International Association of Geodesy Symposia
140

Hansjörg Kutterer
Florian Seitz
Hamza Alkhatib
Michael Schmidt *Editors*

The 1st International Workshop on the Quality of Geodetic Observation and Monitoring Systems (QuGOMS'11)

Proceedings of the 2011 IAG International Workshop, Munich, Germany, April 13–15, 2011

Springer

International Association of Geodesy Symposia

Chris Rizos, Series Editor
Pascal Willis, Assistant Series Editor

International Association of Geodesy Symposia

Chris Rizos, Series Editor
Pascal Willis, Assistant Series Editor

Symposium 101: Global and Regional Geodynamics
Symposium 102: Global Positioning System: An Overview
Symposium 103: Gravity, Gradiometry, and Gravimetry
Symposium 104: Sea SurfaceTopography and the Geoid
Symposium 105: Earth Rotation and Coordinate Reference Frames
Symposium 106: Determination of the Geoid: Present and Future
Symposium 107: Kinematic Systems in Geodesy, Surveying, and Remote Sensing
Symposium 108: Application of Geodesy to Engineering
Symposium 109: Permanent Satellite Tracking Networks for Geodesy and Geodynamics
Symposium 110: From Mars to Greenland: Charting Gravity with Space and Airborne Instruments
Symposium 111: Recent Geodetic and Gravimetric Research in Latin America
Symposium 112: Geodesy and Physics of the Earth: Geodetic Contributions to Geodynamics
Symposium 113: Gravity and Geoid
Symposium 114: Geodetic Theory Today
Symposium 115: GPS Trends in Precise Terrestrial, Airborne, and Spaceborne Applications
Symposium 116: Global Gravity Field and Its Temporal Variations
Symposium 117: Gravity, Geoid and Marine Geodesy
Symposium 118: Advances in Positioning and Reference Frames
Symposium 119: Geodesy on the Move
Symposium 120: Towards an Integrated Global Geodetic Observation System (IGGOS)
Symposium 121: Geodesy Beyond 2000: The Challenges of the First Decade
Symposium 122: IV Hotine-Marussi Symposium on Mathematical Geodesy
Symposium 123: Gravity, Geoid and Geodynamics 2000
Symposium 124: Vertical Reference Systems
Symposium 125: Vistas for Geodesy in the New Millennium
Symposium 126: Satellite Altimetry for Geodesy, Geophysics and Oceanography
Symposium 127: V Hotine Marussi Symposium on Mathematical Geodesy
Symposium 128: A Window on the Future of Geodesy
Symposium 129: Gravity, Geoid and Space Missions
Symposium 130: Dynamic Planet - Monitoring and Understanding . . .
Symposium 131: Geodetic Deformation Monitoring: From Geophysical to Engineering Roles
Symposium 132: VI Hotine-Marussi Symposium on Theoretical and Computational Geodesy
Symposium 133: Observing our Changing Earth
Symposium 134: Geodetic Reference Frames
Symposium 135: Gravity, Geoid and Earth Observation
Symposium 136: Geodesy for Planet Earth
Symposium 137: VII Hotine-Marussi Symposium on Mathematical Geodesy
Symposium 138: Reference Frames for Applications in Geosciences
Symposium 139: Earth on the Edge: Science for a Sustainable Planet

More information about this series at
http://www.springer.com/series/1345

The 1st International Workshop on the Quality of Geodetic Observation and Monitoring Systems (QuGOMS'11)

Proceedings of the 2011 IAG International Workshop, Munich, Germany, April 13 – 15, 2011

Edited by

Hansjörg Kutterer
Florian Seitz
Hamza Alkhatib
Michael Schmidt

Springer

Volume Editors

Hansjörg Kutterer
Bundesamt für Kartographie und Geodäsie
Frankfurt am Main
Germany

Florian Seitz
Technische Universität München
München
Germany

Hamza Alkhatib
Leibniz Universität Hannover
Geodätisches Institut
Hannover
Germany

Michael Schmidt
Deutsches Geodätisches Forschungsinstitut
München
Germany

Series Editors

Chris Rizos
School of Surveying
University of New South Wales
Sydney
Australia

Assistant Series Editors

Pascal Willis
Institut national de l'Information
Geographique et Forestiere
Direction Technique
Saint-Mande
France

ISSN 0939-9585
ISBN 978-3-319-10827-8 ISBN 978-3-319-10828-5 (ebook)
DOI 10.1007/978-3-319-10828-5
Springer Cham Heidelberg New York Dordrecht London

Library of Congress Control Number: 2014956880

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

The 1st international Workshop on the Quality of Geodetic Observation and Monitoring (QuGOMS) was devoted to the general methodology in the field of estimation and filtering with a refined uncertainty modelling emphasizing applications in engineering geodesy and Earth system observation. Geodetic multi-sensor systems and networks using terrestrial and space-borne observation techniques were thematic anchor points.

The QuGOMS workshop has been organized jointly by the study groups IC-SG2 and IC-SG3 of the Intercommission Committee on Theory (ICCT) of the International Association of Geodesy (IAG). Besides its strong relations to all IAG Commissions the workshop was also in the scope of FIG Commissions 5 and 6. Thus, it attracted scientists under the umbrellas of both IAG and FIG. The workshop took place in the rooms of the International Graduate School of Science and Engineering (IGSSE) of the Technische Universität München, Garching/Munich from 13th to 15th April 2011.

The workshop was organized in five regular sessions. To a large extent, the sessions' topics referred to the subjects of the IC-SG2 and IC-SG3 of the ICCT:

- Uncertainty modelling of geodetic data
- Theoretical studies on combination strategies and parameter estimation
- Recursive state-space filtering
- Sensor networks and multi-sensor systems in engineering geodesy
- Multi-mission approaches with view to physical processes in the Earth system

The contributed papers showed quite well the related questions and hence the close connection of methodology in the different fields of application such as global geodesy and engineering geodesy. Without doubt, it is worthwhile to continue this kind of workshop to foster scientific exchange also between scientific organizations.

There were several colleagues who contributed to the success of the workshop and of the proceedings. All editors of this volume acted also as convenors. Hamza Alkhatib coordinated the reviews of the submitted papers and communicated with the symposium editors and with Springer. Otto Heunecke took care of the contributions from engineering geodesy. Various reviewers helped to ensure a valid review process. Florian Seitz acted as local host. All in all this is gratefully acknowledged.

Frankfurt am Main, Germany — Hansjörg Kutterer
12 September 2013

Contents

Part III Recursive State-Space Filtering

Part IV Sensor Networks and Multi Sensor Systems in Engineering Geodesy

Part V Multi-Mission Approaches with View to Physical Processes in the Earth System

Part I

Uncertainty Modeling of Geodetic Data

Modeling Data Quality Using Artificial Neural Networks

Ralf Laufer and Volker Schwieger

Abstract

Managing data quality is an important issue in all technical fields of applications. Demands on quality-assured data in combination with a more diversified quality description are rising with increasing complexity and automation of processes, for instance within advanced driver assistance systems (ADAS). Therefore it is important to use a comprehensive quality model and furthermore to manage and describe data quality throughout processes or sub-processes.

This paper focuses on the modeling of data quality in processes which are in general not known in detail or which are too complex to describe all influences on data quality. As emerged during research, artificial neural networks (ANN) are capable for modeling data quality parameters within processes with respect to their interconnections.

Since multi-layer feed-forward ANN are required for this task, a large number of examples, depending on the number of quality parameters to be taken into account, is necessary for the supervised learning of the ANN, respectively determining all parameters defining the net. Therefore the general usability of ANN was firstly evaluated for a simple geodetic application, the polar survey, where an unlimited number of learning examples could be generated easily. As will be shown, the quality parameters describing accuracy, availability, completeness and consistency of the data can be modeled using ANN. A combined evaluation of availability, completeness or consistency and accuracy was tested as well. Standard deviations of new points can be determined using ANN with sub-mm accuracy in all cases.

To benchmark the usability of ANN for a real practical problem, the complex process of mobile radio location and determination of driver trajectories on the digital road network based on these data, was used. The quality of calculated trajectories could be predicted sufficiently from a number of relevant input parameters such as antenna density and road density. The cross-deviation as an important quality parameter for the trajectories could be predicted with an accuracy of better than 40 m.

R. Laufer • V. Schwieger (✉)
Institute of Engineering Geodesy, University of Stuttgart,
Geschwister-Scholl-Str. 24D, 70174 Stuttgart, Germany
e-mail: volker.schwieger@ingeo.uni-stuttgart.de

H. Kutterer et al. (eds.), *The 1st International Workshop on the Quality of Geodetic Observation and Monitoring Systems (QuGOMS'11)*, International Association of Geodesy Symposia 140,
DOI 10.1007/978-3-319-10828-5_1,

Keywords
Data quality • Neural networks • Propagation of data quality

1 Introduction

Managing data quality requires first of all an adequate and homogenous quality model. The model which is used in the following was developed by Wiltschko (2004) and Wiltschko and Kaufmann (2005) and consists of six inherent quality characteristics. The six quality characteristics of the used quality model are shortly defined as follows:

- Availability: Measure for existence of information at a certain time and at a certain place.
- Timeliness: Measure for correlation of information with temporally changing reality.
- Completeness: Measure for existence of all relevant information to describe reality.
- Consistency: Measure for correlation of information with the information model.
- Correctness: Measure for the correlation of information with reality, timeliness assumed.
- Accuracy: Describes the correlation between determined value and real value.

Depending on the kind of data, these quality characteristics can be concretized each with several quantitative quality parameters. Quality measurement methods are needed as well to determine values for these quantitative quality parameters. With this model all relevant aspects of data quality can be sufficiently described.

Considering a data management process, the occurring input and output data can be described by quality parameters which are part of individual quality models defined for the different data. Since output data is generated from input data, their quality can be determined from the quality of the input data. However the relation between input and output data quality can be quite complex and often cannot be described in an analytical way. In some cases the dependencies are only partly known. Therefore a robust method is necessary which can handle this lack of information. As it turned out, artificial neural networks are more capable to solve this problem than other methods such as Petri nets or Monte-Carlo-simulation (Laufer 2011). An important assumption is that all inputs that influence the output significantly are captured (ideal case).

2 Data Quality Propagation with Artificial Neural Networks

In the following part, ANN will be briefly introduced and their most important advantages with respect to the other methods mentioned in the introduction are presented.

Artificial neural networks are a method which was developed around 70 years before (Zell 1997). The motivation was to copy the human brain in design and principle of operation. The networks consist of neural cells linked by artificial nerve tracts which can transmit information in an intelligent way (e.g. threshold based). Starting with very simple networks of artificial neural cells, more and more complex networks were developed over time, capable to solve more complex problems in combinatorics, pattern recognition, diagnostics and other fields of application where normal computers still get to their limits. A broad variety of applications is listed, for example, in Hagan et al. (1996). The main advantages of ANN with respect to other methods described in Laufer (2011) are:

- Massive parallel work flow (fast),
- Ability to learn and adapt,
- Ability to generalize and associate and
- High fault tolerance.

The artificial neural network can be trained in different ways depending on the kind of network and the application. In this paper only feed-forward networks and therefore supervised learning can be determined successfully. For supervised learning a sufficient amount of learning examples is necessary depending on the complexity of the process as well as the number of different input and output parameters.

In feed-forward networks, all neurons are arranged in several levels. Typically a network consists of 2–3 layers. Figure 1 shows a single artificial neuron with R inputs with the weights $w_{i,j}$ for the R different input parameters p_j of the neuron no. i. b describes the bias of the neuron. The inputs are used in the transfer function Σ to generate the net input n. The activation function f generates the net output a. The flow of information in these kinds of ANN is always fixed in one direction.

To train the network, training data sets representing the whole possible data range are necessary. Validation data sets are used to check the learning status of the net after each iteration. After finishing the training, the net performance must be evaluated independently using new data sets not yet known to the net, so-called evaluation data.

3 Propagation of Data Quality for Polar Point Determination

The polar point determination is a quite simple method to calculate Cartesian point coordinates from measured distance and direction to a new point. In the first step ANN were used

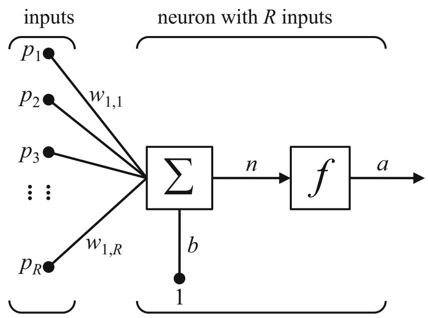

Fig. 1 Artificial neuron with R inputs

to propagate the two quality parameters standard deviation (in the following abbreviated with "deviation") in longitudinal and cross direction (l and c) referring to the direction to the point. These two output parameters describing the accuracy are besides others related to the following input parameters:

- Measured distance (s)
- Standard deviation of distance (σ_s)
- Distance-related correction of measured distance (ppm)

The additional handling of availability of input data was tested in the second step. Unavailable data were flagged with the value "0". For mathematical reasons it was necessary to normalize all values (since the input parameters may vary heavily in magnitude, e.g. distances up to several km measured with a standard deviation of a few mm). For the generation of training and evaluation data, it was necessary to create a fault tree for different possible lacks in completeness (cf. Laufer 2011) for further details.

The additional handling of quality parameters describing availability demands a more complex network. For the single handling of accuracy parameters, a network with one hidden layer containing nine neurons and two output neurons (abbrev.: [9-2]) was sufficient. For additional propagation of availability a [15-15-2]-network with two hidden layers was necessary.

As shown in Fig. 2, the accuracy parameters l and c can be propagated with an uncertainty below 0.1 mm. To avoid the network memorizing the learning data sets, new example data sets not used for training or validation of the network were used to evaluate the trained network. The input data were chosen randomly out of the trained intervals.

Handling availability in addition leads to similar results: l and c were propagated with an accuracy of better than 0.1 mm, the availability with an accuracy of better than 0.03. Hence it is possible to distinguish between small but real values near zero and not unavailable data, flagged with "0".

In a third step the propagation of parameters describing completeness was determined. Therefore it was first of all necessary to arrange the input and output data in data sets,

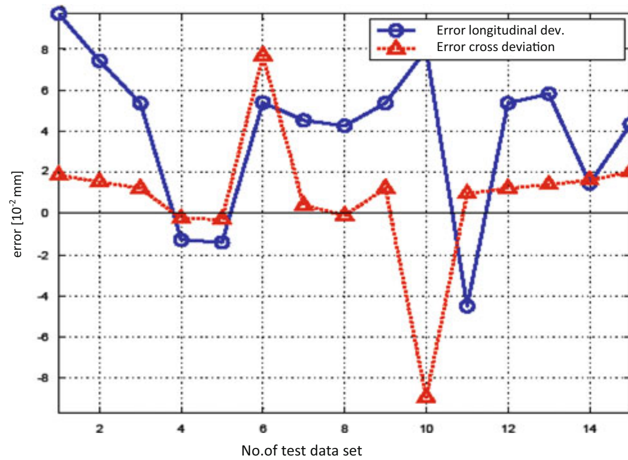

Fig. 2 Polar point determination: Propagation of accuracy (evaluation data)

which can contain one or more single parameters. Otherwise it would not be possible to separate between a lack of completeness and a lack of availability. If a complete data set is missing, it is per definition a lack of availability. If one parameter within a data set is missing, it is a lack of completeness. To simplify matters, in this example all input parameters and all output parameters where defined as data sets.

For propagation of completeness, the input and output parameter vectors were expanded by one flag for each data set to provide information about completeness. The flag is a binary character, where the value "1" stands for "data set is complete". Whereas "0" means "data set is incomplete".

To keep the number of training data sets small and to get the ANN well trained at the same time, it is important to simulate a much higher lack in completeness during training than will occur in reality. Otherwise, in a small amount of training data, there are only a few lacks in completeness occurring. Hence the network would not be able to learn how to handle incomplete data sets.

Again, the accuracy parameters can be propagated with an accuracy of better than ±0.1 mm (Fig. 3). The completeness of the output data sets can be determined better than ±0.04. This means the completeness can be easily propagated by adding new binary parameters to the input and output vectors.

Propagation of consistency and correctness of the data can be handled as well by introducing additional binary characters. The distinction between binary values describing completeness, consistency and correctness happens automatically by the clearly defined and constant position of the parameters within the vectors.

Handling timeliness of data, on the contrary, demands time-sensitive ANN, so-called dynamic networks that can handle time-invariant data. Since the focus of our research

Fig. 3 Polar point determination: Propagation of accuracy and completeness (evaluation data)

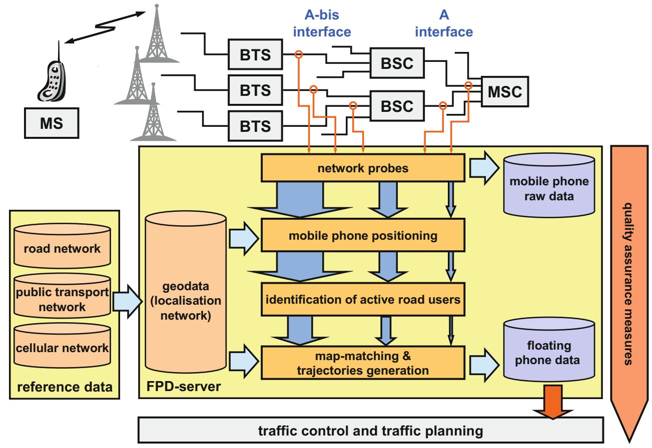

Fig. 4 Process of FPD generation (Wiltschko et al. 2007)

was on static networks, the quality character timeliness has not been investigated so far.

4 Propagation of Data Quality for Mobile Phone Positioning

The project Do-iT (data optimization for integrated telematics) with a duration of almost 4 years was funded by the Federal Ministry of Economics and Technology. The focus of the project was the generation of mobile phone trajectories, so called floating phone data (FPD), of individual motorized road users within the main road network for traffic applications (e. g. congestion detection or traffic planning). Generation of FPD is a complex process consisting of several sub-processes as outlined in Fig. 4.

At the Institute of Engineering Geodesy Stuttgart (IIGS) two different approaches regarding two different interfaces for data recording where developed. The A-interface includes data of all registered mobile phone users, not depending on the status of the phone (active or only switched on). The data on the second interface, the so-called Abis-interface ("bis" is the French word for "bonus"), contain only data of active mobile phones (phones during a call or data connection). These data always have a high spatial and temporal resolution, whereas the data from A-interface only have a medium resolution if the phone is active, otherwise a low resolution.

For practical reasons, the A-data can serve as practical example to evaluate the use of ANN for propagation of data quality. Therefore only the approach for A-bis data will be briefly described in the following.

In a first step the recorded data are sorted and prioritized in near-realtime with respect to the prospected calculability and quality of generated FPD. Static mobile phones (e.g. in buildings beside the roads), phones in trains or tram lines are to be eliminated. In a second step, single positions are generated, using beside other information, the measured signal strengths from neighboring radio cells and the theoretical signal strength propagation maps for each radio cell. The sequence of positions of each user can be converted into a sequence of road elements on the digital road network (thus a trajectory) using map-aiding methods (Czommer 2000). See Ramm and Schwieger (2008) or Do-iT (2009) to read more about the methods.

For evaluation of the generated trajectories, an empirical quality parameter describing the correctness was introduced. The parameter *cross deviation* describes the mean perpendicular deviation of a sequence of positions from the most likely route of the user on the digital road network. In Fig. 5 the situation is shown graphically. The figure is a screenshot from a project evaluation report (Do-iT 2009), therefore a reference trajectory measured by GPS is visible as well.

In the following it will be shown that the quality parameter cross deviation can be propagated sufficiently with ANN near realtime. The propagation of quality can be triggered immediately after the generation of the FPD trajectory. The following influence parameters were identified and investigated in Laufer (2011):

- Length of trajectory in m,
- Duration of trajectory in s,
- Mean antenna density in antennas per km^2,
- Difference in antenna density in antennas per km^2,
- Mean density of road network in road elements per km^2,
- Difference in density of road network in road elements per km^2,
- Number of road elements the trajectory consists of.

Besides these seven input parameters there are more parameters which have a weaker and more or less diffuse

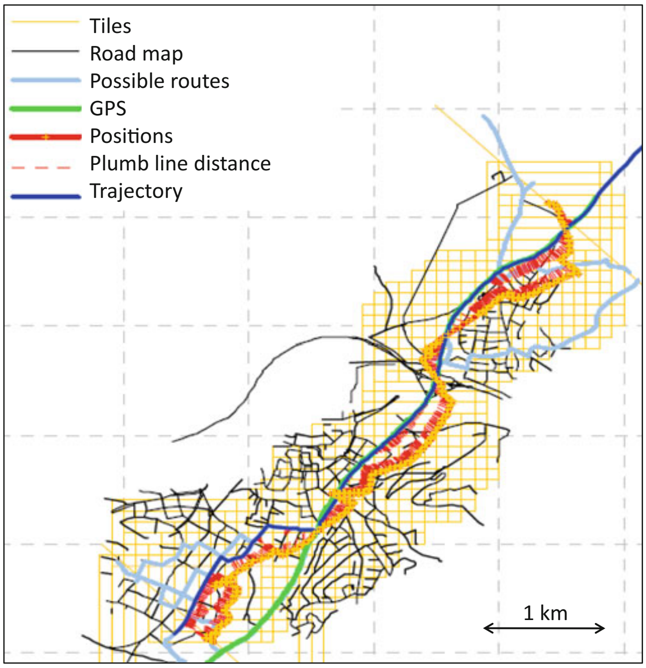

Fig. 5 Evaluation of a generated FPD trajectory with GPS

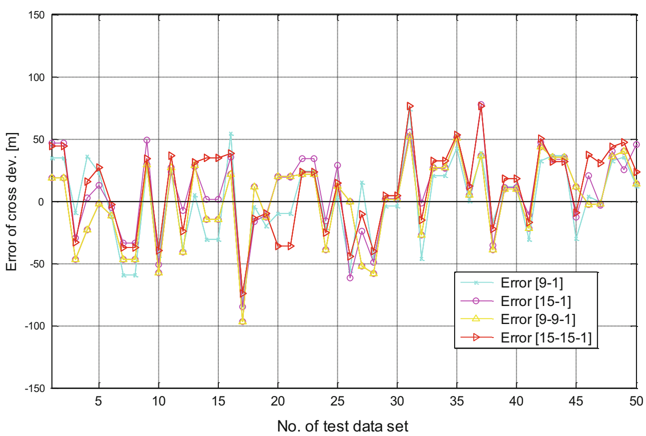

Fig. 6 Propagation of cross deviation (Laufer 2011)

influence on the quality of the FPD. These parameters were not considered within this research.

The data base for training the ANN consists of 8 recorded days. The mean cross deviation reached 180 m with a mean length of trajectories of 6.5 km and a duration of 250 s. All data was recorded in the test area of Karlsruhe and Ettlingen where denser urban road networks dominate. Only some smaller areas on the edges of the area have a more rural character.

To find the best network configuration, different net dimensions were tested as can be seen in Fig. 6 where the results for one single day (March 26th 2009) are displayed. The error of the propagated cross deviation within 50

Table 1 Results for testing different net configurations and different amount of data

Net-configuration	Mean square error	Max. dev. [m]	Standard deviation [m]
9-9-1/15-15-1 best of each day	0.024	93.9	41.0
9-9-1 (all 8 days together)	0.025	103.8	41.0
15-15-1 (all 8 days together)	0.025	108.8	38.2
15-15-1 (all 8 days) 6 extrema eliminated	0.024	80.1	27.9

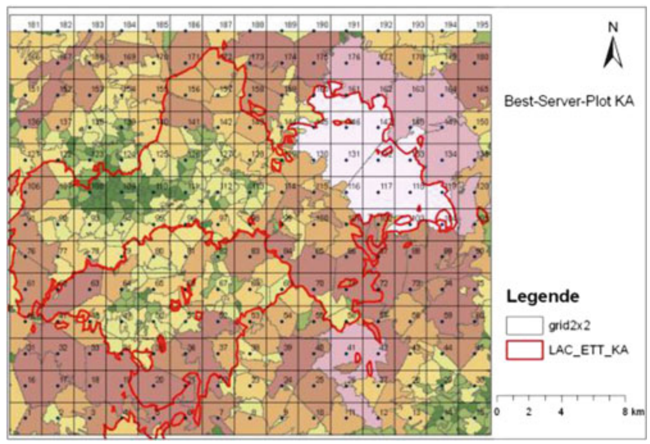

Fig. 7 Best-Server-Plot for Karlsruhe and Ettlingen

randomly chosen test data sets does not exceed 100 m for all networks. The network [15-15-1] does perform best.

As can be seen from Table 1, one single day data is enough to get the ANN trained. The standard deviation of all 50 test data sets reaches 41 m in both cases (single days versus all 8 days together). If a network with 15 neurons in each hidden layer is used, the ANN trained with all 8 days of data together performs a bit better than with a single day and a standard deviation of 38.2 m was reached. Elimination of the six most extremal cross deviations leads to a standard deviation of only 27.9 m.

Since all results are reproducible, ANN seem to be a practical method to propagate cross deviation. Nevertheless, extremal values in cross deviation are more difficult to handle for ANN than more regular values. In a whole, ANN are capable to propagate cross deviation with a standard deviation of 30–40 m. To get a better idea of how good or bad this result is, a closer look into the process of generating FPD is necessary.

Figure 7 shows the test area of Karlsruhe with the radio cells where the Abis-data were available (inside the red surrounded area). The radio cells are very inhomogeneous in size, depending on the density of the road network and the expected amount of radio cells to handle. Since the single positions of road users are extracted from mobile phone data, particularly from the measured signal strength, the standard deviation of this method reaches 560 m for a single position (Do-iT 2009). This means that the identification of the correct route within a denser road network, as in most parts of the test area, is difficult. Furthermore the connection between the correctness of a route and the cross deviation is only based on a small number of test trajectories where GPS as reference was onboard. Therefore, the cross deviation is not a exact quantitative reference, it should be considered as an indicator to evaluate the trajectories.

Regarding these facts, the results reached with ANN are quite satisfactory. It is possible to map the dependencies between input data quality and the cross deviation as output quality parameter.

Conclusion

As could be shown for two examples, polar point determination and mobile phone positioning, ANN are capable for propagation of data quality within processes. However, the results depend on the modeling of the ANN as well as the quality of the training data. Finding the best performing network can only be done empirically by checking, for instance, the mean square error. The use of dynamic ANN provides the opportunity to handle timelines and should be part of further research. To get more experiences in using ANN and to prove their general usability for propagation of data quality more practical examples are necessary.

References

Czommer R (2000) Leistungsfähigkeit fahrzeugautonomer Ortungsverfahren auf der Basis von Map-Matching-Techniken. Bayerische Akademie der Wissenschaften, Verlag C. H. Beck, DGK, Reihe C, Nr. 535, PhD thesis

Do-iT (2009) Qualitätsbewertung Map-Matching gestützter Positionsbestimmungsverfahren auf Abis- und A-Ebene sowie Evaluierung der Positionsbestimmung durch höherwertig Ortungsverfahren. Internal project report, unpublished

Hagan MT, Demuth HB, Beale M (1996) Neural network design. PWS Publishing Company Division of Thomson Learning, USA

Laufer R (2011) Prozedurale Qualitätsmodellierung und-management für Daten – ingenieurtechnische und verkehrstechnische Anwendungen. University of Stuttgart, Institute of Engineering Geodesy, Bayerische Akademie der Wissenschaften, Verlag C. H. Beck, DGK, Reihe C, Nr. 662, PhD thesis

Ramm K, Schwieger V (2008) Mobile positioning for traffic state acquisition. J Location Based Services, Taylor & Francis, London

Wiltschko T (2004) Sichere Information durch infrastrukturgestützte Fahrerassistenzsysteme zur Steigerung der Verkehrssicherheit an Straßenknotenpunkten, VDI, Reihe 12, Nr. 570, VDI-Verlag, Düsseldorf, 2004, PhD thesis

Wiltschko T, Kaufmann T (2005) Modellierung und Bewertung von Prozessen in der Geodatenverarbeitung. In: Proceedings on AGIT Symposium und Fachmesse angewandte Geoinformatik, Universität Salzbug

Wiltschko T, Möhlenbrink W, Schwieger V (2007) Acquisition of traffic state information by mobile phone positioning. In: Proceedings on 6th European congress on ITS, Aalborg, Denmark

Zell A (1997) Simulation Neuronaler Netze, 2nd edn. München, Oldenbourg

Magic Square of Real Spectral and Time Series Analysis with an Application to Moving Average Processes

I. Krasbutter, B. Kargoll, and W.-D. Schuh

Abstract

This paper is concerned with the spectral analysis of stochastic processes that are real-valued, one-dimensional, discrete-time, covariance-stationary, and which have a representation as a moving average (MA) process. In particular, we will review the meaning and interrelations of four fundamental quantities in the time and frequency domain, (1) the stochastic process itself (which includes filtered stochastic processes), (2) its autocovariance function, (3) the spectral representation of the stochastic process, and (4) the corresponding spectral distribution function, or if it exists, the spectral density function. These quantities will be viewed as forming the corners of a square (the "magic square of spectral and time series analysis") with various connecting lines, which represent certain mathematical operations between them. To demonstrate the evaluation of these operations, we will discuss the example of a q-th order MA process.

Keywords

Moving average process • Spectral analysis • Stochastic process • Time series analysis

1 Introduction

The spectral analysis of deterministic functions and the formulation of stochastic processes belong to the well-established statistical tools in various fields within geodesy (see, e.g. Koch and Schmidt 1994; Moritz 1989; Welsch et al. 2000). We found, however, that in particular the nature of the spectral representation of stochastic processes in terms of the stochastic Fourier integral and its relationships with the autocovariance and spectral distribution (or density) function is far less well known than the details of the time-domain and Fourier analyses of deterministic functions. Our motivation for this paper is therefore to take a step towards closing this gap in understanding. We will in particular provide the reader with the key definitions of the involved stochastic processes as well as of their crucial properties (Sect. 2). Then we will state and explain the computational formulae for the spectral analysis of general real-valued covariance-stationary stochastic processes (Sect. 3). This is in contrast to the usual representation of these formulae in the mathematical statistics oriented literature (e.g. Brockwell and Davis 1991; Priestley 2004), where one generally finds only the results for complex-valued stochastic processes, which often complicates their application in practical situations. To aid the understanding of the mathematical relationships of the involved fundamental statistical quantities (stochastic process, autocovariance function, spectral representation of the process, spectral distribution or density function) we will use a corresponding graphical representation in form of a "magic square" (also in Sect. 3). We will conclude this paper with an outlook to extensions to the presented example [moving average (MA) process].

I. Krasbutter • B. Kargoll (✉) • W.-D. Schuh
Institute of Geodesy and Geoinformation, University of Bonn, Nussallee 17, 53115 Bonn, Germany
e-mail: bkargoll@geod.uni-bonn.de

H. Kutterer et al. (eds.), *The 1st International Workshop on the Quality of Geodetic Observation and Monitoring Systems (QuGOMS'11)*, International Association of Geodesy Symposia 140,
DOI 10.1007/978-3-319-10828-5_2,

2 Basic Elements of Stochastic Processes

In this chapter, we will provide a summary of basic definitions (D) and properties of the stochastic processes considered in Sect. 3.

(D1): We say that $\mathcal{X}_T = (\Omega, \mathcal{A}, P, \{\mathcal{X}_t, t \in T\})$ is a **(general) stochastic process** if and only if (iff)

- $(\Omega, \mathcal{A}, P)$ is any probability space (where Ω denotes the sample space, $\mathcal{A}$ a σ-algebra of events, and P a probability measure),
- T is any non-empty set, and
- $\mathcal{X}_t$ is a random variable defined on $(\Omega, \mathcal{A})$ for any $t \in T$.

In this paper, we will restrict our attention to real-valued and one-dimensional stochastic processes as given in the following definition.

(D2): We say that $\mathcal{X}_T$ is a **real-valued (one-dimensional) stochastic process** iff $\mathcal{X}_t : (\Omega, \mathcal{A}) \to (\mathbb{R}, \mathcal{B})$ for any $t \in T$, where $\mathcal{B}$ is the Borel sigma algebra generated by the set of all real-valued, one-dimensional, left-open and right-closed intervals.

In Sect. 3, we will use stochastic processes that have a discrete parameter set T in the time domain as well as processes with a continuous parameter set in the frequency domain. This distinction is made by the following definition.

(D3): We say that $\mathcal{X}_T$ is a

- **discrete-parameter stochastic process** (or stochastic process with discrete parameter) iff $T \subset \mathbb{Z}$. Furthermore, we call $\mathcal{X}_T$ a **discrete-time stochastic process** or **discrete-time time series** iff the elements of T refer to points in time.
- **continuous-parameter stochastic process** (or stochastic process with continuous parameter) iff $T \subset \mathbb{R}$. In addition, we call $\mathcal{X}_T$ a **continuous-frequency stochastic process** iff the elements of T refer to (angular) frequencies, in which case we will also write $T = W$.

As far as discrete-parameter stochastic processes are concerned, we will focus our attention on covariance-stationary processes (in the time domain). The precise meaning of this concept is provided as follows.

(D4): We say that $\mathcal{X}_T$ is **covariance stationary** iff

- $E\{\mathcal{X}_t\} = \mu < \infty$ (i.e. constant/finite) for any $t \in T$,
- $E\{(\mathcal{X}_t - \mu)^2\} = \sigma^2_{\mathcal{X}} < \infty$ (i.e. constant/finite) for any $t \in T$, and
- $\gamma_{\mathcal{X}}(t_1, t_2) = \gamma_{\mathcal{X}}(t_1 + \Delta t, t_2 + \Delta t)$ for any $t_1, t_2 \in T$ and any Δt with $t_1 + \Delta t, t_2 + \Delta t \in T$,

where $E\{.\}$ denotes the expectation operator and $\gamma_{\mathcal{X}}$ the autocovariance function, defined by $\gamma_{\mathcal{X}}(t_1, t_2) = E\{(\mathcal{X}_{t_1} - \mu)((\mathcal{X}_{t_2} - \mu)\}$. For a covariance-stationary stochastic process, we have that $\gamma_{\mathcal{X}}(t_1, t_2) = \gamma_{\mathcal{X}}(t_1 - t_2, 0)$ for any $t_1, t_2 \in T$ (and $0 \in T$) such that also $t_1 - t_2 \in T$; that is, we can always rewrite $\gamma_{\mathcal{X}}$ by using only a single variable argument, the second one taking the constant value 0. In light of this, we redefine the autocovariance function for covariance-stationary processes as

$$\gamma_{\mathcal{X}}(k) := \gamma_{\mathcal{X}}(k, 0) = \gamma_{\mathcal{X}}(t + k, t)$$

for any $k, t \in T$ with $t + k \in T$; the parameter k is called lag (cf. Brockwell and Davis 1991, pp. 11–12).

The fundamental instance of a covariance-stationary process and primary building block for certain other stochastic processes is white noise, defined as follows.

(D5): We say that $\mathcal{E}_T := \mathcal{X}_T$ is **(discrete-parameter) white noise** with mean 0 and variance $\sigma^2_{\mathcal{X}}$ iff

- $T \subset \mathbb{Z}$,
- $E\{\mathcal{X}_t\} = 0$ for any $t \in T$, and
- $\gamma_{\mathcal{X}}(k) = \begin{cases} \sigma^2_{\mathcal{X}} & \text{if } k = 0, \\ 0 & \text{if } k \neq 0 \end{cases}$.

Now let us consider a *non-recursive filter* C, defined by the filter equation $y_t = \sum_{k=-\infty}^{\infty} c_k u_{t-k}$ for any $t \in \mathbb{Z}$, or in lag operator notation $y_t = C(L)u_t$ with $L^k u_t := u_{t-k}$ and $C(L) = \sum_{k=-\infty}^{\infty} c_k L^k$, where $(u_t \mid t \in \mathbb{Z})$ is any filter input sequence and $(y_t \mid t \in \mathbb{Z})$ any filter output sequence (in either case of real numbers or random variables), and $(c_k \mid k \in \mathbb{Z})$ is any sequence of real-valued filter coefficients. If we view the random variables of a white noise process $\mathcal{E}_T$ as filter input to a

- *causal* (i.e. $c_k = 0$ for any $k < 0$),
- either *finite* or *infinite* (i.e. a finite or an infinite number of filter coefficients is non-zero),
- *absolutely summable* (i.e. $\sum_{k=-\infty}^{\infty} |c_k| < \infty$), and
- *invertible* (i.e. there exists an inverse filter $\overline{C}$ with filter coefficients $(\bar{c}_k \mid k \in \mathbb{N}^0)$ such that $[\overline{C}(L)C(L)]u_t = u_t$ where $\overline{C}(L) = \sum_{k=0}^{\infty} \bar{c}_k L^k$)

version of such a non-recursive filter, then we obtain the moving average process as filter output, as explained in the following definition.

(D6): If $\mathcal{E}_T$ with $T \subset \mathbb{Z}$ is (discrete) white noise with mean 0 and variance $\sigma^2_{\mathcal{E}}$, then we say that $\mathcal{L}_T : (\Omega, \mathcal{A}, P, \{\mathcal{L}_t, t \in T\})$ is a **(discrete-parameter) moving average process** of order q (or MA(q) process) (with $q \in \mathbb{N}$) iff the random variables $\mathcal{L}_t$ satisfy, for any $t \in T$, the equation

$$\mathcal{L}_t = \mathcal{E}_t + \beta_1 \mathcal{E}_{t-1} + \ldots + \beta_q \mathcal{E}_{t-q} = \beta(L)\mathcal{E}_t$$

with $\beta(L) = 1 + \beta_1 L + \ldots + \beta_q L^q$. In the limiting case of $q = \infty$, we call $\mathcal{L}_T$ an MA(∞) process.

Treating $\beta(L)$ as a complex polynomial, then, if $\beta(z) \neq 0$ for any $z \in \mathbb{C}$ with $|z| \leq 1$, then the filter β and, hence the MA(q) process, is invertible (cf. Brockwell and Davis 1991, pp. 86–87). Furthermore, whereas any MA(q) process with $q < \infty$ is covariance-stationary (cf. Priestley 2004, p. 137), the $MA(\infty)$ process is covariance-stationary iff the sequence $(\beta_k \mid k \in \mathbb{N}^0)$ of filter coefficients is absolutely summable