quad (40)$$

The right side of the equal sign is stored in vector $\mathbf{b}_{II}$:

$$\mathbf{b}_{II} = [b_{II,1}, \dots, b_{II,4 \cdot E}]^T. \quad (41)$$

A comparable procedure can be found in Unger et al. (2016). Altogether, the nonlinear equality constraints for the state parameters are given by:

$$\mathbf{g}(\mathbf{x}_k) = [\mathbf{g}_I(\mathbf{x}_k); \mathbf{g}_{II}(\mathbf{x}_k)] \quad (42)$$

$$\mathbf{b} = [\mathbf{b}_I; \mathbf{b}_{II}]. \quad (43)$$

2.4.6 Initialization

The initial state parameter vector $\mathbf{x}_0$ is created by the GNSS and IMU observations. We assume zero for the velocity in each coordinate component:

$$\mathbf{v}_0 = \mathbf{0}_{[3 \times 1]}. \quad (44)$$

The initial plane parameters $\mathbf{n}_{e,0}$ and $d_{e,0}$ were estimated from the planes' vertices stored in the city model by the Drixler algorithm (Drixler 1993). We make use of the vertices of the 3D city model given for the initial vertices $\mathbf{x}_{V,0}$:

$$\mathbf{x}_{\text{State},0} = [t_{\text{GNSS},0}; \mathbf{o}_{\text{IMU},0}; \mathbf{v}_0] \quad (45)$$

$$\mathbf{x}_{\text{Plane},0} = [\mathbf{n}_{1,0}; d_{1,0}; \dots; \mathbf{n}_{E,0}; d_{E,0}] \quad (46)$$

$$\mathbf{x}_{V,0} = [\mathbf{V}_{1,0}^{\text{glo}}; \dots; \mathbf{V}_{M,0}^{\text{glo}}] \quad (47)$$

$$\mathbf{x}_0 = [\mathbf{x}_{\text{State},0}; \mathbf{x}_{\text{Plane},0}; \mathbf{x}_{V,0}]. \quad (48)$$

We apply the following equations for building the VCM of the initial state parameter vector $\Sigma_{xx,0}$:

$$\Sigma_{xx,\text{State},0} = \text{diag}(\sigma_{t,0}^2, \sigma_{o,0}^2, \sigma_{v,0}^2) \quad (49)$$

$$\Sigma_{xx,\text{Plane},0} = \text{diag}(\sigma_{n,0}^2, \sigma_{d,0}^2, \dots, \sigma_{n,0}^2, \sigma_{d,0}^2) \quad (50)$$

$$\Sigma_{xx,V,0} = \text{diag}(\sigma_{V,0}^2, \dots, \sigma_{V,0}^2) \quad (51)$$

$$\Sigma_{xx,0} = \begin{bmatrix} \Sigma_{xx,\text{State},0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \Sigma_{xx,\text{Plane},0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \Sigma_{xx,V,0} \end{bmatrix}. \quad (52)$$

The standard deviations of the initial state parameters of translation, orientation, and velocity are denoted by $\sigma_{t,0} = [\sigma_{t,0}, \sigma_{t,0}, \sigma_{t,0}]$, $\sigma_{o,0} = [\sigma_{o,0}, \sigma_{o,0}, \sigma_{o,0}]$, and $\sigma_{v,0} = [\sigma_{v,0}, \sigma_{v,0}, \sigma_{v,0}]$. The standard deviations of the initial plane parameters and vertices are denoted by $\sigma_{n,0} = [\sigma_{n,0}, \sigma_{n,0}, \sigma_{n,0}]$, $\sigma_{d,0}$ and $\sigma_{V,0}$.

2.5 Workflow

The workflow of our algorithm is summarized in Algorithm 1. This is an adaption of the algorithm proposed in Vogel et al. (2018). We highlighted the cross references to equations and sections in blue. The partial derivatives (see lines 9, 22, 23, and 40) were obtained once using

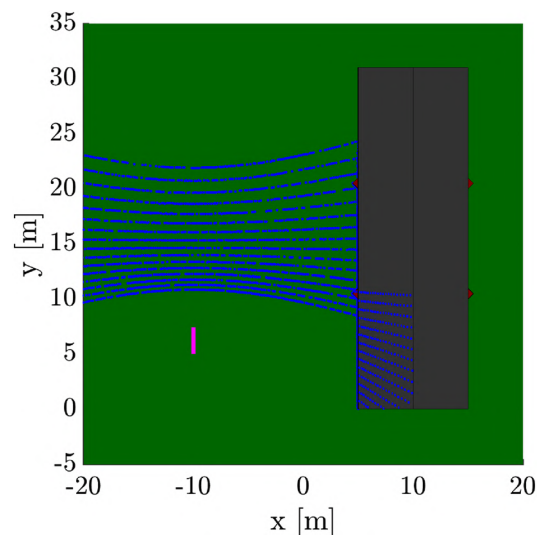


Fig. 6 Top view of the simulation environment with a simulated point cloud (blue dots) of the first epoch and trajectory of the UAS (magenta line)

the symbolic Math toolbox of Matlab[®]. Subsequently, we implemented the partial derivatives in a function. Another possibility to obtain the partial derivatives could be the use of INTLAB (Rump 1999). We fixed the stop criterion c_{stop} to 1×10^{-12} in the measurement update step. This stop criterion was reached after an average of six iterations in the subsequent simulation.

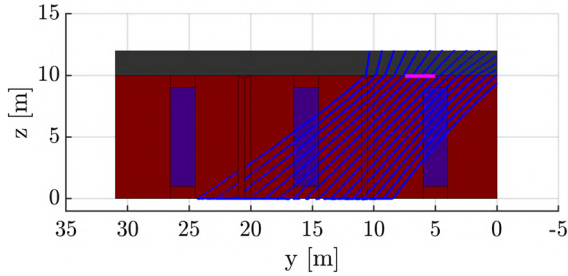


Fig. 7 Side view of the simulation environment with a simulated point cloud (blue dots) of the first epoch and trajectory of the UAS (magenta line)

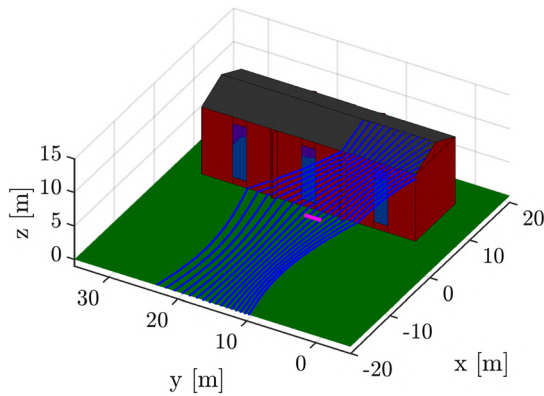


Fig. 8 Oblique view of the simulation environment with a simulated point cloud (blue dots) of the first epoch and trajectory of the UAS (magenta line)

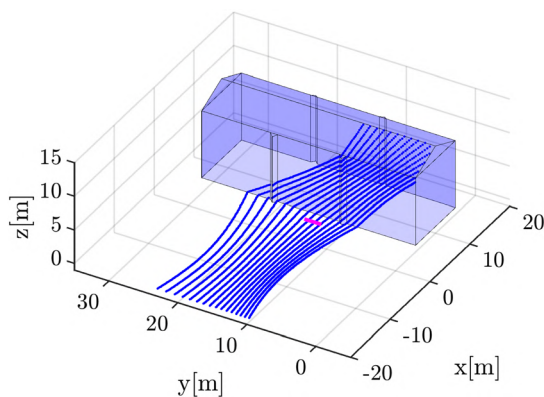


Fig. 9 3D city model with a simulated point cloud (blue dots) of the first epoch and trajectory of the UAS (magenta line)

Algorithm 1: Iterated extended Kalman filter with nonlinear implicit measurement equation and nonlinear equality state constraints

```

1 System model (see Eqs. 20-28):
2  $\mathbf{x}_k = \mathbf{f}(\mathbf{x}_{k-1}, \mathbf{u}_{k-1}) + \mathbf{w}_{k-1}$ ,  $\mathbf{w}_{k-1} \sim N(\mathbf{0}, \Sigma_{ww})$ 
3 Observation model (see Eqs. 29-35):
4  $\mathbf{h}(E(l_k), \mathbf{x}_k) = \mathbf{h}(l_k + \mathbf{v}_k, \mathbf{x}_k) = \mathbf{0}$ ,  $\mathbf{v}_k \sim N(\mathbf{0}, \Sigma_{vv})$ 
5 Initial parameter vector and its VCM (see Eqs. 44-52):
6  $\hat{\mathbf{x}}_0^+ = \mathbf{x}_0$ ,  $\Sigma_{\hat{\mathbf{x}},0}^+ = \Sigma_{xx,0}$ ,  $k = 1$ 
7 while  $k < K$  do
8   Prediction step
9    $\mathbf{F}_{x,k} = \partial \mathbf{f} / \partial \mathbf{x} |_{\hat{\mathbf{x}}_{k-1}^+, \mathbf{u}_{k-1}}$  (see Eqs. 22-24)
10   $\hat{\mathbf{x}}_k^- = \mathbf{f}(\hat{\mathbf{x}}_{k-1}^+, \mathbf{u}_{k-1})$  (see Eqs. 20-21)
11   $\Sigma_{xx,k}^- = \mathbf{F}_{x,k} \Sigma_{xx,k-1}^+ \mathbf{F}_{x,k}^T + \Sigma_{ww}$ 
12  Assignment (see Section 2.3 and Eq. 1)
13   $\mathbf{l}_{Scan,k}^{\text{glo}} = \hat{\mathbf{t}}_k^- + \hat{\mathbf{R}}_k^- \cdot \mathbf{l}_{Scan,k}^{\text{loc}}$ 
14  Assign  $\mathbf{l}_{Scan,k}^{\text{glo}}$  to the planes of the city model
15  Arrange observation vector (see Eqs. 11-19)
16  Store all points, distinctively assigned to a plane,
17  in  $\mathbf{l}_{Scan,k}^{\text{loc}} = [\mathbf{l}_{Scan,1,k}^{\text{loc}}, \dots, \mathbf{l}_{Scan,E,k}^{\text{loc}}]$ 
18  Arrange the observation vector
19   $\mathbf{l}_k = [\mathbf{l}_{Scan,k}^{\text{loc}}, \mathbf{l}_{V,0}, \mathbf{l}_{Pose,k}]^T$  and its VCM
20   $\Sigma_{ll,k}$ 
21  Measurement-update step
22   $\tilde{\mathbf{l}}_{k,0} = \mathbf{l}_k$ ,  $\tilde{\mathbf{x}}_{k,0} = \hat{\mathbf{x}}_k^-$ 
23   $\Delta \mathbf{x} = \infty$ ,  $\Delta \mathbf{l} = \infty$ ,  $m = 0$ 
24  while  $\max(\Delta \mathbf{x}) > c_{\text{stop}} \vee \max(\Delta \mathbf{l}) > c_{\text{stop}}$  do
25     $\mathbf{H}_{x,k,m} = \partial \mathbf{h} / \partial \mathbf{x} |_{\tilde{\mathbf{l}}_{k,m}, \tilde{\mathbf{x}}_{k,m}}$ 
26     $\mathbf{H}_{l,k,m} = \partial \mathbf{h} / \partial \mathbf{l} |_{\tilde{\mathbf{l}}_{k,m}, \tilde{\mathbf{x}}_{k,m}}$ 
27     $\mathbf{O}_{k,m} = \mathbf{H}_{x,k,m} \Sigma_{xx,k}^- \mathbf{H}_{x,k,m}^T$ 
28     $\mathbf{S}_{k,m} = \mathbf{H}_{l,k,m} \Sigma_{vv} \mathbf{H}_{l,k,m}^T$ 
29     $\mathbf{K}_{k,m} = \Sigma_{xx,k}^- \mathbf{H}_{x,k,m}^T \cdot (\mathbf{O}_{k,m} + \mathbf{S}_{k,m})^{-1}$ 
30     $\mathbf{r}_{k,m} =$ 
31     $\mathbf{H}_{l,k,m} \cdot (\mathbf{l}_k - \tilde{\mathbf{l}}_{k,m}) + \mathbf{H}_{x,k,m} \cdot (\tilde{\mathbf{x}}_k^- - \tilde{\mathbf{x}}_{k,m})$ 
32     $\tilde{\mathbf{x}}_{k,m+1} = \tilde{\mathbf{x}}_{k,m} - \mathbf{K}_{k,m} \cdot (\mathbf{h}(\tilde{\mathbf{l}}_{k,m}, \tilde{\mathbf{x}}_{k,m}) + \mathbf{r}_{k,m})$ 
33     $\mathbf{G}_{k,m} = \Sigma_{vv} \mathbf{H}_{l,k,m}^T (\mathbf{O}_{k,m} + \mathbf{S}_{k,m})^{-1}$ 
34     $\tilde{\mathbf{l}}_{k,m+1} = \mathbf{l}_k - \mathbf{G}_{k,m} \cdot (\mathbf{h}(\tilde{\mathbf{l}}_{k,m}, \tilde{\mathbf{x}}_{k,m}) + \mathbf{r}_{k,m})$ 
35     $\Delta \mathbf{x} = \|\tilde{\mathbf{x}}_{k,m+1} - \tilde{\mathbf{x}}_{k,m}\|$ ,  $\Delta \mathbf{l} = \|\tilde{\mathbf{l}}_{k,m+1} - \tilde{\mathbf{l}}_{k,m}\|$ 
36     $m = m + 1$ 
37   $\hat{\mathbf{x}}_k^+ = \tilde{\mathbf{x}}_{k,m}$ 
38   $\hat{\mathbf{l}}_k^+ = \tilde{\mathbf{l}}_{k,m}$ 
39   $\mathbf{L}_k = \mathbf{I}_{j \times j} - \mathbf{K}_{k,m-1} \mathbf{H}_{x,k,m-1}$ 
40   $\Sigma_{\hat{\mathbf{x}},k}^+ = \mathbf{L}_k \Sigma_{xx,k}^- \mathbf{L}_k^T + \mathbf{K}_{k,m-1} \mathbf{S}_{k,m-1} \mathbf{K}_{k,m-1}^T$ 
41   $\mathbf{U}_k = \mathbf{G}_{k,m-1} \cdot \mathbf{H}_{x,k,m-1}$ 
42   $\Sigma_{\hat{\mathbf{v}},k}^+ =$ 
43   $\Sigma_{vv} + \mathbf{G}_{k,m-1} \mathbf{S}_{k,m-1} \mathbf{G}_{k,m-1}^T - \mathbf{U}_k \Sigma_{xx,k}^- \mathbf{U}_k^T$ 
44  Constraint step (see Eqs. 36-43)
45   $\mathbf{D} = \partial \mathbf{g} / \partial \mathbf{x} |_{\hat{\mathbf{x}}_k^-}$ 
46   $\mathbf{d} = \mathbf{b} - \mathbf{g}(\hat{\mathbf{x}}_k^-) + \mathbf{D} \cdot \hat{\mathbf{x}}_k^-$ 
47  Set  $\mathbf{W} = \mathbf{I}_{j \times j}$ 
48   $\hat{\mathbf{x}}_k^+ =$ 
49   $\hat{\mathbf{x}}_k^- - \mathbf{W}^{-1} \mathbf{D}^T (\mathbf{D} \mathbf{W}^{-1} \mathbf{D}^T)^{-1} (\mathbf{D} \hat{\mathbf{x}}_k^- - \mathbf{d})$ 
50   $\Sigma_{\hat{\mathbf{x}},k}^+ =$ 
51   $\Sigma_{\hat{\mathbf{x}},k}^- - \Sigma_{\hat{\mathbf{x}},k}^- \mathbf{D}^T (\mathbf{D} \Sigma_{\hat{\mathbf{x}},k}^- \mathbf{D}^T)^{-1} \mathbf{D} \Sigma_{\hat{\mathbf{x}},k}^-$ 
52  Set  $\hat{\mathbf{x}}_k^+ = \hat{\mathbf{x}}_k^+$  and  $\Sigma_{\hat{\mathbf{x}},k}^+ = \Sigma_{\hat{\mathbf{x}},k}^+$ 

```


Table 1 Assumed systematic effects Δ and standard deviations σ of the added noise for all sensors and observations (Obs.)

Sensor	Obs.	Scenario 1		Scenario 2	
		σ	Δ	σ	Δ
Laser scanner	Wall, roof	0.02 m	0 m	0.02 m	0 m
	Ground	0.02 m	0 m	0.2 m	0 m
	Window	0.02 m	0 m	0.1 m	0.6 m
GNSS	t_x, t_y, t_z	0.5 m	0 m	0.5 m	0 m
IMU	ω, ϕ	0.2°	0°	0.2°	0°
	κ	0.2°	0°	0.2°	0.01° · k

3 Simulation

Computer simulations are a great tool for analyzing and interpreting engineering systems. Here, we intended to evaluate the functionality and performance of the IEKF implemented. Therefore, we focused on two scenarios which basically conform to the testing environment designated for the *UAS Project*. Furthermore, we focused on the challenging initialization phase of the IEKF and analyzed a rather short trajectory of 2.5 m length and with 50 epochs.

Therefore, we created a model of a fictitious building and a ground plane. Subsequently, we determined a fictitious trajectory beside this building. We assumed a constant velocity for the UAS of 1 m s^{-1} and simulated laser scanner measurements with a frequency of 20 Hz. Consequently, the epochs simulated are at constant distances of 0.05 m. Each of the epochs consists of a laser scan and the desired values of position and orientation obtained from the determined trajectory at a certain time.

Figures 6, 7 and 8 depict different views of the building and ground plane created. The building created contains roof planes (in dark grey), wall planes (in red), and window planes (in transparent blue). Figure 9 depicts the corresponding 3D city model of the building. As we can see, the ground plane is not part of the 3D city model. The 3D city model contains the vertices of each plane and is given in the superordinate coordinate system. The simulated and correctly georeferenced point cloud (blue dots) of the first epoch and the trajectory of the UAS (magenta line) is depicted in each figure.

3.1 Scenarios

We simulated two scenarios in which the laser scanner and IMU measurements were generated differently under certain assumptions regarding measurement accuracy and bias. We repeated the simulation 500 times ($S = 500$) for both scenarios. In scenario 1, we only added normally distributed noise on the laser scanner measurements, the position

Table 2 Assumed standard deviations σ for the initial VCM of the state parameters $\Sigma_{xx,0}$, the system noise Σ_{ww} , and the observation vector Σ_{ll}

Initial state noise	$\sigma_{t,0} = 0.5 \text{ m}$
	$\sigma_{o,0} = 0.2^\circ$
	$\sigma_{v,0} = 1 \text{ m s}^{-1}$
	$\sigma_{n,0} = 0.0001$
	$\sigma_{d,0} = 0.001 \text{ m}$
	$\sigma_{V,0} = 0.0001 \text{ m}$
System noise	$\sigma_{t,w} = 3 \cdot \Delta \tau \text{ m}$
	$\sigma_{o,w} = 3 \cdot \Delta \tau^\circ$
	$\sigma_{v,w} = 5 \cdot \Delta \tau \text{ m s}^{-1}$
	$\sigma_{n,w} = 0$
	$\sigma_{d,w} = 0 \text{ m}$
	$\sigma_{V,w} = 0 \text{ m}$
	$\sigma_{LS} = 0.02 \text{ m}$
Measurement noise	$\sigma_t = 0.5 \text{ m}$
	$\sigma_o = 0.2^\circ$
	$\sigma_v = 0.0001 \text{ m}$

(representing the GNSS receiver), and the orientation (representing the IMU). In scenario 2, we systematically perturbed the laser scanner measurements, hitting the windows of the simulated building, and increased the (standard deviation of the) noise for laser scanner measurements, hitting the windows and the ground. The systematic and increased random disturbances of the measurements hitting the windows simulate the infiltration behavior for glass. We included the increased measurement uncertainty and the actual infiltration of the laser into the increased noise of these measurements. The increased noise for measurements hitting the ground simulates possible unevenness and vegetation. Thus, this increased noise is more justified by actual structures in the object space than by an increased uncertainty of the measurement. Furthermore, we added a drift on the κ -component of the IMU. Therefore, we linked the systematic effect Δ to the epoch number $k \in \{1, \dots, K\}$. The κ -component conforms to the heading of the UAS. All assumed systematic effects Δ and standard deviations σ of the added noise are depicted in Table 1. Basically, the assumptions of scenario 1 should be consistent with the manufacturer's specifications of the sensors used or planned for our UAS. The assumptions in scenario 2, especially the ones for the laser scanner, are based on experience from test measurements. In our simulation, we assume that the positions and orientations of all sensors in a platform coordinate system or body frame have already been determined in a calibration process. We also assume that the sensors are properly synchronized. To neglect the effect of generalization in the 3D city model, we used the same model for simulation and in the following IEKF algorithm.

3.2 Values Chosen for the IEKF

We used a constant distance threshold $d_{\text{assign}} = 0.3$ m for the assignment algorithm (see Sect. 2.3) in our simulation. This chosen threshold is a trade-off between the rather imprecise pose in the first epochs and the more precise pose in the last epochs.

The standard deviations used for building the initial VCM $\Sigma_{xx,0}$ (see Eqs. 49–52) are depicted at the top of Table 2. The standard deviations of the initial position $\sigma_{t,0}$ and orientation $\sigma_{o,0}$ conform to the standard deviations of the GNSS and IMU observations. We assume a standard deviation for the initial velocity $\sigma_{v,0}$ which includes the planned maximal velocities of the UAV. For the standard deviations of the initial plane parameters $\sigma_{n,0}$ and $\sigma_{d,0}$, we met the assumption that these values should be very small. Thus, they correspond roughly to the standard deviation of the initial vertices $\sigma_{v,0}$, which, in turn, conform to the standard deviation chosen for the observation vector. If the accuracies of the planes' vertices are known, the standard deviations of the plane parameters could also be determined by means of variance propagation.

The standard deviations chosen for the system noise (see Eqs. 25–28) can be found in the middle of Table 2. We set the standard deviation for translation $\sigma_{t,w}$, orientation $\sigma_{o,w}$ and velocity $\sigma_{v,w}$ depending on the time period $\Delta\tau$ between two epochs and regarding possible unpredictable movements of the UAS in this time period. We set the system noise to zero for $\Sigma_{ww,\text{Plane}}$ and $\Sigma_{ww,V}$, because these parameters are constant over time. Consequently, $\sigma_{n,w}$, $\sigma_{d,w}$, and $\sigma_{v,w}$ are zero.

We used the simulated standard deviation $\sigma_{LS} = 0.02$ m of the laser scanner measurements for the stochastic model of the observation vector (see Eq. 19). We also introduced the simulated standard deviations $\sigma_t = 0.5$ m and $\sigma_o = 0.2^\circ$ for the GNSS and IMU observations. In the case of the initial vertices of the 3D building model, we introduced a small standard deviation of $\sigma_v = 0.1$ mm. The 3D city model is, thus, almost introduced as a fixed feature. We are aware that this assumption is too optimistic and that the accuracy of

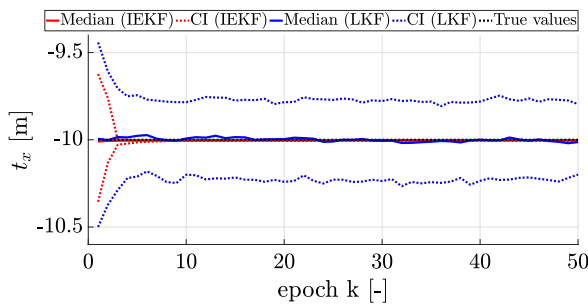


Fig. 10 Median and confidence intervals of t_x in scenario 1

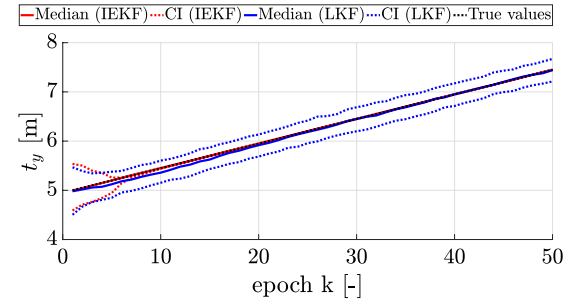


Fig. 11 Median and confidence intervals of t_y in scenario 1

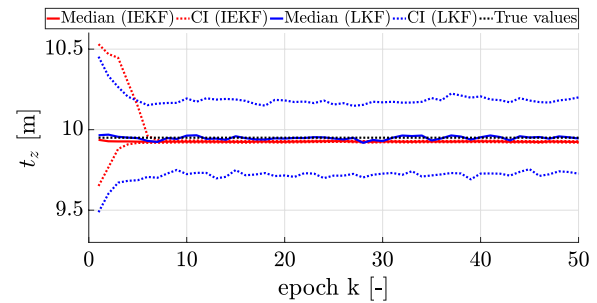


Fig. 12 Median and confidence intervals of t_z in scenario 1

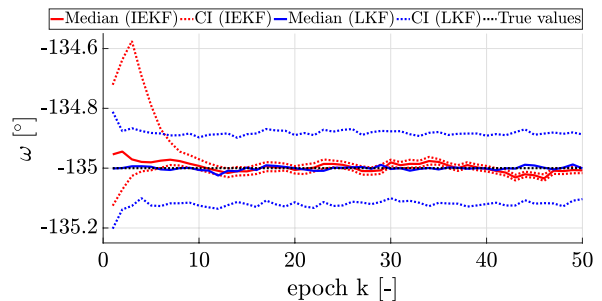


Fig. 13 Median and confidence intervals of ω in scenario 1

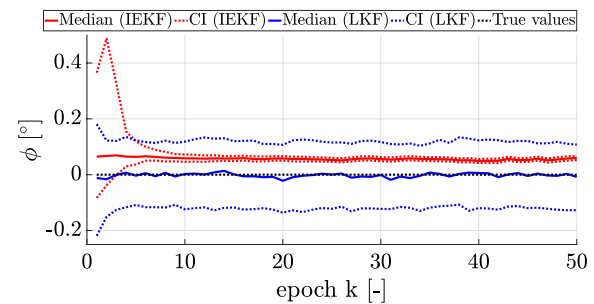


Fig. 14 Median and confidence intervals of ϕ in scenario 1

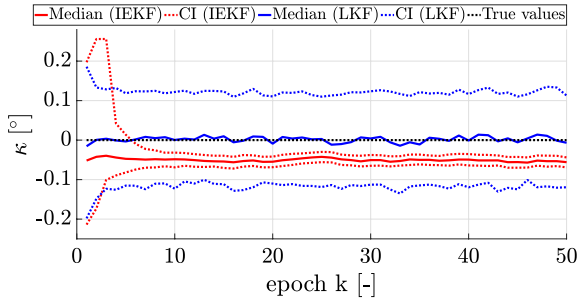


Fig. 15 Median and confidence intervals of κ in scenario 1

the introduced 3D city model is rather in the range of a few centimeters. However, in the context of this paper, the influence of a less precise 3D city model will not be considered in more detail.

We assume that $\Sigma_{ll,k}$ and Σ_{ww} do not contain correlations on their own and among themselves.

4 Results and Discussion

In the following, we depict the results of the IEKF for both scenarios and compare them with the results of a linear Kalman filter (LKF), which only uses the simulated GNSS and IMU observations. The comparison between LKF and IEKF evaluates the benefit of introducing laser scanner observations and a city model. The state parameter vector $\mathbf{x}_k$ of the LKF solely consists of $\mathbf{x}_{\text{State},k}$ (Eq. 6). The observation vector $\mathbf{l}_k$ consists of $\mathbf{l}_{\text{Pose},k}^{\text{glo}}$ (Eq. 15). The transition matrix $\mathbf{F}_{x,k}$ of the LKF is given by $\mathbf{F}_{x,\text{State},k}$ (Eq. 24). Consequently, the VCM of the observation vector and of the system noise have to be adapted. The LKFs' only remaining measurement equation (Eq. 31) has to be transformed in explicit form. For the LKF, no linearization is necessary and no constraints are used. For more information concerning the LKF, see (Simon 2010). Figures 10, 11, 12, 13, 14 and 15 depict the median values (solid lines) and 68% (1σ) confidence intervals (CI) (dashed lines) for the translation parameters t_x , t_y , t_z (Figs. 10, 11, 12) and orientation parameters ω , ϕ , κ (Figs. 13, 14, 15) in scenario 1. The CI are calculated numerically using the total number of simulated runs S according to Alkhatib et al. (2009). The red lines result from the filtered state parameters $\hat{\mathbf{x}}_k^+$ of the IEKF after the constraint step. The blue lines result from the filtered state parameters $\hat{\mathbf{x}}_k^+$ of the LKF, respectively. The true values are plotted in a dashed black line.

The median of the IEKF result for t_x (across flight direction) in Fig. 10 is very close to the true translation, whereas we can see slight random deviations between the median of the LKF results and the true values. The latter can be seen for each pose component and is not explicitly mentioned

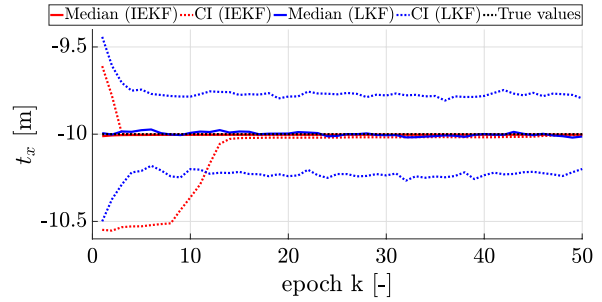


Fig. 16 Median and confidence intervals of t_x in scenario 2

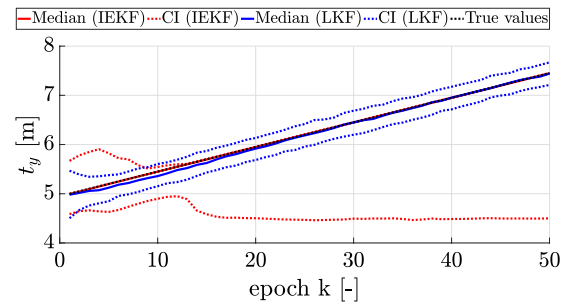


Fig. 17 Median and confidence intervals of t_y in scenario 2

again in the following. The CI of the IEKF and the LKF shrink rapidly in the first 5 epochs. Whereas the CI of the IEKF is very close to the median, the CI of the LKF is significantly broader. The latter can be seen for each pose component and is also not explicitly mentioned again. The IEKF performs well for that pose component because of the measurement constellation. Each measurement distinctively assigned to a plane provides information for that pose component. In other words, each measured plane of the 3D city model is sensitive for that pose component.

The results are similar for t_y (in flight direction) in Fig. 11. Again, the median of the IEKF results is very close to the true translation and its CI shrinks within the first epochs. For this pose component, the shrinking of the CI (IEKF) is slightly slower than for t_x and lasts until epoch 10. A reason for that slight difference can be found in the weaker measurement constellation for that pose component. Only the triangle-shaped protrusions are sensitive for t_y . Consequently, there are significantly fewer observations which provide information for that pose component.

Figure 12 shows the results for t_z (up-direction). The median of the IEKF results is very constant, but seems to deviate from the true values by a small systematic shift of about 3 cm. This systematic shift is caused by ground points wrongly assigned to a wall plane of the building. This results in a slight rotation around the y-axis (see results for ϕ) in combination with a slight descent of t_z . Again, the CI of the

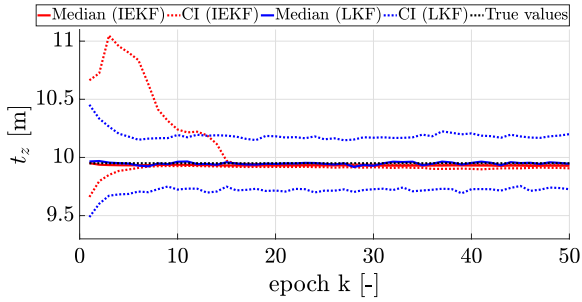


Fig. 18 Median and confidence intervals of t_z in scenario 2

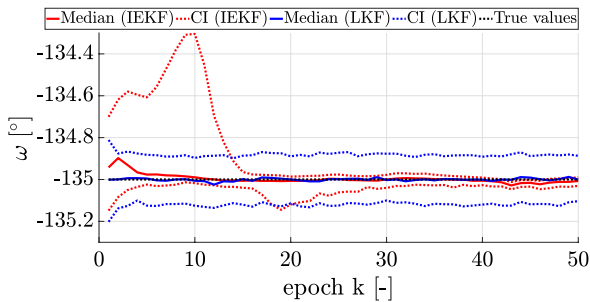


Fig. 19 Median and confidence intervals of ω in scenario 2

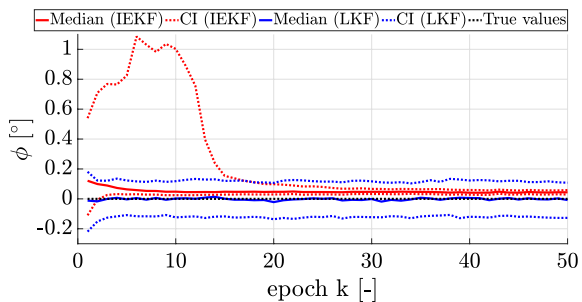


Fig. 20 Median and confidence intervals of ϕ in scenario 2

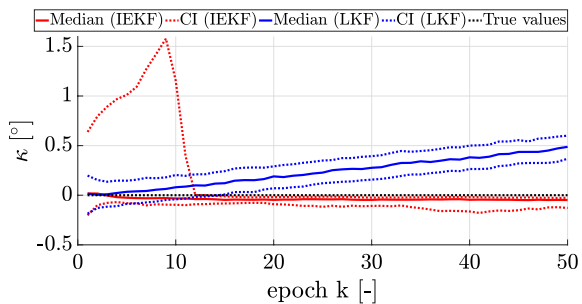


Fig. 21 Median and confidence intervals of κ in scenario 2

IEKF translation shrinks rapidly within the first 6 epochs. The boundaries of the CI in the first epochs are not symmetrical around the median. They are slightly shifted to larger t_z values. That means that the IEKF converges slower towards the true translation with initial t_z values larger than the true translation. An initial t_z value larger than the true translation may cause significantly fewer assignments in the first epochs. Only the roof planes are sensitive for that pose component.

The results for ω , which represents the rotation around the x -axis, are shown in Fig. 13. After 10 epochs, the median of IEKF and LKF obtain comparable results, whereas the median of the IEKF seems to vary slightly more. The CI of the IEKF shrinks within the first 12 epochs. Similar to the t_z pose component, there is a shift of the IEKF CI towards larger ω values. The only sensitive planes for that pose component are the triangle-shaped protrusions and the roof planes. This may explain the larger variations in the median of the IEKF orientation.

Figures 14, 15 show the results for ϕ (rotation around y -axis) and κ (rotation around z -axis), respectively. We can see similar characteristics for these pose components. In both cases, the median of the IEKF results is constant but systematically shifted by a small value of about 0.05° . As has already been mentioned for t_z , the systematic shift in ϕ is caused by wrongly assigned ground points. Because of the skewed alignment of the laser scanner, the wrongly assigned ground points also cause a slight rotation in κ . All planes are sensitive for both pose components.

For scenario 2, we depicted the results in a similar way in Figs. 16, 17, 18, 19, 20 and 21. In Fig. 16, we can see the results for t_x . Similar to scenario 1, the median of the IEKF results comes very close to the true translation. The lower boundary of the CI (IEKF) is farther afiel from the median than the upper boundary. It shrinks until epoch 14. An explanation for the one-sided enlarged CI (IEKF) is the systematically extended laser scanner measurements hitting the windows. With an initial t_x translation, which is farther afiel from the building than the true translation, these systematically extended laser scanner measurements are wrongly assigned to the walls of the building. After some epochs, these false assignments seem to have disappeared. These effects are analyzed later on (see Fig. 29).

The median of the IEKF results for t_y in Fig. 17 is close to the true translation. The upper boundary of the CI (IEKF) shrinks until epoch 13, whereas the lower boundary of the CI (IEKF) clearly departs from the median from epoch 14 on. Again, these effects are analyzed later on (see Figs. 28 and 29).

We can see the results for t_z in scenario 2 in Fig. 18. The results are comparable to the ones obtained in scenario 1 for that pose component. A significant difference can be found in the upper boundary of the CI (IEKF), which

Table 3 RMSE of position and orientation in scenario 1 (first row) and 2 (second row)

Criterion	RMSE _{t_x,R} [m]		RMSE _{t_y,R} [m]		RMSE _{t_z,R} [m]		RMSE _{ω,R} [°]		RMSE _{φ,R} [°]		RMSE _{κ,R} [°]	
	IEKF	LKF	IEKF	LKF	IEKF	LKF	IEKF	LKF	IEKF	LKF	IEKF	LKF
Minimum	0.0033	0.1095	0.0018	0.1121	0.0225	0.1106	0.0112	0.0654	0.0489	0.0629	0.0456	0.0622
	0.0024		0.0023		0.0166		0.0122		0.0384		0.0333	0.1867
Maximum	2.1173	0.2957	3.4644	0.3348	3.0158	0.3085	1.8773	0.1407	2.5912	0.1426	2.1497	0.1466
	2.1955		2.9587		3.2166		2.6897		2.9243		2.3307	0.3233
Mean	0.0654	0.1899	0.1955	0.1941	0.1166	0.1911	0.0694	0.0964	0.1168	0.0973	0.1284	0.0968
	0.1080		0.4496		0.1829		0.1648		0.2008		0.2288	0.2535
Median	0.0079	0.1864	0.0177	0.1898	0.0343	0.1897	0.0245	0.0966	0.0621	0.0980	0.0556	0.0968
	0.0128		0.0424		0.0417		0.0519		0.0727		0.0585	0.2510
SD	0.2517	0.0337	0.5697	0.0362	0.3254	0.0344	0.1578	0.0130	0.2348	0.0123	0.2536	0.0135
	0.3069		0.7712		0.4385		0.2972		0.2933		0.2924	0.0245
↓ CI (68%)	0.0037	0.1561	0.0041	0.1598	0.0246	0.1566	0.0162	0.0834	0.0556	0.0849	0.0501	0.0838
	0.0030		0.0048		0.0194		0.0179		0.0434		0.0404	0.2311
↑ CI (68%)	0.0355	0.2271	0.1010	0.2283	0.0901	0.2254	0.0753	0.1090	0.1050	0.1091	0.1150	0.1091
	0.1498		1.7169		0.2881		0.2400		0.2707		0.4953	0.2783
↓ CI (95%)	0.0034	0.1306	0.0023	0.1320	0.0234	0.1267	0.0136	0.0735	0.0531	0.0723	0.0475	0.0725
	0.0026		0.0028		0.0178		0.0145		0.0402		0.0361	0.2053
↑ CI (95%)	0.7583	0.2614	2.4217	0.2725	0.8801	0.2634	0.5240	0.1229	0.7461	0.1187	1.0262	0.1275
	1.5383		2.3080		1.4151		1.0929		1.0209		0.9829	0.3064
Best RMSE	95.0%	5.0%	89.8%	10.2%	90.2%	9.8%	87.6%	12.4%	80.2%	19.8%	80.6%	19.4%
	90.2%	9.8%	68.2%	31.8%	81.2%	18.8%	61.2%	38.8%	57.4%	42.6%	65.4%	34.6%

converges to the median much slower than in scenario 1. In scenario 2, it does not converge until epoch 15. Again, initial t_z values larger than the true position cause significantly fewer assignments in the first epochs.

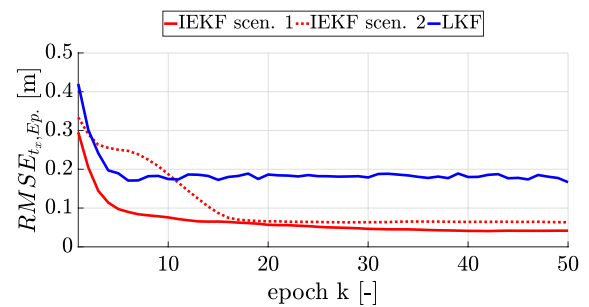
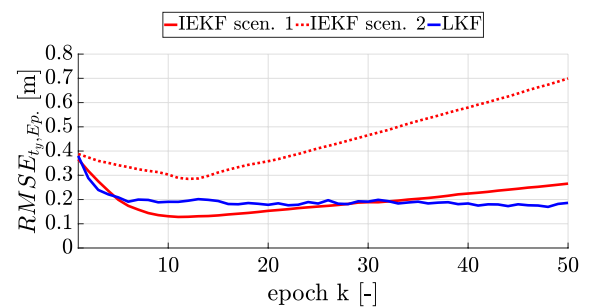
We observe comparable results to scenario 1 for the pose parameters ω and ϕ in Figs. 19 and 20. In both figures, we can observe that one boundary of the CI (IEKF) is farther afield from the median than the other and that the CI (IEKF) shrinks slower than in scenario 1.

The pose component κ is depicted in Fig. 21. The CI (IEKF) shrinks until epoch 12. In addition, we can see that the simulated drift in κ does not affect the IEKF results, whereas the median of the LKF departs from the true orientation continuously as simulated.

A suitable value to evaluate the performances of the filters is the root-mean-square error (RMSE). In our simulation, the RMSE is the error between filtered state parameters and the true state parameters. The RMSE of the pose parameter t_x for run $s \in \{1, \dots, S\}$ is calculated according to:

$$\text{RMSE}_{t_x, R, s} = \frac{1}{K} \sum_{k=1}^K \sqrt{(t_{x,s,k} - \bar{t}_{x,k})^2}. \quad (53)$$

$t_{x,s,k}$ results from the filtered state parameter vector $\hat{\mathbf{x}}_k^+$ from the IEKF or the LKF, respectively. $\bar{t}_{x,k}$ are the true values.

**Fig. 22** RMSE_{t_x,Ep.}**Fig. 23** RMSE_{t_y,Ep.}

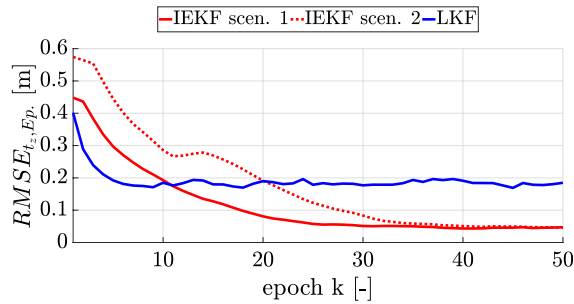
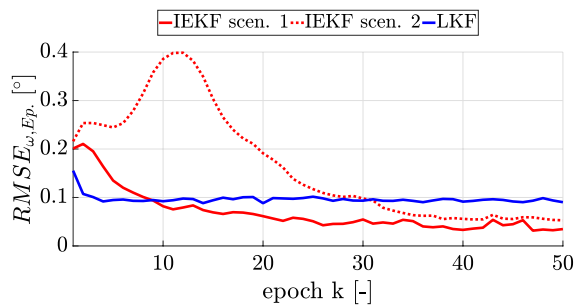
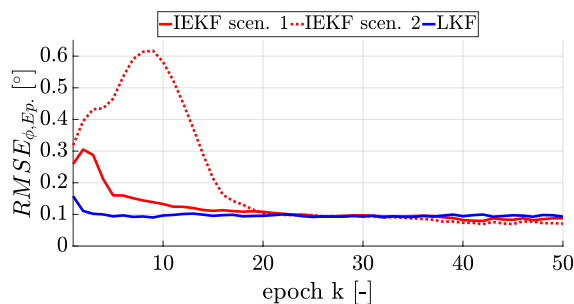
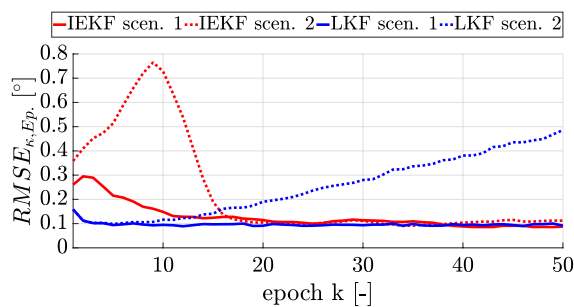
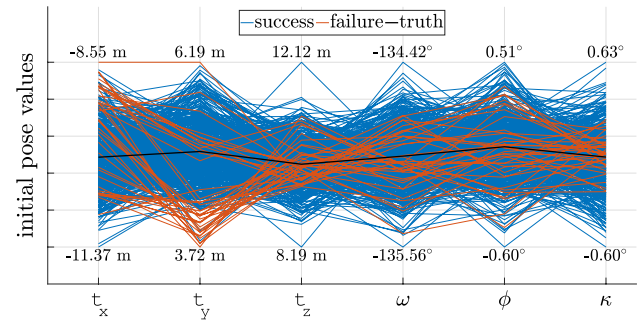
Fig. 24 $RMSE_{t_z, Ep}$.Fig. 25 $RMSE_{\omega, Ep}$.Fig. 26 $RMSE_{\phi, Ep}$.Fig. 27 $RMSE_{\kappa, Ep}$.

Fig. 28 Initial pose values plotted as parallel coordinates in scenario 1. Total failure rate: 7.6%

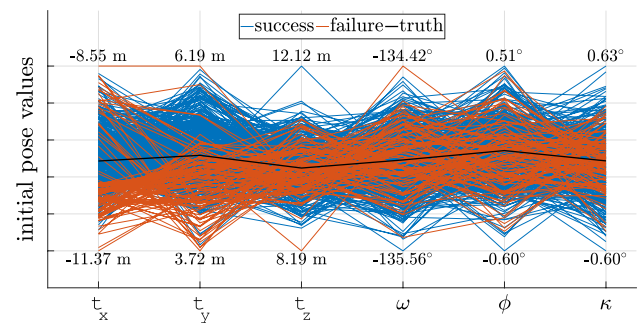


Fig. 29 Initial values plotted as parallel coordinates in scenario 2. Total failure rate: 20.4%

The calculation of the RMSE for the other pose parameters is carried out analogously.

Table 3 depicts the characteristic criteria: Minimum, maximum, mean, median, standard deviation (SD), and the lower bound ($\downarrow$) and upper bound ($\uparrow$) of the 68% and 95% CI of different RMSE. In the last row of Table 3, we display the rate of runs where the IEKF obtained a smaller RMSE than LKF and vice versa. Each criterion is divided into two rows regarding scenario 1 (above) and scenario 2 (below).

Table 3 clearly shows that the IEKF obtains better results than the LKF. The median values especially are significantly smaller in each state parameter and each scenario. By contrast, the range of the IEKF results is significantly larger than the range of the LKF results, which can be seen in the SD and 95% CI values. A possible explanation can be found in a larger number of runs, where the IEKF is far away from the true values. We will subsequently call these runs *failures*. These *failures* can be caused by a disadvantageous initial pose parameter in the first epoch. By this, a large number of laser scanner measurements are assigned to wrong planes and the IEKF *converges* to a wrong pose. The interaction between initial pose parameters and performance of the IEKF is analyzed at the end of this section. The *failures* are also the explanation for the large discrepancy between the mean and median values.

To evaluate the performance of the IEKF over the epochs, we calculated a mean RMSE for each epoch over all S simulation runs. The calculation is as follows for the RMSE of t_x in epoch $k \in \{1, \dots, K\}$:

$$\text{RMSE}_{t_x, \text{Ep}, k} = \frac{1}{S} \sum_{s=1}^S \sqrt{(t_{x,s,k} - \bar{t}_{x,k})^2}. \quad (54)$$

We obtain the RMSE for the other pose parameters analogously. The results can be found in Figs. 22, 23, 24, 25, 26 and 27.

Figures 22, 23, 24, 25, 26 and 27 generally support the interpretations made before. The RMSE of the translation parameters t_x and t_z determined by the IEKF decrease with an increasing number of epochs. At the last epoch, the $\text{RMSE}_{t_x, \text{Ep}}$ and $\text{RMSE}_{t_z, \text{Ep}}$ of the IEKF is significantly smaller than the RMSE obtained by LKF in both scenarios. Again, $\text{RMSE}_{t_y, \text{Ep}}$ differs and increases after epoch 10 with an increasing number of epochs. In both scenarios, especially in scenario 2, the $\text{RMSE}_{t_y, \text{Ep}}$ obtained by the IEKF is larger than the one obtained by LKF in the last epochs. The RMSE of the orientation parameter show different characteristics, but they are intrinsically quite similar. The RMSE obtained by IEKF starts with large values and significantly decreases after epoch 10. In the last epoch, $\text{RMSE}_{\phi, \text{Ep}}$ and $\text{RMSE}_{\kappa, \text{Ep}}$ of IEKF and LKF are quite similar. $\text{RMSE}_{\omega, \text{Ep}}$ of the IEKF undercuts the $\text{RMSE}_{\omega, \text{Ep}}$ of the LKF until epoch 9 (scenario 1), resp. epoch 30 (scenario 2). The $\text{RMSE}_{\kappa, \text{Ep}}$ of LKF in scenario 2 is clearly affected by the simulated IMU drift. It is remarkable that (except in $\text{RMSE}_{t_y, \text{Ep}}$) the RMSE of the IEKF between scenario 1 and scenario 2 is very similar in the last epoch.

For a deeper understanding of the link between initialization and performance of the IEKF, we plotted the initial values of GNSS and IMU of the first epoch in parallel coordinates. The parallel coordinate representation was first used by Inselberg (1985). For this representation, we will show all pose components in six parallel axes. On each axis, all initial values belonging to this component are depicted. This representation is used to identify whether some extreme initial pose values could have had a large impact on the divergence of the IEKF. In addition, we colored the runs where the IEKF failed (*failure*) in orange and the others (*success*) in blue. The distinction in *failure* and *success* was selected based on the RMSE of the position in the last epoch. The run is classified as a *failure* when there is a value larger than 10 cm.

In scenario 1 (Fig. 28), the failure rate is 7.6%. For the initial pose parameters t_z , ω , ϕ , and κ , the *failures* are distributed over the whole value range. For t_x , the *failures* are slightly concentrated in a range between -9.2 and -8.55 m and a second range around -10.6 m. For t_y , the *failures* are even more concentrated in a range between 3.72 and 4.3 m.

Initial values for t_y in a range between 3.72 and 4.3 m lead to most of the *failures*. As previously mentioned, we have a weak measurement constellation for t_y . Furthermore, the trajectory is parallel to the y -axis with a true value of $t_y = 7.5$ m in epoch 50, the consequence being that with initial values for t_y between the true value (5 m) and 6.19 m, the IEKF may get back to the true values in a later epoch. This possibility is not given for initial values for t_y between 3.72 and 4.3 m.

In scenario 2, the failure rate of 20.4% is significantly higher. Again, for the initial pose parameters t_z , ω , ϕ , and κ , the *failures* are distributed over the whole value range. For t_y , the *failures* are mainly in a range between 3.72 and 5 m. The reason for that has already been explained for scenario 1. For t_x , we can see a strong concentration of *failures* in a range between -11.37 and -10.4 m. Thus, especially that range for t_x causes the *failures* in scenario 2. The reason for that can be found in the systematically perturbed laser scanner measurements hitting the windows. These measurements are extended by an additive value $\Delta = 0.6$ m. In combination with a disadvantageous initialization, these perturbed measurements are wrongly assigned to the wall plane and, therefore, cause the *failures*.

We used GNSS and IMU observations in the measurement update step for all the displayed results in this section. We performed the same experiment without using the GNSS and IMU observations in the measurement update step. The differences in the results are insignificantly small. Therefore, we do not show these results here.

5 Conclusions and Outlook

In this paper, we presented the mathematical basics and the workflow of an IEKF which makes it possible to determine the trajectory of a UAS better than 5 cm in position (median value) and 0.08° for orientation (median value). It is not mandatory to have continuous GNSS and IMU observations for the implemented IEKF. The trajectory is mainly obtained using laser scanner measurements of building facades, which are modeled as planes in a 3D city model. The laser scanner measurements and the planes of the 3D city model are combined by implicit measurement equations and nonlinear equality constraints within the IEKF.

To demonstrate the functionality and performance of the IEKF implemented, we developed a simulator, which showed that the IEKF is even suited to handle systematically perturbed observations.

Nevertheless, the algorithm demonstrated may be tuned to deal with disadvantageous initial values. A possible starting point can be found in the threshold of the assignment algorithm. Instead of a constant value, this value should be varied regarding the estimated accuracy of the predicted position.

In the future, we plan to evaluate the performance of the IEKF based on real data, especially regarding effects of generalization of the 3D city model and effects of a 3D city model with larger uncertainties. Thus, we can also check whether the assumptions made for the simulation (see Table 1) are also valid in reality. In particular, the performance of the GNSS receiver can deviate strongly from the simulated data with the same precision.

In the IEKF presented, the plane parameters and vertices are introduced as parameters to preserve the topology of the 3D city model. With an increasing number of planes and vertices, the computation becomes inefficient. Here, we plan to reorganize the IEKF presented by means of a dual-state Kalman filter.

As has already been mentioned, we plan to integrate camera measurements into the IEKF. This enables the stabilization of the trajectory in the long term and the improvement of the IEKF's initial behavior. Therefore, we have to integrate the collinearity equations into our approach. Subsequently, we have to integrate the object coordinates of the tie points into the state parameter vector. This emphasizes the need to use a dual-state Kalman filter. By integrating the camera measurements⁸ into the IEKF, we must deal with observations captured in different epochs: one or more images observing a tie point were taken in a past epoch and only one image is taken in the current epoch.

Finally, we aim to analyze the benefit of integrating further geometries in addition to the planes of the 3D city model. There are many cylindrical-shaped objects in urban areas, such as lantern, traffic lights or street signs, which may be used in the IEKF.

Acknowledgements This work was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation)—NE 1453/5-1, HE 1822/37-1, and as part of the Research Training Group i.c.sens [RTG 2159]. The computations were performed by the compute cluster, which is funded by the Leibniz University Hannover, the Lower Saxony Ministry of Science and Culture (MWK), and DFG.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Abmayr T, Härtl F, Hirzinger G, Burschka D, Fröhlich C (2008) A correlation based target finder for terrestrial laser scanning. *J Appl Geod* 2(3):429. <https://doi.org/10.1515/JAG.2008.015>
- Alkhatib H, Neumann I, Kutterer H (2009) Uncertainty modeling of random and systematic errors by means of Monte Carlo and fuzzy techniques. *J Appl Geod* 3(2):136. <https://doi.org/10.1515/JAG.2009.008>
- Besl PJ, McKay ND (1992) A method for registration of 3-D shapes. *IEEE Trans Pattern Anal Mach Intell* 14(2):239–256. <https://doi.org/10.1109/34.121791>
- Dang T (2007) Kontinuierliche Selbstkalibrierung von Stereokameras. Schriftenreihe/Institut für Mess- und Regelungstechnik, Karlsruher Institut für Technologie vol Nr. 008. Univ.-Verl. Karlsruhe, Karlsruhe
- Dang T (2008) An iterative parameter estimation method for observation models with nonlinear constraints. *Metrol Meas Syst* 15(4):421–432
- Denham WF, Pines S (1966) Sequential estimation when measurement function nonlinearity is comparable to measurement error. *AIAA J* 4(6):1071–1076. <https://doi.org/10.2514/3.3606>
- Dennig D, Bureick J, Link J, Diener D, Hesse C, Neumann I (2017) Comprehensive and highly accurate measurements of crane runways, profiles and fastenings. *Sensors* (Basel, Switzerland). <https://doi.org/10.3390/s17051118>
- Drixler E (1993) Analyse der Form und Lage von Objekten im Raum. Deutsche Geodätische Kommission. Reihe C. Dissertationen vol Heft Nr. 409. Beck, München
- Elkhrachy I, Niemeier W (2006) Stochastic assessment of terrestrial laser scanner. In: Proceedings of the ASPRS annual conference, Reno, United States of America
- Ettlinger A, Neuner H, Burgess T (2018) Development of a Kalman filter in the Gauss-Helmert model for reliability analysis in orientation determination with smartphone sensors. *Sensors* (Basel, Switzerland). <https://doi.org/10.3390/s18020414>
- Forster C, Lynen S, Kneip L, Scaramuzza D (2013) Collaborative monocular slam with multiple micro aerial vehicles. In: 2013 IEEE/RSJ International conference on intelligent robots and systems (IROS), pp 3962–3970. IEEE, Piscataway, NJ. <https://doi.org/10.1109/IROS.2013.6696923>
- Glira P, Pfeifer N, Briesse C, Ressel C (2015) Rigorous strip adjustment of airborne laserscanning data based on the ICP algorithm. *ISPRS Ann Photogramm Remote Sens Spat Inf Sci* 2:73–80. <https://doi.org/10.5194/isprsannals-II-3-W5-73-2015>
- Gröger G, Kolbe TH, Nagel C, Häfele KH (2012) OGC City Geography Markup Language (CityGML) encoding standard, 2.0.0 edn. Open Geospatial Consortium, Wayland
- Hartmann J, Trusheim P, Alkhatib H, Paffenholz JA, Diener D, Neumann I (2018) High accurate pointwise (geo-)referencing of a k-tls based multi-sensor-system. *ISPRS Ann Photogramm Remote Sens Spat Inf Sci* 4:81–88. <https://doi.org/10.5194/isprs-annals-IV-4-81-2018>
- Hebel M, Arens M, Stilla U (2009) Utilization of 3D city models and airborne laser scanning for terrain-based navigation of helicopters and UAVS. *Int Arch Photogramm Remote Sens Spat Inf Sci* 38-3/W4:187–192
- Hol JD, Schön TB, Luinge H, Slycke PJ, Gustafsson F (2007) Robust real-time tracking by fusing measurements from inertial and vision sensors. *J Real Time Image Process* 2(2):149–160. <https://doi.org/10.1007/s11554-007-0040-2>
- Holst C, Kuhlmann H, Paffenholz JA, Neumann I (2015) Tls im statischen, stop & go sowie kinematischen einsatz. In: Terrestrisches Laserscanning 2015 (TLS 2015), Beiträge zum ... DVW-Seminar, vol 147. Wißner-Verlag [Erscheinungsort nicht ermittelbar]
- Inselberg A (1985) The plane with parallel coordinates. *Vis Comput* 1(2):69–91. <https://doi.org/10.1007/BF01898350>
- Jutzi B, Weinmann M, Meidow J (2013) Improved uav-borne 3D mapping by fusing optical and laserscanner data. *ISPRS Int Arch*

⁸ We do not plan to use stereo cameras.

- Photogramm Remote Sens Spat Inf Sci 2:223–228. <https://doi.org/10.5194/isprsarchives-XL-1-W2-223-2013>
- Kaul L, Zlot R, Bosse M (2015) Continuous-time three-dimensional mapping for micro aerial vehicles with a passively actuated rotating laser scanner. *J Field Robot*. <https://doi.org/10.1002/rob.21614>
- Lee KW, Wijesoma S, Guzmán JI (2007) A constrained SLAM approach to robust and accurate localisation of autonomous ground vehicles. *Robot Auton Syst* 55(7):527–540. <https://doi.org/10.1016/j.robot.2007.02.004>
- Li-Chee-Ming J, Armenakis C (2013) Determination of UAS trajectory in a known environment from FPV video. *Int Arch Photogramm Remote Sens Spat Inf Sci* 40-1/W2:247–252
- Luhmann T (2013) Close range photogrammetry: 3D imaging techniques. De Gruyter, Berlin
- Nguyen V, Harati A, Martinelli A, Siegwart R, Tomatis N (2006) Orthogonal SLAM: a step toward lightweight indoor autonomous navigation. In: 2006 IEEE/RSJ International conference on intelligent robots and systems, pp 5007–5012. <https://doi.org/10.1109/IROS.2006.282527>
- Nüchter A, Borrmann D, Koch P, Kühn M, May S (2015) A man-portable, IMU-free mobile mapping system. *ISPRS Ann Photogramm Remote Sens Spat Inf Sci* 3:17–23. <https://doi.org/10.5194/isprs-annals-II-3-W5-17-2015>
- Paffenholz JA (2012) Direct geo-referencing of 3D point clouds with 3D positioning sensors. In: Zugl.: Hannover, Univ., Diss., 2012, Wissenschaftliche Arbeiten der Fachrichtung Geodäsie und Geoinformatik der Leibniz Universität Hannover, vol 302. Univ, Hannover
- Rump S (1999) INTLAB—INTERVAL LABORATORY. In: Csendes T (ed) Developments in reliable computing. Kluwer Academic Publishers, Dordrecht, pp 77–104. <http://www.ti3.tuhh.de/rump/>
- SBG Systems (2019) Miniature high performance inertial sensors: Elipse 2 series. https://www.sbg-systems.com/wp-content/uploads/Ellipse_Series_Leaflet.pdf. Accessed 14 May 2019
- Schuhmacher S, Böhm J (2005) Georeferencing of terrestrial laser scanner data for applications in architectural modeling. <https://doi.org/10.18419/opus-3749>
- Simon D (2010) Kalman filtering with state constraints: a survey of linear and nonlinear algorithms. *IET Control Theory Appl* 4(8):1303–1318. <https://doi.org/10.1049/iet-cta.2009.0032>
- Soloviev A, Bates D, van Graas F (2007) Tight coupling of laser scanner and inertial measurements for a fully autonomous relative navigation solution. *Navigation* 54(3):189–205
- Steffen R (2013) A robust iterative Kalman filter based on implicit measurement equations. *Photogramm Fernerkundung Geoinf* 2013(4):323–332. <https://doi.org/10.1127/1432-8364/2013/0180>
- Steffen R, Beder C (2007) Recursive estimation with implicit constraints. In: Hamprecht FA, Schnörr C, Jähne B (eds) Pattern recognition. Lecture notes in computer science, vol 4713. Springer, Berlin, pp 194–203
- Tailanian M, Paternain S, Rosa R, Canetti R (2014) Design and implementation of sensor data fusion for an autonomous quadrotor. In: 2014 IEEE International instrumentation and measurement technology conference (I2MTC 2014). IEEE, Piscataway, NJ, pp 1431–1436. <https://doi.org/10.1109/I2MTC.2014.6860982>
- Talaya J, Alamus R, Bosch E, Serra A, Kornus W, Baron A (2004) Integration of a terrestrial laser scanner with GPS/IMU orientation sensors. In: Proceedings of the XXth ISPRS congress, Istanbul, Turkey, vol 1223
- Unger J, Rottensteiner F, Heipke C (2016) Integration of a generalised building model into the pose estimation of UAS images. *Int Arch Photogramm Remote Sens Spat Inf Sci* 89:1057–1064. <https://doi.org/10.5194/isprsarchives-XLI-B1-1057-2016>
- Unger J, Rottensteiner F, Heipke C (2017) Assigning tie points to a generalised building model for UAS image orientation. *ISPRS Int Arch Photogramm Remote Sens Spat Inf Sci* 2:385–392. <https://doi.org/10.5194/isprs-archives-XLII-2-W6-385-2017>
- Vogel S, Alkhatib H, Bureick J, Moftizadeh R, Neumann I (2019) Georeferencing of laser scanner-based kinematic multi-sensor systems in the context of iterated extended Kalman filters using geometrical constraints. *Sensors (Basel, Switzerland)*. <https://doi.org/10.3390/s19102280>
- Vogel S, Alkhatib H, Neumann I (2016) Accurate indoor georeferencing with kinematic multi sensor systems. In: 2016 International conference on indoor positioning and indoor navigation (IPIN), pp 1–8. <https://doi.org/10.1109/IPIN.2016.7743601>
- Vogel S, Alkhatib H, Neumann I (2018) Iterated extended kalman filter with implicit measurement equation and nonlinear constraints for information-based georeferencing. In: 2018 International conference on information fusion (FUSION), pp 1209–1216. <https://doi.org/10.23919/ICIF.2018.8455258>

H. Alkhatib, J.-A. Paffenholz, and H. Kutterer

Abstract

The Kalman filter is supposed to be the optimal analytical closed-form solution for the Bayesian space-state estimation problem, if the state-space system is linear and the system noises are additive Gaussian. Unfortunately, except in the above mentioned cases, there is no closed-form solution to the filtering problem. So it is necessary to adopt alternative techniques in order to solve the Bayesian filtering problem. Sequential Monte Carlo (SMC) filtering – or commonly known as particle filter – is a well known approach that allows to reach this goal numerically, and works properly with nonlinear, non-Gaussian state estimation. However, computational difficulties could occur concerning the sufficient number of particles to be drawn. We present in this paper a more efficient approach, which is based on the combination of SMC filter and the extended Kalman filter. We identified the resulting filter as extended Kalman particle filter (EKPF). This filter is applied to a method for the direct geo-referencing of 3D terrestrial laser scans.

Keywords

Nonlinear state estimation • Bayesian Filtering • Sequential Monte Carlo Filtering • GNSS

1 Introduction

Linear filtering theory according to [Kalman and Bucy \(1960\)](#) is optimal only if the system, which consists

of measurement and transition equations, is linear and the error process is Gaussian. Unfortunately, the modeling of reality sometimes differs from these optimal assumptions and nonlinear, non-Gaussian, and non-stationary state estimation should be taken into account. Thus over the years a multitude of approximate nonlinear filters has been proposed; see e.g., [Doucet et al. \(2001\)](#), and [Simon \(2006\)](#). A well known analytical approximation to handle a nonlinear system is to linearize the measurement and the system equations using Taylor series expansions; see e.g., [Simon \(2006\)](#).

However, as pointed out in [Doucet et al. \(2001\)](#) this type of nonlinear filter which includes the first-order

H. Alkhatib (✉) · J.-A. Paffenholz
Geodätisches Institut der Leibniz Universität Hannover,
D-30167 Hannover, Nienburger Str. 1, Germany
e-mail: alkhatib@gih.uni-hannover.de

H. Kutterer
Geodätisches Institut der Leibniz Universität Hannover,
D-30167 Hannover, Nienburger Str. 1, Germany

Bundesamt für Kartographie und Geodäsie
Richard-Strauss-Allee 11, D-60598 Frankfurt am Main

and the higher-order extended Kalman filter (EKF), is prone to diverge if the system equations are highly nonlinear. This gives us the motivation to use other filter techniques such as the sequential Monte Carlo (SMC) approach in order to take the nonlinearities into account. The SMC filter (also known as particle filter (PF)) is a suboptimal filter for implementing the recursive Bayesian filter by Monte Carlo (MC) techniques; see e.g., Doucet et al. (2001) and Ristic et al. (2004). The main idea behind the SMC filter is to approximate the posterior power density function (PDF) of the state parameters by a set of random samples, which can be generated from a known PDF. By means of the drawn particles the mean as well as the variance-covariance information of the state vector are estimated.

In order to obtain an equivalent representation of the posterior PDF a large number of particles should be drawn. Unfortunately, the high computational cost due to the large number of required particles restricts the use of SMC in many applications. In this paper, two filtering techniques are discussed: first, the generic PF, and, second a filtering technique, which can significantly improve the performance of PF, and which reduces the computational cost of the algorithm.

2 Nonlinear State Estimation

2.1 The Mathematical Model

Before describing the different filter algorithms, we briefly introduce the notation and the terminology used throughout this paper. To define the problem of nonlinear filtering, let us consider the state vector $\mathbf{x}_k \in \mathbb{R}^{n_x}$, where n_x is the dimension of the state vector, and k is the time index. The evolution of the state vector $\mathbf{x}_k$ is described by the dynamic model:

$$\mathbf{x}_{k+1} = f(\mathbf{x}_k, \mathbf{u}_k, \mathbf{w}_k) \quad (12.1)$$

where f is a known, in general nonlinear function of $\mathbf{x}_k$, $\mathbf{u}_k$ the vector of known (deterministic) input, and $\mathbf{w}_k$ is the process noise vector, which is caused by mismodeling effects and other disturbances in the motion model. The main aim of filtering is to estimate the optimal state vector $\mathbf{x}_{k+1}$ from the observations $\mathbf{y}_{k+1} \in \mathbb{R}^{n_y}$ and $\mathbf{x}_k$ where n_y is the dimension of the measurement model:

$$\mathbf{y}_{k+1} = h(\mathbf{x}_{k+1}, \mathbf{v}_{k+1}). \quad (12.2)$$

In (12.2) h is a known, in general nonlinear function, and $\mathbf{v}_{k+1}$ is the measurement noise vector, which obeys a known PDF and is mutually independent with the system noise $\mathbf{w}_k$.

2.2 The Bayes Filter

From a Bayesian perspective, the filtering problem is to estimate the state $\mathbf{x}_{k+1}$ recursively given the data $\mathbf{y}_{1:k+1}$ up to time $k+1$. Thus, it is required to evaluate the joint posterior PDF given the hole data. That is:

$$p(\mathbf{x}_{k+1} | \mathbf{y}_{1:k+1}) = \frac{p(\mathbf{y}_{k+1} | \mathbf{x}_{k+1}) \cdot p(\mathbf{x}_{k+1} | \mathbf{y}_{1:k})}{p(\mathbf{y}_{k+1} | \mathbf{y}_{1:k})} \quad (12.3)$$

where the posterior PDF at time k , $p(\mathbf{x}_k | \mathbf{y}_{1:k})$, is first projected forward in time in order to calculate the prior PDF at time $k+1$. This is done by using the probabilistic process model (cf. Simon 2006, pp. 464):

$$p(\mathbf{x}_{k+1} | \mathbf{y}_{1:k}) = \int p(\mathbf{x}_{k+1} | \mathbf{x}_k) \cdot p(\mathbf{x}_k | \mathbf{y}_{1:k}) d\mathbf{x}_k. \quad (12.4)$$

The probabilistic model of the state evolution $p(\mathbf{x}_{k+1} | \mathbf{x}_k)$ is defined by the system described in (12.1) and the known PDF of the noise vector $\mathbf{w}_k$. The term $p(\mathbf{y}_{k+1} | \mathbf{y}_{1:k})$ in (12.3) is a normalizing factor. Figure 12.1 illustrates the k th recursive step of sequential Bayesian filtering, along with inputs and outputs.

Simultaneously with the recursion given jointly by (12.3) and (12.4), we can estimate the current state via a *maximum a posteriori* (MAP) approach (see, for instance, Koch (2007)):

$$\hat{\mathbf{x}}_{k+1} = \max_{\mathbf{x}_{k+1}} p(\mathbf{x}_{k+1} | \mathbf{y}_{1:k+1}). \quad (12.5)$$

Note, that a closed-form solution for the filtering problem presented in (12.3) and (12.5) only exists if the system equations presented in (12.1) and (12.2) are linear, and both the system noise and the observation noise are Gaussian. When these conditions are fulfilled, we obtain the known Kalman filter, which is a special sequential Bayesian filter where the posterior density $p(\mathbf{x}_{k+1} | \mathbf{y}_{1:k+1})$ also becomes Gaussian, refer to (Arulampalam et al. 2002).

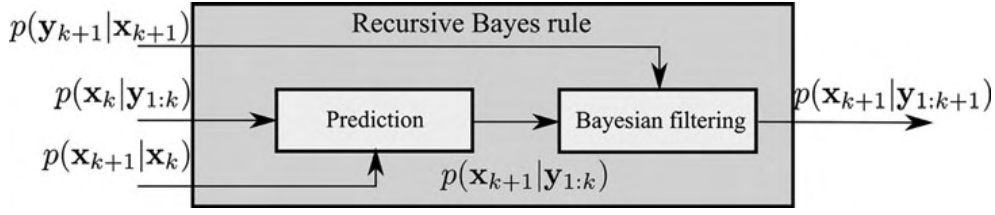


Fig. 12.1 Recursive computation in sequential Bayesian filtering. Each epoch k has two computation steps: the prediction step and the filtering step. These steps are sequential. The prediction unit takes in the motion model density and the posterior density

of previous time step k and outputs predicted posterior density $p(\mathbf{x}_{k+1} | \mathbf{y}_{1:k})$. Next the Bayesian filtering unit takes in this predicted posterior density and the likelihood density to estimate the posterior density $p(\mathbf{x}_{k+1} | \mathbf{y}_{1:k+1})$ for the current time step

2.3 The Generic Particle Filter

In this section, we describe how to approximate the optimal Bayesian solution (see Sect. 2.2) if an analytical solution is unsolvable. The PF is a suboptimal solution to approximate the Bayesian estimator given in (12.3) numerically by means of SMC techniques.

The main idea of SMC is based on particle representation of a PDF. The SMC technique is used to determine the components of the state vector in the nonlinear filtering system given by (12.1) and (12.2). The resulting MC algorithm is known as sequential importance sampling (SIS). This method approximates the posterior PDF by a set of M weighted samples of this density without making any explicit assumption about its form and can thus be used in general nonlinear, non-Gaussian systems.

Let the particle set $\{\mathbf{x}_{0:k}^{(i)}, w_k^{(i)}\}_{i=1}^M$ denote a random measure that characterizes the posterior PDF $p(\mathbf{x}_{0:k} | \mathbf{y}_{1:k})$, where $\{\mathbf{x}_{0:k}^{(i)}, i = 1, \dots, M\}$ is a set of realization points with associated weights $\{w_k^{(i)}, i = 1, \dots, M\}$, and $\{\mathbf{x}_j, 0 = 1, \dots, k\}$ is the set of all states up to epoch k . One can then approximate the posterior density in epoch k as:

$$p(\mathbf{x}_{0:k} | \mathbf{y}_{1:k}) \approx \sum_{i=1}^M w_k^{(i)} \delta(\mathbf{x}_{0:k} - \mathbf{x}_{0:k}^{(i)}) \quad (12.6)$$

where δ represents the Kronecker Delta function, and the associated weights $w_k^{(i)}$ sum up to unity. The weights $w_k^{(i)}$ in (12.6) are chosen using the principle of importance sampling, cf. Doucet et al. (2001). After a lengthy derivation, which will not be given here due to lack of space, the weights are computed recursively

based on the weight update equation (refer to Ristic et al. 2004, pp. 37–39):

$$w_{k+1}^{(i)} = w_k^{(i)} \frac{p(\mathbf{y}_{k+1} | \mathbf{x}_{k+1}^{(i)}) p(\mathbf{x}_{k+1}^{(i)} | \mathbf{x}_k^{(i)})}{\pi(\mathbf{x}_{k+1}^{(i)} | \mathbf{x}_{1:k}^{(i)}, \mathbf{y}_{1:k})} \quad (12.7)$$

where $\pi(\cdot)$ is a known PDF from which it is easy to draw samples, $p(\mathbf{y}_{k+1} | \mathbf{x}_{k+1}^{(i)})$ the evaluated likelihood PDF for each particle and $p(\mathbf{x}_{k+1}^{(i)} | \mathbf{x}_k^{(i)})$ is simply the PDF of the state at epoch $k + 1$ given a specific state at previous epoch k for every particle $\mathbf{x}_k^{(i)}$. The weight update equation (12.7) yields a sequential update of the importance weights, given an appropriate choice of the proposal distribution $\pi(\cdot)$. Doucet et al. (2001) show that the selection of the proposal PDF is one of the most critical issues in the SIS algorithm.

The SIS algorithm starts with M initialization values of the state vector $(\mathbf{x}_0^{(i)})$ with $i \in 1, \dots, M$, which can be randomly generated from the initial PDF $\pi(\mathbf{x}_0)$. These particles are then propagated at each epoch $k = 1, 2, \dots, n$ in forward by substitution in the dynamic equation (12.1). In order to distinguish between this drawn particles (they were indicated in (12.7) as $\mathbf{x}_k^{(i)}$) and the resampled particles in the following step, we rename the resulting particles as $\mathbf{x}_{k,-}^{(i)}$. As the current observations $\mathbf{y}_k$ become available, we compute the conditional likelihood of each particle: $p(\mathbf{y}_k | \mathbf{x}_{k,-}^{(i)})$. The evaluation of the likelihood is based on the known PDF of the measurement noise and on the nonlinear measurement equation. On the basis of (12.7) we recursively compute the relative weights. Before we evaluate the current state obtained by (12.5) and move to the next time step, the particles are resampled.

In other words, we randomly generate new particles $\mathbf{x}_{k,+}^{(i)}$ based on the relative weights. Particles with relatively small weights are eliminated. Otherwise, particles with large weights are duplicated. This resampling is used to avoid the problem of degeneracy of drawn particles.

2.4 The Extended Kalman Particle Filter

In this section, an implementation issue to improve the performance of PF presented in Sect. 2.3 is introduced. A shortcoming of the PF algorithm is the computational cost caused by the increase of the generated particles. A large number of samples should be drawn in order to achieve the convergence of the algorithm, and to estimate the desired state vectors and its covariance matrix. In [Ristic et al. \(2004\)](#) and [Simon \(2006\)](#), several implementation issues are considered for improving the PF algorithm, including degeneracy, the selection of the importance density, and particle filters with an improved sample diversity. Due to lack of space we only discuss the developed approach for enhancement of convergence based on combination with the well known Kalman filter such as the extended Kalman filter (EKF). The novelty of the proposed EKPF algorithm is the update of each particle at every time step k using the EKF, when a new measurement $\mathbf{y}_k$ arrives. In other words, we are running an extra EKF step for i th particle $\mathbf{x}_{k+1,-}^{(i)}$ at the epoch $k + 1$:

$$\begin{aligned}\mathbf{P}_{k+1,-}^{(i)} &= \mathbf{F}_k^{(i)} \mathbf{P}_{k,+}^{(i)} \mathbf{F}_k^{(i)T} + \mathbf{Q}_k \\ \mathbf{K}_{k+1}^{(i)} &= \mathbf{P}_{k+1,-}^{(i)} \mathbf{H}_{k+1}^{(i)T} \left(\mathbf{H}_{k+1}^{(i)} \mathbf{P}_{k+1,-}^{(i)} \mathbf{H}_{k+1}^{(i)T} + \mathbf{R}_{k+1} \right)^{-1} \\ \mathbf{x}_{k+1,+}^{(i)} &= \mathbf{x}_{k+1,-}^{(i)} + \mathbf{K}_{k+1}^{(i)} \left[\mathbf{y}_{k+1} - \mathbf{h}(\mathbf{x}_{k+1,-}^{(i)}) \right] \\ \mathbf{P}_{k+1,+}^{(i)} &= \left(\mathbf{I} - \mathbf{K}_{k+1}^{(i)} \mathbf{H}_{k+1}^{(i)} \right) \mathbf{P}_{k+1,-}^{(i)}.\end{aligned}\quad (12.8)$$

$\mathbf{K}_{k+1}^{(i)}$ represents the Kalman gain of the i th particle, and $\mathbf{P}_{k+1}^{(i)}$ is the appropriate estimation of the state covariance matrix. $\mathbf{Q}_k$ and $\mathbf{R}_{k+1}$ are the covariance matrices of the process noise vector $\mathbf{w}_k$ and the observation noise vector $\mathbf{v}_{k+1}$, respectively. We distinguish in (12.8) between the prior $\mathbf{P}_{k+1,-}^{(i)}$ and the posterior $\mathbf{P}_{k+1,+}^{(i)}$. The transition and design matrices $\mathbf{F}_k^{(i)}$ and $\mathbf{H}_{k+1}^{(i)}$ in (12.8) are defined as: $\mathbf{F}_k^{(i)} = \left. \frac{\partial f}{\partial \mathbf{x}} \right|_{\mathbf{x}=\mathbf{x}_{k,+}^{(i)}}$ and $\mathbf{H}_{k+1}^{(i)} = \left. \frac{\partial h}{\partial \mathbf{x}} \right|_{\mathbf{x}=\mathbf{x}_{k+1,-}^{(i)}}$, respectively.

The Taylor series are evaluated for the transition matrix $\mathbf{F}$ for the particle from the previous epoch $\mathbf{x}_{k,+}^{(i)}$ and for the predicted particle from the current epoch $\mathbf{x}_{k+1,-}^{(i)}$ for the matrix $\mathbf{H}$. Please note that the functions $f(\cdot)$ and $h(\cdot)$ are both time invariant. The key idea behind this approach is the substitution of the possibly nonlinear model given by (12.1) and (12.2) with a linearized model to reduce the variance of the drawn particles in order to get short computing times without increasing the number of samples.

The generated prior particles $\mathbf{x}_{k,-}^{(i)}$ would be transformed to a new set of particles $\mathbf{x}_{k,+}^{(i)}$ using the EKF step given by (12.8). Based on the transformed particles $\mathbf{x}_{k,+}^{(i)}$ and their $\mathbf{P}_{k+1,+}^{(i)}$ we generate and propagate a new set of particles using the Gaussian PDF:

$$\mathbf{x}_{k+1}^{(i)} \sim p\left(\mathbf{x}_{k+1}^{(i)} | \mathbf{x}_k^{(i)}, \mathbf{y}_{k+1}\right) \sim \mathcal{N}\left(\mathbf{x}_{k+1,+}^{(i)}, \mathbf{P}_{k+1,+}^{(i)}\right) \quad (12.9)$$

where $\sim$ in (12.9) means that the particles are drawn from a specific PDF. The remaining computational steps of the EKPF are similar to the generic PF.

3 Numerical Study and Results

In this section an application of the algorithms presented in Sects. 2.3 and 2.4 is shown and the results are discussed. The main goal of the numerical investigation is to derive position and orientation parameters for the transformation of a local sensor-defined coordinate system (denoted by upper index L) to an global earth centered, earth fixed coordinate system (denoted by upper index G). This is a typical task within the direct geo-referencing procedure of 3D terrestrial laser scans. For this purpose, an adapted sensor-driven method based on a multi-sensor system (MSS) has been developed at the Geodetic Institute of the Leibniz Universität Hannover (GIH). The MSS is established by a sensor fusion of a phase-based terrestrial laser scanner (TLS) and additional navigation sensors to observe the parameters.

The above mentioned transformation parameters include the position of the MSS, which is equal to the translation vector and a rotation matrix, which contains the orientation of the three axes of the MSS – roll, pitch and yaw angle, known from aeronautics. The mathematical modeling of the MSS in form of a EKF approach is presented in [Paffenzholz et al. \(2009\)](#).

This approach uses the constant rotation of the TLS about its local vertical axis (z^L) in combination with kinematic GNSS measurements to estimate four of the six degrees of freedom of the transformation – the position vector as well as the orientation in the horizontal plane. Therefore, one GNSS antenna is mounted eccentrically on the TLS. In order to optimize the direct geo-referencing strategy the MSS is enhanced with additional navigation sensors to estimate the residual spatial rotation angles about the x^L - and y^L -axis.

In this MSS application the trajectory can be described by a circle in 3D space. This parameterization is due to the circular motion of the antenna reference point (ARP) caused by the constant rotation of the TLS about the z^L -axis, as already mentioned. The orientation change of the ARP within two time steps is given by the circular arc segment s divided by the radius r_k .

The state vector is expressed by the components:

$$\mathbf{x}_k^G = [\mathbf{X}_k^G \ \alpha_{S,k}^G \ \beta_{S,k}^L \ \gamma_{S,k}^L]^T \quad (12.10)$$

where $\mathbf{X}_k^G$ is the global position of the ARP at the epoch k , $\alpha_{S,k}^G$ describes the azimuthal orientation of the MSS, $\beta_{S,k}^L$ the inclination in scan direction and $\gamma_{S,k}^L$ is perpendicular to the scan direction. The space state model leads to:

$$\mathbf{x}_{k+1} = \begin{bmatrix} \mathbf{X}_k^G + \mathbf{R}_L^G(\lambda, \varphi) \cdot \mathbf{R}_{SN}^{aP}(\alpha_{S,k}^G) \cdot \Delta \mathbf{X}_k^L \\ \alpha_{S,k}^G + \frac{s_k}{r_k} \\ \beta_{S,k}^L \\ \gamma_{S,k}^L \end{bmatrix} + \mathbf{w}_k. \quad (12.11)$$

The term $\Delta \mathbf{X}_k^L$ in (12.11) is given by:

$$\Delta \mathbf{X}_k^L = \begin{bmatrix} r_k \cdot \cos(\gamma_{S,k}^L) \cdot \sin\left(\frac{s_k}{r_k}\right) \\ r_k \cdot \cos(\gamma_{S,k}^L) \cdot \cos\left(\frac{s_k}{r_k}\right) \\ s_k \cdot \sin(\beta_{S,k}^L) \end{bmatrix} - \Delta \mathbf{X}_{SN,k}^{GNSS} \quad (12.12)$$

where $\Delta \mathbf{X}_{SN,k}^{GNSS}$ represents the eccentric position of the GNSS antenna. $\Delta \mathbf{X}_k^L$ in (12.11) is responsible for the high-nonlinearity in the space state model. It should be pointed out, that in Paffenholtz et al. (2009) additional

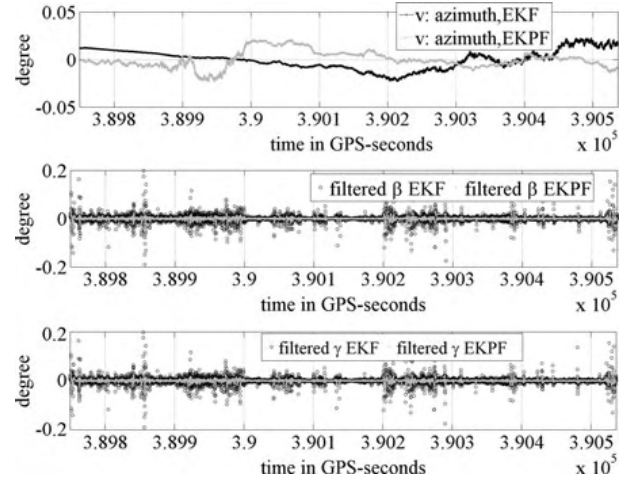


Fig. 12.2 Filtering results of EKF and the EKPF; *top*: the residuals obtained within a linear regression of the orientation α_S^G ; *middle and bottom*: The filtered inclinations β_S^L and γ_S^L . The EKPF approach shows a significant improvement of the filter effect for the filtered inclination whereas the filtered results of the estimated azimuth are comparable

adaptive parameters are considered in the space state model. However, the consideration of such adaptive parameters in the EKPF algorithm (refer to Sect. 2.4) needs a significant modification, which will be shown in future works.

The measurement model is characterized by the position of the GNSS antenna $\mathbf{X}_k^G$, and the measurements of the inclination sensor $\beta_{S,k}^L$ and $\gamma_{S,k}^L$. That yields:

$$\mathbf{y}_{k+1} = \begin{bmatrix} \mathbf{X}_{k+1}^G \\ \beta_{S,k+1}^L \\ \gamma_{S,k+1}^L \end{bmatrix} = \mathbf{H}_{k+1} \cdot \mathbf{x}_{k+1}^G + \mathbf{v}_{k+1}. \quad (12.13)$$

As start value for the EKPF approach (see Sect. 2.4) we randomly drawn 500 particles from $\mathcal{N}(\mathbf{x}_0, \mathbf{P}_0)$ with $\mathbf{x}_0 = \mathbf{0}$ and $\mathbf{P}_0$ the initial covariance matrix which has been chosen equally to the noise covariance matrix in Paffenholtz et al. (2009).

Figure 12.2 presents a subsample of the estimated state parameters by classical EKF algorithm (black circles) and EKPF approach (gray points). The upper part of this figure shows the residuals obtained within a linear regression of the orientation α_S^G . Due to the constant rotation of the TLS about its vertical axis, we expect a linear relationship between α_S^G and time. Therefore, the residuals are quality indicators. The

residuals are in both algorithms comparable, and lead to a metric uncertainty of about 1.5 cm for the azimuth calculation at a distance up to 35 m. The middle and lower part of Fig. 12.2 show a comparison between the filtered inclinations β_S^L and γ_S^L , respectively. Here again, the EKPF effect is noticeable for the filtered inclinations, mainly in case of higher noise level.

Conclusion

In this paper, the newly developed filtering approach EKPF was introduced. It is based on a combination of the SMC technique and an EKF step. The EKPF approach has been applied to derive transformation parameters for the direct geo-referencing of 3D terrestrial laser scans. The results show an improvement of the filter effect. They were compared to the classical EKF approach. The main benefit of the developed approach is the better performance in case of high-nonlinear space state equations. A second important result, which could be not shown in the above example due to lack of

space, is the significant decrease of the number of the generated particles compared to the generic PF.

References

- Arulampalam MS, Maskell S, Gordon N, Clapp T (2002) A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking. *IEEE Trans. Signal Proc.*, vol 50, pp 174–188
- Doucet A, de Freitas N, Gordon N (2001) *Sequential Monte Carlo methods in practice*. Springer, New York, Berlin
- Kalman RE, Bucy RS (1960) New results in linear filtering and prediction theory. *J Basic Eng Trans ASME*, D 83:95–108
- Koch KR (2007) *Introduction to Bayesian statistics*. 2nd edn. Springer, Berlin
- Paffenzholz J-A, Alkhatib H, Brieden P, Kutterer H (2009) Optimized direct geo-referencing strategy for a TLS-based Multi-Sensor-System. In: Grün A, Kahmen H (eds) *Optical 3D-measurement techniques IX*, Vienna, pp 287–292
- Ristic B, Arulampalam S, Gordon N (2004) *Beyond the Kalman filter, particle filters for tracking applications*. Artech House, Boston
- Simon D (2006) *Optimal state estimation*. Wiley, Hoboken

Alternative Nonlinear Filtering Techniques in Geodesy for Dual State and Adaptive Parameter Estimation

H. Alkhatib

Abstract

In many fields of geodesy applications, state and parameter estimation are of major importance within modeling of on-line processes. The fundamental block of such processes is a filter for recursive estimation. Kalman Filter is the well known filter, a simple and efficient algorithm, as an optimal recursive Bayesian estimator for a somewhat restricted class of linear Gaussian problems. However, in the case that state and/or measurement functions are highly non-linear and the density function of process and/or measurement noise are non-Gaussian, classical filters do not yield satisfying estimates. So it is necessary to adopt alternative filtering techniques in order to provide almost optimal results. A number of such filtering techniques will be reviewed in this contribution, but the main focus lays on the sequential Monte Carlo (SMC) estimation. The SMC filter (well known as particle filter) allows to reach this goal numerically, and works properly with nonlinear, non-Gaussian state estimation. The main idea behind the SMC filter is to approximate the posterior PDF by a set of random particles, which can be generated from a known PDF. These particles are propagated through the nonlinear dynamic model. They are then weighted according to the likelihood of the observations. By means of the particles the true mean and the covariance of the state vector are estimated. However, the computational cost of particle filters has often been considered as their main disadvantage. This occurs due to the large, sufficient number of particles to be drawn. Therefore a more efficient approach will be presented, which is based on the combination of SMC filter and the Kalman Filter. The efficiency of the developed filters will be demonstrated through application to a method for direct georeferencing tasks for a multi-sensor system (MSS).

Keywords

Nonlinear filtering in state space • Kalman filter • Bayesian filter • Sequential Monte Carlo filter • Multi-sensor system

1 Introduction

The Kalman filtering technique is used in geodesy especially in applications such as engineering navigation and deforma-

tion analysis. Both kinds of application require a sequential estimation of the system state based on information coming from a (dynamic) model of the system and from external observations. Over the years refined models were introduced for the system description (i.e., the trajectory of a vehicle), and a great variety of sensors was used to observe the system state. This yielded often nonlinearities in the equations of the Kalman filter. In the field of engineering navigation for example, Aussems (1999) describes the vehicle's trajectory by a refined model of consecutive arcs. In this model the

H. Alkhatib (✉)
Geodetic Institute Leibniz Universität Hannover, Nienburger Str. 1,
30167 Hannover, Germany
e-mail: alkhatib@gih.uni-hannover.de

vehicle coordinates are nonlinearly related to other state parameters like the angular velocity in the horizontal plane, the pitch angle or the tangential component of the linear velocity. Another approach that may lead to nonlinear state equations is the adaptive enhancement of the Kalman filter. For the estimation of the thermal diffusivity of an aluminum column Eichhorn (2008) introduces this physical parameter into the state vector. Thus, the initially linear state equations become nonlinear at the order of reciprocal exponentials. A majority of the performed observations of the system are nonlinearly related to its state parameters; refer to Sternberg (2000) for a disquisition on observation equations in engineering navigation. Non-linearities can occur directly due to the nature of the relation between measurements like distances or angles and point coordinates or due to the transformation of the observations in a different coordinate system as is often the case in navigation applications. Hence over the years a multitude of approximate nonlinear filters has been proposed, see e.g. Bar-Shalom et al. (2001), and Simon (2006). These methods can be loosely grouped into the following two main categories:

- *Gaussian approximate methods:* A well known analytical approximation to handle a nonlinear system is to linearize the measurement and the system equations using Taylor series expansions; see Gelb (1974). However, as pointed out in Doucet et al. (2001) this type of nonlinear filter which includes the first-order and the higher-order extended Kalman filter (EKF), is tended to diverge if the system equations are highly nonlinear. Another method to overcome the linearization process is the Unscented Kalman filter (UKF). The UKF has been introduced in Julier and Uhlmann (1997) and approximates the first and the second moment of the posterior PDF rather than to approximate nonlinear functions.
- *Sequential Monte-Carlo methods:* The SMC filter (also known as particle filter) is a suboptimal filter for implementing the recursive Bayesian filter by Monte Carlo (MC) techniques, see e.g. Doucet et al. (2001) and Ristic et al. (2004). The main idea behind the SMC filter is to approximate the posterior PDF by a set of random samples, which can be generated from a known PDF. These generated particles are propagated through the dynamic model. They are then weighted according to the likelihood of the observations. In a resampling step, the new weighted particles are drawn with a probability proportional to the likelihood of the observation. By means of the drawn particle the true mean and the covariance of the state vector are estimated. If we assume that the number of drawn particles is very large, the MC approximation becomes an equivalent representation to the functional description of the posterior PDF.

In this paper nonlinear filtering approaches are considered. We focus only on the second category, namely the

SMC filter. The problem of high computational cost due to the large required number of particles has been solved in Alkhatib et al. (2012). The approach that has been proposed in Alkhatib et al. (2012) to overcome the computational effort and improve the performance of the filtering process was to combine the SMC filter with one of the filter from the Gaussian approximate methods such as EKF. In this paper, an extension of the aforementioned algorithm [proposed in Alkhatib et al. (2012)] to the estimation of state and static or minimal time-varying unknown model parameters (adaptive parameters) is introduced. The estimation of both the dynamic state and the static parameters is commonly known as the dual estimation. Numerous papers have been dealt with the developing of estimation algorithms based on SMC methods, refer, e. g., to Storvic (2002), which considered models with sufficient statistics for the parameters and applied particle filters to an augmented vector of states and sufficient statistics. The used strategy here is based on Storvic (2002) by adding random walk to the parameters, and then expand the state space with the extended parameters for the dual estimation.

The paper is organized as follows. Section 2.1 describes the mathematical model of the system which includes both the system dynamics and the measurement model. Section 2.2 briefly introduces the EKF and the UKF. The SMC filter is presented in Sect. 2.3. Two numerical simulations are demonstrated in Sect. 3. Finally, section “Conclusion” summarizes the results and gives an outlook for future work.

2 Nonlinear State Estimation

2.1 The Probabilistic Inference

Probabilistic inference is the problem of estimating states or parameters of a system in an optimal and consistent approach (using probability theory) given noisy observations. This general framework is shown in Fig. 1. In particular, we will be addressing the sequential (recursive) probabilistic inference problem within discrete-time nonlinear dynamic systems that can be described by a dynamic state-space model (DSSM). The hidden system state x_k , with initial probability density $p(x_0)$, evolves over time (k is the discrete time index) as an indirect or partially observed first order Markov process according to the conditional probability density $p(x_k|x_{k-1})$. The observations y_k are conditionally independent given the state and are generated according to the conditional probability density $p(y_k|x_k)$.

The DSSM can also be written as a set of nonlinear system equations

$$\mathbf{x}_k = f(\mathbf{x}_{k-1}, \mathbf{u}_{k-1}, \mathbf{w}_{k-1}) \quad (1)$$

$$\mathbf{y}_k = h(\mathbf{x}_k, \mathbf{v}_k) \quad (2)$$

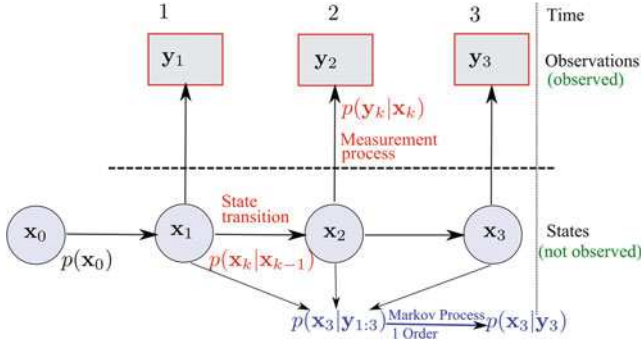


Fig. 1 Graphical model of a probabilistic dynamic state-space model

where $\mathbf{v}_k$ is the process noise that drives the dynamic system through the nonlinear state transition function f , and $\mathbf{n}_k$ is the observation or measurement noise corrupting the observation of the state through the nonlinear observation function h . The state transition density $p(\mathbf{x}_k|\mathbf{x}_{k-1})$ is fully specified by f and the process noise distribution $p(\mathbf{v}_k)$, whereas h and the observation noise distribution $p(\mathbf{n}_k)$ fully specify the observation likelihood $p(\mathbf{y}_k|\mathbf{x}_k)$. The exogenous input to the system, $\mathbf{u}_k$, is assumed known.

2.2 The Bayes Filter

From a Bayesian perspective, the filtering problem is to estimate the state $\mathbf{x}_{k+1}$ recursively given the data $\mathbf{y}_{1:k+1}$ up to time $k+1$. Thus, it is required to evaluate the joint posterior PDF given the hole data. That is:

$$p(\mathbf{x}_{k+1}|\mathbf{y}_{1:k+1}) = \frac{p(\mathbf{y}_{k+1}|\mathbf{x}_{k+1}) \cdot p(\mathbf{x}_{k+1}|\mathbf{y}_{1:k})}{p(\mathbf{y}_{k+1}|\mathbf{y}_{1:k})} \quad (3)$$

where the posterior PDF at time k , $p(\mathbf{x}_k|\mathbf{y}_{1:k})$, is first projected forward in time in order to calculate the prior PDF at time $k+1$. This is done by using the probabilistic process model (cf. Simon 2006, pp. 464):

$$p(\mathbf{x}_{k+1}|\mathbf{y}_{1:k}) = \int p(\mathbf{x}_{k+1}|\mathbf{x}_k) \cdot p(\mathbf{x}_k|\mathbf{y}_{1:k}) d\mathbf{x}_k. \quad (4)$$

The probabilistic model of the state evolution $p(\mathbf{x}_{k+1}|\mathbf{x}_k)$ is defined by the system described in Eq. (1) and the known PDF of the noise vector $\mathbf{w}_k$. The term $p(\mathbf{y}_{k+1}|\mathbf{y}_{1:k})$ in Eq. (3) is a normalizing factor.

2.3 The Extended Kalman Particle Filter

On approach that has been proposed for improving particle filtering is described in Alkhatib et al. (2012) which combine the generic particle filter with the extended Kalman filter. In this approach each particle is updated at the measurement

time using the EKF, and the resampling is performed using the measurement equation. The main goal in this approach was the improving the numerical efficiency of the SMC filter.

The main drawback of the SMC filter is its computational cost which leads to increasing the sufficient number of particles to approximate the statistical moments of the state parameter and the adaptive parameter as well. In Ristic et al. (2004) and Simon (2006), several implementation issues are considered for improving the PF algorithm, including degeneracy, the selection of the importance density, and particle filters with an improved sample diversity. Due to the lack of space we discuss here only the developed approach for enhancement of convergence based on combination with the well known Kalman filter such as the extended Kalman filter (EKF). The novelty of the proposed EKPF algorithm was the update of each particle at every time step k using the EKF, when a new measurement $\mathbf{y}_k$ arrives. In other words, we are running an extra EKF step for every particle i :

$$\mathbf{P}_{k+1,i}^- = \mathbf{F}_{k,i} \mathbf{P}_{k,i}^+ \mathbf{F}_{k,i}^T + \mathbf{Q}_k \quad (5)$$

$$\mathbf{K}_{k+1,i} = \mathbf{P}_{k+1,i}^- \mathbf{H}_{k+1,i}^T (\mathbf{H}_{k+1,i} \mathbf{P}_{k+1,i}^- \mathbf{H}_{k+1,i}^T + \mathbf{R}_{k+1})^{-1}$$

$$\mathbf{x}_{k+1,i}^+ = \mathbf{x}_{0,i}^- + \mathbf{K}_{k+1,i} [\mathbf{y}_{k+1} - \mathbf{h}(\mathbf{x}_{k+1,i}^-)]$$

$$\mathbf{P}_{k+1,i}^+ = (\mathbf{I} - \mathbf{K}_{k+1,i} \mathbf{H}_{k+1,i}) \mathbf{P}_{k+1,i}^-.$$

$\mathbf{K}_{k+1}$ represents the Kalman gain of the i -th particle, and $\mathbf{P}_{k+1,i}$ is the appropriate estimation of the state covariance matrix. We distinguish in Eq. (5) between the a priori $\mathbf{P}_{k+1,i}^-$ and the a posteriori $\mathbf{P}_{k+1,i}^+$. The transition and design matrices $\mathbf{A}_k$ and $\mathbf{H}_k$ in Eq. (5) are defined as:

$$\mathbf{A}_k = \frac{\partial f}{\partial \mathbf{x}} \Big|_{\mathbf{x}=\mathbf{x}_k^+} \quad \text{and} \quad \mathbf{H}_k = \frac{\partial h}{\partial \mathbf{x}} \Big|_{\mathbf{x}=\mathbf{x}_k^-}, \quad (6)$$

respectively. The key idea behind this approach is the substitution of the possibly non-linear model given by Eq. (1) with a linearized model to reduce the variance of the drawn particles in order to get short computing times without increasing the number of samples.

The generated prior particles $\mathbf{x}_{k,i}^-$ would be transformed to a new set of particles $\mathbf{x}_{k,i}^+$ using the EKF step given by Eq. (5). Based on the transformed particles $\mathbf{x}_{k,i}^+$ and their covariance matrix $\mathbf{P}_{k+1,i}^+$ we generate and propagate a new set of particles using the Gaussian importance PDF:

$$\mathbf{x}_{k+1,i} \sim p(\mathbf{x}_{k+1,i}^i|\mathbf{x}_k^i, \mathbf{y}_{k+1}) \sim \mathcal{N}(\mathbf{x}_{k+1,i}^+, \mathbf{P}_{k+1,i}^+). \quad (7)$$

where the symbol $\sim$ in Eq. (7) means that the particles are generated from a specific PDF (in this case the Gaussian PDF). The remaining computational steps of the EKPF are similar to the generic PF, see Alkhatib et al. (2012).

2.4 Combined Parameter and State Estimation in EKPF Algorithm

The algorithm in Sect. 3.2 has to be developed in order to estimate the static unknown parameters. The maximum likelihood parameter estimation should be then performed based on PF and an effective stochastic approximation gradient algorithm is used to optimize cost function. The estimation of static parameters and dynamic state variables is performed simultaneously. The algorithm shown in this paper is adapted from Yang et al. (2008).

The state-space model should be extended by the static parameters $\boldsymbol{\vartheta}$:

$$\mathbf{x}_k \sim p(\mathbf{x}_k | \mathbf{x}_{k-1}, \boldsymbol{\vartheta}) \quad (8)$$

$$\mathbf{y}_k \sim p(\mathbf{y}_k | \mathbf{x}_k, \boldsymbol{\vartheta}) \quad (9)$$

The static parameter $\boldsymbol{\vartheta} \in \mathbb{R}^m$ with m dimensional unknown static parameters vector. The method introduced here focuses rather on the estimation of $\boldsymbol{\vartheta}$ directly by the maximum likelihood method. In other words, the dynamic state parameters are estimated by the SMC filter and static parameters are estimated by recursive ML method online.

The cost function to estimate the likelihood is given by:

$$f(\boldsymbol{\vartheta}) = p(\mathbf{y}_k | \mathbf{y}_{0:k-1}, \boldsymbol{\vartheta}) \quad (10)$$

$$\int p(\mathbf{y}_k | \mathbf{x}_k, \boldsymbol{\vartheta}) p(\mathbf{x}_k | \mathbf{y}_{0:k-1}, \boldsymbol{\vartheta}) d\mathbf{x}_k.$$

A closed-form of the integral given in (10) expect for only fey special cases is impossible. Numerical methods will be approximate the optimal solution. In every filter step, the cost function will be maximized. The problem here is to find the zeros of the gradient $\nabla f(\boldsymbol{\vartheta})$:

$$\boldsymbol{\vartheta}_k = \boldsymbol{\vartheta}_{k-1} + \gamma_k \hat{\nabla} f(\boldsymbol{\vartheta}_{k-1}). \quad (11)$$

In (11) $\hat{\nabla} f(\boldsymbol{\vartheta}_{k-1})$ denotes the estimated value of the gradient in the point $\boldsymbol{\vartheta}_{k-1}$ and γ_k denotes a sequence of decreasing step size. After a sufficient number of iterations, the true value of $\boldsymbol{\vartheta}_k$ will be estimated. The gradient estimate in (11) is obtained by numerical approximation, for more details see Yang et al. (2008).

3 Numerical Applications

We compare in this section the performance of the different nonlinear filter techniques presented in Sect. 2 by two numerical experiments. The first one is a simple target tracking problem, which is typically arises in navigation, and the second is to derive position and orientation parameters for the

transformation of a local sensor-defined coordinate system to a global coordinate system.

3.1 Tracking of a Nonlinear Trajectory

In this scenario we track a vehicle moving along a highly nonlinear trajectory. Consider a base state vector $\mathbf{x} = [x_k \ y_k \ \dot{x}_k \ \dot{y}_k]$, where x_k and y_k specify the position of the vehicle; and $\dot{x}_k$ and $\dot{y}_k$ are the velocities in the Cartesian plane. The system is given by a linear kinematic model and a nonlinear measurement model.

Dynamic system: The dynamic system (cf. Eq. 1) of the vehicle can be modeled with a discretized Wiener velocity model, cf. Bar-Shalom et al. (2001) and Särkkä (2006):

$$\mathbf{x}_{k+1} = \begin{bmatrix} x_{k+1} \\ y_{k+1} \\ \dot{x}_{k+1} \\ \dot{y}_{k+1} \end{bmatrix} = \begin{bmatrix} 1 & 0 & \Delta t & 0 \\ 0 & 1 & 0 & \Delta t \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (12)$$

$$\boldsymbol{\Sigma}_{\mathbf{w}_k \mathbf{w}_k} = \begin{bmatrix} \frac{\Delta t^3}{3} & 0 & \frac{\Delta t^2}{2} & 0 \\ 0 & \frac{\Delta t^3}{3} & 0 & \frac{\Delta t^2}{2} \\ \frac{\Delta t^2}{2} & 0 & \Delta t & 0 \\ 0 & \frac{\Delta t^2}{2} & 0 & \Delta t \end{bmatrix} \cdot q, \quad (13)$$

where $\mathbf{w}_k$ is Gaussian with $E(\mathbf{w}_k) = \mathbf{0}$ and $q = 0.1[m^2/sec^3]$ fixes the spectral density of the noise.

Measurement model: We assume that two sensors provide measurements of distance s_i and horizontal angle ϑ_i to the moving vehicle with a time discretization step of $\Delta t = 0.01 \text{ sec}$:

$$s_{i,k+1} = \sqrt{(x_{k+1} - X_0^i)^2 + (y_{k+1} - Y_0^i)^2} \quad (14)$$

$$\vartheta_{i,k+1} = \arctan\left(\frac{y_{k+1} - Y_0^i}{x_{k+1} - X_0^i}\right) \quad (15)$$

where X_0^i, Y_0^i are the locations of the sensors with $i \in \{1, 2\}$. The noise vector $\mathbf{v}_k \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_{\mathbf{v}})$ is characterized by $\boldsymbol{\Sigma}_{\mathbf{v}} = \text{diag}[\sigma_{s_1}^2 \ \sigma_{\vartheta_1}^2 \ \sigma_{s_2}^2 \ \sigma_{\vartheta_2}^2]$.

Figure 2 shows the final tracking results for three different filtering techniques compared to the real trajectory. The prior standard deviation of the measurements given by $\sigma_{s_i} = 0.05[m]$ for distances and $\sigma_{\vartheta_i} = 0.01[rad]$ for angles. As we can see in Fig. 2 all filtering methods for such optimal measurement accuracies give almost identical estimates of the system states. However, Low Cost sensors have become prevalent in geodetic applications in the last years. Therefore high variances of the measurements have to be taken into account in the analysis process. If a more rigorous analysis of

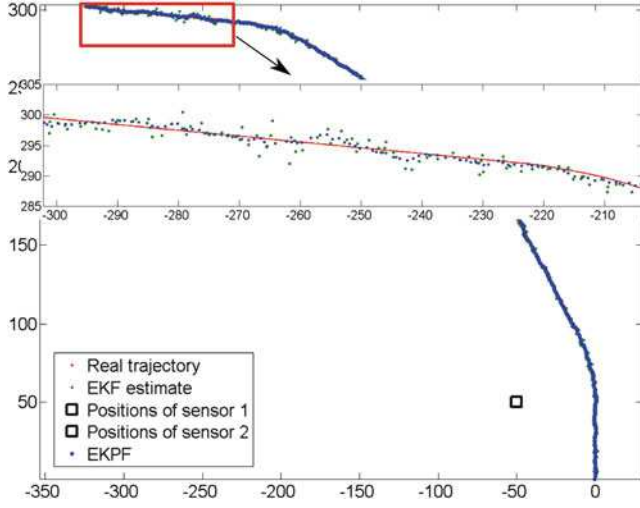


Fig. 2 Nonlinear filtering results of nonlinear trajectory

the measurements of low-cost sensors can be carried out, the acquisition costs of a sensor system can be decreased. From a theoretical point of view, the higher the variances the more important is a very good approximation of the nonlinear function (EKF) or of the PDF of the system (PF, EKPF). One would expect that in cases of very high variances the EKPF produces significantly more accurate results of the estimated system states. For this reason, the variances in this simulation study are increased by a factor of four. The RMSE values for all 500 runs are shown in Fig. 3. It is clearly noticeable that nearly all runs of the PF have a smaller RMSE in comparison to the EKF and UKF.

Naturally, a more realistic dynamic model may reduce the mismodelling effects due to approximation errors of the function or the PDF, but this is not in the scope of the study.

3.2 Tracking of a Nonlinear Trajectory

In this section an application of the algorithm presented in Sect. 2 is shown and the results are discussed. The main goal of the numerical investigation is to derive position and orientation parameters for the transformation of a local sensor-defined coordinate system to a global coordinate system. This is a typical task within the direct georeferencing procedure of static terrestrial laser scans. For this purpose, an adapted sensor-driven method based on a multi-sensor system (MSS) has been developed at the Geodetic Institute of the Leibniz Universität Hannover (GIH). The MSS is established by a sensor fusion of a phase-based terrestrial laser scanner (TLS) and additional navigation sensors to observe the parameters.

The above mentioned transformation parameters include the position of the MSS, which is equal to the translation

vector and a rotation matrix, which contains the orientation of the three axes of the MSS—roll, pitch and yaw angle known from aeronautics. The MSS as well as the mathematical modeling in form of a Kalman filtering approach are presented in details in Paffenholtz et al. (2010).

This approach uses the constant rotation of the TLS about its vertical axis with combination of kinematic GNSS measurements to estimate 4 of the 6° of freedom of the transformation—the position vector as well as the orientation in the horizontal plane. Therefore one GNSS antenna is mounted eccentrically on the TLS. In order to optimize the direct georeferencing strategy the MSS is enhanced with additional navigation sensors—inclinometers—to estimate the residual spatial rotation angles about the x- and y-axis of the TLS.

In this MSS application the trajectory can be described by a circle in 3D space. This parameterization is due to the orbital motion of the antenna reference point (ARP) caused by the constant rotation of the TLS about its vertical axis, as already mentioned. The orientation change of the ARP within two time steps is given by the circular arc segment s divided by the radius r_k .

The state vector is expressed by the following components:

$$\mathbf{x}_k^G = [\mathbf{X}_k^G \ \alpha_{S,k}^G \ \beta_{S,k}^L \ \gamma_{S,k}^L]^T \quad (16)$$

where $\mathbf{X}_k^G$ is the global position of the ARP at the epoch k , $\alpha_{S,k}^G$ describes the azimuth orientation of the MSS, $\beta_{S,k}^L$ the inclination in scan direction and $\gamma_{S,k}^L$ is perpendicular to the scan direction. The indices G and L in Eq. (16) are referred to the global or to the local coordinate system, respectively. The space state model leads to:

$$\mathbf{x}_{k+1} = \begin{bmatrix} \mathbf{X}_k^G + \mathbf{R}_L^G(\lambda, \varphi) \cdot \mathbf{R}_{SN}^{aP}(\alpha_{S,k}^G) \cdot \Delta \mathbf{X}_k^L \\ \alpha_{S,k}^G + \frac{s_k}{r_k} \\ \beta_{S,k}^L \\ \gamma_{S,k}^L \end{bmatrix} \mathbf{w}_k. \quad (17)$$

The term $\Delta \mathbf{X}_k^L$ in Eq. (17) is given by:

$$\Delta \mathbf{X}_k^L = \begin{bmatrix} r_k \cdot \cos(\gamma_{S,k}^L) \cdot \sin\left(\frac{s_k}{r_k}\right) \\ r_k \cdot \cos(\gamma_{S,k}^L) \cdot \cos\left(\frac{s_k}{r_k}\right) \\ s_k \cdot \sin(\beta_{S,k}^L) \end{bmatrix} - \Delta \mathbf{X}_{SN,k}^{GNSS} \quad (18)$$

where $\Delta \mathbf{X}_{SN,k}^{GNSS}$ represents the eccentric position of the GNSS antenna. $\Delta \mathbf{X}_k^L$ in Eq. (17) is responsible for the high-nonlinearity in the space state model. It should be pointed out, that in Paffenholtz et al. (2010) additional adaptive

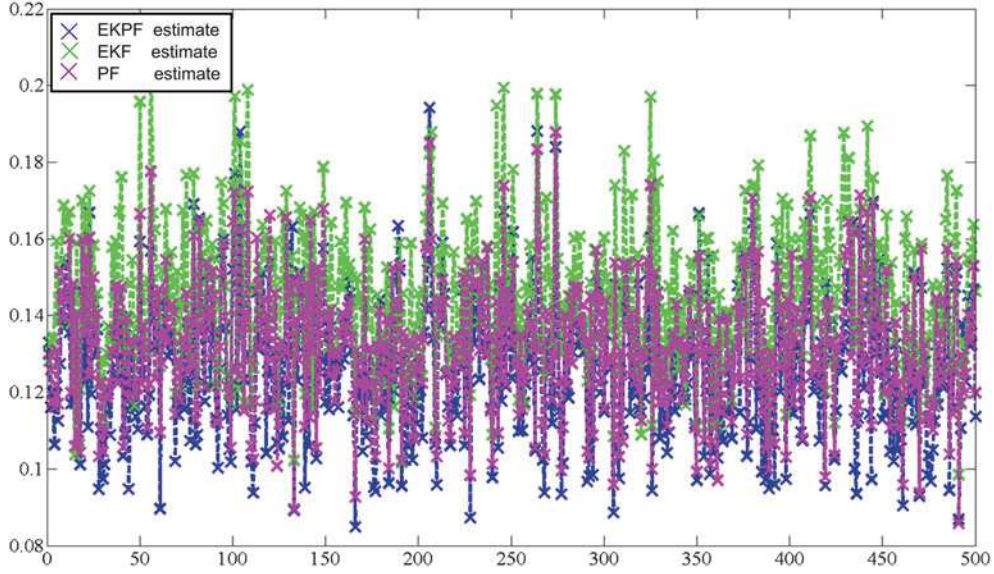


Fig. 3 RMSE values for the nonlinear filtering results of the 500 runs

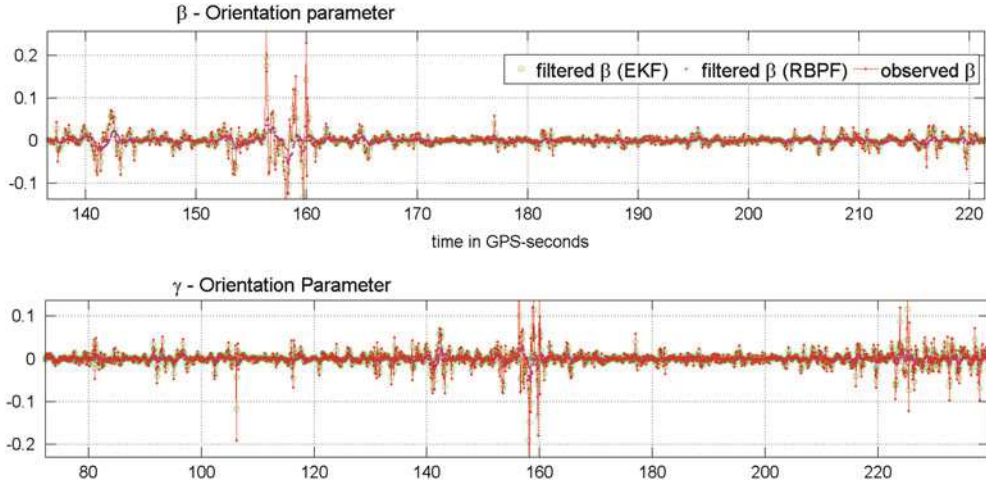


Fig. 4 Filtering results of EKF and the EKPF; *top* the residuals obtained within a linear regression of the orientation α_S^G ; *middle and bottom* the filtered inclinations β_S^L and γ_S^L

parameters are considered in the space state model. However, the consideration of such adaptive parameters in the EKPF algorithm (refer to Sect. 3.2) needs a significant modification, which will be shown in future works.

The measurement model is characterized by the position of the GNSS antenna $\mathbf{X}_k^G$, and the measurements of the inclination sensor $\beta_{S,k}^L$ and $\gamma_{S,k}^L$. This yields to:

$$\mathbf{y}_{k+1} = \begin{bmatrix} \mathbf{X}_{k+1}^G \\ \beta_{S,k+1}^L \\ \gamma_{S,k+1}^L \end{bmatrix} = \mathbf{A}_{k+1} \cdot \mathbf{x}_{k+1}^G + \mathbf{v}_{k+1}. \quad (19)$$

The state vector [refer to Eq. (17)] has been initialized on $\mathbf{x}_0 = \mathbf{0}$ with the initial covariance matrix which has been chosen equally to the noise covariance matrix in Alkhatib et al. (2012):

$$\mathbf{P}_0 = \text{diag} \begin{bmatrix} 100 \text{ mm}^2 & 100 \text{ mm}^2 & 900 \text{ mm}^2 \dots \\ 0.01 \text{ mg rad}^2 & 1 \text{ mg rad}^2 & 1 \text{ mg rad}^2 \end{bmatrix}.$$

As start value for the EKPF, we randomly draw 500 particles from $\mathcal{N}(\mathbf{x}_0, \mathbf{P}_0)$.

Figure 4 presents a subsample of the estimated state parameters by classical EKF algorithm and EKPF. The upper part of this figure shows the residuals obtained within a linear regression of the orientation α_S^G . Due to the constant rotation of the TLS about its vertical axis, we expect a linear relationship between α_S^G and time. Therefore, the residuals are quality indicators. The residuals are in both algorithms comparable, and lead to a metric uncertainty of about 1.5 cm for the azimuth calculation at a distance up to 35 m. The middle and lower part of Fig. 4 show a comparison between

the filtered inclinations β_S^L and γ_S^L , respectively. Here again, the EKPF effect is noticeable for the filtered inclinations, mainly in case of higher noise level.

Conclusion

In this paper, the newly developed EKPF algorithm for non-linear dynamic systems with adaptive static parameter. The algorithm is based on the combination of the newly developed EKPF algorithm and the gradient techniques. The algorithm has been applied to derive transformation parameters for the direct georeferencing of terrestrial laser scans. The EKPF algorithm is based on a combination of the SMC techniques and an EKF step, which guarantees a faster convergence. The results of the developed filter show an improvement of the filter effect. The EKPF with adaptive parameter shows a better performance in case of high-nonlinear space state equations.

References

- Alkhatib H, Paffenzholz J-A, Kutterer H (2012) Sequential Monte Carlo Filtering for nonlinear GNSS trajectories. In: Nico S (ed) VII Hotine-Marussi Symposium on mathematical geodesy. Proceedings of the Symposium in Rome, 6–10 June, 2009, pp.81–86. Springer (International Association of Geodesy Symposia, 137), Berlin/New York
- Aussems T (1999) Positionsschätzung von Landfahrzeugen mittels KALMAN-Filterung aus Satelliten- und Koppelnavigationsbeobachtungen. Veröffentlichungen des Geodätischen Instituts der Rheinisch-Westfälischen Technischen Hochschule Aachen, Nr. 55, Aachen
- Bar-Shalom Y, Li XR, Kirubarajan T, Li X-R (2001) Estimation with applications to tracking and navigation. Theory algorithms and software. Wiley, New York
- Doucet A, Freitas N, Gordon N (2001) Sequential Monte Carlo methods in practice. Springer, New York
- Eichhorn A (2008) Analysis of dynamic deformation processes with adaptive Kalman-filtering. J Appl Geodesy 1(1):9–15
- Gelb A (1974) Applied optimal estimation. MIT, Cambridge
- Julier SJ, Uhlmann JK (1997) A new extension of the Kalman filter to nonlinear systems. In: SPIE Proceedings of AeroSense. The 11th International Symposium on Aerospace/Defense Sensing, Simulation and Controls. SPIE, Orlando, FL, USA
- Paffenzholz J-A, Alkhatib H, Kutterer H (2010) Direct georeferencing of a static terrestrial laser scanner. J Appl Geodesy 4(3):115–126
- Ristic B, Arulampalam S, Gordon N (2004) Beyond the Kalman filter. Particle filters for tracking applications. Artech House, Boston
- Särkkä S (2006) Recursive Bayesian inference on stochastic differential equations. Ph.D. thesis, Helsinki University of Technology
- Simon D (2006) Optimal state estimation. Kalman, H infinity, and nonlinear approaches // Kalman, H [infinity] and nonlinear approaches. Wiley, Hoboken
- Sternberg H (2000) Zur Bestimmung der Trajektorie von Landfahrzeugen mit einem hybriden Messsystem. Schriftenreihe des Studienganges Geodäsie und Geoinformation, Universität der Bundeswehr München, No. 67, Neubiberg
- Storvik G (2002) Particle filters in state space models with the presence of unknown static parameters. IEEE Trans Signal Process 50(2):281–289
- Yang X, Xing K, Shi K, Pan Q (2008) Joint parameter and state estimation in particle filtering and stochastic optimization. J Control Theory Appl 6(2):215–220

Article

Ellipsoidal and Gaussian Kalman Filter Model for Discrete-Time Nonlinear Systems

Ligang Sun ^{1,*}, Hamza Alkhatib ¹, Boris Kargoll ², Vladik Kreinovich ³ and Ingo Neumann ¹

¹ Geodätisches Institut Hannover, Leibniz Universität Hannover, 30167 Hannover, Germany; alkhatib@gih.uni-hannover.de (H.A.); neumann@gih.uni-hannover.de (I.N.)

² Institut für Geoinformation und Vermessung Dessau, Hochschule Anhalt, 06846 Dessau, Germany; boris.kargoll@hs-anhalt.de

³ Department of Computer Science, University of Texas at El Paso, El Paso, TX 79968, USA; vladik@utep.edu

* Correspondence: sunligang@ieee.org

Received: 9 September 2019; Accepted: 20 November 2019; Published: 3 December 2019



Abstract: In this paper, we propose a new technique—called Ellipsoidal and Gaussian Kalman filter—for state estimation of discrete-time nonlinear systems in situations when for some parts of uncertainty, we know the probability distributions, while for other parts of uncertainty, we only know the bounds (but we do not know the corresponding probabilities). Similarly to the usual Kalman filter, our algorithm is iterative: on each iteration, we first predict the state at the next moment of time, and then we use measurement results to correct the corresponding estimates. On each correction step, we solve a convex optimization problem to find the optimal estimate for the system’s state (and the optimal ellipsoid for describing the systems’s uncertainty). Testing our algorithm on several highly nonlinear problems has shown that the new algorithm performs the extended Kalman filter technique better—the state estimation technique usually applied to such nonlinear problems.

Keywords: Ellipsoidal and Gaussian Kalman filter; state estimation; unknown but bounded uncertainty; nonlinear programming; convex optimization

1. Introduction

State estimation is important for virtually all areas of engineering and science. Every discipline which uses the mathematical modeling of its systems needs state estimation; this includes electrical engineering, mechanical engineering, chemical engineering, aerospace engineering, robotics, dynamical systems’ control and many others.

We estimate the system’s state based on the measurement results. In this estimation, we need to take into account that measurements are never absolutely accurate, the measurement results contain inaccuracy (“noise”)—e.g., due to inevitable imperfection of the measuring instruments. Also, our understanding of the system’s dynamics is also usually approximate. In addition to the internal factors (which are described by the system’s state) and the known external factors, there are usually also many other factors that affect the system—and which, from the viewpoint of the model, constitute the noise. The presence of noise affects our ability to predict the system’s behavior and to control the system. It is therefore desirable to minimize the effect of the noise—in particular, to minimize the effect of noise on the state estimation. In engineering, traditionally, techniques for decreasing the effect of noise—i.e., for separating (“filtering”) signal from noise—are known as filtering; because of this, the state estimation problem can be viewed as an important particular case of filtering.

To formulate the state estimation problem in precise terms, we need to have a mathematical model of the actual system, and we need to have a model describing the system’s and measurement

uncertainties (noise). To select the best filtering technique, we also need to select a numerical measure for describing the remaining inaccuracy of state estimation. Once this measure is selected, we need, given the measurement results, to find the state estimates for which the selected measure of inaccuracy attains the smallest possible value; see, e.g., [1]. In other words, from the mathematical viewpoint, estimating a state means solving the corresponding optimization problem.

The quality of the resulting state estimates depends on how well our models—i.e., our assumptions about the system and about the noise—describe the *actual* system and the actual noise. The most widely used state estimation techniques are *stochastic* techniques, i.e., techniques based on the assumption that we know the probability distribution of the noise (or, more generally, that we have some information about this distribution). In most techniques, it is assumed that the noise is Gaussian, but other distributions have also been considered. Stochastic techniques have been actively developed for dozens (if not hundreds) of years.

The most widely used stochastic state estimation techniques are the Kalman Filter (KF) techniques [2] and the Extended Kalman Filter (EKF) techniques [3,4]. In most practical applications, the traditional Kalman filters are used, which are based on the assumption that the system is linear (and that the noise is Gaussian). In practice, however, many real-life systems are non-linear (and the actual noise distribution is sometimes non-Gaussian).

Because of the ubiquity of non-linear systems, several filtering techniques have been developed for the non-linear case. Many applications use the finite sum approximation techniques such as the Gaussian sum filter (see, e.g., [5]) or linearization techniques such as EKF (see, e.g., [6]). These techniques work well for many non-linear systems. Their main limitation is that they assume that both the system noise and the measurement noise are normally distributed. As a result, sometimes, when the actual distributions are non-Gaussian—e.g., when they are multi-modal—these techniques do not perform well.

Another widely used filtering technique applicable to non-linear systems is the *Unscented Kalman Filter* (UKF), first proposed in [7]. The UKF technique is based on using a special non-linear transformation of the data—called Unscented Transform (UT)—that, crudely speaking, transforms the actual probability distributions into simpler ones. The main advantage of UKF is that, in contrast to EKF, UKF techniques do not assume that the required nonlinear transformations are smooth (differentiable). As a result, UKF does not involve time-consuming computation of the corresponding derivatives—i.e., Jacobian and Hessian matrices. On the other hand, because of their more general nature, UKF algorithms are more complex and, in general, require more computation time.

Yet another widely filtering technique for nonlinear (and non-Gaussian, and non-stationary) situations is the Particle Filter (PF) technique [8]. The main idea of this technique is that we simulate each probability distribution by selecting several sample points (called *particles*). We start with selecting several states distributed according to the known prior distribution of the system's state. To each of the selected states, we apply the system's dynamics—simulating the corresponding system's noise and measurement noise. The resulting sample of states (and sample of expected measurement results) represent the distributions corresponding to the next moment of time. We can then use the actual observation results obtained at this moment of time to update the distributions. This method works very well in many practical applications. For example, in geodesy, the PF techniques have been successfully used to accurately determine the trajectory on a moving vehicle based on Laser-scanner-based multi-sensor measurements [9,10].

While these techniques work well in many practical situations, they all have a common limitation: they assume that we have a reasonably detailed information about the corresponding probability distributions. In practice, often, we do not have that information. Sometimes, some approximate distributions are provided, but the actual frequencies of different noise values are very different from what these approximate distributions predict. This is a frequent situation, e.g., for measuring instruments, when the manufacturer provides us only with an upper bound on the systematic error

component (or even on the overall measurement error) without providing any information on the probabilities of different values within the given bounds; see, e.g., [11].

Such situations are known as situations with Unknown But Bounded uncertainty (UBB). Techniques for state estimation under such uncertainty have been developed since the 1960s; see, e.g., [12–14].

In the case of UBB uncertainty, for each noise component, once we know the upper bound Δ on its absolute value, the only information that we have about the actual noise value is that this value is located in the interval $[-\Delta, \Delta]$; we do not know the probabilities of different values from this interval. Once we know the bounds Δ_i on each noise value n_i ($1 \leq i \leq k$), we can therefore conclude that the tuple $n = (n_1, \dots, n_k)$ formed by these noise values belongs to the box

$$[-\Delta_1, \Delta_1] \times \dots \times [-\Delta_n, \Delta_n],$$

methods for processing such information are given, e.g., in [15–17].

We may also have some additional information about the relation between different noise values—e.g., the upper bound on the difference between noise values at two consequent moments of time. As a result, in addition to the interval containing each noise value, we may know a bounded close (hence compact) set containing the tuples of possible noise values.

There exist several techniques for dealing with such uncertainty. The most accurate techniques are the ones that use generic polytopes [18,19]—since by using polytopes, we can approximate any reasonable compact set with any given accuracy. However, to get an accurate approximation, we need to use a large number of parameters—especially in multi-dimensional case. As a result, in practice, these general methods are only efficient in low-dimensional situations.

Somewhat faster techniques emerge when we limit ourselves to a special class of polytopes called *zonotopes*; see, e.g., [1,20–23]. Mathematically, zonotopes are defined as Minkowski sums of intervals (for those who are not familiar with this notion, the definition of Minkowski sum is given in the next section). These methods are more efficient, but still mostly limited to low-dimensional cases.

The only efficient general techniques available in the higher-dimensional cases are techniques based on using ellipsoids; see, e.g., [24–26]. If we use ellipsoids, then some problems become easy to solve: e.g., if the system is linear and we know the ellipsoid that contains its initial state, then the set of all possible states at the next moment of time is also an ellipsoid, and this ellipsoid can be easily computed. Other problems are not so easy. For example, if we have two different sources of noise and each of them is described by an ellipsoid, then the set of possible values of the overall noise is no longer an ellipsoid—using the term which is explained on the next section, it is the Minkowski sum of the two original ellipsoids. To apply the ellipsoid technique to this situation, we need to enclose this Minkowski sum into an ellipsoid. We want this enclosing ellipsoid to represent the sum as accurately as possible—so, to compute this ellipsoid, we need to solve the corresponding optimization problem (this will also be described in the next section).

The above techniques take care of the situations in which we either know all the corresponding probability distributions or we do not know the probabilities at all—we only know the bounds on the noises. In practice, often, we have the probability information about some noise components, and we only know bounds on other components of the noise. For such situations, we need state estimation techniques that would take both types of uncertainty into account. For the linear case, such techniques have been developed by Benjamin Noack in his PhD dissertation [27]. In this paper, we extend these techniques to the general nonlinear case.

Following Noack, we use ellipsoids to describe the UBB uncertainty. At this stage of our research, we limit ourselves to the situations when all the probability distributions are Gaussian. Thus, we call our new filtering techniques the Ellipsoidal and Gaussian Kalman Filter (EGKF, for short).

For random uncertainties, the newly proposed method keeps the Kalman filter's recursive framework, and thus, with respect to this uncertainty, is as efficient as the original KF. Of course, since

we also use ellipsoidal uncertainty, we need to solve an optimization problem on each step, as a result of which our method requires somewhat more computation time than the usual KF.

The structure of the paper is as follows. Section 2 describes the mathematical definitions and results that are used in the following text—including the general definitions and results about the corresponding dynamical systems and filters. In Section 3, we analyze the corresponding problem, derive the formulas of the resulting EGKF algorithm, and, finally, present this algorithm in a practical ready-to-use form. In Section 4, we show, on several test cases, how the new method works—and we show that, in most cases, it indeed performs better than EKF. The last section contains conclusions and future work.

2. Mathematical Model

This section contains the mathematical definitions and results that are used in the following text—including the general definitions and results about the corresponding dynamical systems and filters. It also includes the formulation of the problem in precise terms.

2.1. Mathematical Definitions and Results Used in This Paper

The following definitions, theorems and corollaries are required for the derivation of the new filter model EGKF. Proofs of these results can be found, e.g., in [26].

Definition 1. Let n be a positive integer. By a bounded ellipsoid in $\mathbb{R}^n$ with nonempty interior (or simply ellipsoid, for short) we mean a set

$$\mathcal{E} = \mathcal{E}(\mathbf{c}, S) = \{\mathbf{x} \in \mathbb{R}^n \mid (\mathbf{x} - \mathbf{c})^T S^{-1} (\mathbf{x} - \mathbf{c}) \leq 1\}, \quad (1)$$

where $\mathbf{c} \in \mathbb{R}^n$ and S is a positive definite matrix (this will be denoted by $S > 0$). $\mathbf{c}$ is called the center of the ellipsoid $\mathcal{E}$, and S is called the shape matrix.

The shape matrix specifies the size and orientation of the ellipsoid.

Definition 2. Let A and B be sets in the Euclidean space $\mathbb{R}^n$. By the Minkowski sum $A + B$ of A and B , we mean the set of all possible values obtained by adding elements of A and B :

$$A + B = \{a + b \mid a \in A, b \in B\}. \quad (2)$$

In general, for K ellipsoids in $\mathbb{R}^n$

$$\mathcal{E}_k = \mathcal{E}(\mathbf{c}_k, S_k), (k = 1, 2, \dots, K), \quad (3)$$

their Minkowski sum

$$\mathcal{U}_K = \sum_{k=1}^K \mathcal{E}_k \quad (4)$$

is not an ellipsoid anymore; however, it is still a convex set.

In this paper, we will look for the ellipsoid with the smallest trace $\text{tr}(S)$ that contains the Minkowski sum. We selected this criterion since it corresponds to minimizing the mean square error in the probabilistic case. In principle, we could instead minimize the volume of the ellipsoid—which corresponds to minimizing the determinant $|S|$ of the shape matrix, or the largest eigenvalue of the shape matrix $\lambda_M(S)$. Minimizing the largest eigenvalue $\lambda_M(S)$ makes the ellipsoid more like a ball (in 2D case, more like a circle). Minimizing the volume sometimes leads to oblate ellipsoids, with large uncertainty in some directions.

For minimizing the trace, we will denote the corresponding minimization problem by T^+ :

$$\mathcal{E}^* = \arg \min_{\mathcal{U}_K \subseteq \mathcal{E}} \text{tr}(\mathbf{S}) \quad (\text{Problem } T^+). \quad (5)$$

It is known that the optimal ellipsoid $\mathcal{E}^*$ always exists and is unique [26].

Theorem 1. The center of the optimal ellipsoid $\mathcal{E}^*$ for Problem T^+ is given by

$$\mathbf{c}^* = \sum_{k=1}^K \mathbf{c}_k. \quad (6)$$

Theorem 2. Let $\mathcal{D}$ be the convex set of all vectors $\alpha \in \mathbb{R}^K$ for which $\alpha_k > 0$ for all k and $\sum_{k=1}^K \alpha_k = 1$. Then, for any $\alpha \in \mathcal{D}$, the ellipsoid $\mathcal{E}_\alpha = \mathcal{E}^+(\mathbf{c}^*, S_\alpha)$, with $\mathbf{c}^*$ defined by (6) and

$$S_\alpha = \sum_{k=1}^K \alpha_k^{-1} S_k \quad (7)$$

contains the Minkowski sum $\mathcal{U}_K$.

Corollary 1. (Special case of Theorem 2) When $K = 2$, we have $\alpha_1 + \alpha_2 = 1$, so the formula for S_α can be rewritten as

$$S_\alpha = \frac{1}{\alpha_1} S_1 + \frac{1}{\alpha_2} S_2 = \left(1 + \frac{1}{\beta}\right) S_1 + (1 + \beta) S_2, \quad (8)$$

where β can be any positive real number.

Proof of Corollary 1. Indeed, for each $\beta > 0$, we can take $\alpha_1 = \frac{\beta}{1+\beta}$ and $\alpha_2 = \frac{1}{1+\beta}$. $\square$

Theorem 3. In the family $\mathcal{E}_\alpha = \mathcal{E}^+(\mathbf{c}^*, S_\alpha)$, the minimal-trace ellipsoid containing the Minkowski sum of the ellipsoids $\mathcal{E}_k = \mathcal{E}^+(\mathbf{c}_k, S_k)$, $k = 1, 2, \dots, K$ is obtained for

$$S_{\alpha^*} = \left(\sum_{k=1}^K \sqrt{\text{tr}(S_k)} \right) \cdot \left(\sum_{k=1}^K \frac{S_k}{\sqrt{\text{tr}(S_k)}} \right). \quad (9)$$

Corollary 2. (Special case of Theorem 3) When $K = 2$, we have

$$S_{\alpha^*} = \left(1 + \frac{1}{\beta^*}\right) S_1 + (1 + \beta^*) S_2, \text{ where } \beta^* = \sqrt{\frac{\text{tr}(S_1)}{\text{tr}(S_2)}}. \quad (10)$$

2.2. Dynamical Systems and Linearization

In this paper, we consider the following general discrete-time nonlinear system:

$$\mathbf{x}_{k+1} = f_k(\mathbf{x}_k, \mathbf{u}_k, \mathbf{w}_k, \mathbf{a}_k), \quad (11a)$$

$$\mathbf{y}_k = h_k(\mathbf{x}_k, \mathbf{v}_k, \mathbf{b}_k), \quad (11b)$$

where:

- $\mathbf{x}_k$ is a n -dimensional state vector at time t_k ,
- $\mathbf{u}_k$ is the known input vector,
- $\mathbf{w}_k \sim \mathcal{N}(\mathbf{0}, C_k^u)$ is a Gaussian system noise with covariance matrix C_k^u ,
- $\mathbf{a}_k \in \mathcal{E}(\mathbf{0}, S_k^u)$ is an UBB perturbation with shape matrix S_k^u ,

- $\mathbf{v}_k \sim N(\mathbf{0}, C_k^z)$ is a Gaussian measurement noise with covariance matrix C_k^z ,
- $\mathbf{b}_k \in \mathcal{E}(\mathbf{0}, S_k^z)$ is an UBB perturbation with shape matrix S_k^z —all at the time t_k .

To make notations clearer, in this paper, parameters related to the first (system) equation will be denoted by u and parameters related to the second (measurement) equation will be denoted by z .

Similarly to EKF, we start with linearization: namely, we expand the right-hand side of the above equations in Taylor series and keep only linear terms in the corresponding expansion. For the system Equation (11a), we perform Taylor expansion around the point $(\mathbf{x}_k = \hat{\mathbf{x}}_k^+, \mathbf{u}_k = \mathbf{u}_k, \mathbf{w}_k = \mathbf{0}, \mathbf{a}_k = \mathbf{0})$ and get the following result:

$$\begin{aligned} \mathbf{x}_{k+1} &= f_k(\hat{\mathbf{x}}_k^+, \mathbf{u}_k, \mathbf{0}, \mathbf{0}) + \left. \frac{\partial f_k}{\partial \mathbf{x}_k} \right|_{(\hat{\mathbf{x}}_k^+, \mathbf{u}_k, \mathbf{0}, \mathbf{0})} (\mathbf{x}_k - \hat{\mathbf{x}}_k^+) + \left. \frac{\partial f_k}{\partial \mathbf{w}_k} \right|_{(\hat{\mathbf{x}}_k^+, \mathbf{u}_k, \mathbf{0}, \mathbf{0})} \mathbf{w}_k \\ &\quad + \left. \frac{\partial f_k}{\partial \mathbf{a}_k} \right|_{(\hat{\mathbf{x}}_k^+, \mathbf{u}_k, \mathbf{0}, \mathbf{0})} \mathbf{a}_k + \dots \\ &\approx f_k(\hat{\mathbf{x}}_k^+, \mathbf{u}_k, \mathbf{0}, \mathbf{0}) + F_{x,k}(\mathbf{x}_k - \hat{\mathbf{x}}_k^+) + F_{w,k} \mathbf{w}_k + F_{a,k} \mathbf{a}_k \\ &= F_{x,k} \mathbf{x}_k + [f_k(\hat{\mathbf{x}}_k^+, \mathbf{u}_k, \mathbf{0}, \mathbf{0}) - F_{x,k} \hat{\mathbf{x}}_k^+] + F_{w,k} \mathbf{w}_k + F_{a,k} \mathbf{a}_k \\ &= F_{x,k} \mathbf{x}_k + \tilde{\mathbf{u}}_k + F_{w,k} \mathbf{w}_k + F_{a,k} \mathbf{a}_k. \end{aligned} \quad (12)$$

Here, we denoted $\tilde{\mathbf{u}}_k = f_k(\hat{\mathbf{x}}_k^+, \mathbf{u}_k, \mathbf{0}, \mathbf{0}) - F_{x,k} \hat{\mathbf{x}}_k^+$ and $F_{x,k}$, $F_{w,k}$, and $F_{a,k}$ denote the corresponding derivatives. Note that, in spite of the fact that our uncertainty model is different from the purely probabilistic model underlying the EKF technique, the resulting formulas are similar to the corresponding EKF formulas—since at this stage, we did not yet use any information about the uncertainty.

For the measurement Equation (11b), a similar Taylor expansion around the natural point $(\mathbf{x}_k = \hat{\mathbf{x}}_k^-, \mathbf{v}_k = \mathbf{0}, \mathbf{b}_k = \mathbf{0})$ leads to the following formula:

$$\begin{aligned} \mathbf{y}_k &= h_k(\hat{\mathbf{x}}_k^-, \mathbf{0}, \mathbf{0}) + \left. \frac{\partial h_k}{\partial \mathbf{x}_k} \right|_{(\hat{\mathbf{x}}_k^-, \mathbf{0}, \mathbf{0})} (\mathbf{x}_k - \hat{\mathbf{x}}_k^-) + \left. \frac{\partial h_k}{\partial \mathbf{v}_k} \right|_{(\hat{\mathbf{x}}_k^-, \mathbf{0}, \mathbf{0})} \mathbf{v}_k \\ &\quad + \left. \frac{\partial h_k}{\partial \mathbf{b}_k} \right|_{(\hat{\mathbf{x}}_k^-, \mathbf{0}, \mathbf{0})} \mathbf{b}_k + \dots \\ &\approx h_k(\hat{\mathbf{x}}_k^-, \mathbf{0}, \mathbf{0}) + H_{x,k}(\mathbf{x}_k - \hat{\mathbf{x}}_k^-) + H_{v,k} \mathbf{v}_k + H_{b,k} \mathbf{b}_k \\ &= H_{x,k} \mathbf{x}_k + \tilde{\mathbf{z}}_k + H_{v,k} \mathbf{v}_k + H_{b,k} \mathbf{b}_k. \end{aligned} \quad (13)$$

Here, we denoted $\tilde{\mathbf{z}}_k = h_k(\hat{\mathbf{x}}_k^-, \mathbf{0}, \mathbf{0}) - H_{x,k} \hat{\mathbf{x}}_k^-$. Notice that if the measurement equation is linear, then $\tilde{\mathbf{z}}_k = \mathbf{0}$.

Thus, from the original system (11), we get the following linearized system:

$$\mathbf{x}_{k+1} = F_{x,k} \mathbf{x}_k + \tilde{\mathbf{u}}_k + F_{w,k} \mathbf{w}_k + F_{a,k} \mathbf{a}_k, \quad (14a)$$

$$\mathbf{y}_k = H_{x,k} \mathbf{x}_k + \tilde{\mathbf{z}}_k + H_{v,k} \mathbf{v}_k + H_{b,k} \mathbf{b}_k. \quad (14b)$$

These equations describe the change in the (unknown) true state $\mathbf{x}_k$. In practice, we only know the *estimates* for the state. Following the general idea of Kalman filter, at each moment of time t_k , we will consider:

- an *a priori* estimate of this state—i.e., what we predict based on the information available at the previous moment of time—and
- an *a posteriori* estimate—what we get after taking into account the results of measurements performed at this moment of time.

In this paper, we will denote the *a priori* estimate by $\hat{\mathbf{x}}_k^-$ and the *a posteriori* estimate by $\hat{\mathbf{x}}_k^+$.

Since we consider the case when the system has both the random and the unknown-but-bounded error components (14), both a priori and a posteriori estimates should have the same form:

$$\hat{\mathbf{x}}_k^- = \mathbf{c}_k^- + \mathbf{g}_k^- + \mathbf{e}_k^-, \quad (15)$$

$$\hat{\mathbf{x}}_k^+ = \mathbf{c}_k^+ + \mathbf{g}_k^+ + \mathbf{e}_k^+, \quad (16)$$

where $\mathbf{c}_k^-$ and $\mathbf{c}_k^+$ are points, $\mathbf{g}_k^- \sim N(\mathbf{0}, C_k^-)$ and $\mathbf{g}_k^+ \sim N(\mathbf{0}, C_k^+)$ are random variables with Gaussian distribution, $\mathbf{e}_k^- \in \mathcal{E}(\mathbf{0}, S_k^-)$ and $\mathbf{e}_k^+ \in \mathcal{E}(\mathbf{0}, S_k^+)$ are two points within two ellipsoids with shape matrices S_k^- and S_k^+ respectively.

Figure 1 illustrates shows how a state $\mathbf{x}_0$ can be decomposed into the sum of the point estimate $\mathbf{c}_0$ and two perturbations: the UBB noise component $\mathbf{e}_0$ and the Gaussian noise component $\mathbf{g}_0$.

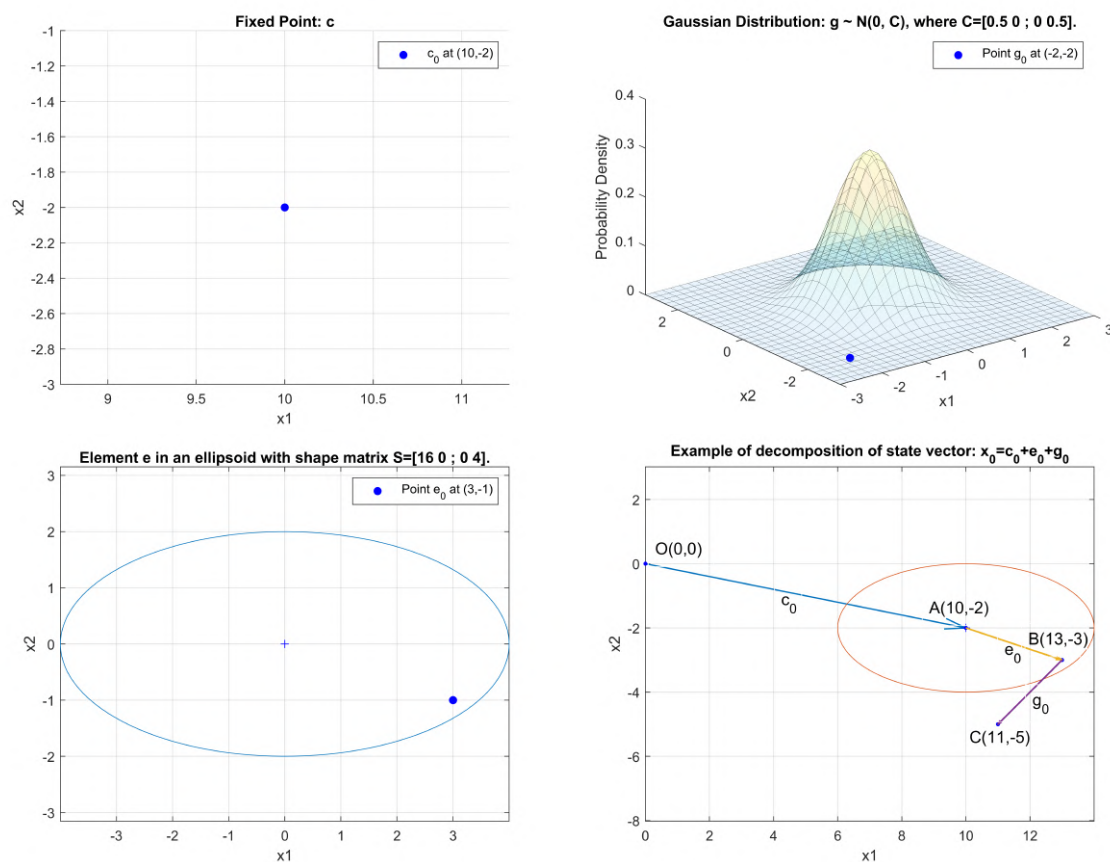


Figure 1. Decomposition of state vector $\mathbf{x}_0 = \mathbf{c}_0 + \mathbf{e}_0 + \mathbf{g}_0$.

Our objective is to use the measurement results to generate the a posteriori estimate $\hat{\mathbf{x}}_0^+$, i.e., in precise terms, to compute the point estimate $\mathbf{c}_k^+$ and the matrices C_k^+ and S_k^+ .

3. Derivation of Ellipsoidal and Gaussian Kalman Filter

In this section we derive the EGKF model. The derivation is based on the results presented in Section 2.1. The derivation is summarized by the two theorems presented in the following subsection.

3.1. Prediction

As we have mentioned, our algorithm will describe, moment by moment, how our knowledge of the system's state changes. At the first moment of time, we have information $\hat{\mathbf{x}}_0^+$ about the initial state of the system. For each moment $t_k, k = 0, 1, \dots$, our algorithm will describe the transition from moment t_k to the next moment t_{k+1} . So, at moment t_1 , after applying our algorithm, we will know the

a posteriori estimate $\hat{\mathbf{x}}_1^+$ for the state of the system at this moment. Then, at moment t_2 , we will get an a posteriori estimate $\hat{\mathbf{x}}_2^+$ for the state of the system at moment t_2 , etc.

Let us describe the transition from moment t_k to the next moment t_{k+1} . At moment t_k , we know the a posteriori estimate of the state, i.e., we know that

$$\hat{\mathbf{x}}_k^+ = \mathbf{c}_k^+ + \mathbf{g}_k^+ + \mathbf{e}_k^+, \quad (17)$$

where $\mathbf{g}_k^+ \sim N(\mathbf{0}, C_k^+)$, $\mathbf{e}_k^+ \in \mathcal{E}(\mathbf{0}, S_k^+)$ and $\mathbf{c}_k^+, C_k^+, S_k^+$ are given.

As we showed in the previous section, the a priori estimate at the next moment of time has the form

$$\hat{\mathbf{x}}_{k+1}^- = F_{x,k} \hat{\mathbf{x}}_k^+ + \tilde{\mathbf{u}}_k + F_{w,k} \mathbf{w}_k + F_{a,k} \mathbf{a}_k. \quad (18)$$

The following theorem describes this a prior estimate in the desired form (15).

Theorem 4. The a priori estimate $\hat{\mathbf{x}}_{k+1}^-$, defined by the Formula (18), is equal to: $\hat{\mathbf{x}}_{k+1}^- = \mathbf{c}_{k+1}^- + \mathbf{g}_{k+1}^- + \mathbf{e}_{k+1}^-$, where

$$\mathbf{c}_{k+1}^- = F_{x,k} \mathbf{c}_k^+ + \tilde{\mathbf{u}}_k, \quad (19)$$

$$\mathbf{g}_{k+1}^- \sim N(\mathbf{0}, C_{k+1}^-), \quad (20)$$

$$\mathbf{e}_{k+1}^- \in \mathcal{E}(\mathbf{0}, S_{k+1}^-), \quad (21)$$

$$C_{k+1}^- = F_{x,k} C_k^+ F_{x,k}^T + F_{w,k} C_k^u F_{w,k}^T, \quad (22)$$

and

$$S_{k+1}^- = \left(\sqrt{\text{tr}(F_{x,k} S_k^+ F_{x,k}^T)} + \sqrt{\text{tr}(F_{a,k} S_k^u F_{a,k}^T)} \right) \cdot \left(\frac{F_{x,k} S_k^+ F_{x,k}^T}{\sqrt{\text{tr}(F_{x,k} S_k^+ F_{x,k}^T)}} + \frac{F_{a,k} S_k^u F_{a,k}^T}{\sqrt{\text{tr}(F_{a,k} S_k^u F_{a,k}^T)}} \right). \quad (23)$$

Proof. From (17) and (18), we get

$$\begin{aligned} \hat{\mathbf{x}}_{k+1}^- &= F_{x,k} (\mathbf{c}_k^+ + \mathbf{g}_k^+ + \mathbf{e}_k^+) + \tilde{\mathbf{u}}_k + F_{w,k} \mathbf{w}_k + F_{a,k} \mathbf{a}_k \\ &= F_{x,k} \mathbf{c}_k^+ + F_{x,k} \mathbf{g}_k^+ + F_{x,k} \mathbf{e}_k^+ + \tilde{\mathbf{u}}_k + F_{w,k} \mathbf{w}_k + F_{a,k} \mathbf{a}_k \\ &= \mathbf{c}_{k+1}^- + \mathbf{g}_{k+1}^- + \mathbf{e}_{k+1}^-, \end{aligned} \quad (24)$$

where we denoted:

$$\mathbf{c}_{k+1}^- = F_{x,k} \mathbf{c}_k^+ + \tilde{\mathbf{u}}_k, \quad (25)$$

$$\mathbf{g}_{k+1}^- = F_{x,k} \mathbf{g}_k^+ + F_{w,k} \mathbf{w}_k, \quad (26)$$

$$\mathbf{e}_{k+1}^- = F_{x,k} \mathbf{e}_k^+ + F_{a,k} \mathbf{a}_k. \quad (27)$$

Here,

$$\mathbf{g}_k^+ \sim N(\mathbf{0}, C_k^+), \mathbf{w}_k \sim N(\mathbf{0}, C_k^u), \quad (28)$$

therefore according to the properties of Gaussian distributions and (26),

$$\mathbf{g}_{k+1}^- \sim N(\mathbf{0}, C_{k+1}^-), \quad (29)$$

where

$$C_{k+1}^- = F_{x,k} C_k^+ F_{x,k}^T + F_{w,k} C_k^u F_{w,k}^T. \quad (30)$$

Also, here,

$$\mathbf{e}_k^+ \in \mathcal{E}(\mathbf{0}, S_k^+), \mathbf{a}_k \in \mathcal{E}(\mathbf{0}, S_k^u), \quad (31)$$

therefore according to Corollary 2:

$$\mathbf{e}_{k+1}^- \in \mathcal{E}(\mathbf{0}, S_{k+1}^-), \quad (32)$$

where

$$S_{k+1}^- = \left(\sqrt{\text{tr}(F_{x,k} S_k^+ F_{x,k}^T)} + \sqrt{\text{tr}(F_a S_k^u F_a^T)} \right) \cdot \left(\frac{F_{x,k} S_k^+ F_{x,k}^T}{\sqrt{\text{tr}(F_{x,k} S_k^+ F_{x,k}^T)}} + \frac{F_a S_k^u F_a^T}{\sqrt{\text{tr}(F_a S_k^u F_a^T)}} \right). \quad (33)$$

The proof is complete. $\square$

3.2. Filtering

In the previous section, we show how to generate the a priori estimate for the system's state. As a result, at each moment t_k , we have the following estimate:

$$\hat{\mathbf{x}}_k^- = \mathbf{c}_k^- + \mathbf{g}_k^- + \mathbf{e}_k^-, \quad (34)$$

where $\mathbf{g}_k^- \sim N(\mathbf{0}, C_k^-)$, $\mathbf{e}_k^+ \in \mathcal{E}(\mathbf{0}, S_k^-)$ and $\mathbf{c}_k^-$, C_k^- , S_k^- are known.

Let us now describe how we take the observation results $\mathbf{y}_k$ into account, and thus, transform the a priori estimate into an a posteriori one. In line with the general KF idea, to compensate for the prediction error $\mathbf{y}_k - h_k(\hat{\mathbf{x}}_k^-, \mathbf{0}, \mathbf{0})$, we add, to the a priori estimate, a term proportional to this error, i.e., compute the following a posteriori estimate

$$\hat{\mathbf{x}}_k^+ = \hat{\mathbf{x}}_k^- + K_k[\mathbf{y}_k - h_k(\hat{\mathbf{x}}_k^-, \mathbf{0}, \mathbf{0})], \quad (35)$$

for an appropriate matrix K_k . In this section, we consider what happens for a general selection of the gain factor K_k . In the next subsection, we will show how to select the optimal gain factor.

Let us show how the resulting a posteriori estimate can be represented in the desired form (15). To come up with this form, we introduce a new parameter β ; the optimal value of this parameter will also be described in the next subsection.

Theorem 5. Let $\beta > 0$. Then, an a posteriori estimate $\hat{\mathbf{x}}_k^+$, as defined by the Formula (35), has the form $\hat{\mathbf{x}}_k^+ = \mathbf{c}_k^+ + \mathbf{g}_k^+ + \mathbf{e}_k^+$, where:

$$\mathbf{c}_k^+ = (I - K_k H_{x,k}) \mathbf{c}_k^- + K_k(\mathbf{y}_k - \tilde{\mathbf{z}}_k), \quad (36)$$

$$\mathbf{g}_k^+ \sim N(\mathbf{0}, C_k^+), \quad (37)$$

$$\mathbf{e}_k^+ \in \mathcal{E}(\mathbf{0}, S_k^+), \quad (38)$$

$$C_k^+ = (I - K_k H_{x,k}) C_k^- (I - K_k H_{x,k})^T + K_k H_{v,k} C_k^z H_{v,k}^T K_k^T, \quad (39)$$

and

$$S_k^+(\beta) = \left(1 + \frac{1}{\beta} \right) (I - K_k H_{x,k}) S_k^- (I - K_k H_{x,k})^T + (1 + \beta) K_k H_{b,k} S_k^z H_{b,k}^T K_k^T. \quad (40)$$

Proof. From (34) and (35) we get

$$\begin{aligned}
 \hat{\mathbf{x}}_k^+ &= \hat{\mathbf{x}}_k^- + K_k[\mathbf{y}_k - h_k(\hat{\mathbf{x}}_k^-, \mathbf{0}, \mathbf{0})] \\
 &= \hat{\mathbf{x}}_k^- + K_k[\mathbf{y}_k - H_{x,k}\hat{\mathbf{x}}_k^- - \tilde{\mathbf{z}}_k] \\
 &= (I - K_k H_{x,k})\hat{\mathbf{x}}_k^- + K_k(\mathbf{y}_k - \tilde{\mathbf{z}}_k) \\
 &= (I - K_k H_{x,k})(\mathbf{c}_k^- + \mathbf{g}_k^- + \mathbf{e}_k^-) + K_k(H_{x,k}\mathbf{x}_k + H_{v,k}\mathbf{v}_k + H_{b,k}\mathbf{b}_k) \\
 &= \mathbf{c}_k^+ + \mathbf{g}_k^+ + \mathbf{e}_k^k,
 \end{aligned} \tag{41}$$

where we denoted

$$\mathbf{c}_k^+ = (I - K_k H_{x,k})\mathbf{c}_k^- + K_k H_{x,k}\mathbf{x}_k, \tag{42}$$

$$\mathbf{g}_k^+ = (I - K_k H_{x,k})\mathbf{g}_k^- + K_k H_{v,k}\mathbf{v}_k, \tag{43}$$

$$\mathbf{e}_k^+ = (I - K_k H_{x,k})\mathbf{e}_k^- + K_k H_{b,k}\mathbf{b}_k. \tag{44}$$

For the point estimate part, we have $\mathbf{v}_k = \mathbf{0}$ and $\mathbf{b}_k = \mathbf{0}$, therefore from (42) we get

$$\mathbf{c}_k^+ = (I - K_k H_{x,k})\mathbf{c}_k^- + K_k(\mathbf{y}_k - \tilde{\mathbf{z}}_k). \tag{45}$$

Here,

$$\mathbf{g}_k^- \sim N(\mathbf{0}, C_k^-), \mathbf{v}_k \sim N(\mathbf{0}, C_k^z), \tag{46}$$

therefore according to properties of Gaussian distributions and (37),

$$\mathbf{g}_k^+ \sim N(\mathbf{0}, C_k^+), \tag{47}$$

where

$$C_k^+ = (I - K_k H_{x,k})C_k^- (I - K_k H_{x,k})^T + F_{v,k}C_k^z F_{v,k}^T. \tag{48}$$

Also here,

$$\mathbf{e}_k^- \in \mathcal{E}(\mathbf{0}, S_k^-), \mathbf{b}_k \in \mathcal{E}(\mathbf{0}, S_k^z), \tag{49}$$

then according to Corollary 1, for any $\beta > 0$, we get

$$\mathbf{e}_k^+ \in \mathcal{E}(\mathbf{0}, S_k^+), \tag{50}$$

where

$$S_k^+(\beta) = \left(1 + \frac{1}{\beta}\right) (I - K_k H_{x,k})S_k^- (I - K_k H_{x,k})^T + (1 + \beta)K_k H_{b,k}S_k^z H_{b,k}^T K_k^T. \tag{51}$$

The proof is complete. $\square$

3.3. Optimization Problem

To finalize our algorithm, the only thing left is to find the optimal gain factor K_k . Similar to KF and to EKF, we will find the gain factor that minimizes the mean square error of the posteriori estimation.

In our case, this error has two components: the statistical component and the UBB component. To combine these errors into a single objective function, we need to decide how much weight we give to each component. This relative weight will be described by a parameter $\eta \in [0, 1]$, so that the resulting cost function will have the following form:

$$J(\beta) = (1 - \eta)\text{tr}(C_k^+) + \eta\text{tr}[S_k^+(\beta)]. \tag{52}$$

This expression represents the overall uncertainty of the posteriori estimation. We will find the gain factor K_k and the value β that minimize this cost function.

Plugging (48) and (51) into (52) we get:

$$\begin{aligned}
 J(\beta) &= (1-\eta) \text{tr} \left[(I - K_k H_{x,k}) C_k^- (I - K_k H_{x,k})^T \right] \\
 &\quad + (1-\eta) \text{tr} \left[K_k H_{v,k} C_k^z H_{v,k}^T K_k^T \right] \\
 &\quad + \eta \left(1 + \frac{1}{\beta} \right) \text{tr} \left[(I - K_k H_{x,k}) S_k^- (I - K_k H_{x,k})^T \right] \\
 &\quad + \eta(1+\beta) \text{tr} \left(K_k H_{b,k} S_k^z H_{b,k}^T K_k^T \right) \\
 &\triangleq (1-\eta) \text{tr} \left[(I - K_k H_{x,k}) C_k^- (I - K_k H_{x,k})^T \right] \\
 &\quad + (1-\eta) \text{tr} \left[K_k H_{v,k} C_k^z H_{v,k}^T K_k^T \right] + \eta \left(1 + \frac{1}{\beta} \right) M + \eta(1+\beta)N,
 \end{aligned} \tag{53}$$

where we denoted $M = \text{tr} \left[(I - K_k H_{x,k}) S_k^- (I - K_k H_{x,k})^T \right]$ and $N = \text{tr} \left(K_k H_{b,k} S_k^z H_{b,k}^T K_k^T \right)$.

Let us first find the optimal value β . Since $M > 0$ and $N > 0$, taking into account that the arithmetic mean is always larger than or equal that the geometric mean, we conclude that

$$\left(1 + \frac{1}{\beta} \right) M + (1+\beta)N = M + N + \frac{1}{\beta}M + \beta N \geq M + N + 2\sqrt{MN}. \tag{54}$$

In particular, when $\frac{1}{\beta}M = \beta N$, i.e., when $M = \beta^2 N$ and $\beta = \sqrt{\frac{M}{N}}$, we have:

$$\left(1 + \frac{1}{\beta} \right) M + (1+\beta)N = M + N + 2\sqrt{MN} = \left(\sqrt{M} + \sqrt{N} \right)^2. \tag{55}$$

Thus, the smallest possible value of the cost function is attained when $\beta = \sqrt{\frac{M}{N}}$. In this case, the cost function $J(\beta)$, as described by the Formula (52), has the following form:

$$\begin{aligned}
 J(\beta) &= (1-\eta) \text{tr} \left[(I - K_k H_{x,k}) C_k^- (I - K_k H_{x,k})^T \right] \\
 &\quad + (1-\eta) \text{tr} \left[K_k H_{v,k} C_k^z H_{v,k}^T K_k^T \right] + \eta \left(\sqrt{M} + \sqrt{N} \right)^2.
 \end{aligned} \tag{56}$$

To find the optimal value of the gain factor K_k , let us differentiate the resulting expression by K_k and equate the resulting derivative to 0.

Notice that

$$\begin{aligned}
 \frac{\partial \sqrt{M}}{\partial K_k} &= \frac{1}{2} M^{-\frac{1}{2}} \frac{\partial M}{\partial K_k} = -\frac{1}{2\sqrt{M}} (I - K_k H_{x,k}) \left(S_k^{-T} + S_k^- \right) H_{x,k}^T \\
 &= -\frac{1}{\sqrt{M}} (I - K_k H_{x,k}) S_k^- H_{x,k}^T, \\
 \frac{\partial \sqrt{N}}{\partial K_k} &= \frac{1}{2} N^{-\frac{1}{2}} \frac{\partial N}{\partial K_k} = \frac{1}{2\sqrt{N}} K_k H_{b,k} \left(S_k^{z,T} + S_k^z \right) H_{b,k}^T \\
 &= \frac{1}{\sqrt{N}} K_k H_{b,k} S_k^z H_{b,k}^T.
 \end{aligned} \tag{57}$$

Therefore,

$$\begin{aligned}
 \frac{\partial J}{\partial K_k} &= 2(1-\eta)(K_k H_{x,k} - I) C_k^- H_{x,k}^T + 2(1-\eta) K_k H_{v,k} C_k^z H_{v,k}^T \\
 &\quad + 2\eta \left(1 + \frac{1}{\beta} \right) (K_k H_{x,k} - I) S_k^- H_{x,k}^T + 2\eta(1+\beta) K_k H_{b,k} S_k^z H_{b,k}^T.
 \end{aligned} \tag{58}$$

Equating the resulting derivative $\frac{\partial J}{\partial K_k}$ to 0, we get the following explicit formula for the optimal gain K_k :

$$K_k = A^{-1} \cdot \left[(1 - \eta) C_k^- H_{x,k}^T + \eta \left(1 + \frac{1}{\beta} \right) S_k^- H_{x,k}^T \right], \quad (59)$$

where we denoted:

$$A = (1 - \eta) H_{x,k} C_k^- H_{x,k}^T + (1 - \eta) H_{v,k} C_k^z H_{v,k}^T + \eta \left(1 + \frac{1}{\beta} \right) H_{x,k} S_k^- H_{x,k}^T + \eta (1 + \beta) H_{b,k} S_k^z H_{b,k}^T. \quad (60)$$

The only thing left to determine is the parameter β . For each β , substituting the expression (48), (51) and (59) into the formula for the cost function, we get the corresponding value of $J(\beta)$. We thus need to find the value β^* that minimizes this cost function, i.e., that solves the following minimization problem:

$$\begin{aligned} \min_{\beta} J(\beta) \\ \text{s.t. } \beta \in (0, +\infty) \subset \mathbb{R}^1. \end{aligned} \quad (61)$$

This optimization problem is more complex than the quadratic optimization problems which usually occur in filtering. Indeed, for quadratic objective functions, derivatives are linear and thus, when we equate all the derivatives to 0, we get an easy-to-solve system of linear equations. In contrast, for non-quadratic objective functions, equating derivatives to 0 leads to a difficult-to-solve system of nonlinear equations. Another reason why non-quadratic optimization problems are often difficult to solve is that traditional methods for solving these problems—such as gradient descent—often lead to a local minimum and not to the desired global minimum of the function. Good news is that in our case, similarly to [27], we can show that the cost function is convex. For convex objective functions, every local minimum is also a global minimum; see, e.g., [28]. Because of this, there exist many efficient techniques for solving convex optimization problems; for example, we can use the corresponding algorithm from the *INTLAB* toolbox [28].

Once we find the optimal value β^* , we can substitute it into the Formulas (48), (51) and (59), and get the desired a posteriori state estimate $\hat{x}_k^+$.

It should be mentioned that this estimate depends on the parameter η —which was introduced to balance the random uncertainty and UBB uncertainty. Specifically, we should mention three important cases—as described, e.g., in [26]:

- (1) When $\eta = \frac{1}{2}$, the stochastic uncertainty and UBB uncertainty have the same weight. and the optimal gain factor K_k contains no η in this case. This setting is recommended to users when there is no expert-based information about η available.
- (2) When $\eta = 0$, we get:

$$K_k(\beta) = C_k^- H_{x,k}^T \cdot \left[H_{x,k} C_k^- H_{x,k}^T + H_{v,k} C_k^z H_{v,k}^T \right]^{-1}. \quad (62)$$

In this case the filtering model contains no UBB uncertainty and the EGKF algorithm reduces to EKF [6].

- (3) When $\eta = 1$, the model now only contains UBB uncertainty. In this case,

$$K_k(\beta) = \left(1 + \frac{1}{\beta} \right) S_k^- H_{x,k}^T \cdot \left[\left(1 + \frac{1}{\beta} \right) H_{x,k} S_k^- H_{x,k}^T + (1 + \beta) H_{b,k} S_k^z H_{b,k}^T \right]^{-1}. \quad (63)$$

Based on Theorems 4 and 5, and Equations (59) and (61), we can summarize the EGKF model into the following Algorithm 1:

Algorithm 1 Ellipsoidal and Gaussian Kalman filter model.1: **Input:**

- System information: $f_k, h_k, C_k^u, C_k^z, S_k^u, S_k^z$.
- Initial values: c_0^+, C_0^+, S_0^+ .
- Parameters: n, η .

2: **for** $k=1,2,\dots,K$ **do**3: **Calculation:**

$$\mathbf{c}_k^- = F_{x,k-1} \mathbf{c}_{k-1}^+ + \tilde{\mathbf{u}}_{k-1}. \quad (64)$$

$$C_k^- = F_{x,k-1} C_{k-1}^+ F_{x,k-1}^T + F_{w,k-1} C_{k-1}^u F_{w,k-1}^T. \quad (65)$$

$$B = \left(\sqrt{\text{tr} \left(F_{x,k-1} S_{k-1}^+ F_{x,k-1}^T \right)} + \sqrt{\text{tr} \left(F_{a,k-1} S_{k-1}^u F_{a,k-1}^T \right)} \right). \quad (66)$$

$$S_k^- = B \cdot \left(\frac{F_{x,k-1} S_{k-1}^+ F_{x,k-1}^T}{\sqrt{\text{tr} \left(F_{x,k-1} S_{k-1}^+ F_{x,k-1}^T \right)}} + \frac{F_{a,k-1} S_{k-1}^u F_{a,k-1}^T}{\sqrt{\text{tr} \left(F_{a,k-1} S_{k-1}^u F_{a,k-1}^T \right)}} \right). \quad (67)$$

$$A = (1 - \eta) H_{x,k} C_k^- H_{x,k}^T + (1 - \eta) H_{v,k} C_k^z H_{v,k}^T + \eta \left(1 + \frac{1}{\beta} \right) H_{x,k} S_k^- H_{x,k}^T + \eta (1 + \beta) H_{b,k} S_k^z H_{b,k}^T. \quad (68)$$

$$K_k = A^{-1} \cdot \left[(1 - \eta) C_k^- H_{x,k}^T + \eta \left(1 + \frac{1}{\beta} \right) S_k^- H_{x,k}^T \right]. \quad (69)$$

$$\mathbf{c}_k^+ = (I - K_k H_{x,k}) \mathbf{c}_k^- + K_k (\mathbf{y}_k - \tilde{\mathbf{z}}_k). \quad (70)$$

$$C_k^+ = (I - K_k H_{x,k}) C_k^- (I - K_k H_{x,k})^T + F_{v,k} C_k^z F_{v,k}^T. \quad (71)$$

$$S_k^+(\beta) = \left(1 + \frac{1}{\beta} \right) (I - K_k H_{x,k}) S_k^- (I - K_k H_{x,k})^T + (1 + \beta) K_k H_{b,k} S_k^z H_{b,k}^T K_k^T. \quad (72)$$

$$\beta^* = \arg \min_{\beta > 0} \{ (1 - \eta) \text{tr} [C_k^+(\beta)] + \eta \text{tr} [S_k^+(\beta)] \}. \quad (73)$$

4: **Output:**

$$\mathbf{c}_k^+, C_k^+, \text{ and } S_k^+.$$

5: **end for****4. Applications**

This section contains the result of two test of applying EGKF to simulated data sets. We also applied EKF to both problems, and compared the root-mean-square error (RMSE) of the two methods.

- In the first test, EGKF was applied on a highly nonlinear benchmark problem. To compare the result, we performed 100 simulations.
- In the second test, we applied EGKF on a two-dimensional trajectory estimation problem. For this problem, we performed 1000 Monte Carlo runs.

To simulate the UBB uncertainty, we used uniform distributions on the corresponding ellipsoids.

4.1. Example 1: A Highly Nonlinear Benchmark Example

We consider the following example:

$$\mathbf{x}_{k+1} = \frac{1}{2}\mathbf{x}_k + \frac{25\mathbf{x}_k}{1 + \mathbf{x}_k^2} + 8 \cos[1.2(k-1)] + \mathbf{w}_k + \mathbf{a}_k, \quad (74)$$

$$\mathbf{y}_k = \frac{1}{20}\mathbf{x}_k^2 + \mathbf{v}_k + \mathbf{b}_k, \quad (75)$$

where:

- $\mathbf{x}_k$ is a 1-dimensional scalar,
- $\mathbf{u}_k = 8 \cos[1.2(k-1)]$ is the input vector,
- $\mathbf{w}_k \sim N(\mathbf{0}, 1)$ is Gaussian noise,
- $\mathbf{w}_{i,k} \in \mathcal{E}(\mathbf{0}, 9)$ is the unknown but bounded perturbation; in this 1-D case the ellipsoid is simply the interval $[-3, 3]$,
- $\mathbf{v}_k \sim N(\mathbf{0}, 1)$ is Gaussian measurement noise,
- $\mathbf{b}_k \in \mathcal{E}(\mathbf{0}, 4)$ is the unknown but bounded perturbation—in this case, the ellipsoid is the interval $[-2, 2]$.

Here:

- the initial true state is $\mathbf{x}_0 = 0.1$,
- the initial state estimate is $\hat{\mathbf{x}}_0 = \mathbf{x}_0$,
- the initial estimate for covariance matrix is $\mathbf{C}_0^+ = 2$,
- the initial estimate for the shape matrix is $\mathbf{S}_0^+ = 1 \times 10^{-3}$.

This discrete-time dynamical system is a known benchmark in the nonlinear estimation theory; see, e.g., [6,8,29]. A high degree of nonlinearity in both process and measurement equations makes state estimation problem for this system very difficult.

We use this example to show that the new EGKF method behaves better than the traditional first order EKF when UBB uncertainties are taken into account.

We used a simulation length of 50 time steps. Following recommendations from the previous section, we selected the weight parameter $\eta = 0.5$.

We repeated our simulation 100 times. Figure 2 shows the results of applying both EGKF and EKF for several of the 100 runs, namely, for runs 25, 50, 75, and 100.

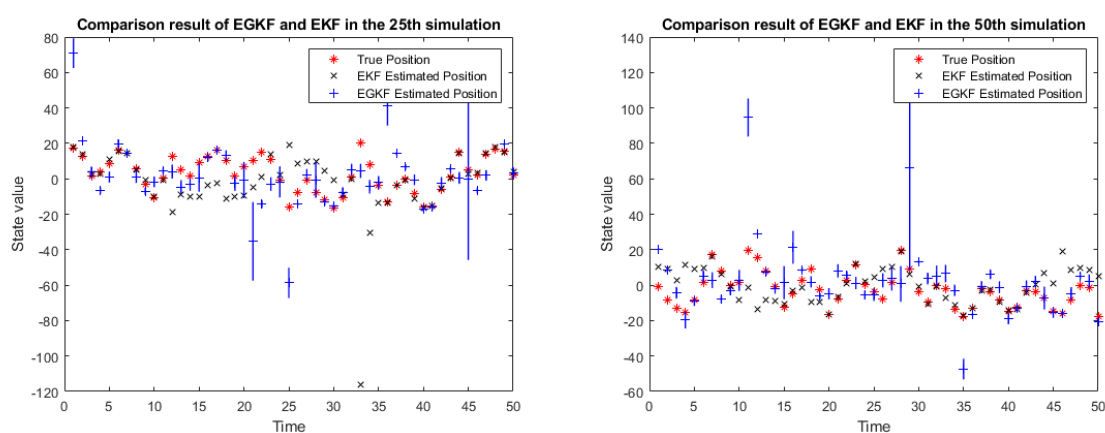


Figure 2. Cont.

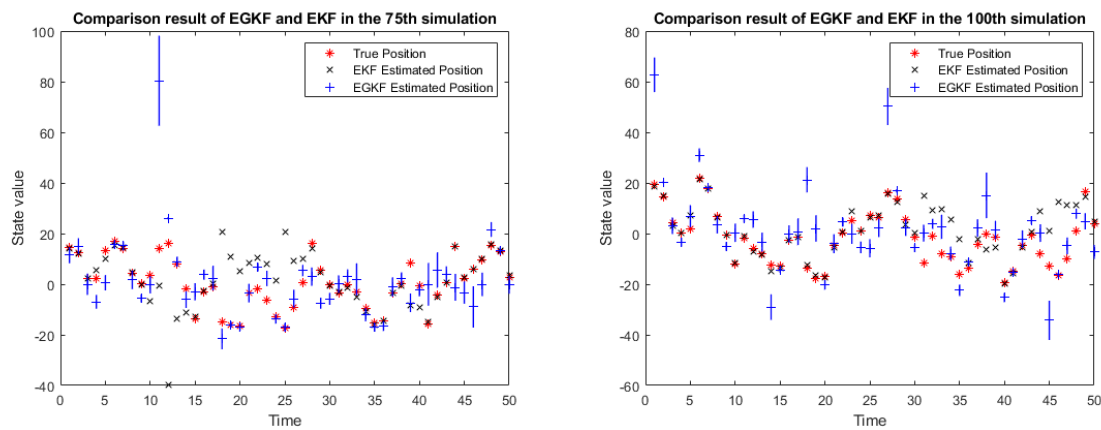


Figure 2. Comparison results in 4 simulations.

In the above figure:

- the red stars denote the true state,
- the black crosses represent the EKF estimates of the state,
- the blue lines represent the UBB ellipsoids estimated by EGKF (in this 1D case they are intervals),
- the blue plus signs mark the centers of the output ellipsoids.

One can see that the centers of the ellipsoids are, in general, different given from the EKF estimates.

For each of the two methods and for each of the 100 iterations, at each moment of time, we can calculate the difference between the actual state and the corresponding estimate. (As estimates corresponding to EGKF, we took the centers of the corresponding a posteriori ellipsoids.)

Based on 50 moments of time, we get a vector consisting of 50 such actual-state-vs.-estimated-state differences. To compare the vectors corresponding to two different methods, we calculated the l_2 norms of these vectors—this is equivalent to comparing the root-mean-square estimation errors. The resulting values are presented in Table 1 for simulations 1, 11, ..., 91.

Table 1. Comparison of Ellipsoidal and Gaussian Kalman Filter (EGKF) and Extended Kalman Filter (EKF) in 10 simulations.

k	1	11	21	31	41	51	61	71	81	91
EGKF	53.59	116.38	70.38	102.95	100.16	75.81	82.11	48.71	66.30	81.55
EKF	72.49	116.64	230.08	234.69	80.73	31.32	109.53	79.69	8.63	64.95

In most cases, the EGKF leads to a smaller mean squared estimation error. Over all 100 simulations:

- the average l_2 norm of the EGKF estimates is 148.70, while
- for the extended Kalman filter, the average l_2 norm is much higher: it is equal to 192.29.

Thus, we conclude that, on average, the new EGKF algorithm performs much better than EKF. We got a similar conclusion when, instead of comparing the l_2 norms, we compared:

- the l_1 norms—which correspond to comparing mean absolute values of the estimation errors, and
- the l_∞ norms—which correspond to comparing the largest estimation errors.

4.2. Example 2: Two-Dimensional Trajectory Estimation

In the second test, we used 1000-times Monte Carlo simulations to compare the performance of EGKF and EKF on a 2D trajectory estimation problem from [10].

In this problem, a vehicle moves on a plane, following a curved trajectory. The state vector $\mathbf{x} = (x, y, v_x, v_y)$ contains positions and velocities of the vehicle, in x-direction and y-direction, respectively. After linearization, we get the following equation:

$$\mathbf{x}_{k+1} = F_k \mathbf{x}_k + \mathbf{w}_k + \mathbf{a}_k. \quad (76)$$

Here:

- $\mathbf{x}_k = (x_k, y_k, v_{x,k}, v_{y,k})$ is the state vector at time t_k ;
- the transition matrix F_k has the form:

$$F_k = \begin{pmatrix} 1 & 0 & \Delta t & 0 \\ 0 & 1 & 0 & \Delta t \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}; \quad (77)$$

- $\mathbf{w}_k$ is Gaussian noise with covariance matrix C_k^u ,
- $\mathbf{a}_k$ is the unknown but bounded uncertainty, which is bounded by an ellipsoid with shape matrix S_k^u .

In total, in each of the 1000 simulations, we observed 500 epochs with time step $\Delta t = 0.1$ s. In our formulas, as units of time, distance, and angle, we chose second, meter and degree, respectively.

In this experiment, two observation stations located at points $S_1 = [s_{12}, s_{12}]$ and $S_2 = [s_{21}, s_{22}]$ performed the measurements. Each station measured the distance to the vehicle and the direction to the vehicle, as described by the following equation:

$$\mathbf{y}_k = \begin{pmatrix} d_1 \\ d_2 \\ \theta_1 \\ \theta_2 \end{pmatrix} = \begin{pmatrix} \sqrt{[x - s_{11}]^2 + [y - s_{12}]^2} \\ \sqrt{[x - s_{21}]^2 + [y - s_{22}]^2} \\ \arctan[(y - s_{12}) / (x - s_{11})] \\ \arctan[(y - s_{22}) / (x - s_{21})] \end{pmatrix} + \mathbf{v}_k + \mathbf{b}_k. \quad (78)$$

Here:

- $\mathbf{v}_k$ represents random uncertainty; it is Gaussian with covariance matrix C_k^z ,
- $\mathbf{b}_k$ represents the unknown but bounded uncertainty; it is bounded by an ellipsoid with shape matrix S_k^z .

The initial state estimate, and initial estimates for the covariance matrix and for the shape matrix are:

- $\mathbf{x}_0 = (0, 0, 0, 0)$,
- $C_0^+ = \text{diag}(0.01, 0.01, 0.01, 0.01)$,
- $S_0^+ = \text{diag}(10^{-6}, 10^{-6}, 10^{-6}, 10^{-6})$.

The initial covariance matrices in process equation and measurement equation are:

$$C_0^u = \begin{pmatrix} 0.0033 & 0 & 0.005 & 0 \\ 0 & 0.0033 & 0 & 0.005 \\ 0.005 & 0 & 0.01 & 0 \\ 0 & 0.005 & 0 & 0.01 \end{pmatrix}, C_0^z = \begin{pmatrix} 0.05^2 & 0 & 0 & 0 \\ 0 & 0.05^2 & 0 & 0 \\ 0 & 0 & 0.01^2 & 0 \\ 0 & 0 & 0 & 0.01^2 \end{pmatrix}. \quad (79)$$

The initial shape matrices of UBB uncertainties in process equation and measurement equation are:

$$S_0^u = \begin{pmatrix} 1^2 & 0 & 0 & 0 \\ 0 & 1^2 & 0 & 0 \\ 0 & 0 & 0.5^2 & 0 \\ 0 & 0 & 0 & 0.5^2 \end{pmatrix}, S_0^z = \begin{pmatrix} 0.01^2 & 0 & 0 & 0 \\ 0 & 0.01^2 & 0 & 0 \\ 0 & 0 & (\frac{\pi}{180})^2 & 0 \\ 0 & 0 & 0 & (\frac{\pi}{180})^2 \end{pmatrix}. \quad (80)$$

As in the first example, we select the value $\eta = 0.5$ of the weighting parameter.

Figure 3 shows an example of trajectory estimates by using EGKF and EKF obtained on one of the 100 iterations—namely, on the 5th iteration. To see it clearer, Figure 4 shows a more detailed information for several portions of the trajectory.

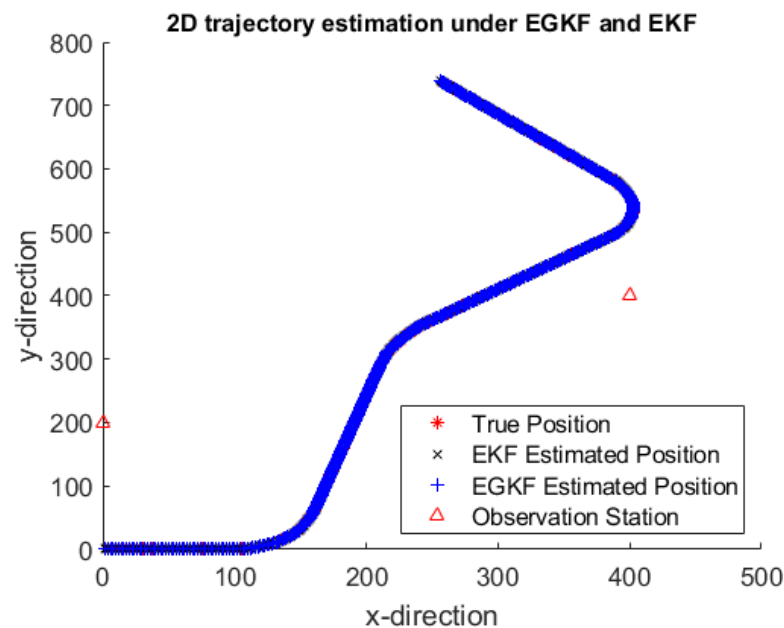


Figure 3. 2D trajectory estimated results of the 5th simulation.

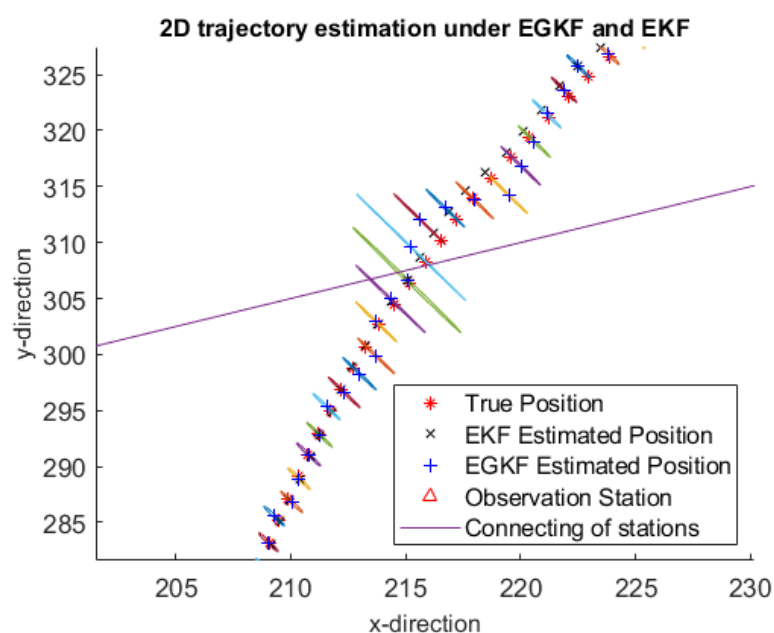


Figure 4. 2D trajectory estimated results of the 5th simulation (zoom in).

Here:

- the red stars mark the true positions,
- black crosses mark the EKF estimates,
- the blue plus signs are the centers of the ellipsoids computed by EGKF.

The two observation stations are marked by red triangles; they are connected by a straight line.

Figure 5 shows the comparison of RMSE estimation errors EGKF and EKF averaged over all 1000 Monte-Carlo runs.

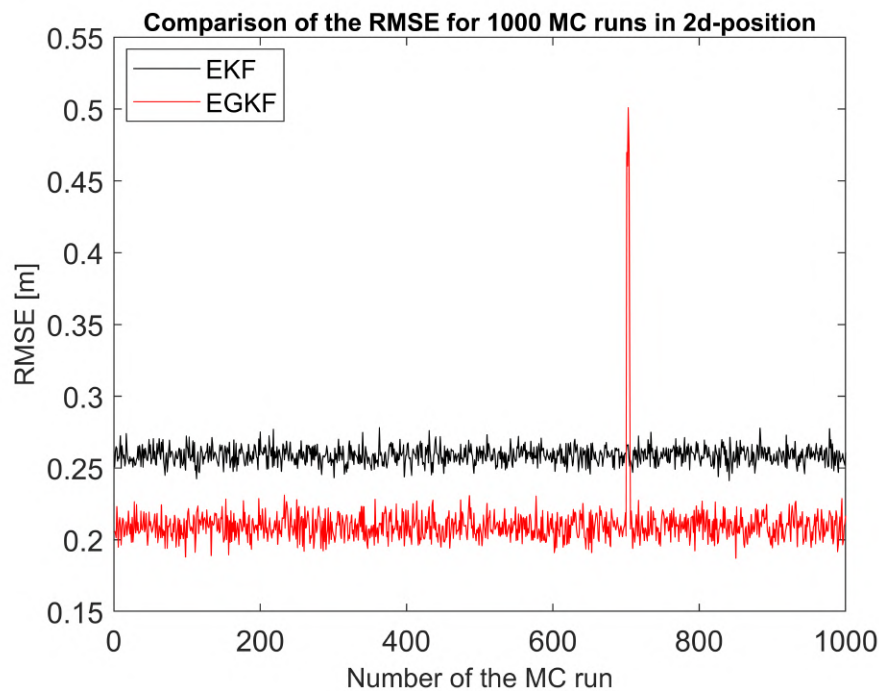


Figure 5. Root-mean-square error (RMSE) comparison of EGKF and EKF in 1000 runs.

The peak occurred in the figure between [600, 800] is probably caused by the random uncertainty, just like what EKF behaves in nonlinear applications sometimes. In this experiment we introduced new UBB uncertainty but we still kept the influence of random uncertainty ($\eta = 0.5$). The new EGKF can provide a better global estimation than EKF but a few exceptions may still happen. This unexpected jump also caused a larger standard deviation value in EGKF.

In almost all simulations (namely, in 997 cases out of 1000), the RMSE of EGKF was significantly smaller than for the EKF. We can therefore conclude that the new EGKF techniques provides better estimation for this nonlinear system than EKF. Detailed statistics of the comparison can be found in Table 2.

Table 2. RMSE Comparison of EGKF and EKF.

RMSE [m]	EGKF	EKF
Min	0.187	0.241
Max	0.501	0.278
Mean	0.210	0.259
Median	0.209	0.259
Standard Deviation	0.019	0.006

In Figure 6, we compared the average traces (over all 1000 Monte-Carlo runs) of EGKF UBB shape matrix S_k^+ , EGKF covariance matrix C_k^+ and EKF covariance matrix, in every epoch.

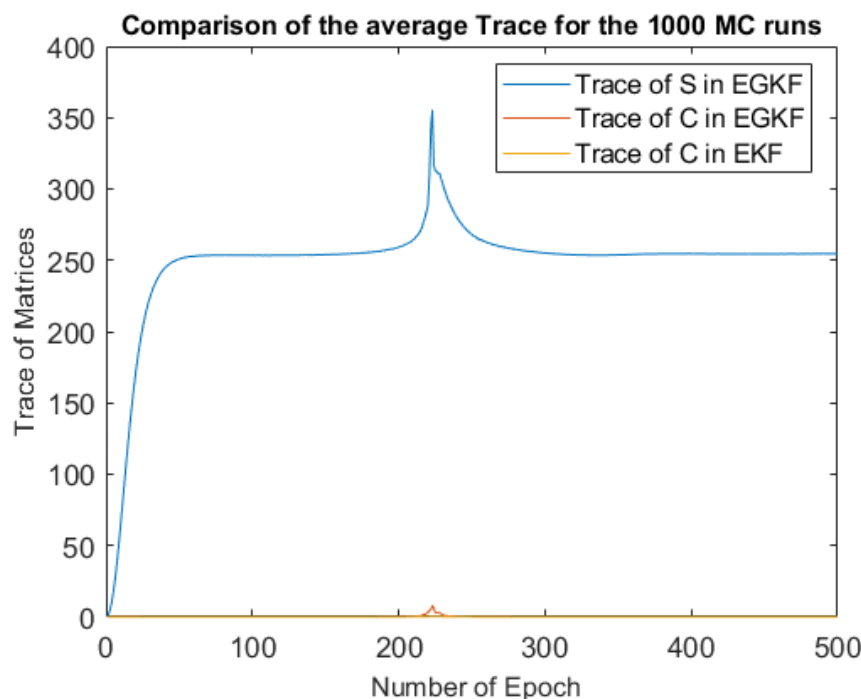


Figure 6. Average trace comparison for 1000 MC runs of EGKF and EKF in every epoch.

Notice that the trace of EGKF shape matrix is much larger than the traces of covariance matrices.

We can also see that the UBB uncertainty grows in the starting part of the trajectory and when the trajectory crosses the straight line connecting the observation stations—this corresponds to $k \in [210, 220]$. This phenomenon is easy to explain: in this case, both angle measurements measure, in effect, the same quantity, so from the measurements, we get fewer information than in other parts of the trajectory.

The EGKF algorithm has been applied to a data set which is obtained from a real world experiment see, e.g., [30]. In that experiment, taken from the scope of georeferencing of terrestrial laser scanner, a multi-sensor system has collected the trajectory information of two GNSS antennas. EGKF was applied to estimate the positions and velocities of those two antennas and the results were compared with EKF.

5. Conclusions and Future Work

In this paper, we propose a new method for state estimation in situations when, in addition to the probabilistic uncertainty, we also have unknown-but-bounded errors. Our testing shows that, on average, the new method leads to much more accurate estimates than the standard extended Kalman filter.

While our method has been shown to be successful, there is room for improvement.

First, the efficiency of our method depends on the initial selection of the parameters describing the system uncertainty and the measurement uncertainties. In some practical situations, we have prior information about these uncertainties, but in other practical cases, we do not have this information. Similarly to the usual Kalman filter, our method eventually converges to the correct uncertainty estimate, but this convergence may be somewhat slow. It would be nice to come up with recommendations on how to select the initial uncertainty parameters that would guarantee faster convergence.

Second, in our method, we assumed that measurements are reasonably frequent, so that during the time interval between the two measurements, we can safely ignore terms quadratic in terms of the corresponding changes—and thus consider a linearized system of equations. This is true in many practical situations—e.g., in the example of GNSS-based navigation that we considered in this paper. However, in some practical situations, measurements are rarer, and so, in the interval between the two measurements, we can no longer ignore terms which are quadratic in terms of changes. To cover such situations, it is desirable to extend our technique to second-order Kalman filtering.

Third, in this paper, we followed the usual Kalman filter techniques in assuming that the corresponding probability distributions are Gaussian. As we mentioned in Section 1, in practice, sometimes distributions are not Gaussian. It is therefore desirable to extend our method to the general non-Gaussian case – e.g., by using ideas of unscented Kalman filtering.

Fourth, in our method, we minimized the mean square estimation error—which, for our method, corresponds to selecting an ellipsoid with the smallest possible trace of the corresponding shape matrix. While in many practical problems, minimizing the mean square error is a reasonable idea, in some problems, it may be more reasonable to minimize, e.g., the largest possible estimation error. In this case, as we have mentioned, instead of minimizing the trace, we should be minimizing the largest eigenvalue of the shape matrix. It is this desirable to extend our method to this—and other—possible criteria.

Fifth, to make computations feasible, we approximated the set of possible states by an ellipsoid. In principle, we can get more accurate estimates if we use families of sets with more parameters for such approximation—e.g., zonotopes or, more generally, generic polytopes. From this viewpoint, it is desirable to come up with efficient algorithms for processing zonotope (and, more generally, polytope) uncertainty.

Finally, it is desirable to analyze (and, if necessary, improve) the stability of our method (and of other state estimation techniques). While stability is very important for practical problems, it is not even clear how to formulate this problem in precise mathematical terms, since the state estimation problems are usually ill-posed—as most inverse problems (see, e.g., [31])—and thus, strictly speaking, not stable.

Author Contributions: Conceptualization, I.N.; Data curation, L.S. and H.A.; Formal analysis, L.S. and V.K.; Funding acquisition, I.N.; Investigation, L.S. and H.A.; Methodology, L.S., B.K. and V.K.; Project administration, I.N.; Resources, H.A. and I.N.; Software, B.K.; Supervision, V.K. and I.N.; Validation, L.S.; Visualization, L.S.; Writing—original draft, L.S.; Writing—review & editing, H.A., B.K., V.K. and I.N.

Funding: This work was funded by the German Research Foundation (DFG) as a part of the Research Training Group i.c.sens (GRK2159).

Acknowledgments: The authors acknowledge the discussions with Steffen Schön, Franz Rottensteiner and Claus Brenner from Leibniz University Hannover. The authors also thank Sergey Grigorian from Department of Mathematics in University of Texas Rio Grande Valley for his comments and suggestions.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Le, V.T.H.; Stoica, C.; Alamo, T.; Camacho, E.F.; Dumur, D. Zonotopic guaranteed state estimation for uncertain systems. *Automatica* **2013**, *49*, 3418–3424. [[CrossRef](#)]
2. Kalman, R.E. A new approach to linear filtering and prediction problems. *J. Basic Eng.* **1960**, *82*, 35–45. [[CrossRef](#)]
3. Smith, G.L.; Schmidt, S.F.; McGee, L.A. *Application of Statistical Filter Theory to the Optimal Estimation of Position and Velocity on Board a Circumlunar Vehicle*; NASA: Washington, DC, USA, 1962.
4. McElhoe, B.A. An assessment of the navigation and course corrections for a manned flyby of mars or venus. *IEEE Trans. Aerosp. Electron. Syst.* **1966**, *4*, 613–623. [[CrossRef](#)]
5. Sorenson, H.W.; Alspach, D.L. Recursive Bayesian estimation using Gaussian sums. *Automatica* **1971**, *7*, 465–479. [[CrossRef](#)]
6. Simon, D. *Optimal State Estimation: Kalman, H Infinity, and Nonlinear Approaches*; JohnWiley & Sons, Inc.: Hoboken, NJ, USA, 2006.

7. Julier, S.; Uhlmann, A. New extension of the Kalman filter to nonlinear systems. In Proceedings of the AeroSense, Orlando, FL, USA, 28 July 1997.
8. Gordon, N.; Salmond, D.; Smith, A.F.M. Novel approach to nonlinear/non-gaussian Bayesian state estimation. *IEE Proc. F Radar Signal Process.* **1993**, *140*, 107–113. [[CrossRef](#)]
9. Alkhatib, H.; Paffenhof, J.-A.; Kutterer, H. Sequential Monte Carlo filtering for nonlinear GNSS trajectories. In Proceedings of the Symposium on VII Hotine-Marussi Symposium on Mathematical Geodesy, Rome, Italy, 6–10 June 2009; Sneeuw, N., Novák, P., Crespi, M., Sansò, F., Eds.; Springer: Berlin/Heidelberg, Germany, 2012; pp. 81–86.
10. Alkhatib, H. Alternative Nonlinear Filtering Techniques in Geodesy for Dual State and Adaptive Parameter Estimation. In *the 1st International Workshop on the Quality of Geodetic Observation and Monitoring Systems (QuGOMS'11)*; Kutterer, H., Seitz, F., Alkhatib, H., Schmidt, M., Eds.; Springer: Cham, Switzerland, 2015; Volume 140.
11. Milanese, M.; Vicino, A. Optimal estimation theory for dynamic systems with set membership uncertainty: An overview. *Automatica* **1996**, *27*, 997–1009. [[CrossRef](#)]
12. Witsenhausen, H.S. Sets of possible states of linear systems given perturbed observations. *IEEE Trans. Autom. Control* **1968**, *13*, 556–558. [[CrossRef](#)]
13. Schweppe, F. Recursive state estimation: Unknown but bounded errors and system inputs. *IEEE Trans. Autom. Control* **1968**, *13*, 22–28. [[CrossRef](#)]
14. Schweppe, F.C. *Uncertain Dynamic Systems*; Prentice Hall: Englewood Cliffs, NJ, USA, 1973.
15. Kreinovich, V.; Lakeyev, A.; Rohn, J.; Kahl, P. *Computational Complexity and Feasibility of Data Processing and Interval Computations*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2013; Volume 10.
16. Ferson, S.; Kreinovich, V.; Hajagos, J.; Oberkampf, W.; Ginzburg, L. *Experimental Uncertainty Estimation and Statistics for Data Having Interval Uncertainty*; Report SAND2007-0939,162; Sandia National Laboratories: Albuquerque, NM, USA, 2007.
17. Kutterer, H.; Neumann, I. Recursive least-squares estimation in case of interval observation data. *Int. J. Reliab. Saf.* **2011**, *5*, 229–249. [[CrossRef](#)]
18. Vicino, A.A.; Zappa, G. Sequential approximation of feasible parameter sets for identification with set membership uncertainty. *IEEE Trans. Autom. Control* **1996**, *41*, 774–785. [[CrossRef](#)]
19. Walter, E.; Piet-Lahanier, H. Exact recursive polyhedral description of the feasible parameter set for bounded-error models. *IEEE Trans. Autom. Control* **1989**, *34*, 911–915. [[CrossRef](#)]
20. Combastel, C. A state bounding observer for uncertain non-linear continuous-time systems based on zonotopes. In Proceedings of the 44th IEEE Conference on Decision and Control, Seville, Spain, 15 December 2005; pp. 7228–7234.
21. Alamo, T.; Bravo, J.M.; Camacho, E.F. Guaranteed state estimation by zonotopes. *Automatica* **2005**, *41*, 1035–1043. [[CrossRef](#)]
22. Althoff, M.; Stursberg, O.; Buss, M. Safety assessment for stochastic linear systems using enclosing hulls of probability density functions. In Proceedings of the 2009 European Control Conference, Budapest, Hungary, 23–26 August 2009; pp. 625–630.
23. Schön, S.; Kutterer, H. Using zonotopes for overestimation-free interval least-squares—some geodetic applications. *Reliab. Comput.* **2005**, *11*, 137–155. [[CrossRef](#)]
24. Bertsekas, D.P.; Rhodes, I.B. On the minimax reachability of target sets and target tubes. *Automatica* **1971**, *7*, 233–247. [[CrossRef](#)]
25. Polyak, B.T.; Nazin, S.A.; Durieu, C.; Walter, E. Ellipsoidal parameter or state estimation under model uncertainty. *Automatica* **2004**, *40*, 1171–1179. [[CrossRef](#)]
26. Durieu, C.; Walter, E.; Polyak, B. Multi-input multi-output ellipsoidal state bounding. *J. Optim. Theory Appl.* **2001**, *111*, 273–303. [[CrossRef](#)]
27. Noack, B. *State Estimation for Distributed Systems with Stochastic and Set-Membership Uncertainties*; KIT Scientific Publishing: Karlsruhe, Germany, 2014; Volume 14.
28. Rump, S.M. INTLAB—INTERVAL LABORATORY. In *Developments in Reliable Computing*; Csendes, T., Ed.; Kluwer Academic Publishers: Amsterdam, The Netherlands, 1999; pp. 77–104.
29. Kitagawa, G. Non-gaussian state—Space modeling of nonstationary time series. *J. Am. Stat. Assoc.* **1987**, *82*, 1032–1041.

30. Sun, L.; Alkhatib, H.; Paffenholz, J.A.; Neumann, I. Geo-Referencing of a Multi-Sensor System Based on Set-membership Kalman Filter. In Proceedings of the 21st International Conference on Information Fusion (FUSION), Cambridge, UK, 10–13 July 2018; pp. 1–5.
31. Luise, B. State estimation analysed as inverse problem. In *Assessment and Future Directions of Nonlinear Model Predictive Control*; Springer: New York, NY, USA, 2007; pp. 335–346.



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

Geo-Referencing of a Multi-Sensor System Based on Set-membership Kalman Filter

1st Ligang Sun
Geodetic Institute
Leibniz Universität Hannover
Hannover, Germany
ligang.sun@gih.uni-hannover.de

2nd Hamza Alkhatib
Geodetic Institute
Leibniz Universität Hannover
Hannover, Germany
alkhatib@gih.uni-hannover.de

3rd Jens-André Paffenholtz
Geodetic Institute
Leibniz Universität Hannover
Hannover, Germany
paffenholtz@gih.uni-hannover.de

4th Ingo Neumann
Geodetic Institute
Leibniz Universität Hannover
Hannover, Germany
neumann@gih.uni-hannover.de

Abstract—In this paper, a novel set-membership Kalman filter is applied on a data set which is obtained from a real world experiment. In this experiment, taken from the scope of geo-referencing of terrestrial laser scanner, a multi-sensor system has captured the trajectory of two GNSS antennas. The dynamical system contains the random uncertainty and set-membership uncertainty simultaneously. Both estimated results from classic extended Kalman filter and novel set-membership Kalman filter are shown and compared. Detailed analysis of the set-membership Kalman filter is given in the end.

Index Terms—set-membership Kalman filter, state estimation, multi-sensor system.

I. INTRODUCTION

The Kalman filter (KF), as developed in [1], has been widely used in several disciplines such as engineering navigation and deformation analysis. Most of the applications require a sequential estimation of the system states based on information coming from a process equation itself and external observations. Over the years refined models were introduced for the system description (e.g., the kinematic model of a vehicle), and a great variety of sensors was used to observe the system state. For example [2] and [3] described an adaptive extended Kalman filter (EKF) approach for geo-referencing tasks to a multi-sensor system (MSS), which is a fusion of a phase-measuring terrestrial laser scanner (TLS) with Global Navigation Satellite System (GNSS) equipment and inclinometers.

According to [1] the estimated result is optimal only if the system (which consists of transition and measurement equations) is linear and the error process is Gaussian. Unfortunately, the modelling of reality diverges from these assumptions, and nonlinear, non-Gaussian, non-stationary state estimation should be taken into account. Thus over the years a multitude of approximate nonlinear filters has been proposed, see e.g. [4] and [5].

A well known analytical approximation to handle a nonlinear system with probabilistic assumptions about the system

uncertainties is to linearize the measurement and the system equations using Taylor series expansions, see [6]. However, as pointed out in [4] this type of nonlinear filter which includes the first-order and the higher-order EKF, is prone to diverge if the system equations are highly nonlinear.

All above mentioned approaches are based on the assumption that system and measurement noises are of probabilistic nature and obey a specific probability distribution. The stochastic formulations of the uncertainties provide an appropriate solution for the state estimation especially when the errors are unbounded and some outliers are existing. However, in some cases this assumption may be vague (see, e.g. [7]) and may lead to biased estimates of the state vector and their corresponding uncertainties. In these cases, set-membership uncertainty state estimation has the advantage that no specific error behavior needs to be assumed, expect for defining the bounds of the uncertainties. In [9] a new filter model called set-membership Kalman filter (SKF) was developed. On one hand, the SKF takes both random and set-membership uncertainties into account in the system, on the other hand, this filter can deal with state estimation problems for nonlinear systems. The unknown but bounded (UBB) uncertainties are based on ellipsoidal bounding [10]. This filter can be regarded as an extension of the filter developed in [8], where the linear case was investigated.

The main work of this paper is to apply the new SKF on real data sets from a MSS, for the first time. However, due to the missing of group truth, it is not possible to compare the estimated results with the true states, even though the comparison of EKF and SKF is discussed. This paper is organized as follows. Section 2 briefly describes the SKF model and summarizes the main steps of the developed filter. The used MSS as well as the system and measurement model are presented in Section 3. The numerical results are demonstrated in Section 4. Finally, Section 5 summarizes the results and gives an outlook for future work.

II. SET-MEMBERSHIP KALMAN FILTER MODEL AND ALGORITHM

Consider the following nonlinear dynamical system:

$$x_{k+1} = f_k(x_k, u_k, w_k, a_{1,k}, a_{2,k}, \dots, a_{I,k}) \quad (1a)$$

$$y_k = h_k(x_k, v_k, b_k) \quad (1b)$$

where x_k is a n -dimensional state vector, u_k is the known input vector, $w_k \sim N(0, C_k^u)$ is a Gaussian system noise with covariance matrix C_k^u , $a_{i,k} \in \mathcal{E}(0, S_{i,k}^u)$ is the unknown but bounded perturbation with shape matrix $S_{i,k}^u$, $i = 1, 2, \dots, I$, denotes the i th set-membership perturbation in the prediction equation. $v_k \sim N(0, C_k^z)$ is a Gaussian measurement noise with covariance matrix C_k^z , and $b_k \in \mathcal{E}(0, S_k^z)$ is the unknown but bounded perturbation with shape matrix S_k^z .

Based on Taylor expansion we get the following linearized system of (1):

$$x_{k+1} = F_{x,k}x_k + \tilde{u}_k + F_{w,k}w_k + A_k \quad (2a)$$

$$y_k = H_{x,k}x_k + \tilde{z}_k + H_{v,k}v_k + H_{b,k}b_k \quad (2b)$$

where $A_k = \sum_{i=1}^I F_{a_i,k}a_{i,k}$, $F_{x,k}x_k$, $F_{w,k}$ and $F_{a_i,k}$ are the first order derivatives of nonlinear function f with respect to the variables x_k , w_k and $a_{i,k}$. $H_{x,k}$, $H_{v,k}$ and $H_{b,k}$ are the first order derivatives of the nonlinear function h with respect to the variables x_k , v_k and b_k . $\tilde{u}_k = f_k(\hat{x}_k^+, u_k, 0, 0) - F_{x,k}\hat{x}_k^+$. $\tilde{z}_k = h_k(\hat{x}_k^-, 0, 0) - H_{x,k}\hat{x}_k^-$. Notice that $\tilde{z}_k = 0$ if the measurement equation is linear.

Our objective is to calculate the explicit expressions of $\hat{x}_k^{c-}$, C_k^- , S_k^- and $\hat{x}_k^{c+}$, C_k^+ , S_k^+ . And the new SKF algorithm is given in the following. Detailed derivation process can be found in [9].

Algorithm 1 Set-membership Kalman filter model

1: **Initialization:**

- 1) Initial state midpoint $\hat{x}_0^+ = x_0$.
- 2) Initial estimated random covariance matrix C_0^+ .
- 3) Initial estimated set-membership shape matrix S_0^+ .

2: **for** $k=1, 2, \dots, K$ **do**

3: **Input of Prediction Step:**

- 1) Point post-estimation $\hat{x}_k^+$, with estimated covariance C_k^+ and shape matrix S_k^+ .
- 2) Nonlinear system model

$$x_{k+1} = f_k(x_k, u_k, w_k, a_{1,k}, a_{2,k}, \dots, a_{I,k}), \quad (3)$$

where $w_k \sim N(0, C_k^u)$ and $a_{i,k} \in \mathcal{E}(0, S_{i,k}^u)$, $i = 1, 2, \dots, I$.

- 3) Control input u_k , random noise covariance C_k^u and shape matrices $S_{i,k}^u$, $i = 1, 2, \dots, I$, for set-membership uncertainty.

4: **Calculation of Prediction Step:**

- 1) Calculate the error covariance matrix C_{k+1}^- according to

$$C_{k+1}^- = F_{x,k}C_k^+F_{x,k}^T + F_{w,k}C_k^uF_{w,k}^T. \quad (4)$$

- 2) Calculate the center of the priori ellipsoid:

$$\hat{x}_{k+1}^{c-} = F_{x,k}\hat{x}_k^{c+} + \tilde{u}_k. \quad (5)$$

- 3) Calculate the shape matrix of the priori ellipsoid:

$$\begin{aligned} S_{k+1}^- = & (\sqrt{\text{tr}(F_{x,k}S_k^+F_{x,k}^T)} \\ & + \sum_{i=1}^I \sqrt{\text{tr}(F_{a_i,k}S_{i,k}^uF_{a_i,k}^T)}) \\ & \cdot \left(\frac{F_{x,k}S_k^+F_{x,k}^T}{\sqrt{\text{tr}(F_{x,k}S_k^+F_{x,k}^T)}} \right. \\ & \left. + \sum_{i=1}^I \frac{F_{a_i,k}S_{i,k}^uF_{a_i,k}^T}{\sqrt{\text{tr}(F_{a_i,k}S_{i,k}^uF_{a_i,k}^T)}} \right). \end{aligned} \quad (6)$$

The predicted point estimate $\hat{x}_{k+1}^-$ is characterized by the random error C_{k+1}^- and the set-membership error by $\hat{x}_{k+1}^{c-}$ and S_{k+1}^- . $\text{tr}(\cdot)$ denotes the trace of a matrix.

5: **Output of Prediction Step:**

Priori estimated state: $\hat{x}_k^{c-}$, C_k^- , and S_k^- .

6: **Input of Filtering Step:**

- 1) Priori or predicted estimate $\hat{x}_k^-$ with error covariance matrix C_k^- and ellipsoid center $\hat{x}_k^{c-}$ and shape matrix S_k^- .
- 2) Nonlinear measurement model:

$$y_k = h_k(x_k, v_k, b_k), \quad (7)$$

where $v_k \sim N(0, C_k^z)$ and $b_k \in \mathcal{E}(0, S_k^z)$.

- 3) Observation y_k , sensor noise with random covariance C_k^z and set-membership shape matrix S_k^z .
- 4) $\tilde{z}_k(\hat{x}_k^-) = h_k(\hat{x}_k^-, 0, 0) - H_{x,k}\hat{x}_k^-$.
- 5) Weighting parameter η . $0 \leq \eta \leq 1$ is a weighting parameter of random uncertainty and set-membership uncertainty, which is usually based on the information given by experts.

7: **Calculation of Filtering Step:**

- 1) For a given weighting parameter η , the optimal gain factor K_k is

$$\begin{aligned} K_k(\beta) = & [(1-\eta)C_k^-H_{x,k}^T + \eta(1+\frac{1}{\beta})S_k^-H_{x,k}^T] \\ & \cdot [(1-\eta)H_{x,k}C_k^-H_{x,k}^T \\ & + (1-\eta)H_{v,k}C_k^zH_{v,k}^T \\ & + \eta(1+\frac{1}{\beta})H_{x,k}S_k^-H_{x,k}^T \\ & + \eta(1+\beta)H_{b,k}S_k^zH_{b,k}^T]^{-1}. \end{aligned} \quad (8)$$

- 2) Calculate the center of updated estimate $\hat{x}_k^+$ by means of

$$\hat{x}_k^{c+} = (I - K_kH_{x,k})\hat{x}_k^{c-} + K_k[y - \tilde{z}_k(\hat{x}_k^{c-})]. \quad (9)$$

- 3) Calculate the updated error covariance matrix C_k^+ by

$$C_k^+(\beta) = (I - K_k H_{x,k}) C_k^- (I - K_k H_{x,k})^T + K_k H_{v,k} C_k^z H_{v,k}^T K_k^T. \quad (10)$$

- 4) Update the shape matrix S_k^+ by

$$S_k^+(\beta) = (1 + \frac{1}{\beta})(I - K_k H_{x,k}) S_k^- (I - K_k H_{x,k})^T + (1 + \beta) K_k H_{b,k} S_k^z H_{b,k}^T K_k^T. \quad (11)$$

- 5) The optimal parameter β^* can be solved by

$$\beta^* = \arg \min \{ (1 - \eta) \text{tr}[C_k^+(\beta)] + \eta \text{tr}[S_k^+(\beta)] \}. \quad (12)$$

The updated point estimate $\hat{x}_k^+$ is characterized by random error characteristic $C_k^+ = C_k^+(\beta)$ and set-membership error description $S_k^+ = S_k^+(\beta)$. Put β^* into above 4 functions to get the optimal output.

8: **Output of Filtering Step:**

Posteriori estimated state: $\hat{x}_k^{c+}$, $C_k^+(\beta^*)$, and $S_k^+(\beta^*)$.

9: **end for**

The crucial part of the algorithm is to solve an optimization problem in each step. The optimization problem can be rewritten as the following standard form [9]:

$$\begin{aligned} \min_{\beta} \quad & J(\beta) = (1 - \eta) \text{tr}(C_k^+) + \eta \text{tr}(S_k^+(\beta)). \\ \text{s.t.} \quad & \beta \in [0, +\infty) \subset \mathbb{R}^1. \end{aligned} \quad (13)$$

Problem (13) is a nonlinear programming problem since the objective function J is nonlinear. The solution of above optimization problem was denoted by β^* . Furthermore, problem (13) is a convex optimization problem [8]. Therefore, any existing local minimum is a global minimum. Usually, it is hard to solve a nonlinear programming problem due to the constrained equations or inequalities. MATLAB function *fminsearch* is an efficient way to solve the problem (13). Further, an advanced toolbox *INTLAB* [12] can also be used.

Both random and set-membership uncertainties were considered in here. The novel SKF takes UBB uncertainties into account in both process equation and measurement equation, therefore it has a better uncertainty measures. It also keeps the recursive framework of random uncertainties from KF, thus the advantages of KF are reserved during the propagation process. A better estimation under these more reliable assumptions is calculated based on solving an optimization problem in each step. Different with standard KF, where the output is usually an gaussian distribution and the mean of the distribution was regarded as the estimated point, in SKF, a set containing all the mean values of possible distributions was put out. Any point in the ellipsoid $\mathcal{E}(\hat{x}_{c+}, S_k^+)$ can be regarded as the optimal estimation of state x_k .

III. MULTI-SENSOR SYSTEM AND ITS KINEMATIC MODEL

In this section, the MSS which was developed in [2] and used in the real world experiment would be introduced. The aim of the MSS is to directly provide the transformation parameters and their uncertainty information for the purpose of direct geo-referencing of 3D point clouds and validate the new SKF. GNSS equipments and an optional inclinometer were installed on the top of a terrestrial laser scanner, see Figure 1. It should be noticed that the inclinometer is not used here.

A. Conceptual design and realization of a multi-sensor system

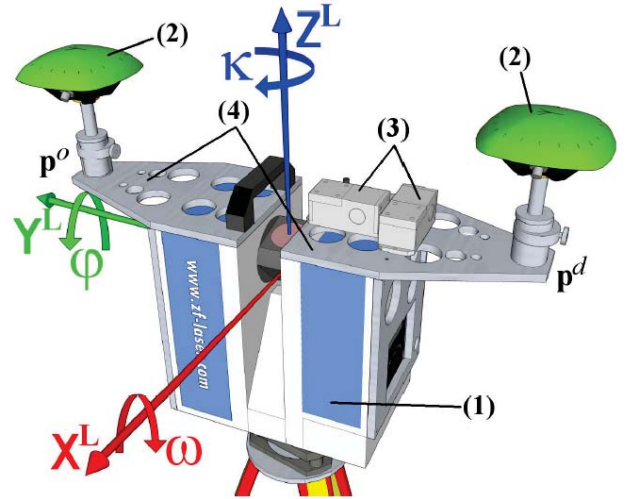


Fig. 1. Multi-sensor system: (1) Terrestrial laser scanner Z+F IMAGER 5006, (2) Javed GrAnt G3T GNSS antenna, (3) Schaevitz LSOC-1° inclinometer (not used in the real world experiments), (4) Wing support structure. [2]

The terrestrial laser scanner was used to generate 3D point clouds (composed of several 2D vertical profiles). Here the phase-measuring TLS Zoller+Fröhlich (Z+F) IMAGER 5006 was chosen. The requirement of exploitation the 360° rotation as a function of time and orientation reference can be fulfilled by this panoramic scanner type.

The GNSS antennas were installed to measure the 3D positions, which are represented by the antenna reference points and describe the circular motion of the MSS. The fulfillment of the uncertainty requirement is a challenging task due to the short period (approx. 15 min) of kinematic GNSS data acquisition for a 360° rotation of the laser scanner. The typically used data rate is 10-20 Hz. Two identical Leica LEIAX1202GG antennas were connected to two similar Javad TRE G3T Delta receivers, and all the GNSS antennas were individually and absolutely calibrated. For further details of the used data set obtained in the real world experiment the reader is referred to [11].

B. Process equation

Based on the process equation of [2, p. 89, Eq. (5.8)], the modified nonlinear process equation after introducing set-membership uncertainty of this MSS is

$$x_{k+1} = f_k(x_k, u_k, w_k, a_k) = \begin{pmatrix} X'_k + s_k \Delta t \sin(2\pi u_k t_k) \\ Y'_k - s_k \Delta t \cos(2\pi u_k t_k) \\ Z'_k \\ X''_k - s_k \Delta t \sin(2\pi u_k t_k) \\ Y''_k + s_k \Delta t \cos(2\pi u_k t_k) \\ Z''_k \\ s_k \end{pmatrix} + w_k + a_k. \quad (14)$$

According to [2], the state vector is

$$x_k = (X'_k, Y'_k, Z'_k, X''_k, Y''_k, Z''_k, s_k)^T, \quad (15)$$

where (X'_k, Y'_k, Z'_k) and (X''_k, Y''_k, Z''_k) are the coordinates of the two antennas respectively, s_k is the magnitude of velocity of the MSS. The unit of position component is meter [m], and the unit of time is second [s]. The significant frequency of the sine oscillation is introduced as known input vector u_k . In this experiment, $u_k = 0.001265$ for the 360° rotation. This value is empirically obtained by the analyses of several independent data acquisitions during the MSS's calibration by [2]. $w_k \sim N(0, C_k^u)$ is Gaussian system noise with covariance matrix C_k^u , $a_k \in \mathcal{E}(0, S_k^u)$ is an unknown but bounded perturbation with shape matrix S_k^u in the process equation.

After Taylor expansion for the nonlinear process equation (14), we get the following linearized process equation:

$$x_{k+1} = T_k x_k + B_k u_k + w_k + a_k, \quad (16)$$

where

$$T_k = \frac{\partial f_k}{\partial x_k} = \begin{pmatrix} I_{6 \times 6} & T_{6 \times 1} \\ 0_{1 \times 6} & I_{1 \times 1} \end{pmatrix}, \quad (17)$$

$$T_{6 \times 1} = [\Delta t \cdot \sin(2\pi u_k t), -\Delta t \cdot \cos(2\pi u_k t), 0, -\Delta t \cdot \sin(2\pi u_k t), \Delta t \cdot \cos(2\pi u_k t), 0], \quad (18)$$

and

$$B_k = \frac{\partial f_k}{\partial u_k} = \begin{pmatrix} 2\pi t s_k \Delta t \cos(2\pi u_k t) \\ 2\pi t s_k \Delta t \sin(2\pi u_k t) \\ 0 \\ -2\pi t s_k \Delta t \cos(2\pi u_k t) \\ -2\pi t s_k \Delta t \sin(2\pi u_k t) \\ 0 \\ 0 \end{pmatrix}. \quad (19)$$

C. Measurement equation

The observation vector is defined by the used sensors in this MSS. In this experiment, 3D GNSS positions in a local topo-centric NEU system or in the ERTS89 combined with the UTM projection with 10 Hz data rate are assumed ($\Delta t = 0.1s$). The observation vector is composed of the 3D positions $y_k = (X'_k, Y'_k, Z'_k, X''_k, Y''_k, Z''_k)$. Based on measurement equation of [2, p. 90, Eq. (5.14)], the modified measurement equation is:

$$y_k = h_k(x_k, v_k, b_k) = H_k x_k + v_k + b_k, \quad (20)$$

where $v_k \sim N(0, C_k^z)$ is the Gaussian measurement noise with covariance matrix C_k^z , and $b_k \in \mathcal{E}(0, S_k^z)$ is an UBB perturbation with shape matrix S_k^z .

IV. EXPERIMENTAL RESULTS

In this section, three estimated trajectories of the two antennas on the MSS are given, based on EKF and SKF respectively.

The position part of the initial state is set directly by the initial observation:

$$x_0 = (-2502.7285, -6038.9481, -26.4036, -2502.4626, -6039.8838, -26.4139, 0.0023)^T. \quad (21)$$

A. EKF ($\eta = 0$)

According to the solution in [9], EKF can be regarded as a special case of SKF when $\eta = 0$.

The initial covariance matrix of the random uncertainty in process equation is set as a diagonal matrix:

$$C_0^u = 10^{-6} \cdot \text{diag}(1, 1, 1, 1, 1, 1, 4). \quad (22)$$

We assume that the observations, i.e., 3d GNSS positions of each antenna, are independent and therefore the initial covariance matrix of the measurement equation is in the form:

$$C_0^z = \begin{pmatrix} Q_{3 \times 3}^{(1)} & 0_{3 \times 3} \\ 0_{3 \times 3} & Q_{3 \times 3}^{(2)} \end{pmatrix}, \quad (23)$$

where $Q_{3 \times 3}^{(1)}$ and $Q_{3 \times 3}^{(2)}$ are two covariance matrices from the measurement instrument, and they are symmetric.

The covariance matrix of initial state x_0^{c+} follows by:

$$C_0^{c+} = \begin{pmatrix} Q_{3 \times 3}^{(1)} & & \\ & Q_{3 \times 3}^{(2)} & \\ & & 0.005 \end{pmatrix}. \quad (24)$$

The estimated states under classic EKF are given in Figure. 2. The following 7 figures show the estimated positions of the two antennas and the velocity of the MSS in one period. In the first 6 figures, the direct observed states, predicted states and filtered states are compared in every epoch.

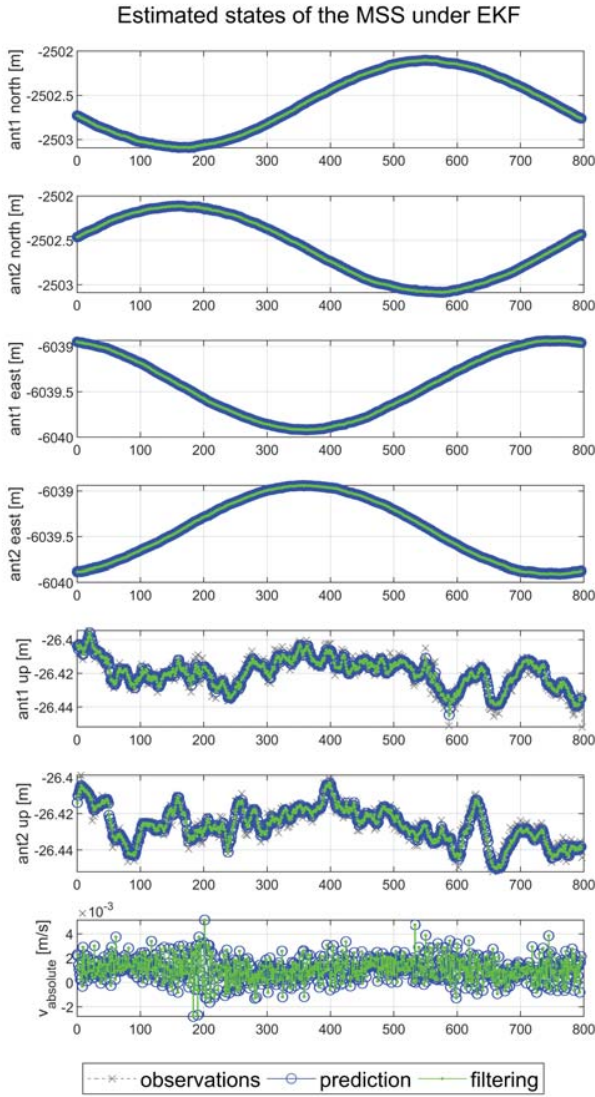


Fig. 2. 7 Estimated State Parameters of the MSS under EKF

The 3D estimated trajectories of the two antennas under EKF were plotted in Figure 3. It is easy to notice that both of the two trajectories of the two antennas are not closed circles, since usually the antennas do not return to their initial positions exactly after one period due to the uncertainty in the measurement.

Estimated Trajectories of 2 GNSS Antennas in a MSS under EKF

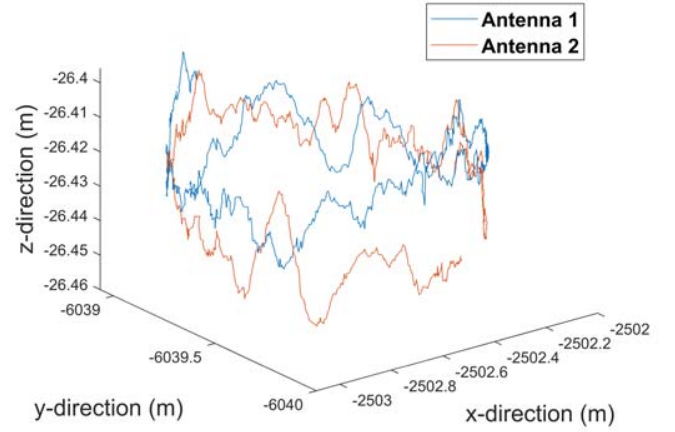


Fig. 3. 3D Estimated Trajectories under EKF

According to [9], EKF actually is a special case of SKF when the weighting parameter $\eta = 0$.

B. SKF ($\eta = 0.5$)

$\eta = 0.5$ is the normal case for SKF when there is no expert-based information available.

The observation data is processed by GNSS analysis and techniques, therefore UBB uncertainty may occur and propagate in the whole procedure, which is an important reason why the set-membership uncertainty is introduced at first and the SKF is applied later. Except for UBB uncertainty bounded by ellipsoid, an alternative approach based on interval mathematics can be found in [13].

Based on the standard variance of each variable in process equation and measurement equation, and also the expert-based information from [2], the UBB of x, y, z, v in process equation are bounded by $[-1 \times 10^{-3}, 1 \times 10^{-3}]$, $[-1 \times 10^{-3}, 1 \times 10^{-3}]$, $[-5 \times 10^{-3}, 5 \times 10^{-3}]$, $[-3 \times 10^{-3}, 3 \times 10^{-3}]$, respectively. And the UBB of x, y, z in measurement equation are bounded by $[-4 \times 10^{-3}, 4 \times 10^{-3}]$, $[-4 \times 10^{-3}, 4 \times 10^{-3}]$, $[-7 \times 10^{-3}, 7 \times 10^{-3}]$, respectively. Also assume that the UBB uncertainties are independent in different epochs. Thus, the initial shape matrices of set-membership uncertainties in process equation and measurement equation are set up as follows:

$$S_0^u = 10^{-6} \cdot \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 25 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 25 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 9 \end{pmatrix}, \quad (25)$$

and

$$S_0^z = 10^{-6} \cdot \begin{pmatrix} 16 & 0 & 0 & 0 & 0 & 0 \\ 0 & 16 & 0 & 0 & 0 & 0 \\ 0 & 0 & 49 & 0 & 0 & 0 \\ 0 & 0 & 0 & 16 & 0 & 0 \\ 0 & 0 & 0 & 0 & 16 & 0 \\ 0 & 0 & 0 & 0 & 0 & 49 \end{pmatrix}. \quad (26)$$

The authors have confidence on the UBB of the initial state therefore a very small shape matrix of the initial state $\hat{x}_0^{c+}$ was chosen:

$$S_0^+ = 10^{-10} \cdot I, \quad (27)$$

where I is a 7-dimensional unit matrix.

Due to the lack of expert-based information, the weight parameter here is $\eta = 0.5$.

The 7 estimated states under the novel SKF are given in Figure. 4-10, respectively.

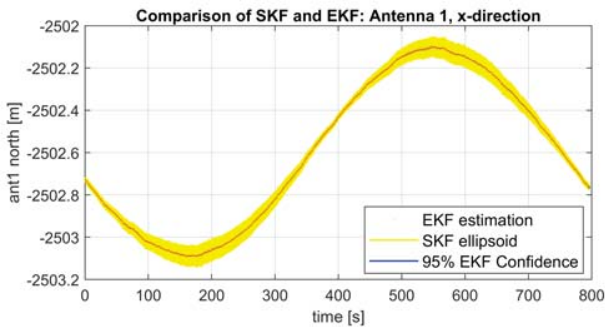


Fig. 4. 99.95% EKF points are located in SKF, antenna 1, x-direction.

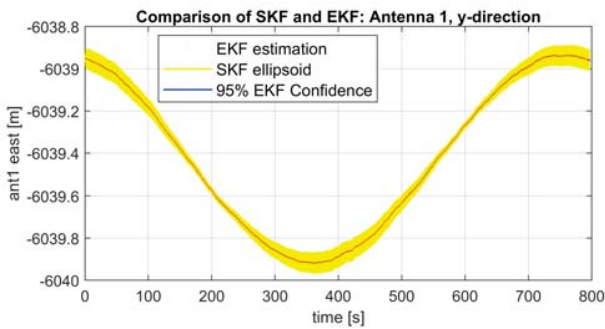


Fig. 5. 100% EKF points are located in SKF, antenna 1, y-direction.

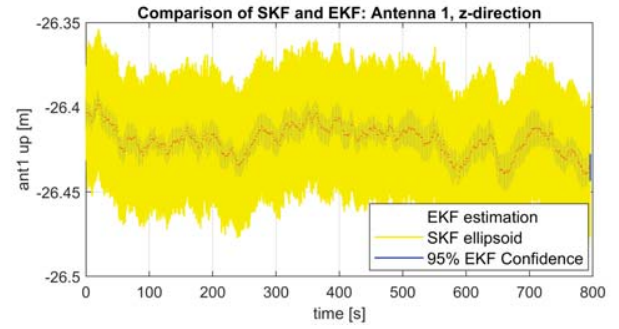


Fig. 6. 100% EKF points are located in SKF, antenna 1, z-direction.

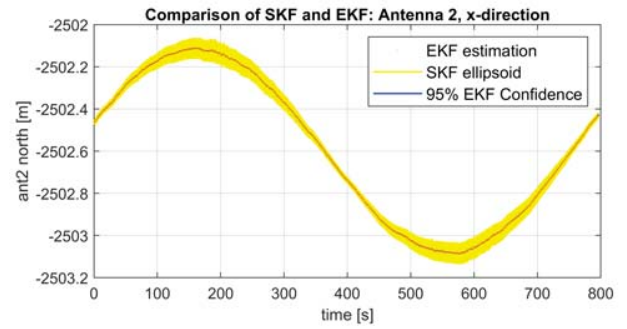


Fig. 7. 99.98% EKF points are located in SKF, antenna 2, x-direction.

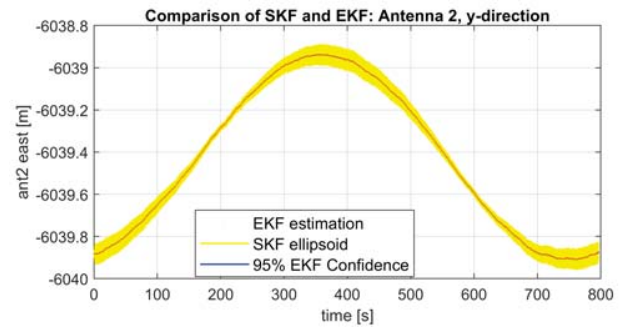


Fig. 8. 100% EKF points are located in SKF, antenna 2, y-direction.

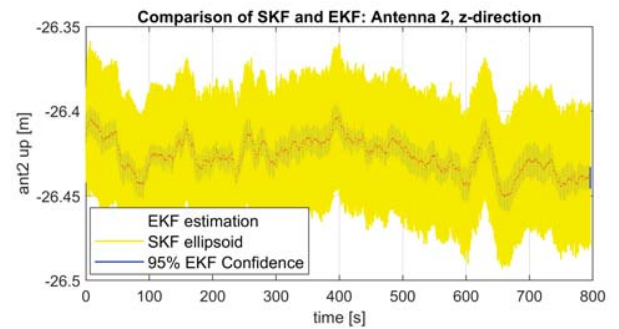


Fig. 9. 100% EKF points are located in SKF, antenna 2, z-direction.

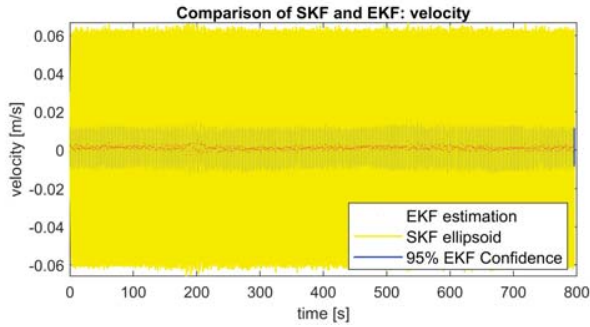


Fig. 10. 100% EKF points are located in SKF, velocity.

In each above figure, the yellow part contains all the output one-dimensional ellipsoids (intervals). The previous EKF estimated points are also plotted in the same figure, and the percentage of the EKF points located inside of SKF ellipsoids is calculated. Almost every EKF estimated point is within an estimated ellipsoid under the novel SKF. The blue one-dimensional ellipsoid (interval) is the 95% confidence area for each EKF estimated point, which is quite small compared to the estimated set-membership part.

The estimated ellipsoid is much larger than the EKF 95% confidence area in every epoch. The geometry of the series of ellipsoids highly relies on the position of the two antennas. The estimated ellipsoid is larger when the velocity is smaller. Furthermore, it is obvious to notice that there are larger uncertainties in z-direction under SKF, which is consistent with the following 3D trajectory, see Figure 11.

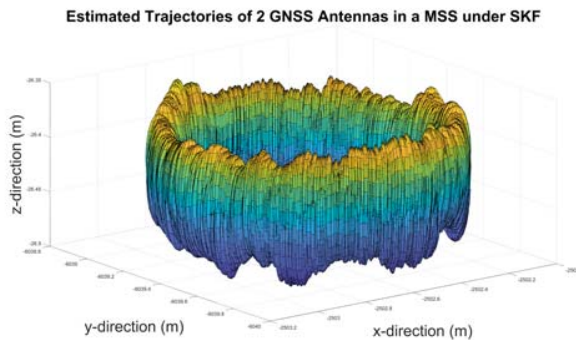


Fig. 11. 3D Estimated Trajectories under SKF

As what we expected to the SKF, in every epoch we get two ellipsoids containing the filtered positions of the two antennas, respectively. A major difference with the standard KF is that, the estimated states are ellipsoids instead of single points, and every inner point of one ellipsoid has the same estimation status. But one still can choose a series of specific points in these ellipsoids if necessary. It is noteworthy that the estimated ellipsoids have larger uncertainty in z-direction than x,y-direction, and the main reason causing this phenomena is that GNSS analysis keeps larger uncertainty in the altitude in

both process equation and measurement equation, see (25) and (26) at the beginning.

C. SKF ($\eta = 1$)

When $\eta = 1$, the SKF reduces to a totally set-based estimation method since there is not any random uncertainty anymore in the system. The estimated results are given in the following figures. Due to Antenna 2 has a very similar performance with Antenna 1, here only the estimated positions of Antenna 1 and the velocity are shown.

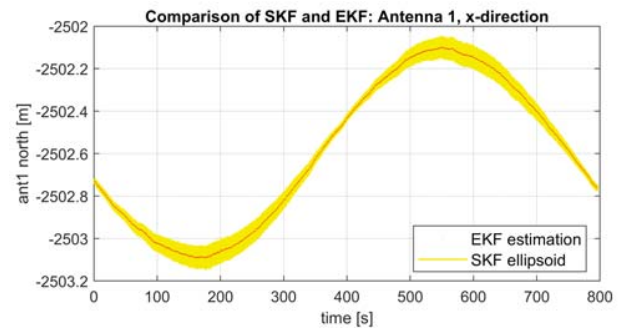


Fig. 12. 100% EKF points are located in SKF, antenna 1, x-direction.

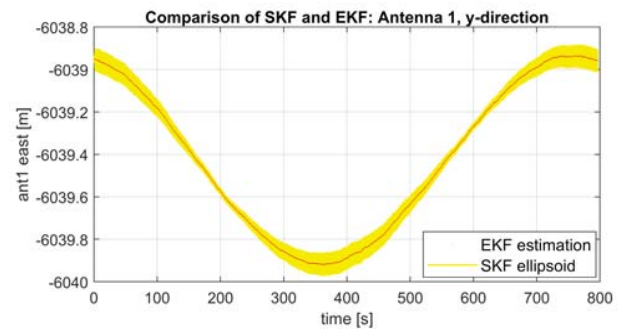


Fig. 13. 100% EKF points are located in SKF, antenna 1, y-direction.

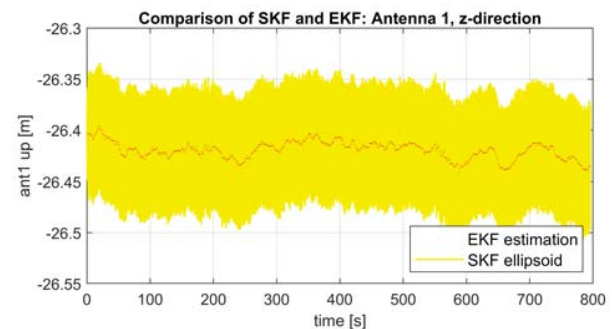


Fig. 14. 100% EKF points are located in SKF, antenna 1, z-direction.

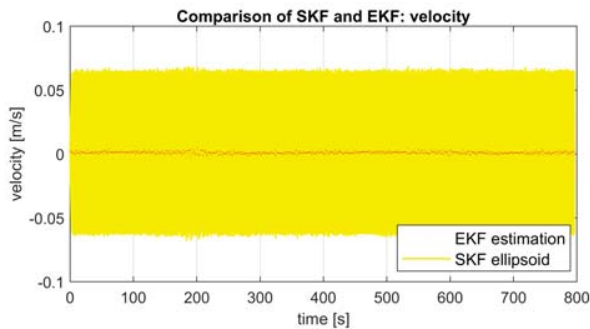


Fig. 15. 100% EKF points are located in SKF, velocity.

V. CONCLUSION AND FUTURE WORK

This paper shows how the novel ellipsoid-based SKF works on a real world experimental data set, and the SKF is one reasonable and applicable model when some unknown but bounded uncertainties were included in the nonlinear system. Under the statistical and UBB assumption, state estimation problem can be solved properly and one can also carry reachability analysis for the nonlinear systems. The MSS was designed to do further direct geo-referencing of 3D point clouds, which is closely relative to information fusion problem but beyond the scope of this paper.

It is worthy to notice that there is no reference data that can be used as true states of this system in this experiment. If the reference data was available, then different criteria can be used to compare the efficiency and accuracy of EKF and SKF. In the future, the data sets containing not only measurement information but also reference data will be used to validate if the SKF is much more reliable or efficient than the EKF. Further, detailed analysis of the propagation process would be done to find the specific reasons that the filtered ellipsoids are in such size scales. In the future this SKF will also be applied on some real data sets which were collected from dynamical sensor networks.

ACKNOWLEDGMENT

This work was supported by the German Research Foundation (DFG) as part of the Research Training Group i.c.sens (RTG 2159). The authors greatly acknowledge Dr. Yi Zhang in Leibniz Universität Hannover for his helpful suggestion on the computation of the experiment. The authors also thank the five anonymous reviewers for their valuable comments.

REFERENCES

- [1] R.E. Kalman, "A new approach to linear filtering and prediction problems," *Journal of Basic Engineering*, vol. 82.1, pp. 35-45, 1960.
- [2] J.-A. Paffenholtz, "Direct geo-referencing of 3D point clouds with 3D positioning sensors," PhD thesis, German Geodetic Commission (DGK), Series C, No. 689, Munich, 2012.
- [3] J.-A. Paffenholtz, H. Alkhatib, H. Kutterer, "Direct geo-referencing of a static terrestrial laser scanner," *Journal of Applied Geodesy*, vol. 4, No. 3, pp. 115-126, 2010.
- [4] A. Doucet, N. de Freitas, N. Gordon, "Sequential Monte Carlo Methods in Practice," Springer, New York, Berlin, 2001.
- [5] D. Simon, "Optimal State Estimation," Wiley, Hoboken, New Jersey, 2006.
- [6] A. Gelb, "Applied Optimal Estimation," MIT Press, Cambridge, MA, 1974.
- [7] M. Milanese, A. Vicino, "Optimal estimation theory for dynamic systems with set-membership uncertainty: An overview," *Automatica*, 27(6), pp.997-1009, 1991.
- [8] B. Noack, "State estimation for distributed systems with stochastic and set-membership uncertainties," KIT Scientific Publishing, vol.14, 2014.
- [9] L. Sun, H. Alkhatib, B. Kargoll, V. Kreinovich, I. Neumann, "A New Kalman Filter Model for Nonlinear Systems Based on Ellipsoidal Bounding," arXiv:1802.02970, 2018.
- [10] C. Durieu, E. Walter, B. Polyak, "Multi-input multi-output ellipsoidal state bounding," *Journal of optimization theory and applications*, 111(2): 273-303, 2001.
- [11] N. Nadarajah, J.-A. Paffenholtz, P. J. Teunissen, "Integrated GNSS Attitude Determination and Positioning for Direct Geo-Referencing," in *Sensors*, vol. 14, no. 7, pp. 12715-12734. DOI: 10.3390/s140712715.
- [12] S. Rump, "INTLAB - INTerval LABoratory," in: *Developments in Reliable Computing*, Springer, Dordrecht, pp. 77-104, 1999.
- [13] S. Schön, H. Kutterer, "Uncertainty in GPS Networks due to Remaining Systematic Errors: The Interval Approach," *Journal of Geodesy*, 80(3): 150-162, 2006.

7.3 Originalpublikationen zur Kapitel 4

- P#14 Alkhatib, H.;** Neumann, I.; Kutterer, H. (2009): Uncertainty modeling of random and systematic errors by means of Monte Carlo and fuzzy techniques, In: Journal of Applied Geodesy, Jg. 3, Nr. 2/2009, S. 67–79. DOI: 10.1515/JAG.2009.008
- P#15 Alkhatib, H.;** Kutterer, H. (2013): Estimation of Measurement Uncertainty of kinematic TLS Observation Process by means of Monte-Carlo Methods, In: Journal of Applied Geodesy, Jg. 7, Nr. 2/2013, S. 125–134. DOI: 10.1515/jag-2013-0044
- P#16 Alkhatib, H.;** Kargoll, B.; Paffenholz, J.A. (2018): Further results on robust multivariate time series analysis in nonlinear models with autoregressive and t-distributed errors, In: Rojas, I.; Pomares, H.; Valenzuela, O. (Hrsg.) Time Series Analysis and Forecasting. ITISE 2017. Contributions to Statistics, S. 25-38. Springer, Cham. DOI: 10.1007/978-3-319-96944-2_3, ISBN: 978-3-319-96943-5
- P#17 Zhao, X.;** Kargoll, B.; Omidalizarandi, M.; Xu, X.; **Alkhatib, H.** (2018): Model selection for parametric surfaces approximating 3D point clouds for deformation analysis, Remote Sensing, Jg. 10, Heft 4, Nr. 634, Special Issue: "3D Modelling from Point Clouds: Algorithms and Methods. DOI: 10.3390/rs10040634
- P#18 Alkhatib, H.;** Omidalizarandi, M.; Kargoll, B. (2019): A Bootstrap Approach to Testing for Time-Variability of AR Process Coefficients in Regression Time Series with t-Distributed White Noise Components. In: International Association of Geodesy Symposia. Springer, Berlin, Heidelberg. DOI: 10.1007/1345_2019_78

Uncertainty modeling of random and systematic errors by means of Monte Carlo and fuzzy techniques

Hamza Alkhatib, Ingo Neumann and Hansjörg Kutterer

Abstract. The standard reference in uncertainty modeling is the “Guide to the Expression of Uncertainty in Measurement (GUM)”. GUM groups the occurring uncertain quantities into “Type A” and “Type B”. Uncertainties of “Type A” are determined with the classical statistical methods, while “Type B” is subject to other uncertainties which are obtained by experience and knowledge about an instrument or a measurement process. Both types of uncertainty can have random and systematic error components. Our study focuses on a detailed comparison of probability and fuzzy-random approaches for handling and propagating the different uncertainties, especially those of “Type B”. Whereas a probabilistic approach treats all uncertainties as having a random nature, the fuzzy technique distinguishes between random and deterministic errors. In the fuzzy-random approach the random components are modeled in a stochastic framework, and the deterministic uncertainties are treated by means of a range-of-values search problem. The applied procedure is outlined showing both the theory and a numerical example for the evaluation of uncertainties in an application for terrestrial laserscanning (TLS).

Keywords. GUM, Monte Carlo methods, fuzzy methods.

1. Introduction

The “Guide to the Expression of Uncertainty in Measurement (GUM)” is a standard reference in uncertainty modeling in engineering and mathematical sciences, cf. ISO (1995). GUM groups the occurring uncertain quantities into “Type A” and “Type B”. Uncertainties of “Type A” are determined with the classical statistical methods, while “Type B” is subject to other uncertainties which are obtained by experience and knowledge about an instrument or a measurement process. Whereas the uncertainties of the quantities of “Type A” can be estimated based on repeated measurement of the quantity of interest, the estimated uncertainties of the quantities of “Type B” are based on expert knowledge, e.g., the technical knowledge about an instrumental error source. Both types of uncertainty can have random and systematic error components:

- A random error ε arises from non predictable variations of some influence factors under seemingly the same actual conditions (non reproducible effects), see, e.g., Bandemer (2006, pp. 63 ff).

- A systematic error δ is due to non controllable effects during the measurement and the preprocessing steps of the measurement, it biases the output quantity y . Although systematic errors are unknown, they bias the measurement result in one direction (reproducible, but unknown effects), see, e.g., Grabe (2005).

GUM defines an output quantity y as a function of input quantities $\mathbf{z}$. The input quantities can be considered as influence parameters which, e.g., can be relevant in pre-processing steps:

$$y = f(z_1, z_2, \dots, z_n) = f(\mathbf{z}) \quad (1)$$

with $f(\cdot)$ the observation model and n the number of input quantities $\mathbf{z}$:

- “..., whose values and uncertainties are directly determined in the current measurement (original measurement).”
- “..., whose values and uncertainties are brought into the measurement from external sources, like the values from a calibration for an instrument (influence factor) (ISO 1995, Chapter 4.1.3).”

Please note that in general the input quantities z_i may be a measurement result y themselves. In order to have a clear representation, only the case where z_i is a measurement or an influence factor is treated in this paper. The quantity z_i can be carrier of both random and systematic errors. GUM proposes to treat random and systematic errors in a stochastic framework. It introduces variances to describe their uncertainties.

Let us denote the function $f(\cdot)$ from equation (1) as observation model and divide the uncertain influence factors into three groups: additional information, sensor parameters, and model constants. Whereas the uncertainty of the original measurement is usually of “Type A”, the uncertainty of the influence factors can be of “Type A” or “Type B”. Figure 1 shows the interaction between the measurement, the influence factors and the observation model. Systematic errors of the input quantities are meaningful by many reasons:

- The model constants are only partially representative for the given situation (e.g., the model constants for the refraction index for distance measurements).
- The number of additional information (measurements) may be too small to estimate reliable distributions for a random treatment.
- Measurement results are affected by rounding errors.

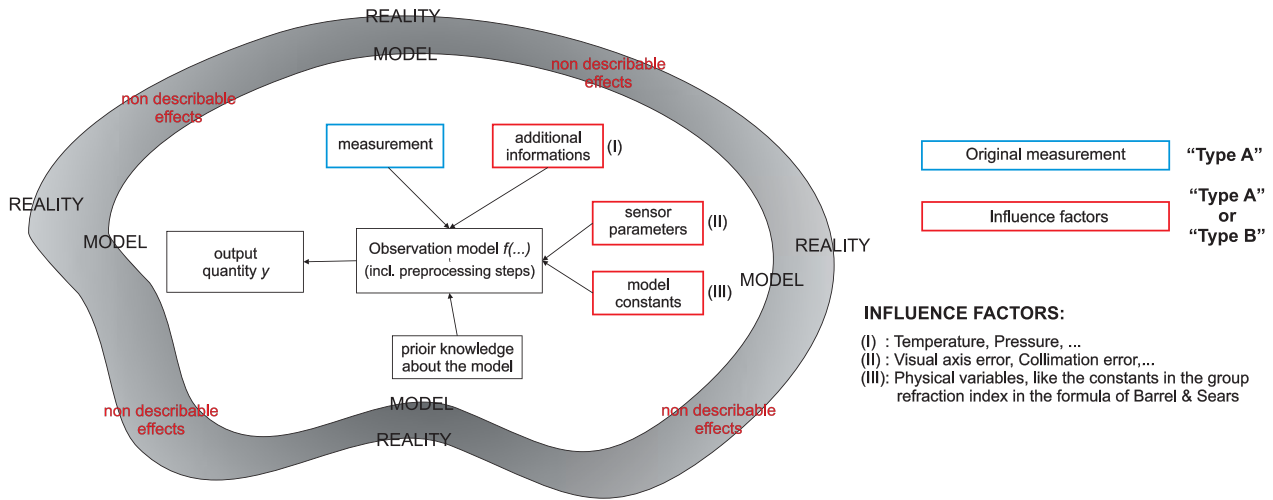


Figure 1: Interaction between input quantities, the observation model and the output quantities.

The GUM, cf. ISO (1995, p. 19), describes an approach for determining the standard uncertainty $u(y)$ of the measurement result y from the standard uncertainties and correlation coefficients associated with the input estimates using the law of propagation of uncertainties (LOP):

$$u^2(y) = \sum_{i=1}^n c_i^2 u^2(z_i) + 2 \sum_{i=1}^{n-1} \sum_{j=i+1}^n c_i c_j u(z_i) u(z_j) r(z_i, z_j), \quad (2)$$

where the quantities $c_1, c_2, \dots, c_n$ are partial derivatives of $y = f(\mathbf{z})$ with respect to the random inputs $Z_1, Z_2, \dots, Z_n$ evaluated at the realization $z_1, z_2, \dots, z_n$ respectively, and $r(z_i, z_j)$ is the correlation coefficient between Z_i and Z_j for $i, j = 1, \dots, n$ and $i \neq j$. The equation (2) delivers the combined standard uncertainty $u(y)$ of the measurement result y from the standard uncertainties and correlation coefficients associated with the input estimates using the LOP of uncertainties.

Equation (2) can be reformulated in matrix notation, cf., e.g., Koch (1999, p. 100):

$$u^2(y) = \mathbf{A} \boldsymbol{\Sigma}_{zz} \mathbf{A}^T, \quad (3)$$

where the matrix $\mathbf{A}$ contains the partial derivatives $y = f(\mathbf{z})$ with respect to $Z_1, Z_2, \dots, Z_n$, that is $\mathbf{A} = \left[\frac{\partial f}{\partial z_1}, \dots, \frac{\partial f}{\partial z_n} \right]$, and $\boldsymbol{\Sigma}_{zz}$ is the uncertainty matrix of the input quantities $\mathbf{z}$.

Unfortunately, in many practical applications, the models are neither linear nor can be approximated by a linearized model using Taylor series expansions. As a consequence of the high non-linearity, the LOP could be difficult to use Hennes (2007). For cases where the model for evaluating the uncertainty is strongly nonlinear or highly complicated to linearize (using Taylor series expansions), the GUM framework will not be satisfied. That is because the standard GUM framework uses only a measurement

model linearized about the best available estimates of the input quantities. For this reason the *Extension of GUM* (ISO 2007) recommends the propagation of uncertainties using a probabilistic approach. Within the mentioned approach the propagation of uncertainties is numerically treated by Monte Carlo (MC) techniques. The difference between the GUM (ISO 1995) and the extension of GUM (ISO 2007) (the so called probabilistic approach) in case of nonlinearity and/or Non-Gaussianity will be not significantly differ in the first and the second central moments but rather in the estimate of the confidence region, which are reflected in the non-Gaussian PDF of the output quantities.

The acceptance of MC techniques has significantly increased during the last decade. Consequently, it's widely used within many scientific disciplines. Hennes (2007) suggested to use MC simulations instead of the treatment of the combined uncertainties by applying the LOP. Siebert and Sommer (2004) recommended a MC based method to evaluate the measurement uncertainties in non-linear models. Koch (2008a) suggested the determination of the uncertainty according to GUM by a Bayesian confidence interval using MC simulation. The approach has been explained in detail and applied to the results of terrestrial laserscanning (TLS). Furthermore, the approach has been extended in Koch (2008b) to evaluate uncertainties of correlated measurements by another application in TLS.

GUM assumes that random and systematic uncertainties are both appropriately handled by means of power density functions (PDFs). Neumann (2009, pp. 24 ff) and the references therein show that a "pure" probabilistic approach in some cases can lead to a too optimistic evaluation of uncertainties. Too optimistic evaluation means that, e.g., the confidence intervals of output quantities are too narrow in comparison to the true values. This shall be highlighted with two references. On the one hand McNish (1962) presents in his paper a too optimistic estima-

tion of confidence intervals for the measurement of the speed of light. On the other hand Stigler (1996) and Youden (1972) illustrate a difference between confidence intervals with the true values for the astronomical unit. Hence, in this paper an alternative approach for evaluating the uncertainties will be introduced. In this approach the systematic component is evaluated using fuzzy techniques (in contrast to the “pure” probabilistic approach). The fuzzy techniques have proven to be an appropriate solution for the description of uncertainties and were applied in different science and engineering applications.

The basic idea in this paper is related to the so-called Fuzzy-Random-Variables (FRV) which are based on a combination of probability theory and fuzzy theory, see Kwakernaak (1978) and Kwakernaak (1979) for more details. For the theory of FRV the reader is referred to Terán (2007). Further discussions of fuzzy techniques in the context of GUM can be also found in Mauris et al. (2001). The description of the systematic uncertainty component of the input quantities with the fuzzy-random approach leads to a more pessimistic uncertainty of the output quantity. In this paper the probabilistic approach and fuzzy-random approach shall be thoroughly compared to each other from a theoretical point of view and using a practical example in TLS.

The paper is organized as follows: First we will describe the general idea of MC techniques to describe measurement uncertainties in the context of GUM; second a fuzzy-random approach to handle these measurement uncertainties is introduced. Then, the computation steps of both approaches are compared with the GUM approach. In the fifth section all three methods are applied to TLS and the obtained results are critically compared to each other. The paper finishes with a discussion and an outlook for further research.

2. Uncertainty modeling with probabilistic approach

In MC techniques, both the random and the systematic components of the uncertainty are treated as having a random nature. Please note that not the systematic component itself is modeled as random, it is the knowledge about the systematic component for which a PDF is introduced (Koch 2007). The GUM suggested in some cases to select the PDF of the input quantities as rectangular, triangular, and trapezoidal (ISO 1995). In these cases, it is hard or even impossible to obtain the estimate of the uncertainty for the output quantity in a closed mathematical form. An alternative to modeling and propagating first and second moments according to equation (2) or (3) is propagating PDFs of the observation model from equation (1) by MC simulations:

$$Y = f(Z_1, Z_2, \dots, Z_n) = f(\mathbf{Z}). \quad (4)$$

Here Y represents a random output quantity and $Z_1, Z_2, \dots, Z_n$ are the n random inputs.

2.1. Sampling from PDF

Any MC simulation requires random numbers. Random numbers are generated on a computer by means of deterministic procedures. Mostly, rectangular distributed random numbers are generated, which may then in turn be transformed into random numbers of random variables having other PDFs (Gentle 2003).

Starting from pseudo-random numbers $u_1, u_2, \dots, u_n$, generated by one of the standard methods such as linear congruence method (Koch 2007, p. 183), some random numbers $x_1, x_2, \dots, x_n$ may be generated which may be viewed as realizations of random variables $X_1, X_2, \dots, X_n$ with another PDF, for instance as realizations $x_1, x_2, \dots, x_n$ of normally distributed random variables. This process makes in particular use of the so-called inversion method or acceptance-rejection method. A far more comprehensive discussion of such algorithms can be found, e.g., in Gentle (2003) or Koch (2007). Multivariate uniform, trapezoidal and triangular PDFs might be needed for evaluating the uncertainties of the measurements according to GUM. In this paper, the inversion method is used to generate rectangular-distributed and triangular-distributed random numbers. The main reason for the selection of the inversion method is its implementation simplicity. Detailed explanation of this method can be found, e.g., in Gentle (2003).

To demonstrate the modeling of uncertainties with a MC simulation in Section 5, the generation algorithms of random numbers from rectangular, triangular and normal PDF will be shortly described according to Koch (2007):

- *Generation of rectangular-distributed random numbers:* In a rectangular PDF, the values of a random variable lie within an interval $[a_-, a_+]$, where a_- and a_+ are the distribution parameters, the upper and the lower limits of the rectangular PDF. That means, there is no specific knowledge about the values within the interval, cf. Koch (2008a) (typical examples are error bounds, rounding errors and digitalization errors). The PDF for a continuous random variable of the uniform PDF, can be defined by (cf., e.g., Koch (2007, p. 20)):

$$p(x | a_-, a_+) = \begin{cases} \frac{1}{a_+ - a_-} & \text{if } a_- \leq x \leq a_+, \\ 0 & \text{if } x < a_- \text{ and } x > a_+. \end{cases} \quad (5)$$

The expected value $E(X)$ and the variance $V(X)$ of the rectangular PDF are given by (cf. Koch (2008a)):

$$E(X) = \frac{a_+ + a_-}{2}, \quad V(X) = \frac{(a_+ - a_-)^2}{12}. \quad (6)$$

- *Generation of triangular-distributed random numbers:* The symmetric triangular PDF is a special

case of the symmetric trapezoidal PDF. As an illustrative example for this type of distributions, the triangular PDF is suggested as a distribution for an addition constant of an instrument. The symmetric triangular cumulative distribution function (CDF) is then given by:

$$F(x | a_-, a_+) = \begin{cases} \frac{(x-a_-)^2}{2a^2} & \text{if } a_- \leq x \leq \frac{a_+ + a_-}{2}, \\ 1 - \frac{(a_+ - x)^2}{2a^2} & \text{if } \frac{a_+ + a_-}{2} \leq x \leq a_+, \end{cases} \quad (7)$$

where $a = \frac{a_+ - a_-}{2}$. In order to generate random numbers $x_1, x_2, \dots, x_n$ from the triangular distribution, the inverse function of the CDF in equation (7) is derived. The method is similar to the one used to generate random variables from the uniform distribution. One starts with one-dimensional vector of standard uniform-distributed random numbers, $H \sim \mathcal{U}(0, 1)$. These can be transformed to the desired distribution using the inverse function of the CDF (7):

$$F^{-1}(h) = \begin{cases} a_- + \sqrt{2ha^2} & \text{if } 0 \leq h \leq \frac{1}{2}, \\ a_+ - \sqrt{2a^2(1-h)} & \text{if } 0 \leq (1-h) \leq \frac{1}{2}. \end{cases} \quad (8)$$

The expected value and the variance of the triangular PDF are given by (cf. Koch (2008a)):

$$E(X) = \frac{a_+ + a_-}{2}, \quad V(X) = \frac{(a_+ - a_-)^2}{24}. \quad (9)$$

- *Generation of correlated normally-distributed random numbers:* The normal PDF is the most widely used PDF in statistics. The one-dimensional normal PDF is fully described by its expected value μ and its variance σ^2 , that is $X \sim \mathcal{N}(\mu, \sigma^2)$. The multinormal PDF is a building block for the some input quantities of the random variables in equation (4). It is well known that the multinormal PDF is fully characterized by its expected value $\boldsymbol{\mu} \in \mathbb{R}^n$ and its variance-covariance matrix $\boldsymbol{\Sigma} \in \mathbb{R}^{n \times n}$, where n is the length of the random vector. Consider a vector $\mathbf{Z} = (Z_1, \dots, Z_n)^T$ of independent standard-normally distributed random variables $\mathbf{Z} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$. As $\boldsymbol{\Sigma}$ is positive definite then there exists the Cholesky decomposition $\boldsymbol{\Sigma} = \mathbf{R}^T \mathbf{R}$, where $\mathbf{R}$ is an $n \times n$ upper triangular matrix. The standard normal numbers can be then transformed to the desired distribution (that is $\mathbf{X} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$) using $\mathbf{X} = \mathbf{R}^T \mathbf{Z} + \boldsymbol{\mu}$. Another possibility to generate multivariate normal random variables is to factorize the covariance matrix $\boldsymbol{\Sigma}$ using singular value decomposition, cf., e.g., Koch (2007, p. 196).

2.2. MC technique to evaluate uncertainty

The MC techniques are of great importance for uncertainty evaluation. With a set of generated samples the PDF for the value of the output quantity Y in equation (4) will be numerically approximated. MC approaches to estimate the uncertainty include the following steps:

- *Step 1:* A set of random samples $z_1, z_2, \dots, z_n$, which have the size n , is generated from the PDF for each random input quantity $Z_1, Z_2, \dots, Z_n$. The sampling procedure is repeated M times for every input quantity.
- *Step 2:* The output quantities y will be then calculated by:

$$y^{(i)} = f(z_1^{(i)}, z_2^{(i)}, \dots, z_n^{(i)}) = f(\mathbf{z}^{(i)}), \quad (10)$$

where $i = 1, \dots, M$ are the generated samples of the random output quantity Y .

- *Step 3:* Particularly relevant estimates of any statistical quantities can be calculated:

(1) the expectation of the output quantity

$$\hat{E}(f(\mathbf{z})) = \hat{E}(y) = \frac{1}{M} \sum_{i=1}^M f(\mathbf{z}^{(i)}), \quad (11)$$

(2) the estimate of the variance of the output quantity (Alkhatib 2007)

$$\hat{\mathbf{D}}(y) = \frac{1}{M} \sum_{i=1}^M (f(\mathbf{z}^{(i)}) - \hat{E}(f(\mathbf{z}))) \cdot (f(\mathbf{z}^{(i)}) - \hat{E}(f(\mathbf{z})))^T. \quad (12)$$

2.3. Confidence intervals in the MC case

To compute the confidence interval by MC simulation of the output quantity with the significance level of γ (it may be denoted as $y_{\text{conf}, \text{MC}} = [\underline{y}, \bar{y}]$), one has to order the independent variates $y^{(i)}$ from the smallest to the largest. An numerically computed $100(1 - 2\gamma)\%$ confidence interval for the random variable Y is then given (cf. Buckland (1984)):

$$y_{\text{conf}, \text{MC}} = [\underline{y} = y_j, \bar{y} = y_k], \quad (13)$$

where $j = (M + 1)\gamma$ and $k = (M + 1)(1 - \gamma)$. The resulting j and k are in general not integer numbers. In order to round to the nearest integer the linear interpolation between as example y_i and y_{i+1} , where $i < j < i + 1$ is used. Another way to compute the MC confidence region is described in Koch (2008a).

3. A fuzzy-random approach to uncertainty modeling

In this section, a fuzzy-random approach to uncertainty modeling in the context of GUM is introduced. Fuzzy techniques have proven to be an appropriate solution for the description of uncertainties. Recently, many procedures have been introduced in different engineering applications, cf., e.g., Fellin et al. (2005), Kreinovich et al. (2006), Möller and Beer (2004). This includes discussions about combined approaches in fuzzy theory, interval mathematics and probability theory (Ferson et al. 2002) and linguistic motivations, see, e.g., Kreinovich (2007).

In the fuzzy-random approach we distinguish between random and systematic errors in the propagation process of the uncertainties of the input quanti-

ties to the output quantity y . Whereas the random part is treated with the LOP or with the MC technique, systematic errors are propagated by means of a range-of-values search problem (see Section 3.2). Both types of uncertainty are modeled in a comprehensive way, using fuzzy intervals (see Section 3.1). This procedure is in full accordance with the recommendations of the GUM. However, the difference is in the treatment of the uncertainty about the systematic errors, for which no variances but interval radii are introduced.

A comprehensive modeling of uncertainties from the observations to the parameters of interest (including statistical tests) with the approach used here is given in Neumann and Kutterer (2006) and Neumann and Kutterer (2009).

3.1. Uncertainty modeling using fuzzy intervals

The random and systematic components of the uncertainties are characterized with a special case of fuzzy theory, the so called Fuzzy-Randomness (Möller and Beer 2004, Viertl 1996). Each uncertain quantity z_i is exclusively modeled in terms of fuzzy intervals. A fuzzy interval $\tilde{A}$ is uniquely defined by its membership function $m_{\tilde{A}}(x)$ over the set $\mathbb{R}$ of real numbers with a membership degree between 0 and 1:

$$\tilde{A} := \{(x, m_{\tilde{A}}(x)) \mid x \in \mathbb{R}\} \quad \text{with } m_{\tilde{A}}(x) : \mathbb{R} \rightarrow [0, 1]. \quad (14)$$

The membership function of a fuzzy interval can be described by its left $L(\cdot)$ and right $R(\cdot)$ reference function (see also Figure 2):

$$m_{\tilde{A}}(x) = \begin{cases} L\left[\frac{X_m - x - r}{c_l}\right] & \text{for } x < X_m - r, \\ 1 & \text{for } X_m - r \leq x \leq X_m + r, \\ R\left[\frac{x - X_m - r}{c_r}\right] & \text{for } x > X_m + r, \end{cases} \quad (15)$$

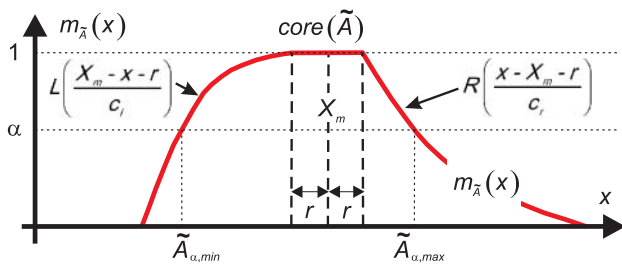


Figure 2: Fuzzy interval and its α -cut.

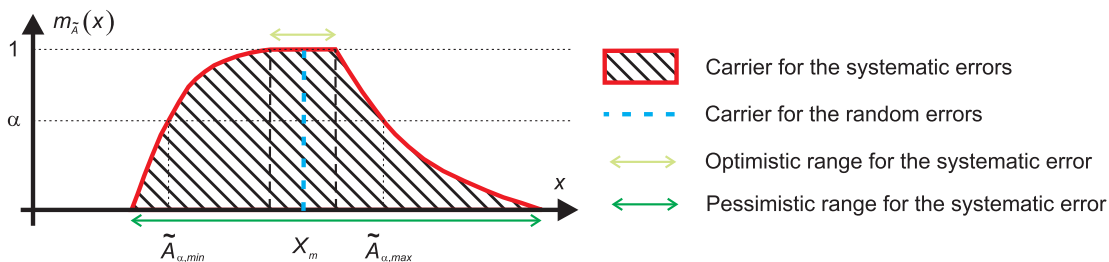


Figure 3: Interpretation of a fuzzy interval.

with X_m denoting the midpoint, r its radius, and c_l, c_r the spread parameters of the monotonously decreasing reference functions (convex fuzzy intervals).

The α -cut $\tilde{A}_\alpha$ with $\alpha \in (0, 1]$ of a fuzzy interval $\tilde{A}$ is defined by

$$\tilde{A}_\alpha := \{x \in \mathbb{R} \mid m_{\tilde{A}}(x) \geq \alpha\}. \quad (16)$$

Each α -cut represents in case of monotonously decreasing reference functions an interval. The lower $\tilde{A}_{\alpha, \min}$ and the upper bound $\tilde{A}_{\alpha, \max}$ of an α -cut and its radius $\tilde{A}_{\alpha, r}$ are

$$\begin{aligned} \tilde{A}_{\alpha, \min} &= \min(\tilde{A}_\alpha), \\ \tilde{A}_{\alpha, \max} &= \max(\tilde{A}_\alpha), \\ \tilde{A}_{\alpha, r} &= \frac{\tilde{A}_{\alpha, \max} - \tilde{A}_{\alpha, \min}}{2}. \end{aligned} \quad (17)$$

The α -cut with $\alpha = 1$ is called the core of $\tilde{A}$ (see Figure 2) with $\text{core}(\tilde{A}) := \{x \in \mathbb{R} \mid m_{\tilde{A}}(x) = 1\}$, and the set with a positive membership function is the support of $\tilde{A}$, that is $\text{supp}(\tilde{A}) := \{x \in \mathbb{R} \mid m_{\tilde{A}}(x) > 0\}$. The membership function can also be constructed on the α -cuts:

$$m_{\tilde{A}}(x) = \sup_{\alpha \in (0, 1]} \alpha \cdot m_{\tilde{A}_\alpha}(x). \quad (18)$$

In this concept fuzzy intervals serve as basic quantities; their midpoints X_m are considered in the following as random variables and their spreads c_l, c_r describe the range of the uncertainty about systematic errors. If one component has random uncertainty only, then this input quantity only consists of a single midpoint as core with radius $r = 0$ and without a left and right reference function. In contrast to the probabilistic approach, the membership function of a fuzzy interval cannot be interpreted in a probabilistic meaning. Therefore the propagation of the systematic uncertainties has to be modified accordingly (see Section 3.2). In the fuzzy case, we model the influence of a systematic component of the uncertainty on the output quantity y . Figure 3 shows the interpretation of a fuzzy interval in the here presented approach. The construction of the membership function can be based on expert knowledge. Each expert provides a range-of-values (an interval) for a systematic error which he considers as realistic. The core of a fuzzy interval describes the range-of-values where all experts agree that these values are possible (most

optimistic case). The support represents the most pessimistic expert opinion for the range-of-values for the systematic error. The above described procedure to construct fuzzy intervals is based on the theory of nested sets, see Nguyen and Kreinovich (1996) for a detailed explanation and Neumann (2009) in the context of uncertainty propagation in parameter estimation, and in the context of hypothesis tests.

3.2. Uncertainty propagation with the extension principle

The propagation process of the random and systematic errors is separated in two parts. Whereas the random components are treated with the LOP (GUM, Chapter 5.2) or with the MC technique (see Section 2), the propagation of the systematic errors is a range-of-values search problem. The extension principle introduced by Zadeh (1965) serves as basic rule to propagate both types of uncertainty.

Therefore the mapping $\tau : y = f(\mathbf{z}), \mathbf{z} \in \mathbb{R}^n \rightarrow y \in \mathbb{R}$ can be extended to the case of fuzzy intervals $\tilde{\tau} : \mathcal{J}(\mathbb{R}^n) \rightarrow \mathcal{J}(\mathbb{R})$ with $\mathcal{J}(\mathbb{R}^n)$ the fuzzy intervals over the real numbers in $\mathbb{R}^n$ and $\mathcal{J}(\mathbb{R})$ the fuzzy intervals in $\mathbb{R}$ (Dubois and Prade 1980, p. 37):

$$\tilde{y}_{\alpha, \min} = \min_{z_i \in [\tilde{z}_{i\alpha}]} f(\mathbf{z}), \quad \tilde{y}_{\alpha, \max} = \max_{z_i \in [\tilde{z}_{i\alpha}]} f(\mathbf{z}), \quad (19)$$

with

$$\tilde{y}_{\alpha} = [\tilde{y}_{\alpha, \min}, \tilde{y}_{\alpha, \max}], \quad m_{\tilde{y}}(x) = \sup_{\alpha \in (0,1]} \alpha \cdot m_{\tilde{y}_{\alpha}}(x). \quad (20)$$

The computation of the membership function for the measurement results is based on the α -cuts $\tilde{\mathbf{z}}_{\alpha}$ of the input quantities, within the optimization problem of equation (19). The approximate midpoint of the fuzzy interval for the output quantity y_m is

$$y_m = f(z_{1m}, z_{2m}, \dots, z_{nm}) = f(\mathbf{z}_m). \quad (21)$$

In general, the solution in equation (21) is only correct for sufficiently linear functions $f(\cdot)$. For practical applications linear reference functions for the membership function of the input quantities and a linear function $f(\cdot)$ are of interest. In these cases, the propagation process only needs to be applied for the α -cuts with $\alpha = 0$ and $\alpha = 1$.

3.3. Confidence intervals in the fuzzy case

The computation of confidence intervals in the fuzzy-random approach is also based on the extension principle. Let $k(Z_1, Z_2, \dots, Z_n) = k(\mathbf{Z})$ be a confidence function for the output quantity Y and the realizations of the random variables $\mathbf{Z}$ are given by fuzzy intervals $\tilde{z}_1, \tilde{z}_2, \dots, \tilde{z}_n$. Then the membership function $m_{\tilde{y}_{\text{conf, fuzzy}}}$ of the fuzzy confidence interval $\tilde{y}_{\text{conf, fuzzy}}$ for

the output quantity is given by

$$\begin{aligned} \tilde{y}_{\text{conf, fuzzy}} &= \tilde{k}(\tilde{z}_1, \dots, \tilde{z}_n) = \tilde{k}(\tilde{\mathbf{z}}) = m_{\tilde{y}_{\text{conf, fuzzy}}}(y) \\ &= \sup_{\substack{(z_1, \dots, z_n) \in \mathbb{R}_1 \times \dots \times \mathbb{R}_n \\ y \in k(z_1, \dots, z_n)}} \min(m_{\tilde{z}_1}(z_1), \dots, m_{\tilde{z}_n}(z_n)) \\ &\quad \forall y \in \mathbb{R}. \end{aligned} \quad (22)$$

A geometric interpretation of equation (22) for non-linear functions is difficult. In case of sufficiently linear functions $f(\mathbf{z}_m)$ the fuzzy confidence interval $\tilde{y}_{\text{conf, fuzzy}}$ for the output quantity based on equation (22) is obtained by the combination of both uncertainty components (Kutterer 2002):

$$\begin{aligned} \tilde{y}_{\alpha, \text{conf, fuzzy}} &[\underline{y} - \tilde{y}_{\alpha, r}, \bar{y} + \tilde{y}_{\alpha, r}] \\ \text{with } \tilde{y}_{\alpha, r} &= \frac{\tilde{y}_{\alpha, \max} - \tilde{y}_{\alpha, \min}}{2}. \end{aligned} \quad (23)$$

Whereas the α -level of zero corresponds to the pessimistic case, the optimistic case is obtained for $\alpha = 1$. Please note that only the random uncertainty components from the input quantities $\mathbf{z}$ contribute to the lower and upper bound of the MC confidence interval $\tilde{y}_{\text{conf, MC}} = [\underline{y}, \bar{y}]$.

4. Compilation of the GUM, probabilistic, and fuzzy-random approach

In this section a short compilation of the calculation steps in the GUM, probabilistic, and fuzzy-random approach is given. Figure 4 shows a diagram with the treatment of random and systematic errors of each technique.

In GUM (the left column of Figure 4) the output quantity Y (cf. equation (4)) is treated as a random variable with a specified PDF. The probability function is assumed (or approximated) to be the normal (Gaussian) PDF. The probabilistic approach of y is based on propagating PDFs by numerical MC simulation of the observation model (4) (the middle column of Figure 4). Only if the assumption is fulfilled that the PDFs for all random input quantities $X_1, \dots, X_n$ are full specified, a PDF of the output quantity Y can be computed numerically. On the basis of the generated PDF, the expected value $E(y)$ and its uncertainty $u_{\text{MC}}(y)$ can be determined. The systematic error as well as the random error are fully described in the probabilistic approach by means of PDFs/CDFs. In contrast to the probabilistic approach, the fuzzy-random approach distinguishes between random and systematic errors within the uncertainty propagation process. Whereas the random component is treated as random variable with, e.g., the MC simulations, the systematic component is propagated as a range-of-value search problem. The two different propagation principles of random and systematic errors in the fuzzy-random approach can be covered with the so called FRV from Section 3.

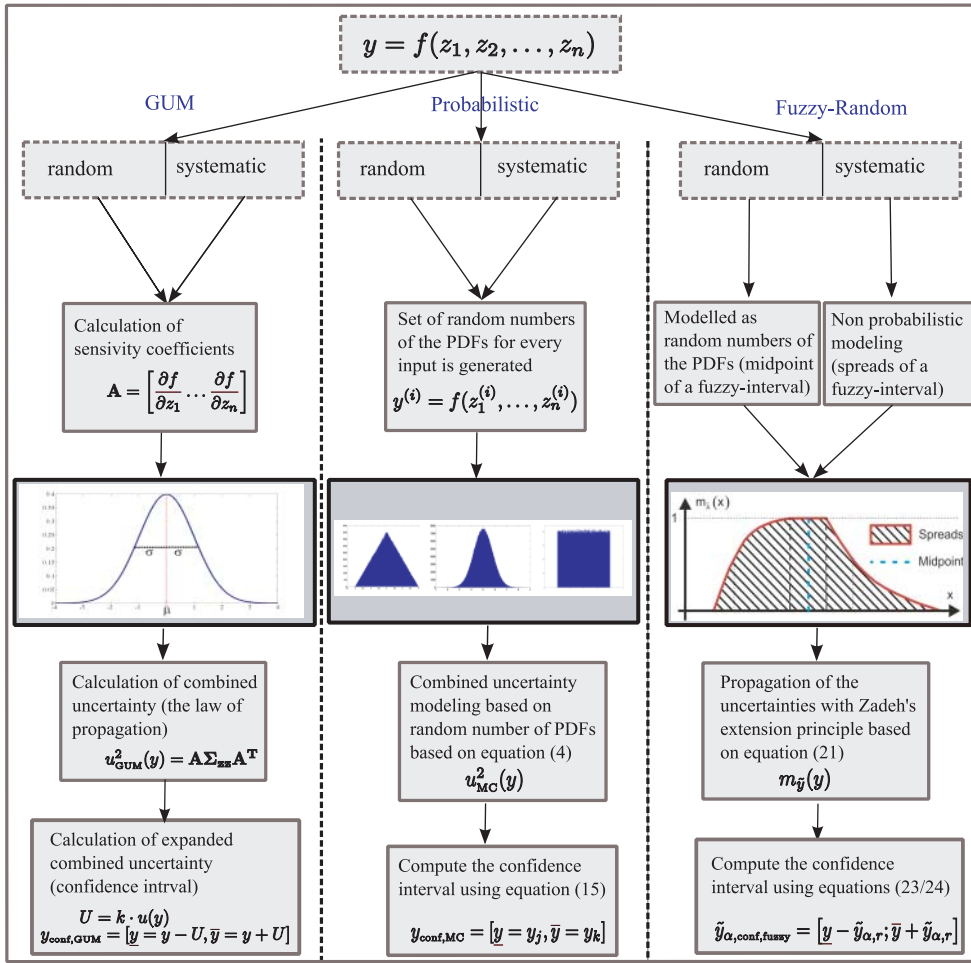


Figure 4: Comparison of the GUM, the probabilistic, and the fuzzy-random approach.

This leads to the uncertainty propagation given in equation (19).

5. Numerical example for an application to TLS

In this section a short numerical example for the comparison of the two approaches from Section 2 and 3 is presented. The aim was to detect the vertical displacements of the bridge under load, e.g., due to traffic or train crossings (Strübing 2007). For this reason, a laserscanner of type Leica HDS 4500 was placed beneath the bridge; the measurements in the “Profiler Mode” span the green plane in Figure 5. The discrepancies to the standard case of normal distributed measurements are meaningful by many reasons (see also Section 1). The laserscanner carries out very fast measurements and the measurements are influenced by vibrations due to the traffic load of the bridge. The time series of the vertical height h_t of the bridge at the stations 7.28 m and 21.90 m can be expressed in the local coordinate system of the laserscanner by the following equation:

$$h_t = s_t \cdot \cos(\zeta_t), \quad (24)$$

with the slope distance s_t and the zenith angle ζ_t , measured by the laserscanner. The number of measured epochs q is 100. The vertical displacements w_t of the bridge are obtained by subtracting the mean height of the bridge from the time series in equation (24):

$$w_t = h_t - \frac{1}{q} \sum_{t=1}^q h_t = s_t \cdot \cos(\zeta_t) - \frac{1}{q} \sum_{t=1}^q s_t \cdot \cos(\zeta_t). \quad (25)$$

The equation (25) is evaluated for two points. The first point 1831 is at station 7.28 m and point 8987 at station 21.90 m (see Figure 5).

5.1. Uncertainties for the measurements and influence factors

The uncertainty of output quantity $y \triangleq w_t$ depends on the following input quantities z_i :

- accuracy of the distance (z_1 , Type A), and its additional constant (z_2 , Type B),
- distance depending term for the accuracy of the distance measurement (z_3 , Type B),
- spatial direction of incidence angle of the measured distance under the bridge with respect to the object surface (z_4 , Type B),

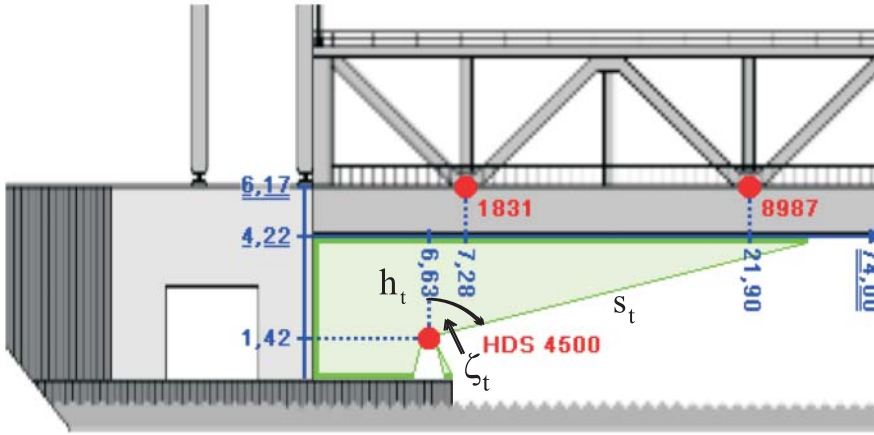


Figure 5: Position of the laserscanner beneath the bridge (Strübing 2007).

Table 1: Uncertainties for the input quantities z .

Input quantity	Error component	PDF/ $m_{\tilde{y}}(z_i)$	Type	Uncertainty	$u(z_i)$ for GUM
z_1	random	normal	A	$\sigma = 3$ mm	$u(z_1) = 3$ mm
z_2	systematic	triangular	B	$(a_+ - a_-)/2 = 3$ mm $\tilde{z}_{x=0,r} = 3$ mm	$u(z_2) = 1.225$ mm
z_3	random	normal	B	$\sigma = 0.2$ mm (1831) $\sigma = 0.9$ mm (8987)	$u(z_3) = 0.2$ mm (1831) $u(z_3) = 0.9$ mm (8987)
z_4	random	normal	B	$\sigma = 2.6$ mm (1831) $\sigma = 7.2$ mm (8987)	$u(z_4) = 2.6$ mm (1831) $u(z_4) = 7.2$ mm (8987)
z_5	random	normal	A	$\sigma = 20$ mgon	$u(z_5) = 20$ mgon
z_6	systematic	triangular	B	$(a_+ - a_-)/2 = 20$ mgon $\tilde{z}_{x=0,r} = 20$ mgon	$u(z_6) = 8.165$ mgon
z_7	systematic	rectangular	B	$(a_+ - a_-)/2 = 10$ mgon $\tilde{z}_{x=0,r} = 10$ mgon	$u(z_7) = 5.774$ mgon

- accuracy of the zenith angle (z_5 , Type A) and the vertical index error (z_6 , Type B),
- vertical resolution for the zenith angle (the step width of the motor) (z_7 , Type B).

The uncertainties described by the PDF/membership function for the input quantities z_i are given in Table 1. The assumptions for the uncertainties of z_1 , z_5 , and z_6 are based on the technical data from the manufacturer and for the uncertainties of z_2 , z_3 , and z_4 on Schulz and Ingensand (2004) and for z_7 on Reshetnyuk (2006). The input quantities z_3 and z_4 have a correlation of 0.5, according to Koch (2008a). The numbers 8987 and 1831 in the brackets represent the point number (see Figure 5). In order to have an easier representation, each input quantity is modeled either as random or as systematic. In general the uncertainty budget of each input quantity may consist of both a random and systematic component. Please note that the uncertainties of the input quantities of z_2 , z_6 , and z_7 are fully described with their distribution parameters (in this case the upper limit a_+ and the lower limit a_-).

5.2. Specification and discussion of the numerical results

This study focuses on the comparison of the three different techniques to model and propagate the occurring uncertainties in Table 1. The PDFs and the order of magnitude of the uncertainties from Table 1 are in our opinion realistic regarding typical situations in the applications. Their description must be carefully examined in future work, but this is not part of the paper. First the LOP-technique (according to GUM) is used to propagate the uncertainties (Section 5.2.1). Then the MC- and the fuzzy technique from Sections 2 and 3 were applied.

5.2.1. Uncertainties obtained by the GUM approach

The propagation process of the uncertainties in the GUM approach is separated into two steps. In the first step, the combined uncertainty

$$u_{h_t}^2 = \mathbf{A} \Sigma_{zz} \mathbf{A}^T, \quad (26)$$

for the vertical height h_t of the bridge is computed.

The matrices from equation (3) are given by

$$\mathbf{A} = \begin{bmatrix} \cos \zeta_t \\ \cos \zeta_t \\ \cos \zeta_t \\ \cos \zeta_t \\ -s_t \sin \zeta_t \\ -s_t \sin \zeta_t \\ -s_t \sin \zeta_t \end{bmatrix},$$

$$\Sigma_{zz} = \begin{bmatrix} u^2(z_1) & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & u^2(z_2) & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & u^2(z_3) & 0.5u(z_3)u(z_4) & 0 & 0 & 0 \\ 0 & 0 & 0.5u(z_4)u(z_3) & u^2(z_4) & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & u^2(z_5) & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & u^2(z_6) & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & u^2(z_7) \end{bmatrix}. \quad (27)$$

For an easy presentation we assume that the heights of different epochs t are uncorrelated and their uncertainties are assumed to be equal. Using the LOP from equation (26) for equation (25) then leads to:

$$\begin{aligned} u_{w_t}^2 &= \left(\frac{q-1}{q} \right)^2 u_{h_t}^2 + (q-1) \frac{u_{h_t}^2}{q^2} \\ &= \frac{q^2 - 2q + 1 + (q-1)}{q^2} u_{h_t}^2 \\ &= \frac{q-1}{q} u_{h_t}^2 \end{aligned} \quad (28)$$

for the combined uncertainty $u_{w_t}^2$ of the vertical displacements w_t . The uncertainty u_{w_t} using equation (28) are evaluated regarding to the uncertainties of the input quantities given in Table 1. We obtain 0.0043 m and 0.0059 m for the points 1831 and 8987, respectively.

5.2.2. Uncertainties obtained by the probabilistic approach

In the probabilistic approach the random and systematic components from Table 1 are treated as having a random nature. In order to compare the performance of the MC technique to compute the uncertainties for the output quantities, the statistical moments of the PDF for every input quantities z_i can be computed either analytically based on the statistical moments or can be estimated numerically from the generated MC samples. The analytical moments of input quantities, which are rectangular or triangular distributed are evaluated analytically according to equation (6) and (9), respectively. The expected values and variances of the normally distributed input quantities are given according to a prior knowledge or expert statements. As above mentioned, the statistical moments of all input quantities are computed also with the aid of generated MC samples (refer to equations (11) and (12)). The results are given by Table 2.

The confidence intervals of every input quantity are computed only by MC simulations (refer to Section 2.2) for $\gamma = 5\%$. We identify 100 000 MC runs as necessary to obtain the estimates of the statistical moments of the PDF with at least three or four significant digits. The same results were numerically proven in Koch (2008a).

According to Section 2 we obtain the uncertainty and the confidence interval of the output quantity $y \triangleq w_t$ for $M = 100\,000$ runs. The results are given in Table 3.

Comparison of uncertainties, calculated by GUM, with the result of probabilistic approach for the points 1831 and 8987, exhibits small differences of the uncertainty values and confidence intervals. The confidence bounds for point 8987 is slightly asym-

Table 2: Comparison of the statistical values (expected values, variances and confidence intervals) of the analytical PDF and the empirical PDF using MC simulations for the input quantities z .

Input quantity	PDF	$E(z_i)$ [m or gon]		$V(z_i)$ [m ² or gon ²]		95% confidence interval	
		analytical	empirical	analytical	empirical	lower limit	upper limit
z_1	normal	2.87400000	2.87400001	0.00000900	0.00000918	2.879800	2.868000
z_2	triangular	0.00000000	0.00000325	0.00000150	0.00000151	-0.002322	0.0023398
z_3	normal	0.00000000	0.00000000	0.00000003	0.00000003	-0.000340	0.0003435
z_4	normal	0.00000000	-0.00000001	0.00000670	0.00000674	-0.005021	0.0051029
z_5	normal	0.00000000	0.00000000	0.00048400	0.00047041	-0.042459	0.0427390
z_6	triangular	14.5240000	14.5239607	0.00006667	0.00006613	14.508000	14.539000
z_7	rectangular	0.00000000	-0.00000562	0.00083333	0.00083499	-0.047558	0.047073

Table 3: Uncertainties and confidence interval obtained by the MC-technique.

MC-result	Point 1831 (7.28 m)	Point 8987 (21.90 m)
$\hat{\sigma}_y$	4.4 mm	5.9 mm
$y_{\text{conf, MC}} = [\underline{y}, \bar{y}]$ with $\gamma = 5\%$	[-8.6 mm, 8.6 mm]	[-11.6 mm, 11.7 mm]

Table 4: Rang of values of the systematic uncertainty obtained by the fuzzy technique.

fuzzy result (systematic component)	Point 1831 (7.28 m)	Point 8987 (21.90 m)
$\tilde{y}_{\alpha=1,r} = (\tilde{y}_{\alpha=1,\max} - \tilde{y}_{\alpha=1,\min})/2$	0.2 mm	4.8 mm
$\tilde{y}_{\alpha=0,r} = (\tilde{y}_{\alpha=0,\max} - \tilde{y}_{\alpha=0,\min})/2$	10.3 mm	16.1 mm

Table 5: Random component of the uncertainty obtained by the fuzzy technique.

fuzzy result (random component)	Point 1831 (7.28 m)	Point 8987 (21.90 m)
$\hat{\sigma}_y$	3.9 mm	5.4 mm
$y_{\text{conf, MC}} = [\underline{y}, \bar{y}]$ with $\gamma = 5\%$	$[-7.6 \text{ mm}, 7.6 \text{ mm}]$	$[-10.6 \text{ mm}, 10.7 \text{ mm}]$

Table 6: Confidence interval obtained by the fuzzy technique.

fuzzy result (confidence interval)	Point 1831 (7.28 m)	Point 8987 (21.90 m)
$y_{\text{conf, fuzzy}} = [\underline{y} - \tilde{y}_{\alpha=1,r}; \bar{y} + \tilde{y}_{\alpha=1,r}]$	$[-7.8 \text{ mm}, 7.8 \text{ mm}]$	$[-15.4 \text{ mm}, 15.5 \text{ mm}]$
$y_{\text{conf, fuzzy}} = [\underline{y} - \tilde{y}_{\alpha=0,r}; \bar{y} + \tilde{y}_{\alpha=0,r}]$	$[-17.9 \text{ mm}, 17.9 \text{ mm}]$	$[-26.7 \text{ mm}, 26.8 \text{ mm}]$

metric due to deviation of the evaluated PDF of the output quantity from the normal PDF and due to propagating of the MC samples through the nonlinear observation model. In the probabilistic approach the systematic component of the uncertainty can be reduced by repeated measurements. In order to specify this statement an illustrative example will be helpful. In general by a simple computation of the mean value the variance may be reduced if the measurement process is repeated. For more details, the reader is referred to Neumann (2009).

5.2.3. Uncertainties obtained by the fuzzy-random approach

In the fuzzy-random approach the treatment of the random and systematic components in the propagation process of the uncertainties is different to the probabilistic approach, see Section 3. According to Section 3 we obtain the systematic uncertainty of the output quantity $y \triangleq w_t$ for $\alpha = 0$ and $\alpha = 1$ with equations (17), (19) and (21). The results are given in Table 4.

The α -level of zero refers to the pessimistic case and the α -level of one to the optimistic case (see Figure 3). Within the propagation process of the systematic component, the radius $\tilde{z}_{\alpha,r}$ of all random components z_i from Table 1 is zero. In the presented propagation process a systematic error component cannot be reduced by repeated measurements. The range-of-values for the systematic uncertainties cannot be decreased by mean value computation due to the mathematic rules of fuzzy theory, see Neumann (2009). The small systematic error for the Point 1831 is due to the small influence of the systematic errors of the zenith angle. For the propagation process of the random components with the methods described in Section 2, the uncertainty of the input quantities with a systematic error component are set to zero.

As a result, we obtain the uncertainty and the confidence interval of the output quantity $y \triangleq w_t$ for $M = 100\,000$ runs. The results are summarized in Table 5.

Finally, we obtain the confidence interval for the fuzzy-random approach with equation (23) for $\alpha = 0$ and $\alpha = 1$. The results are shown in Table 6.

In comparison to Table 3 the random component of the uncertainties of the output quantity in the fuzzy-random approach are smaller. This is due to the reason that only the input quantities with a random uncertainty component contribute to the random uncertainty of the output quantity. These input quantities which are only affected by systematic uncertainties were not taken into account in the propagation process of random uncertainties.

In this example the reference value for the vertical height of the bridge is obtained by the mean value of the heights from each epoch. As discussed in this section in the probabilistic approach the systematic uncertainty component for the output quantity can be reduced by a mean value computation. In the fuzzy-random approach the uncertainty of a systematic component cannot be reduced by the mean value computation. This leads to a significant larger confidence interval in the fuzzy-random approach from Table 6 in comparison to Table 3. This can especially be highlighted for the pessimistic case with $\alpha = 1$.

5.3. A best and worst case scenario

The last example deals with a case study for different magnitudes for the occurring uncertainties. In order to evaluate the consequences of changing magnitudes for the uncertainties, two scenarios are computed: A best case scenario with a small uncertainty of the input quantity $z_7 : (a_+ - a_-)/2 = 20 \text{ mgon}$ (the step width of the motor, see Table 1), and a worst case

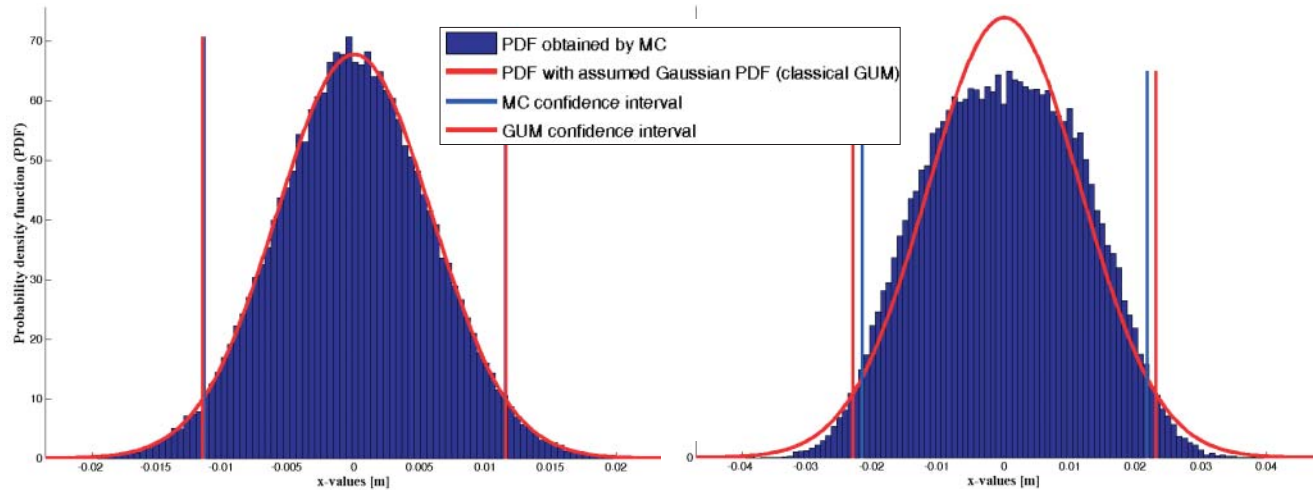


Figure 6: Comparison between the numerically computed MC-PDF and the Gaussian-PDF of the output quantity $y \triangleq w_t$, left the best case scenario with small uncertainties and right the worst case scenario with high uncertainties of the input quantity z_7 .

scenario with high uncertainties of the input quantity z_7 : $(a_+ - a_-)/2 = 50$ mgon. The assumption about the high uncertainties in the worst case scenario are realistic for very fast measurements of the laser-scanner (Zoller+Fröhlich 2007).

5.3.1. Comparing the probabilistic approach with the GUM

Figure 6 shows the MC simulations resulting histogram for the best and the worst case scenario with corresponding theoretical normal fit and the 95% confidence interval. In the best case scenario, the numerical results of the MC-technique and of the LOP don't show any significant differences. In the left part of Figure 6 the resulting MC PDF of the output quantity $y \triangleq w_t$ is compared with the GUM PDF, which is assumed to be Gaussian. The worst case scenario is introduced to find out how much the computed MC PDF and its confidence interval differ from the classical Gaussian PDF and its confidence interval. A comparison pointed out a significant difference between the uncertainty calculated by means of LOP in GUM and by means of MC techniques. The maximum difference in the worst case scenario is obtained in the confidence interval: $\max[|y_{MC} - y_{GUM}|, |\bar{y}_{MC} - \bar{y}_{GUM}|] = 1.25$ mm. Additionally, as we can see from Figure 6, the variation in the form of both PDFs (numerical MC and Gaussian PDF) is perceivable.

This difference between the GUM and the probabilistic approach is due to the nonlinear observation model and non-Gaussian PDF of the output quantity. As mentioned before, this difference will not be significantly in the mean and the variance but rather in the estimate of the confidence interval. Overall, the probabilistic approach can be considered as an accurate and reliable approach for evaluating the uncertainty of the output quantities and its confidence

interval, if the PDFs for the input quantities are known and the observation model (1), which interrelates input and output quantities, is nonlinear.

5.3.2. Comparing the fuzzy-random approach with the GUM

A geometrical interpretation of the systematic error of the height difference $y \triangleq w_t$ (output quantity) in the fuzzy-random approach is given in Figure 7. The range of values for the systematic error can be seen as a shift in the distribution of the random component. For a clear representation the distribution of the random errors of the output quantity (obtained by the MC-technique) is shown at the lower and upper bound of the systematic error. The results for the α -levels 0 (pessimistic case) and 1 (optimistic case) are shown. When comparing the results of the fuzzy confidence interval with the results of the GUM one can clearly see, that the fuzzy confidence intervals are significantly larger. This is due to the reason that in the here presented example the systematic errors dominate the uncertainty budget. The larger the systematic errors are in comparison to the random errors, the stronger is the difference between the GUM confidence interval and the fuzzy confidence interval. This is clearly visible in Figure 7, where the left part shows the example with the best case scenario and the right part the worst case scenario. The difference of the GUM confidence interval and the fuzzy confidence interval is significantly larger in the worst case scenario.

6. Summary and outlook

In this paper, a measurement equation was analyzed with multidimensional input quantities and a one-dimensional output quantity. The input quantities are carrier of random and systematic errors. A probabilistic and a fuzzy-random approach were introduced to handle and to propagate both types of un-

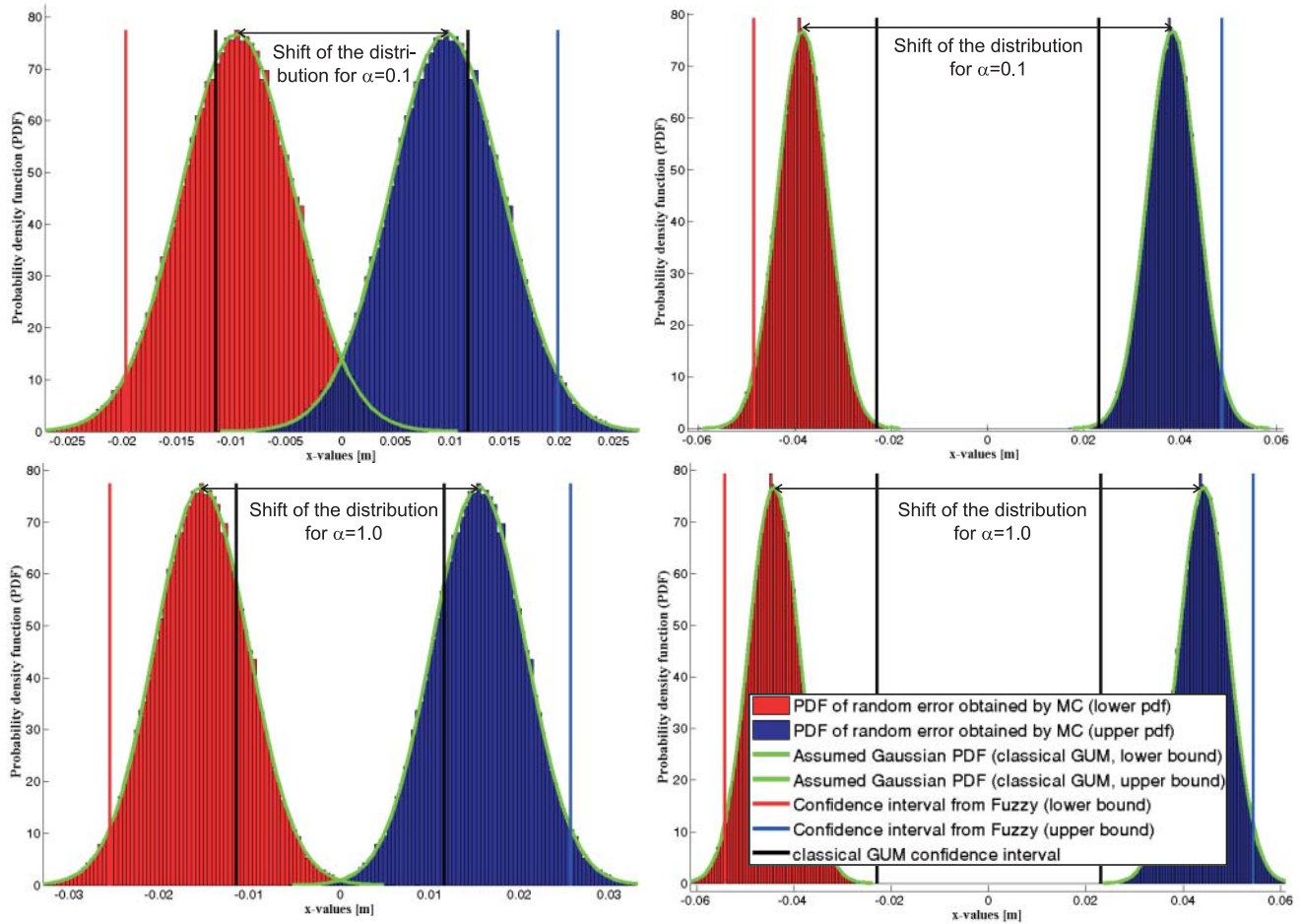


Figure 7: Comparison between the confidence intervals of the fuzzy-random approach and the GUM for the quantity $y \triangleq w_t$, left the best case scenario with small uncertainties and right the worst case scenario with high uncertainties of the input quantity z_7 .

certainties. In the probabilistic approach both types of errors were handled as having a random nature, whereas in fuzzy-random approach the propagation process distinguishes between the random and the systematic error components. Due to the random nature in the probabilistic approach, the uncertainty of the systematic component can be reduced by averaged/repeated measurements. Therefore, it is a more optimistic representation of the uncertainties than in the fuzzy-random approach, where the systematic component of the uncertainties cannot be reduced by averaged/repeated measurements due to the mathematic rules of fuzzy theory.

Three important information can be stated from the results obtained in the paper: First, in case of strong differences from a Gaussian PDF for the random errors, the need for a more sophisticated error propagation process than in the GUM is beneficial. Second, the fuzzy-random approach allows to deal with a pessimistic and optimistic outcome for the uncertainty of the output quantity. Third, both techniques provide a rigorous consideration of all the information contained in a PDF within the propagation process of the random uncertainties to the output quantity.

Further work has to deal with an extended discussion of the presented fuzzy-random approach with input quantities having both types of uncertainties, a random and systematic component. Additionally, the bias of the output quantity resulting from the evaluation of non-linear functions has to be discussed in detail, especially in the fuzzy-random approach. Furthermore, extensions of probabilistic and fuzzy-random approaches to handle with multidimensional correlated uncertainties in the input as well as in the output quantities have to be taken into consideration. Especially, the construction of multidimensional MC- and fuzzy confidence regions needs further studies.

Acknowledgments

The presented paper shows in parts results and new ideas developed during the research project KU 1250/4-2 “Geodätische Deformationsanalysen unter Berücksichtigung von Beobachtungsunpräzision und Objektunschärfe”, which is funded by the German Research Foundation (DFG). This is gratefully acknowledged by the authors.

References

- Alkhatib, H., On Monte Carlo methods with applications to the current satellite gravity missions, PhD thesis, Institute for Geodesy and Geoinformation of University of Bonn, 2007.
- Bandemer, H., Mathematics of Uncertainty – Ideas, Methods, Application Problems, Springer, Berlin, Heidelberg, New York, 2006.
- Buckland, S. T., MC Confidence Intervals, *Biometrika* 40 (1984), 811–817.
- Dubois, D. J. and Prade, H. M., Fuzzy Sets and Systems: theory and applications, Academic Press, London, New York, 1980.
- Fellin, W., Lessmann, H., Oberguggenberger, M., and Vieider, R., editors, Analysing Uncertainty in Civil Engineering, Springer, Berlin, Heidelberg, New York, 2005.
- Ferson, S., Ginzburg, L., Kreinovich, V., Nguyen, H., and Starks, S., Uncertainty in risk analysis: towards a general second-order approach combining interval, probabilistic, and fuzzy techniques, in: Proceedings of the 2002 IEEE International Conference on Fuzzy Systems, 2002.
- Gentle, J. E., Random Number Generation and Monte Carlo Methods, Wiley, Springer, 2nd edition, 2003.
- Grabe, M., Measurement Uncertainties in Science and Technology, Springer, Berlin, New York, 2005.
- Hennes, M., Konkurrierende Genauigkeitsmasse: Potential und Schwächen aus der Sicht des Anwenders, *AVN* 4 (2007), 136–146.
- ISO, Guide to the expression of uncertainty in measurements (GUM), International Organization for Standardization, Geneva, 1995.
- ISO, Evaluation of measurement data – Supplement 1 to the “Guide to the expression of uncertainty in measurement” – Propagation of distributions using a Monte Carlo method, Joint Committee for Guides in Metrology, Bureau International des Poids et Mesures, Geneva, 2007.
- Koch, K. R., Parameter Estimation and Hypothesis Testing in Linear Models, Springer, Berlin, Heidelberg, New York, 2nd edition, 1999.
- Koch, K. R., Introduction to Bayesian Statistics, Springer, Berlin, Heidelberg, New York, 2nd edition, 2007.
- Koch, K. R., Evaluation of uncertainties in measurements by Monte Carlo simulations with an application for laserscanning, *J. Applied Geodesy* 2 (2008a), 67–77.
- Koch, K. R., Determining uncertainties of correlated measurements by Monte Carlo simulations applied to laserscanning, *J. Applied Geodesy* 2 (2008b), 139–147.
- Kreinovich, V., Why Intervals? Why Fuzzy Numbers? Towards a new Justification, in: Proceedings of the 2007 IEEE Symposium on Foundations of Computational Intelligence (FOCI), 2007.
- Kreinovich, V., Starks, S. A., Araiza, R., Xiang, G., Velasco, A. A., Averill, M. G., and Keller, G. R., Towards Combining Probabilistic, Interval, Fuzzy Uncertainty, and Constraints: An Example Using the Inverse Problem in Geophysics, in: Proceedings of the 12th GAMM – IMACS International Symposium on Scientific Computing, Computer Arithmetic and Validated Numerics (SCAN), 2006.
- Kutterer, H., Zum Umgang mit Ungewissheit in der Geodäsie – Bausteine für eine neue Fehlertheorie, volume 553, Deutsche Geodätische Kommission, Reihe C, München, 2002.
- Kwakernaak, H., Fuzzy random variables – I: definitions and theorems, *Informations Sciences* 15 (1978), 1–29.
- Kwakernaak, H., Fuzzy random variables – II: algorithms and examples for the discrete case, *Informations Sciences* 17 (1979), 253–278.
- Mauris, G., Lasserre, V., and Foulloy, L., A fuzzy approach for the expression of uncertainty in measurement, *Measurement* 29:3 (2001), 165–177.
- McNish, A. G., The speed of light, *IRE Transaction on Instrumentation* 1:11 (1962), 138–148.
- Möller, B. and Beer, M., Fuzzy Randomness: Uncertainty in Civil Engineering and Computational Mechanics, Springer, Berlin, New York, 2004.
- Neumann, I., Modeling of an extended uncertainty budget in parameter estimation and hypothesis testing, treated exemplarily with the geodetic deformation analysis (in German), PhD thesis, Geodetic Institute, Leibniz University of Hannover, 2009, to appear.
- Neumann, I. and Kutterer, H., Geodetic Deformation Analysis with respect to Observation Imprecision, in: Proceedings of the XXIII International FIG Congress, Munich, Germany, 2006.
- Neumann, I. and Kutterer, H., The probability of type I and type II errors in imprecise hypothesis testing with an application to geodetic deformation analysis, *International Journal of Reliability and Safety*, 2009, to appear.
- Nguyen, H. T. and Kreinovich, V., Nested Intervals and Sets: Concepts, Relations to Fuzzy Sets, and Applications, in: Kearfott, R. B. and Kreinovich, V., editors, Applications of Interval Computations, Kluwer Verlag, Dordrecht, 245–290, 1996.
- Reshetyuk, Y., Investigation and calibration of pulsed time-of-flight terrestrial laser scanners, Licentiate thesis in Geodesy, Sweden, 2006.
- Schulz, T. and Ingensand, H., Influencing Variables, Precision and Accuracy of Terrestrial Laser Scanners, in: Intergeo East, Bratislava, Slovakia, 2004.
- Siebert, B. R. L. and Sommer, K.-D., Weiterentwicklung des GUM und Monte-Carlo-Techniken, *Technisches Messen* 71 (2004), 67–80.
- Stigler, S. M., Statistics and the Question of Standards, *Journal of Research of the National Institute of Standards and Technology* 101:6 (1996), 779–789.
- Strübing, T., Realization of a sensor integrated measuring system for the acquisition of deformation processes of structures, (in German), Geodetic Institute, Leibniz University of Hannover, 2007.
- Terán, P., Probabilistic foundations for measurement modelling with fuzzy random variables, *Fuzzy Sets and Systems* 158:9 (2007), 973–986.
- Viertl, R., Statistical Methods for Non-Precise DATA, CRC Press, New York, London, 1996.
- Youden, W. J., Enduring values, *Technometrics* 14:1 (1972), 1–11.
- Zadeh, L., Fuzzy sets, *Information Control* 8 (1965), 338–353.
- Zoller+Fröhlich, Technical datasheet IMAGER 5006, Zoller+Fröhlich GmbH, 2007.

Received: Jan 2, 2009

Accepted: Mar 5, 2009

Author information

Hamza Alkhatib and Hansjörg Kutterer
Geodetic Institute
Leibniz University of Hannover
Nienburger Str. 1
30167 Hannover, Germany
E-mail: alkhatib@gih.uni-hannover.de

Ingo Neumann
Institute of Geodesy—Geodetic Lab
University FAF Munich
Werner Heisenberg Weg 39
85579 Neubiberg, Germany

Estimation of Measurement Uncertainty of kinematic TLS Observation Process by means of Monte-Carlo Methods

Hamza Alkhatib^{1,*} and Hansjörg Kutterer²

¹ Geodetic Institute, Engineering Geodesy and Geodetic Data Analysis Section, Gottfried Wilhelm Leibniz Universität Hannover, Germany

² Bundesamt für Kartographie und Geodäsie, Frankfurt am Main

Abstract. In many cases, the uncertainty of output quantities may be computed by assuming that the distribution represented by the result of measurement and its associated standard uncertainty is a Gaussian. This assumption may be unjustified and the uncertainty of the output quantities determined in this way may be incorrect. One tool to deal with different distribution functions of the input parameters and the resulting mixed-distribution of the output quantities is given through the Monte Carlo techniques. The resulting empirical distribution can be used to approximate the theoretical distribution of the output quantities. All required moments of different orders can then be numerically determined. To evaluate the procedure of derivation and evaluation of output parameter uncertainties outlined in this paper, a case study of kinematic terrestrial laserscanning (k-TLS) will be discussed. This study deals with two main topics: the refined simulation of different configurations by taking different input parameters with diverse probability functions for the uncertainty model into account, and the statistical analysis of the real data in order to improve the physical observation models in case of k-TLS. The solution of both problems is essential for the highly sensitive and physically meaningful application of k-TLS techniques for monitoring of, e. g., large structures such as bridges.

Keywords. Monte Carlo simulation, kinematic terrestrial laserscanning, uncertainty modeling.

1 Introduction

The main tasks of an engineer include design, produce, and test of structures devices, and processes. These tasks

Corresponding author: Hamza Alkhatib, Geodetic Institute, Engineering Geodesy and Geodetic Data Analysis Section, Gottfried Wilhelm Leibniz Universität Hannover, Germany.
E-mail: alkhatib@gih.uni-hannover.de.

Received: September 18, 2012. Accepted: January 30, 2013.

will be involved with mathematical and physical modeling of the different phenomena. In the constructed mathematical/physical model some information about constants parameters and functional variables are needed. Uncertainties evolve into his modeling together with these information. The sources of uncertainties result from the data or the measurements, from the statistical evaluation of the model and from the physical model.

The Guide to the Expression of Uncertainty in Measurement (GUM) is the standard reference in uncertainty modeling in engineering and mathematical science, cf. [6]. GUM groups the occurring uncertain quantities into Type A and Type B. Uncertainties of Type A are determined with the classical statistical methods, while Type B is subject to other uncertainties like knowledge about an instrument. Whereas the uncertainties of Type A can be estimated based on the measurement itself, the estimated uncertainties of Type B are based on expert knowledge, e.g., the technical knowledge about an instrumental error source.

The extension of GUM [7] recommends the propagation of uncertainties using a probabilistic approach. Within the mentioned approach the propagation of uncertainties is numerically treated by Monte Carlo (MC) techniques. The difference between the GUM [6] and the extension of GUM [7] in case of nonlinearity and/or non-Gaussianity will not significantly differ in the first and the second central moments but rather in the estimate of the confidence region, which is reflected in the non-Gaussian PDF of the output quantities. The acceptance of MC techniques has significantly increased during the last decade. Consequently, it's widely used within many scientific disciplines. [5] suggested to use MC simulations instead of the treatment of the combined uncertainties by applying the law of error propagation (LOP). [15] recommended a MC based method to evaluate the measurement uncertainties in nonlinear models. [9] suggested the determination of the uncertainty according to GUM by a Bayesian confidence interval using MC simulation. The approach has been explained in detail and applied to the results of terrestrial laserscanning (TLS). Furthermore, the approach has been extended in [10] to evaluate uncertainties of correlated measurements by another application in TLS. In [11] the 2D case of kinematic TLS was studied where repeated profile scans are observed from a fixed station with a high repetition frequency for monitoring purposes. The focus was put on a refined modeling of the uncertainty of both the observations and the derived positions of the profile points. MC simula-

tion techniques were applied to provide numerical results for discussion and validation.

The paper is organized as follows: First we will describe the general idea of MC techniques to describe measurement uncertainties in the context of GUM. The application example to kinematic TLS is given and discussed in the following sections.

2 UNCERTAINTY MODELING WITH MONTE CARLO TECHNIQUES

In MC techniques, both, the random and the systematic components of the uncertainty, are treated as having a random nature. Please note that not the systematic component itself is modelled as random, it is the knowledge about the systematic component for which a probability distribution is introduced [8].

The GUM suggests in some cases to select the probability distribution function (PDF) of the input quantities as rectangular, triangular, and trapezoidal [6]. In these cases, it is hard/impossible to obtain the estimate of the uncertainty for the output quantity in a closed mathematical form. An alternative to modeling and propagating uncertainties is propagating distributions by MC simulations of the observation model from (1):

$$y = f(p_1, p_2, \dots, p_n) = f(\mathbf{p}). \quad (1)$$

Here y represents a random output quantity and $p_1, p_2, \dots, p_n$ are the n random inputs.

2.1 MC Approach to Evaluate Uncertainty

The MC techniques are of great importance for uncertainty evaluation. With a set of generated samples the distribution function for the value of the output quantity y in (1) will be numerically approximated. In general, the functional relations between the basic influence parameters, referring to (1), the observations and the parameters of interest are non-linear, and the normal distribution is not the adequate probability density function. In such case, MC simulation is a suitable way to approximately derive the stochastic properties of the quantities of interest (output quantities). It is assumed that the functional model is completely formulated relating the output quantities with the input quantities: the observations and the basic influence parameters, respectively. It is further assumed that the probability densities of the considered input quantities are a priori known. Then, a sample vector of the input quantities can be drawn repeatedly using random number generators. Random numbers are generated on a computer by means of deterministic procedures. In particular, rectangular distributed random numbers are generated, which may then in turn be transformed into random numbers of random variables having

other distributions, for instance, into numbers of a normally distributed random variable [3].

For each input sample vector the corresponding values of the output quantities are calculated by using the corresponding functional relation. The set of output sample vectors yields an empirical distribution which can be used to approximate the correct random distribution of the output quantities. All required measures (expectation values, variances and covariances) as well as higher order central moments such as skewness and kurtosis can then be derived. To sum up, MC approaches to estimate the uncertainty include the following steps:

- *Step 1:* A set of random samples, which have the size n , is generated from the PDF for each random input quantity $p_1, p_2, \dots, p_n$. The sampling procedure is repeated M times for every input quantity.
- *Step 2:* The output quantities y be will then calculated by:

$$y^{(i)} = f(p_1^{(i)}, p_2^{(i)}, \dots, p_n^{(i)}) = f(\mathbf{p}^{(i)}), \quad (2)$$

where $i = 1, \dots, M$ are the generated samples of the random input quantity p .

- *Step 3:* Particularly relevant estimates of any statistical quantities can be calculated as the expected value of the output quantity:

$$\hat{E}(f(\mathbf{p})) = \hat{E}(y) = \frac{1}{M} \sum_{i=1}^M f(\mathbf{p}^{(i)}), \quad (3)$$

the estimate of the variance of the output quantity [1]:

$$\hat{D}(y) = \frac{1}{M} \sum_{i=1}^M \left(f(\mathbf{p}^{(i)}) - \hat{E}(f(\mathbf{p})) \right)^2 \quad (4)$$

and the skewness of univariate output quantities $y^{(1)}, y^{(2)}, \dots, y^{(M)}$

$$\hat{s}k(y) = \frac{1}{(M-1)\hat{s}^3} \sum_{i=1}^M \left(f(\mathbf{p}^{(i)}) - \hat{E}(f(\mathbf{p})) \right)^3 \quad (5)$$

where $\hat{E}(f(\mathbf{p}))$ is the mean value of the output quantities (see (3)), s is its standard deviation $s = \sqrt{\hat{D}(y)}$ (refer to (5) and (6)), and M is the number of generated samples. The skewness for a normal distribution is zero, and any symmetric data should have a skewness near zero, see, e. g., [12].

For univariate output quantities the kurtosis can be

estimated by

$$\hat{k}u(y) = \frac{1}{(M-1)\hat{s}^4} \sum_{i=1}^M \left(f(\mathbf{p}^{(i)}) - \hat{E}(f(\mathbf{p})) \right)^4. \quad (6)$$

The kurtosis for a standard normal distribution is 3.

Figure 1 shows a diagram with the main steps of uncertainty modeling with a different treatment of the random and systematic uncertainties.

In [9] and [10] the above mentioned MC algorithm in case of TLS uncertainty assessment have been discussed. [2] apply it to k-TLS vertical profile scans and they combine it with a deterministic approach based on fuzzy sets. Here, only MC techniques will be considered but it is extended to the discussion of the properties of the derived time series and of their validation using real k-TLS observation data.

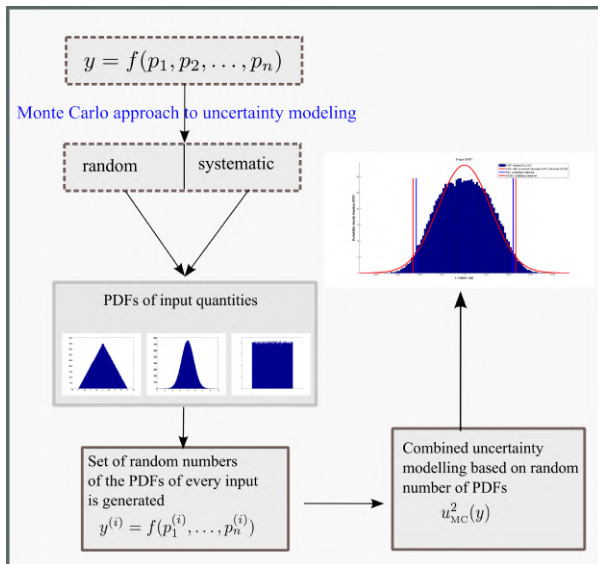


Figure 1. Treatment of uncertainty components in MC approach.

3 APPLICATION OF THE MONTE CARLO-APPROACH TO K-TLS

3.1 Object and Setup

In this section a short numerical example for the approach, presented in Section 2, is shown. The aim of the application is to detect the vertical displacements of a bridge under load, e.g., due to car traffic or train crossings. For this reason, a laserscanner of type Zoller+Fröhlich Imager 5006 scanner was placed beneath the bridge which is located in the southern part of Germany. [11] give a detailed description of the bridge, of the loading tests with different trucks, of the applied observation procedures and of the derived

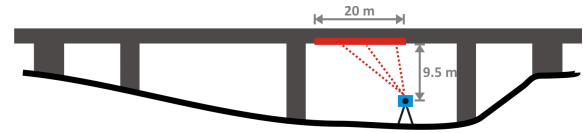


Figure 2. Bridge and scanner.

data; see Figure 3 for a graphical representation of the object and the location of the laser scanner. The horizontal section in along-track direction of the bridge (y -axis) considered here has a length of 20 m with a shortest distance between scanner and bridge of about 9.5 m. [16] show new analysis results of the k-TLS observations. Moreover, they compare these results with strain gauge observations and with numerical simulations based on finite-element models. Note that the consistency of all three kinds of data is better than 1 mm. Here, the unloaded state of the Autobahn bridge is studied which was repeatedly observed in order to get a reference geometry for the analysis of the load-induced deformations. For the observation of the vertical profiles a repetition rate of 12.5 profiles per second was used while the repetition frequency of the distance measurements was 500 kHz. For the vertical angle this yields an increment of 10 mgon. There are 7216 points per epoch within the observed section; 500 profiles representing the unloaded state were considered in total.

The profile data were processed according to the procedure outlined in Section 2. In order to prepare a meaningful model for the MC simulations the observed and processed data were roughly checked regarding the level of variance and with respect to possible quantities of influence. Within this analysis also an effect in the original data was detected which seems to be due to a different temporal resolution of distance and vertical angle observation; it is indicated in Figure 3. This effect has been modeled accordingly for the simulations.

3.2 SIMULATION OF K-TLS PROFILES

The functional model, which was used in [2], has been established for the simulations. The time series of the vertical height z of every point of the bridge can be expressed in the local coordinate system of the laserscanner by the following equation:

$$z = d \cdot \cos(\zeta), \quad \zeta = \zeta_0 + \Delta\zeta \quad (7)$$

with d the observed distance between laser scanner and object point which induces a constant and a distance-proportional effect, ζ_0 the observed zenith angle with a constant angular effect, and $\Delta\zeta$ the discretization term which is induced by the angular increment of the vertical servomotor. In this study seven uncertainty components were modeled:

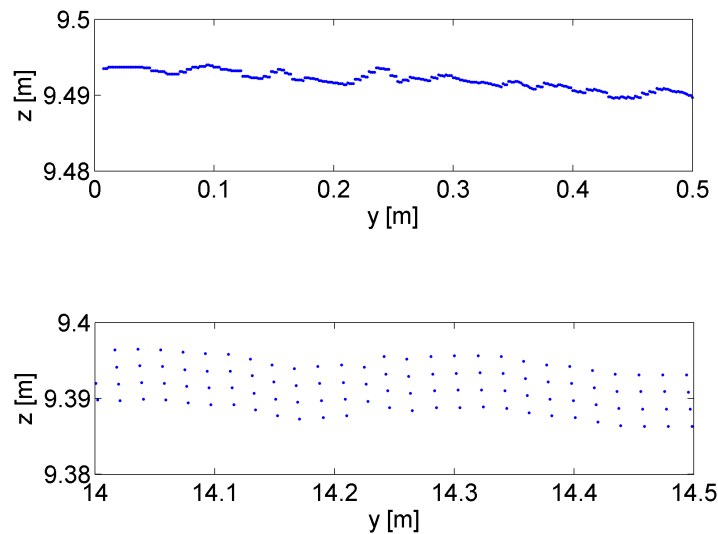


Figure 3. Two zooms into a representative profile scan – y -coordinate (along bridge) and z -coordinate (height with inflated scale); upper figure: section directly above the laser scanner in a spatial distance of about 9.5 m with orthogonal angle of incidence; lower figure: section in a spatial distance of about 17 m with oblique angle of incidence.

- Uncertainty of the distance (p_1 , Type A), and their additional constant (p_2 , Type B),
- distance depending term for the uncertainty of the distance measurement (p_3 , Type B, see [9]),
- incidence angle of the measured distance under the bridge (p_4 , Type B),
- uncertainty of the zenith angle (p_5 , Type A) and the vertical index error (p_6 , Type B),
- vertical resolution for the zenith angle (the step width of the motor) (p_7 , Type B).

The uncertainties and the PDF for the input quantities $p_i, i \in \{1, \dots, 7\}$ are given in Table 1.

The symbols μ and σ^2 in Table 1 denote the expectation value and the variance of the random variable, respectively; the uniform and the triangular distribution are defined by the lower bound p_{il} and the upper bound p_{iu} of the interval with positive values of the density function. The assumptions for the uncertainties of p_1 , p_5 and p_6 are based on the technical data from the manufacturer and for the uncertainties of p_2 , p_3 and p_4 on [14] and for p_7 on [13].

In the following, the results of three different MC simulation runs are shown and discussed which were calculated for the bridge section described in Section 3.1. In all simulations 100,000 samples were drawn for each random quantity; the obtained values were processed according to the model described in equation (1). 100,000 MC runs are generated as necessary to obtain the estimates of the statistical moments of the PDF with at least three significant digits.

The same results were numerically proven in [9]. Three different class widths were selected for the simulations: one / five / ten observation values per class and epoch. As representative value for each class and epoch the respective arithmetic mean of the single class values was used; this is reasonable because of the yet small class widths. Thus, only a minor spatial filtering was applied but not a temporal filter. The temporal sequences of these representative class values define the time series or data series, respectively, which are analyzed further.

Due to the unloaded state of the bridge all profiles can be considered as stationary. Therefore three central moments of the underlying probability density functions are determined empirically: standard deviation (of the single value), skewness and kurtosis. Note that expectation value and standard deviation are necessary and sufficient in order to uniquely define a normal distribution. The skewness of a normally distributed random variable equals 0, and the kurtosis equals 3 (In order to refer the kurtosis of an arbitrary density to the normal distribution the value 3 can be subtracted; then the kurtosis of the normal distribution equals 0). Hence, skewness and kurtosis are well-suited to detect violations of the normal distribution assumption.

Simulation I: For this simulation, the three input quantities (p_1 , p_3 and p_5) were considered for uncertainty modeling: the constant and distance-proportional effect of the distance observation, and the constant angular effect of the zenith angle observation. The input quantities for Simulation I are defined in the left three columns of Table 2. The

p_i	Er. Com.	PDF	PDF Type
p_1	random	$p_1 \sim N(\mu_{p_1}, \sigma_{p_1}^2)$	normal
p_2	systematic	$p_2 \sim T(p_{2l}, p_{2u})$	triangular
p_3	random	$p_3 \sim N(\mu_{p_3}, \sigma_{p_3}^2)$	normal
p_4	random	$p_4 \sim N(\mu_{p_4}, \sigma_{p_4}^2)$	normal
p_5	random	$p_5 \sim N(\mu_{p_5}, \sigma_{p_5}^2)$	normal
p_6	systematic	$p_6 \sim T(p_{6l}, p_{6u})$	triangular
p_7	systematic	$p_7 \sim U(p_{7l}, p_{7u})$	uniform

Table 1. Uncertainties for the input quantities.

three central moments of the empirical distributions of the respective representative class values obtained as results of the Simulation I are presented in Figure 4.

Simulation II: For the second simulation the same input quantities were used as in the first simulation; in addition, the uncertainty induced by the angular increment of the vertical servo-motor (p_7) was modeled. The three central moments of the empirical distributions derived as results of the Simulation I are presented in Figure 5.

Simulation III: For the last simulation all input quantities described in Table 1 were used; the result is shown in Figure 6.

Looking at the standard deviations shown in Figure 4 and Figure 5, the distance-proportional effect on the standard deviations of the representative profile points is obvious. Moreover, the square-root law $\sigma_{\bar{x}}/\sqrt{n}$ for the standard deviation of the mean value $\bar{x}$ with respect to the standard deviation of the single values by the number n of sample values can clearly be seen. In addition, the skewness is insignificant in both simulations. The difference lies in the kurtosis. Whereas in Figure 4 the normal distribution assumption seems to hold, it is clearly violated in Figure 5. The assumption of a Gaussian distribution in Figure 6 is not obvious. Therefore, the rigorous mathematical assessment of the discussion about this assumption has to be referred to a suitable hypothesis test. For this purpose the Kolmogorov-Smirnov-Test (KS-Test) is used. The KS-Test is a form of minimum distance estimation used to compare a data set with a reference probability distribution. The test quantifies a distance between the empirical distribution function of the data set and the cumulative distribution function of the reference distribution. By modifying the KS-Test it can serve as a goodness of fit test, see, e. g. [4]. In the case of testing for normality of the distribution, the samples are standardized and compared with a standard normal distribution. As a result of the performed hypothesis, we were able to approve that only in Simulation I the normal distribution holds.

Due to the convolution different probability distributions (normal, triangular and uniform, respectively) the resulting

distributions of Simulation II and III are not normal distribution, this can be validated by means of the KS-Test. Moreover (especially for Simulation II), the kurtosis values decrease from 3 (which is valid for observations directly in vertical direction and which does not contradict to the normal distribution assumption) to about 2 in a horizontal distance of about 20 m. There are two effects which superpose each other: one from the uniform distribution and the other from the (non-linear) cosine function. In case of increasing the class width, the effect on the kurtosis is significantly mitigated possibly due to the central limit theorem of probability theory.

4 VALIDATION OF THE SIMULATION RESULTS

In order to assess and to validate the simulation results, actually observed profile data were processed and analyzed as well in full accordance with the procedure applied for the two simulation runs. Figure 7 shows the obtained results; like in Section 3 the standard deviations, the skewness and the kurtosis of the individual classes of height coordinates are given. The standard deviations show again a clear dependence on the horizontal distance between the scanner and the profile points; this dependence is reduced when the class width is increased. However, in contrast to the simulated data, the mentioned square-root law does not fully apply – neither for small values of the y -coordinate nor for large values. For small values the reduction of the variance induced by averaging is smaller than expected, for large value the reduction effect is larger than expected. Like in the simulations, the skewness of the empirical distributions of the individual classes does not significantly differ from 0; note that the visible variability of the values decreases when y increases. Hence, the empirical distributions are symmetric – independent of the class width. However, the decrease of the kurtosis with respect to increasing values of y is remarkable. On the one hand, there is a systematic and significant decrease of the values from 3 (what is expected in case of normal distribution) to a value slightly below 2. This indicates clearly the violation of the normal distribution assumption. On the other hand however, this effect is

Simulation I: without vertical increment		
Input quantity	PDF	Num. value (std. dev)
p_1	Normal	0.5 mm
p_3	Normal	30 ppm
p_5	Normal	10 mgon
Simulation II: with vertical increment		
Input quantity	PDF	Num. value (std. dev)
p_1	Normal	0.3 mm
p_3	Normal	30 ppm
p_5	Normal	5 mgon
p_7	Uniform	20 mgon
Simulation III: with all input quantities		
Input quantity	PDF	Num. value (std. dev)
p_1	Normal	0.5 mm
p_2	Triangular	0.4 mm
p_3	Normal	30 ppm
p_4	Normal	1 mm
p_5	Normal	10 mgon
p_6	Triangular	8 mgon
p_7	Uniform	10 mgon

Table 2. MC simulation: input quantities for the uncertainty models (type of probability densities and numerical values of the standard deviations).

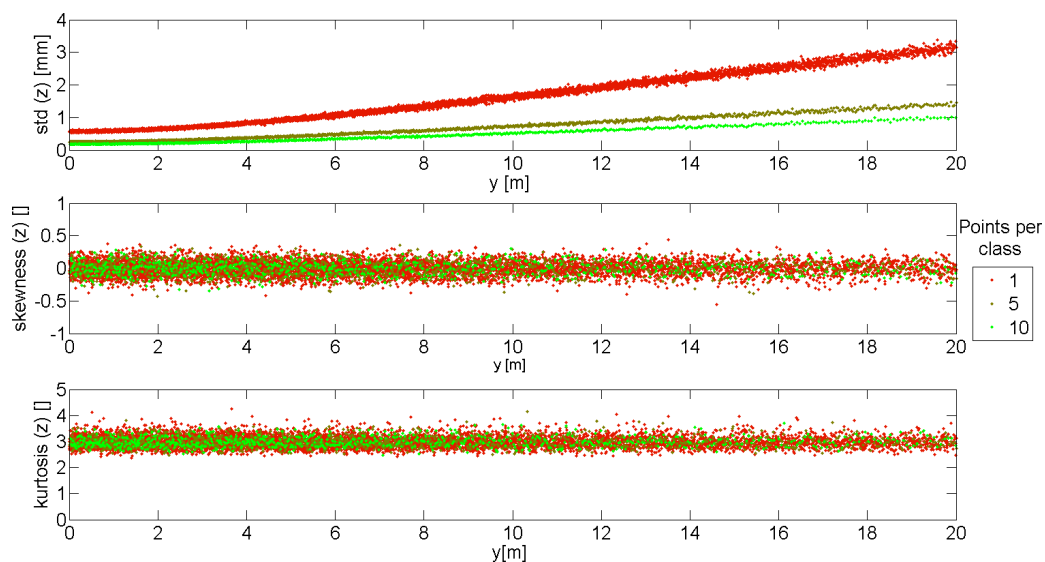


Figure 4. Simulation I – without vertical motor increment uncertainty: analysis of the simulated k-TLS profiles for three different class widths: standard deviations, skewness, and kurtosis.

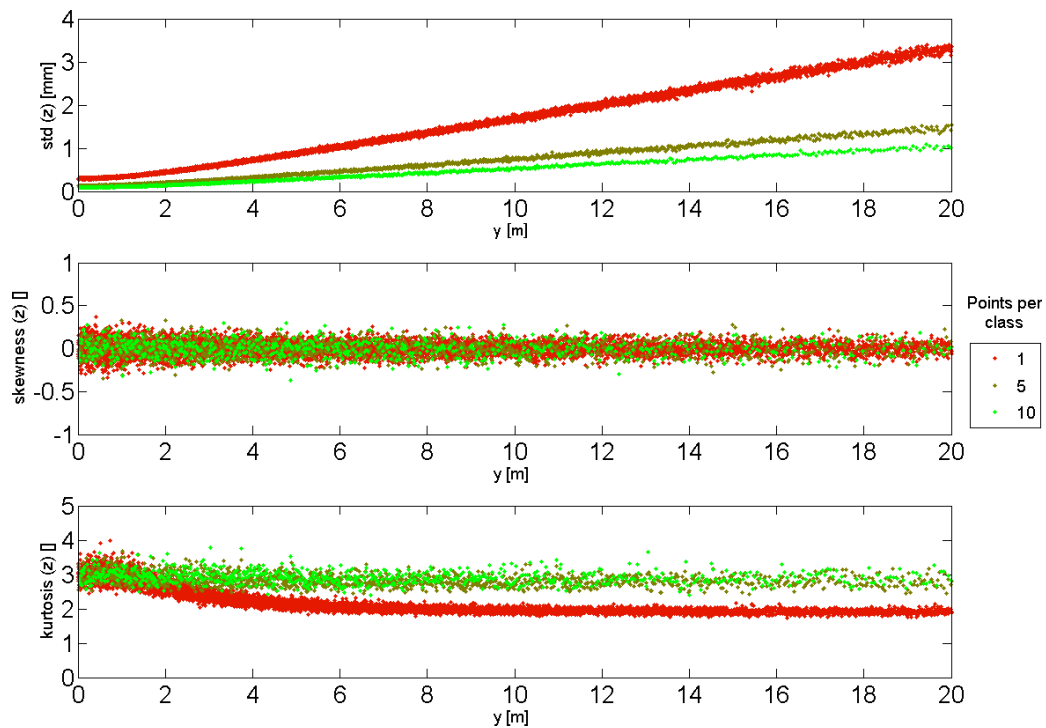


Figure 5. Simulation II – with vertical motor increment uncertainty: analysis of the simulated k-TLS profiles for three different class widths: standard deviations, skewness, and kurtosis.

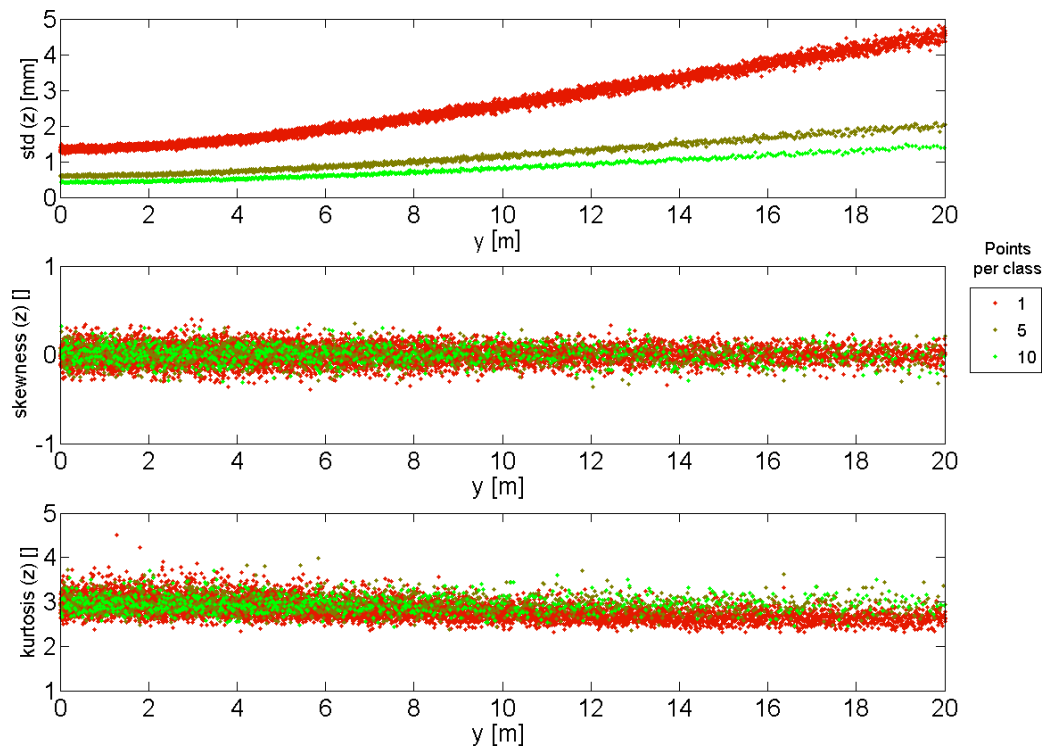


Figure 6. Simulation III - with all input quantities listed in Tab. 1: analysis of the simulated k-TLS profiles for three different class widths: standard deviations, skewness, and kurtosis.

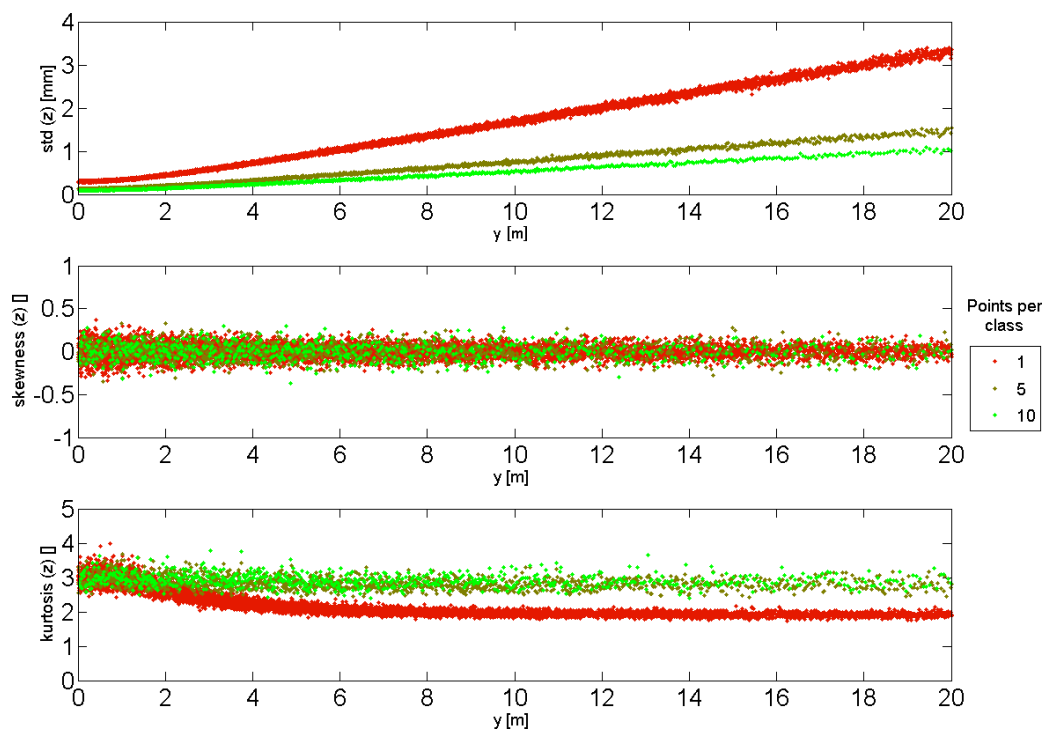


Figure 7. Real data: analysis of the observed k-TLS profiles for three different class widths: standard deviations, skewness, and kurtosis.

mitigated in case of wider classes. Both effects were also obtained in Simulation II shown in Figure 5 by modeling of a uniformly distributed uncertainty component for the angular increment of the vertical servo-motor. Note that the visible variability of the values decreases when y increases.

Obviously, the real-data results fit quite well to the results of Simulation II which could be obtained using a rather basic uncertainty model with a few input parameters only. In addition to the simulations there are some further effects in the real data which could not be modeled up to now. Looking, e. g., at the subfigures of Figure 7 in total, some regions of horizontal distances y can be identified where the values of the central moments are obviously disturbed. This holds in particular for the standard deviations like, e. g., between 16 m and 17 m; there are also some periodic characteristics. A following study is required which aims at a refined statistical modeling and analysis of the k-TLS profile time series.

5 CONCLUSIONS

In this paper the 2D case of kinematic TLS was studied where repeated profile scans are observed from a fixed station with a high repetition frequency for monitoring purposes. The focus was put on a refined modeling of the uncertainty of both the observations and the derived positions of the profile points. In order to take into account the

complete data processing chain, the strategy for generating and analyzing time series was considered. MC simulation techniques were applied to provide numerical results for discussion and validation. It turned out that a rather small number of input parameters for the uncertainty model are required to obtain simulation results which fit quite well to actually observed data. These real data were observed on the occasion of loading tests at an Autobahn bridge in southern Germany.

Further work has to address two main topics: the more refined simulation of more complex configurations by taking more parameters for the uncertainty model into account, and the rigorous and thorough statistical analysis of the real data in order to improve the physical observation models in case of k-TLS. The solution of both problems is essential for the highly sensitive and physically meaningful application of k-TLS techniques for monitoring of, e. g., large structures such as bridges.

References

- [1] Alkhatib H., *On Monte Carlo methods with applications to the current satellite gravity missions*, PhD thesis, Institute for Geodesy and Geoinformation of University of Bonn, 2007.
- [2] Alkhatib H., Neumann I. and Kutterer H., Uncertainty modeling of random and systematic errors by means of Monte Carlo and fuzzy techniques, *Journal of Applied Geodesy* **3** (2009), 67–81.

- [3] Gentle J. E., *Random Number Generation and Monte Carlo Methods*, Wiley, Springer, 2nd edition, 2003
- [4] Gibbons J. D. and Chakraborti S., *Nonparametric statistical inference*, volume 198 of Statistics, Chapman & Hall/CRC Press, Boca Raton and Fla, 5th edition, 2011.
- [5] Hennes M., Konkurrierende Genauigkeitsmasse: Potential und Schwächen aus der Sicht des Anwenders, *AVN* **4** (2007), 136–146.
- [6] ISO, Guide to the expression of uncertainty in measurements (GUM), International Organization for Standardization, Geneva, 1995.
- [7] ISO, Evaluation of measurement data - Supplement 1 to the Guide to the expression of uncertainty in measurement - Propagation of distributions using a Monte Carlo method., Joint Committee for Guides in Metrology, Bureau International des Poids et Mesures, Geneva, 2007.
- [8] Koch K. R., *Introduction to Bayesian Statistics*, Springer, Berlin - Heidelberg - New York, 2nd edition, 2007.
- [9] Koch K. R., Evaluation of uncertainties in measurements by Monte Carlo simulations with an application for laserscanning, *J. Applied Geodesy* **2** (2008), 67–77.
- [10] Koch K. R., Determining uncertainties of correlated measurements by Monte Carlo simulations applied to laserscanning, *J. Applied Geodesy* **2** (2008), 139–147.
- [11] Kutterer H., Alkhatib H., Paffenholz J.-A. and Vennegeerts H., Monte-Carlo simulation of profile scans from kinematic TLS, in: Proceedings of the FIG Congress 2010, <http://www.g.net/pub/?g2010/papers/ts03d%5Cts03dkuttereralkhatibetal3795.pdf>.
- [12] Martinez W. L. and Martinez A. R., *Computational statistics handbook with MATLAB*, volume 8 of Computer science and data analysis series, Chapman & Hall/CRC, Boca Raton, 2nd edition, 2008.
- [13] Reshetyuk Y., *Investigation and calibration of pulsed time-of-flight terrestrial laser scanners*, Licentiate thesis in Geodesy, Sweden, 2006.
- [14] Schulz T. and Ingensand H., *Influencing Variables, Precision and Accuracy of Terrestrial Laser Scanners*, in: Intergeo East, Bratislava, Slovakia, 2004.
- [15] Siebert B. R. L. and Sommer K.-D., Weiterentwicklung des GUM und Monte-Carlo-Techniken, *Technisches Messen* **71** (2004), 67–80.
- [16] Vennegeerts H., Liebig J. P., Hansen M., Neuner H., P., J.-A., Grünberg J., and Kutterer H., *Monitoring eines Brückentragwerks Vergleichende Messungen mit einem terrestrischen Laserscanner und Sensoren der Baumesstechnik*, in: Wunderlich T., editor, Ingenieur-Vermessung 2010, Wichmann, Heidelberg, 2010.

Further Results on a Robust Multivariate Time Series Analysis in Nonlinear Models with Autoregressive and t-Distributed Errors



Hamza Alkhatib , Boris Kargoll and Jens-André Paffenholtz

Abstract We investigate a time series model which can generally be explained as the additive combination of a multivariate, nonlinear regression model with multiple univariate, covariance stationary autoregressive (AR) processes whose white noise components obey independent scaled t-distributions. These distributions enable the stochastic modeling of heavy tails or outlier-afflicted observations and present the framework for a partially adaptive, robust maximum likelihood (ML) estimation of the deterministic model parameters, of the AR coefficients, of the scale parameters, and of the degrees of freedom of the underlying t-distributions. To carry out the ML estimation, we derive a generalized expectation maximization (GEM) algorithm, which takes the form of linearized, iteratively reweighted least squares. In order to derive a quality assessment of the resulting estimates, we extend this GEM algorithm by a Monte Carlo based bootstrap algorithm that enables the computation of the covariance matrix with respect to all estimated parameters. We apply the extended GEM algorithm to a multivariate global navigation satellite system (GNSS) time series, which is approximated by a three-dimensional circle while taking into account the colored measurement noise and partially heavy-tailed white noise components. The precision of the circle model fitted by the GEM algorithm is superior to that of the previous standard estimation approach.

H. Alkhatib (✉) · B. Kargoll · J.-A. Paffenholtz
Geodätisches Institut, Leibniz Universität Hannover, Nienburger Str. 1,
30167 Hannover, Germany
e-mail: alkhatib@gih.uni-hannover.de
URL: <https://www.gih.uni-hannover.de>

B. Kargoll
e-mail: kargoll@gih.uni-hannover.de

J.-A. Paffenholtz
e-mail: paffenholtz@gih.uni-hannover.de

Keywords Multivariate time series · Nonlinear regression model · AR process
Scaled t-distribution · Partially adaptive estimation · Robust parameter estimation
GEM algorithm · Bootstrapping · GNSS time series

1 Introduction

The estimation of the parameters of a measurement process is an important part of modeling and monitoring in many applications. Unfortunately, the measured data may contain multiple outliers. In such a case, the probability density function (pdf) of the random deviations is presumed to be heavy-tailed. Robust parameter estimation is then used to minimize the influence of outliers and to avoid the removal of measurements. Robust approaches utilize estimators that are less affected by outliers than the least-squares (LS) estimator. Maximum likelihood (ML) type estimators (M-estimators) have been developed to be robust to outliers, e.g., the Huber M-estimator [1]. These estimators basically take advantage of using reduced weights for outliers. In addition, (Student's) t regression models were introduced, which assume scaled t-distributed errors in robust ML estimation (cf. [2]). The computation of corresponding ML estimates can be easily realized by the iterative reweighted least-square (IRLS) approach, as shown in [3]. The variances of the random deviations are rescaled by means of the resulting weights conforming to their locations under the density function. This procedure, a so-called (partially) adaptive estimator, allows for estimating the unknown degree of freedom (d.o.f.) of the t-distribution jointly with the regression parameters and the scale parameter.

Multivariate multiple regression is a technique that estimates more than one (possibly nonlinear) regression model with more than one outcome variable (deterministic regression function) and with random deviations that are generally assumed to obey the normal distribution. Alternatively, [4] presumed the multivariate t-distribution with unknown scale matrix and unknown d.o.f., and demonstrated different expectation maximization (EM) algorithms to estimate the unknown model parameters jointly with the scale matrix and the d.o.f. efficiently. In previous contributions, [5, 6] demonstrated that the expectation *conditional* maximization (ECM) and the ECM *either* (ECME) algorithms accelerate the convergence of the EM algorithm noticeably. To deal with models that do not allow for closed-form solutions by EM, the optimization principle of *generalized* expectation maximization (GEM) was suggested by [3]. This approach approximates the maximum inside of every EM step instead of reaching the maximum completely. GEM algorithms applying Newton–Raphson steps have been used commonly (see e.g. [7]). To handle nonlinear functional models, [2, 8] presented an IRLS algorithm for GEM with Gauss–Newton steps.

Many multivariate time series are expected to be correlated over time, which makes (partially adaptive) parameter estimation more challenging. Autocorrelations appear in many types of sensor data. For instance, inertial sensor data, satellite gravity gradiometry data, and global navigation satellite system (GNSS) data have random deviations that show colored noise characteristics (see, e.g., [9–12]). Such sensors can also produce multiple outliers, so that robust parameter estimation is preferable over the classical LS approach. In order to take the autocorrelatedness of random deviations in time series into account, [13] extended the aforementioned partially adaptive estimator for linear regression models based on the scaled t-distribution. They assumed autoregressive (AR) random deviations in a univariate time series, with the white noise components of the AR process following independently and identically a scaled t-distribution.

We demonstrated in [14] the extension of the preceding univariate, linear model to a multivariate and nonlinear (differentiable) regression model. Regarding the setup of the AR model, we simplified the complex general case of stochastically dependent time series to the case where each time series component is associated with a univariate AR process of individual order, independently of the AR processes of the other components. Thus, cross-correlations between the different colored noise processes are currently not considered. In this paper, we add two important issues that have not been addressed in [14]. First, we replace the algorithm regarding the estimation of the d.o.f. of the underlying t-distribution by an interval-based search approach, which generally is more reliable than the previously used conventional search algorithm.

Second, we extend the previous GEM algorithm by a Monte Carlo (MC) bootstrap algorithm (in the sense of [15, 16]) for computing the covariance matrix of all unknown model parameters. In the context of geodetic regression models with autocorrelated measurements, [16] employed for this purpose MC strategies, as extensions of an iterative preconditioned conjugate gradients multiple adjustment (PCGMA) algorithm, to estimate an accurate covariance matrix for the model parameters (including optimal weighting and regularization parameters). Bootstrap models are also well suited to time series analysis with combined regression and AR models [17]. The method used in the current paper is based on an MC bootstrap approach to covariance matrix estimation in the context of EM algorithms for missing data models, as explained in [7].

The remainder of the paper is structured as follows. In Sect. 2, we describe the general time series model. In Sect. 3, we derive a GEM algorithm for parameter estimation and provide a bootstrap algorithm concerning the associated covariance matrix. In Sect. 4, we analyze three-dimensional (3D) time series of GNSS observations, where the regression model is given by a 3D circle.

2 The Observation Model

The time series model we consider in the following is

$$Y_{k,t} = h_{k,t}(\xi_1, \dots, \xi_m) + E_{k,t}, \quad (1)$$

$$E_{k,t} = \alpha_{k,1} E_{k,t-1} + \dots + \alpha_{k,p_k} E_{k,t-p_k} + U_{k,t}, \quad (2)$$

$$U_{k,t} \stackrel{\text{ind.}}{\sim} t_{v_k}(0, \sigma_k^2), \quad (3)$$

where $k \in \{1, \dots, N\}$ is a sensor component and $t \in \{1, \dots, n\}$ a time instance. In the first equation, $Y_{k,t}$ is an observable or measurand described by a possibly nonlinear function $h_{k,t}(\xi_1, \dots, \xi_m)$ of unknown parameters $\boldsymbol{\xi} = [\xi_1 \dots \xi_m]^T$ and a random deviation or measurement error $E_{k,t}$. Thus, the random vectors $\mathbf{Y}_t = [Y_{1,t} \dots Y_{N,t}]^T$ form a multivariate time series. For each sensor component, the random deviations are modeled by a covariance stationary AR(p_k) process (2) with component-dependent, given order p_k and unknown coefficients $\boldsymbol{\alpha}_k = [\alpha_{k,1} \dots \alpha_{k,p_k}]^T$. By employing different AR processes, we allow for different forms of colored measurement noise in the various components, while assuming cross-correlations between errors of different sensor components to be negligible. Furthermore, according to (3), the error variables $U_{k,1}, \dots, U_{k,n}$ of each component independently and identically follow a t-distribution with component-dependent, unknown scale factor σ_k^2 and unknown d.o.f. v_k . Thus, we allow these white noise components to have different (unknown) levels of variability and individual (unknown) tail or outlier characteristics.

For brevity of expressions, we treat the regression part as the vector-valued functions $\mathbf{h}_t(\boldsymbol{\xi}) = [h_{1,t}(\boldsymbol{\xi}) \dots h_{N,t}(\boldsymbol{\xi})]^T$. Let us also stack all of the unknown model parameters $\boldsymbol{\xi}, \boldsymbol{\alpha}_1, \dots, \boldsymbol{\alpha}_N, \sigma_1^2, \dots, \sigma_N^2, v_1, \dots, v_N$ in the single vector $\boldsymbol{\theta}$. In addition, we write the colored and white noise components as the multivariate time series $\mathbf{E}_t = [E_{1,t} \dots E_{N,t}]^T$ and $\mathbf{U}_t = [U_{1,t} \dots U_{N,t}]^T$, respectively. The lower case $\mathbf{y}_t, \mathbf{e}_t$, and $\mathbf{u}_t$ constitute real-valued vectors as certain numerical realizations of the random vectors $\mathbf{Y}_t, \mathbf{E}_t$ and $\mathbf{U}_t$, respectively.

To estimate the parameters $\boldsymbol{\theta}$, we construct a likelihood function given each white noise path $\mathbf{u}_k$. Using the definition of the scaled t-distribution (cf. [2]) and the independence assumption in (3), we can write the joint pdf of each white noise series in the factorized form

$$f(\mathbf{u}_k) = \prod_{t=1}^n \frac{\Gamma(\frac{v_k+1}{2})}{\sqrt{v_k \pi \sigma_k^2} \Gamma(\frac{v_k}{2})} \left[1 + \left(\frac{u_{k,t}}{\sigma_k} \right)^2 / v_k \right]^{-\frac{v_k+1}{2}}, \quad (4)$$

where Γ is the gamma function. Furthermore, due to the lack of cross-correlations, the product of these pdfs yields the joint pdf $f(\mathbf{u}) = f(\mathbf{u}_1) \dots f(\mathbf{u}_N)$. The idea is now to introduce the functional model (1) and the AR processes (2) into that pdf. Assuming the AR processes to be invertible, we can write (2) also as

$$u_{k,t} = e_{k,t} - \alpha_{k,1}e_{k,t-1} - \cdots - \alpha_{k,p_k}e_{k,t-p_k}. \quad (5)$$

Introducing the lag operator notation $L^j e_t = e_{t-j}$ and the lag polynomial $\alpha_k(L) = 1 - \alpha_{k,1}L - \cdots - \alpha_{k,p_k}L^{p_k}$, we can abbreviate the right-hand side of (5) by $\alpha_k(L)e_{k,t}$. Here, we may view $\alpha_k(L)$ as a *decorrelation filter* since it transforms the colored noise series $e_{k,1}, \dots, e_{k,n}$ into the white noise series $u_{k,1}, \dots, u_{k,n}$. Let us now “invert” also the functional model (1) and substitute the resulting error equations $E_{k,t} = Y_{k,t} - h_{k,t}(\xi_1, \dots, \xi_m)$ into (5). Then, the natural logarithm of the (factorized) pdf $f(\mathbf{u}_1) \dots f(\mathbf{u}_N)$ in (4) takes the form

$$\begin{aligned} \log f(\mathbf{u}) = & \sum_{k=1}^N \left(n \log \left[\frac{\Gamma(\frac{\nu_k+1}{2})}{\sqrt{\nu_k \pi \sigma_k^2} \Gamma(\frac{\nu_k}{2})} \right] \right. \\ & \left. - \frac{\nu_k+1}{2} \sum_{t=1}^n \log \left[1 + \left(\frac{\alpha_k(L)(y_{k,t} - h_{k,t}(\xi))}{\sigma_k} \right)^2 / \nu_k \right] \right), \end{aligned} \quad (6)$$

which (as a function of all unknown model parameters and observations) we define to be the log-likelihood function $\log \mathcal{L}(\theta; \mathbf{y})$. As the inverted AR process (5) involves time instances $t = 0, -1, \dots$ at which no data are available, we make the assumption that the associated observations and noise realizations all take the value 0. When fixing the “initial conditions” in this standard manner, $\log \mathcal{L}$ is sometimes referred to as a *conditional log-likelihood function*.

This function, however, will only be used directly for the ML estimation of the degrees of freedom ν_k as the likelihood equations for all other parameters are too complicated and too cumbersome to solve. We can obtain closed-form expressions for the parameters ξ , α_k , and σ_k^2 if we replace the distributional assumption (3) by the equivalent model (cf. [2])

$$U_{k,t} \stackrel{\text{ind.}}{\sim} N(0, \sigma_k^2/w_{k,t}) \mid w_{k,t}, \quad (7)$$

$$W_{k,t} \stackrel{\text{ind.}}{\sim} \frac{\chi_{\nu_k}^2}{\nu_k}. \quad (8)$$

Here, the random variables $W_{k,t}$ are unobservable data in the form of *latent variables*, which later take the role of observation weights in an IRLS algorithm. Small weights are associated with errors in the tails (i.e., with outliers) and lead to an increase of the variance through its rescaling $\sigma_k^2/w_{k,t}$. For this reason, the outlier model (7)–(8)

is sometimes called a *variance-inflation model*. The two independence assumptions in that model allow for factorization of the joint pdf of the white noise and latent variables, so that we obtain for the logarithm

$$\begin{aligned} \log f(\mathbf{u}, \mathbf{w}) = & \text{const.} - \frac{n}{2} \sum_{k=1}^N \log(\sigma_k^2) + \frac{n}{2} \sum_{k=1}^N v_k \log\left(\frac{v_k}{2}\right) - n \sum_{k=1}^N \log \Gamma\left(\frac{v_k}{2}\right) \\ & - \sum_{k=1}^N \sum_{t=1}^n \frac{1}{2} \left[v_k + \left(\frac{\alpha_k(L)(y_{k,t} - h_k(\xi))}{\sigma_k} \right)^2 \right] w_{k,t} + \sum_{k=1}^N \sum_{t=1}^n \frac{1}{2} (v_k - 1) \log w_{k,t} \end{aligned} \quad (9)$$

(see [14] for details). This function serves in the sequel as the proxy log-likelihood function $\log \mathcal{L}(\boldsymbol{\theta}; \mathbf{y}, \mathbf{w})$, instead of $\log \mathcal{L}(\boldsymbol{\theta}; \mathbf{y})$. The idea of replacing or imputing the “missing data” $w_{k,t}$ by conditional expectations based on the stochastic model for $W_{k,t}$ leads us to the following GEM algorithm.

3 Generalized EM Algorithm

EM Algorithm for Estimating Unknown Parameters

In [14], we developed a GEM algorithm, which consists of an E- and an M-step. The Q-function needed in the E-step is defined as the conditional expectation of the foregoing log-likelihood function, given measurement results $\mathbf{y}$ and parameter values in the i th iteration step $\boldsymbol{\theta}^{(i)}$, that is,

$$Q(\boldsymbol{\theta} | \boldsymbol{\theta}^{(i)}) = E_{\mathbf{W} | \mathbf{y}; \boldsymbol{\theta}^{(i)}} \{ \log \mathcal{L}(\boldsymbol{\theta}; \mathbf{y}, \mathbf{W}) \}. \quad (10)$$

In order to carry out the E-step (required for the determination of the weights within IRLS), initial parameter values are needed. In addition, we choose within the first iteration step equal weights $w_{k,t}^{(0)} = 1$ for each component of the multivariate time series.

To carry out the M-step, we maximize the Q-function given in (10) by determining the first partial derivatives of the Q-function with regard to the unknown parameters ξ , α_k , σ_k^2 , and v_k grouped in $\boldsymbol{\theta}$ and subsequently setting these derivatives equal to zero. The current iteration step is denoted by $(i + 1)$ and leads to the parameter solution $\boldsymbol{\theta}^{(i+1)}$, which replaces the solution $\boldsymbol{\theta}^{(i)}$ of the previous iteration step. As

the determination of the derivatives with respect to the parameters ξ involves the linearization of the functions $h_{k,t}$, the maximum is approximated by taking a Gauss–Newton step (of length γ). Thus, we have a GEM algorithm, which is summarized in Algorithm 1. Important steps are tagged with a circle around a letter–number combination. These steps are described in detail in the following lines:

- N.1** This update is added entirely or partially to the trial solution (in the sense of a Gauss–Newton step with step size $\gamma \in (0, 1]$). In case the Q-function is decreased by the current step length (which is initialized by $\gamma = 1$ in each iteration step), the step length is halved and the Q-function evaluated again at the current estimates. The step length is reduced until the Q-function increases and thus approaches the maximum, as required by GEM.
- N.2** Since we aim for covariance stationary and invertible AR processes, it is necessary to determine whether all roots of $\alpha_k^{(i+1)}(z) = 0$ are located within the unit circle. In case this is not true, we stabilize the preceding polynomial by mirroring all roots with magnitude exceeding 1 into the unit circle (cf. [18]), using MATLAB’s `polystab` routine.
- N.3** The estimates $v_1^{(i+1)}, \dots, v_N^{(i+1)}$ constitute the zeros of these equations, which are to be found numerically. We use a reliable zero search based on the one-dimensional interval Newton method described in [19].

Bootstrapping Algorithm for Determining the Covariance Matrix

We have presented in section “[EM Algorithm for Estimating Unknown Parameters](#)” a GEM algorithm, which allows for the estimation of the deterministic model parameters, of the AR coefficients, of the scale parameters, and of the d.o.f.s of the underlying t-distributions. Oftentimes, not only the estimated parameter is of interest, but also their (variance-)covariance matrix. This covariance matrix, which we denote by $\Sigma\{\hat{\theta}\}$, characterizes the quality of the estimated parameters. As the estimator $\hat{\theta}$ is a complex nonlinear function of random deviations, propagation of variance–covariance information cannot be carried out directly. Bootstrapping, in combination with MC techniques, can however be applied in this situation without any difficulty. The calculation process to derive the covariance matrix of the estimated parameters is summarized in Algorithm 2.

Algorithm 1: GEM Algorithm

Input : $y_{k,t}, h_{k,t}(\xi), p_k, \text{itermax}, \varepsilon, \varepsilon_v$ ($k = 1, \dots, N; t = 1, \dots, n$)

Output: $\hat{\xi}, \hat{\sigma}_k^2, \hat{v}_k, \hat{e}_k, \hat{u}_k, \hat{\mathbf{W}}_k, \hat{\alpha}_k$

Initialization: $\xi^{(0)}; v_k^{(0)}; \sigma_k^{(0)}; \alpha_k^{(0)}$

for $i = 0 \dots \text{itermax}$ **do**

E-Step:

$$w_{k,t}^{(i)} = \frac{v_k^{(i)} + 1}{v_k^{(i)} + \left(\frac{\alpha_k^{(i)}(L)(y_{k,t} - h_{k,t}(\xi^{(i)}))}{\sigma_k^{(i)}} \right)^2} \rightsquigarrow \mathbf{W}_k^{(i)} = \begin{bmatrix} w_{k,1}^{(i)} & 0 & \dots & 0 \\ 0 & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & w_{k,n}^{(i)} \end{bmatrix}$$

M-Step:

$$\Delta y_{k,t}^{(i)} = y_{k,t} - h_{k,t}(\xi^{(i)}), \mathbf{A}_{k,t}^{(i)} = \frac{\partial h_{k,t}(\xi^{(i)})}{\partial \xi}$$

$$\overline{\Delta y}_{k,t}^{(i)} = \alpha_k^{(i)}(L) \Delta y_{k,t}^{(i)}, \quad \overline{\mathbf{A}}_{k,t}^{(i)} = \alpha_k^{(i)}(L) \mathbf{A}_{k,t}^{(i)}$$

$$\Delta \xi^{(i+1)} = \left(\sum_{k=1}^N \frac{1}{(\sigma_k^2)^{(i)}} \overline{\mathbf{A}}_k^{(i),T} \mathbf{W}_k^{(i)} \overline{\mathbf{A}}_k^{(i)} \right)^{-1} \sum_{k=1}^N \frac{1}{(\sigma_k^2)^{(i)}} \overline{\mathbf{A}}_k^{(i),T} \mathbf{W}_k^{(i)} \overline{\Delta \mathbf{y}}_k^{(i)}$$

$$\gamma = 1$$

$$\xi^{(i+1)} = \xi^{(i)} + \gamma \Delta \xi^{(i+1)} \text{ (halve } \gamma \text{ if necessary)} \quad \textcircled{\text{N.1}}$$

$$e_{k,t}^{(i+1)} = y_{k,t} - h_{k,t}(\xi^{(i+1)}) \rightsquigarrow \mathbf{E}_k^{(i+1)} = \begin{bmatrix} e_{k,0}^{(i+1)} & \dots & e_{k,1-p_k}^{(i+1)} \\ \vdots & & \vdots \\ e_{k,n-1}^{(i+1)} & \dots & e_{k,n-p_k}^{(i+1)} \end{bmatrix}$$

$$\alpha_k^{(i+1)} = \left(\mathbf{E}_k^{(i+1),T} \mathbf{W}_k^{(i)} \mathbf{E}_k^{(i+1)} \right)^{-1} \mathbf{E}_k^{(i+1),T} \mathbf{W}_k^{(i)} \mathbf{e}_k^{(i+1)}$$

$$\text{(stabilize } \alpha_k^{(i+1)} \text{ if necessary)} \quad \textcircled{\text{N.2}}$$

$$u_{k,t}^{(i+1)} = e_{k,t}^{(i+1)} - \alpha_{k,1}^{(i+1)} e_{k,t-1}^{(i+1)} - \dots - \alpha_{k,p_k}^{(i+1)} e_{k,t-p_k}^{(i+1)} = \alpha_k^{(i+1)}(L) e_{k,t}^{(i+1)}$$

$$(\sigma_k^2)^{(i+1)} = \frac{1}{n} \sum_{t=1}^n w_{k,t}^{(i)} \left(u_{k,t}^{(i+1)} \right)^2 = \frac{\mathbf{u}_k^{(i+1),T} \mathbf{W}_k^{(i)} \mathbf{u}_k^{(i+1)}}{n}$$

$$0 = 1 + \log v_k^{(i+1)} - \psi \left(\frac{v_k^{(i+1)}}{2} \right) + \psi \left(\frac{v_k^{(i+1)} + 1}{2} \right) - \log \left(v_k^{(i+1)} + 1 \right) \\ + \frac{1}{n} \sum_{t=1}^n \left(\log \frac{v_k^{(i+1)} + 1}{v_k^{(i+1)} + \left(u_{k,t}^{(i+1)} / \sigma_k^{(i+1)} \right)^2} - \frac{v_k^{(i+1)} + 1}{v_k^{(i+1)} + \left(u_{k,t}^{(i+1)} / \sigma_k^{(i+1)} \right)^2} \right) \quad \textcircled{\text{N.3}}$$

if $\max_{j,k} (|\xi_j^{(i)} - \xi_j^{(i+1)}|, |(\sigma_k^2)^{(i)} - (\sigma_k^2)^{(i+1)}|) < \varepsilon$ **and**

$\max_k (|v_k^{(i)} - v_k^{(i+1)}|) < \varepsilon_v$ **then**

break

Algorithm 2: Bootstrapping Algorithm

Input : $\hat{\xi}, \hat{\sigma}_k^2, \hat{v}_k, \hat{e}_k, \hat{u}_k, \hat{W}_k, \hat{\alpha}_k$

Output: $\Sigma\{\hat{\theta}\}$

For every component $k = 1, \dots, N$, every time instance $t = 1, \dots, n$ and every $b = 1, \dots, B$,

- Generate the white noise samples using the estimated t-distributions:

$$u_{k,1}^{(b)}, \dots, u_{k,n}^{(b)} \stackrel{\text{ind.}}{\sim} t_{\hat{v}_k}(0, \hat{\sigma}_k^2).$$

- Compute the colored noise samples using the estimated AR models:

$$e_{k,t}^{(b)} = \hat{\alpha}_{k,1} e_{k,t-1}^{(b)} + \dots + \hat{\alpha}_{k,p_k} e_{k,t-p_k}^{(b)} + u_{k,t}^{(b)}.$$

- Compute the observation samples using the fitted deterministic functions:

$$\ell_{k,t}^{(b)} = h_{k,t}(\hat{\xi}) + e_{k,t}^{(b)}.$$

- Use Algorithm 1 to compute the bootstrap solutions:

$$\hat{\theta}^{(1)}, \dots, \hat{\theta}^{(b)}.$$

and compute the mean bootstrap solution:

$$\bar{\theta} = \frac{1}{B} \sum_{b=1}^B \hat{\theta}^{(b)}.$$

- Compute the bootstrap covariance matrix: $\Sigma\{\hat{\theta}\} \approx \frac{1}{B} \sum_{b=1}^B (\hat{\theta}^{(b)} - \bar{\theta})(\hat{\theta}^{(b)} - \bar{\theta})^T.$
-

4 Field Experiment Setup and Its Results

The multi-sensor system (MSS) considered for geo-referencing in this contribution consists of a terrestrial laser scanner (TLS) and two GNSS antennas/receivers. As shown in Fig. 1, the TLS is the core sensor of the MSS, which rotates about its vertical axis with a constant angular velocity. The GNSS receivers are connected to two eccentric GNSS antennas, which are mounted such that the centroid of antenna reference points (ARPs) coincides with the TLS's rotating axis. In addition to these GNSS receivers, we assume to have a nearby reference GNSS station with known position. During the data acquisition, the MSS performs a complete 360° rotation about its vertical axis while collecting both TLS data (i.e., a 3D point cloud) and GNSS measurements, which are synchronized through GNSS receiver event marker (see [20] for details regarding the MSS).

The objective of the navigation system is to provide the position (the centroid of ARPs) and the pointing direction (heading) of the laser scanner. In [20], the standard real-time kinematic (RTK) positioning [21] was used to estimate individual rotating antenna positions. Then, a constrained nonlinear filtering method, in particular an extended Kalman filter, was used to obtain the above parameters. Nadarajah et al.



Fig. 1 Prototypical realization of the MSS formed by a laser scanner (blue) and two eccentrically mounted GNSS antennas (green). This MSS was used for the field experiment

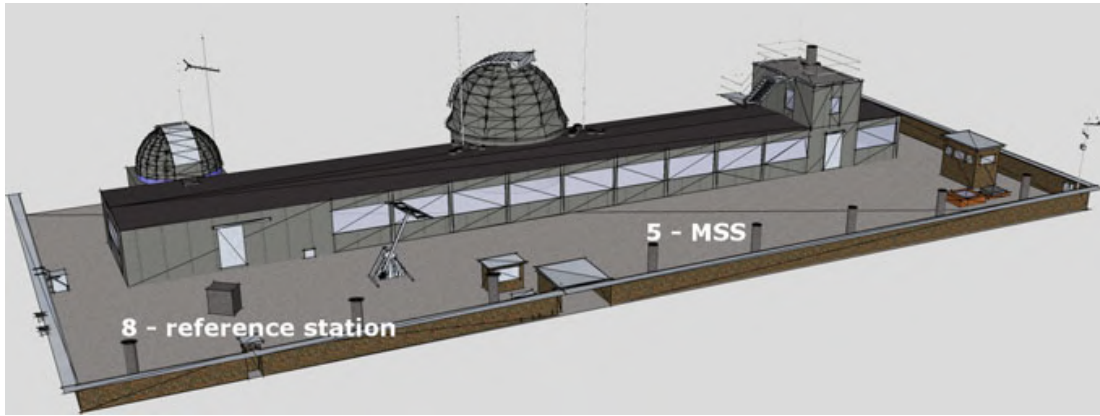


Fig. 2 Location of the practical experiment on the roof of the building of the Geodetic Institute (Messdach) at the Leibniz Universität Hannover, Germany. The MSS from Fig. 1 is mounted on pillar 5 while reference station is located at pillar 8

[22] utilized a constrained integer least-squares and array-aided positioning enabling improved ambiguity resolution and improved positioning accuracy.

For numerical analyses, we used the GNSS observations from a field experiment on the building of the Geodetic Institute (Messdach) at the Leibniz Universität Hannover, Germany. The MSS was mounted on pillar 5 (see Fig. 2) and equipped with a TLS Z+F *Imager 5006* as well as two individually and absolutely calibrated *LEIAX1202GG* GNSS antennas about 0.6 m apart. These antennas are connected to two dual frequency GNSS receivers *JAVAD TRE_G3TH DELTA*. The reference station (and thus the origin of the coordinates) is located at pillar 8 (about 20 m from the MSS) and equipped with a *JAVAD TRE_G3TH DELTA* GNSS receiver and *LEIAR25.R3 LEIT* antenna. A full 360° rotation consists of approximately 7609 points (acquired with a data rate of 10 Hz) with respect to one antenna.

We applied the GEM algorithm to approximate a measured and preprocessed 3D GNSS time series (see [20, 22]) by a 3D circle with the three components North (X_t), East (Y_t), and Up (Z_t):

$$\begin{pmatrix} X_t \\ Y_t \\ Z_t \end{pmatrix} = \begin{pmatrix} -r \cos(T_t) \sin(\Phi) + r \sin(T_t) \cos(\theta) \cos(\Phi) + C_x \\ r \cos(T_t) \cos(\Phi) + r \sin(T_t) \cos(\theta) \sin(\Phi) + C_y \\ -r \sin(T_t) \sin(\theta) + C_z \end{pmatrix} + \begin{pmatrix} E_{1,t} \\ E_{2,t} \\ E_{3,t} \end{pmatrix} \quad (11)$$

The multivariate (three-dimensional) nonlinear regression model in (11) contains six model parameters: two for the orientation (azimuth angle $\Phi \in [-\pi, \pi]$ and zenith angle $\theta \in [0, \pi]$) of its unit normal vector, one for the radius (r), and three for the circle center (C_x, C_y, C_z) (cf. [23]). In this application, $n = 7896$ time instances in (11) were registered. Concerning the random deviations $\mathbf{E}_t$, we determined three different AR models. To identify a correct and computationally inexpensive model for the given real data set, we applied the GEM algorithm (Algorithm 1) for different AR orders concerning the North-, East-, and Up-time series, beginning with small orders. We used the maximum cumulated periodogram test described in [13] for testing if each estimated noise series $\hat{u}_{k,1}, \dots, \hat{u}_{k,n}$ differs significantly from theoretical white noise. In this way, we determined for the North- and East-component an AR(15) model each, and for the Up-component an AR(18) model (all of which passed the white noise tests). In Fig. 3, the estimated cumulated periodograms of the decorrelated residuals for the North- and East-component are depicted.

Figure 4 shows the adjusted circle and the observed 3D points. Having obtained an estimated d.o.f. of 120 and 88, respectively, for the North- and East-component, we conclude that the white noise components of these GNSS series are almost normally

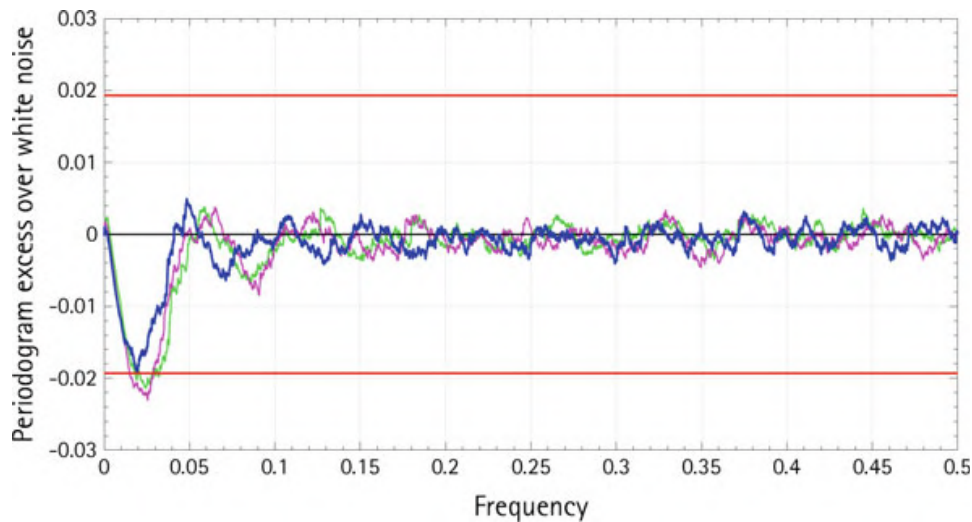


Fig. 3 Excess of the estimated periodogram of the decorrelated residuals for the North- and East-component AR(15) model (blue and green) and for the Up-component AR(18) model (magenta) with respect to the theoretical white noise periodogram (black) and 99% significance bounds (red)

Fig. 4 3D view of observed (black points) and adjusted circle (red line) for $n = 7827$ real three-dimensional GNSS measurements taken from [22], displayed in a North East Up (NEU) coordinate system

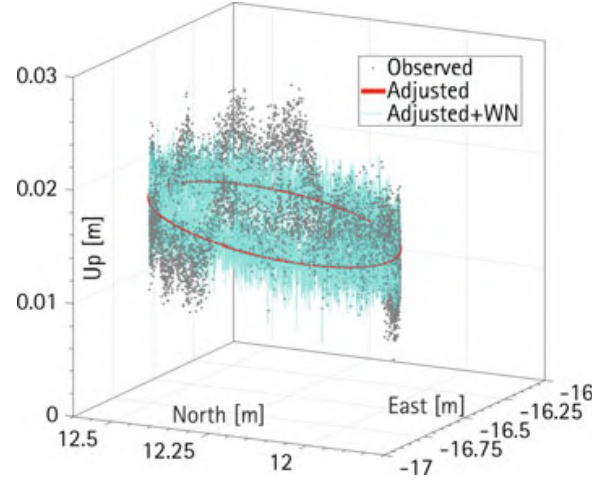


Table 1 Results of the bootstrap algorithm using 500 replications. The rows 1–6 of this table indicate the means of estimated model parameters, their standard deviations, and the 95% confidence intervals. The rows 7–9 give the results for estimated scale factors for the North- (X), East- (Y), and Up- (Z) components

	Mean	Std	95% confidence interval
r (m)	0.2971	0.00022	[0.2968 0.2974]
Φ (rad)	−0.00078	0.00061	[−0.00183 0.0003]
θ (rad)	−0.00665	0.00283	[−0.0112 − 0.00217]
C_x (m)	12.2340	0.00013	[12.2338 12.2342]
C_y (m)	−16.6317	0.00038	[−16.6321 − 16.6312]
C_z (m)	0.01628	0.00138	[0.01458 0.01776]
σ_X^2	7.3×10^{-7}	1.2×10^{-8}	$[7.1 \times 10^{-7} \ 7.6 \times 10^{-7}]$
σ_Y^2	1.4×10^{-6}	2.2×10^{-8}	$[1.3 \times 10^{-6} \ 1.4 \times 10^{-6}]$
σ_Z^2	3.0×10^{-6}	4.0×10^{-8}	$[2.9 \times 10^{-6} \ 3.1 \times 10^{-6}]$

distributed. In contrast, the d.o.f. for the Up-component turned out to be 33, which indicates that the corresponding white noise residuals are moderately heavy-tailed.

We used Algorithm 2 of section “[Bootstrapping Algorithm for Determining the Covariance Matrix](#)” to illustrate the bootstrap sampling with the GEM algorithm and to derive the full variance–covariance matrix of all estimated parameters. For this purpose, we generated $B = 500$ samples. In Table 1, we present 95% confidence interval alongside the means and standard deviations with respect to both the six estimated circle parameters and the three scale factors. The metric components (radius and center point) of the circle model are estimated with standard deviations at submillimeter level. For the azimuthal orientation (Φ), the model parameter is estimated in centi-degree range, resulting in a metric uncertainty of about 0.018 m at a distance of 30 m. In comparison to the approach in [20], these results constitute an improvement of the estimated parameters of main interest within the direct geo-referencing of 3D point clouds. The estimated zenith angle (θ) is dominated by the Up-component of the GNSS observations. Since this Up-component is typically characterized by a higher measurement noise than the horizontal components

(North and East), we expected a higher standard deviation for the zenith angle. The same behavior can be seen for the estimated scale factors and corresponding confidence intervals, whereas the horizontal components perform better in contrast to the Up-component. In case of a carefully performed leveling process of the MSS, the variations of the zenith angle can be minimized and, therefore, would play a secondary role.

5 Conclusions and Outlook

In this paper, we considered multivariate regression time series with both AR colored noise and outlier-afflicted/heavy-tailed white noise components. The latter were modeled by scaled t-distributions with estimated d.o.f., in which distributions are heavy-tailed and thus frequently used as a way to achieve a data-adaptive robust ML estimator. We described the theory and implementation of a GEM algorithm, in which the deterministic model parameters, the AR coefficients, the scale factors, and the d.o.f. for the multiple time series can be estimated in the form of IRLS. In order to derive the covariance matrix of these parameters as a quality measure for the estimated parameters, we demonstrated an MC-based bootstrap algorithm, which allows also for the computation of confidence intervals concerning all parameters for a given error probability. The presented algorithm was also tested in a real data experiment using GNSS measurements. A model selection with respect to the order of the AR process was performed by employing a periodogram-based white noise test. Finally, the analysis of the estimated d.o.f. throughout the different multivariate time series showed that the white noise residuals of the Up-component of the GNSS time series deviate from a normal distribution. In the future, we intend to extend the GEM algorithm by modeling in addition cross-correlations between the individual time series.

References

1. Huber, P.J.: Robust estimation of a location parameter. *Ann. Math. Stat.* **35**, 73–101 (1964)
2. Lange, K.L., Little, R.J.A., Taylor, J.M.G.: Robust statistical modeling using the t-distribution. *J. Am. Stat. Assoc.* **84**, 881–896 (1989)
3. Dempster, A.P., Laird, N.M., Rubin, D.B.: Maximum likelihood from incomplete data via the EM algorithm. *J. R. Stat. Soc. (Ser. B)* **39**, 1–38 (1977)
4. Liu, C.H.: ML estimation of the multivariate t distribution and the EM algorithm. *J. Multivar. Anal.* **63**, 296–312 (1997)
5. Liu, C.H., Rubin, D.B.: The ECME algorithm: a simple extension of EM and ECM with faster monotone convergence. *Biometrika* **81**, 633–648 (1994)
6. Meng, X., Rubin, D.B.: Maximum likelihood estimation via the ECM algorithm: a general framework. *Biometrika* **80**, 267–278 (1993)
7. McLachlan, G.J., Krishnan, T.: *The EM Algorithm and Extensions*. Wiley, Hoboken, New Jersey (2008)

8. Phillips, R.F.: Least absolute deviations estimation via the EM algorithm. *Stat. Comput.* **12**, 281–285 (2002)
9. Wang, K., Xiong, S., Li, Y.: Modeling with noises for inertial sensors. In: *Position Location and Navigation Symposium (PLANS) 2012*, IEEE/ION, pp. 625–632 (2012)
10. Park, M., Gao, Y.: Error and performance analysis of MEMS-based inertial sensors with a low-cost GPS receiver. *Sensors* **8**, 2240–2261 (2008)
11. Schuh, W.D.: The processing of band-limited measurements; filtering techniques in the least squares context and in the presence of data gaps. *Space Sci. Rev.* **108**, 67–78 (2003)
12. Luo, X.: *GPS Stochastic Modelling: Signal Quality Measures and ARMA Processes*. Springer, Berlin, Heidelberg (2013)
13. Kargoll, B., Omidalizarandi, M., Loth, I., Paffenholz, J.A., Alkhatib, H.: An iteratively reweighted least-squares approach to adaptive robust adjustment of parameters in linear regression models with autoregressive and t-distributed deviations. *J. Geodesy.* (2017). <https://doi.org/10.1007/s00190-017-1062-6>
14. Alkhatib, H., Kargoll, B., Paffenholz, J.A.: Robust multivariate time series analysis in nonlinear models with autoregressive and t-distributed errors. In: Valenzuela, O., Rojas, F., Pomares, H., Rojas, I. (eds.) *Proceedings ITISE 2017—International work-conference on Time Series*, vol. 1, pp. 23–36 (2017). ISBN: 978-84-17293-01-7
15. Efron, B.: Bootstrap methods: another look at the jackknife. *Ann. Stat.* **7**(1), 1–26 (1997)
16. Alkhatib, H.: On Monte Carlo methods with applications to the current satellite gravity missions, Dissertation. Universitt Bonn, Bonn. Institute for Geodesy and Geoinformation (2008). <http://hss.ulb.uni-bonn.de/2007/1078/1078.pdf>
17. Li, H., Maddala, G.S.: Bootstrapping time series models. *Economet. Rev.* **15**, 115–158 (1996)
18. Porat, B.: *Digital Processing of Random Signals*. Dover, Mineola/New York (1994)
19. Hargreaves, G.I.: *Interval Analysis in MATLAB*. Numerical Analysis Report No. 416, Manchester Centre for Computational Mathematics, The University of Manchester (2002). ISSN 1360-1725
20. Paffenholz, J.A.: Direct Geo-Referencing of 3D Point Clouds with 3D Positioning Sensors. German Geodetic Commission, Series C. Ph.D. thesis, No. 689, Munich (2012). https://dgk.badw.de/fileadmin/user_upload/Files/DGK/docs/c-689.pdf
21. Wübbena, G., Bagge, A., Schmitz, M.: RTK networks based on Geo++ GNSMART-concepts, implementation, results
22. Nadarajah, N., Paffenholz, J.A., Teunissen, P.J.G.: Integrated GNSS attitude determination and positioning for direct geo-referencing. *Sensors* **14**(7), 12715–12734 (2014). <https://doi.org/10.3390/s140712715>
23. Moon, P., Spencer, D.E.: *Field Theory Handbook—Including Coordinate Systems, Differential Equations, and Their Solutions*, 3rd edn. Springer, New York (1988)

Article

Model Selection for Parametric Surfaces Approximating 3D Point Clouds for Deformation Analysis

Xin Zhao * , Boris Kargoll , Mohammad Omidalizarandi , Xiangyang Xu and Hamza Alkhatib 

Geodetic Institute, Leibniz Universität Hannover, Nienburger Str. 1, 30167 Hannover, Germany; kargoll@gih.uni-hannover.de (B.K.); zarandi@gih.uni-hannover.de (M.O.); xu@gih.uni-hannover.de (X.X.); alkhatib@gih.uni-hannover.de (H.A.)

* Correspondence: zhao@gih.uni-hannover.de; Tel.: +49-511-7622466

Received: 20 February 2018; Accepted: 16 April 2018; Published: 19 April 2018



Abstract: Deformation monitoring of structures is a common application and one of the major tasks of engineering surveying. Terrestrial laser scanning (TLS) has become a popular method for detecting deformations due to high precision and spatial resolution in capturing a number of three-dimensional point clouds. Surface-based methodology plays a prominent role in rigorous deformation analysis. Consequently, it is of great importance to select an appropriate regression model that reflects the geometrical features of each state or epoch. This paper aims at providing the practitioner some guidance in this regard. Different from standard model selection procedures for surface models based on information criteria, we adopted the hypothesis tests from D.R. Cox and Q.H. Vuong to discriminate statistically between parametric models. The methodology was instantiated in two numerical examples by discriminating between widely used polynomial and B-spline surfaces as models of given TLS point clouds. According to the test decisions, the B-spline surface model showed a slight advantage when both surface types had few parameters in the first example, while it performed significantly better for larger numbers of parameters. Within B-spline surface models, the optimal one for the specific segment was fixed by Vuong's test whose result was quite consistent with the judgment of widely used Bayesian information criterion. The numerical instabilities of B-spline models due to data gap were clearly reflected by the model selection tests, which rejected inadequate B-spline models in another numerical example.

Keywords: terrestrial laser scanning; surface modeling; B-spline; polynomial; Gauss-Markov model; simulation-based Cox's test; Vuong's test

1. Introduction

Deformation monitoring of engineering structures such as bridges, tunnels, dams, and tall buildings is a common application of engineering surveying [1]. As summarized in Mukupa et al. [2], deformation analysis can be based on different comparison objects, namely, point-to-point, point-to-surface, or surface-to-surface. The point-to-point-based analysis is a common approach to describe deformations that are captured by conventional point-wise surveying techniques. Examples of such techniques are the total station and the global navigation satellite system; however, in many cases, these have been surpassed by the use of LiDAR technology, especially terrestrial laser scanning (TLS) [3,4]. Although the single-point precision of TLS is in the sub-centimeter range (± 2 to ± 25 mm), the high redundancy of the scanning observations facilitates a higher precision via the application of least-squares based curve or surface estimation and, hence, an adequate precision of the estimated deformation parameters [5].

A point-to-surface-based analysis is carried out to represent a deformation by the distance between the point cloud in one epoch and the surface estimated from measurements in another epoch. Such a surface can be constructed as a polygonal model (mesh) [6–8] or as a regression model (e.g., a polynomial or B-spline surface model) [9–13]. The procedure of a surface-to-surface-based deformation analysis, which is appropriate in certain situations, is to divide the point clouds into cells and to compare the parameters of fitted planes based on cell points in two epochs. This method is applied in Lindenbergh et al. [14], where the different positions of the laser scanner and strong wind contribute to the change of the coordinate system. The aforementioned three approaches to deformation analysis are complemented by the “point-cloud-based” approach, in which a deformation is reflected by the parameters of a coordinate transformation between sets of point clouds in various epochs. The common algorithm for determining the transformation matrix is the iterative closest point algorithm. The authors of Girardeau-Montaut et al. [15] presented three simple cloud-to-cloud comparison techniques for detecting changes in building sites or indoor facilities within a certain time.

Aiming at rigorous deformation detection from scatter point clouds, it is crucial to describe the geometrical features of the object accurately by an appropriate curve or surface regression model, and emphasis is put on the latter model in this paper. The purpose of surface fitting is to estimate the continuous model function from the scatter point samples, which can be implemented by approximation in the case of redundant measurements. There are many approximation approaches for working with surfaces based on an implicit, explicit, or parametric form. Parametric models are usually employed to fit point cloud data in applications such as deformation monitoring and reverse engineering. Different parametric models perform differently in terms of accuracy and number of coefficients when fitted to a dataset. Among the many methods utilized in various applications for approximating point clouds, polynomial model fitting is usually applied to smooth and regular objects due to its simple operation. In [16], the authors assumed a concrete arch as regular and analyzed the deformation behavior through comparing fitted second-degree polynomial surfaces. The more involved fitting of B-splines and non-uniform rational basis splines is often preferred for modeling geometrically complicated objects. In this context, much research has focused on the optimization of the mathematical and stochastic models. In Bureick et al. [17], the authors optimized free-form curve approximation by means of an optimal selection of the knot vector. Furthermore, in Harmening and Neuner [18], the authors improved the parametrization process in B-spline surface fitting by using an object-oriented approach instead of focusing on a superior coordinates system, thereby enabling the generated parameters to reflect the features of the object realistically. Moreover, in Zhao et al. [19], the authors suggested a new stochastic model for TLS measurements and used the resulting covariance matrix within the least-squares estimation of a B-spline curve.

The need for model selection and statistical validation was emphasized in Wunderlich et al. [20], the authors of which described the deficiencies in current areal deformation analysis and presented possible strategies to improve this situation. Typically, the selection of surface model depends on the object features—for example, whether the surface is regular or irregular. However, in most cases, it is unclear whether the object is smooth enough to be described by a simple model (e.g., as a low-order, global polynomial surface) or not. This limitation serves as the motivation for discriminating between estimated surface models in order to select the most appropriate one. In the context of model selection, Harmening and Neuner [21,22] investigated statistical methods based on information criteria and statistical learning theory for selecting the optimal number of control points within B-spline surface estimation. Another possibility is to compare the (log-)likelihoods of competing models directly by means of the general testing principle by D.R. Cox [23]. In Williams [24], the authors improved the Cox’s test based on the use of Monte-Carlo simulation, which is straightforward to implement. This kind of test has already been used in Zhao et al. [19] to select the best fitting stochastic model for B-spline curve estimation. In Vuong [25], the authors use likelihood-ratio-based statistics to discriminate the competing models based on Kullback–Leibler information criterion. Such hypothesis

tests offer the advantage that significant probabilistic differences between models can be detected, which is information that has not been provided by the methods mentioned previously.

The motivation of this paper lies in the selection of the most parsimonious, yet sufficiently accurate, parametric description of structure based on TLS measurements, whose model is applied to reflect the surface-based deformation of measured objects. Different from standard model selections procedures based on information criteria, we introduce two likelihood-ratio tests from D.R. Cox and Q.H. Vuong, which are instantiated in numerical examples to discriminate statistically between widely used polynomial and B-spline surfaces as models of given TLS point clouds. The selected surface model's performance in reflecting deformation is compared with the result of the block-means approach. The paper is organized as follows. In Section 2, the methodology of surface approximation and model selection is reviewed and explained. This methodology is instantiated in two numerical examples by discriminating between widely used polynomial and B-spline surfaces. The evaluation of approximated surface models as well as their performance in deformation analysis are given as results in Section 3. The subsequent Section 4 provides a further discussion on the results and a comparison with results obtained by some well-known penalization information criterion approaches. Finally, conclusions are drawn in Section 5.

2. Methodology

2.1. Experiment Design

An experiment was conducted jointly with the Institute of Concrete Construction of the Leibniz Universität Hannover to probe the load-caused behavior and ultimate bearing capacity of a concrete arch structure with a length of about 2 m and thickness of 0.1 m. Loads were placed on top of the arch's surface for 13 epochs, within which the load was exerted with a uniform speed (2 kN/min) of about 20 min followed by a break of approximately 10 min for data capture [9]. The weight of load was increased continuously and reached 520 kN at the end of the 13th epoch.

A multi-sensor-system (MSS) consisting of a TLS (here Z+F Imager 5006), laser tracker (here Leica AT960LR) and digital camera (here Nikon D750) were used to acquire the data from the deformable arch structure. The positions of the MSS relative to the arch structure are shown in Figure 1 (see also [16]). TLS data were acquired in "super high" resolution mode with normal quality. The vertical and horizontal resolution was 0.0018° and vertical and horizontal accuracy was 0.007° rms. The TLS scanned the top and the side surfaces of the arch. The laser tracker was used as a reference sensor for the validation purpose, which allows sub-millimeter accuracy with a maximum permissible error of $15\ \mu\text{m} + 6\ \mu\text{m}/\text{m}$ [26]. In addition, a digital camera was used to capture the feature points with a high resolution (thus exploiting their strength in discrete feature point extraction). Targets were mounted in the surroundings and on the arch (see Figure 1) to perform the external calibration between the sensors.

Among various datasets, the focus of this literature lies in capturing the data by the TLS to use a large number of 3D point clouds with a high accuracy in approximating a surface model, which is important in rigorous deformation monitoring. In our experiment, the top surface of the arch is of great interest since it is under load pressure in the consecutive 13 epochs and has obvious movements compared to the other parts of the structure. However, as can be seen from Figure 1, the top surface is partially occluded by the steel beams. Consequently, it is necessary to extract the top-surface points in order to enable an accurate surface model. As a preliminary step to separate the obstructions and the arch-shape part of the object, the reflectance image was generated by using the reflectivity values of the raw TLS data. It was performed by assigning the reflectivity values of each point cloud to one pixel based on the scan resolution [27] (see Figure 2). Since the occluded objects such as the beams on top of the arc shape object were darker compared to the arc shape part, it could be discarded by means of the OpenCV threshold function and by setting the threshold value manually to 80 from a range of 0–255. Therefore, those values greater than 80 were set to 0. However, before performing the

thresholding, OpenCV GaussianBlur function with a size of 5 was applied to reduce the image noise. Next, the morphological opening and closing filters were applied to discard the very small segments. As previously, the position and orientation of the global coordinate system was defined in Figure 1, the Z-axis was in the zenith direction. Therefore, the Pass Through filter of the Point Cloud Library was applied to cut off those 3D point clouds below or greater than the predefined threshold in the Z-axis direction. Then the “setFilterLimits” member function of this filter was set to (−4.50 m, −3.75 m) to select the 3D point clouds within the boundaries. Subsequently, the StatisticalOutlierRemoval filter of the Point Cloud Library was applied to remove the outliers of the 3D point clouds. In this filter, the k nearest neighbor points were used to estimate the mean distance. Therefore, its “setMeanK” member function was set to 20. Next, the “setStddevMulThresh” member function of this filter was set to 3.0 to reject the outliers by means of the 3σ test. The extracted arch surface data is shown in Figure 3. As an example, only two representative segments of the point cloud within the red boundary are separately investigated since the middle area has significant deviations compared to the other parts of the surface. The same methodology is applicable in modeling other segments.

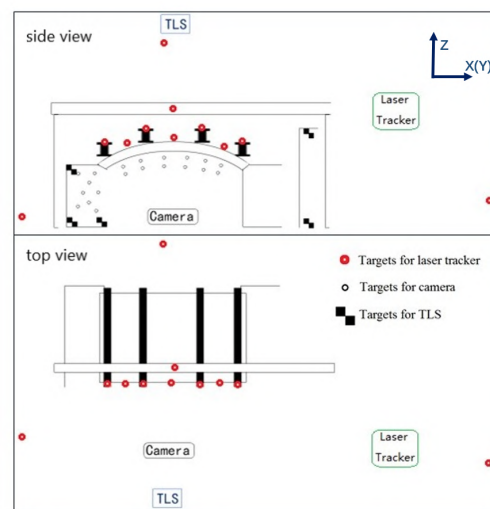


Figure 1. Sketch map of the experimental design concerning the locations of the instruments and relevant targets in side view (**upper**) and top view (**bottom**) [16].

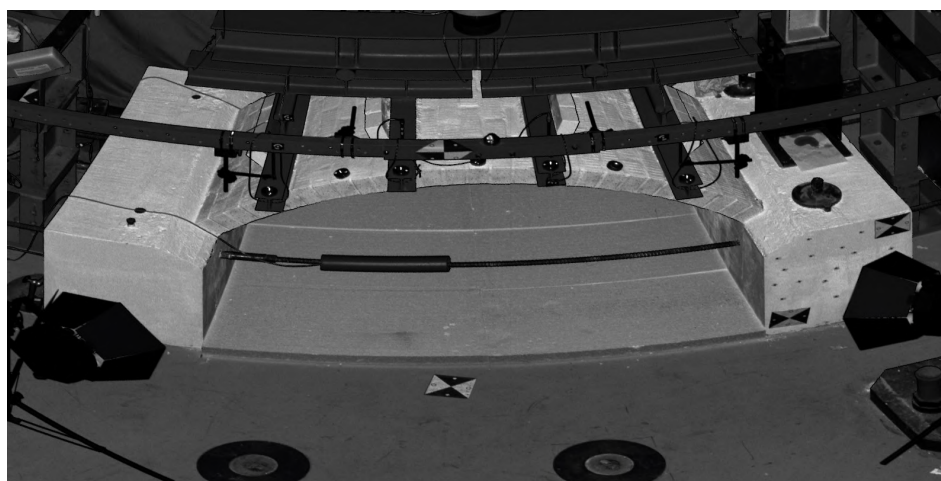


Figure 2. Reflectance image generated by reflectivity values of TLS data [16].

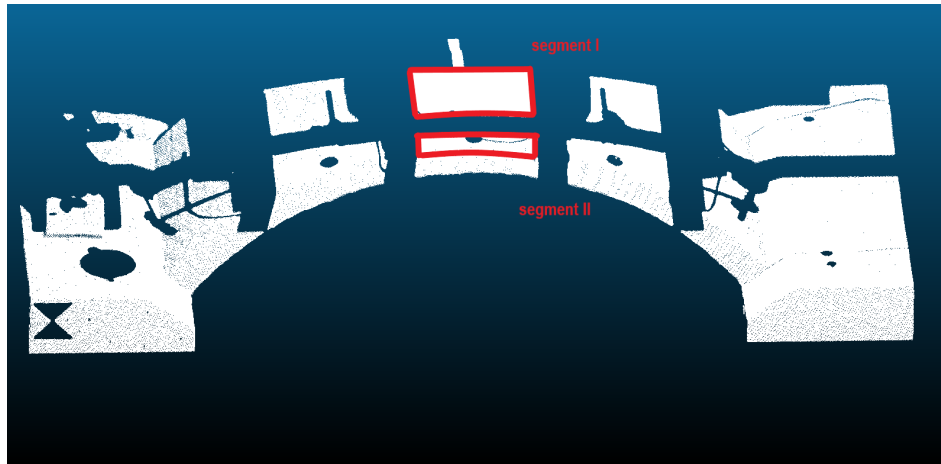


Figure 3. Extracted Arc-shape object and the target segments in our numerical example (within the red boundary) shown by the software CloudCompare.

2.2. Surface Approximation

In surface-based deformation analysis, an appropriate surface model is required as a representation of each deformation status. Among the various parametric models, polynomial surface fitting are the most common approach due to its easy implementation [9,16], while free-form surfaces, especially B-splines, have become relevant to deformation analysis due to their capacity to accurately model more detailed geometrical features including sharp edges, cusps, and leaps [28]. The functional relationships behind B-splines and polynomials, as two of the most commonly used types of surface model, as well as the corresponding approximation steps, are described in this subsection.

2.2.1. B-Spline Surface Approximation

The mathematical description of a 3D point

$$\mathbf{S} = [X \ Y \ Z] = \sum_{i=0}^n \sum_{j=0}^m \mathbf{N}_{i,p}(\bar{u}) \mathbf{N}_{j,q}(\bar{v}) \mathbf{P}_{i,j} \quad (1)$$

on a B-spline surface is based on the bidirectional combination of basis functions $\mathbf{N}_{i,p}(\bar{u})$, $\mathbf{N}_{j,q}(\bar{v})$ and 3D control points $\mathbf{P}_{i,j} = [P_{X_{i,j}} \ P_{Y_{i,j}} \ P_{Z_{i,j}}]$, which are located on a bidirectional net with the number of $n + 1$ and $m + 1$ in u - and v -directions.

B-spline surface approximation builds upon B-spline curve fitting in the two directions. Following Bureick et al. [17], this procedure can be carried out in three steps:

1. Parametrization of the measurements with s rows and t columns with respect to the u - and v -direction.
2. Determination of the knot vectors $\mathbf{U}$ and $\mathbf{V}$ in the u - and v -direction.
3. Estimation of the control points by means of a linear Gauss–Markov model.

The first two steps consist of the parameterization and computation of knot vectors, which serve as input parameters for the final estimation. Since the calculation of B-spline parameters are beyond the scope of this paper, the interested reader is referred to Bureick et al. [17] and Piegl and Tiller [29].

The final step of B-spline approximation is to estimate the positions of the control points, which is done essentially by adjusting a linear Gauss–Markov model (cf. [29,30]). Given measured points located on a grid defined by s rows and t columns, they also can be arranged in matrix form as

$$\mathbf{l} = \begin{bmatrix} \mathbf{l}_1 \\ \vdots \\ \mathbf{l}_{s,t} \end{bmatrix} = \begin{bmatrix} X_1 & Y_1 & Z_1 \\ \vdots & \vdots & \vdots \\ X_{s,t} & Y_{s,t} & Z_{s,t} \end{bmatrix}. \quad (2)$$

The addition of a corresponding matrix $\mathbf{v}$ of residuals to the observation matrix yields the adjusted observations, which can be represented by the functional B-spline model in Equation (1). We thus have for a particular observation the equation

$$\mathbf{l}_{k_u k_v} + \mathbf{v}_{k_u k_v} = \mathbf{S} = \sum_{i=0}^n \sum_{j=0}^m \mathbf{N}_{i,p}(\bar{u}_{k_u}) \mathbf{N}_{j,q}(\bar{v}_{k_v}) \mathbf{P}_{i,j} \quad (3)$$

where $k_u = 2, \dots, s$ and $k_v = 2, \dots, t$. We can write all of the equations jointly in the form $\mathbf{l} + \mathbf{v} = \mathbf{A}_0 \mathbf{x}$ with a design matrix within which the basis functions $\mathbf{N}_{i,p}(\bar{u})$, $\mathbf{N}_{j,q}(\bar{v})$ are calculated based on the parameterization and knot vectors

$$\mathbf{A}_0 = \begin{bmatrix} N_{0,p}(\bar{u}_1) \cdot N_{0,q}(\bar{v}_1) & \cdots & N_{n,p}(\bar{u}_1) \cdot N_{m,q}(\bar{v}_1) \\ \vdots & & \vdots \\ N_{0,p}(\bar{u}_s) \cdot N_{0,q}(\bar{v}_t) & \cdots & N_{n,p}(\bar{u}_s) \cdot N_{m,q}(\bar{v}_t) \end{bmatrix} \quad (4)$$

and (unknown) parameter matrix $\mathbf{x}$. We assume all of the measured point coordinates to have identical accuracies and to be uncorrelated, so that we obtain for the least squares estimates of these parameters:

$$\hat{\mathbf{x}} = \begin{bmatrix} \hat{P}_{X_{0,0}} & \hat{P}_{Y_{0,0}} & \hat{P}_{Z_{0,0}} \\ \vdots & \vdots & \vdots \\ \hat{P}_{X_{n,m}} & \hat{P}_{Y_{n,m}} & \hat{P}_{Z_{n,m}} \end{bmatrix} = (\mathbf{A}_0^T \mathbf{A}_0)^{-1} \mathbf{A}_0^T \mathbf{l}. \quad (5)$$

2.2.2. Polynomial Surface Approximation

Denoting a generic surface point by $\mathbf{S} = (X, Y, Z)$, the Z -component can be expressed as the two-fold linear combination

$$\begin{aligned} Z = \sum_{i=0}^p \sum_{j=0}^q a_{i,j} X^i Y^j &= a_{0,0} + a_{1,0}X + a_{0,1}Y + a_{2,0}X^2 + a_{1,1}XY + a_{0,2}Y^2 + \dots \\ &+ a_{p,0}X^p + a_{p-1,1}X^{p-1}Y + \dots + a_{1,q-1}XY^{q-1} + a_{0,q}Y^q \end{aligned} \quad (6)$$

where $\mathbf{a} = [a_{0,0}, a_{1,0}, \dots, a_{p,q}]$ is the coefficient vector having $(p+1)(q+1) - 1$ elements, and where p and q represent the polynomial degrees with respect to the X - and Y -components, respectively.

Polynomial surface fitting consists of the estimation of the coefficient vector $\mathbf{a}$, which we carry out again in the least-squares sense by minimizing the sum of squared (“vertical”) residuals

$$\Omega^2 = \sum_{n=1}^N (l_n - \sum_{i=0}^p \sum_{j=0}^q a_{i,j} X_n^i Y_n^j)^2 \quad (7)$$

where $\mathbf{l} = [l_1, l_2, \dots, l_N]$ is now an observation vector consisting of the N measured Z -coordinates. The design matrix

$$\mathbf{A}_1 = \begin{bmatrix} X_1^0 \cdot Y_1^0 & \cdots & X_1^i \cdot Y_1^j & \cdots & X_1^p \cdot Y_1^q \\ \vdots & & & & \vdots \\ X_N^0 \cdot Y_N^0 & \cdots & X_N^i \cdot Y_N^j & \cdots & X_N^p \cdot Y_N^q \end{bmatrix} \quad (8)$$

is computed by exponentiation (where $i = 0, 1, 2, \dots, p$, $j = 0, 1, 2, \dots, q$) and multiplication of the X- and Y-components, which are considered as error-free. Under the assumption again of homoskedastic and uncorrelated measurements, the estimated parameters are given by

$$\tilde{\mathbf{x}} = \begin{bmatrix} \tilde{a}_{0,0} \\ \vdots \\ \tilde{a}_{p,q} \end{bmatrix} = (\mathbf{A}_1^T \mathbf{A}_1)^{-1} \mathbf{A}_1^T \mathbf{l}. \quad (9)$$

2.2.3. Parameter Number of Competing Models

The approximation quality of a surface model is related to its complexity embodied in the number of parameters. Hence, when choosing pairs of models to be compared, we should pay attention to the parameter numbers.

In the initial comparison, the target segments of the point cloud of the first epoch are modeled by means of polynomial and B-spline surface functions with similar numbers of parameters in order to minimize the effect of model complexity. As the most basic description of a surface, the second degree polynomial function ($p = q = 2$) with six unknown parameters is approximated. A B-spline surface with the same parameter number ($n = 1, m = 2$) is modeled as a competitor. In order to facilitate a comprehensive analysis, polynomial functions of higher degrees are adjusted and compared to other estimated B-spline surface models. According to the general polynomial model Equation (6), a third-degree polynomial model is based on the specification of $p = q = 3$, resulting in 10 unknown parameters to be estimated. It is reasonable to compare this model with the adjusted B-spline surface involving nine parameters ($m = n = 2$). Further comparisons are carried out between the fourth-degree polynomial and B-spline surface models with 15 and 16 unknown parameters, respectively. Table 1 lists the candidate surface models mentioned above, where N_{poly} and N_B represent the number of parameters of the polynomial and of the B-spline models, respectively.

Table 1. The numbers of parameters for the various employed polynomial and B-spline surface models.

Pairs	Polynomial Model		B-Spline Model	
	Degree	N_{poly}	n, m	N_B
I	2nd	6	$n = 1, m = 2$	6
II	3rd	10	$n = 2, m = 2$	9
III	4th	15	$n = 3, m = 3$	16

It should be mentioned that polynomial functions of degrees higher than four are useless in our numerical example, since the resulting normal equation matrices within parameter estimation Equation (9) would be ill-conditioned. In this case, on the one hand, it is quite interesting to compare the best-fitting fourth-degree polynomial model with a higher-quality yet more complex B-spline surfaces when considering the complexity of models as penalization. It is predicted that the latter would be superior to the former in initial comparison pairs, but the superiority is expected to be offset by penalization due to increasing parameters. The comparison results of Segment I will be presented in Section 3.1 in Table 3 and in Appendix in Table A1. It helps to judge in which situation the B-spline models are recommended compared with the polynomial model. On the other hand, in practice, among the recommended B-spline models, we need the optimal one for further deformation analysis, which motivates the comparison within B-spline models. The comparison results of Segment I will be shown in Section 3.2 in Table 5 and in Appendix in Table A2.

2.3. Model Selection Method

The aim of model selection is to find a balance between the parsimony of the model and its approximation quality [21]. Unlike the trial-and-error procedures and information theoretic criterion

model selection approaches, we adopt the likelihood-ratio-based hypotheses testing framework to discriminate between the competing models. The likelihood ratio tests are generally used to compare two nested models; however, in our case, the polynomial surface model and B-spline model are non-nested because neither model can be reduced to the other by imposing a set of parametric restrictions or limiting process.

Regarding the non-nested models selection problem, in Cox [23] and Vuong [25], the authors proposed respective approaches to extend the likelihood ratio test into non-nested cases. In this subsection, both a simulation-based version of Cox's test and Vuong's test, which will be instantiated later with the experiment data, are explained.

2.3.1. Simulation-Based Version of Cox's Test

Under the assumption of normally distributed, uncorrelated and homoskedastic random deviations, the observation models can be defined in terms of the generic log-likelihood function

$$\begin{aligned} L(\mathbf{x}, \sigma^2; \mathbf{1}) &= \ln \prod_{n=1}^N \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left\{ -\frac{1}{2} \left(\frac{l_n - \mathbf{A}_n \mathbf{x}}{\sigma} \right)^2 \right\} \\ &= -\frac{N}{2} \ln(2\pi) - \frac{N}{2} \ln(\sigma^2) - \frac{1}{2} \sum_{n=1}^N \frac{(l_n - \mathbf{A}_n \mathbf{x})^2}{\sigma^2} \end{aligned} \quad (10)$$

where the variance factor σ^2 is treated as an unknown parameter alongside the functional parameters. Let us define $L_0(\mathbf{x}, \sigma^2; \mathbf{1})$, and $L_1(\mathbf{x}, \sigma^2; \mathbf{1})$ to be the specific log-likelihood functions with respect to the design matrices Equations (4) and (8), respectively. Both types of design matrix define different types of functions where neither is a special case of the other one. Thus, the two sets of multivariate normal distributions defined by L_0 and L_1 are non-nested, so that the likelihood ratio test cannot be applied in its usual form ([31] cf.) (pp. 276–278).

According to Cox [23], we may, however, use the logarithmized likelihood ratio

$$\begin{aligned} L_{0,1} &= L_0(\hat{\mathbf{x}}, \hat{\sigma}^2; \mathbf{1}) - L_1(\tilde{\mathbf{x}}, \tilde{\sigma}^2; \mathbf{1}) \\ &= -\frac{N}{2} \ln(\hat{\sigma}^2) - \frac{1}{2\hat{\sigma}^2} \sum_{n=1}^N (l_n - \mathbf{A}_{0n} \hat{\mathbf{x}})^2 + \frac{N}{2} \ln(\tilde{\sigma}^2) + \frac{1}{2\tilde{\sigma}^2} \sum_{n=1}^N (l_n - \mathbf{A}_{1n} \tilde{\mathbf{x}})^2 \end{aligned} \quad (11)$$

for testing the adequacy of the polynomial model against the B-spline model. Note that the substituted least squares solutions Equations (5) and (9) are identical to the maximum likelihood estimates; furthermore, the two occurring maximum likelihood estimates of the variance factor σ^2 are given by $\hat{\sigma}^2 = \frac{1}{N} \sum_{n=1}^N (l_n - \mathbf{A}_{0n} \hat{\mathbf{x}})^2$ and $\tilde{\sigma}^2 = \frac{1}{N} \sum_{n=1}^N (l_n - \mathbf{A}_{1n} \tilde{\mathbf{x}})^2$ ([31] cf.) (pp. 161). The statistic $L_{0,1}$ follows approximately a normal distribution

1. with certain expectation μ_0 and standard deviation σ_0 if the polynomial model is true, and
2. with certain expectation μ_1 and standard deviation σ_1 if the B-spline model is true.

Thus, we may calculate the approximately standard normally distributed statistics $T_0 = (L_{0,1} - \mu_0)/\sigma_0$ and $T_1 = (L_{0,1} - \mu_1)/\sigma_1$ for carrying out two separate tests of the hypotheses—

1. H_0 : the polynomial model is true;
2. H_1 : the B-spline model is true—

at significance level α . We may determine the means μ_0 and μ_1 as well as the standard deviations σ_0 and σ_1 conditionally on the two parameter solutions $(\hat{\mathbf{x}}, \hat{\sigma}^2)$, and $(\tilde{\mathbf{x}}, \tilde{\sigma}^2)$ through a Monte Carlo simulation in analogy to the approach taken in Williams [24].

According to that approach, we start by generating a large number M of observation vectors $\mathbf{l}^{(1)}, \dots, \mathbf{l}^{(M)}$ randomly from the N -dimensional Gaussian distribution $N(\mathbf{A}\hat{\mathbf{x}}, \hat{\sigma}^2 \mathbf{I}_N)$. Based on these samples, we compute the corresponding solutions $(\hat{\mathbf{x}}^{(1)}; \hat{\sigma}^{(1)}), \dots, (\hat{\mathbf{x}}^{(M)}; \hat{\sigma}^{(M)})$ with respect to the

polynomial model and $(\tilde{\mathbf{x}}^{(1)}; \tilde{\sigma}^{(1)}), \dots, (\tilde{\mathbf{x}}^{(M)}; \tilde{\sigma}^{(M)})$ with respect to the B-spline model. We use the first solution set to evaluate the corresponding log-likelihood functions $L_0^{(1)}, \dots, L_0^{(M)}$, and the second set to evaluate $L_1^{(1)}, \dots, L_1^{(M)}$, so that we may compute the realizations $L_{0,1}^{(1)}, \dots, L_{0,1}^{(M)}$ of Equation (11). Thus, the arithmetic mean and empirical standard deviation of these sampled logarithmized likelihood ratios serve as estimates of μ_0 and σ_0 , leading to the standardized Gaussian test statistic T_0 under the currently assumed polynomial model.

The second test statistic T_1 (with respect to the test of the B-spline model) is computed in analogy to the first one, sampling now M observation vectors from $N(\mathbf{A}\tilde{\mathbf{x}}, \tilde{\sigma}^2 \mathbf{I}_N)$, and using the two new sets of parameter solutions (regarding the polynomial and B-spline model) to compute the M realizations of the log-likelihood ratios, as well as the resulting estimates of μ_1 and σ_1 .

Since Cox [23] suggests applying the one-sided decision rules,

1. reject H_0 if $T_0 < k_\alpha^{N(0,1)}$, and
2. reject H_1 if $T_1 > k_{1-\alpha}^{N(0,1)}$

(where $k_\alpha^{N(0,1)}$ is the α -quantile and $k_{1-\alpha}^{N(0,1)}$ the $1 - \alpha$ -quantile of the standard normal distribution), the execution of the two tests may result in four mutually exclusive decisions:

1. The polynomial model is rejected and the B-spline model is not rejected in the case of

$$T_0 < k_\alpha^{N(0,1)} \wedge T_1 \leq k_{1-\alpha}^{N(0,1)}. \quad (12)$$

2. The B-spline model is rejected and the polynomial model is not rejected in the case of

$$T_0 \geq k_\alpha^{N(0,1)} \wedge T_1 > k_{1-\alpha}^{N(0,1)}. \quad (13)$$

3. Both the polynomial and the B-spline models are rejected in the case of

$$T_0 < k_\alpha^{N(0,1)} \wedge T_1 > k_{1-\alpha}^{N(0,1)}. \quad (14)$$

4. Neither the polynomial nor the B-spline model is rejected in the case of

$$T_0 \geq k_\alpha^{N(0,1)} \wedge T_1 \leq k_{1-\alpha}^{N(0,1)}. \quad (15)$$

2.3.2. Vuong's Non-Nested Hypothesis Test

Vuong's test is based on the Kullback–Leibler information criterion (KLIC), which measures the closeness of two models and uses the likelihood-ratio-based statistics to test the null hypothesis that the competing models are equally close to the true data generating process against the alternative hypothesis that one model is closer [25]. Specifically, the two competing models are given as $\mathbf{F}_\theta = \{f(l; \theta); \theta \in \Theta\}$ and $\mathbf{G}_\gamma = \{g(l; \gamma); \gamma \in \Gamma\}$, l denotes variables, and θ and γ are their respective parameters. As defined by Vuong, the two models' Kullback–Leibler distances from the true density $h_0(l)$ are $E_0[\ln h_0(l)] - E_0[\ln f(l; \theta_*)]$ and $E_0[\ln h_0(l)] - E_0[\ln g(l; \gamma_*)]$, respectively, where E_0 denotes the expectation under the true model, and θ_* and γ_* are the pseudo-true values of θ and γ . It is clear that the model with a minimum KLIC value is closer to the truth, which is, however, hard to quantify. Thus, an equivalent selection criterion can be based on the quantities $E_0[\ln f(l; \theta_*)]$ and $E_0[\ln g(l; \gamma_*)]$, the better model being the one with larger quantity.

There are three possible cases when comparing, and we propose the null hypothesis, as the two models have equal expectation values so that they are equivalent. One alternative hypothesis is $E_0[\ln f(l; \theta_*)] > E_0[\ln g(l; \gamma_*)]$, which means $\mathbf{F}_\theta$ is the better model. The other alternative hypothesis is $E_0[\ln f(l; \theta_*)] < E_0[\ln g(l; \gamma_*)]$, meaning $\mathbf{G}_\gamma$ is better. Since the quantity $E_0[\ln f(l; \theta_*)] - E_0[\ln g(l; \gamma_*)]$ is still hard to quantify, Vuong consistently estimates it by $(1/n)$ times the likelihood ratio statistic.

In our specific case to discriminate between the polynomial and B-spline surface models, with independent and identically distributed observations $\mathbf{l}$, the probability density functions of competing models are given as $\mathbf{P}_{x,\sigma} = \{p(l; \mathbf{x}, \sigma^2)\}$ and $\mathbf{B}_{x,\sigma} = \{b(l; \mathbf{x}, \sigma^2)\}$, in which the parameter matrix $\mathbf{x}$ is estimated in its respective Gauss–Markov models and the variance factor σ^2 for competing models is calculated in the same way as in Cox’s test: $\hat{\sigma}^2 = \frac{1}{N} \sum_{n=1}^N (l_n - \mathbf{A}_{0n}\hat{\mathbf{x}})^2$ and $\hat{\sigma}^2 = \frac{1}{N} \sum_{n=1}^N (l_n - \mathbf{A}_{1n}\hat{\mathbf{x}})^2$. Thus, we propose the hypothesis in discriminating between models as follows:

1. H_0 : the polynomial and B-spline models are equally close to the truth;
2. H_p : the polynomial model is better since it is closer to the truth than the B-spline model is;
3. H_b : the B-spline model is better since it is closer to the truth than the Polynomial model.

Similar to Cox’s test, the statistic of Vuong’s test is also based on likelihood ratio. If we define $L_0(\hat{\mathbf{x}}, \hat{\sigma}^2; \mathbf{l})$ and $L_1(\hat{\mathbf{x}}, \hat{\sigma}^2; \mathbf{l})$ as log-likelihood functions for competing polynomial and B-spline surface models, the logarithmized likelihood ratio $L_{0,1}$ is calculated as (11). Vuong’s test is potentially sensitive to the number of estimated parameters on condition that the logarithmized likelihood ratio $L_{0,1}$ is adjusted by a correction factor K .

$$\tilde{L}_{0,1} \equiv L_{0,1} - K. \quad (16)$$

Vuong [25] suggests that K corresponds to either Akaike’s information criteria (AIC) or Bayesian information criteria (BIC). According to the former, $K = p_0 - p_1$, and, according to the latter, $K = (p_0/2) \ln N - (p_1/2) \ln N$, where p_0 and p_1 are numbers of parameters in competing models. The BIC generally penalizes free parameters more strongly than AIC. Here, we prefer the BIC correction factor in order to avoid an over-fitting problem.

Then, the adjusted likelihood ratio $\tilde{L}_{0,1}$ is rescaled in statistic as $T_V = N^{-1/2} \tilde{L}_{0,1} / \hat{w}$, where $\hat{w}^2$ is the variance calculated as

$$\hat{w}^2 = \frac{1}{N} \sum_{n=1}^N \left[\ln \frac{p(l_n; \mathbf{A}_{0n}\hat{\mathbf{x}}, \hat{\sigma}^2)}{b(l_n; \mathbf{A}_{1n}\hat{\mathbf{x}}, \hat{\sigma}^2)} \right]^2 - \left[\frac{1}{N} \sum_{n=1}^N \ln \frac{p(l_n; \mathbf{A}_{0n}\hat{\mathbf{x}}, \hat{\sigma}^2)}{b(l_n; \mathbf{A}_{1n}\hat{\mathbf{x}}, \hat{\sigma}^2)} \right]^2. \quad (17)$$

According to Vuong [25], when N is reasonably large, the statistic T_V converges, asymptotically to a standard normal distribution, $N(0, 1)$. In the decision-making process, practically, we compare T_V against the quantiles of a standard normal distribution, $C_{\alpha/2}^{N(0,1)}$, for significance level α . The models are discriminated through the following decision rules:

1. The polynomial and B-spline models are equally close to the truth in case of

$$C_{\alpha/2}^{N(0,1)} \leq T_V \leq C_{1-\alpha/2}^{N(0,1)}. \quad (18)$$

2. The polynomial model is better since it is closer to the truth than B-spline model in case of

$$T_V > C_{1-\alpha/2}^{N(0,1)}. \quad (19)$$

3. The B-spline model is better since it is closer to the truth than Polynomial model in case of

$$T_V < C_{\alpha/2}^{N(0,1)}. \quad (20)$$

2.4. Deformation Analysis

To probe the selected surface model’s performance in deformation analysis, each target segment of the 1st and 13th epochs are approximated by surface models. In the specific application, loads were exerted perpendicular to the ground (in the Z-direction) so that the deformation (Δ) is defined as the difference in approximated Z-coordinates of the two epochs ($\hat{\mathbf{Z}}_{13}, \hat{\mathbf{Z}}_1$), that is,

$$\Delta = \hat{\mathbf{Z}}_{13} - \hat{\mathbf{Z}}_1. \quad (21)$$

The evaluation criterion of surface models' performance is whether they are able to reflect the actual deformation, which are recorded by the point cloud. Since it is impossible to get an exact mutual spatial referencing of points in the different epochs, we compared the point clouds through the block-mean approach used in Paffenholz et al. [32]. In this application, the blocks had a size of $5 \text{ mm} \times 5 \text{ mm}$ involving 2–9 points, for which the medians of the Z-coordinates were computed as representative values. The high-density block-means between the two epochs were used to approximate the point-wise changes.

3. Results

3.1. Evaluation of Competing Polynomial and B-Spline Models

In order to statistically discriminate between the aforementioned polynomial and B-spline surface models listed in Table 1, the result of Cox's and Vuong's test for the two segments are given in Tables 2 and 4, respectively. It is noticeable that the observations are assumed to be following identical and independent normal distribution, which satisfy the prerequisite of both tests. In addition, Figure 4 demonstrates that the 10,000 log-likelihood ratio values of Equation (11) sampled with respect to Cox's test follow approximately a Gaussian distribution under both the polynomial and B-spline surface models. Thus, it is justified to standardize the log-likelihood ratio computed from the actual measurements by means of the sample mean and standard deviation under each of the two stochastic models (resulting in the values for T_0 and T_1 shown in Tables 2 and 4).

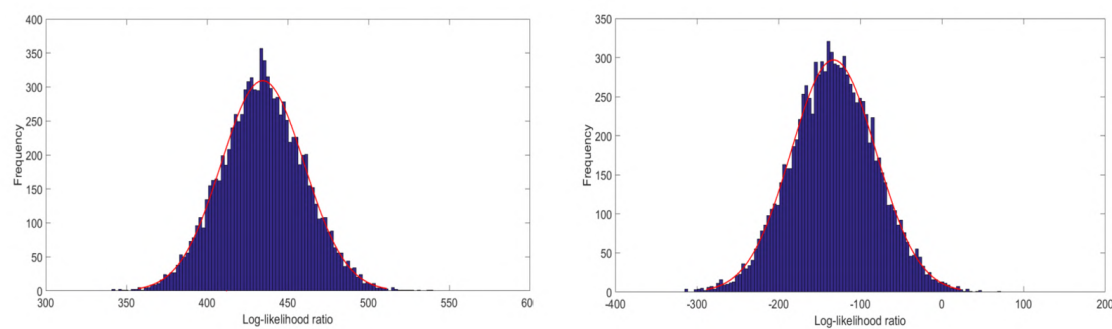


Figure 4. Histogram of the sampled log-likelihood ratio $L_{0,1}$ under the polynomial (left) and B-spline (right) surface model, approximated by a Gaussian density functions (in red).

The statistics are compared to the critical value at type-I error rate $\alpha = 0.05$. In Cox's test, the critical values are $k_{0.05}^{N(0,1)} = -1.64$ for statistic T_0 , and $k_{0.95}^{N(0,1)} = 1.64$ for statistic T_1 . In Vuong's test, the critical values are $C_{0.025}^{N(0,1)} = -1.96$ and $C_{0.975}^{N(0,1)} = 1.96$.

It can be clearly seen in Table 2 that, within the first pair of models, the B-spline surface model with six parameters is preferred over the second-degree polynomial model, since, in Vuong's test, the former is better verified, and, in Cox's test, the latter is rejected. This result indicates, with minor parameters, B-spline models are superior to the equivalent polynomial one. This conjecture is validated. In the second pair of models, neither the third-degree polynomial model nor the B-spline model is rejected or selected by tests, whose findings indicate that there is no significant difference between the two models. Next, further comparisons are carried out within Pair III between a fourth-degree polynomial and B-spline models with 15 and 16 unknown parameters, respectively. According to the tests results, Vuong's test indicates there is no significant superiority between the two, while the polynomial model is rejected by the Cox's test.

As mentioned before, polynomial functions of degrees greater than four become numerically unstable due to the appearance of singular matrices, so that they cannot be recommended. By contrast, approximated B-spline surface models for Segment I can describe the target segment of the point cloud

increasingly well when the parameters are increased, without producing such numerical difficulties. When comparing the increasingly accurate B-spline models with the fourth-degree polynomial surface model of Segment I by means of both hypothesis tests, as previous predicted, we find that the Cox's test always rejects the models with fewer parameters, which leads to the problem of over-fitting. By contrast, Vuong's test initially tends to prefer B-spline models due to higher approximation quality until Pair 36, in which large quantities of parameters are set ($N_B = 1600$). Table 3 offers the last five comparison pairs to show the aforementioned change in test decision. The complete test results for discriminating between fourth-degree polynomial and B-spline surface models are shown in Table A1.

Table 2. Results for Segment I of Cox's test for discriminating between polynomial and B-spline surface models at type-I error rate $\alpha = 0.05$.

Pair	Cox's Test				Vuong's Test	
	T_0	Rejected	T_1	Rejected	T_V	Preferred
I	−39.93	polynomial	−23.44	no	−29.95	B-spline
II	−0.68	no	1.44	no	0.19	no
III	−14.85	polynomial	−1.21	no	0.37	no

Table 3. Partial results for Segment I of Vuong's test and Cox's test for discriminating between fourth-degree polynomial and B-spline surface models at type-I error rate $\alpha = 0.05$.

Pair	Competing Models		Cox's Test				Vuong's Test	
	Polynomial	B-Spline	T_0	Rejected	T_1	Rejected	T_V	Preferred
32	$N_{poly} = 15$	$N_B = 1296$	−236.65	polynomial	−70.33	no	−8.99	B-spline
33	$N_{poly} = 15$	$N_B = 1369$	−213.58	polynomial	−72.86	no	−9.28	B-spline
34	$N_{poly} = 15$	$N_B = 1444$	−199.14	polynomial	−70.94	no	−5.00	B-spline
35	$N_{poly} = 15$	$N_B = 1521$	−192.05	polynomial	−84.49	no	−3.14	B-spline
36	$N_{poly} = 15$	$N_B = 1600$	−180.07	polynomial	−80.21	no	−0.36	no

The testing results for Segment II are listed in Table 4. It is indicated by the first comparison pair that both the polynomial and the B-spline surface models are rejected by Cox's test, while Vuong's test considers the polynomial model to be closer to the truth than its competitor. Within the second and third pairs, B-spline models are judged as insufficient by Cox's test, whereas the polynomial models are preferred by Vuong's test.

Table 4. Results for segment II of Cox's test for discriminating between polynomial and B-spline surface models at type-I error rate $\alpha = 0.05$.

Pair	Cox's Test				Vuong's Test	
	T_0	Rejected	T_1	Rejected	T_V	Preferred
I	−3.96	polynomial	4.46	B-spline	5.95	polynomial
II	−1.58	no	5.89	B-spline	7.43	polynomial
III	0.87	no	10.65	B-spline	11.31	polynomial

The testing result differs greatly between the two segments. These differences can be explained by the fact that Segment II contains large data gaps, which substantially distort the B-spline model estimation, in contrast to the polynomial model estimation.

3.2. Evaluation of Competing B-Spline Models with Various Parameters

The results in Table 3 motivate us to evaluate B-spline surface models with increasing parameters by means of Vuong's test in search of a balance between model complexity and its approximation

quality. Two B-spline models with different degrees or control points (Model I and Model II) are non-nested because the parameters in Model I are not a subset of the parameters in Model II. The modification of the degree or the number of control points (see Equation (1)) leads to a change in the number of knots, resulting in different basis functions. The comparison of pair setting and evaluation results are shown in Table A2, where $N_{B1} = (m_1 + 1) \cdot (n_1 + 1)$ denotes the parameter number of the first B-spline model and $N_{B2} = (m_2 + 1) \cdot (n_2 + 1)$ denotes that of the second model in the competing pair. The statistic value and test decision are shown in the last two columns. Here, we present the comparison results of B-spline models with $m_1 = n_1 = i$ and $m_2 = n_2 = i + 2$ ($i = 1, 2, 3 \dots 37$) instead of neighbor models ($m_1 = n_1 = i, m_2 = n_2 = i + 1$). There were large oscillations in the testing results of the pairwise neighbor models. These oscillations, which were due to the similar parameter number and model quality between neighbor models, served as noise and would confuse the result. The testing results became stable when we chose comparison models as $m_1 = n_1 = i$ and $m_2 = n_2 = i + 2$.

According to the results in Table A2, B-Spline Model II with more parameters is preferred initially; however, due to the increasing penalized term, B-Spline Model I is preferred in the final pair. The middle comparison pairs are considered to be in the overlapping region, where the test decisions swing between the two competing models. Table 5 shows the comparison pairs in the overlapping region. It can be considered that the balance sought of the model's complexity and its approximation quality is located in this region. Figure 5 offers a direct view of test statistic T_V in comparison with critics boundaries $C_{0.025}^{N(0,1)} = -1.96$ and $C_{0.975}^{N(0,1)} = 1.96$.

Table 5. Partial results of Vuong's test for discriminating B-spline surface models at type-I error rate $\alpha = 0.05$ (overlapping region).

Pair	B-Spline Model I		B-Spline Model II		Vuong's Test	
	m_1, n_1	N_{B1}	m_2, n_2	N_{B2}	T_V	Preferred
13	$m_1 = n_1 = 14$	225	$m_2 = n_2 = 16$	289	−0.57	no
14	$m_1 = n_1 = 15$	256	$m_2 = n_2 = 17$	324	−6.57	model 2
15	$m_1 = n_1 = 16$	289	$m_2 = n_2 = 18$	361	−4.06	model 2
16	$m_1 = n_1 = 17$	324	$m_2 = n_2 = 19$	400	−3.89	model 2
17	$m_1 = n_1 = 18$	361	$m_2 = n_2 = 20$	441	−4.15	model 2
18	$m_1 = n_1 = 19$	400	$m_2 = n_2 = 21$	484	3.02	model 1
19	$m_1 = n_1 = 20$	441	$m_2 = n_2 = 22$	529	9.09	model 1
20	$m_1 = n_1 = 21$	484	$m_2 = n_2 = 23$	576	0.41	no

We consider the B-spline model with 361 parameters ($n = 18, m = 18$), which lies roughly in the middle of the overlapping region, as the optimal one. Figure 6 shows the side- and top-view of this surface model.

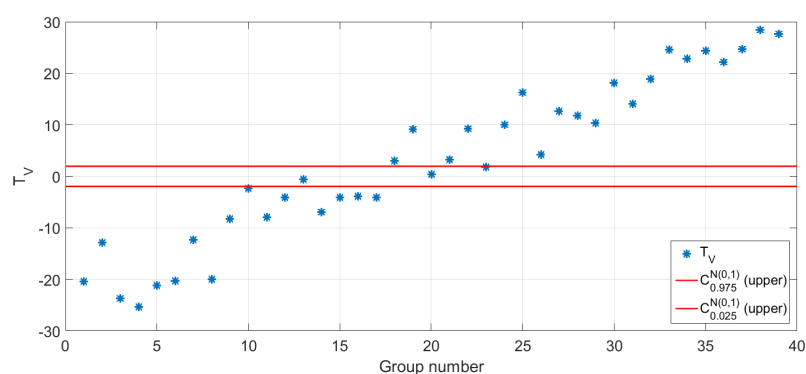


Figure 5. Statistic values of Vuong's test in comparison with critics.

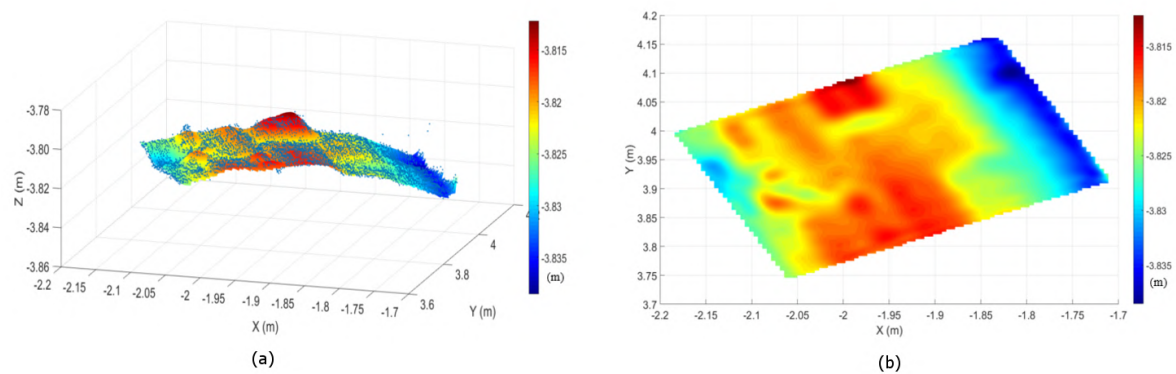


Figure 6. Side-view (a) and top-view (b) of approximated B-spline surface ($n = 18, m = 18$) with the original measurements (blue points).

3.3. Performance in Deformation Analysis

3.3.1. Deformation of Segment I

According to the model evaluation results for Segment I, the fourth-degree model is best-fitting among polynomial surface models, while the B-spline surface model with 361 parameters is optimal among all candidate models. We approximate both types of surface models for the point cloud in the 13th epoch with the same number of parameters.

Because traditional polynomial models are still the most widely used regression method in deformation analysis due to their simple operating, while the B-spline model has the potential to describe geometrically complicated objects, it is of significance to compare the performance in this numerical example between two surface models with their best-fitting parameters in reflecting deformation. In Figure 7a shows the modeled fourth-degree polynomial surfaces for the 1st (upper) and 13th (lower) epochs, while Figure 7b shows the equivalent epochs approximated by means of B-spline surfaces with 361 parameters. It is obvious that the B-spline surfaces in Figure 7b describe more detailed geometrical features than polynomial model in Figure 7a. The arch's deformation in Z-coordinates between the two epochs are shown in Figures 7c,d, by means of approximated polynomial surfaces of Figure 7a and B-spline models of Figure 7b, respectively. In order to validate that the reflected changes are the real arch's deformation recorded by the points instead of regression models, we compare the two epochs' point cloud through the block-mean approach (see Figure 8 for the differences between the two epoch's point clouds). By comparison, it is obvious that the deformation shown in Figure 7d for B-spline models reflect these differences precisely, especially in Areas A and B; in contrast, the fourth-degree polynomial surfaces in Figure 7c fail to show this deformation due to their global smoothing effect. The preceding difference and model deformation are also shown pointwise in Figure 9. The green asterisks denote the point-wise differences recorded by block-means, while the red and blue asterisks are that reflected by the fourth-degree polynomial and B-spline surfaces, respectively.

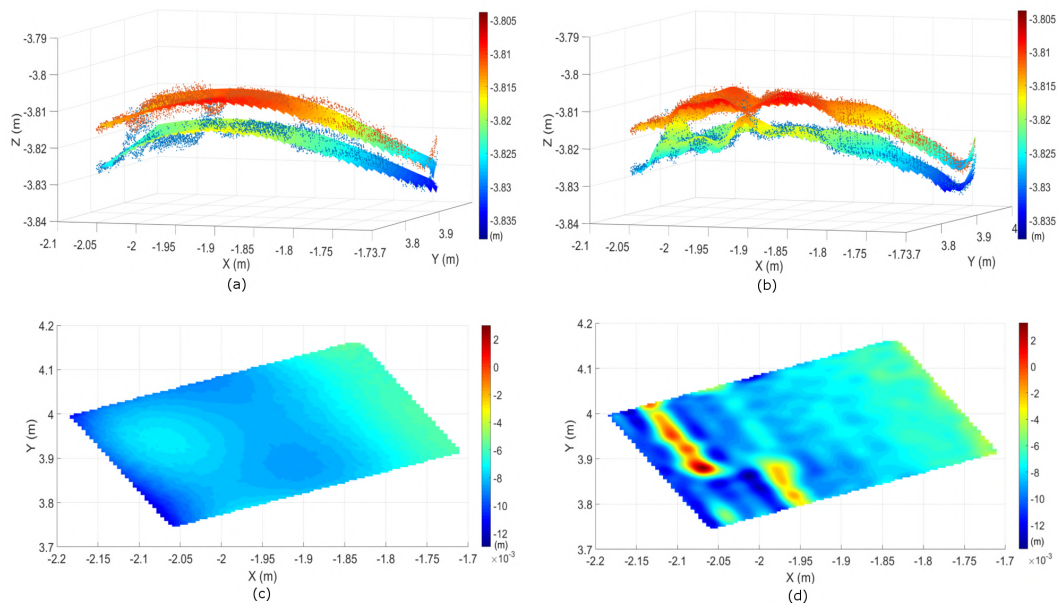


Figure 7. Polynomial (a,c) and B-spline surface models (b,d) in terms of differences of the 1st and 13th epochs in Segment I.

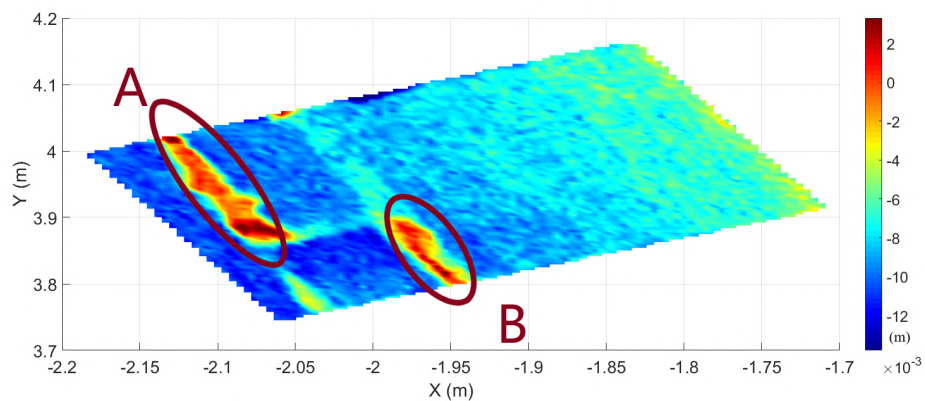


Figure 8. Deformation of segment I reflected by block means of the point cloud differences based on the 1st and 13th epochs.

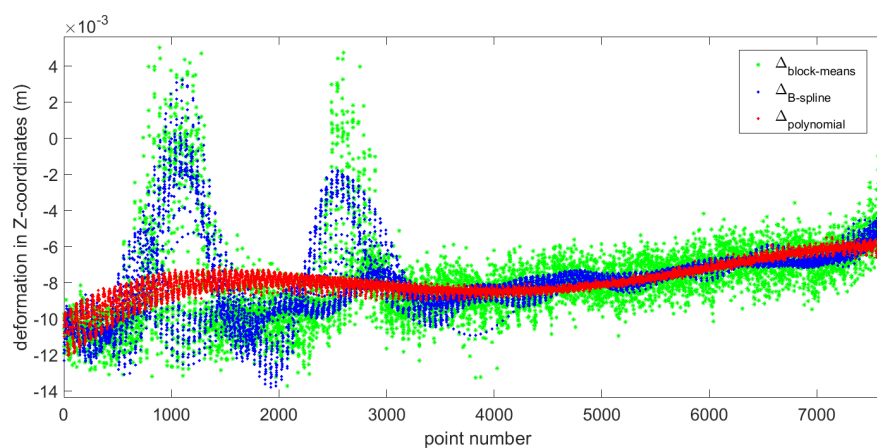


Figure 9. Deformation of Segment I between 1st and 13th epochs reflected by various approaches.

3.3.2. Deformation of Segment II

In Figure 10a shows the modeled fourth-degree polynomial surfaces for the 1st (upper) and 13th (lower) epochs, while Figure 10b shows the same epochs approximated by means of B-spline surfaces with 16 parameters.

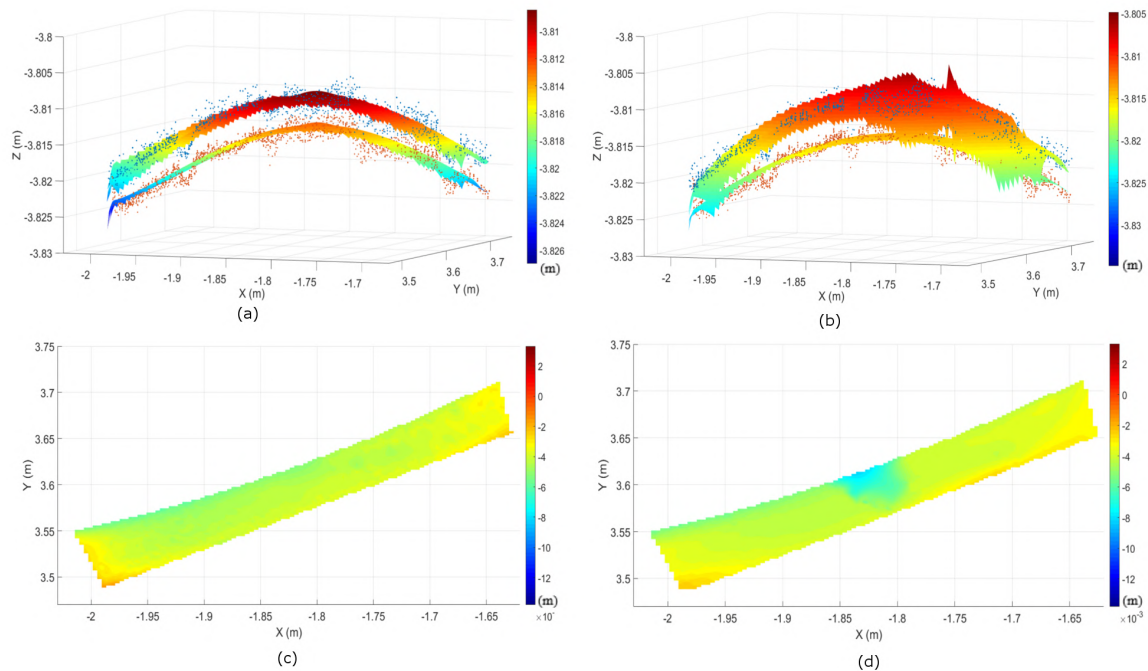


Figure 10. Polynomial (a,c) and B-spline surface models (b,d) in reflecting deformation of segment II based on the 1st and 13th epochs.

The missing data lead to oscillation, especially at the edges of the data gap (see the bounded area in Figure 11). Thus, it can be found from Figure 10d that the deformation reflected by B-splines is far from the truth. Here, the estimated B-spline surfaces clearly show the aforementioned numerical instabilities, which are caused by an inadequate specification of the knot locations with the applied classical approach to knot vector determination in Piegl and Tiller [29].

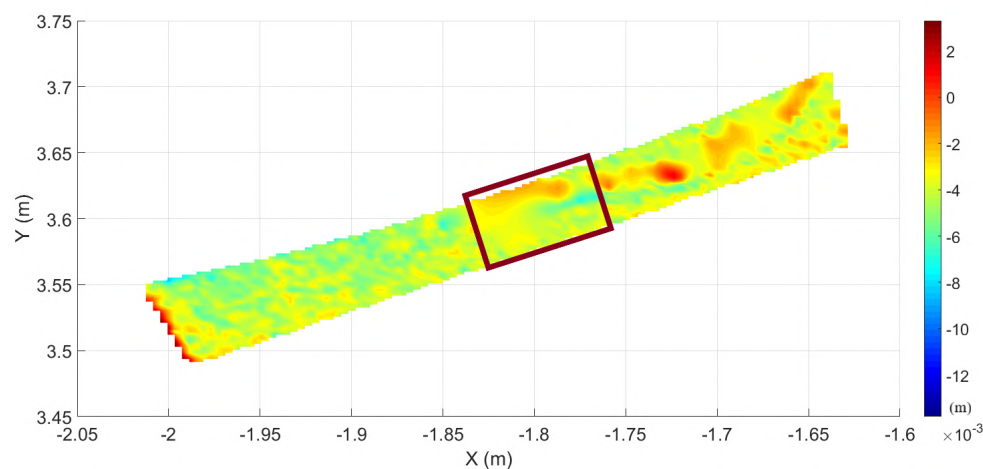


Figure 11. Deformation of Segment II reflected by block means of the point cloud differences based on the 1st and 13th epochs.

4. Discussion

In our numerical example of Segment I, the different results between the two hypothesis tests are caused by the penalized term regarding the parameter numbers. In Table 2, for example, Pair III has various test results. In Cox's test, the fourth-degree polynomial model is rejected because of the relatively poor accuracy, while the Vuong's test result recommends neither, because the improved accuracy is offset by the punishment of increasing parameters. In parallel, in Table 3, the test decision initially shows in the consistency of both tests that the B-spline models are better compared to the fourth-degree polynomial model. However, as the number of parameters increase, the improvement of model accuracy declines. Finally, in Vuong's test, the advantage of the model's quality is offset again by the large penalized value and, consequently, shows results that are different from Cox's test in Pair 36.

Although Cox's test without penalized term is limited to discriminate models with similar parameters, it is practically very straightforward to implement and are able to offer more reliable decisions by simulating the test distribution, especially when the sample size is small [33]. We expect to improve the simulation-based version of Cox's test by adding a proper correction factor similar to that in Vuong's non-nested hypothesis test, which would be one of our future research projects.

Since previous geodetic literatures [21,22,28] has solved the model selection problem through well-known penalization information criteria: the AIC and BIC, it is necessary in this section to compare Vuong's non-nested hypothesis test with this widely used approach. It is noticeable that there are close connections between AIC, BIC, and Vuong's test. Taking the BIC as an example, the value of model 1 is calculated as

$$BIC_1 = -2 \ln L_1 - p_1 \cdot \ln N \quad (22)$$

where L_1 is the maximum value of the likelihood function for Model 1, p_1 denotes the parameter quantity, and N is the number of measurements. The different BIC value between two models is calculated as

$$BIC_{12} = -2 \ln \frac{L_1}{L_2} - (p_1 - p_2) \cdot \ln N \quad (23)$$

where the first term in the right part contains logarithmized likelihood ratio $L_{0,1}$ in Equation (16), so that BIC_{12} is equal to the (un-normalized) adjusted test statistic $\tilde{L}_{0,1}$ for Vuong's test. The main difference is that Vuong's test makes judgments in a framework of likelihood ratio hypothesis testing, which offers the advantage that significant probabilistic differences between models can be detected, which is not provided by classical penalization information criterion methods. We compared the Vuong's test results with both the AIC and BIC to discriminate between B-spline surface models, and the result is shown in Figure 12. According to the BIC's curve, the B-spline model with 361 parameters ($n = 18, m = 18$) is optimal, since it is associated with the smallest value. This result is quite consistent with the judgment of Vuong's test, because the BIC penalized term is used in our adjusted test statistic. By contrast, the AIC tends to prefer much larger models.

Furthermore, the performance of best-fitting polynomial and B-spline surfaces in reflecting deformation were compared. The superior model was the one able to reflect the deformation details recorded by the point clouds. In order to get an exact mutual spatial referencing of points in the different epochs, we used the block-mean approach to approximate the point-wise changes. The comparison results of Segment I indicated that the selected B-spline surfaces can reflect the actual deformation, especially in Areas A and B of Figure 8, while the best-fitting polynomial model failed to offer this information due to its global smooth effect. However, in the case of Segment II, B-spline models failed to reflect the actual deformation values, especially at the edges of the data gap.

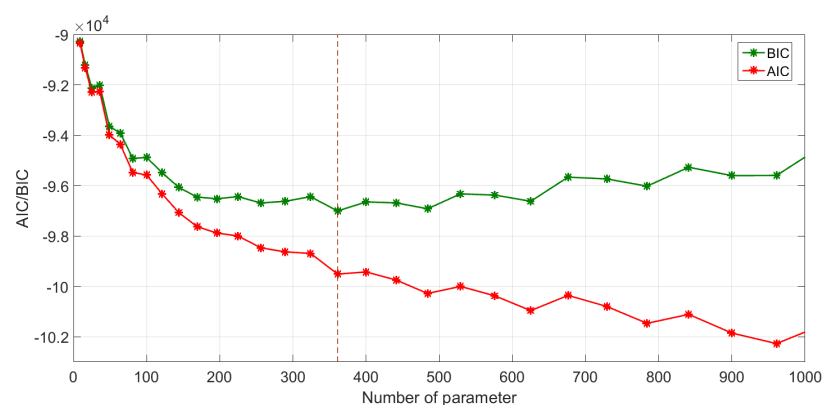


Figure 12. AIC (red) and BIC (green) values with an increasing number of parameters.

5. Conclusions

In this paper, we approximated point cloud data of a surveyed arch structure by two common surface models: polynomials and B-splines. Subsequently, we compared different adjusted surface models via Cox's test and Vuong's test to select an appropriate parametric model, which was sufficient to describe the geometrical features of the two target segments.

Regarding Segment I, in the initial comparison between lower degree polynomial and B-spline models, none of the B-spline models investigated was rejected, but only the polynomial model with degree 3 was found to be adequate in Cox's test, while Vuong's test indicated no significant difference in Pairs II and III. However, none of these models could reflect detailed geometrical features of target segments. Since it was not possible to increase the degree of polynomial approximation (due to numerical instability of the normal equations) for modeling geometrical details, B-splines were recommended in the field of applications presented. That motivated us to search for an optimal model balancing between approximation quality and its complexity. According to Vuong's test decisions, the B-spline surface model with $N_B = 361$ was considered as the optimal one in the specific numerical example.

The model selection testing results of Segment II were quite different from that of Segment I. All the B-spline models were rejected by Cox's test, while in Pairs II and III, the equivalent polynomial surfaces were preferred by Vuong's test, as a consequence of the aforementioned numerical instabilities with the knot vector determination and the resulting oscillation effects. Such deficiencies were clearly reflected by the model selection tests, which rejected inadequate B-spline models.

A consistent model selection result was obtained by comparing Vuong's test decision with the widely used BIC in discriminating B-spline surface models. Thus, it is concluded that the alternative model selection methodology elaborated in this paper, in parallel with well-known penalization information criteria, can effectively guide practitioners in selecting a parsimonious and accurate model for structures, such as the arch in the numerical example presented. The main difference is that Vuong's test makes judgments in a framework of likelihood ratio hypothesis testing, which can detect the significant probabilistic differences between models. It was proved here that the models selected by the model selection tests have good performance in reflecting actual deformation.

The model selection methodology is applicable not only to TLS data but also to point clouds obtained by other LiDAR technology, such as airborne laser scanning and mobile laser scanning. There are also distribution-free hypothesis tests, such as Clarke's test [34], available for mixed distribution observations.

Acknowledgments: The publication of this article was funded by the Open Access fund of Leibniz Universität Hannover.

Author Contributions: Xin Zhao conducted the statistical analysis and wrote the manuscript; Boris Kargoll contributed to the analysis approach and provide advice with respect to the manuscript; Mohammad

Omidalizarandi contributed to the data processing; Xiangyang Xu helped to conceive the study and motivate the paper; Hamza Alkhatib organized the study and commented on the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AIC	Akaike information criterion
BIC	Bayesian information criterion
KLIC	Kullback-Leibler information criterion
MSS	multi-sensor-system
RMSD	root-mean-square deviation
TLS	terrestrial laser scanning

Appendix A

Table A1. Results for Segment I of Vuong’s test and Cox’s test for discriminating between fourth-degree polynomial and B-spline surface models at type-I error rate $\alpha = 0.05$.

Pair	Competing Models		Cox’s Test				Vuong’s Test	
	Polynomial	B-Spline	T_0	Rejected	T_1	Rejected	T_V	Preferred
1	$N_{poly} = 15$	$N_B = 25$	−96.88	polynomial	−6.31	no	−14.92	B-spline
2	$N_{poly} = 15$	$N_B = 36$	−82.69	polynomial	−5.30	no	−14.59	B-spline
3	$N_{poly} = 15$	$N_B = 49$	−210.65	polynomial	−17.54	no	−13.99	B-spline
4	$N_{poly} = 15$	$N_B = 64$	−324.26	polynomial	−19.77	no	−13.37	B-spline
5	$N_{poly} = 15$	$N_B = 81$	−306.37	polynomial	−31.21	no	−12.58	B-spline
6	$N_{poly} = 15$	$N_B = 100$	−323.49	polynomial	−28.37	no	−12.75	B-spline
7	$N_{poly} = 15$	$N_B = 121$	−338.20	polynomial	−33.64	no	−13.23	B-spline
8	$N_{poly} = 15$	$N_B = 144$	−359.43	polynomial	−38.19	no	−11.79	B-spline
9	$N_{poly} = 15$	$N_B = 169$	−325.74	polynomial	−42.33	no	−10.44	B-spline
10	$N_{poly} = 15$	$N_B = 196$	−341.49	polynomial	−41.76	no	−9.68	B-spline
11	$N_{poly} = 15$	$N_B = 225$	−294.75	polynomial	−48.19	no	−11.53	B-spline
12	$N_{poly} = 15$	$N_B = 256$	−333.69	polynomial	−50.22	no	−10.59	B-spline
13	$N_{poly} = 15$	$N_B = 289$	−279.57	polynomial	−46.54	no	−11.23	B-spline
14	$N_{poly} = 15$	$N_B = 324$	−318.33	polynomial	−51.20	no	−10.02	B-spline
15	$N_{poly} = 15$	$N_B = 361$	−317.28	polynomial	−52.92	no	−9.90	B-spline
16	$N_{poly} = 15$	$N_B = 400$	−275.54	polynomial	−59.22	no	−7.43	B-spline
17	$N_{poly} = 15$	$N_B = 441$	−276.37	polynomial	−52.10	no	−8.55	B-spline
18	$N_{poly} = 15$	$N_B = 484$	−313.79	polynomial	−66.44	no	−9.12	B-spline
19	$N_{poly} = 15$	$N_B = 529$	−262.48	polynomial	−54.70	no	−7.99	B-spline
20	$N_{poly} = 15$	$N_B = 576$	−270.20	polynomial	−56.54	no	−10.10	B-spline
21	$N_{poly} = 15$	$N_B = 625$	−285.65	polynomial	−49.27	no	−6.99	B-spline
22	$N_{poly} = 15$	$N_B = 676$	−268.69	polynomial	−68.30	no	−10.59	B-spline
23	$N_{poly} = 15$	$N_B = 729$	−267.04	polynomial	−62.88	no	−8.99	B-spline
24	$N_{poly} = 15$	$N_B = 784$	−241.21	polynomial	−67.39	no	−9.54	B-spline
25	$N_{poly} = 15$	$N_B = 841$	−268.29	polynomial	−68.11	no	−8.99	B-spline
26	$N_{poly} = 15$	$N_B = 900$	−280.65	polynomial	−75.17	no	−8.70	B-spline
27	$N_{poly} = 15$	$N_B = 961$	−217.72	polynomial	−69.20	no	−6.99	B-spline
28	$N_{poly} = 15$	$N_B = 1024$	−258.19	polynomial	−84.32	no	−7.37	B-spline
29	$N_{poly} = 15$	$N_B = 1089$	−227.56	polynomial	−80.89	no	−8.67	B-spline
30	$N_{poly} = 15$	$N_B = 1156$	−240.33	polynomial	−82.54	no	−9.77	B-spline
31	$N_{poly} = 15$	$N_B = 1225$	−232.69	polynomial	−72.30	no	−7.59	B-spline
32	$N_{poly} = 15$	$N_B = 1296$	−236.65	polynomial	−70.33	no	−8.99	B-spline
33	$N_{poly} = 15$	$N_B = 1369$	−213.58	polynomial	−72.86	no	−9.28	B-spline
34	$N_{poly} = 15$	$N_B = 1444$	−199.14	polynomial	−70.94	no	−5.00	B-spline
35	$N_{poly} = 15$	$N_B = 1521$	−192.05	polynomial	−84.49	no	−3.14	B-spline
36	$N_{poly} = 15$	$N_B = 1600$	−180.07	polynomial	−80.21	no	−0.36	no

Table A2. Results of Vuong’s test for discriminating B-spline surface models at type-I error rate $\alpha = 0.05$.

Pair	B-Spline Model I		B-Spline Model II		Vuong’s Test	
	m_1, n_1	N_{B1}	m_2, n_2	N_{B2}	T_V	Preferred
1	$m_1 = n_1 = 2$	9	$m_2 = n_2 = 4$	25	−20.46	model 2
2	$m_1 = n_1 = 3$	16	$m_2 = n_2 = 5$	36	−12.84	model 2
3	$m_1 = n_1 = 4$	25	$m_2 = n_2 = 6$	49	−23.72	model 2
4	$m_1 = n_1 = 5$	36	$m_2 = n_2 = 7$	64	−25.30	model 2
5	$m_1 = n_1 = 6$	79	$m_2 = n_2 = 8$	81	−21.23	model 2
6	$m_1 = n_1 = 7$	64	$m_2 = n_2 = 9$	100	−20.32	model 2
7	$m_1 = n_1 = 8$	81	$m_2 = n_2 = 10$	121	−12.28	model 2
8	$m_1 = n_1 = 9$	100	$m_2 = n_2 = 11$	144	−20.03	model 2
9	$m_1 = n_1 = 10$	121	$m_2 = n_2 = 12$	169	−8.31	model 2
10	$m_1 = n_1 = 11$	144	$m_2 = n_2 = 13$	196	−2.30	model 2
11	$m_1 = n_1 = 12$	169	$m_2 = n_2 = 14$	225	−7.92	model 2
12	$m_1 = n_1 = 13$	196	$m_2 = n_2 = 15$	256	−4.11	model 2
13	$m_1 = n_1 = 14$	225	$m_2 = n_2 = 16$	289	−0.57	no
14	$m_1 = n_1 = 15$	256	$m_2 = n_2 = 17$	324	−6.57	model 2
15	$m_1 = n_1 = 16$	289	$m_2 = n_2 = 18$	361	−4.06	model 2
16	$m_1 = n_1 = 17$	324	$m_2 = n_2 = 19$	400	−3.89	model 2
17	$m_1 = n_1 = 18$	361	$m_2 = n_2 = 20$	441	−4.15	model 2
18	$m_1 = n_1 = 19$	400	$m_2 = n_2 = 21$	484	3.02	model 1
19	$m_1 = n_1 = 20$	441	$m_2 = n_2 = 22$	529	9.09	model 1
20	$m_1 = n_1 = 21$	484	$m_2 = n_2 = 23$	576	0.41	no
21	$m_1 = n_1 = 22$	529	$m_2 = n_2 = 24$	625	3.21	model 1
22	$m_1 = n_1 = 23$	576	$m_2 = n_2 = 25$	676	9.23	model 1
23	$m_1 = n_1 = 24$	625	$m_2 = n_2 = 26$	729	1.97	model 1
24	$m_1 = n_1 = 25$	676	$m_2 = n_2 = 27$	784	10.03	model 1
25	$m_1 = n_1 = 26$	729	$m_2 = n_2 = 28$	841	16.27	model 1
26	$m_1 = n_1 = 27$	784	$m_2 = n_2 = 29$	900	4.23	model 1
27	$m_1 = n_1 = 28$	841	$m_2 = n_2 = 30$	961	12.7	model 1
28	$m_1 = n_1 = 29$	900	$m_2 = n_2 = 31$	1024	11.81	model 1
29	$m_1 = n_1 = 30$	961	$m_2 = n_2 = 32$	1089	10.33	model 1
30	$m_1 = n_1 = 31$	1024	$m_2 = n_2 = 33$	1156	18.17	model 1
31	$m_1 = n_1 = 32$	1089	$m_2 = n_2 = 34$	1225	14.05	model 1
32	$m_1 = n_1 = 33$	1156	$m_2 = n_2 = 35$	1296	18.83	model 1
33	$m_1 = n_1 = 34$	1225	$m_2 = n_2 = 36$	1369	24.62	model 1
34	$m_1 = n_1 = 35$	1296	$m_2 = n_2 = 37$	1444	22.86	model 1
35	$m_1 = n_1 = 36$	1369	$m_2 = n_2 = 38$	1521	24.40	model 1
36	$m_1 = n_1 = 37$	1444	$m_2 = n_2 = 39$	1600	22.14	model 1
37	$m_1 = n_1 = 38$	1521	$m_2 = n_2 = 40$	1681	24.68	model 1
38	$m_1 = n_1 = 39$	1600	$m_2 = n_2 = 41$	1764	24.86	model 1
39	$m_1 = n_1 = 40$	1681	$m_2 = n_2 = 42$	1849	27.61	model 1

References

1. Han, J.Y.; Guo, J.; Jiang, Y.S. Monitoring tunnel profile by means of multi-epoch dispersed 3D LiDAR point clouds. *Tunn. Undergr. Space Technol.* **2013**, *33*, 186–192.
2. Mukupa, W.; Roberts, G.W.; Hancock, C.M.; Al-Manasir, K. A review of the use of terrestrial laser scanning application for change detection and deformation monitoring of structures. *Surv. Rev.* **2017**, *49*, 99–116.
3. Gordon, S.J.; Lichti, D.D. Modeling terrestrial laser scanner data for precise structural deformation measurement. *J. Surv. Eng.* **2007**, *133*, 72–80.
4. Lovas, T.; Barsi, A.; Detrekoi, A.; Dunai, L.; Csak, Z.; Polgar, A.; Berenyi, A.; Kibedy, Z.; Szocs, K. Terrestrial laser scanning in deformation measurements of structures. *Int. Arch. Photogramm. Remote Sens.* **2008**, *37*, 527–531.
5. Monserrat, O.; Crosetto, M. Deformation measurement using terrestrial laser scanning data and least squares 3D surface matching. *ISPRS J. Photogramm. Remote Sens.* **2008**, *63*, 142–154.

6. Tsakiri, M.; Lichti, D.; Pfeifer, N. Terrestrial laser scanning for deformation monitoring. In Proceedings of the 12th FIG Symposium, Baden, Germany, 22–24 May 2006.
7. Lane, S.N.; Westaway, R.M.; Murray Hicks, D. Estimation of erosion and deposition volumes in a large, gravel-bed, braided river using synoptic remote sensing. *Earth Surf. Process. Landf.* **2003**, *28*, 249–271.
8. Barnhart, T.B.; Crosby, B.T. Comparing Two Methods of Surface Change Detection on an Evolving Thermokarst Using High-Temporal-Frequency Terrestrial Laser Scanning, Selawik River, Alaska. *Remote Sens.* **2013**, *5*, 2813–2837.
9. Yang, H.; Xu, X.; Neumann, I. Deformation behavior analysis of composite structures under monotonic loads based on terrestrial laser scanning technology. *Compos. Struct.* **2018**, *183*, 594–599.
10. Ingensand, H.; Ioannidis, C.; Valani, A.; Georgopoulos, A.; Tsiligiris, E. 3D model Generation for Deformation Analysis using Laser Scanning data of a Cooling Tower. In Proceedings of the 12th FIG Symposium, Baden, Germany, 22–24 May 2006.
11. Van Gosliga, R.; Lindenbergh, R.; Pfeifer, N. Deformation analysis of a bored tunnel by means of terrestrial laser scanning. In Proceedings of the International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, Dresden, Germany, 25–27 September 2006.
12. Vežočanik, R.; Ambrožič, T.; Sterle, O.; Bilban, G.; Pfeifer, N.; Stopar, B. Use of terrestrial laser scanning technology for long term high precision deformation monitoring. *Sensors* **2009**, *9*, 9873–9895.
13. You, L.; Tang, S.; Song, X.; Lei, Y.; Zang, H.; Lou, M.; Zhuang, C. Precise Measurement of Stem Diameter by Simulating the Path of Diameter Tape from Terrestrial Laser Scanning Data. *Remote Sens.* **2016**, *8*, 717.
14. Lindenbergh, R.; Uchanski, L.; Bucksch, A.; Van Gosliga, R. Structural monitoring of tunnels using terrestrial laser scanning. *Rep. Geod.* **2009**, *87*, 231–238.
15. Girardeau-Montaut, D.; Roux, M.; Marc, R.; Thibault, G. Change detection on points cloud data acquired with a ground laser scanner. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2005**, *36*, 30–35.
16. Yang, H.; Omidalizand, M.; Xu, X.; Neumann, I. Terrestrial laser scanning technology for deformation monitoring and surface modeling of arch structures. *Compos. Struct.* **2017**, *169*, 173–179.
17. Bureick, J.; Alkhatib, H.; Neumann, I. Robust Spatial Approximation of Laser Scanner Point Clouds by Means of Free-form Curve Approaches in Deformation Analysis. *J. Appl. Geod.* **2016**, *10*, 27–35.
18. Harmening, C.; Neuner, H. A constraint-based parameterization technique for B-spline surfaces. *J. Appl. Geod.* **2015**, *9*, 143–161.
19. Zhao, X.; Alkhatib, H.; Kargoll, B.; Neumann, I. Statistical evaluation of the influence of the uncertainty budget on B-spline curve approximation. *J. Appl. Geod.* **2017**, *11*, 215–230.
20. Wunderlich, T.; Niemeier, W.; Wujanz, D.; Holst, C.; Neitzel, F.; Kuhlmann, H. Areal deformation analysis from TLS point clouds—The challenge. *Allgemeine Vermessungs Nachrichten* **2016**, *123*, 340–351.
21. Harmening, C.; Neuner, H. Choosing the Optimal Number of B-spline Control Points (Part 1: Methodology and Approximation of Curves). *J. Appl. Geod.* **2016**, *10*, 139–157.
22. Harmening, C.; Neuner, H. Choosing the optimal number of B-spline control points (Part 2: Approximation of surfaces and applications). *J. Appl. Geod.* **2017**, *11*, 43–52.
23. Cox, D.R. Tests of Separate Families of Hypotheses. In Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Berkeley, CA, USA, 20 June–30 July 1961.
24. Williams, D.A. Discrimination between Regression Models to Determine the Pattern of Enzyme Synthesis in Synchronous Cell Cultures. *Biometrics* **1970**, *26*, 23–32.
25. Vuong, Q.H. Likelihood ratio tests for model selection and non-nested hypotheses. *Econ. J. Econom. Soc.* **1989**, *8*, 307–333.
26. Hexagon Metrology. Leica Absolute Tracker AT960 Brochure. Available online: <http://www.hexagonmi.com/products/laser-tracker-systems/leica-absolute-tracker-at960#loren> (accessed on 2015).
27. Meierhold, N.; Spehr, M.; Schilling, A.; Gumhold, S.; Maas, H. Automatic feature matching between digital images and 2D representations of a 3D laser scanner point cloud. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2010**, *38*, 446–451.
28. Bureick, J.; Neuner, H.; Harmening, C.; Neumann, I. Curve and Surface Approximation of 3D Point Clouds. *Allgemeine Vermessungs-Nachrichten* **2016**, *123*, 315–327.
29. Piegl, L.; Tiller, W. *The NURBS Book*; Springer Science & Business Media: Berlin, Germany, 1997.
30. Koch, K. Fitting Free-Form Surfaces to Laserscan Data by NURBS. *Allgemeine Vermessungs-Nachrichten* **2009**, *116*, 134–140.

31. Koch, K. *Parameter Estimation and Hypothesis Testing in Linear Models*; Springer: Berlin, Germany, 1999.
32. Paffenholz, J.A.; Vennegeerts, H.; Kutterer, H. High frequency terrestrial laser scans for monitoring kinematic processes. In Proceedings of the INGeo 2008—4th International Conference on Engineering Surveying, Bratislava, Slovakia, 23–24 October 2008.
33. Lewis, F.; Butler, A.; Gilbert, L. A unified approach to model selection using the likelihood ratio test. *Methods Ecol. Evol.* **2011**, *2*, 155–162.
34. Clarke, K.A. A simple distribution-free test for nonnested model selection. *Polit. Anal.* **2007**, *15*, 347–363.



© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

A Bootstrap Approach to Testing for Time-Variability of AR Process Coefficients in Regression Time Series with t-Distributed White Noise Components

Hamza Alkhatib, Mohammad Omidalizarandi, and Boris Kargoll

Abstract

In this paper, we intend to test whether the random deviations of an observed regression time series with unknown regression coefficients can be described by a covariance-stationary autoregressive (AR) process, or whether an AR process with time-variable (say, linearly changing) coefficients should be set up. To account for possibly present multiple outliers, the white noise components of the AR process are assumed to follow a scaled (Student) t-distribution with unknown scale factor and degree of freedom. As a consequence of this distributional assumption and the nonlinearity of the estimator, the distribution of the test statistic is analytically intractable. To solve this challenging testing problem, we propose a Monte Carlo (MC) bootstrap approach, in which all unknown model parameters and their joint covariance matrix are estimated by an expectation maximization algorithm. We determine and analyze the power function of this bootstrap test via a closed-loop MC simulation. We also demonstrate the application of this test to a real accelerometer dataset within a vibration experiment, where the initial measurement phase is characterized by transient oscillations and modeled by a time-variable AR process.

Keywords

Bootstrap test · EM algorithm · Monte Carlo simulation · Regression time series · Scaled t-distribution · Time-variable autoregressive process

1 Introduction

Reliable and precise estimation of geodetic time series models remains a challenging task as they frequently involve huge numbers of auto-correlated and outlier-afflicted measurements. On the one hand, a parsimonious model that allows both for the description and the estimation of auto-correlations is given by autoregressive (AR) processes (cf. Schuh 2003). On the other hand, a flexible approach to modeling multiple outliers (or more generally a heavy-tailed error

law) is enabled by the assumption that the random deviations follow a scaled t-distribution (cf. Koch and Kargoll 2013).

Since adjustment techniques based on least squares are sensitive to misspecifications of the functional and stochastic observation model (cf. Kutterer 1999), as well as sensitive to outliers (cf. Baarda 1968), frequently encountered data features such as functional non-linearity, colored measurement noise and heavy-tailed error distribution should be adequately taken into account. Modern geodetic sensors often involve a data sampling at a high rate, thus producing significantly auto-correlated noise (cf. Kuhlmann 2001), in potentially huge numbers of observations. In such cases, the use of a covariance matrix easily exceeds the memory of the computer. Instead, an AR process can often be used for modeling (auto-)correlations more parsimoniously (cf. Schuh 2003). Moreover, the error law of geodetic measurements has frequently been found to be heavy-tailed, in which cases

H. Alkhatib (✉) · M. Omidalizarandi
Geodetic Institute, Leibniz University Hannover, Hannover, Germany
e-mail: alkhatib@gih.uni-hannover.de

B. Kargoll
Institut für Geoinformation und Vermessung Dessau, Anhalt
University of Applied Sciences, Dessau-Roßlau, Germany

robust M-estimation can be applied to the aforementioned models (cf. Wiśniewski 2014).

Kargoll et al. (2018a) recently dealt with the case where both the coefficients of the AR model of the random deviations in a linear functional model and the shape parameters of the heavy-tailed error law are unknown. As suggested by Koch and Kargoll (2013) in a geodetic context, the family of scaled t-distributions was used to model the error law. Here, the degree of freedom is a shape parameter, which controls the thickness of the tails, and which can be estimated from the given measurements jointly with the other (functional and stochastic) model parameters, in the sense of a *self-tuning* robust estimator (cf. Parzen 1979). With this kind of estimator, the unknown parameters of the functional and the AR model can conveniently be computed via iteratively reweighted least squares (IRLS).

This is not only possible when the AR process is covariance-stationary, but also when the AR coefficients are modeled as time-variable quantities through a linear regression (Kargoll et al. 2018b). Such models have been found useful in describing non-stationary effects in time series measurements that cannot be properly described as part of the deterministic model at the level of the observation equations, e.g., local distortions in Gravity Field and Steady-State Ocean Circulation Explorer (GOCE) satellite gravity gradient data (Schuh and Brockmann 2016) or transient oscillations in terrestrial accelerometer measurements (Kargoll et al. 2018b). An important issue that has not been addressed in these expositions is how such an AR error model with t-distributed white noise components can be tested for time-variability (this testing problem is defined in Sect. 2). Whereas a corresponding suitable test statistic is easily identified from standard testing theory, the associated probability distribution is not readily available due to the complexity of the model.

With difficult-to-handle problems, Monte Carlo (MC) simulation may give an adequate solution (see, e.g., Koch 2018a,b). In particular, MC bootstrap approaches (in the sense of Efron 1979) can often be applied in situations where an approximate probability distribution of a test statistic cannot be derived analytically. Besides a few geodetic fields of application (e.g., Teunissen 1998; Neuner et al. 2014; Angrisano et al. 2018; Lösler et al. 2018), bootstrap methods have been devised for rather general time series models (cf. Li and Maddala 1996; Politis 2003) and in the context of Expectation Maximization (EM) algorithms for missing-data models (cf. McLachlan and Krishnan 2008). As our self-tuning robust estimator is based on both of these, we can develop a bootstrap approach to solving the aforementioned problem of testing for time-variability of AR coefficients. In Sect. 3 we outline this procedure in the general case of testing against a linearly modeled time-variability of the coefficients of an AR(p) process, employed to model the

random deviations of a possibly nonlinear regression time series. In Sect. 4, we simulate and investigate the size and power function of the test.

2 General Testing Problem

We assume that observations $\ell = [\ell_1, \dots, \ell_n]^T$ can be described by a nonlinear, vector-valued ('deterministic model') function f and random deviations $e = [e_1, \dots, e_n]^T$ through the so-called observation equations $\ell = f(\mathcal{E}) + e$, where $\mathcal{E} = [x_1, \dots, x_q]^T$ constitute unknown parameters. We assume the random deviations to follow an AR(p) process

$$e_t = \sum_{j=1}^p \alpha_{t,j} e_{t-j} + u_t \quad (t = 1, \dots, n) \quad (1)$$

with the time-dependent coefficients $\alpha_{t,j} = \mathbf{B}_t \mathbf{y}_j$ being described by linear combinations involving fixed vectors $\mathbf{B}_t$ and unknown parameters $\mathbf{y}_j = [y_{j,1}, \dots, y_{j,m}]^T$. We assume the random variables $\mathbf{u} = [u_1, \dots, u_n]^T$ in (1) to be independently and identically $t_\nu(0, \sigma^2)$ -distributed with center 0, unknown scale parameter σ^2 and unknown degree of freedom ν . It is well known that this stochastic model is equivalent to the model $u_t | w_t \sim N(0, \sigma^2/w_t)$ involving gamma-distributed latent variables $w_t \sim G(\nu/2)$ (cf. Koch and Kargoll 2013). Common designs of the time-variability models include polynomials defined by $\mathbf{B}_t = [1 \ \tau_t^1 \ \dots \ \tau_t^{m-1}]$. Clearly, time-dependency is eliminated either for $m = 1$ (in which case we may write $\alpha_{t,j} = y_{j,1} =: \alpha_j$), or for $y_{j,2} = \dots = y_{j,m} = 0$ ($j = 1, \dots, p$). If it is not known whether an AR model is time-variable or not, it makes more sense to keep these parameters $\mathbf{Y} = [y_{1,2}, \dots, y_{1,m}, \dots, y_{p,2}, \dots, y_{p,m}]^T$ in the model and to test their joint significance. For this purpose, we define the null and the alternative hypothesis by

$$H_0 : \mathbf{Y} = \mathbf{0} \quad \text{vs.} \quad H_1 : \mathbf{Y} \neq \mathbf{0}. \quad (2)$$

A natural choice for a test statistic that measures deviations from H_0 (i.e., the degree of $\mathbf{Y}$ being nonzero) in a cumulative manner is given by the weighted sum $T = \hat{\mathbf{Y}}^T \hat{\Sigma}_{\hat{\mathbf{Y}}}^{-1} \hat{\mathbf{Y}}$ of squares of the estimates of $\mathbf{Y}$, involving the corresponding joint (*a posteriori*) covariance matrix. In simple cases, it might be sufficient to test against a linear drift of an AR(1) process, i.e., to expand the time-variable AR(p) process up to $p = 1$ and $m = 2$, and to test $H_0 : y_2 = 0$ versus $H_1 : y_2 \neq 0$, writing more simply $\alpha_t = y_1 + \tau_t \cdot y_2$ instead of $\alpha_{t,1} = y_{1,1} + \tau_t \cdot y_{1,2}$. Then, the previous test statistic would simplify to $T = \hat{y}_2^2 / \hat{\sigma}_{y_2}^2$. However, even in this simple setting, it is not known how well the probability distribution of this

statistic can be approximated by a standard distribution (say, a t-distribution), due to the intricacy of the entire observation model. Therefore, we pursue the solution of this testing problem not via the derivation of an approximate distribution (which might even be intractable), but via the following bootstrap approach.

3 Bootstrap Approach to Solving the Testing Problem

Since the distribution of the test statistic about the time-variability parameters is unknown, we resort to simulation-based bootstrap testing. The general idea is to generate a large number of sampled values for the test statistic under the assumption that H_0 is true, and to define the critical value for the (one-sided) test as that value which is exceeded by only α % of the sampled values. This allows us then to assess whether the value of the test statistic computed from the given measurements is too large as to support H_0 , which is the case when that value exceeds the previously simulated critical value. The individual steps of this approach are outlined in the following for the general problem of testing against a time-variability model with specified polynomial order m and AR model order p with given measurement results ℓ , deterministic model $f(x)$, time-variability design matrix B and significance level α .

Estimation Step The modified EM algorithm described in Kargoll et al. (2018b, Section 3) was extended by the linearization procedure explained in Alkhatib et al. (2018, Section 3). This algorithm outputs the estimates $\hat{x}$ of the functional parameters, the estimates $\hat{y}$ of the parameters of the time-variable AR process alongside their joint covariance matrix $\hat{\Sigma}_{\hat{y}\hat{y}}$, the estimate $\hat{\sigma}^2$ of the scale factor as well as the estimate $\hat{\nu}$ of the degree of freedom of the underlying t-distribution, and estimates $\hat{u}$ of the white noise residuals.

Testing Step The value $T = \hat{Y}^T \hat{\Sigma}_{\hat{y}\hat{y}}^{-1} \hat{Y}$ of the test statistic is computed.

Generation Step The generation scheme begins with the sampling of the white noise components $u_t^{(k)}$ for the time instances $t = 1, \dots, n$ and the MC iterations $k = 1, \dots, B$ (where B is the total number of bootstrap samples). For this purpose, we consider the following two alternatives.

- (1) *Parametric bootstrapping*: random numbers with respect to the Student $t_{\hat{\nu}}(0, \hat{\sigma}^2)$ -distribution are independently generated for the white noise component $u_t^{(k)}$.
- (2) *Nonparametric bootstrapping*: random numbers $\lambda_t^{(k)}$ with respect to the discrete uniform distribution $\mathcal{U}(1, n)$ are drawn with replacement to define $u_t^{(k)} = \hat{u}_{\lambda_t^{(k)}}$

(thereby re-using the residuals $\hat{u}$ from the Estimation Step).

To ensure that the B measurement series are generated under H_0 , that is, under the assumption of a time-constant AR process, we take the time-wise means $\bar{\alpha}_j = \frac{1}{n} \sum_{t=1}^n \hat{\alpha}_{t,j}$ ($j = 1, \dots, p$) of the estimated AR coefficients $\hat{\alpha}_{t,j} = B_t \hat{y}_j$. For when H_0 is true, then all of the AR coefficients are constant throughout time, and the estimated means can be expected to approximate the true constant coefficient values. Now, the previously generated white noise components are correlated recursively through evaluation of the time-constant AR model equation $e_t^{(k)} = \sum_{j=1}^p \bar{\alpha}_j e_{t-j}^{(k)} + u_t^{(k)}$, using the initial values $e_0^{(k)} = \dots = e_{1-p}^{(k)} = 0$. Adding to these colored noise components the estimated deterministic model yields the sampled measurements $\ell_t^{(k)} = f_t(\hat{x}) + e_t^{(k)}$. The resulting measurement series $\ell^{(k)}$ is adjusted in exactly the same way as the actual measurement series ℓ within the previous Estimation Step, which gives the sampled estimates $\hat{x}^{(k)}$, $\hat{y}^{(k)}$, $\hat{\Sigma}_{\hat{y}\hat{y}}^{(k)}$, $(\hat{\sigma}^2)^{(k)}$ and $\hat{\nu}^{(k)}$. The sampled test statistic is obtained by $T^{(k)} = (\hat{Y}^{(k)})^T (\hat{\Sigma}_{\hat{y}\hat{y}}^{(k)})^{-1} \hat{Y}^{(k)}$, as in the Testing Step.

Evaluation Step To determine how extreme the test value T is in comparison to the values $T^{(1)}, \dots, T^{(B)}$ generated under H_0 , we estimate the p -value by

$$\widehat{pv} = \frac{1}{B} \sum_{k=1}^B I(T^{(k)} > T), \quad (3)$$

according to McKinnon (2007, Section 2). Here, I is the indicator function, which takes the value 1 in case the argument is true, and the value 0 if the argument is false.

Decision Step A large p -value indicates a rather large support of H_0 by the data. Thus, we reject H_0 if the estimated p -value is less than the predefined significance level α .

Note that when the random deviations of the measurements can safely be assumed to be normally distributed, then the bootstrap tests can be carried out in a similar way as described. The main differences are that

1. the degree of freedom of the t-distribution is not estimated but fixed a priori within the Estimation Steps by setting it to a large value, e.g. to 120, for which value the t-distribution closely approximates a normal distribution.
2. random numbers with respect to the normal distribution $\mathcal{N}(0, \hat{\sigma}^2)$ are independently generated for the white noise components within the parametric bootstrapping of the Generation Step.

The form of the test statistic $T = \hat{Y}^T \hat{\Sigma}_{\hat{y}\hat{y}}^{-1} \hat{Y}$, however, remains unchanged.

4 Monte Carlo Simulation

We consider the linear regression time series model

$$\ell_t = \frac{a_0}{2} + \sum_{j=1}^M a_j \cos(2\pi f_j \tau_t) + b_j \sin(2\pi f_j \tau_t) + e_t \quad (4)$$

($t = 1, \dots, n$), consisting of sine and cosine basis functions with unknown (“Fourier”) coefficients $a_0, a_1, \dots, a_M$ and $b_1, \dots, b_M$ (collected in the parameter vector $\mathbf{x}$), and of random deviations. The Fourier frequencies

$$f_j = 8 \cdot j \text{ [Hz]} \quad (j = 1, \dots, M) \quad (5)$$

are treated as error-free quantities. We set the number of frequencies in this simulation to $M = 12$ as a typical number encountered in vibration analysis experiments involving real accelerometer measurements (see Sect. 5 and Kargoll et al. 2018a). Furthermore, $\tau_1, \dots, \tau_n$ are given time instances sampled with constant sampling rate $\Delta\tau = 0.00512$ s, beginning at time $\tau_1 = 67.68128$ s. The number of observations is $n = 10,000$. This functional model is linear, so that the design matrices $\mathbf{A}_{[n \times 25]}$ is immediately obtained (without linearization). Concerning the colored noise e_t , we specified a time-variable AR(1)-process using the global polynomial $\alpha_t = y_1 + y_2 \cdot \tau_t$ of degree 1. For $y_2 = 0$, the AR(1)-process is time-constant. The simulation of the p -values ($\widehat{pv}^{(i)}$) and the power function consists for every repetition $i \in \{1, \dots, 100\}$ of the following steps:

- Generate the white noise $u_t^{(i)}$ from the true t-distribution $t_3(0, 10^{-6})$ for $t = 1, \dots, 10,000$.
- Correlate the white noise by means of the AR(1) process $e_t^{(i)} = \alpha_{t,1} e_{t-1}^{(i)} + u_t^{(i)}$ with $\alpha_{t,1} = y_1 + y_2 \cdot \tau_t$. We fix here the true offset parameter $y_1 = -0.5$, and we vary the true slope parameters y_2 in steps $[0 : 0.0001 : 0.003]$.
- Add generated colored noise to a specified oscillation model (4) to determine observations $\ell^{(i)}$.

- Do the Estimation Step to obtain $\hat{\mathbf{x}}^{(i)}, \hat{\mathbf{y}}^{(i)}, \hat{\Sigma}_{\hat{\mathbf{y}}\hat{\mathbf{y}}}^{(i)}, \hat{\sigma}_{\hat{\mathbf{y}}}^{(i)}, \hat{v}^{(i)}$ using exactly the same functional and time-variable AR model as described before.
- Carry out the Testing Step to compute $T^{(i)} = (\hat{y}_2^{(i)} / \hat{\sigma}_{\hat{y}_2}^{(i)})^2$.
- Carry out the Generation Step (parametric/nonparametric) to compute $T^{(i,k)}$ for $k = 1, \dots, B$. In order to demonstrate the performance of the bootstrap test, we use exactly the same functional and time-variable AR model as in the previous Estimation Step. In particular, the model orders p and m with respect to the, respectively, AR and time-variability model are maintained.
- Carry out the Evaluation Step to compute $\widehat{pv}^{(i)}$.

This simulation was carried out for the bootstrap sample sizes $B = 19, B = 99$ and $B = 999$. According to Davidson and MacKinnon (2000), B should be chosen such that $\alpha \cdot (B + 1)$ is an integer. In this paper we fixed $\alpha = 0.05$, so that $B = 19$ is the least possible value. Since the loss of power of a bootstrap test is proportional to $1/B$ according to Davidson and MacKinnon (2000), we investigated the largest number $B = 999$ possible with the computer hardware available in our experiment. With this number, we can already be quite sure that the resulting p -values are not overly dependent on the particular sequence of random numbers. The results of the estimated $\widehat{pv}^{(i)}$ -values for the case $y_2 = 0$ (i.e., a time-constant AR(1) process) are shown in Fig. 1. Somewhat surprisingly, varying B does not change the p -value very much within a MC run. Most importantly, the size $\alpha = 0.05$ is reproduced well on the average since the H_0 -rejection rate for all 100 replications, defined by

$$\frac{1}{100} \sum_{i=1}^{100} \mathbf{I}(\widehat{pv}^{(i)} < 0.05)$$

turned out to be 0.04 for parametric and 0.05 for non-parametric bootstrapping. Figure 2 shows the empirical power function (sensitivity of the hypothesis tests) evaluated at $y_2 = [0 : 0.0001 : 0.003]$. We see that a linear

Fig. 1 Comparison of estimated $\widehat{pv}^{(i)}$ -value (rejecting the null hypothesis whenever $\widehat{pv}^{(i)} < 0.05$) under the time-constant AR(1) model ($y_2 = 0$) for different bootstrap sample sizes

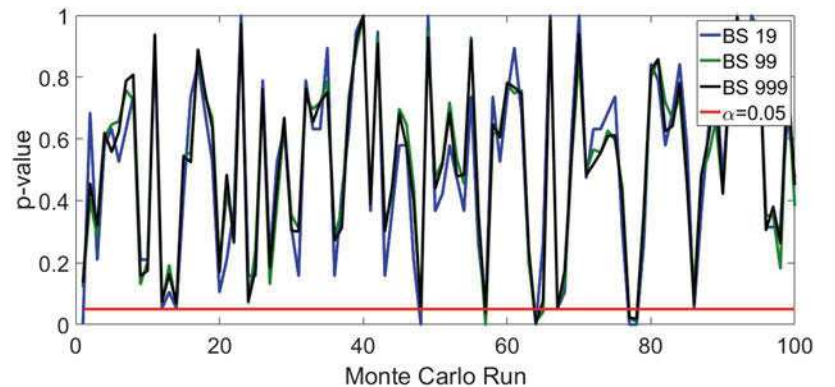
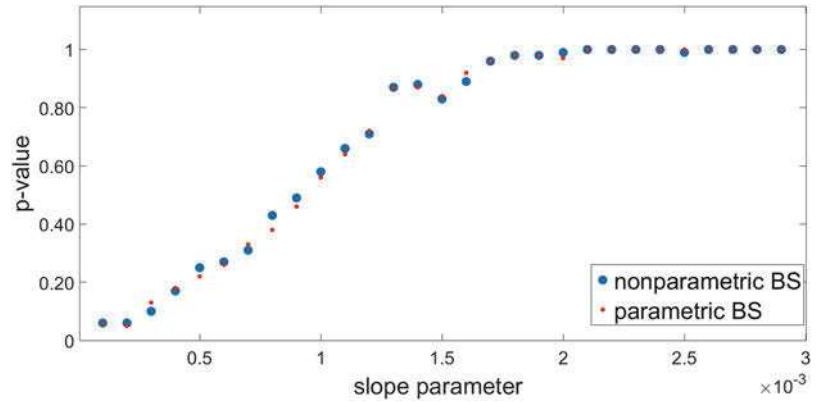


Fig. 2 Comparison of estimated $\hat{p}v^{(i)}$ -value (rejecting the null hypothesis whenever $\hat{p}v^{(i)} < 0.05$) under the AR(1) model with time variability parameter values $y_2 = [0 : 0.0001 : 0.003]$ for parametric and nonparametric bootstrapping



time-variability with a slope of 0.0015 is detected with a relative frequency of about 0.80. This slope value may be viewed as an empirical analogue to the size of an outlier detectable with a specified probability of 0.80 within Baarda's reliability theory.

5 An Application to Vibration Analysis

We applied the bootstrap approach to testing for time-variability of an AR process to a vibration dataset measured by means of a highly accurate single-axis PicoCoulomb (PCB) Piezotronics accelerometer. As part of a vibration experiment, carried out at the Institute of Concrete Construction at the Leibniz Universität Hannover, that sensor was mounted on a shaker table, which was set to an oscillation frequency of 16 Hz. This frequency is well below Nyquist frequency of half the sampling frequency of the accelerometer (195/2 Hz) and thus detectable. The dataset, which spans about 45 min of measurements, is characterized by initial transient oscillations with irregular amplitudes but stable frequency (approximately throughout the first 1,500 data values), followed by a stationary oscillation with stable amplitudes. The dataset excluding the initial phase was previously modeled in Kargoll et al. (2018a) by the truncated Fourier series (4) with $M = 12$, with AR random deviations e_t , and with t-distributed white noise components u_t . The frequencies were treated as the fixed quantities $f_j = j \cdot 8$ Hz ($j = 1, \dots, M$). Besides the main frequency $f_2 = 16$ Hz, 11 other frequencies at multiples of 8 Hz with associated small amplitudes were identified within the discrete Fourier transform. These were suspected to arise from the physical properties of the shaker table and thus modeled deterministically. The Fourier coefficients $a_0, a_1, \dots, a_{12}$ and $b_1, \dots, b_{12}$ were treated as unknowns and collected within the parameter vector x . In Kargoll et al. (2018a), the initial phase was modeled by employing the observation model (4) in connection with a time-variable AR(6) model involving cubic polynomials defined by $B_t = [1 \ \tau_t^1 \ \tau_t^2 \ \tau_t^3]$.

This model was found by trying out different AR as well as different polynomial model orders until the periodogram excess of the estimated white noise residuals appeared to be small in comparison to theoretical white noise. The reasonable choices $m \in \{2, 3, 4, 5\}$ correspond to polynomial models defined by

- $B_t = [1 \ \tau_t^1]$ for $m = 2$ (linear drift),
- $B_t = [1 \ \tau_t^1 \ \tau_t^2]$ for $m = 3$ (quadratic polynomial),
- $B_t = [1 \ \tau_t^1 \ \tau_t^2 \ \tau_t^3]$ for $m = 4$ (cubic polynomial),
- $B_t = [1 \ \tau_t^1 \ \tau_t^2 \ \tau_t^3 \ \tau_t^4]$ for $m = 5$.

Figure 3 shows for the fixed AR model order $p = 6$ that the time-variability model with $m = 4$ results in the acceptance of the white noise test since its cumulated periodogram excess over theoretical white noise lies completely within the approximate 99% significance bounds. In contrast, the white noise hypothesis is rejected for the choices $m = 2$, $m = 3$ and $m = 5$. However, as the white noise components contain outliers when the degree of freedom of the underlying t-distribution is relatively small, the periodogram estimate might be affected and contaminated by outliers. We therefore seek to improve the model selection step with respect to the polynomial order of the time-variability model by carrying out the bootstrap test given in Sect. 3. For this purpose, we retained within the current study the AR model order of $p = 6$ and considered the testing problem (2) for $m \in \{2, 3, 4, 5\}$. Assembling the complete B -matrix from one of these choices at a time, the corresponding p -value estimate was computed (as explained in Sect. 3) both under parametric and nonparametric bootstrapping, as well as for both $B = 99$ and $B = 999$ bootstrap samples (see Table 1). Since the averaging (3) yields increasingly precise p -value estimate with increasing number B of bootstrap sample, and since some of the estimated p -values change considerable when taking $B = 999$ instead of $B = 99$ bootstrap samples, we conclude that B should be at least $B = 999$. This confirms the finding of McKinnon (2007) that “it might be dangerous to use a value of B less than 999”. For that number of bootstrap samples, we see that the

Fig. 3 Excess of the estimated periodograms of the decorrelated residuals for $m = 2$ (dotted red line), $m = 3$ (dotted blue line), $m = 4$ (solid blue line) and $m = 5$ (solid red line) with respect to the AR(6) model over the theoretical white noise periodogram (equal to 0); approximate 99% significance bounds are displayed as the two heavy black lines

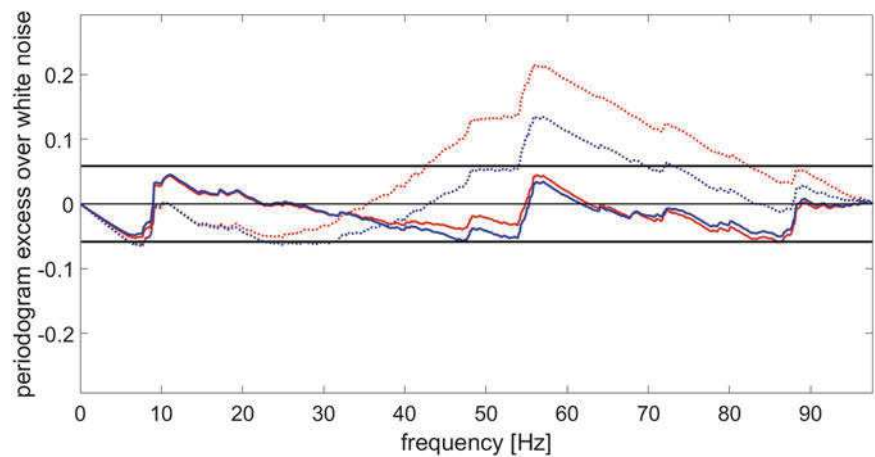


Table 1 Estimated p -values for the parametric and the non-parametric version of the bootstrap test for time-variability of an AR(6) process, modeling random deviations of the observation model (4) for the initial segment of the analyzed accelerometer dataset

m	2	3	4	5
Parametric ($B = 99$)	0	0	0.03	0.17
($B = 999$)	0	0	0.02	0.20
Non-parametric ($B = 99$)	0	0	0.07	0.10
($B = 999$)	0	0	0.03	0.19

p -values obtained by non-parametric bootstrapping are very close to the values obtained parametrically by generating random numbers from the fitted t-distribution. We therefore conclude that this choice can safely be made by the user according to computational convenience. Most importantly, the p -values (for $B = 999$) are well below the standard significance level $\alpha = 0.05$ for the cubic polynomial model ($m = 4$), which previously was found to be the only adequate one. As the p -values with respect to the linear drift model ($m = 2$) and for the quadratic polynomial model ($m = 3$) are zero, the null hypothesis of ‘no time-variability of the AR model’ is not supported by the data, as for the test against the adequate cubic polynomial model. In a practical situation it might be desirable to formulate the more general alternative hypothesis ‘The AR model is time variable.’ The previous results show that any of the aforementioned specific alternative models (linear drift, quadratic, cubic polynomials) could be used within the Estimation Step and the Generation Step because each of these models implies the correct rejection of the null hypothesis. Increasing the optimal polynomial order of $m = 4$ to $m = 5$ apparently desensitizes the test since the p -values now exceed any reasonable choice for the significance level, resulting in the inadequate acceptance of H_0 . In a previous study, we found that assuming an AR model order too small or too large often results in inferior model estimates, which behavior was documented by an unstable acceptance rate of a white noise test (see Kargoll et al. 2018a, Figure 9). Thus, estimated AR models have a

tendency to be acceptable only within certain ranges of order. We suspect that a similar phenomenon might occur for the order of the time-variability model, but we cannot prove this finding, yet.

6 Summary, Conclusions, and Outlook

We presented a statistical procedure based on MC bootstrapping to test the null hypothesis that the random deviations of a regression time series follow a time-constant, fixed-order AR process. The alternative hypothesis may be specified by an arbitrary linear model that forces each AR coefficient to lie exactly on a time-dependent deterministic function. To take a potentially large number of outliers of unknown frequency and magnitudes into account, the input white noise to that AR process was modeled by means of a t-distribution with estimable scale factor and degree of freedom. Since the exact and even approximate test distribution are unknown, the MC bootstrapping yields an estimate of the p -value, which may be compared to a specified significance level to arrive at the test decision. A closed-loop simulation showed that the specified significance level of the bootstrap test is reproduced closely. Moreover, the application of the bootstrap test to an initial segment of an accelerometer measurement series, which was previously modeled by means of cubic polynomials with respect to a time-variable AR(6) process, showed that the null hypothesis of no time-variability should indeed be rejected in favor of linear, quadratic or cubic polynomials. We may therefore conclude that the model of linearly drifting AR coefficients may be used in a test against the general alternative that the ‘AR model is time-variable’. This conjecture, however, should be further investigated in the future. The real-data study also demonstrated that at least 999 bootstrap samples should be generated to obtain adequate p -values. Both the closed-loop simulation and the real-data analysis showed that parametric bootstrapping (using the t-distribution estimated from the

given dataset) and non-parametric bootstrapping (drawing at random with replacement from the estimated white noise of the real-data adjustment) in order to generate the white noise samples result in very similar p -value estimates. It appears that the presented bootstrap approach may be adapted to similar testing problems in the context of time series analysis involving intricate parametric models.

Acknowledgements Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation)—386369985. The presented application of the PCB Piezotronics accelerometer within the vibration analysis experiment was performed as a part of the collaborative project “Spatio-temporal monitoring of bridge structures using low cost sensors” with ALLSAT GmbH, which is funded by the German Federal Ministry for Economic Affairs and Energy (BMWi) and the Central Innovation Programme for SMEs (ZIM Kooperationsprojekt, ZF4081803DB6). In addition, the authors acknowledge the Institute of Concrete Construction (Leibniz Universität Hannover) for providing the shaker table and the reference accelerometer used within this experiment.

References

- Alkhatib H, Kargoll B, Paffenholz JA (2018) Further results on robust multivariate time series analysis in nonlinear models with autoregressive and t-distributed errors. In: Rojas I, Pomares H, Valenzuela O (eds) Time series analysis and forecasting. ITISE 2017. Contributions to statistics. Springer, Cham, pp 25–38. https://doi.org/10.1007/978-3-319-96944-2_3
- Angrisano A, Maratea A, Gagliione S (2018) A resampling strategy based on bootstrap to reduce the effect of large blunders in GPS absolute positioning. *J Geod* 92:81–92. <https://doi.org/10.1007/s00190-017-1046-6>
- Baarda W (1968) A testing procedure for use in geodetic networks. Publications on Geodesy (New Series), vol 2, no 5, Netherlands Geodetic Commission, Delft
- Davidson R, MacKinnon JG (2000) Bootstrap tests: How many bootstraps? *Economet Rev* 19:55–68. <https://doi.org/10.1080/07474930008800459>
- Efron B (1979) Bootstrap methods: another look at the jackknife. *Ann Stat* 7:1–26. <https://doi.org/10.1214/aos/1176344552>
- Kargoll B, Omidalizarandi M, Loth I, Paffenholz JA, Alkhatib H (2018a) An iteratively reweighted least-squares approach to adaptive robust adjustment of parameters in linear regression models with autoregressive and t-distributed deviations. *J Geod* 92:271–297. <https://doi.org/10.1007/s00190-017-1062-6>
- Kargoll B, Omidalizarandi M, Alkhatib H, Schuh WD (2018b) Further results on a modified EM algorithm for parameter estimation in linear models with time-dependent autoregressive and t-distributed errors. In: Rojas I, Pomares H, Valenzuela O (eds) Time series analysis and forecasting. ITISE 2017. Contributions to statistics. Springer, Cham, pp 323–337. https://doi.org/10.1007/978-3-319-96944-2_22
- Koch KR (2018a) Bayesian statistics and Monte Carlo methods. *J Geod Sci* 8:18–29. <https://doi.org/10.1515/jogs-2018-0003>
- Koch KR (2018b) Monte Carlo methods. *Int J Geomath* 9:177–143. <https://doi.org/10.1007/s13137-017-0101-z>
- Koch KR, Kargoll B (2013) Expectation-maximization algorithm for the variance-inflation model by applying the t-distribution. *J Appl Geod* 7:217–225. <https://doi.org/10.1515/jag-2013-0007>
- Kuhlmann H (2001) Importance of autocorrelation for parameter estimation in regression models. In: 10th FIG international symposium on deformation measurements, pp 354–361
- Kutterer H (1999) On the sensitivity of the results of least-squares adjustments concerning the stochastic model. *J Geod* 73(7):350–361. <https://doi.org/10.1007/s001900050253>
- Li H, Maddala GS (1996) Bootstrapping time series models. *Econom Rev* 15:115–158. <https://doi.org/10.1080/07474939608800344>
- Lösler M, Eschelbach C, Haas R (2018) Bestimmung von Messunsicherheiten mittels Bootstrapping in der Formanalyse. *Zeitschrift für Geodäsie, Geoinformatik und Landmanagement zfv* 143:224–232
- McKinnon J (2007) Bootstrap hypothesis testing. Queen’s Economics Department Working Paper No. 1127, Queen’s University, Kingston, Ontario, Canada
- McLachlan GJ, Krishnan T (2008) The EM algorithm and extensions, 2nd edn. Wiley, Hoboken
- Neuner H, Wieser A, Krähenbühl N (2014) Bootstrapping: Moderne Werkzeuge für die Erstellung von Konfidenzintervallen. In: Neuner H (ed) Zeitabhängige Messgrößen - Ihre Daten haben (Mehr-)Wert. Schriftenreihe des DVW, vol 74, pp 151–170
- Parzen E (1979) A density-quantile function perspective on robust estimation. In: Launer L, Wilkinson GN (eds) Robustness in statistics. Academic, New York, pp 237–258. <https://doi.org/10.1016/B978-0-12-438150-6.50019-4>
- Politis DN (2003) The impact of bootstrap methods on time series analysis. *Statist Sci* 18:219–230. <https://doi.org/10.1214/ss/1063994977>
- Schuh WD (2003) The processing of band-limited measurements; filtering techniques in the least squares context and in the presence of data gaps. *Space Sci Rev* 108:67–78. <https://doi.org/10.1023/A:1026121814042>
- Schuh WD, Brockmann JM (2016) Refinement of the stochastic model for GOCE gravity gradients by non-stationary decorrelation filters. ESA Living Planet Symposium 2016, Prag, Poster 2382
- Teunissen (1998) Success probability of integer GPS ambiguity rounding and bootstrapping. *J Geod* 72:606–612. <https://doi.org/10.1007/s001900050199>
- Wiśniewski Z (2014) M -estimation with probabilistic models of geodetic observations. *J Geod* 88(10):941–957. <https://doi.org/10.1007/s00190-014-0735-7>

Literaturverzeichnis

- [Abmayr et al. 2008] ABMAYR, T. ; HÄRTL, F. ; HIRZINGER, G. ; BURSCHKA, D. ; FRÖHLICH, C.: A correlation based target finder for terrestrial laser scanning. In: *Journal of Applied Geodesy* 2 (2008), Nr. 3, S. 429
- [Acko und Godina 2005] ACKO, B. ; GODINA, A.: Verification of the conventional measuring uncertainty evaluation model with Monte Carlo simulation. In: *International Journal of Simulation Modelling* 4 (2005), Nr. 2, S. 76–84. – ISSN 1726-4529
- [Alkhatib 2015] ALKHATIB, H.: Alternative Nonlinear Filtering Techniques In Geodesy For Dual State And Adaptive Parameter Estimation. In: ALKHATIB, Hamza (Hrsg.) ; KUTTERER, Hansjörg (Hrsg.) ; SCHMIDT, Michael (Hrsg.) ; SEITZ, Florian (Hrsg.): *The 1st International Workshop on the Quality of Geodetic Observation and Monitoring Systems (QuGOMS'11)*. Cham : Springer International Publishing, 2015 (International Association of Geodesy Symposia), S. 107–113. – ISBN 978-3-319-10827-8
- [Alkhatib et al. 2017a] ALKHATIB, H. ; KARGOLL, B. ; NEUMANN, I. ; KREINOVICH, V.: Normalization-Invariant Fuzzy Logic Operations Explain Empirical Success of Student Distributions in Describing Measurement Uncertainty. In: MELIN, P. (Hrsg.) ; CASTILLO, O. (Hrsg.) ; KACPRZYK, J. (Hrsg.) ; REFORMAT, M. (Hrsg.) ; MELEK W. (Hrsg.): *Fuzzy Logic in Intelligent System Design*, Springer, Cham, Switzerland, 2017, S. 300–306. – ISBN 978-3-319-67136-9
- [Alkhatib et al. 2017b] ALKHATIB, H. ; KARGOLL, B. ; PAFFENHOLZ, J.-A.: Robust multivariate time series analysis in nonlinear models with autoregressive and t-distributed errors. In: VALENZUELA, O. (Hrsg.) ; ROJAS, F. (Hrsg.) ; POMARES, H. (Hrsg.) ; ROJAS, I. (Hrsg.): *Proceedings ITISE 2017 - International work-conference on Time Series*, 2017
- [Alkhatib et al. 2018a] ALKHATIB, H. ; KARGOLL, B. ; PAFFENHOLZ, J.-A.: Further results on robust multivariate time series analysis in nonlinear models with autoregressive and t-distributed errors. In: ROJAS, Ignacio (Hrsg.) ; POMARES, Héctor (Hrsg.) ; VALENZUELA, Olga (Hrsg.): *Contributions to Statistics*, Springer, 2018
- [Alkhatib et al. 2018b] ALKHATIB, H. ; KARGOLL, B. ; PAFFENHOLZ, J.-A.: Further results on robust multivariate time series analysis in nonlinear models with autoregressive and t-distributed errors. In: ROJAS, Ignacio (Hrsg.) ; POMARES, Héctor (Hrsg.) ; VALENZUELA, Olga (Hrsg.): *Contributions to Statistics*, Springer, 2018
- [Alkhatib und Kutterer 2013] ALKHATIB, H. ; KUTTERER, H.: Estimation of Measurement Uncertainty of kinematic TLS Observation Process by means of Monte-Carlo Methods. In: *Journal of Applied Geodesy* 7 (2013), S. 125–133
- [Alkhatib et al. 2009] ALKHATIB, H. ; NEUMANN, I. ; KUTTERER, H.: Uncertainty modeling of random and systematic errors by means of Monte Carlo and fuzzy techniques. In: *Journal of Applied Geodesy* 3 (2009), Nr. 2, S. 67–79. – ISSN 18629016
- [Alkhatib et al. 2019] ALKHATIB, H. ; OMIDALIZARANDI, M. ; KARGOLL, B.: A Bootstrap Approach to Testing for Time-Variability of AR Process Coefficients in Regression Time Series with t-Distributed White Noise Components. Berlin, Heidelberg : Springer Berlin Heidelberg, 2019 (International Association of Geodesy Symposia)

- [Alkhatib et al. 2012] ALKHATIB, H. ; PAFFENHOLZ, J.-A. ; KUTTERER, H.: Sequential Monte Carlo filtering for nonlinear GNSS trajectories. In: SNEEUW, Nico (Hrsg.) ; NOVÁK, P. (Hrsg.) ; CRESPI, M. (Hrsg.) ; SANSÒ, F. (Hrsg.): *VII Hotine-Marussi Symposium on Mathematical Geodesy*. Berlin and Heidelberg : Springer, 2012 (International Association of Geodesy Symposia), S. 81–86
- [Alkhatib und Schuh 2007] ALKHATIB, H. ; SCHUH, W.-D.: Integration of the Monte Carlo Covariance Estimation Strategy into Tailored Solution Procedures for Large-scaled Least Squares Problems. In: *Journal of Geodesy* 81 (2007), S. 53–66
- [Alkhatib und Weitkamp 2012] ALKHATIB, H. ; WEITKAMP, A.: Bayesischer Ansatz zur Integration von Expertenwissen in die Immobilienbewertung, Teil 1. In: *Zeitschrift für Geodäsie, Geoinformation und Landmanagement (zfv)* 02 (2012), S. 93–102
- [Alkhatib und Weitkamp 2013] ALKHATIB, H. ; WEITKAMP, A.: Robust Bayesian Regression Approach For Areas With Small Numbers Of Purchases. In: ROYAL INSTITUTION OF CHARTERED SURVEYORS (Hrsg.): *Proceedings of COBRA Conference. New Delhi*, 2013
- [Antoniou und Lu 2007] ANTONIOU, A. ; LU, W.-S.: *Practical Optimization*. Boston, MA : Springer US, 2007. – ISBN 978-0-387-71106-5
- [Bandemer 2006] BANDEMER, H.: *Mathematics of Uncertainty – Ideas, Methods, Application Problems*. Berlin, Heidelberg, New York : Springer, 2006
- [Bar-Shalom et al. 2001] BAR-SHALOM, Y. ; LI, X. R. ; KIRUBARAJAN, T. ; LI, X.-R.: *Estimation with applications to tracking and navigation: Theory algorithms and software*. New York, NY : John Wiley & Sons, Inc, 2001
- [Bishop 2009] BISHOP, Christopher M.: *Pattern recognition and machine learning*. Corrected at 8th printing 2009. New York, NY : Springer, 2009 (Information science and statistics). – ISBN 9781493938438
- [Box und Andersen 1955] BOX, G. E. P. ; ANDERSEN, S. L.: Permutation Theory in the Derivation of Robust Criteria and the Study of Departures from Assumption. In: *Journal of the Royal Statistical Society. Series B (Methodological)* 17 (1955), Nr. 1, S. 1–34. – URL <http://www.jstor.org/stable/2983783>. – ISSN 00359246
- [Bureick et al. 2016a] BUREICK, J. ; ALKHATIB, H. ; NEUMANN, I.: Robust spatial approximation of laser scanner point clouds by means of free-form curve approaches in deformation analysis. In: *Journal of Applied Geodesy* 10 (2016), Nr. 1, S. 27–35. – ISSN 1862-9024
- [Bureick et al. 2019a] BUREICK, J. ; ALKHATIB, H. ; NEUMANN, I.: Fast converging elitist genetic algorithm for knot adjustment in B-spline curve approximation. In: *Journal of Applied Geodesy* (2019). – ISSN 1862-9024
- [Bureick et al. 2016b] BUREICK, J. ; NEUNER, H. B. ; HARMENING, C. ; NEUMANN, I.: Curve and surface approximation of 3D point clouds. In: *Allgemeine Vermessungs-Nachrichten (AVN)* 123 (2016), Nr. 11-12, S. 315–327
- [Bureick et al. 2019b] BUREICK, J. ; VOGEL, I. ; UNGER, J. ; ALKHATIB, H.: Georeferencing of an Unmanned Aerial System by Means of an Iterated Extended Kalman Filter Using a 3D City Model. In: *PFG – Journal of Photogrammetry, Remote Sensing and Geoinformation Science* 2 (2019), Nr. 3, S. 429. – ISSN 2512-2789
- [Bureick et al. 2019c] BUREICK, J. ; VOGEL, S. ; NEUMANN, I. ; H., Alkhatib: Geo-Referenzierung von Unmanned Aerial Systems über Laserscannermessungen und 3D-Gebäudemodelle. In: *DVW*

- e. V. Hg. *Terrestrisches Laserscanning 2019 (TLS 2019) DVW-Schriftenreihe*, Wißner-Verlag, Augsburg Bd. 96. 2019, S. 53–66
- [C. 2005] C., Christophe: A State Bounding Observer for Uncertain Non-linear Continuous-time Systems based on Zonotopes. In: *Proceedings of the 44th IEEE Conference on Decision and Control* (2005), S. 7228–7234
- [Candy 2016] CANDY, J. V.: *Bayesian signal processing: Classical, modern, and particle filtering methods*. Second edition. Hoboken, New Jersey : John Wiley & Sons Inc, 2016. – ISBN 1119125499
- [Casella und Robert 2002] CASELLA, G. ; ROBERT, C. P.: *Monte Carlo statistical methods*. Corrected 3rd printing. New York : Springer, 2002 (Springer Texts in Statistics). – ISBN 0-387-98707-X
- [Caspary 2013] CASPARY, W.: *Fehlertolerante auswertung von messdaten: Daten, und modellanalyse, robuste schätzung*. München, Germany : Oldenbourg Verlag München, 2013. – URL <http://search.ebscohost.com/login.aspx?direct=true&scope=site&db=nlebk&db=nlabk&AN=754003>. – ISBN 9783486727715
- [Chang und Guo 2005] CHANG, X.-W. ; GUO, Y.: Huber’s M-estimation in relative GPS positioning: computational aspects. In: *Journal of Geodesy* 79 (2005), Nr. 6-7, S. 351–362. – ISSN 1432-1394
- [Chiang et al. 2002] CHIANG, Y.-T. ; WANG, L.-S. ; CHANG, F.-R.: Filtering method for nonlinear systems with constraints. In: *IEE Proceedings - Control Theory and Applications* 149 (2002), Nr. 6, S. 525–531. – ISSN 13502379
- [Cox 1961] COX, D. R.: Tests of Separate Families of Hypotheses. In: *Proceedings of the fourth Berkeley symposium on mathematical statistics and probability*, 1961
- [Cox 1972] COX, M. G.: The numerical evaluation of B-Splines. In: *IMA Journal of Applied Mathematics* 10 (1972), Nr. 2, S. 134–149. – ISSN 0272-4960
- [Dang 2007] DANG, T.: *Kontinuierliche Selbstkalibrierung von Stereokameras*. Karlsruhe, KIT, Dissertation, 2007
- [de Boor 1972] DE BOOR, C.: On calculating with B-splines. In: *Journal of Approximation Theory* 6 (1972), Nr. 1, S. 50–62. – ISSN 00219045
- [Dempster et al. 1977] DEMPSTER, A. P. ; LAIRD, N. M. ; RUBIN, D. B.: Maximum Likelihood from Incomplete Data via the EM Algorithm. In: *Journal of the Royal Statistical Society. Series B (Methodological)* 39 (1977), Nr. 1, S. 1–38. – URL <http://www.jstor.org/stable/2984875>. – ISSN 00359246
- [Dennig et al. 2017] DENNIG, D. ; BUREICK, J. ; LINK, J. ; DIENER, D. ; HESSE, C. ; NEUMANN, I.: Comprehensive and Highly Accurate Measurements of Crane Runways, Profiles and Fastenings. In: *Sensors* 17 (2017), Nr. 5. – ISSN 1424-8220
- [Dierckx 1993] DIERCKX, P.: *Curve and Surface Fitting with Splines*. Repr. Oxford [i.a.] : Clarendon Press, 1993 (Oxford science publications). – ISBN 9780198534402
- [Dorndorf et al. 2019] DORNDORF, A. ; KARGOLL, B. ; PAFFENHOLZ, J.-A. ; ALKHATIB, H.: A Bayesian Nonlinear Regression Model Based on t-Distributed Errors. In: *International Association of Geodesy Symposia*. Springer Berlin Heidelberg, 2019
- [Doucet et al. 2001] DOUCET, A. ; FREITAS, N. de ; GORDON, N.: *Sequential Monte Carlo methods in practice*. New York, NY : Springer, 2001

- [Durieu et al. 2001] DURIEU, C. ; WALTER, É. ; POLYAK, B.: Multi-Input Multi-Output Ellipsoidal State Bounding. In: *Journal of Optimization Theory and Applications* 111 (2001), Nr. 2, S. 273–303
- [Efron 1979] EFRON, B.: Bootstrap Methods: Another Look at the Jackknife. In: *Ann. Statist.* 7 (1979), Nr. 1, S. 1–26
- [Efron und Hastie 2016] EFRON, B. ; HASTIE, T.: *Institute of Mathematical Statistics monographs*. Bd. 5: *Computer age statistical inference: Algorithms, evidence, and data science*. New York : Cambridge University Press, 2016. – ISBN 9781107149892
- [Efron und Tibshirani 1993] EFRON, B. ; TIBSHIRANI, R. J.: *An Introduction to the Bootstrap*. Boston, MA : Springer US, 1993. – ISBN 978-0-412-04231-7
- [Efron und Tibshirani 1991] EFRON, B. ; TIBSHIRANI, Robert J.: Statistical Data Analysis in the Computer Age. In: *Science* 253 (1991), Nr. 5018, S. 390–395
- [Elkhrachy und Niemeier 2006] ELKHRACHY, I. ; NIEMEIER, W.: Stochastic assessment of terrestrial laser scanner. In: *Proceedings of the ASPRS Annual Conference, Reno, United States of America*, 2006
- [Evensen 1994] EVENSEN, G.: Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics. In: *Journal of Geophysical Research* 99 (1994), Nr. C5, S. 10143. – ISSN 0148-0227
- [Förstner und Wrobel 2016] FÖRSTNER, W. ; WROBEL, B. P.: *Photogrammetric Computer Vision*. Bd. 11. Cham : Springer International Publishing, 2016. – ISBN 978-3-319-11549-8
- [Fraser 1976] FRASER, D. A. S.: Necessary Analysis and Adaptive Inference. In: *Journal of the American Statistical Association* 71 (1976), Nr. 353, S. 99. – ISSN 01621459
- [Gálvez et al. 2015] GÁLVEZ, A. ; IGLESIAS, A. ; AVILA, A. ; OTERO, C. ; ARIAS, R. ; MANCHADO, C.: Elitist clonal selection algorithm for optimal choice of free knots in B-spline data fitting. In: *Applied Soft Computing* 26 (2015), S. 90–106. – ISSN 15684946
- [Gelb 1974] GELB, Arthur (Hrsg.): *Applied optimal estimation*. Cambridge and Mass : MIT Press, 1974
- [Gelman 2013] GELMAN, A.: *Bayesian data analysis*. Third edition. 2013 (Chapman & Hall/CRC texts in statistical science). – ISBN 978-1439840955
- [Gentle 2003] GENTLE, J. E.: *Random Number Generation and Monte Carlo Methods*. 2. Springer : Wiley, 2003. – ISBN B006VE0T5A
- [Geweke 1993] GEWEKE, J.: Bayesian treatment of the independent student-t linear model. In: *Journal of Applied Econometrics* 8 (1993), Nr. S1, S. S19–S40. – ISSN 08837252
- [Ghilani und Wolf 2010] GHILANI, C. D. ; WOLF, P. R.: *Adjustment computations: Spatial data analysis*. 5. Hoboken, NJ : John Wiley & Sons, Inc, 2010
- [Givens und Hoeting 2013] GIVENS, G. H. ; HOETING, J. A.: *Computational statistics*. 2nd ed. Hoboken, N.J. : Wiley, 2013 (Wiley series in computational statistics). – ISBN 9781118555552
- [Glira et al. 2015] GLIRA, P. ; PFEIFER, N. ; BRIESE, C. ; RESSL, C.: Rigorous strip adjustment of airborne laserscanning data based on the ICP algorithm. In: *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences* II-3/W5 (2015), S. 73–80. – ISSN 2194-9050

- [Gordon et al. 1993] GORDON, N. J. ; SALMOND, D. J. ; SMITH, A.F.M.: Novel approach to nonlinear/non-Gaussian Bayesian state estimation. In: *IEE Proceedings F Radar and Signal Processing* 140 (1993), Nr. 2, S. 107. – ISSN 0956375X
- [Götzelmann et al. 2006] GÖTZELMANN, M. ; KELLER, W. ; REUBELT, T.: Gross Error Compensation for Gravity Field Analysis Based on Kinematic Orbit Data. In: *Journal of Geodesy* 80 (2006), Nr. 4, S. 184–198. – ISSN 1432-1394
- [Grabe 2005] GRABE, M.: *Measurement Uncertainties in Science and Technology*. Berlin-New York : Springer, 2005
- [Guo et al. 2010] GUO, J. ; OU, J. ; WANG, H.: Robust estimation for correlated observations: two local sensitivity-based downweighting strategies. In: *Journal of Geodesy* 84 (2010), Nr. 4, S. 243–250. – ISSN 1432-1394
- [Hargreaves 2002] HARGREAVES, G. I. ; MANCHESTER CENTRE FOR COMPUTATIONAL MATHEMATICS, THE UNIVERSITY OF MANCHESTER (Hrsg.): *Interval Analysis in MATLAB. Numerical Analysis Report No. 416*. 2002
- [Harmening und Neuner 2015a] HARMENING, C. ; NEUNER, H. B.: A constraint-based parameterization technique for B-spline surfaces. In: *Journal of Applied Geodesy* 9 (2015), Nr. 3, S. 143–161. – ISSN 1862-9024
- [Harmening und Neuner 2015b] HARMENING, C. ; NEUNER, H. B.: A constraint-based parameterization technique for B-spline surfaces. In: *Journal of Applied Geodesy* 9 (2015), Nr. 3, S. 143–161. – ISSN 1862-9024
- [Harmening und Neuner 2016] HARMENING, C. ; NEUNER, H. B.: Choosing the optimal number of B-spline control points (Part 1: Methodology and approximation of curves). In: *Journal of Applied Geodesy* 10 (2016), Nr. 3, S. 139–157. – ISSN 1862-9024
- [Hartmann et al. 2017] HARTMANN, J. ; PAFFENHOLZ, J.-A. ; STRÜBING, T. ; NEUMANN, I.: Determination of Position and Orientation of LiDAR Sensors on Multisensor Platforms. In: *Journal of Surveying Engineering* 143 (2017), Nr. 4, S. 04017012. – ISSN 0733-9453
- [Hartmann et al. 2018] HARTMANN, J. ; TRUSHEIM, P. ; ALKHATIB, H. ; PAFFENHOLZ, J.-A. ; DIENER, D. ; NEUMANN, I.: HIGH ACCURATE POINTWISE (GEO-)REFERENCING OF A K-TLS BASED MULTI-SENSOR-SYSTEM. In: *ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences* IV-4 (2018), S. 81–88
- [Hastings 1970] HASTINGS, W. K.: Monte Carlo Sampling Methods Using Markov Chains and Their Applications. In: *Biometrika* 57 (1970), Nr. 1, S. 97. – ISSN 0006-3444
- [Hebel et al. 2009] HEBEL, M. ; ARENS, M. ; STILLA, U.: Utilization of 3D city models and airborne laser scanning for terrain-based navigation of helicopters and UAVs. In: *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* XXXVIII (2009), S. Part 3/W4
- [Hekimoglu und Berber 2003] HEKIMOGLU, S. ; BERBER, M.: Effectiveness of robust methods in heterogeneous linear models. In: *Journal of Geodesy* 76 (2003), Nr. 11-12, S. 706–713. – ISSN 1432-1394
- [Helmert 1907] HELMERT, FR.: *Die Ausgleichungsrechnung nach der Methode der kleinsten Quadrate*. 2. Teubner, Leipzig, Berlin, 1907
- [Hennes 2007] HENNES, M.: Konkurrierende Genauigkeitsmasse: Potential und Schwächen aus der Sicht des Anwenders. In: *AVN* 4 (2007), S. 136–146

- [Huber 1964] HUBER, P. J.: Robust Estimation of a Location Parameter. In: *Ann. Math. Statist.* 35 (1964), 03, Nr. 1, S. 73–101. – URL <https://doi.org/10.1214/aoms/1177703732>
- [ISO 1995] ISO: Guide to the expression of uncertainty in measurements (GUM). Geneva : International Organization for Standardization, 1995
- [ISO 2007] ISO: Evaluation of measurement data - Supplement 1 to the “Guide to the expression of uncertainty in measurement” - Propagation of distributions using a Monte Carlo method. Geneva : Joint Committee for Guides in Metrology, Bureau International des Poids et Mesures, 2007
- [ISO 2008] ISO: Evaluation of measurement data - Supplement 1 to the "Guide to the expression of uncertainty in measurement Propagation of distributions using a Monte Carlo method. Geneva : Joint Committee for Guides in Metrology, Bureau International des Poids et Mesures, 2008
- [Jäger et al. 2016] JÄGER, R. ; MÜLLER, T. ; SALER, H.: *Klassische und robuste Ausgleichungsverfahren: Ein Leitfaden für Ausbildung und Praxis von Geodäten und Geoinformatikern.* 2., 02. Auflage, neu bearb. + erw. Berlin and s.l. : Wichmann H, 2016. – ISBN 9783879076154
- [Jäger et al. 2005] JÄGER, R. R. ; MÜLLER, T. ; SALER, H. ; SCHWÄBLE, R.: *Klassische und robuste Ausgleichungsverfahren: Ein Leitfaden für Ausbildung und Praxis von Geodäten und Geoinformatikern.* Heidelberg : Wichmann, 2005
- [JCGM 2008] JCGM: Evaluation of measurement data - Supplement 1 to the "Guide to the expression of uncertainty in measurement Propagation of distributions using a Monte Carlo method., Joint Committee for Guides in Metrology (JCGM/WG 1), 2008
- [Julier und Uhlmann 1997] JULIER, S. J. ; UHLMANN, J. K.: A new extension of the Kalman filter to nonlinear systems. In: *SPIE Proceedings of AeroSense*, 1997
- [Junhuan 2005] JUNHUAN, P.: The asymptotic variance–covariance matrix, Baarda test and the reliability of L1-norm estimates. In: *Journal of Geodesy* 78 (2005), Nr. 11-12, S. 668–682. – ISSN 1432-1394
- [Kalman 1960] KALMAN, R. E.: A new approach to linear filtering and prediction problems. In: *Journal of Basic Engineering* 82 (1960), Nr. 1, S. 35–45. – ISSN 00219223
- [Kargoll et al. 2018a] KARGOLL, B. ; OMIDALIZARANDI, H. ; SCHUH, W.-D.: Further Results on a Modified EM Algorithm for Parameter Estimation in Linear Models with Time-Dependent Autoregressive and t-Distributed Errors. In: ROJAS, Ignacio (Hrsg.) ; POMARES, Hector (Hrsg.): *Time Series Analysis and Forecasting* Bd. 92. Cham : Springer, 2018, S. 323–337. – ISBN 978-3-319-96943-5
- [Kargoll et al. 2018b] KARGOLL, B. ; OMIDALIZARANDI, M. ; LOTH, I. ; PAFFENHOLZ, J.-A. ; ALKHATIB, H.: An iteratively reweighted least-squares approach to adaptive robust adjustment of parameters in linear regression models with autoregressive and t-distributed deviations. In: *Journal of Geodesy* 92 (2018), Nr. 3, S. 271–297. – ISSN 1432-1394
- [Katzfuss et al. 2016] KATZFUSS, M. ; STROUD, J. R. ; WIKLE, C. K.: Understanding the Ensemble Kalman Filter. In: *The American Statistician* 70 (2016), Nr. 4, S. 350–357. – ISSN 0003-1305
- [Keller 2013] KELLER, H.: Multi-Sensor Platform for Indoor Mobile Mapping: System Calibration and Using a Total Station for Indoor Applications. In: *Remote Sensing* 5 (2013), Nr. 11, S. 5805–5824

- [Koch 1999] KOCH, K.-R.: *Parameter estimation and hypothesis testing in linear models*. 2. Berlin : Springer, 1999
- [Koch 2000] KOCH, K.-R.: *Einführung in die Bayes-Statistik*. Berlin - Heidelberg - New York : Springer, 2000. – ISBN 3540666702
- [Koch 2009a] KOCH, K. R.: Fitting free-form surfaces to laserscan data by NURBS. In: *Allgemeine Vermessungs-Nachrichten (AVN)* 116 (2009), Nr. 4, S. 134–140
- [Koch 2009b] KOCH, K.-R.: Identity of simultaneous estimates of control points and of their estimates by the lofting method for NURBS surface fitting. In: *The International Journal of Advanced Manufacturing Technology* 44 (2009), Nr. 11-12, S. 1175–1180. – ISSN 0268-3768
- [Koch 2010] KOCH, K.-R.: Three-dimensional NURBS surface estimated by lofting method. In: *The International Journal of Advanced Manufacturing Technology* 49 (2010), Nr. 9-12, S. 1059–1068. – ISSN 0268-3768
- [Koch 2013] KOCH, K.-R.: Comparison of two robust estimations by expectation maximization algorithms with Huber's method and outlier tests. In: *Journal of Applied Geodesy* 7 (2013), Nr. 2. – ISSN 18629016
- [Koch 2018] KOCH, K.-R.: Bayesian statistics and Monte Carlo methods. In: *Journal of Geodetic Science* 8 (2018), Nr. 1, S. 18–29
- [Koop 2010] KOOP, G.: *Bayesian econometrics*. Reprinted with corr., [Nachdr.]. Chichester : Wiley, 2010. – ISBN 9780470845677
- [Kroese et al. 2011] KROESE, D. P. ; TAIMRE, Th. ; BOTEV, Z. I.: *Wiley series in probability and statistics*. Bd. 706: *Handbook of Monte Carlo methods*. Hoboken and N.J : Wiley, 2011. – URL <http://site.ebrary.com/lib/academiccompletetitles/home.action>. – ISBN 978-0-470-17793-8
- [Kutterer und Schön 2004] KUTTERER, H. ; SCHÖN, S.: Alternativen bei der Modellierung der Unsicherheit beim Messen. In: *Zeitschrift für Geodäsie, Geoinformation und Landmanagement (zfv)* 129 (2004), Nr. 6, S. 389–398
- [Lange et al. 1989] LANGE, K. L. ; LITTLE, R. J. A. ; TAYLOR, J. M. G.: Robust Statistical Modeling Using the t Distribution. In: *Journal of the American Statistical Association* 84 (1989), Nr. 408, S. 881. – ISSN 01621459
- [Le et al. 2013] LE, Vu Tuan H. ; STOICA, Cristina ; ALAMO, T. ; CAMACHO, E.F. ; DUMUR, Didier: Zonotopic guaranteed state estimation for uncertain systems. In: *Automatica* 49 (2013), Nr. 11, S. 3418–3424. – URL <https://hal-supelec.archives-ouvertes.fr/hal-00876455>
- [Lehmann 2013] LEHMANN, R.: 3σ -Rule for Outlier Detection from the Viewpoint of Geodetic Adjustment. In: *Journal of Surveying Engineering* 139 (2013), Nr. 4, S. 157–165. – ISSN 0733-9453
- [Lehmann 2015] LEHMANN, R.: Observation error model selection by information criteria vs. normality testing. In: *Studia Geophysica et Geodaetica* 59 (2015), Nr. 4, S. 489–504. – ISSN 0039-3169
- [Liu 1997] LIU, C.: ML Estimation of the Multivariate Distribution and the EM Algorithm. In: *Journal of Multivariate Analysis* 63 (1997), Nr. 2, S. 296–312. – ISSN 0047259X
- [Liu und Rubin 1994] LIU, C. ; RUBIN, D. B.: The ECME algorithm: A simple extension of EM and ECM with faster monotone convergence. In: *Biometrika* 81 (1994), Nr. 4, S. 633–648. – ISSN 0006-3444

- [Luo 2013] LUO, X.: *GPS Stochastic Modelling: Signal Quality Measures and ARMA Processes*. Dordrecht : Springer, 2013 (Springer Theses, Recognizing Outstanding Ph.D. Research). – URL <http://gbv.eblib.com/patron/FullRecord.aspx?p=1082846>. – ISBN 9783642348358
- [Ma und Kruth 1995] MA, W. ; KRUTH, J. P.: Parameterization of randomly measured points for least squares fitting of B-spline curves and surfaces. In: *Computer-Aided Design* 27 (1995), Nr. 9, S. 663–675. – ISSN 00104485
- [MacKinnon 2007] MACKINNON, J.: *Bootstrap Hypothesis Testing*. 2007
- [Marshall 2002] MARSHALL, J.: L 1 -norm pre-analysis measures for geodetic networks. In: *Journal of Geodesy* 76 (2002), Nr. 6-7, S. 334–344. – ISSN 1432-1394
- [Martinez und Martinez 2015] MARTINEZ, W. L. ; MARTINEZ, A. R.: *Chapman and Hall/CRC Computer Science and Data Analysis Ser. Bd. v.22: Computational Statistics Handbook with MATLAB, Third Edition*. 3rd ed. Boca Raton : Chapman and Hall/CRC, 2015. – ISBN 9781466592735
- [Marx 2013] MARX, C.: On resistant Lp-Norm Estimation by means of iteratively reweighted least Squares. In: *Journal of Applied Geodesy* 7 (2013), Nr. 1, S. 1–10. – ISSN 18629016
- [McLachlan und Krishnan 2008] MCLACHLAN, G. J. ; KRISHNAN, T.: *The EM algorithm and extensions*. 2nd ed. Hoboken, N.J. : Wiley-Interscience, 2008. – ISBN 978-0-471-20170-0
- [Meng und Rubin 1993] MENG, X.-L. ; RUBIN, D. B.: Maximum Likelihood Estimation via the ECM Algorithm: A General Framework. In: *Biometrika* 80 (1993), Nr. 2, S. 267. – ISSN 0006-3444
- [Mikhail und Gracie 1981] MIKHAIL, E. M. ; GRACIE, G.: *Analysis and adjustment of survey measurements*. New York : Van Nostrand Reinhold, 1981. – ISBN 0442253699
- [Milanese und Vicino 1991] MILANESE, M. ; VICINO, A.: Optimal estimation theory for dynamic systems with set membership uncertainty: An overview. In: *Automatica* 27 (1991), Nr. 6, S. 997–1009. – URL <http://www.sciencedirect.com/science/article/pii/000510989190134N>. – ISSN 0005-1098
- [Nadarajah 2009] NADARAJAH, S.: The product t density distribution arising from the product of two Student's t PDFs. In: *Statistical Papers* 50 (2009), Nr. 3, S. 605–615. – ISSN 0932-5026
- [Neumann 2009] NEUMANN, I.: *Reihe C. Bd. 634: Zur Modellierung eines erweiterten Unsicherheitshaushaltes in Parameterschätzung und Hypothesentests: PhD thesis*. München : DGK, 2009
- [Niemeier 2008] NIEMEIER, W.: *Ausgleichsrechnung: Statistische Auswertemethoden*. Tübingen : de Gruyter, 2008 (De Gruyter Lehrbuch). – URL <http://gbv.eblib.com/patron/FullRecord.aspx?p=3044705>. – ISBN 9783110190557
- [Noack 2014] NOACK, B.: *State Estimation for Distributed Systems with Stochastic and Set-membership Uncertainties*. 2014
- [Nocedal und Wright 2006] NOCEDAL, J. ; WRIGHT, S. J.: *Numerical Optimization*. Second Edition. New York, NY : Springer Science+Business Media LLC, 2006 (Springer Series in Operations Research and Financial Engineering). – URL <http://site.ebrary.com/lib/alltitles/docDetail.action?docID=10228772>. – ISBN 9780387303031
- [Nowel und Kamiński 2014] NOWEL, K. ; KAMIŃSKI, W.: Robust estimation of deformation from observation differences for free control networks. In: *Journal of Geodesy* 88 (2014), Nr. 8, S. 749–764. – ISSN 1432-1394

- [Orlov 1991] ORLOV, Aleksandr I.: How often are the observations normal, 1991
- [Paffenholz 2012] PAFFENHOLZ, J.-A.: *Direct geo-referencing of 3D point clouds with 3D positioning sensor*. Hannover, Leibniz Universität hannover, Dissertation, 2012
- [Paffenholz et al. 2010] PAFFENHOLZ, J.-A. ; ALKHATIB, H. ; KUTTERER, H.: Direct geo-referencing of a static terrestrial laser scanner. In: *Journal of Applied Geodesy* 4 (2010), Nr. 03, S. 115–126
- [Paffenholz et al. 2017] PAFFENHOLZ, J.-A. ; ALKHATIB, H. ; STENZ, U. ; NEUMANN, I.: Aspekte der Qualitätssicherung von Multi-Sensor-Systemen. In: *allgemeine vermessungs-nachrichten (avn)* 124 (2017), Nr. 4, S. 79–91
- [Park und Lee 2007] PARK, H. ; LEE, J. H.: B-spline curve fitting based on adaptive curve refinement using dominant points. In: *Computer-Aided Design* 39 (2007), Nr. 6, S. 439–451. – ISSN 00104485
- [Park und Gao 2008] PARK, M. ; GAO, Y.: Error and Performance Analysis of MEMS-based Inertial Sensors with a Low-cost GPS Receiver. In: *Sensors* 8 (2008), Nr. 4, S. 2240–2261. – ISSN 1424-8220
- [Parzen 1962] PARZEN, E.: On Estimation of a Probability Density Function and Mode. In: *The Annals of Mathematical Statistics* 33 (1962), Nr. 3, S. 1065–1076. – ISSN 0003-4851
- [Phillips 2002] PHILLIPS, R. F.: Least absolute deviations estimation via the EM algorithm. In: *Statistics and Computing* 12 (2002), Nr. 3, S. 281–285. – ISSN 1573-1375
- [Piegl und Tiller 1997] PIEGL, L. A. ; TILLER, W.: *The NURBS Book*. 2nd ed. Berlin and New York : Springer, 1997 (Monographs in visual communications). – ISBN 3-540-61545-8
- [Polyak et al. 2004] POLYAK, B. T. ; NAZIN, S. A. ; DURIEU, C. ; WALTER, E.: Ellipsoidal parameter or state estimation under model uncertainty. In: *Automatica* 40 (2004), Nr. 7, S. 1171–1179. – ISSN 0005-1098
- [Ristic et al. 2004] RISTIC, B. ; ARULAMPALAM, S. ; GORDON, N.: *Beyond the Kalman filter: Particle filters for tracking applications*. Boston and Mass : Artech House, 2004
- [Rosenblatt 1956] ROSENBLATT, M.: Remarks on Some Nonparametric Estimates of a Density Function. In: *The Annals of Mathematical Statistics* 27 (1956), Nr. 3, S. 832–837. – ISSN 0003-4851
- [Schaffrin 1987] SCHAFFRIN, B.: Approximating the Bayesian estimate of the standard deviation in a linear model. In: *Journal of Geodesy* 61 (1987), Nr. 3, S. 276–280. – ISSN 1432-1394
- [Schmitt und Neuner 2015] SCHMITT, C. ; NEUNER, Hans: Knot estimation on B-Spline curves. In: *Vermessung & Geoinformation* (2015), Nr. 2-3/2015, S. 188–197
- [Schön 2003] SCHÖN, S.: *Analyse und Optimierung geodätischer Messanordnungen unter besonderer Berücksichtigung des Intervallansatzes*. Karlsruhe, KIT, Promotion, 2003
- [Schuh 2003] SCHUH, W.-D.: The Processing of Band-Limited Measurements; Filtering Techniques in the Least Squares Context and in the Presence of Data Gaps. In: *Space Science Reviews* 108 (2003), Nr. 1, S. 67–78. – ISSN 1572-9672
- [Schweppe 1968] SCHWEPPE, F.: Recursive state estimation: Unknown but bounded errors and system inputs. In: *IEEE Transactions on Automatic Control* 13 (1968), Nr. 1, S. 22–28. – ISSN 0018-9286

- [Siebert und Sommer 2004] SIEBERT, B. R. L. ; SOMMER, K.-D.: Weiterentwicklung des GUM und Monte-Carlo-Techniken. In: *Technisches Messen* 71 (2004), S. 67–80
- [Simon 2006] SIMON, D.: *Optimal state estimation: Kalman, H infinity, and nonlinear approaches* // *Kalman, H [infinity] and nonlinear approaches*. Hoboken, NJ : John Wiley & Sons, Inc, 2006
- [Simon 2010] SIMON, D.: Kalman filtering with state constraints: a survey of linear and nonlinear algorithms. In: *IET Control Theory & Applications* 4 (2010), Nr. 8, S. 1303–1318. – ISSN 17518644
- [Simon und Simon 2010] SIMON, D. ; SIMON, D. L.: Constrained Kalman filtering via density function truncation for turbofan engine health estimation. In: *International Journal of Systems Science* 41 (2010), Nr. 2, S. 159–171. – ISSN 0020-7721
- [Soloviev et al. 2007] SOLOVIEV, A. ; BATES, D. ; VAN GRAAS, F: Tight coupling of laser scanner and inertial measurements for a fully autonomous relative navigation solution. In: *Navigation* 54 (2007), Nr. 3, S. 189–205
- [Steffen und Beder 2007] STEFFEN, R. ; BEDER, C.: Recursive Estimation with Implicit Constraints. In: HAMPRECHT, Fred A. (Hrsg.) ; SCHNÖRR, Christoph (Hrsg.) ; JÄHNE, Bernd (Hrsg.): *Pattern Recognition*. Berlin, Heidelberg : Springer Berlin Heidelberg, 2007, S. 194–203. – ISBN 978-3-540-74936-3
- [Sun et al. 2019] SUN, L. ; ALKHATIB, H. ; KARGOLL, B. ; KREINOVICH, V. ; NEUMANN, I.: Ellipsoidal and Gaussian Kalman Filter Model for Discrete-Time Nonlinear Systems. In: *Mathematics* 7 (2019), Nr. 12, S. 1168
- [Sun et al. 2018] SUN, L. ; ALKHATIB, H. ; PAFFENHOLZ, J.-A. ; NEUMANN, I.: Geo-Referencing of a Multi-Sensor System Based on Set-Membership Kalman Filter. In: FUSION, International Conference on I. (Hrsg.): *2018 21st International Conference on Information Fusion (FUSION)*. Piscataway, NJ : IEEE, 2018, S. 1–5. – ISBN 978-0-9964527-6-2
- [Thrun et al. 2006] THRUN, S. ; BURGARD, W. ; FOX, D.: *Probabilistic robotics*. Cambridge and Mass : MIT Press, 2006 (Intelligent robotics and autonomous agents). – ISBN 0262201623
- [Unger et al. 2017] UNGER, J. ; ROTTENSTEINER, F. ; HEIPKE, C.: ASSIGNING TIE POINTS TO A GENERALISED BUILDING MODEL FOR UAS IMAGE ORIENTATION. In: *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XLII-2/W6* (2017), S. 385–392. – ISSN 2194-9034
- [Vicino und Zappa 1996] VICINO, A. ; ZAPPA, G.: Sequential approximation of feasible parameter sets for identification with set membership uncertainty. In: *IEEE Transactions on Automatic Control* 41 (1996), Nr. 6, S. 774–785. – ISSN 0018-9286
- [Vogel 2020] VOGEL, S: *Kalman Filtering with State Constraints Applied to Multi-sensor Systems and Georeferencing*. Hannover, Leibniz Universität hannover, Dissertation, 2020
- [Vogel et al. 2019] VOGEL, S. ; ALKHATIB, H. ; BUREICK, J. ; MOFTIZADEH, R. ; NEUMANN, I.: Georeferencing of Laser Scanner-Based Kinematic Multi-Sensor Systems in the Context of Iterated Extended Kalman Filters Using Geometrical Constraints. In: *Sensors* 19 (2019), Nr. 10. – ISSN 1424-8220
- [Vogel et al. 2016] VOGEL, S. ; ALKHATIB, H. ; NEUMANN, I.: Accurate Indoor Georeferencing with Kinematic Multi Sensor Systems. In: *IEEE International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, 2016, S. 8–

- [Vogel et al. 2018] VOGEL, S. ; ALKHATIB, H. ; NEUMANN, I.: Iterated Extended Kalman Filter with Implicit Measurement Equation and Nonlinear Constraints for Information-Based Georeferencing. In: FUSION, International Conference on I. (Hrsg.): *2018 21st International Conference on Information Fusion (FUSION)*. Piscataway, NJ : IEEE, 2018, S. 1209–1216. – ISBN 978-0-9964527-6-2
- [Voss 2014] VOSS, J.: *An introduction to statistical computing: A simulation-based approach*. 1st ed. Chichester, West Sussex : Wiley, 2014. – ISBN 978-1-118-35772-9
- [Vuong 1989] VUONG, Q. H.: Likelihood ratio tests for model selection and non-nested hypotheses. In: *Econometrica: Journal of the Econometric Society* 8 (1989), S. 307–333
- [Walter und Piet-Lahanier 1989] WALTER, E. ; PIET-LAHANIER, H.: Exact recursive polyhedral description of the feasible parameter set for bounded-error models. In: *IEEE Transactions on Automatic Control* 34 (1989), Nr. 8, S. 911–915. – ISSN 0018-9286
- [Wegman 1988] WEGMAN, E.: Computational Statistics: A New Agenda for Statistical Theory and Practice. In: *Journal of Washington Academy of Sciences* (1988), Nr. 78, S. 310–322
- [Weitkamp und Alkhatib 2012] WEITKAMP, A. ; ALKHATIB, H.: Bayesischer Ansatz zur Integration von Expertenwissen in die Immobilienbewertung, Teil 2. In: *Zeitschrift für Geodäsie, Geoinformation und Landmanagement (zfv)* 02 (2012), S. 103–114
- [Weitkamp und Alkhatib 2014] WEITKAMP, A. ; ALKHATIB, H.: Die Bewertung kaufpreisarmer Lagen mit multivariaten statistischen Verfahren - Möglichkeiten und Grenzen robuster Methoden bei der Auswertung weniger Kauffälle. In: *allgemeine vermessungs-nachrichten (avn)* (2014), Nr. 1, S. 3–12
- [Williams 1970] WILLIAMS, D. A.: Discrimination between Regression Models to Determine the Pattern of Enzyme Synthesis in Synchronous Cell Cultures. In: *Biometrics* 26 (1970), Nr. 1, S. 23. – ISSN 0006341X
- [Wiśniewski 2014] WIŚNIEWSKI, Z.: M-estimation with probabilistic models of geodetic observations. In: *Journal of Geodesy* 88 (2014), Nr. 10, S. 941–957. – ISSN 1432-1394
- [Witsenhausen 1968] WITSENHAUSEN, H.: Sets of possible states of linear systems given perturbed observations. In: *IEEE Transactions on Automatic Control* 13 (1968), Nr. 5, S. 556–558. – ISSN 0018-9286
- [Wu et al. 2000] WU, Z. ; SEAH, H. S. ; LIN, F.: NURBS Volume for Modelling Complex Objects. In: CHEN, Min (Hrsg.) ; KAUFMAN, Arie E. (Hrsg.) ; YAGEL, Roni (Hrsg.) ; CHEN, M. (Hrsg.) ; KAUFMAN, Arie (Hrsg.): *Volume graphics* Bd. 22. London : Springer, 2000, S. 159–167. – ISBN 978-1-85233-192-4
- [Wunderlich et al. 2016] WUNDERLICH, T. ; NIEMEIER, W. ; WUJANZ, D. ; HOLST, C. ; NEITZEL, F. ; KUHLMANN, H.: Areal deformation analysis from TLS point clouds – the challenge. In: *Allgemeine Vermessungs Nachrichten* 123 (2016), Nr. 11–12, S. 340–351
- [Xu 2005] XU, P.: Sign-constrained robust least squares, subjective breakdown point and the effect of weights of observations on robustness. In: *Journal of Geodesy* 79 (2005), Nr. 1-3, S. 146–159. – ISSN 0949-7714
- [Xu et al. 2017a] XU, X. ; BUREICK, J. ; YANG, H. ; NEUMANN, I.: TLS-based composite structure deformation analysis validated with laser tracker. In: *Composite Structures* in press (2017). – ISSN 02638223

- [Xu et al. 2018] XU, X. ; KARGOLL, B. ; BUREICK, J. ; YANG, H. ; ALKHATIB, H. ; NEUMANN, I.: TLS-based profile model analysis of major composite structures with robust B-spline method. In: *Composite Structures* 184 (2018), S. 814–820. – ISSN 02638223
- [Xu et al. 2017b] XU, X. ; ZHAO, X. ; YANG, H. ; NEUMANN, I.: TLS-based feature extraction and 3D modeling for arch structures. In: *Sensors* 2017 (2017), S. 1–8. – ISSN 1687-725X
- [Yang et al. 2005] YANG, C. ; BAKICH, M. ; BLASCH, E.: Nonlinear constrained tracking of targets on roads. In: *8th International Conference on Information Fusion, 2005*. Piscataway, NJ : IEEE Service Center, 2005, S. 8 pp. – ISBN 0-7803-9286-8
- [Yang et al. 2017] YANG, H. ; OMIDALIZARANDI, M. ; XU, X. ; NEUMANN, I.: Terrestrial laser scanning technology for deformation monitoring and surface modeling of arch structures. In: *Composite Structures* 169 (2017), S. 173–179
- [Yoshimoto et al. 2003] YOSHIMOTO, F. ; HARADA, T. ; YOSHIMOTO, Y.: Data fitting with a spline using a real-coded genetic algorithm. In: *Computer-Aided Design* 35 (2003), Nr. 8, S. 751–760. – ISSN 00104485
- [Yuanxi 1991] YUANXI, Y.: Robust bayesian estimation. In: *Bulletin Géodésique* 65 (1991), Nr. 3, S. 145–150
- [Zhao et al. 2018] ZHAO, X. ; KARGOLL, B. ; OMIDALIZARANDI, M. ; XU, X. ; ALKHATIB, H.: Model Selection for Parametric Surfaces Approximating 3D Point Clouds for Deformation Analysis. In: *Remote Sensing* 10 (2018), Nr. 4, S. 634
- [Zhao et al. 2011] ZHAO, X. ; ZHANG, C. ; YANG, B. ; LI, P.: Adaptive knot placement using a GMM-based continuous optimization algorithm in B-spline curve approximation. In: *Computer-Aided Design* 43 (2011), Nr. 6, S. 598–604. – ISSN 00104485
- [Zhu et al. 2005] ZHU, J. ; SANTERRE, R. ; CHANG, X.-W.: A Bayesian method for linear, inequality-constrained adjustment and its application to GPS positioning. In: *Journal of Geodesy* 78 (2005), Nr. 9, S. 528–534. – ISSN 0949-7714

Danksagung

Abschließend möchte ich mich bei allen Personen bedanken, die zur Fertigstellung dieser Arbeit beigetragen haben.

An erster Stelle möchte ich mich bei Prof. Dr.-Ing. Ingo Neumann bedanken. Ich schätze es sehr, dass er mir durch seine Ratschläge und Ermutigung eine spannende und abwechslungsreiche Postdoktorandenzeit am Geodätischen Institut Hannover ermöglicht hat. Seine unabdingbare Unterstützung hat diese Arbeit erst möglich gemacht. Außerdem möchte ich mich bei den Referenten Prof. Dr.-Ing. Michael Beer, Prof. Dr.-Ing. habil. Hansjörg Kutterer und Prof. Dr.-Ing. Nico Sneeuw dafür, dass sie sich die Zeit genommen haben, diese Arbeit zu begutachten und wertvolle Hinweise zu geben. Ich danke auch Herrn Prof. Dr.-Ing. Steffen Schön herzlich für die Bereitschaft, den Vorsitz der Habilitationskommission zu übernehmen.

Einen sehr großen Anteil am erfolgreichen Abschluss trägt Prof. Dr.-Ing. Boris Kargoll; für die vielen intensiven Diskussionen, wertvollen Anregungen zur Fokussierung der Inhalte und die persönliche Unterstützung während dieser Zeit möchte ich mich herzlich bei ihm bedanken.

Darüber hinaus möchte ich mich bei allen Kolleginnen und Kollegen im GIH bedanken, die mich in vielfältiger Weise begleitet und unterstützt haben. Besonderer Dank gebührt Herrn Dr.-Ing. Sören Vogel für seine enorme Unterstützung und für die vielen hilfreichen Diskussionen.

Weiterhin möchte ich mich bei allen Kolleginnen und Kollegen im Graduiertenkolleg i.c.sens 2159 und im Projekt *Bayessche Adaptive Robuste Ausgleichung von multivariaten geodätischen Messprozessen mit Datenlücken und nichtstationärem farbigem Rauschen* (BARA) bedanken. Die Möglichkeit, an diesen Projekten mitzuwirken, ist für mich in vielerlei Hinsicht sehr wertvoll; ich habe von zahlreichen Diskussionen sehr profitiert. In diesem Zusammenhang möchte ich mich auch bei der DFG für die Finanzierung des Graduiertenkolleg i.c.sens und des Projekts BARA mit der Nummer 386369985 bedanken.

Mein letzter, aber größter Dank geht jedoch an meiner Familie: Taghrid, Khaled und Salma. Meine Frau Taghrid hat es mir ermöglicht, meinen schwierigen Weg zu gehen und dabei immer sich auf ihre Unterstützung und Liebe verlassen zu können. Sie hat mir stets Rückhalt und Kraft gegeben.

Lebenslauf

Persönliche Daten

Name: Hamza Alkhatib
Geburstag: 05.06.1969
Geburtsort: Damaskus, Syrien
Familienstand: verheiratet

Berufliche Laufbahn

08.2007 bis heute Wissenschaftlicher Mitarbeiter am Geodätischen Institut der Leibniz Universität Hannover
01.2002 bis 07.2007 Wissenschaftlicher Mitarbeiter am Institut für Geodäsie und Geoinformation, Universität Bonn
09.1993 bis 06.1997 Katasteramt am Ministerium für lokale Administration und Umwelt Damaskus (Syrien)

Ausbildung

01.2002 bis 06.2007 Promotion am Institut für Geodäsie und Geoinformation, Universität Bonn, Titel: On Monte Carlo Methods with Applications to the Current Satellite Gravity Missions, Referent: Prof. Dr. techn. Wolf-Dieter Schuh, Korreferent: Prof. Dr.-Ing. Jürgen Kusche, Abschluss: Dr.-Ing.
04.1998 bis 09.2001 Studium Vermessungswesen am Karlsruher Institut für Technologie, Abschluss: Dipl.-Ing.
08.1988 bis 09.1993 Studium Bauingenieurwesen an der Universität Damaskus, Abschluss: Bachelor of Science
1982 bis 1988 Oberschule Saada Damaskus, Abschluss: Abitur

