Boris Kargoll*, Alexander Dorndorf, Mohammad Omidalizarandi, Jens-André Paffenholz, and Hamza Alkhatib

Adjustment models for multivariate geodetic time series with vector-autoregressive errors

Abstract: In this contribution, a vector-autoregressive (VAR) process with multivariate t-distributed random deviations is incorporated into the Gauss-Helmert model (GHM), resulting in an innovative adjustment model. This model is versatile since it allows for a wide range of functional models, unknown forms of auto- and cross-correlations, and outlier patterns. Subsequently, a computationally convenient iteratively reweighted least squares method based on an expectation maximization algorithm is derived in order to estimate the parameters of the functional model, the unknown coefficients of the VAR process, the cofactor matrix, and the degree of freedom of the t-distribution. The proposed method is validated in terms of its estimation bias and convergence behavior by means of a Monte Carlo simulation based on a GHM of a circle in two dimensions. The methodology is applied in two different fields of application within engineering geodesy: In the first scenario, the offset and linear drift of a noisy accelerometer are estimated based on a Gauss-Markov model with VAR and multivariate t-distributed errors, as a special case of the proposed GHM. In the second scenario real laser tracker measurements with outliers are adjusted to estimate the parameters of a sphere employing the proposed GHM with VAR and multivariate t-distributed errors. For both scenarios the estimated parameters of the fitted VAR model and multivariate t-distribution are analyzed for evidence of auto- or cross-correlations and deviation from a normal distribution regarding the measurement noise.

1 Introduction

Geodetic observation models of surveyed phenomena often include unknown quantities in terms of parameters to be estimated by adjustment computations. The most general structure of observation model is described by nonlinear condition equations in which multiple observations and unknown parameters may be linked to each other [1]. This general case of adjustment calculus was called Gauss-Helmert model (GHM) by Wolf [2]. There exist numerous special cases of this model that have been studied extensively, e. g., the errors-in-variables (EIV) model adjusted by the method of total least squares [3] or by its various generalizations [4, 5, 6, 7]. Another special case of the GHM is the Gauss-Markov model (GMM), in which the observations are still linked to functional model parameters but occur in separate equations [8, 9, 10]. Such models may also contain variance-covariance information about the measurements; this information defines the so-called stochastic model. Since adjustment calculus has been elaborated on the basis of matrix calculus, variance-covariance information regarding the observables is usually arranged as a matrix, called the "variance-covariance matrix" (VCM).

Since the dimension of the VCM equals the number of observations, the storage and processing of this matrix in the course of the adjustment computations can be challenging, especially because modern sensors such as accelerometers, laser scanners, laser tracker or satellite-based sensors tend to provide huge data sets consisting of millions or even billions of observations even within relatively short periods of time. It is usually not an option to omit the variance-covariance information within the adjustment computations since the standard parameter estimators lose desirable quality characteristics when this

*Corresponding author: Boris Kargoll, Institute of Geoinformation and Surveying, Anhalt University of Applied Sciences, Seminarplatz 2a, 06846 Dessau-Roßlau, Germany, e-mail:
boris.kargoll@hs-anhalt.de, ORCID:
https://orcid.org/0000-0003-3496-6391
Alexander Dorndorf, Mohammad Omidalizarandi, Hamza Alkhatib, Geodetic Institute, Leibniz University Hannover, Nienburger Straße 1, 30167 Hannover, Germany, e-mails: dorndorf@gih.uni-hannover.de, zarandi@gih.uni-hannover.de, alkhatib@gih.uni-hannover.de, ORCID: https://orcid.org/0000-0002-4396-2694 (A. Dorndorf), https://orcid.org/0000-0002-9897-4473 (M. Omidalizarandi), https://orcid.org/0000-0002-4480-1067 (H. Alkhatib)
Jens-André Paffenholz, Institute of Geo-Engineering, TU Clausthal, Erzstraße 18, 38678 Clausthal-Zellerfeld, Germany, e-mail: jens-andre.paffenholz@tu-clausthal.de, ORCID: https://orcid.org/0000-0003-1222-5568

information is neglected. For instance, the least-squares-estimator in the GMM ceases to be the best linear unbiased estimator (BLUE) when the VCM is dropped in the estimation equation [11]. Furthermore, it is often desirable in geodetic data analysis to obtain a realistic description of the uncertainties or variance-covariance information of the estimated parameters, which information is unavailable when an adequate stochastic model of the observables is missing. The "curse of dimensionality" concerning the VCM is aggravated by the increasingly practiced combination of multiple observation series stemming, e. g., from 3D sensors or from sensor networks. Time series of such measurements, e. g., GPS and laser scanner measurements, frequently have been found to be correlated in such a way that the VCM is fully populated, e. g., [12, 13, 17, 16, 14, 15].

To alleviate the computational burden or even computational infeasibility in connection with the storage and processing of a huge VCM, an alternative approach to the stochastic modeling of variance-covariance information can be applied. The goal of that approach is to fully decorrelate the observables by transforming the observation model in such a way that the VCM of the transformed observables becomes a diagonal matrix, e. g., the identity matrix [18]. A diagonal VCM can easily be handled within the adjustment procedure [19], e. g., through vectorization of the occurring matrix products. In geodetic applications the correlations of the observables may frequently be described as colored noise [20], in which case the stochastic model can be expressed as an autoregressive (AR) or autoregressive moving average (ARMA) process and the transformation of the observation equations achieved by means of a computationally efficient digital de-correlation filter [21, 22]. Such processes have also been estimated by means of total least squares based on an EIV model [23]. To model auto- and cross-correlations of multivariate time series, the aforementioned AR and ARMA processes have been extended to vector-autoregressive (VAR) and vector-autoregressive moving average (VARMA) processes [24]. The model order of these recursive processes can be specified to define how far the correlations reach into the past. Thus, VAR(MA) processes can be used to model quite detailed patterns of auto- and cross-correlations. The innovation of the current contribution is to incorporate VAR processes into the GHM as colored noise models.

The white-noise components of (V)AR or (V)ARMA processes are usually assumed to be normally distributed. However, this assumption may be unrealistic or at least questionable in a practical situation, e. g., due to numerous outliers afflicting the measurements. It may then be safer to replace the normal distribution by a larger family of distributions defined by probability density functions having thicker tails than the "ideal" normal distributions. For instance, the usage of generalized t-distributions and of scaled t-distributions has been proposed in connection with ARMA processes [25]. The scaled, multivariate t-distributions have also been applied in connection with VAR processes within rather simple GMM [26, 27]. A key feature of using multivariate t-distributions is that the associated degree of freedom (df) can be estimated alongside the functional parameters, the VAR coefficients and the scale or cofactor matrix; such a method may be called a *self-tuning* estimator [28] since it does not involve a tuning constant for classifying the in- and outliers, in contrast to for instance Huber's M-estimator [29]. The df provides evidence of deviations of the measurements' actual probability distribution from a normal distribution: For large df the multivariate t-distribution is similar to a multivariate normal distribution, whereas the t-distribution has much thicker tails than the latter for a small df, which may indicate the presence of numerous outliers.

The estimation procedure is implemented as an iteratively reweighted least-squares (IRLS) method. The weights are treated as random variables, which are numerically determined by the expectation step within an expectation maximization (EM) algorithm [30]. The various types of model parameters are estimated group-by-group within the maximization step. To deal with the condition equations within the GHM, constrained EM in the spirit of [31] is applied. Due to the down-weighting of observations with extreme errors,[1] this method can be expected to yield a robust estimator for the GHM. Koch introduced the GHM with t-distributed errors and subsequently extended the model to include variance components [32, 33]. This type of GHM is now extended to include a VAR model with multivariate-t-distributed errors. This yields an approach complementary to that of modeling a small number of outliers as additive, unknown offsets within the afflicted observations, as done in [34, 35]. Therefore, the proposed adjustment model is directed at applications in which rather large numbers of outliers can be expected. To model the auto- and cross-correlations, we do not employ the more general VARMA processes since we found the estimation of the moving average part by means of an EM algorithm to diverge.

The paper is structured as follows. In Section 2, we specify the multivariate observation model, the correla-

---

1 We use the term "error" in the sense of "random deviation" throughout the paper.

tion model, and the stochastic model, which jointly define the GHM with VAR and multivariate t-distributed errors. In Section 3, we formulate the optimization principle employed to adjust this model, and we derive the normal equations to be solved within the EM algorithm. We show that this algorithm estimates the functional parameters, the VAR coefficients and the parameters defining the multivariate t-distribution essentially via IRLS. The Monte Carlo simulation described in Section 4 serves the purpose of validating the algorithm, by showing the biases to be expected in the practical situation of data approximation by a circle based on a GHM with VAR and multivariate t-distributed errors. In Section 5 we demonstrate the application of the algorithm for the adjustment of a real, multivariate accelerometer data set, whose offset and drift are modeled as part of a GMM (viewed as a special case of the GHM) with VAR and multivariate t-distributed errors. In a second applied scenario real laser tracker measurements with outliers are adjusted to estimate the parameters of a sphere employing the proposed GHM with VAR and multivariate t-distributed errors. Finally, some limitations and consequently potential extensions of the presented methodology to be explored in the future are outlined in the conclusions and outlook.

2 Adjustment models for multivariate time series

The purpose of this section is to develop adjustment models suitable for the fitting of multivariate geodetic time series. According to the standard categorization in geodetic adjustment theory these models have two main components: the deterministic model and the stochastic model. The following sub-sections demonstrate some possible structures of these models that have been found useful in geodetic time series analysis. These structures allow for certain generalizations of the standard GHM and GHM by allowing their random deviations to be auto- and cross-correlated through (V)AR processes while potential deviations from a normal distribution are modeled stochastically by a multivariate t-distribution.

2.1 Structures of the deterministic observation model

2.1.1 Deterministic part of the Gauss-Helmert model

We suppose that N time series are observed simultaneously without gaps, so that each series consists of n measurements. The measurements can then be doubly indexed in the form $\ell_{k,t}$ with $k = 1, \ldots, N$ and $t = 1, \ldots, n$, where k represents the group index and t a time-related index. We further suppose that the measurements are used to estimate a specified parameter-dependent functional model, which describes spatial objects possibly varying throughout time. In the most general case, each observation is modeled as the sum of an observation-specific location parameter $\mu_{k,t}$ and a corresponding random deviation or "error" $e_{k,t}$, that is,

$$\ell_{k,t} = \mu_{k,t} + e_{k,t} \quad (k = 1, \ldots, N; t = 1, \ldots, n). \quad (1)$$

Each type of quantity in these equations can be vectorized in different ways to obtain more compact notations. The rule will be that all column vectors are symbolized by bold-faced, lower-case Greek letters (in case of unknown parameters) or Roman letters. Firstly, the quantities of one type occurring throughout the time series $k = 1, \ldots, N$ may be collected for every time instance $t = 1, \ldots, n$ via

$$\boldsymbol{\ell}_t = \begin{bmatrix} \ell_{1,t} \\ \vdots \\ \ell_{N,t} \end{bmatrix}, \quad \boldsymbol{\mu}_t = \begin{bmatrix} \mu_{1,t} \\ \vdots \\ \mu_{N,t} \end{bmatrix}, \quad \mathbf{e}_t = \begin{bmatrix} e_{1,t} \\ \vdots \\ e_{N,t} \end{bmatrix}, \quad (2)$$

to define the multivariate observation time series

$$\boldsymbol{\ell}_t = \boldsymbol{\mu}_t + \mathbf{e}_t \quad (t = 1, \ldots, n). \quad (3)$$

Secondly, the observations, location parameters and random deviations can be grouped series-by-series for $k = 1, \ldots, N$ according to

$$\boldsymbol{\ell}_{k:} = \begin{bmatrix} \ell_{k,1} \\ \vdots \\ \ell_{k,n} \end{bmatrix}, \quad \boldsymbol{\mu}_{k:} = \begin{bmatrix} \mu_{k,1} \\ \vdots \\ \mu_{k,n} \end{bmatrix}, \quad \mathbf{e}_{k:} = \begin{bmatrix} e_{k,1} \\ \vdots \\ e_{k,n} \end{bmatrix}, \quad (4)$$

to define the observation series

$$\boldsymbol{\ell}_{k:} = \boldsymbol{\mu}_{k:} + \mathbf{e}_{k:} \quad (k = 1, \ldots, N). \quad (5)$$

Note that the additional colons are used to distinguish the two types of vectors, e. g., the observations $\boldsymbol{\ell}_1$ at time instance $t = 1$ and the observations $\boldsymbol{\ell}_{1:}$ within time series $k = 1$. Finally, it will be convenient to combine the vectors indexed by time to the "complete" vectors

$$\boldsymbol{\ell} = \begin{bmatrix} \boldsymbol{\ell}_1 \\ \vdots \\ \boldsymbol{\ell}_n \end{bmatrix}, \quad \boldsymbol{\mu} = \begin{bmatrix} \boldsymbol{\mu}_1 \\ \vdots \\ \boldsymbol{\mu}_n \end{bmatrix}, \quad \mathbf{e} = \begin{bmatrix} \mathbf{e}_1 \\ \vdots \\ \mathbf{e}_n \end{bmatrix}, \quad (6)$$

so that (3) turns into the single vector equation

$$\boldsymbol{\ell} = \boldsymbol{\mu} + \mathbf{e}. \quad (7)$$

Note that the relationship $\boldsymbol{\mu} = \boldsymbol{\ell} - \mathbf{e}$ obtained from (7) would be written as $\widehat{\boldsymbol{\ell}} = \boldsymbol{\ell} + \mathbf{v}$ in the usual notation of the GHM, where $\widehat{\boldsymbol{\ell}}$ are the (unknown) "adjusted observations". The choice of terminology ("location parameter") and notation ($\boldsymbol{\mu}$) is motivated by our intention to describe the observations in such a way that maximum likelihood (ML) estimation via an EM algorithm can easily be established.

Note also that the vectors $\boldsymbol{\ell}_t$, $\boldsymbol{\mu}_t$ and $\mathbf{e}_t$ can be retrieved from the corresponding complete vectors $\boldsymbol{\ell}$, $\boldsymbol{\mu}$ and $\mathbf{e}$ via

$$\boldsymbol{\ell}_t = \mathbf{J}_t \boldsymbol{\ell}, \quad \boldsymbol{\mu}_t = \mathbf{J}_t \boldsymbol{\mu}, \quad \mathbf{e}_t = \mathbf{J}_t \mathbf{e}, \tag{8}$$

where the $(N \times Nt)$-matrix $\mathbf{J}_t$ consists of rows

$$N \cdot (t-1) + 1, \ldots, N \cdot t$$

of the (Nn)-dimensional identity matrix $\mathbf{I}_{Nn}$. This matrix can be written as the block matrix

$$\mathbf{J}_t = \begin{bmatrix} \mathbf{0}_N & \cdots & \mathbf{0}_N & \mathbf{I}_N & \mathbf{0}_N & \cdots & \mathbf{0}_N \end{bmatrix} \quad (t = 1, \ldots, n) \tag{9}$$

beginning with $t-1$ square zero matrices $\mathbf{0}_N$ of dimension N, followed by the N-dimensional identity matrix $\mathbf{I}_N$, and ending with $n-t$ further zero blocks $\mathbf{0}_N$.

In practice the goal usually is to fit a specific functional model to the given measurements, e. g., a circle. Such a model generally depends on unknown parameters $\boldsymbol{\beta} = [\beta_1 \cdots \beta_m]^T$, e. g., the center coordinates and the radius of a circle. To incorporate this additional model structure it is assumed that the functional parameters $\boldsymbol{\beta}$ and the location parameters $\boldsymbol{\mu}$ fulfill condition equations

$$h_1(\boldsymbol{\beta}, \boldsymbol{\mu}) = 0 \tag{10}$$

$$\vdots \tag{11}$$

$$h_r(\boldsymbol{\beta}, \boldsymbol{\mu}) = 0, \tag{12}$$

which can be written in vectorized form as

$$\mathbf{h}(\boldsymbol{\beta}, \boldsymbol{\mu}) = \mathbf{0}. \tag{13}$$

This equation and (7) form the deterministic part of the GHM, which was defined similarly in [32]. The given somewhat unusual notation, especially the introduction of location parameters, in the setup of the GHM is more aligned with the method of constrained ML estimation than with the usual method of LS. The former estimation method will be more suitable than the latter in connection with the estimation of VAR models and usage of a non-Gaussian pdf.

2.1.2 Deterministic part of the Gauss-Markov model

With certain kinds of adjustment problems each location parameter $\mu_{k,t}$ can be expressed directly as the value of a function $f_{k,t}$ depending on the functional parameters $\boldsymbol{\beta}$, through the relationship

$$\mu_{k,t} = f_{k,t}(\boldsymbol{\beta}). \tag{14}$$

Then, the observation equations (1) take the form

$$\ell_{k,t} = f_{k,t}(\boldsymbol{\beta}) + e_{k,t} \quad (k = 1, \ldots, N; t = 1, \ldots, n). \tag{15}$$

Vectorizing the (unknown) function values $f_{k,t}(\boldsymbol{\beta})$ similarly as in (2) and (4), these observation equations can be arranged by time as

$$\boldsymbol{\ell}_t = \mathbf{f}_t(\boldsymbol{\beta}) + \mathbf{e}_t \quad (t = 1, \ldots, n), \tag{16}$$

or by time series as

$$\boldsymbol{\ell}_{k:} = \mathbf{f}_{k:}(\boldsymbol{\beta}) + \mathbf{e}_{k:} \quad (k = 1, \ldots, N), \tag{17}$$

or collectively, by vectorizing $\mathbf{f}_t(\boldsymbol{\beta})$ for $t = 1, \ldots, n$ as in (6), as

$$\boldsymbol{\ell} = \mathbf{f}(\boldsymbol{\beta}) + \mathbf{e}. \tag{18}$$

This structure of observation equation defines the GMM. Using (14), the condition equations (13) are then simply taken as $\mathbf{h}(\boldsymbol{\beta}, \boldsymbol{\mu}) = \boldsymbol{\mu} - \mathbf{f}(\boldsymbol{\beta}) = \mathbf{0}$. In adjustment computations it is more common to state the observation equations in the equivalent form

$$\boldsymbol{\ell} + \mathbf{v} = \mathbf{f}(\boldsymbol{\beta}), \tag{19}$$

where the "corrections" $\mathbf{v}$ are related to the random deviations $\mathbf{e}$ via sign reversal. In view of the additional stochastic model imposed on the observation model later on, the representation (18) is preferred in this paper. In case the function $\mathbf{f}$ can be written as a matrix-vector product $\mathbf{X}\boldsymbol{\beta}$, we speak of a "linear model", where $\mathbf{X}$ is called the "design matrix"; we then write

$$\boldsymbol{\ell} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e} \tag{20}$$

instead of (18). Linear models have been studied thoroughly in mathematical statistics and geodesy [36, 18]. The parameters $\boldsymbol{\beta}$ can be estimated without any additional assumptions, e. g., by applying "ordinary LS". However, if the observables are correlated or heteroscedastic, the LS estimator will not be BLUE if the correlations and unequal variances present are not taken into account by incorporating a corresponding stochastic model into the observation model and the estimator.

2.2 Stochastic modeling

2.2.1 Modeling of correlations using vector-autoregressive processes

The simplest case of a stochastic model occurs for uncorrelated observables with identical variances, so that the VCM $\boldsymbol{\Sigma}$ is the rescaled identity matrix $\sigma^2\mathbf{I}$, where the variance factor σ^2 may be a known number or an unknown parameter. Geodesists have frequently pointed out that observables are usually correlated (e. g., [37]) and that it is of great practical importance to take correlations into account in geodetic data analysis (e. g., [38]). In case of correlated measurements the VCM $\boldsymbol{\Sigma}$ typically is either fully populated or has a band-limited structure (see [21]). Its decomposition therefore yields $\boldsymbol{\Sigma} = \sigma^2\mathbf{Q}$, where the so-called "cofactor matrix" $\mathbf{Q}$ has the same population characteristics as $\boldsymbol{\Sigma}$. When a univariate time series is surveyed the correlations between the various random variables are called "auto-correlations". Oftentimes these can be modeled by means of an AR process

$$\mathcal{E}_t = \alpha_1\mathcal{E}_{t-1} + \cdots + \alpha_p\mathcal{E}_{t-p} + \mathcal{U}_t \tag{21}$$

for some $p \in \mathbb{N}$ (the "model order" of the AR process) and some $\alpha_1,\ldots,\alpha_p$ (the "coefficients" of the AR process). $(\mathcal{U}_t)_{t\in\mathbb{Z}}$ is a white noise series that recursively generates an auto-correlated time series $(\mathcal{E}_t)_{t\in\mathbb{Z}}$ (see, e. g., [42] for details). In case of multivariate observables, additional correlations between the random variables of different time series, so-called "cross-correlations", may occur. To take both auto- and cross-correlations into account, an error process $(\mathcal{E}_t)_{t\in\mathbb{Z}}$ can be defined to be a VAR(p) process by the equation (cf. Sect. 8.4 in [24])

$$\mathcal{E}_t = \mathbf{A}_1\mathcal{E}_{t-1} + \cdots + \mathbf{A}_p\mathcal{E}_{t-p} + \mathcal{U}_t, \tag{22}$$

where $(\mathcal{U}_t)_{t\in\mathbb{Z}}$ is a multivariate form of white noise. The coefficients of this VAR process are now arranged within the $(N \times N)$-matrices

$$\mathbf{A}_j = \begin{bmatrix} \alpha_{j,1,1} & \cdots & \alpha_{j,1,N} \\ \vdots & \ddots & \vdots \\ \alpha_{j,N,1} & \cdots & \alpha_{j,N,N} \end{bmatrix} \quad (j = 1,\ldots,p). \tag{23}$$

2.2.2 Modeling of observational weights using multivariate t-distributions

Another standard decomposition of a the VCM $\boldsymbol{\Sigma}$ is given in terms of the "weight matrix" $\mathbf{P}$ by

$$\boldsymbol{\Sigma} = \sigma^2\mathbf{P}^{-1}. \tag{24}$$

The main diagonal of the VCM $\boldsymbol{\Sigma}$ consists of the variances of the individual observables, whereas the main diagonal elements of $\mathbf{P}$ are commonly referred to as "weights". The meaning of the weights is twofold in practice: On the one hand, weights may be interpreted as "repetition numbers", i. e., as the number of times that the corresponding observations have occurred. On the other hand, weights can be assigned to observations in order to diminish or increase their relative importance and influence in statistical inference. Abnormal or extreme observations ("outliers"), located in the tails of the probability density function (pdf) defining their probability distribution, are frequently encountered in geodetic data analysis. One standard approach to handling such observations consists in their "down-weighting" by giving them lesser weights than non-outliers ("inliers"). Alternatively, it is possible to carry out statistical tests in order to identify and delete outlying observations (e. g., [39, 40, 41]). However, the deletion of observations creates data gaps, which prohibit the usage of recursive time series models, such as the VAR processes considered in Sect. 2.2.1. Therefore, we restrict ourselves to the approach based on the down-weighting of outlying observations, which avoids their elimination and consequential data gaps. Since the data sets analyzed nowadays tend to be huge and oftentimes have stochastic properties unknown to the practitioner, it is necessary to have a data-adaptive and computationally efficient procedure for computing weights at hand. Mathematically, such a procedure is based on a function (the "weight function"), which corresponds to a certain pdf of the random deviations.

For the purpose of multivariate modeling including potential cross-correlations, the well-known Student's t-distribution can be generalized to an N-variate t-distribution $t(\boldsymbol{\mu},\boldsymbol{\Psi},\nu)$ with location parameter vector $\boldsymbol{\mu}$, scale matrix $\boldsymbol{\Psi}$ and df ν, defined by the pdf (cf. [43])

$$p(\mathbf{x}) = \frac{\Gamma\left(\frac{\nu+N}{2}\right)}{(\sqrt{\nu\pi})^N \sqrt{\det\boldsymbol{\Psi}}\,\Gamma\left(\frac{\nu}{2}\right)}$$

$$\times \left(1 + \frac{(\mathbf{x} - \boldsymbol{\mu})^T\boldsymbol{\Psi}^{-1}(\mathbf{x} - \boldsymbol{\mu})}{\nu}\right)^{-\frac{\nu+N}{2}}. \tag{25}$$

Since the scale matrix $\boldsymbol{\Psi}$ is related to the VCM via $\boldsymbol{\Sigma} = \frac{\nu}{\nu-2}\cdot\boldsymbol{\Psi}$ for df $\nu > 2$, it may be viewed as a "cofactor matrix". Then, a multivariate t-distributed form of white noise of a VAR process can be defined by

$$\mathcal{U}_t \overset{\text{ind.}}{\sim} t(\mathbf{0},\boldsymbol{\Psi},\nu) \tag{26}$$

(see also [44]). The joint pdf of n such white noise vectors $\mathcal{U}_1, \ldots, \mathcal{U}_n$ is therefore given by

$$p(\mathbf{u}_1, \ldots, \mathbf{u}_n) = \prod_{t=1}^{n} p(\mathbf{u}_t) = \left(\frac{\Gamma\left(\frac{\nu+N}{2}\right)}{(\sqrt{\nu\pi})^N \sqrt{\det \mathbf{\Psi}} \, \Gamma\left(\frac{\nu}{2}\right)} \right)^n$$

$$\times \prod_{t=1}^{n} \left(1 + \frac{\mathbf{u}_t^T \mathbf{\Psi}^{-1} \mathbf{u}_t}{\nu} \right)^{-\frac{\nu+N}{2}}. \tag{27}$$

This pdf is more intricate and therefore more difficult to handle in maximum likelihood (ML) estimation than the pdf of a multivariate normal distribution. However, it is well known that a multivariate t-distributed random vector $\mathcal{X} \sim t(\boldsymbol{\mu}, \mathbf{\Psi}, \nu)$ has the property of being a conditionally normally distributed random vector $\mathcal{X} \sim N(\boldsymbol{\mu}, \mathbf{\Psi}/w)$ where w is a given realization of a gamma-distributed random variable $\mathcal{W} \sim G(\nu/2, \nu/2)$ $\mathcal{W}$ (cf. [43]). Due to this property, the distributional model (26) can be replaced by the assumptions

$$\mathcal{U}_t \overset{\text{ind.}}{\sim} N(\mathbf{0}, \mathbf{\Psi}/w_t) \tag{28}$$

and

$$\mathcal{W}_t \overset{\text{ind.}}{\sim} G\left(\frac{\nu}{2}, \frac{\nu}{2}\right). \tag{29}$$

It can be seen in (28) that the white noise vector $\mathcal{U}_t$ has the VCM $\mathbf{\Sigma}_t = \mathbf{\Psi}/w_t$, which is a rescaled version of the cofactor matrix $\mathbf{\Psi}$. For a small value w_t, the cofactor matrix $\mathbf{\Psi}$ may be viewed as being "inflated"; in the univariate case $N = 1$, this rescaling approach therefore has been called "variance inflation model" (e. g., [41]). The idea of this type of outlier model is to assign a relatively large variance to an error which is located in the tails of the pdf, in comparison to an error near the center of the pdf.

Under the distributional assumptions (28)–(29) the joint pdf of white noise vectors $\mathcal{U}_1, \ldots, \mathcal{U}_n$ and the random variables $\mathcal{W}_1, \ldots, \mathcal{W}_n$ assigned to n measurements can be shown to result in

$$p(\mathbf{u}_1, \ldots, \mathbf{u}_n, \mathbf{w}) = \left(\frac{\left(\frac{\nu}{2}\right)^{\frac{\nu}{2}}}{\Gamma\left(\frac{\nu}{2}\right)} \right)^n \prod_{t=1}^{n} w_t^{\frac{\nu}{2}-1} \exp\left(-\frac{\nu}{2}w_t\right)$$

$$\times \frac{1}{\sqrt{(2\pi)^N \det(\mathbf{\Psi}/w_t)}}$$

$$\times \exp\left(-\frac{w_t}{2} \mathbf{u}_t^T \mathbf{\Psi}^{-1} \mathbf{u}_t\right), \tag{30}$$

and (27) defines the corresponding marginal distribution of $\mathcal{U}_1, \ldots, \mathcal{U}_n$. Although the joint pdf (30) appears to be even more complicated than the marginal pdf defining the multivariate t-distribution, considerable simplifications can be achieved by taking a conditional expectation of the log-likelihood function defined by this pdf. This

step, which exploits the fact that the pdf (30) now involves natural exponential functions exp(.) in contrast to (27), corresponds to the "E-step" of an EM algorithm shown in Sect. 3. We first proceed by defining particular instances of the aforementioned log-likelihood functions which include the deterministic model of the GHM or GMM, as well as the VAR process.

2.3 Representations of the adjustment models as likelihood functions

A likelihood function L can be defined by a pdf by fixing an element of its domain (the observation space) and by letting its parameters be "variables". The pdf (30) directly involves as parameters the entries of the cofactor matrix $\mathbf{\Psi}$ and the df ν. When certain realizations $\mathbf{u}_1, \ldots, \mathbf{u}_n$ of the white-noise vector are given, they lead to realizations $\mathbf{e}_1, \ldots, \mathbf{e}_n$ of the VAR process (22). These computational relationships can inversely be written as

$$\mathbf{u}_t = \mathbf{e}_t - \mathbf{A}_1 \mathbf{e}_{t-1} - \ldots - \mathbf{A}_p \mathbf{e}_{t-p}, \tag{31}$$

where $\mathbf{e}_0, \ldots, \mathbf{e}_{1-p}$ are conventionally taken to be zero vectors. Moreover, the realizations $\mathbf{e}_1, \ldots, \mathbf{e}_n$ of a VAR process occur within the condition equations (3) of the GHM, which can inversely be written as

$$\mathbf{e}_t = \boldsymbol{\ell}_t - \boldsymbol{\mu}_t. \tag{32}$$

Consequently, (31) can be recast as

$$\mathbf{u}_t = \mathbf{I}(\boldsymbol{\ell}_t - \boldsymbol{\mu}_t) - \mathbf{A}_1(\boldsymbol{\ell}_{t-1} - \boldsymbol{\mu}_{t-1}) - \ldots - \mathbf{A}_p(\boldsymbol{\ell}_{t-p} - \boldsymbol{\mu}_{t-p})$$

$$= \mathbf{I}(\boldsymbol{\ell}_t - \boldsymbol{\mu}_t) - \mathbf{A}_1 L^1(\boldsymbol{\ell}_t - \boldsymbol{\mu}_t) - \ldots - \mathbf{A}_p L^p(\boldsymbol{\ell}_t - \boldsymbol{\mu}_t)$$

$$= (\mathbf{I} - \mathbf{A}_1 L^1 - \ldots - \mathbf{A}_p L^p)(\boldsymbol{\ell}_t - \boldsymbol{\mu}_t)$$

$$= \mathbf{A}(L)(\boldsymbol{\ell}_t - \boldsymbol{\mu}_t). \tag{33}$$

Here, for brevity of expressions, we defined the so-called lag polynomial $\mathbf{A}(L)$ (see [24], p. 243) based on the lag operator L, whose powers are used to shift the time index t according to the rule $L^j \mathbf{Z}_t := \mathbf{Z}_{t-j}$ with $j \in \{1, 2, \ldots\}$ (see Chap. 2 in [45]).

Thus, the pdf (30) indirectly depends also on the VAR coefficients in the matrices $\mathbf{A}_1, \ldots, \mathbf{A}_p$ and on the location parameters $\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_n$ of the deterministic model in case of a GHM. In the case of the more structured GMM, $\mathbf{f}_t(\boldsymbol{\beta})$ is substituted for $\boldsymbol{\mu}_t$ in (32) and (33) in view of (16), so that the pdf (30) then depends on functional model parameters $\boldsymbol{\beta}$ rather than on $\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_n$.

To fix the dimensions of the parameter space, the kth rows of the VAR coefficient matrices are combined to column vectors of length $p \cdot N$

$$\boldsymbol{\alpha}_k = [(\mathbf{A}_1)_k \; \cdots \; (\mathbf{A}_p)_k]^T \quad (k = 1, \ldots, N), \tag{34}$$

which in turn are combined to the single column vector

$$\boldsymbol{\alpha} = [\boldsymbol{\alpha}_1^T \; \cdots \; \boldsymbol{\alpha}_N^T]^T \qquad (35)$$

of length $p \cdot N^2$. Similarly, the row-wise vectorization of the inverted cofactor matrix $\boldsymbol{\Psi}^{-1}$ yields a column vector of length N^2, which is denoted by $\bar{\boldsymbol{\psi}}$. Then, the parameters that the pdf for the GHM and the GMM with VAR and multivariate t-distributed errors directly and indirectly depend upon can be grouped according to

$$\boldsymbol{\theta}_{\mathrm{GHM}} = [\boldsymbol{\mu}^T \; \boldsymbol{\alpha}^T \; \bar{\boldsymbol{\psi}}^T \; \nu]^T \qquad (36)$$

and

$$\boldsymbol{\theta}_{\mathrm{GMM}} = [\boldsymbol{\beta}^T \; \boldsymbol{\alpha}^T \; \bar{\boldsymbol{\psi}}^T \; \nu]^T. \qquad (37)$$

Taking the natural logarithm of the pdf (30) with $\mathbf{u}_t$ given by (33), the log-likelihood functions of this GHM can be obtained, see (81) in the Appendix. The log-likelihood function for the GHM describes only the observation model, so that the constraints $\mathbf{h}(\boldsymbol{\beta}, \boldsymbol{\mu}) = \mathbf{0}$ in (13) need to be provided for in addition when the model parameters $\boldsymbol{\theta}_{\mathrm{GHM}}$ are estimated. Thus, the complete parameter vector consists of the groups

$$\boldsymbol{\Phi} = [\boldsymbol{\beta}^T \; \boldsymbol{\theta}_{\mathrm{GHM}}^T] = [\boldsymbol{\beta}^T \; \boldsymbol{\mu}^T \; \boldsymbol{\alpha}^T \; \bar{\boldsymbol{\psi}}^T \; \nu]^T. \qquad (38)$$

A similar log-likelihood involving multiple (univariate) AR processes instead of a single (multivariate) VAR process was established in [46]. To obtain the log-likelihood function $\log L(\boldsymbol{\theta}_{\mathrm{GMM}}; \boldsymbol{\ell})$ for the GMM, which was employed in [26], the location parameters $\boldsymbol{\mu}_t$ are simply replaced by $\mathbf{f}_t(\boldsymbol{\beta})$, in which the aforementioned constraints are omitted since the location parameters $\boldsymbol{\mu}$ are not modeled explicitly.

3 Estimation

Maximum likelihood estimation of the model parameters $\boldsymbol{\theta}_{\mathrm{GHM}}$ based on the log-likelihood function (81) requires values for the unobservable data $\mathbf{w}$. Since a stochastic model has been specified for these quantities, such values can be determined as part of an EM algorithm [30]. A similar approach has previously been carried out for numerous models that can be identified as special cases of the current GHM. In all previous instances the EM algorithm was found to be suitable in view of its stable convergence and its computationally simple form as IRLS (see, e. g., [47, 48]), where the entries of $\mathbf{w}$ play the role of the weights. The iterations are indexed by s in the following.

In the expectation step ("E-step") of the EM or IRLS algorithm the weights are computed as conditional expectations of the random variables $\mathcal{W}$ based on their stochastic model, using the current solution $\widehat{\boldsymbol{\theta}}_{\mathrm{GHM}}^{(s)}$. In the subsequent maximization step ("M-step") a new solution $\widehat{\boldsymbol{\theta}}_{\mathrm{GHM}}^{(s+1)}$ is computed by maximizing the conditional expectation of the log-likelihood function ("Q-function") using the current weights.

The Q-function for the GHM with VAR and multivariate t-distributed errors is given by (82) in the Appendix. Evaluation of the conditional expectation of the random weight variables $\mathcal{W}$ yields the computational formula

$$\widetilde{w}_t^{(s)} = \frac{\widehat{\nu}^{(s)} + N}{\widehat{\nu}^{(s)} + \mathbf{u}_t^T (\widehat{\boldsymbol{\Psi}}^{(s)})^{-1} \mathbf{u}_t} \qquad (39)$$

The tilde on the conditional expectation $\widetilde{w}_t^{(s)}$ is used to distinguish this "theoretical" type of quantity from estimates. To maximize the resulting Q-function (85) under the constraints the Lagrangian is defined as

$$F(\boldsymbol{\phi}, \boldsymbol{\lambda} | \widehat{\boldsymbol{\theta}}_{\mathrm{GHM}}^{(s)}) = Q(\boldsymbol{\theta}_{\mathrm{GHM}} | \widehat{\boldsymbol{\theta}}_{\mathrm{GHM}}^{(s)}) + \boldsymbol{\lambda}^T \mathbf{h}(\boldsymbol{\beta}, \boldsymbol{\mu}), \qquad (40)$$

which involves as additional parameters an $(r \times 1)$-vector $\boldsymbol{\lambda}$ of unknown Lagrange multipliers. This approach leads to a "constrained EM algorithm" [31]. In the frequently encountered case that $\mathbf{h}(\boldsymbol{\beta}, \boldsymbol{\mu})$ cannot be expressed "linearly" in the form of

$$\mathbf{X}\boldsymbol{\beta} + \mathbf{B}(\boldsymbol{\mu} - \boldsymbol{\ell}) + \mathbf{m} = \mathbf{0} \qquad (41)$$

for some numerically fixed matrix $\mathbf{X}$, matrix $\mathbf{B}$ and vector $\mathbf{m}$, linearization

$$\mathbf{h}(\boldsymbol{\beta}, \boldsymbol{\mu}) \approx \mathbf{X}\Delta\boldsymbol{\beta} + \mathbf{B}(\boldsymbol{\mu} - \boldsymbol{\ell}) + \overline{\mathbf{m}}. \qquad (42)$$

is applied as shown for instance in [32]. Here, we define the $(m \times 1)$-vector of "incremental parameters"

$$\Delta\boldsymbol{\beta} = \boldsymbol{\beta} - \boldsymbol{\beta}^{(s)}, \qquad (43)$$

the $(r \times 1)$-vector of "misclosures"

$$\mathbf{m}^{(s)} = \mathbf{h}(\boldsymbol{\beta}^{(s)}, \boldsymbol{\mu}^{(s)}), \qquad (44)$$

the $(r \times 1)$-vector of "pseudo-misclosures"

$$\overline{\mathbf{m}}^{(s)} = \mathbf{m} + \mathbf{B}(\boldsymbol{\ell} - \boldsymbol{\mu}^{(s)}), \qquad (45)$$

the $(r \times m)$-Jacobi matrix

$$\mathbf{X}^{(s)} = \frac{\partial \mathbf{h}(\boldsymbol{\beta}^{(s)}, \boldsymbol{\mu}^{(s)})}{\partial \boldsymbol{\beta}}, \qquad (46)$$

and the $(r \times Nn)$-Jacobi matrix

$$\mathbf{B}^{(s)} = \frac{\partial \mathbf{h}(\boldsymbol{\beta}^{(s)}, \boldsymbol{\mu}^{(s)})}{\partial \boldsymbol{\mu}}. \tag{47}$$

Whereas $\Delta\boldsymbol{\beta}$ is unknown, the quantities (44)–(47) are computable with the given parameter estimates of iteration step s. For notational brevity, the dependency of the latter quantities on the iteration step will not be indicated in the following; thus, we write $\mathbf{m}$, $\overline{\mathbf{m}}$, $\mathbf{X}$ and $\mathbf{B}$ instead of $\mathbf{m}^{(s)}$, $\overline{\mathbf{m}}^{(s)}$, $\mathbf{X}^{(s)}$ and $\mathbf{B}^{(s)}$. Forming now the partial derivatives of the linearized Lagrangian (86) given in the Appendix with respect to the different groups of parameters and setting these equal to zero, one obtains

1. the normal equations for the incremental functional parameters $\Delta\boldsymbol{\beta}$:

$$\mathbf{X}^T \boldsymbol{\lambda} = \mathbf{0}_{[m \times 1]}; \tag{48}$$

2. the normal equations for the location parameters $\boldsymbol{\mu}$:

$$\sum_{t=1}^{n} \bar{\mathbf{J}}_t^T \widetilde{w}_t^{(s)} \boldsymbol{\Psi}^{-1} [\bar{\boldsymbol{\ell}}_t - \bar{\mathbf{J}}_t \boldsymbol{\mu}] + \mathbf{B}^T \boldsymbol{\lambda} = \mathbf{0}_{[Nn \times 1]} \tag{49}$$

denoting

$$\bar{\mathbf{J}}_t = \mathbf{A}(L) \mathbf{J}_t, \tag{50}$$

$$\bar{\boldsymbol{\ell}}_t = \mathbf{A}(L) \boldsymbol{\ell}_t; \tag{51}$$

3. the normal equations for the VAR coefficient matrices $\mathbf{A}_1, \ldots, \mathbf{A}_p$:

$$\mathbf{E}\widetilde{\mathbf{W}}^{(s)} \begin{bmatrix} \mathbf{e}_1^T \\ \vdots \\ \mathbf{e}_n^T \end{bmatrix} - \mathbf{E}\widetilde{\mathbf{W}}^{(s)} \mathbf{E}^T \begin{bmatrix} \mathbf{A}_1^T \\ \vdots \\ \mathbf{A}_p^T \end{bmatrix} = \begin{bmatrix} \mathbf{0}_N \\ \vdots \\ \mathbf{0}_N \end{bmatrix}, \tag{52}$$

denoting

$$\mathbf{E} = \begin{bmatrix} \mathbf{e}_0 & \cdots & \mathbf{e}_{n-1} \\ \vdots & & \vdots \\ \mathbf{e}_{1-p} & \cdots & \mathbf{e}_{n-p} \end{bmatrix} \tag{53}$$

with $\mathbf{e}_i = \mathbf{0}_{3 \times 1}$ for all $i \leq 0$ and letting $\widetilde{\mathbf{W}}^{(s)}$ be the n-dimensional diagonal matrix with diagonal entries $\widetilde{w}_1^{(s)}, \ldots, \widetilde{w}_n^{(s)}$;

4. the normal equations for the inverse cofactor matrix $\boldsymbol{\Psi}^{-1}$:

$$\boldsymbol{\Psi} - \frac{1}{n} \sum_{t=1}^{n} \widetilde{w}_t^{(s)} [\mathbf{A}(L)(\boldsymbol{\ell}_t - \boldsymbol{\mu}_t)][\mathbf{A}(L)(\boldsymbol{\ell}_t - \boldsymbol{\mu}_t)]^T = \mathbf{0}_N; \tag{54}$$

5. the normal equation for the df ν (cf. Sect. 5.8.2 [47]):

$$-\mathrm{dg}\left(\frac{\nu}{2}\right) + \log\left(\frac{\nu}{2}\right) + 1 + \frac{1}{n} \sum_{t=1}^{n} \left(\log \widetilde{w}_t^{(s)} - \widetilde{w}_t^{(s)}\right)$$
$$+ \mathrm{dg}\left(\frac{\nu^{(s)} + N}{2}\right) - \log\left(\frac{\nu^{(k)} + N}{2}\right) = 0, \tag{55}$$

involving the digamma function dg;

6. the normal equations for the Lagrange multipliers $\boldsymbol{\lambda}$:

$$\mathbf{X}\Delta\boldsymbol{\beta} + \mathbf{B}(\boldsymbol{\mu} - \boldsymbol{\ell}) + \overline{\mathbf{m}} = \mathbf{0}_{[r \times 1]}. \tag{56}$$

Since the location parameters are obtained from (49) as

$$\boldsymbol{\mu} = \left(\sum_{t=1}^{n} \bar{\mathbf{J}}_t^T \widetilde{w}_t^{(s)} \boldsymbol{\Psi}^{-1} \bar{\mathbf{J}}_t\right)^{-1}$$
$$\times \left(\sum_{t=1}^{n} \bar{\mathbf{J}}_t^T \widetilde{w}_t^{(s)} \boldsymbol{\Psi}^{-1} \bar{\boldsymbol{\ell}}_t + \mathbf{B}^T \boldsymbol{\lambda}\right), \tag{57}$$

(supposing that the inverse exists), these parameters can be eliminated from (56) via substitution, which results in

$$\mathbf{B}\left(\sum_{t=1}^{n} \bar{\mathbf{J}}_t^T \widetilde{w}_t^{(s)} \boldsymbol{\Psi}^{-1} \bar{\mathbf{J}}_t\right)^{-1} \mathbf{B}^T \boldsymbol{\lambda} + \mathbf{X}\Delta\boldsymbol{\beta}$$
$$= \mathbf{B}\boldsymbol{\ell} - \overline{\mathbf{m}} - \mathbf{B}\left(\sum_{t=1}^{n} \bar{\mathbf{J}}_t^T \widetilde{w}_t^{(s)} \boldsymbol{\Psi}^{-1} \bar{\mathbf{J}}_t\right)^{-1}$$
$$\times \sum_{t=1}^{n} \bar{\mathbf{J}}_t^T \widetilde{w}_t^{(s)} \boldsymbol{\Psi}^{-1} \bar{\boldsymbol{\ell}}_t. \tag{58}$$

Together with (48), this gives the normal equation system

$$\begin{bmatrix} \mathbf{B}\left(\sum\limits_{t=1}^{n} \bar{\mathbf{J}}_t^T \widetilde{w}_t^{(s)} \boldsymbol{\Psi}^{-1} \bar{\mathbf{J}}_t\right)^{-1} \mathbf{B}^T & \mathbf{X} \\ \mathbf{X}^T & \mathbf{0}_m \end{bmatrix} \begin{bmatrix} \boldsymbol{\lambda} \\ \Delta\boldsymbol{\beta} \end{bmatrix}$$
$$= \begin{bmatrix} \mathbf{B}\boldsymbol{\ell} - \overline{\mathbf{m}} - \mathbf{B}\left(\sum\limits_{t=1}^{n} \bar{\mathbf{J}}_t^T \widetilde{w}_t^{(s)} \boldsymbol{\Psi}^{-1} \bar{\mathbf{J}}_t\right)^{-1} \sum\limits_{t=1}^{n} \bar{\mathbf{J}}_t^T \widetilde{w}_t^{(s)} \boldsymbol{\Psi}^{-1} \bar{\boldsymbol{\ell}}_t \\ \mathbf{0}_{[m \times 1]} \end{bmatrix}. \tag{59}$$

The parameter groups $(\boldsymbol{\lambda}, \Delta\boldsymbol{\beta})$, $(\mathbf{A}_1, \ldots, \mathbf{A}_p)$, $\boldsymbol{\Psi}^{-1}$ and ν are estimated in that order by successively solving (59), (52), (54) and (55). The parameter group $\boldsymbol{\mu}$ can be determined through (57) after solving for $(\boldsymbol{\lambda}, \Delta\boldsymbol{\beta})$. While equation systems (59), (52) and (54) are linear, the solution of (55) requires a zero search. This step-wise solution approach defines an ECM algorithm, which extends previously established ECM algorithms in simpler models.

Having completed s iterations, the first ECM-step of the new iteration $s + 1$ yields a new solution

$$
\begin{bmatrix} \boldsymbol{\lambda}^{(s+1)} \\ \Delta\boldsymbol{\beta}^{(s+1)} \end{bmatrix} = \begin{bmatrix} \mathbf{B}\left(\sum\limits_{t=1}^{n} \bar{\mathbf{J}}_t^T \widetilde{w}_t^{(s)}\boldsymbol{\Psi}^{-1}\bar{\mathbf{J}}_t\right)^{-1}\mathbf{B}^T & \mathbf{X} \\ \mathbf{X}^T & \mathbf{0}_m \end{bmatrix}^{-1}
$$
$$
\times \begin{bmatrix} \mathbf{B}\boldsymbol{\ell} - \overline{\mathbf{m}} - \mathbf{B}\left(\sum\limits_{t=1}^{n}\bar{\mathbf{J}}_t^T\widetilde{w}_t^{(s)}\boldsymbol{\Psi}^{-1}\bar{\mathbf{J}}_t\right)^{-1}\sum\limits_{t=1}^{n}\bar{\mathbf{J}}_t^T\widetilde{w}_t^{(s)}\boldsymbol{\Psi}^{-1}\bar{\boldsymbol{\ell}}_t \\ \mathbf{0}_{[m\times 1]} \end{bmatrix}
$$
(60)

from (59), where the unknown matrices $\boldsymbol{\Psi}$ and

$$
\bar{\mathbf{J}}_t = \mathbf{A}(L)\mathbf{J}_t = \mathbf{J}_t - \mathbf{A}_1\mathbf{J}_{t-1} - \cdots - \mathbf{A}_p\mathbf{J}_{t-p}
$$

are fixed by substituting (i. e., "conditioning on") the previous solutions $\boldsymbol{\Psi}^{(s)}$ and $\mathbf{A}_1^{(s)}, \ldots, \mathbf{A}_p^{(s)}$. Substituting then in addition $\boldsymbol{\lambda}^{(s+1)}$ for the variables $\boldsymbol{\lambda}$ in (57) gives the new solution

$$
\boldsymbol{\mu}^{(s+1)} = \left(\sum_{t=1}^{n}(\bar{\mathbf{J}}_t^{(s)})^T\widetilde{w}_t^{(s)}(\boldsymbol{\Psi}^{(s)})^{-1}\bar{\mathbf{J}}_t^{(s)}\right)^{-1}
$$
$$
\times \left(\sum_{t=1}^{n}(\bar{\mathbf{J}}_t^{(s)})^T\widetilde{w}_t^{(s)}(\boldsymbol{\Psi}^{(s)})^{-1}\bar{\boldsymbol{\ell}}_t + \mathbf{B}^T\boldsymbol{\lambda}^{(s+1)}\right).
$$
(61)

Moreover, the new solution

$$
\boldsymbol{\beta}^{(s+1)} = \boldsymbol{\beta}^{(s)} + \Delta\boldsymbol{\beta}^{(s+1)}
$$
(62)

via substitution into the rearranged equation (43).

The solution $(\boldsymbol{\beta}^{(s+1)}, \boldsymbol{\mu}^{(s+1)})$ might not fulfill the constraints $\mathbf{h}(\boldsymbol{\beta}, \boldsymbol{\mu}) = \mathbf{0}$ exactly due to linearization errors. Therefore, we apply the principle of constrained EM in the spirit of [31] and iterate the estimation of these parameters until either a maximum number of iteration steps is reached or the maximum absolute misclosure falls below a specified threshold (say, 10^{-14}). To do this, the quantities $\mathbf{m}$, $\overline{\mathbf{m}}$, $\mathbf{X}$ and $\mathbf{B}$ in (44)–(47) are recomputed using $\boldsymbol{\beta}^{[0]} := \boldsymbol{\beta}^{(s+1)}$ and $\boldsymbol{\mu}^{[0]} := \boldsymbol{\mu}^{(s+1)}$, and new solutions $\boldsymbol{\lambda}^{[1]}$, $\boldsymbol{\beta}^{[1]}$, $\boldsymbol{\mu}^{[1]}$ are computed via evaluation of (60)–(62) by substituting again the currently available estimates $\boldsymbol{\Psi}^{(s)}$ and $\mathbf{A}_1^{(s)}, \ldots, \mathbf{A}_p^{(s)}$. When one of the stop criteria is fulfilled, say after t additional iterations, one redefines $\boldsymbol{\lambda}^{(s+1)} := \boldsymbol{\lambda}^{[t]}$, $\boldsymbol{\beta}^{(s+1)} := \boldsymbol{\beta}^{[t]}$ and $\boldsymbol{\mu}^{(s+1)} := \boldsymbol{\mu}^{[t]}$.

Next, the new solution

$$
\begin{bmatrix} \mathbf{A}_1^{(s+1)} & \cdots & \mathbf{A}_p^{(s+1)} \end{bmatrix} = \begin{bmatrix} \mathbf{e}_1^{(s+1)} & \cdots & \mathbf{e}_n^{(s+1)} \end{bmatrix} \widetilde{\mathbf{W}}^{(s)}
$$
$$
\times (\mathbf{E}^{(s+1)})^T \left(\mathbf{E}^{(s+1)}\widetilde{\mathbf{W}}^{(s)}(\mathbf{E}^{(s+1)})^T\right)^{-1}
$$
(63)

of the VAR coefficient matrices is computed from (52), based on the new correlated residuals

$$
\mathbf{e}^{(s+1)} = \boldsymbol{\ell} - \boldsymbol{\mu}^{(s+1)}
$$
(64)

computed from the model (7) and the corresponding residual matrix

$$
\mathbf{E}^{(s+1)} = \begin{bmatrix} \mathbf{e}_0^{(s+1)} & \cdots & \mathbf{e}_{n-1}^{(s+1)} \\ \vdots & & \vdots \\ \mathbf{e}_{1-p}^{(s+1)} & \cdots & \mathbf{e}_{n-p}^{(s+1)} \end{bmatrix}
$$
(65)

with the usual boundary values $\mathbf{e}_i^{(s+1)} = \mathbf{0}_{3\times 1}$ for all $i \le 0$. The current VAR model defines a new de-correlation filter $\mathbf{A}^{(s+1)}(L)$, by which the correlated residuals $\mathbf{e}^{(s+1)}$ are transformed into the white noise residuals

$$
\mathbf{u}^{(s+1)} = \mathbf{A}^{(s+1)}(L)\mathbf{e}_t^{(s+1)}
$$
(66)

in view of (33) and (64). Conditional on these values, equation system (54) gives rise to the solution for the cofactor matrix

$$
\boldsymbol{\Psi}^{(s+1)} = \frac{1}{n}\sum_{t=1}^{n}\widetilde{w}_t^{(s)}\mathbf{u}^{(s+1)}(\mathbf{u}^{(s+1)})^T.
$$
(67)

Finally, $\nu^{(s+1)}$ may be found by a zero search to satisfy (55). It is generally known that the ECM algorithm can be sped up by solving instead the equation

$$
0 = \log\nu^{(s+1)} + 1 - \psi\left(\frac{\nu^{(s+1)}}{2}\right) + \psi\left(\frac{\nu^{(s+1)} + N}{2}\right)
$$
$$
- \log\left(\nu^{(s+1)} + N\right) + \frac{1}{n}\sum_{t=1}^{n}\left(\log\widetilde{w}_t^{(s+1)} - \widetilde{w}_t^{(s+1)}\right)
$$
(68)

where

$$
\widetilde{w}_t^{(s+1)} = \frac{\nu^{(s+1)} + N}{\nu^{(s+1)} + (\mathbf{u}_t^{(s+1)})^T(\boldsymbol{\Psi}^{(s+1)})^{-1}\mathbf{u}_t^{(s+1)}}.
$$
(69)

Equation (68) is obtained by defining the log-likelihood function directly on the joint pdf (30) of the white-noise series based on the multivariate t-distribution, and setting its partial derivative with respect to ν equal to zero. This modification of the ECM algorithm has been called ECM *either* (ECME) [49, 50]. The zero search can be performed with mathematical standard software such as the INTLAB library ([51]) or using a grid search. The estimation of the df concludes the current M-step. The E- and M-step (in the form of an ECME algorithm) are repeated until certain stop criteria are fulfilled, e. g., a combination of a maximum number of iterations and thresholds

$$
d = \max(\max|\widehat{\boldsymbol{\xi}}^{(s)} - \widehat{\boldsymbol{\xi}}^{(s+1)}|, \max|\widehat{\boldsymbol{\alpha}}^{(s)} - \widehat{\boldsymbol{\alpha}}^{(s+1)}|,
$$
$$
\max|\widehat{\boldsymbol{\Sigma}}^{(s)} - \widehat{\boldsymbol{\Sigma}}^{(s+1)}|),
$$
$$
d_\nu = \max|\widehat{\nu}^{(s)} - \widehat{\nu}^{(s+1)}|
$$

for the minimum absolute difference between the current and the previous solution. We defined the threshold d_v by a greater value than the other threshold since the df v fluctuates quite strongly and is the parameter most challenging to estimate. In the following, the solution resulting from the final iteration step is indicated by using "hats", that is, $\widehat{\boldsymbol{\beta}}, \widehat{\boldsymbol{\mu}}$, etc. This EM algorithm was tested in the Monte Carlo simulation of Sect. 4.1. In the special case of a GMM with VAR and multivariate t-distributed errors, a similar unconstrained algorithm can be derived [26]. The methodology is applied in the real data study of Sect. 4.2 and Sect. 4.3.

4 Results

4.1 Gauss-Helmert model: Adjustment of simulated data of a 2D circle

In the first part of this section, the performance of the EM algorithm derived in previous Sect. 3 in producing correct estimates within a Monte Carlo (MC) simulation is studied. For this purpose, generated measurements of $N = 2$ time series are approximated by a 2D circle defined by the nonlinear equation

$$h_t(\boldsymbol{\beta}, \boldsymbol{\mu}) = (x_t - c_x)^2 + (y_t - c_y)^2 - r^2 = 0. \tag{70}$$

On the one hand, the vector $\boldsymbol{\beta}$ of functional parameters consists of the circle center coordinates c_x, c_y and radius r, whose simulated true values are specified by

$$\boldsymbol{\beta} = \begin{bmatrix} c_x \\ c_y \\ r \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix}. \tag{71}$$

On the other hand, the vector $\boldsymbol{\mu} = [\boldsymbol{\mu}_1^T, \ldots, \boldsymbol{\mu}_n^T]^T$ of location parameters consists of n measured points

$$\boldsymbol{\mu}_t = \begin{bmatrix} \mu_{x,t} \\ \mu_{y,t} \end{bmatrix} = \begin{bmatrix} x_t \\ y_t \end{bmatrix},$$

where n was set to 250, 500, 2500 and 5000 in different simulation runs. For greater clarity of notation, we write $\mu_{x,t}$, $\mu_{y,t}$ instead of $\mu_{1,t}$, $\mu_{2,t}$ in the sequel. The true point coordinates were defined by

$$x_t = c_x + r \sin \Phi_t$$
$$y_t = c_y + r \cos \Phi_t$$

with

$$\Phi_t = t \cdot \frac{10}{n} \quad (t \in \{1, \ldots, n\}).$$

Thus, the measurements are equidistantly distributed along the circle, and one half of the circle was measured twice. The numerical realizations $\mathbf{e}_t = [e_{x,t}, e_{y,t}]^T$ of the corresponding random errors are assumed to be determined by a VAR(1) process

$$\begin{bmatrix} e_{x,t} \\ e_{y,t} \end{bmatrix} = \begin{bmatrix} 0.5653 & -0.0197 \\ -0.0431 & 0.7577 \end{bmatrix} \begin{bmatrix} e_{x,t-1} \\ e_{y,t-1} \end{bmatrix} + \begin{bmatrix} u_{x,t} \\ u_{y,t} \end{bmatrix}, \tag{72}$$

where random numbers $\mathbf{u}_t = [u_{x,t}, u_{y,t}]^T$ for the white-noise components were generated from the multivariate t-distribution

$$\begin{bmatrix} \mathcal{U}_{x,t} \\ \mathcal{U}_{y,t} \end{bmatrix} \overset{\text{ind.}}{\sim} t(\mathbf{0}, \boldsymbol{\Psi}, v) = t\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \sigma_0^2 \cdot \begin{bmatrix} \boldsymbol{\Psi}_x & \boldsymbol{\Psi}_{x,y} \\ \boldsymbol{\Psi}_{x,y} & \boldsymbol{\Psi}_y \end{bmatrix}, v \right)$$

with $\sigma_0^2 = 0.001^2$, $\boldsymbol{\Psi}_x = 1.0$, $\boldsymbol{\Psi}_y = 4.0$ and $\boldsymbol{\Psi}_{x,y} = 0.5$. The given cofactor matrix reflects the assumption that the entries of $\mathcal{U}_t$ are pairwise correlated by correlation coefficient $\rho = 0.25$. Simulations were carried out for the df $v = 4$ (referred to as "stochastic model A") and $v = 7$ ("stochastic model B").

In total 500 MC samples of the white-noise components for stochastic model A and B as well as for the different sample sizes n were randomly generated using MATLAB's mvtrnd routine. The resulting simulated observations in models A and B were then adjusted by means of the EM algorithm. The maximum number of iterations was set to itermax = 50, and the convergence thresholds to $\varepsilon = 10^{-8}$ and $\varepsilon_v = 0.001$. The unknown parameters $\boldsymbol{\beta}$ and $\boldsymbol{\Psi}$ of the simulated observations of Model A were additionally estimated using the classical least-squares approach to the GHM ("LS"), without iterative re-weighting and without imposing a VAR model (cf. [18]). The comparison between the results of LS and EM algorithm demonstrates the effect of assuming a simplified stochastic model.

4.1.1 Analysis of the simplified stochastic model on the estimation results

The results of this comparison are visualized in Fig. 1 and Fig. 2 by means of box plots. In each box plot, the middle line shows the median. The bottom and the top border of the boxes are the 25th percentile (Q1) and 75th percentile (Q3), respectively. The distance Q3-Q1 is the inter-quartile range (IQR). The additional lines above and below the box are the whiskers defined by $Q1 - 1.5 \cdot \text{IQR}$ and $Q3 + 1.5 \cdot \text{IQR}$, respectively. The red crosses indicate the outliers in the sense that their value is more than 1.5 times the IQR away from Q1 and Q3. The results of Fig. 1 show that the precision of the estimated parameters $\widehat{\boldsymbol{\beta}}$ decreases with increasing number of observations. Overall, the mean values of

the estimated parameters are similar for LS and EM. However, Fig. 1 shows for numbers of observations less than 5000 that LS produces numerous extreme MC radius estimates outside of the inter-quartile range far away from the true radius, whereas all radius values estimated by EM lie within that range. Thus, EM is significantly more accurate than LS as far as the estimation of the radius is concerned. Figure 2 shows the estimated scale parameters $\widehat{\boldsymbol{\Psi}}$ for 500 MC runs by means of EM and LS estimation. The number of outliers (red crosses) resulting from the LS estimate is significantly larger than the number of outliers from the EM algorithm. It can be clearly seen that the bias in the EM-estimated scale parameters is much smaller compared to the LS-estimated scale parameters. As the number of observations increases, the bias of the estimated scale parameters decreases and the dispersion becomes smaller. The bias of all estimated parameter can be calculated for each MC run by

$$\Delta\hat{\boldsymbol{\Phi}}^{(i)} = \hat{\boldsymbol{\Phi}}^{(i)} - \boldsymbol{\Phi} \quad (i = 1,\ldots,500). \tag{73}$$

The statistical moments of the calculated bias $\Delta\boldsymbol{\beta}$ and $\Delta\boldsymbol{\Psi}$ of the simulations with 5000 observations are presented in Table 1. For the functional parameters $\boldsymbol{\beta}$ the true value lies in the 95 % confidence interval for all estimates (LS, EM model A and EM model B). In contrast, the true values of $\boldsymbol{\Psi}$ lie outside the estimated 95 %-confidence interval of the estimated cofactor matrix $\widehat{\boldsymbol{\Psi}}$ when LS is employed. Applying EM, this statement is valid only for the 95 %-confidence interval for $\widehat{\boldsymbol{\Psi}}_x$ under stochastic model A and for $\widehat{\boldsymbol{\Psi}}_y$ under stochastic model B. Here, the biases of the estimator $\widehat{\boldsymbol{\Psi}}$ with respect to the true simulation parameter values applying EM is less than the corresponding bias when employing LS. The greater deviation concerning the VCM estimated by LS is the result of not taking the correlations and the heavy-tailedness of the noise into account.

4.1.2 Analysis of the degree of freedom on the estimation results

The influence of the df on the estimation results using the EM algorithm is displayed in Fig. 3 and Fig. 4. Figure 3 shows the estimated $\widehat{\boldsymbol{\beta}}_x$, estimated scale parameters $\widehat{\boldsymbol{\Psi}}$ and the differences of the estimated ν to the true values for 500 MC runs by means of EM model A and EM model B estimation. The values of the estimated functional parameters $\widehat{\boldsymbol{\beta}}$ under stochastic model A and stochastic model B show no significant differences. It can be clearly seen that the bias in the EM model B estimated scale parameters is

much smaller compared to EM model A estimated scale parameters. The number of outliers (red crosses) resulting from the EM model A estimate is significantly larger than the number of outliers from the EM model B algorithm. For both models (A and B) applies that if the number of observations is increased, the bias and the dispersion becomes smaller. The bias of the estimated df $\hat{\nu}$ under model A is approximately 2.25, which is significantly larger than the bias under model B (approximately 0.44). It can clearly be seen that increasing the number of observations has no significant influence on the bias of the estimated df $\hat{\nu}$ for both models. The estimated VAR coefficients for both models (EM model A and EM model B) are shown in Fig. 4. The estimated MC solutions for both models show similar spreading behavior around the true value. The number of observations only influences the precision of the estimation but does not lead to a reduction of the bias. The statistical moments of the calculated bias $\Delta\hat{\boldsymbol{\alpha}}$ and $\Delta\hat{\nu}$ of the simulations with 5000 observations are presented in Table 2. For the VAR parameters only $\alpha_{1;2,2}$ lies in the 95 % confidence interval.

4.2 Gauss-Markov model: Adjustment of 3D MEMS accelerometer data

In the second part of this section, the performance of the EM algorithm elaborated in Sect. 3 (which we also refer to as "VAR-multivariate algorithm" in the following) is studied on real data sets, which were measured by an accelerometer. Cost-effective Micro-Electro-Mechanical-Systems (MEMS) accelerometers have received increasing attention in numerous applications such as vibration analysis of bridge structures, e. g., [52, 53, 54]. In such applications, it is desirable to select a suitable accelerometer by considering, for instance, measurement and sensitivity ranges, uncertainty of measurements, cost, and sampling rate. [55] proposed a two-step scenario to accomplish the suitability analysis for MEMS accelerometers. Firstly, the calibration of the MEMS accelerometers was carried out for different positions and over different temperature ranges. Therefore, the characteristics of the calibration parameters such as biases, scale factors and non-orthogonalities between axes have been inspected during different temperature changes. Secondly, a controlled excitation experiment was conducted using a shaker to compare and validate estimated harmonic oscillation parameters including frequency, amplitude and phase shift with other MEMS accelerometers as well as a high-end reference accelerometer. The second step additionally allowed
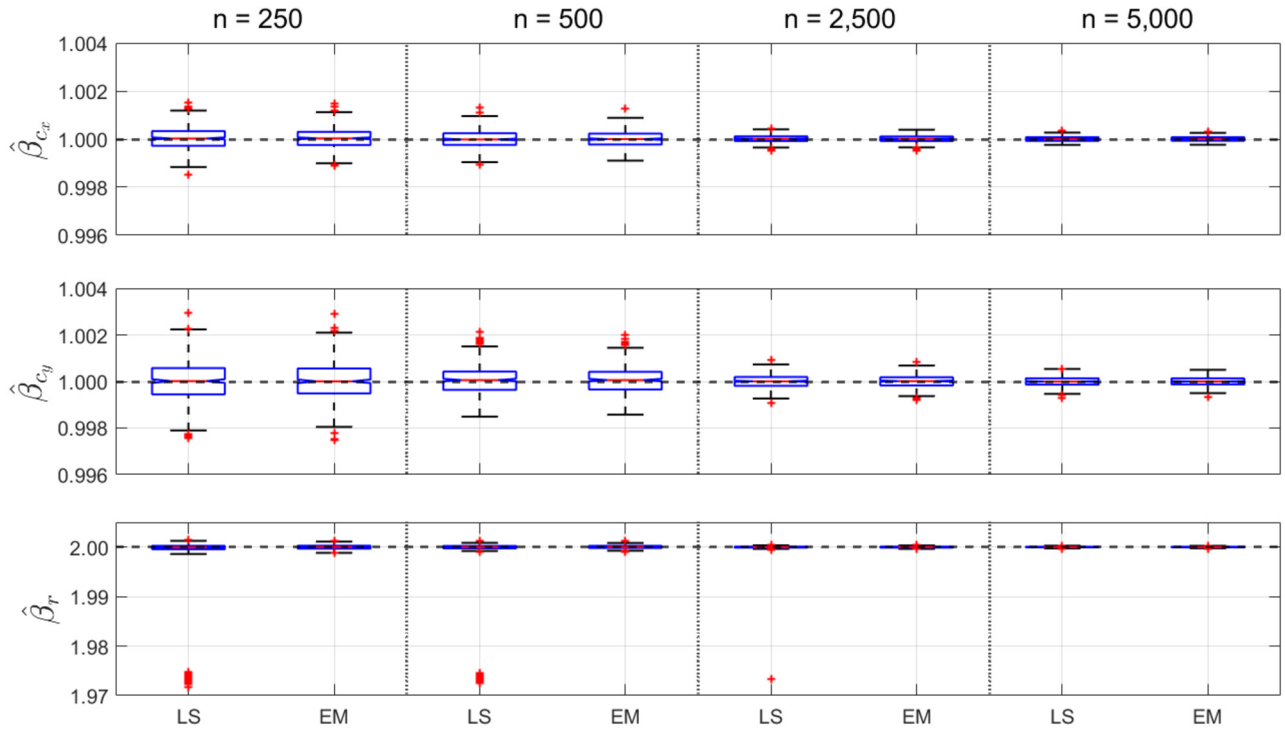
Figure 1: MC simulation results of the estimated parameters for stochastic model A. Dash line: true value of simulation.
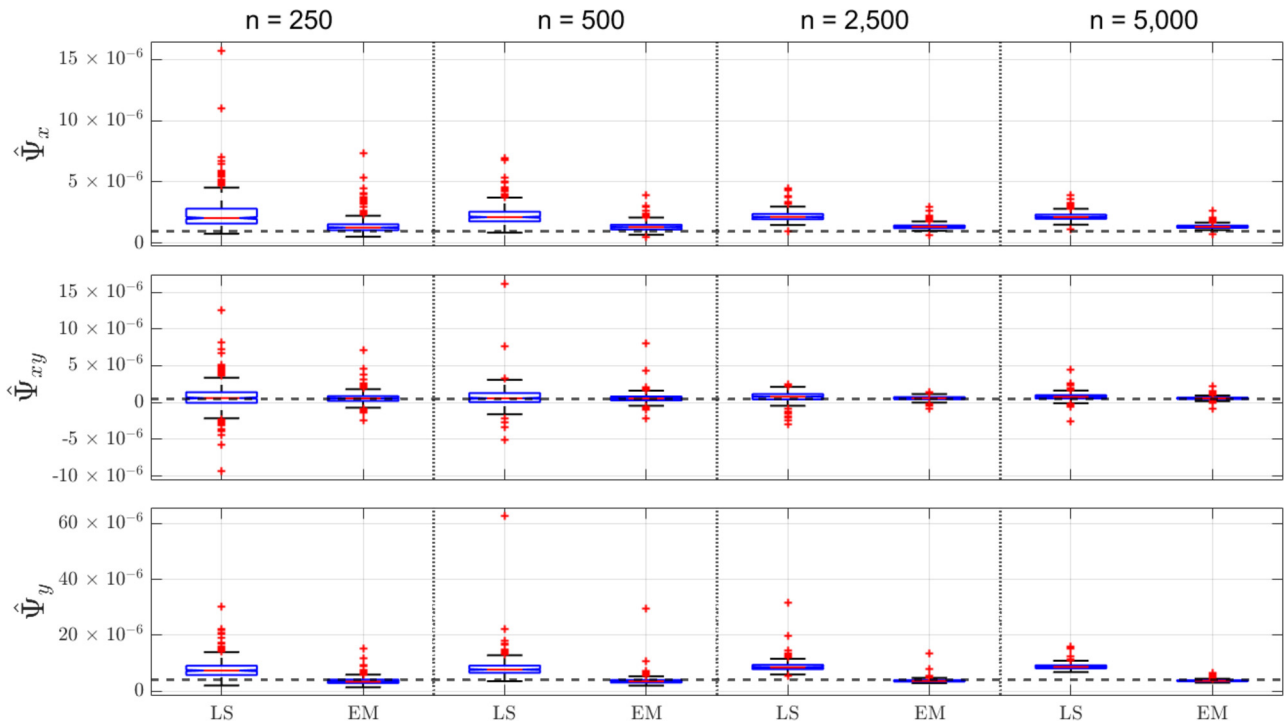

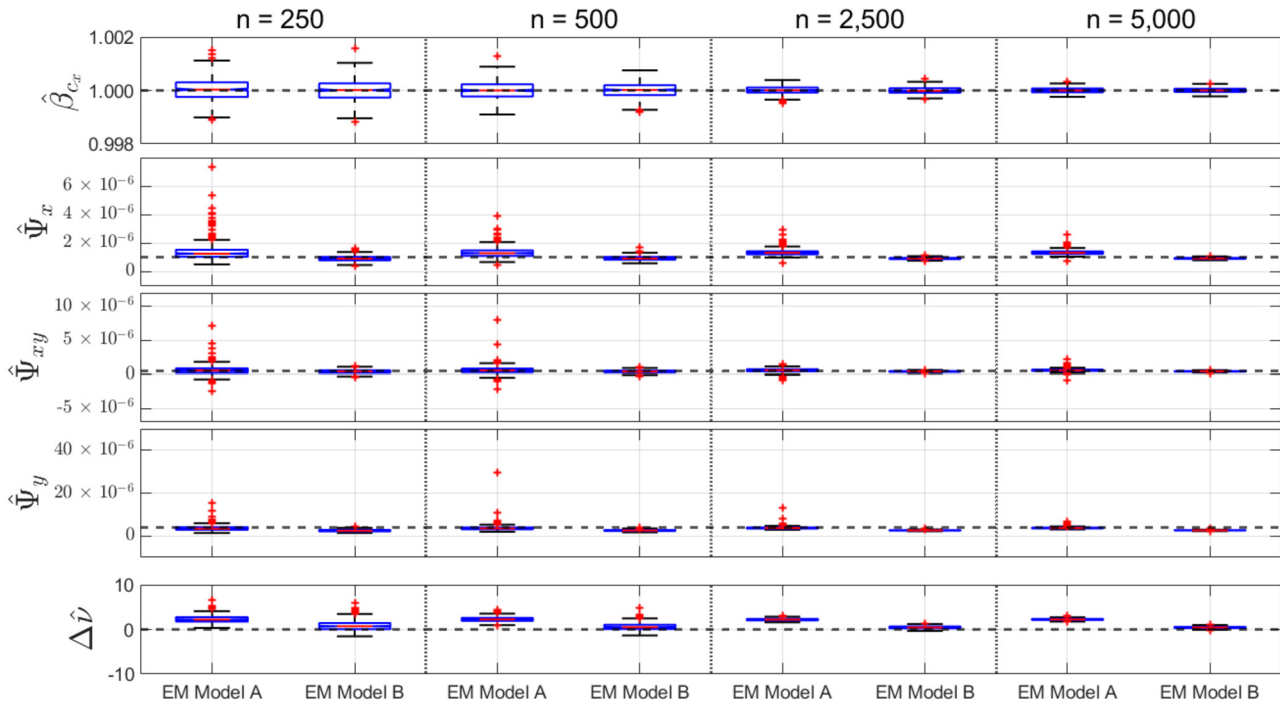**Figure 2:** MC simulation results of the estimated VCM for stochastic model A.

Figure 3: Comparison of the MC simulation results between the stochastic model A and B. The maximum deviation of $\Delta\hat{\nu}$ for the stochastic model B with $n = 250$ is 113.
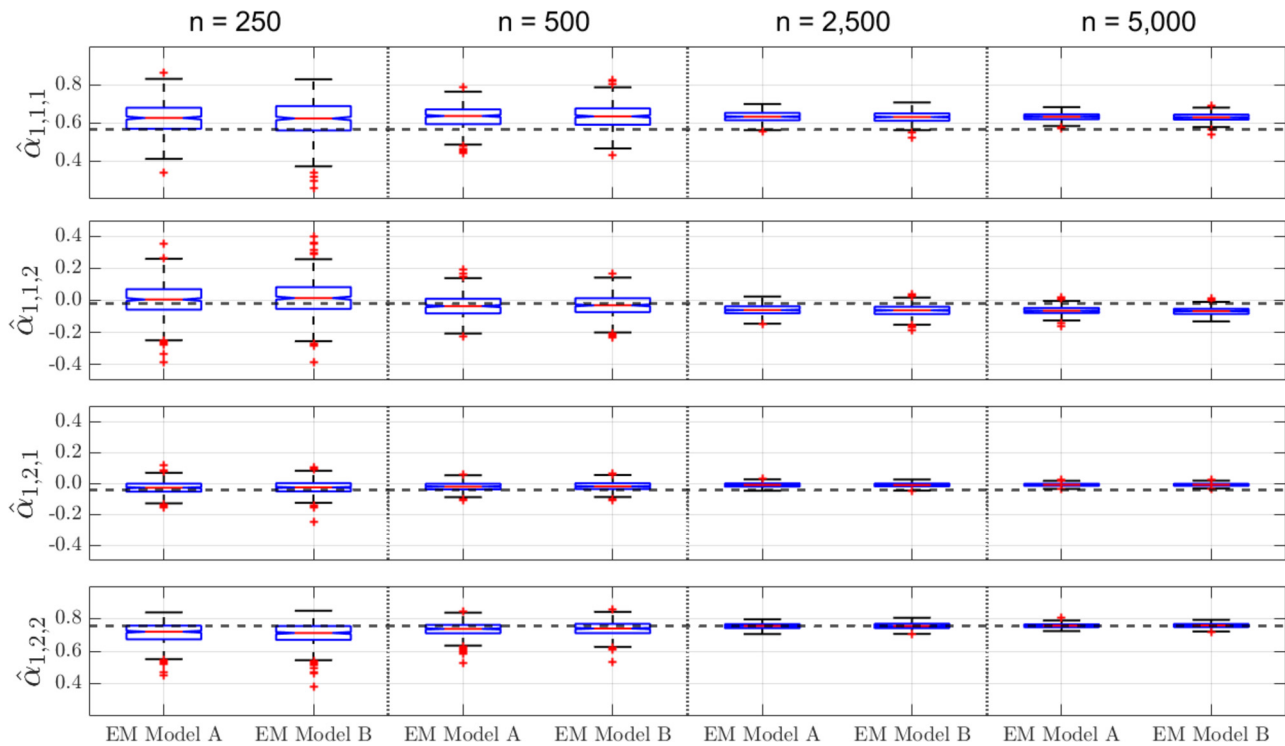

Figure 4: Comparison of VAR coefficients between the MC simulation results of stochastic model A and B.

Table 1: Statistics of MC simulation results for the bias of the estimated parameters $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{\Psi}}$. The results were calculated for the simulation with $n = 5000$ observations. The boundaries of the 95 % interval estimates are defined by the empirical 2.5 and 97.5 percentiles.

	LS Model A	EM Model A	EM Model B
$\Delta\widehat{\beta}_{c_x}$			
mean	$3.7 \cdot 10^{-6}$	$3.0 \cdot 10^{-6}$	$1.2 \cdot 10^{-6}$
std	$1.1 \cdot 10^{-4}$	$9.8 \cdot 10^{-5}$	$9.0 \cdot 10^{-5}$
minimum	$-2.4 \cdot 10^{-4}$	$-2.4 \cdot 10^{-4}$	$-2.2 \cdot 10^{-4}$
maximum	$3.5 \cdot 10^{-4}$	$3.1 \cdot 10^{-4}$	$2.5 \cdot 10^{-4}$
95 % interval	$[-2.0 \cdot 10^{-4},\ 2.1 \cdot 10^{-4}]$	$[-1.9 \cdot 10^{-4},\ 1.9 \cdot 10^{-4}]$	$[-1.7 \cdot 10^{-4},\ 1.8 \cdot 10^{-4}]$
$\Delta\widehat{\beta}_{c_y}$			
mean	$-6.9 \cdot 10^{-7}$	$6.0 \cdot 10^{-7}$	$7.1 \cdot 10^{-6}$
std	$2.0 \cdot 10^{-4}$	$1.9 \cdot 10^{-4}$	$1.7 \cdot 10^{-4}$
minimum	$-7.1 \cdot 10^{-4}$	$-6.5 \cdot 10^{-4}$	$-4.7 \cdot 10^{-4}$
maximum	$5.5 \cdot 10^{-4}$	$5.0 \cdot 10^{-4}$	$4.2 \cdot 10^{-4}$
95 % interval	$[-3.9 \cdot 10^{-4},\ 4.1 \cdot 10^{-4}]$	$[-3.8 \cdot 10^{-4},\ 3.9 \cdot 10^{-4}]$	$[-3.3 \cdot 10^{-4},\ 3.4 \cdot 10^{-4}]$
$\Delta\widehat{\beta}_r$			
mean	$6.3 \cdot 10^{-6}$	$3.3 \cdot 10^{-6}$	$3.8 \cdot 10^{-7}$
std	$9.9 \cdot 10^{-5}$	$9.0 \cdot 10^{-5}$	$8.3 \cdot 10^{-5}$
minimum	$-3.1 \cdot 10^{-4}$	$-2.7 \cdot 10^{-4}$	$-2.1 \cdot 10^{-4}$
maximum	$2.8 \cdot 10^{-4}$	$2.6 \cdot 10^{-4}$	$2.7 \cdot 10^{-4}$
95 % interval	$[-1.9 \cdot 10^{-4},\ 1.9 \cdot 10^{-4}]$	$[-1.9 \cdot 10^{-4},\ 1.7 \cdot 10^{-4}]$	$[-1.5 \cdot 10^{-4},\ 1.6 \cdot 10^{-4}]$
$\Delta\widehat{\Psi}_x$			
mean	$1.2 \cdot 10^{-6}$	$3.5 \cdot 10^{-7}$	$-6.1 \cdot 10^{-8}$
std	$2.8 \cdot 10^{-7}$	$1.6 \cdot 10^{-7}$	$4.8 \cdot 10^{-8}$
minimum	$1.2 \cdot 10^{-7}$	$-2.5 \cdot 10^{-7}$	$-1.9 \cdot 10^{-7}$
maximum	$2.9 \cdot 10^{-6}$	$1.6 \cdot 10^{-6}$	$1.1 \cdot 10^{-7}$
95 % interval	$[7.0 \cdot 10^{-7},\ 1.8 \cdot 10^{-6}]$	$[1.2 \cdot 10^{-7},\ 6.7 \cdot 10^{-7}]$	$[-1.5 \cdot 10^{-7},\ 3.2 \cdot 10^{-8}]$
$\Delta\widehat{\Psi}_{x,y}$			
mean	$2.8 \cdot 10^{-7}$	$7.4 \cdot 10^{-8}$	$-9.2 \cdot 10^{-8}$
std	$4.7 \cdot 10^{-7}$	$2.1 \cdot 10^{-7}$	$6.5 \cdot 10^{-8}$
minimum	$-3.1 \cdot 10^{-6}$	$-1.4 \cdot 10^{-6}$	$-3.1 \cdot 10^{-7}$
maximum	$4.0 \cdot 10^{-6}$	$1.7 \cdot 10^{-6}$	$1.1 \cdot 10^{-7}$
95 % interval	$[-4.6 \cdot 10^{-7},\ 1.3 \cdot 10^{-6}]$	$[-3.0 \cdot 10^{-7},\ 4.6 \cdot 10^{-7}]$	$[-2.2 \cdot 10^{-7},\ 3.5 \cdot 10^{-8}]$
$\Delta\widehat{\Psi}_y$			
mean	$4.8 \cdot 10^{-6}$	$-2.6 \cdot 10^{-7}$	$-1.4 \cdot 10^{-6}$
std	$9.4 \cdot 10^{-7}$	$3.6 \cdot 10^{-7}$	$1.3 \cdot 10^{-7}$
minimum	$2.8 \cdot 10^{-6}$	$-9.7 \cdot 10^{-7}$	$-1.8 \cdot 10^{-6}$
maximum	$1.2 \cdot 10^{-5}$	$2.6 \cdot 10^{-6}$	$-7.1 \cdot 10^{-7}$
95 % interval	$[3.4 \cdot 10^{-6},\ 6.7 \cdot 10^{-6}]$	$[-7.5 \cdot 10^{-7},\ 5.1 \cdot 10^{-7}]$	$[-1.6 \cdot 10^{-6},\ -1.2 \cdot 10^{-6}]$

to check the time synchronization between the MEMS accelerometers. Later on [54] extended the aforementioned two-step scenario to three-step scenario. In the third-step, a static test experiment was accomplished over a long period of time, which allows to estimate an offset and a drift of the measurements as well as auto-correlations and underlying distributional parameters for each axis individually.

As part of this study and to make the third step of the aforementioned three-step scenario more comprehensive, the (unknown) auto- as well as cross-correlations and the (unknown) distributional characteristics of the acceleration measurements recorded from three different axes are analyzed by employing VAR models with t-distributed er-

rors. For this purpose, acceleration measurements were recorded from a cost-effective MEMS accelerometer of type BNO055 with Arduino UNO-Board and 9 axes motion shield (produced from Bosch GmbH), which is a so-called NAMS. The measurements were carried out in three different axes over a period of 24 hours (Fig. 5) with a sampling frequency of 0.3789 Hz. Although it is possible to perform the measurements with a higher sampling frequency (e. g., up to 200 Hz), the aforementioned sampling frequency is specified to speed up the processing.

An observation model consisting of (1) a functional model based on a linear drift with an unknown offset (intercept) and a linear drift coefficient, (2) an auto- and cross-correlation model based on a VAR process with un-

Table 2: Statistics of MC simulation results for the bias of the estimated parameters $\hat{v}$ and $\hat{\alpha}$. The results were calculated for the simulation with $n = 5000$ observations. The boundaries of the 95 % interval estimates are defined by the empirical 2.5 and 97.5 percentiles.

	EM Model A	EM Model B
$\Delta\hat{v}$		
mean	2.25	0.44
std	0.21	0.21
minimum	1.70	−0.19
maximum	3.12	1.16
95 % interval	[1.87, 2.70]	[0.03, 0.89]
$\Delta\hat{\alpha}_{1;1,1}$		
mean	$6.6 \cdot 10^{-2}$	$6.3 \cdot 10^{-2}$
std	$1.9 \cdot 10^{-2}$	$2.1 \cdot 10^{-2}$
minimum	$0.8 \cdot 10^{-2}$	$-2.6 \cdot 10^{-2}$
maximum	$11.7 \cdot 10^{-2}$	$12.4 \cdot 10^{-2}$
95 % interval	$[2.9 \cdot 10^{-2},\ 10.1 \cdot 10^{-2}]$	$[2.2 \cdot 10^{-2},\ 10.3 \cdot 10^{-2}]$
$\Delta\hat{\alpha}_{1;1,2}$		
mean	$-4.4 \cdot 10^{-2}$	$-4.8 \cdot 10^{-2}$
std	$2.4 \cdot 10^{-2}$	$2.4 \cdot 10^{-2}$
minimum	$-14.2 \cdot 10^{-2}$	$-11.2 \cdot 10^{-2}$
maximum	$3.9 \cdot 10^{-2}$	$3.0 \cdot 10^{-2}$
95 % interval	$[-8.7 \cdot 10^{-2},\ -0.2 \cdot 10^{-2}]$	$[-9.5 \cdot 10^{-2},\ -0.3 \cdot 10^{-2}]$
$\Delta\hat{\alpha}_{1;2,1}$		
mean	$3.3 \cdot 10^{-2}$	$3.4 \cdot 10^{-2}$
std	$1.0 \cdot 10^{-2}$	$1.0 \cdot 10^{-2}$
minimum	$4.8 \cdot 10^{-3}$	$6.5 \cdot 10^{-3}$
maximum	$6.9 \cdot 10^{-2}$	$6.8 \cdot 10^{-2}$
95 % interval	$[1.2 \cdot 10^{-2},\ 5.3 \cdot 10^{-2}]$	$[1.6 \cdot 10^{-2},\ 5.2 \cdot 10^{-2}]$
$\Delta\hat{\alpha}_{1;2,2}$		
mean	$-7.3 \cdot 10^{-4}$	$6.3 \cdot 10^{-4}$
std	$1.3 \cdot 10^{-2}$	$1.4 \cdot 10^{-2}$
minimum	$-3.4 \cdot 10^{-2}$	$-3.9 \cdot 10^{-2}$
maximum	$4.6 \cdot 10^{-2}$	$3.5 \cdot 10^{-2}$
95 % interval	$[-2.9 \cdot 10^{-2},\ 2.4 \cdot 10^{-2}]$	$[-2.7 \cdot 10^{-2},\ 2.8 \cdot 10^{-2}]$
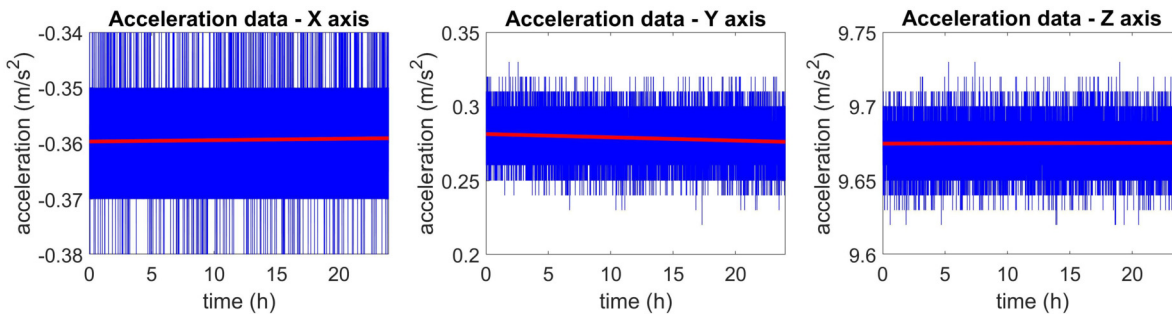
Figure 5: Raw acceleration data recorded from the MEMS accelerometer of type NAMS (blue) and fitted linear model (red) estimated from VAR-multivariate algorithms.

known model order and coefficients, and (3) a stochastic model based on the multivariate t-distribution with unknown scale matrix and df. This observation model has the structure of the GMM with VAR and t-distributed errors. The maximum number of iterations is set to itermax = 200, the convergence thresholds to $\varepsilon = 10^{-12}$, and $\varepsilon_v = 0.001$.

The model parameters estimated by means of the VAR-multivariate algorithm are then compared with the estimates resulting from the "AR-univariate algorithm" (see [54] and [56] for details) where each axis of the MEMS accelerometers is modeled by an individual AR process. The static acceleration measurements $\ell_1, \ldots, \ell_n$ ($n = 32743$) are

Table 3: Statistics of the estimated unknown parameters in the functional model based on the linear drift model, the auto- and cross-correlation model based on the VAR process and the stochastic model based on the centered and scaled t-distribution with an unknown degree of freedom and unknown scale factor from the AR-univariate and VAR-multivariate algorithms.

Sensor	Method	Axis	$\hat{c}_0$ [m/s^2]	$\hat{c}_1$ [m/s^3]	$\hat{\sigma}_l$ [m/s^2]	p [−]	$\hat{\alpha}_l$ [−]	WNT [−]	$\hat{v}$ [−]
NAMS	multivar.	x	−0.3598	4.80e-09	0.0224	1	−	yes	2.1
		y	0.2810	−6.11e-08	0.0493		−		
		z	9.6752	4.63e-09	0.0534		−		
	univar.	x	−0.3597	6.81e-09	0.0066	1	−6.2927e-04	yes	16.87
	univar.	y	0.2811	−6.12e-08	0.0132	1	0.0013	yes	120
	univar.	z	9.6746	6.65e-09	0.0141	1	0.0087	yes	120

modeled based on a linear drift, which is described by the model equations

$$\ell_t = \mathbf{f}_t(\boldsymbol{\beta}) + e_t$$
$$= c_0 + c_1 x_t + e_t \quad (t = 1, \ldots, t_n). \tag{74}$$

where the vector $\boldsymbol{\beta}$ of unknown functional parameters consists of the offset c_0 in [m/s^2] and the linear drift coefficient c_1 in [m/s^3]. The quantities $x_1, \ldots, x_n$ represent equidistant time instances and $e_1, \ldots, e_n$ the random deviations contaminated with colored noise. The preceding observation equations can be written in the form of

$$\ell_t = \mathbf{X}_t \boldsymbol{\beta} + e_t, \tag{75}$$

where $\mathbf{X}_t$ are the rows of the design matrix

$$\mathbf{X} = \begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}, \tag{76}$$

which has full rank.

Table 3 summarizes the statistics of the estimated parameters from the VAR-multivariate and AR-univariate algorithms based on the GMM. $\hat{c}_0$ [m/s^2] is the estimated offset and $\hat{c}_1$ the estimated linear drift coefficient [m/s^3]. p is the VAR model order in case of the VAR-multivariate algorithm and the AR model order in case of the AR-univariate algorithm. WNT stands for "white-noise test" and indicates the acceptance ("yes") or rejection ("no") of the multivariate portmanteau test introduced by [57] and adapted by [26]. $\hat{\sigma}_l$ is the estimated standard deviation of the white noise components. In this study, the multivariate portmanteau test is applied to select an adequate VAR model order by testing the residuals of the VAR model if they are uncorrelated or not. Therefore, the p-value (pv) is calculated based on different VAR model orders $p = 1, \ldots, 10$. Beforehand, the significance level is defined as $\alpha = 0.05$, and the maximum lag is set to 20 similar to the research of [26]. The WNT is accepted if
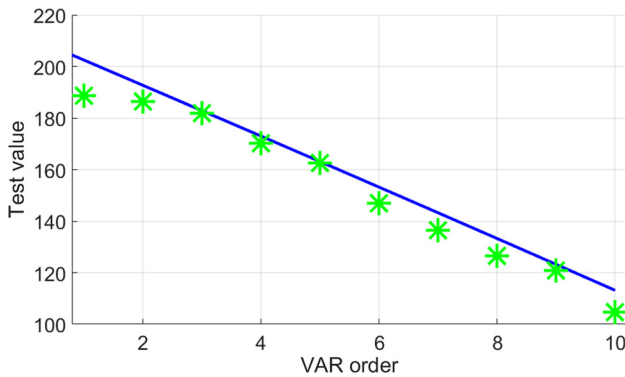
$pv > 0.05$, which corresponds to the acceptance of the white noise hypothesis, otherwise it is rejected. To estimate the df of the underlying t-distribution a zero search based on the reliable interval Newton method as described in [58] is performed by using the INTLAB library [51]. The estimated unknown parameters including offset and drift in both VAR-multivariate and AR-univariate algorithms are rather similar to each other and no significant differences are observable. The df estimate initially was below 2 with the application of the VAR-multivariate algorithm, so that the covariance matrix $\boldsymbol{\Sigma}$ of the white noise is theoretically undefined. We also found some numerical instability with the estimation of the parameters of the scale matrix, so that we set the estimate $\hat{v}$ to the small but numerically stable value 2.1 and repeated the algorithm with this fixed df. In contrast, the AR-univariate algorithm produced much larger df estimates (16.87, 120, and 120) for the individual axes. The value 120 is the maximum value of the zero search interval; if the zero is not found within the search interval, it is automatically set to the threshold 120 (which happened in the present case). This threshold indicates the fitted t-distribution already is very close to a normal distribution. We may therefore conclude that the combination of a VAR process with a multivariate t-distribution yields quite different results than the three individual AR processes with univariate t-distributions, which ignore the cross-correlations in the noise. The reason for why the latter combination leads to increased df estimates needs to be investigated further. The standard deviations of the white noise components for both VAR-multivariate and AR-univariate algorithms are estimated, see Table 3. As it could be seen, the uncertainties in the x-axis of the NAMS acceleration data are less than for the two other axes.

Fig. 6 (a) illustrates the estimated p-values for different VAR model orders. The estimated p-values are all above the significance level, and thus, the VAR model order 1 is selected. Fig. 7 (b) shows the calculated test values that are compared with the cumulative distribution function of the
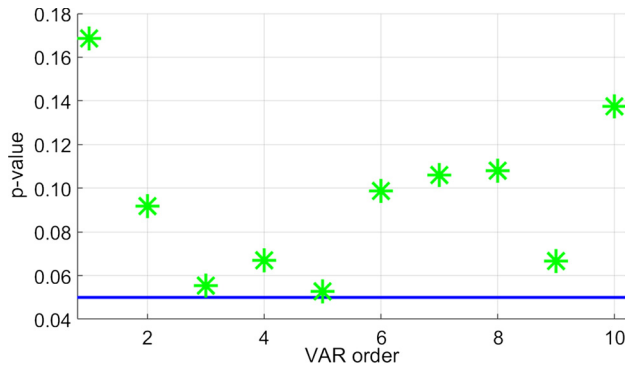
Figure 6: Results of the portmanteau test (p-value) for VAR model orders $p = 1, \ldots, 10$.

Figure 7: Results of the test values for VAR model orders $p = 1, \ldots, 10$.

Figure 8: Representation of the estimated colored and white noise residuals (x, y, and z axes in a sequence) obtained from NAMS acceleration data for VAR model order 1.
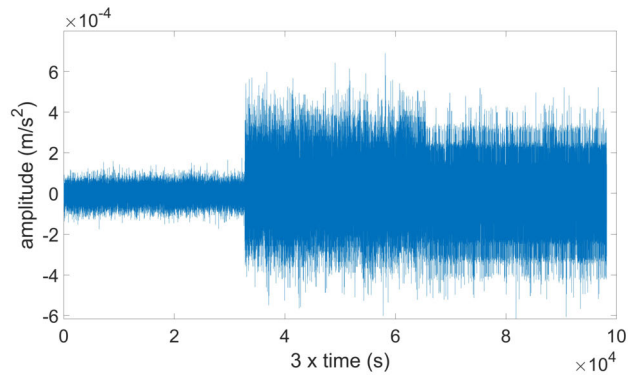
Figure 9: Differences of the estimated residuals (x, y, and z axes in a sequence) (b).

chi-square distribution ($F_{\chi^2(4^2 \cdot (20-p))}$) with $df = 4^2 \cdot (20 - p)$ [26]. As it could be seen, all test values are below the critical values, and thus, the WNT are accepted.

The estimated VCM ($\mathbf{\Sigma}$) and its corresponding correlation matrix ($\mathbf{R}$) based on the VAR-multivariate algorithm are

$$\widehat{\mathbf{\Sigma}} = \begin{bmatrix} 0.000501 & 2.0464e-05 & 9.7799e-06 \\ 2.0464e-05 & 0.002429 & -4.8095e-05 \\ 9.7799e-06 & -4.8095e-05 & 0.002862 \end{bmatrix}$$
(77)

$$\widehat{\mathbf{R}} = \begin{bmatrix} 1 & 0.018551 & 0.008167 \\ 0.018551 & 1 & -0.018241 \\ 0.008167 & -0.018241 & 1 \end{bmatrix}$$
(78)

The latter matrix contains the correlation coefficients between the different axes. As it could be seen, the correlation coefficients are very small and are negligible. Additionally, the coefficients of the VAR model order 1 is arranged within the (3×3)-matrix

$$\widehat{\mathbf{A}}_1 = \begin{bmatrix} 5.7887e-04 & -0.0012 & 0.0032 \\ 0.0090 & -0.0020 & -0.0090 \\ -0.0120 & -2.0993e-04 & 0.0083 \end{bmatrix},$$
(79)

according to the definition of the VAR coefficient matrix (23). As it could be seen, the VAR model coefficients corresponding to y and z axes are slightly greater than those for the x-axis. In addition, the estimated colored and white noise residuals obtained from NAMS acceleration data are illustrated in Fig. 8. The subtraction of the aforementioned residuals are provided (Fig. 9) to have a better realization of the amplitude of the colored noise residuals at each axis. As it could be seen, the y and z-axes of the NAMS accelerometer have stronger colored measurement noise in comparison to the x-axis.

The proposed approach in this study assists us to select a proper MEMS accelerometer for, e. g., the purpose of vibration analysis of bridge structures. It is applied as a third step of the aforementioned scenario in the suitability analysis of the MEMS accelerometer by providing information about unknown offset and drift coefficients, VAR model order and stochastic model parameters. Therefore, a sensor with possibly less offset and drift coefficients over a long period as well as lower uncertainty of the mea-

surements could be selected. In addition, a minimum VAR model order between different MEMS accelerometers reveals less cross-correlation between their axes, which can also be considered as an important influencing factor in the selection procedure. Moreover, it is observed that the MEMS accelerometers used in this study have minimum standard deviations in the x-axis compared to the two other axes. Therefore, the sensors should be mounted with their x-axis in the main observations direction during the monitoring of bridge structures.

4.3 Gauss-Helmert model: Adjustment of a sphere from continuous laser tracker measurements

In the third part of this section, the performance of the EM algorithm in a real-world application estimating the parameters $\boldsymbol{\beta} = \begin{bmatrix} c_x & c_y & c_z & r \end{bmatrix}^T$ of sphere is studied. The practical relevance of this sphere parameter estimation can be seen in the field of surface fitting applications as well as referencing applications of surface and point-wise measuring sensors, i. e., laser scanner and laser tracker measurements. The 3D sphere is defined by the nonlinear equation

$$h_t(\boldsymbol{\beta},\boldsymbol{\mu}) = (x_t - c_x)^2 + (y_t - c_y)^2 + (z_t - c_z)^2 - r^2 = 0, \quad (80)$$

which is an extension in 3D space by considering the z-component in addition to the nonlinear equation of the 2D circle in (70).

To illustrate and discuss the adjustment of a sphere, 3D coordinates on the surface of a sphere with known radius were obtained by continuous measurements with 10 Hz using a laser tracker of type Leica AT960-LR. The extended uncertainty of the x-, y- and z-coordinate is given by $U_{x,y,z} = \pm 15\,\mu m + 6\,\mu m/m$ (MPE according to ISO 10360-10) while using a Leica Red Ring Reflector 1.5" (RRR) in single measurement mode [60]. Here, the data acquisition is performed in continuous measurement mode at 10 Hz using the software tool Tracker Pilot. In the acquisition process, the RRR was moved randomly over the sphere surface from only one laser tracker setup; thus, only a half-sphere is measured. To capture the full sphere surface at least two standpoints of the laser tracker are required. In this case, the measurements have to be transformed into a common coordinate system. This leads to the presence of referencing uncertainties besides sensor-related uncertainties. In order to focus on sensor uncertainties only, the sphere is only observed from one standpoint. The distance between the laser tracker and the sphere was approx-
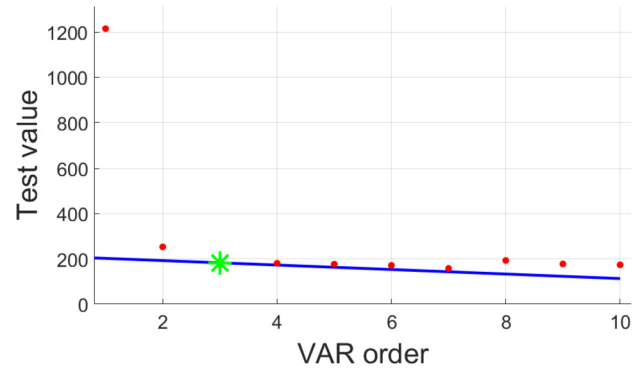
Figure 10: Results of the test values for VAR model orders $p = 1,\ldots,10$. Blue line indicates the critical value for the significance level $\alpha = 0.05$ and the maximum lag is set to 20. The red dots show the test values of the weighted white noise, which are larger than the critical value. The green star represents the test value matching the null hypothesis H_0.

imately 4 m and the y-axis of the local coordinate system of the laser tracker was oriented towards the sphere. The Cartesian coordinates obtained yield a time series of dimension $N = 3$ for a period of approximately 90 s, resulting in $n = 888$ 3D points. There are no data gaps, which results in a $\Delta t = 0.1\,s$ for subsequent measurement. The environmental temperature during the experiment was constant 19.3 °C.

4.3.1 Analysis of the measurements

In analogy to the analysis scheme for the simulated 2D circle in Sect. 4.1, the sphere parameters are obtained on the one hand by means of a LS estimation, i. e. a classical GHM, and on the other hand by means of the proposed EM algorithm in this contribution. The thresholds used within the EM algorithm are similar to the simulation. In this real-world application study, the multivariate portmanteau test is applied to select an adequate VAR model order p by testing the estimated residuals $\hat{\mathbf{u}}_t$ of the VAR model if they are uncorrelated (i. e., by checking for the presence of white noise). This VAR model order selection for $p = 0,\ldots,10$ was similarly applied in Sect. 4.2 using the significance level $\alpha = 0.05$, and the maximum lag is set to 20 similar to the research of [26]. For the VAR model order $p = 0$ no multivariate VAR process is estimated. Instead, the GHM with stochastically independent t-distributed errors is evaluated. The EM algorithm estimates the unknown parameters $\boldsymbol{\beta}$, $\boldsymbol{\Psi}$, ν and $\mathbf{A}_j$, which yields – depending on the VAR model order – in total $4 + 6 + 1 + 9 \cdot p$ parameters to estimate.

Table 4: Estimated sphere parameters for the LS model and the VAR model orders $p = 0$ and $p = 3$.

|	**LS Model**	**EM Model $p = 0$**	**EM Model $p = 3$**
$\widehat{\beta}_{c_x}$ [mm]	99.317	99.317	99.317
$\widehat{\beta}_{c_y}$ [mm]	4014.130	4014.129	4014.125
$\widehat{\beta}_{c_z}$ [mm]	4.009	4.009	4.006
$\widehat{\beta}_r$ [mm]	68.943	68.943	68.939
$\widehat{\Psi}$	–	$\begin{bmatrix} 29.24 & -0.17 & 3.09 \\ -0.17 & 20.86 & 7.07 \\ 3.09 & 7.07 & 5.85 \end{bmatrix} 10^{-5}$	$\begin{bmatrix} 1.15 & 0.69 & 0.31 \\ 0.69 & 0.79 & 0.38 \\ 0.31 & 0.38 & 0.35 \end{bmatrix} 10^{-5}$
$\widehat{v}$	–	2.60	2.10
$\widehat{\Sigma}_{uu}$	$\begin{bmatrix} 60.64 & -3.24 & 3.42 \\ -3.24 & 57.09 & 19.55 \\ 3.42 & 19.55 & 14.29 \end{bmatrix} 10^{-5}$	$\begin{bmatrix} 61.15 & -3.27 & 3.22 \\ -3.27 & 57.06 & 19.63 \\ 3.22 & 19.63 & 14.24 \end{bmatrix} 10^{-5}$	$\begin{bmatrix} 4.21 & 3.14 & 1.31 \\ 3.14 & 3.98 & 1.89 \\ 1.31 & 1.89 & 1.54 \end{bmatrix} 10^{-5}$
$\widehat{R}$	$\begin{bmatrix} 1.00 & -0.06 & 0.12 \\ -0.06 & 1.00 & 0.68 \\ 0.12 & 0.68 & 1.00 \end{bmatrix}$	$\begin{bmatrix} 1.00 & -0.06 & 0.11 \\ -0.06 & 1.00 & 0.69 \\ 0.11 & 0.69 & 1.00 \end{bmatrix}$	$\begin{bmatrix} 1.00 & 0.77 & 0.51 \\ 0.77 & 1.00 & 0.76 \\ 0.51 & 0.76 & 1.00 \end{bmatrix}$

Before discussing the results of the parameter estimation, we will discuss the results of the portmanteau test in Fig. 10. The blue line indicate the critical value for the significance level $\alpha = 0.05$. The red dots show the test values of the weighted white noise, which are larger than the critical value. The green star represents the test value that is smaller than the critical value, which means that the test is passed. It can be seen that the test is passed for order $p = 3$, while for orders $p = 4 \ldots 7$ the test value is close to the critical value. The test for order $p = 0$ failed and is, due to its significantly larger numerical value, not shown in Fig. 10. Based on these portmanteau test results, we will focus in the further discussion on the results of the LS model and the VAR model orders $p = 0$ and $p = 3$.

4.3.2 Discussion of the results

The estimated parameters $\boldsymbol{\beta}$ for the LS model and the VAR model orders $p = 0$ and $p = 3$ shown in Table 4 are almost the same, with the maximum differences of $5\,\mu$m being in the range of the uncertainty of the laser tracker used. The small value for the df indicates that the residuals $\mathbf{u}_t$ are not normally distributed. Differences in the estimated parameters of the three models can be found in the scale matrix and the VCM of the residuals. $\widehat{\Sigma}_{uu}$ is equivalent to the classical estimate of the VCM of the estimated white noise $\mathbf{u}$. For the LS model, $\widehat{\Sigma}_{uu}$ is estimated by means of the estimated residuals $\mathbf{e}$. $\widehat{\Sigma}_{uu}$ for VAR model order $p = 3$ clearly shows, how the values of the VCM of the observations on the main diagonal are reduced by means of the VAR process. The fact that the matrices $\widehat{\Sigma}_{uu}$ for the LS model and the VAR model order $p = 0$ are almost identical can be seen in the residuals shown in Fig. 11.

Figure 11 depicts the estimated colored and white noise residuals obtained from laser tracker measurements for VAR model order $p = 3$. The blue line indicates the colored noise, which shows several significant peaks, which can be interpreted as potential outliers. These may stem from the imperfect surface of the sphere or even can be associated with the roughness of the surface. Furthermore, these significant peaks can result from the measurement procedure since for the data acquisition the RRR is moved individually by hand along the surface. If the contact to the surface is slightly lost, this will also produce peaks as can be seen in the blue line. Due to the orientation of the local laser tracker coordinate system, individual peaks in the y-component (middle part in Fig. 11) can be associated with an uplift of the RRR towards the laser tracker. The minor colored noise level of the z-component also results from the measurement setup, i. e. the orientation of the local laser tracker coordinate system, and shows the uncertainty differences for the angular and the distance measurement unit. Since the estimated parameters $\boldsymbol{\beta}$ for the LS model and the VAR model orders $p = 0$ and $p = 3$ are almost identical (see Table 4), the depicted colored noise is also nearly identical. For the special case of VAR model order $p = 0$ the residual estimates $\hat{\mathbf{e}}_t$ and $\hat{\mathbf{u}}_t$ are equal. This implies equality of $\widehat{\Sigma}_{uu}$ for the LS model and the VAR model order $p = 0$. Furthermore, the residuals in Fig. 11 show an equal distribution of positive and negative peaks (outliers) for the individual coordinate components. Due to the symmetric geometry of the sphere the positive and negative peaks compensate each other in the estimation of the parameters $\boldsymbol{\beta}$.

The red circles in Fig. 11 indicate white noise and the green dots indicate the weighted white noise, i. e. white
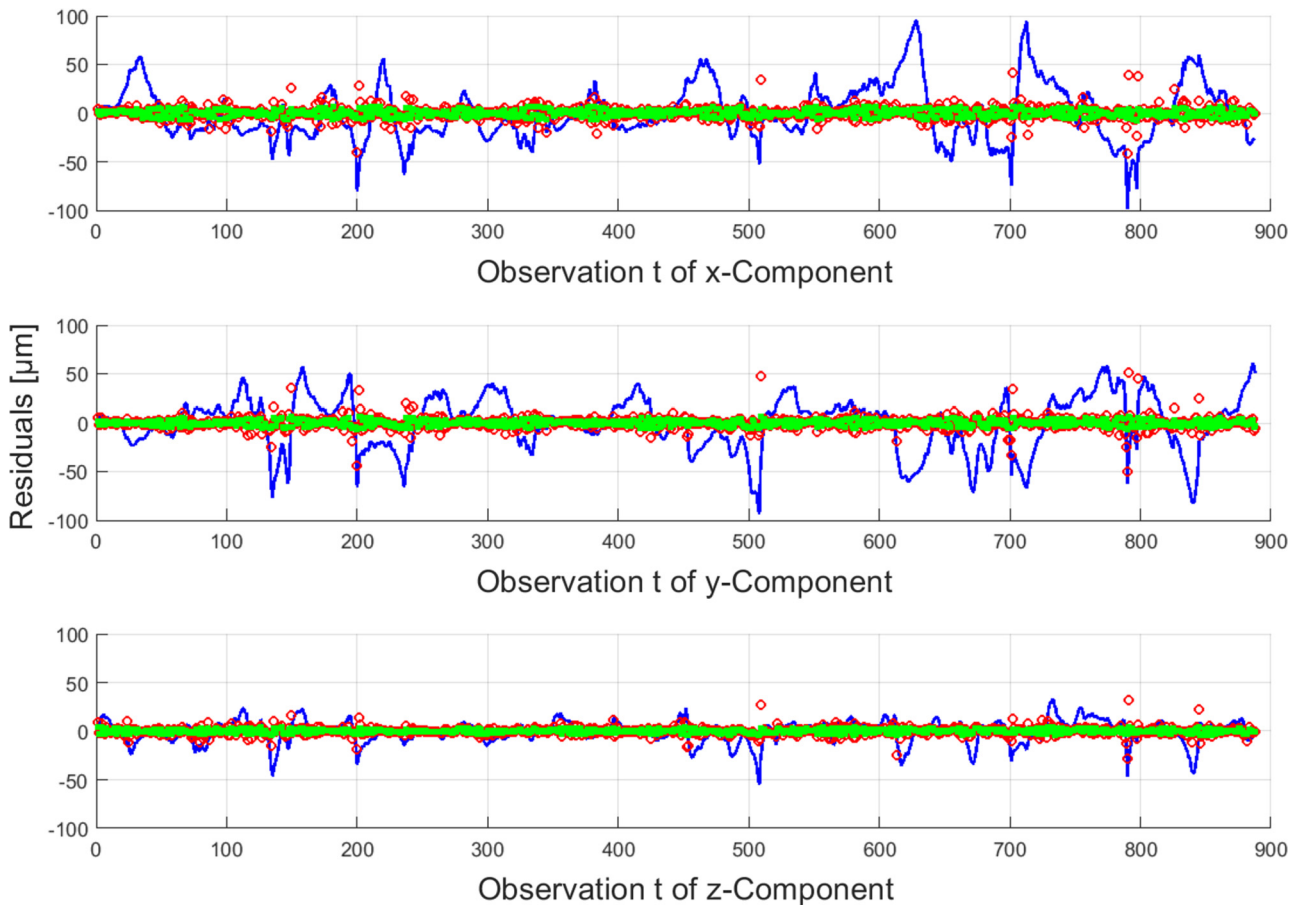
Figure 11: Estimated colored and white noise residuals obtained from laser tracker measurements for VAR model order $p = 3$. Blue line indicates colored noise, red circles indicate white noise and green dots indicate the weighted white noise.

noise values multiplied with their weights $\hat{\mathbf{w}}$. It can be clearly seen that the VAR process is unable to fully smooth the colored noise. First, the substitution of t-distribution instead of the normal distribution allow for reducing the influence of the remaining outliers in the parameter estimation. Furthermore, only by using the weighted white noise (colored in green) the portmanteau test can be passed, which is not applicable for the white noise (colored in red).

Table 5 shows the results of the VAR coefficients for the model orders $p = 1, 2, 3$. To judge about any possible correlations for the colored noise of the x-, y- and z-component, we can confirm correlations based on the shown VAR coefficients in Table 5. Since the off-diagonal elements are significantly different from zero, this indicates the presence of correlations. As an example, $\widehat{\mathbf{A}}_3$ shows similar magnitudes for the x-component with −0.263 and correlations with the y-component by a value of 0.203. Furthermore, the alternating positive and negative sign of the coefficients indicate the negative correlation for the coordinate components.

Taking a look at the VAR coefficients for the model order $p = 1$ one can see small off-diagonal elements, which shows that the correlation of the coordinate components is hardly reduced. This can also be seen by the significant difference of the test value in Fig. 10 for $p = 1$ with respect to the critical value. For the VAR model of order $p = 2$ it is firstly possible to estimate the correlations of the colored noise. On the one hand, in Table 5 the off-diagonal elements for $p = 2$ are obviously larger compared to $p = 1$. On the other hand, the test value is approaching the critical value, which can be seen in Fig. 10.

5 Conclusions and outlook

The nonlinear Gauss-Helmert model constitutes the most general adjustment model, which is widely used in geodesy. This model consists of condition equations (which link the observations and the functional parameters occurring in the mathematical model employed to approximate the measurements) and of a stochastic model.

Table 5: Results of the VAR coefficients for the VAR models of order $p = 1, 2, 3$.

	EM Model $p = 1$			EM Model $p = 2$			EM Model $p = 3$		
$\widehat{\mathbf{A}}_1$	0.986	0.006	0.001	1.412	−0.148	0.108	1.244	0.051	0.006
	−0.023	1.007	−0.021	−0.265	1.622	−0.132	−0.218	1.555	−0.149
	−0.009	0.007	0.982	−0.103	0.090	1.355	−0.118	0.108	1.284
$\widehat{\mathbf{A}}_2$	−			−0.427	0.153	−0.151	−0.002	−0.245	0.091
				0.262	−0.623	0.083	0.159	−0.439	0.125
				0.108	−0.078	−0.430	0.148	−0.114	−0.220
$\widehat{\mathbf{A}}_3$	−			−			−0.263	0.203	−0.159
							0.058	−0.115	−0.043
							−0.025	0.022	−0.164

The latter usually is defined by a VCM, which tends to be huge when the number of measurements is large. We therefore applied and investigated, in the context of multivariate time series, a more manageable type of stochastic model defined by a vector autoregressive (VAR) process. This model takes both auto- and cross-correlations into account and can be estimated alongside the functional model parameters. By employing a multivariate t-distribution with data-adaptable scale matrix and degree of freedom, moderate deviations from the usual assumption of normally distributed noise are possible. To fuse this t-distribution with the VAR process and the constraints, we applied the idea of formulating a generalized Gauss-Helmert model in terms of a likelihood function, which is maximized under the constraints. The proposed model allows for a computationally convenient iteratively reweighted least-squares method based on a constrained expectation maximization algorithm. This methodology extends the previously established approach involving univariate autoregressive models that ignored potential cross-correlations between the time series. Analysis of the biases by means of Monte Carlo simulations showed that the functional model parameters tend to be highly accurate when the number of observations is increased, whereas a significant bias of the degree of freedom of the multivariate t-distribution persists, in particular for smaller degrees of freedom. The accuracy of the estimated VAR coefficients and of the scale parameters generally improves with larger datasets. The methodology is also applicable to functional models that take the form of a Gauss-Markov model, as shown by an analysis of real accelerometer data modeled by an offset and a linear drift. For another real-world application, which utilizes continuous measurements of a laser tracker on a sphere, the applicability of the proposed Gauss-Helmert model was also demonstrated. Furthermore, the analyses of the re-

sults for this application demonstrated that the estimated VAR process coefficients and the estimated parameters of the multivariate t-distribution allows for detailed evaluations of the auto- and cross-correlation as well as the distributional characteristics of the measurement noise. Thus, actual deviations of the noise from expected white-noise and normal-distribution behavior become visible in the course of adjusting real-world datasets. Due to its flexibility the proposed innovative type of Gauss-Helmert model with VAR and t-distributed errors could be useful in various fields of geodesy such as engineering geodesy or Earth system modeling where multivariate, functionally complex and correlated time series are analyzed. Since such datasets often contain gaps a further extension of the model and EM algorithms such that the E-step yields also values for missing measurement values appears to be useful. It is also conceivable that the VAR processes of the current model are replaced by other stochastic processes that might match the correlation pattern of a given set of observables better.

Appendix A. Derivations

Log-likelihood and Q-functions

$$\log L(\boldsymbol{\theta}_{\mathrm{GHM}}; \boldsymbol{\ell}, \mathbf{w}) = -\frac{Nn}{2} \log(2\pi) - \frac{n}{2} \log \det \boldsymbol{\Psi}$$

$$+ \frac{n\nu}{2} \log \left(\frac{\nu}{2}\right) - n \log \Gamma \left(\frac{\nu}{2}\right) + \left(\frac{N}{2} - 1\right) \sum_{t=1}^{n} \log w_t$$

$$- \frac{1}{2} \sum_{t=1}^{n} w_t [\mathbf{A}(L)(\boldsymbol{\ell}_t - \boldsymbol{\mu}_t)]^T \boldsymbol{\Psi}^{-1} [\mathbf{A}(L)(\boldsymbol{\ell}_t - \boldsymbol{\mu}_t)]$$

$$+ \frac{\nu}{2} \sum_{t=1}^{n} (\log w_t - w_t). \tag{81}$$

$$Q(\boldsymbol{\theta}_{\text{GHM}}|\widehat{\boldsymbol{\theta}}_{\text{GHM}}^{(s)}) = E_{\mathcal{W}|\mathbf{u};\widehat{\boldsymbol{\theta}}_{\text{GHM}}^{(s)}} \{\log L(\boldsymbol{\theta}_{\text{GHM}}; \boldsymbol{\ell}, \mathcal{W})\}$$

$$= -\frac{Nn}{2}\log(2\pi) - \frac{n}{2}\log\det\boldsymbol{\Psi} + \frac{n\nu}{2}\log\left(\frac{\nu}{2}\right) - n\log\Gamma\left(\frac{\nu}{2}\right)$$

$$- \sum_{t=1}^{n} \frac{1}{2}\left(\nu + [\mathbf{A}(L)(\boldsymbol{\ell}_t - \boldsymbol{\mu}_t)]^T \boldsymbol{\Psi}^{-1}[\mathbf{A}(L)(\boldsymbol{\ell}_t - \boldsymbol{\mu}_t)]\right)$$

$$\times E_{\mathcal{W}|\mathbf{u};\widehat{\boldsymbol{\theta}}_{\text{GHM}}^{(s)}} \{\mathcal{W}_t\} + \sum_{t=1}^{n} \frac{1}{2}(\nu-1)E_{\mathcal{W}|\mathbf{u};\widehat{\boldsymbol{\theta}}_{\text{GHM}}^{(s)}} \{\log \mathcal{W}_t\}. \tag{82}$$

Here, the conditional expectations $E_{\mathcal{W}|\mathbf{u};\widehat{\boldsymbol{\theta}}_{\text{GHM}}^{(s)}} \{\mathcal{W}_t\}$ and $E_{\mathcal{W}|\mathbf{u};\widehat{\boldsymbol{\theta}}_{\text{GHM}}^{(s)}} \{\log \mathcal{W}_t\}$ are essentially the same as the ones occurring in Eq. (54) in [26] with respect to the GMM with VAR and multivariate-distributed errors. The only difference lies in the different parametrization of the deterministic model, which is independent of the stochastic model for $\mathcal{W}_t$. Consequently,

$$\widetilde{w}_t^{(s)} := E_{\mathcal{W}|\mathbf{u};\widehat{\boldsymbol{\theta}}_{\text{GHM}}^{(s)}} \{\mathcal{W}_t\} = \frac{\widehat{\nu}^{(s)} + N}{\widehat{\nu}^{(s)} + \mathbf{u}_t^T(\widehat{\boldsymbol{\Psi}}^{(s)})^{-1}\mathbf{u}_t} \tag{83}$$

and

$$E_{\mathcal{W}|\mathbf{u};\widehat{\boldsymbol{\theta}}_{\text{GHM}}^{(s)}} \{\log \mathcal{W}_t\} = \log \widetilde{w}_t^{(s)} + \psi\left(\frac{\widehat{\nu}^{(s)} + N}{2}\right)$$

$$- \log\left(\frac{\widehat{\nu}^{(s)} + N}{2}\right), \tag{84}$$

with $\mathbf{u}_t = \widehat{\mathbf{A}}^{(s)}(L)(\boldsymbol{\ell}_t - \boldsymbol{\mu}_t^{(s)})$, so that the Q-function (82) can be written as

$$Q(\boldsymbol{\theta}_{\text{GHM}}|\widehat{\boldsymbol{\theta}}_{\text{GHM}}^{(s)}) = -\frac{n}{2}\log\det\boldsymbol{\Psi}$$

$$- \frac{1}{2}\sum_{t=1}^{n} \widetilde{w}_t^{(s)}[\mathbf{A}(L)(\boldsymbol{\ell}_t - \boldsymbol{\mu}_t)]^T \boldsymbol{\Psi}^{-1}[\mathbf{A}(L)(\boldsymbol{\ell}_t - \boldsymbol{\mu}_t)]$$

$$+ \frac{n\nu}{2}\log\nu - n\log\Gamma\left(\frac{\nu}{2}\right) + \frac{n\nu}{2}\left[\psi\left(\frac{\widehat{\nu}^{(s)} + N}{2}\right)\right.$$

$$\left. - \log\left(\widehat{\nu}^{(s)} + N\right) + \frac{1}{n}\sum_{t=1}^{n}\left(\log\widetilde{w}_t^{(s)} - \widetilde{w}_t^{(s)}\right)\right]. \tag{85}$$

The following normal equations can be derived from the linearized Lagrangian (obtained by combining (40) with (42))

$$F(\boldsymbol{\phi}, \boldsymbol{\lambda}|\widehat{\boldsymbol{\theta}}_{\text{GHM}}^{(s)}) \approx -\frac{n}{2}\log\det\boldsymbol{\Psi}$$

$$- \frac{1}{2}\sum_{t=1}^{n} \widetilde{w}_t^{(s)}[\mathbf{A}(L)(\boldsymbol{\ell}_t - \boldsymbol{\mu}_t)]^T \boldsymbol{\Psi}^{-1}[\mathbf{A}(L)(\boldsymbol{\ell}_t - \boldsymbol{\mu}_t)]$$

$$+ \frac{n\nu}{2}\log\nu - n\log\Gamma\left(\frac{\nu}{2}\right) + \frac{n\nu}{2}\left[\psi\left(\frac{\widehat{\nu}^{(s)} + N}{2}\right)\right.$$

$$\left. - \log\left(\widehat{\nu}^{(s)} + N\right) + \frac{1}{n}\sum_{t=1}^{n}\left(\log\widetilde{w}_t^{(s)} - \widetilde{w}_t^{(s)}\right)\right]$$

$$+ \boldsymbol{\lambda}^T \left[\mathbf{X}\Delta\boldsymbol{\beta} + \mathbf{B}(\boldsymbol{\mu} - \boldsymbol{\ell}) + \overline{\mathbf{m}}\right] \tag{86}$$

by applying slight modifications to the steps concerning the GMM with VAR and multivariate t-distributed errors given in Appendix B in [26], using (8).

Normal equations for the location parameters

Applying (8) to eliminate the time dependence of the location parameters $\boldsymbol{\mu}_t$ allows one to form (and subsequently simplify) the partial derivatives with respect to the complete vector $\boldsymbol{\mu}$ of location parameters as follows.

$$\mathbf{0} = \frac{\partial F(\boldsymbol{\phi}, \boldsymbol{\lambda}|\widehat{\boldsymbol{\theta}}_{\text{GHM}}^{(s)})}{\partial\boldsymbol{\mu}}$$

$$= -\frac{1}{2}\sum_{t=1}^{n} \widetilde{w}_t^{(s)}\frac{\partial}{\partial\boldsymbol{\mu}}[\mathbf{A}(L)(\boldsymbol{\ell}_t - \boldsymbol{\mu}_t)]^T \boldsymbol{\Psi}^{-1}[\mathbf{A}(L)(\boldsymbol{\ell}_t - \boldsymbol{\mu}_t)] + \mathbf{B}^T\boldsymbol{\lambda}$$

$$= -\frac{1}{2}\sum_{t=1}^{n} \widetilde{w}_t^{(s)}\frac{\partial}{\partial\boldsymbol{\mu}}[\mathbf{A}(L)(\boldsymbol{\ell}_t - \mathbf{J}_t\boldsymbol{\mu})]^T \boldsymbol{\Psi}^{-1}[\mathbf{A}(L)(\boldsymbol{\ell}_t - \mathbf{J}_t\boldsymbol{\mu})] + \mathbf{B}^T\boldsymbol{\lambda}$$

$$= -\sum_{t=1}^{n} \widetilde{w}_t^{(s)}[\mathbf{A}(L)(\boldsymbol{\ell}_t - \mathbf{J}_t\boldsymbol{\mu})]\boldsymbol{\Psi}^{-1}\cdot\frac{\partial}{\partial\boldsymbol{\mu}}[-\mathbf{A}(L)(\mathbf{J}_t\boldsymbol{\mu})] + \mathbf{B}^T\boldsymbol{\lambda}$$

$$= \sum_{t=1}^{n}[\mathbf{A}(L)\mathbf{J}_t]^T\left[\widetilde{w}_t^{(s)}\boldsymbol{\Psi}^{-1}\right][\mathbf{A}(L)\boldsymbol{\ell}_t - \mathbf{A}(L)\mathbf{J}_t\boldsymbol{\mu}] + \mathbf{B}^T\boldsymbol{\lambda}$$

$$= \sum_{t=1}^{n}\bar{\mathbf{J}}_t^T\widetilde{w}_t^{(s)}\boldsymbol{\Psi}^{-1}(\bar{\boldsymbol{\ell}}_t - \bar{\mathbf{J}}_t\boldsymbol{\mu}) + \mathbf{B}^T\boldsymbol{\lambda}$$

Normal equations for the VAR coefficient matrices

Substituting (32) into (86) for brevity of expressions and applying also Eq. (88) in [59],

$$\mathbf{0}_N = \frac{\partial F(\boldsymbol{\phi}, \boldsymbol{\lambda}|\widehat{\boldsymbol{\theta}}_{\text{GHM}}^{(s)})}{\partial\mathbf{A}_j}$$

$$= -\frac{1}{2}\sum_{t=1}^{n} \widetilde{w}_t^{(s)}\frac{\partial}{\partial\mathbf{A}_j}[\mathbf{A}(L)\mathbf{e}_t]^T \boldsymbol{\Psi}^{-1}[\mathbf{A}(L)\mathbf{e}_t]$$

$$= -\frac{1}{2}\sum_{t=1}^{n} \widetilde{w}_t^{(s)}\frac{\partial}{\partial\mathbf{A}_j}[\mathbf{e}_t - \mathbf{A}_1\mathbf{e}_{t-1} - \cdots - \mathbf{A}_p\mathbf{e}_{t-p}]^T$$

$$\times \boldsymbol{\Psi}^{-1}[\mathbf{e}_t - \mathbf{A}_1\mathbf{e}_{t-1} - \cdots - \mathbf{A}_p\mathbf{e}_{t-p}]$$

$$= -\frac{1}{2}\sum_{t=1}^{n} \widetilde{w}_t^{(s)}\frac{\partial}{\partial\mathbf{A}_j}\left[\left(\mathbf{e}_t - \sum_{i=1,i\neq j}^{p}\mathbf{A}_i\mathbf{e}_{t-i}\right) - \mathbf{A}_j\mathbf{e}_{t-j}\right]^T$$

$$\times \boldsymbol{\Psi}^{-1}\left[\left(\mathbf{e}_t - \sum_{i=1,i\neq j}^{p}\mathbf{A}_i\mathbf{e}_{t-i}\right) - \mathbf{A}_j\mathbf{e}_{t-j}\right]$$

$$= -\frac{1}{2}\sum_{t=1}^{n} \widetilde{w}_t^{(s)}(-2)\cdot\boldsymbol{\Psi}^{-1}\left[\left(\mathbf{e}_t - \sum_{i=1,i\neq j}^{p}\mathbf{A}_i\mathbf{e}_{t-i}\right) - \mathbf{A}_j\mathbf{e}_{t-j}\right]\mathbf{e}_{t-j}^T$$

$$= \sum_{t=1}^{n} \widetilde{w}_t^{(s)}\boldsymbol{\Psi}^{-1}[\mathbf{A}(L)\mathbf{e}_t]\mathbf{e}_{t-j}^T$$

$$= \sum_{t=1}^{n} \widetilde{w}_t^{(s)}\mathbf{e}_{t-j}[\mathbf{A}(L)\mathbf{e}_t]^T(\boldsymbol{\Psi}^{-1})^T$$

$$= \sum_{t=1}^{n} \widetilde{w}_t^{(s)}\mathbf{e}_{t-j}[\mathbf{A}(L)\mathbf{e}_t]^T(\boldsymbol{\Psi}^{-1})^T\boldsymbol{\Psi}^T = \sum_{t=1}^{n} \widetilde{w}_t^{(s)}\mathbf{e}_{t-j}[\mathbf{A}(L)\mathbf{e}_t]^T$$

The joint normal equation system for all of the VAR coefficient matrices $\mathbf{A}_1, \ldots, \mathbf{A}_p$ then becomes

$$
\begin{bmatrix} \mathbf{0}_N \\ \vdots \\ \mathbf{0}_N \end{bmatrix} = \begin{bmatrix} \sum_{t=1}^{n} \widetilde{w}_t^{(s)} \mathbf{e}_{t-1}[\mathbf{A}(L)\mathbf{e}_t]^T \\ \vdots \\ \sum_{t=1}^{n} \widetilde{w}_t^{(s)} \mathbf{e}_{t-p}[\mathbf{A}(L)\mathbf{e}_t]^T \end{bmatrix}
$$

$$
= \begin{bmatrix} \widetilde{w}_1^{(s)} \mathbf{e}_{1-1} & \cdots & \widetilde{w}_n^{(s)} \mathbf{e}_{n-1} \\ \vdots & & \vdots \\ \widetilde{w}_1^{(s)} \mathbf{e}_{1-p} & \cdots & \widetilde{w}_n^{(s)} \mathbf{e}_{n-p} \end{bmatrix}
$$

$$
\times \begin{bmatrix} [\mathbf{e}_1 - \mathbf{A}_1\mathbf{e}_{1-1} - \cdots - \mathbf{A}_p\mathbf{e}_{1-p}]^T \\ \vdots \\ [\mathbf{e}_n - \mathbf{A}_1\mathbf{e}_{n-1} - \cdots - \mathbf{A}_p\mathbf{e}_{n-p}]^T \end{bmatrix}
$$

$$
= \begin{bmatrix} \mathbf{e}_0 & \cdots & \mathbf{e}_{n-1} \\ \vdots & & \vdots \\ \mathbf{e}_{1-p} & \cdots & \mathbf{e}_{n-p} \end{bmatrix} \widetilde{\mathbf{W}}^{(s)}
$$

$$
\times \left(\begin{bmatrix} \mathbf{e}_1^T \\ \vdots \\ \mathbf{e}_n^T \end{bmatrix} - \begin{bmatrix} \mathbf{e}_0^T & \cdots & \mathbf{e}_{1-p}^T \\ \vdots & & \vdots \\ \mathbf{e}_{n-1}^T & \cdots & \mathbf{e}_{n-p}^T \end{bmatrix} \begin{bmatrix} \mathbf{A}_1^T \\ \vdots \\ \mathbf{A}_p^T \end{bmatrix} \right)
$$

$$
= \mathbf{E}\widetilde{\mathbf{W}}^{(s)} \begin{bmatrix} \mathbf{e}_1^T \\ \vdots \\ \mathbf{e}_n^T \end{bmatrix} - \mathbf{E}\widetilde{\mathbf{W}}^{(s)}\mathbf{E}^T \begin{bmatrix} \mathbf{A}_1^T \\ \vdots \\ \mathbf{A}_p^T \end{bmatrix}.
$$

Normal equations for the inverse scale matrix

Substituting (33) into (86) for brevity of expressions and applying the arguments given in Sect. 373 in [18], one obtains

$$
\mathbf{0} = \frac{\partial F(\boldsymbol{\phi}, \boldsymbol{\lambda}|\widehat{\boldsymbol{\theta}}_{\text{GHM}}^{(s)})}{\partial \boldsymbol{\Psi}^{-1}}
$$

$$
= \frac{n}{2} \cdot \frac{\partial}{\partial \boldsymbol{\Psi}^{-1}}(-\log \det \boldsymbol{\Psi}) - \frac{1}{2}\sum_{t=1}^{n} \widetilde{w}_t^{(s)} \frac{\partial}{\partial \boldsymbol{\Psi}^{-1}} \mathbf{u}_t^T \boldsymbol{\Psi}^{-1}\mathbf{u}_t
$$

$$
= \frac{n}{2}\boldsymbol{\Psi} - \frac{1}{2}\sum_{t=1}^{n}\widetilde{w}_t^{(s)}\mathbf{u}_t\mathbf{u}_t^T
$$

$$
= \boldsymbol{\Psi} - \frac{1}{n}\sum_{t=1}^{n}\widetilde{w}_t^{(s)}[\mathbf{A}(L)(\boldsymbol{\ell}_t - \boldsymbol{\mu}_t)][\mathbf{A}(L)(\boldsymbol{\ell}_t - \boldsymbol{\mu}_t)]^T.
$$

References

[1]   Helmert FR. Die Ausgleichungsrechnung nach der Methode der kleinsten Quadrate. Leipzig, Teubner, 1872.

[2]   Wolf H. Das geodätische Gauß-Helmert-Modell und seine Eigenschaften. Z Vermessungswesen 1978, 103, 41–43.

[3]   Golub GH, Van Loan CH. An analysis of the total least squares problem. SIAM J Numer Anal 1980, 17, 883–893.

[4]   Schaffrin B, Wieser A. On weighted total least-squares adjustment for linear regression. J Geod 2008, 82, 415–421.

[5]   Neitzel F. Generalization of total least-squares on example of unweighted and weighted 2D similarity transformation. J Geod 2010, 84, 751–762.

[6]   Xu P, Liu J, Shi C. Total least squares adjustment in partial errors-in-variables models: algorithm and statistical analysis. J Geod 2012, 86, 661–675.

[7]   Amiri-Simkooei A, Jazaeri S. Weighted total least squares formulated by standard least squares theory. J Geodetic Sci 2012, 2, 113–124.

[8]   Gauß CF. Theoria Motus Corporum Coelestium. Hamburg, Perthes und Besser, 1809.

[9]   Gauß CF. Theoria Combinationis Observationum. Göttingen, Dieterich, 1823.

[10] Markov AA. Wahrscheinlichkeitsrechnung. Leipzig, Teubner, 1912.

[11] Aitken AC. On least squares and linear combinations of observations. Proc Royal Soc Edinburgh 1935, 55, 42–48.

[12] Amiri-Simkooei AR. Noise in multivariate GPS position time-series. J Geod 2009, 83, 175–187.

[13] Koch KR, Kuhlmann H, Schuh WD. Approximating covariance matrices estimated in multivariate models by estimated auto- and cross-covariances. J Geod 2010, 84, 383–397.

[14] Kermarrec G, Schön S. Fully populated VCM or the hidden parameter. J Geod Sci 2017, 7, 151–161.

[15] Kermarrec G, Schön S. A priori fully populated covariance matrices in least-squares adjustment—case study: GPS relative positioning. J Geod 2017, 91, 465–484.

[16] Kauker S, Schwieger V. A synthetic covariance matrix for monitoring by terrestrial laser scanning. J Appl Geod 2017, 11, 77–87.

[17] Kauker S, Holst C, Schwieger V, Kuhlmann H, Schön S. Spatio-temporal correlations of terrestrial laser scanning. AVN 2016, 123, 170–182.

[18] Koch KR. Parameter Estimation and hypothesis testing in linear models. Berlin, Springer, 1999.

[19] Kermarrec G, Schön S. Taking correlations in GPS least squares adjustments into account with a diagonal covariance matrix. J Geod 2016, 90, 793–805.

[20] Williams SDP. The effect of coloured noise on the uncertainties of rates estimated from geodetic time series. J Geod 2003, 76, 483–494.

[21] Schuh WD. The processing of band-limited measurements; filtering techniques in the least squares context and in the presence of data gaps. Space Sci Rev 2003, 108, 67–78.

[22] Klees R, Ditmar P, Broersen P. How to handle colored observation noise in large least-squares problems. J Geod 2003, 76, 629–640.

[23] Zeng W, Fang X, Lin Y, Huang X, Zhou Y. On the total least-squares estimation for autoregressive model. Surv Rev 2018, 50, 186–190.

[24] Brockwell PJ, Davis RA. Introduction to time series and forecasting. 3rd ed. Switzerland, Springer, 2016.

[25] McDonald JB. Partially adaptive estimation of ARMA time series models. Int J Forecasting 1989, 5, 217–230.

[26] Kargoll B, Kermarrec G, Korte J, Alkhatib H. Self-tuning robust adjustment within multivariate regression time series models with vector-autoregressive random errors. J Geod 2020, 94, 512020.

[27] Nduka UC, Ugah TE, Izunobi CH. Robust estimation using multivariate t innovations for vector autoregressive models via ECM algorithm. J Appl Stat 2021, 48, 693–711.

[28] Parzen E. A density-quantile function perspective on robust estimation. In: Launer L, Wilkinson GN (Eds.) Robustness in Statistics. Academic Press, 1979, 237–258.

[29] Huber PJ. Robust estimation of a location parameter. Ann Math Stat 1964, 35, 73–101.

[30] Dempster AP, Laird NM, Rubin DB. Maximum likelihood from incomplete data via the EM algorithm. J Royal Stat Soc B 1977, 39, 1–38.

[31] Takai K. Constrained EM algorithm with projection method. Comput Stat 2012, 27, 701–714.

[32] Koch KR. Robust estimations for the nonlinear Gauss Helmert model by the expectation maximization algorithm. J Geod 2014, 88, 263–271.

[33] Koch KR. Outlier detection for the nonlinear Gauss Helmert model with variance components by the expectation maximization algorithm. J Appl Geod 2014, 8, 185–194.

[34] Koch KR. Beispiele zur Parameterschätzung, zur Festlegung von Konfidenzregionen und zur Hypothesenprüfung. Mitteilungen aus den Geodätischen Instituten der Rheinischen Friedrich-Wilhelms-Universität, No. 87, Bonn, 2000.

[35] Ettlinger A, Neuner H. Assessment of inner reliability in the Gauss-Helmert model. J Appl Geod 2019, 14, 13–28.

[36] Rao CR. Linear statistical inference and its applications. New York, Wiley, 1973.

[37] Tienstra JM. An extension of the technique of the methods of least squares to correlated observations. Bull Geodesique 1947, 6, 301–335.

[38] Kuhlmann H. Importance of autocorrelation for parameter estimation in regression models. In: Proceedings of the 10th FIG International Symposium on Deformation Measurements, FIG 354–61.

[39] Baarda W. A testing procedure for use in geodetic networks. Publications on Geodesy (New Series), Vol. 2, No. 5. Delft, The Netherlands, Netherlands Geodetic Commission, 1968.

[40] Pope AJ. The statistics of residuals and the detection of outliers. NOAA Technical Report NOS65 NGS1. Rockville, Maryland, USA, US Department of Commerce, National Geodetic Survey, 1976.

[41] Lehmann R, Lösler M, Neitzel F. Mean-shift versus variance-inflation approach for outlier detection – a comparative study. Mathematics 2020, 8, 991.

[42] Priestley MB. Spectral analysis and time series. Academic Press, 1981.

[43] Lange KL, Little RJA, Taylor JMG. Robust statistical modeling using the t distribution. J Am Stat Assoc 1989, 84, 881–896.

[44] Liu Y, Sang R, Liu S. Diagnostic analysis for a vector autoregressive model under Student's t-distributions. Stat Neerl 2017, 71, 86–114.

[45] Hamilton JD. Time Series Analysis. Princeton University Press, 1994.

[46] Kargoll B, Omidalizarandi M, Alkhatib H. Adjustment of Gauss-Helmert Models with autoregressive and Student errors. In: Novák P, Crespi M, Sneeuw N, Sansò F (Eds.) IX Hotine-Marussi Symposium on Mathematical Geodesy. International Association of Geodesy Symposia, Vol. 151. Cham, Springer, 2020, 79–87.

[47] McLachlan GJ, Krishnan T. The EM algorithm and extensions. 2nd ed. Hoboken, John Wiley & Sons, 2008.

[48] Rubin DB. Iteratively reweighted least squares. In: Kotz S, Johnson NL, Read CB (Eds.) Encyclopedia of Statistical Sciences, Vol. 4, New York, Wiley, 1983, 272–275.

[49] Liu CH, Rubin DB. The ECME algorithm: a simple extension of EM and ECM with faster monotone convergence. Biometrika 1994, 81, 633–648.

[50] Liu CH, Rubin DB. ML estimation of the t distribution using EM and its extensions, ECM and ECME. Stat Sin 1995, 5, 19–39.

[51] Rump SM. INTLAB–interval laboratory. In: Csendes T (Ed.) Developments in Reliable Computing. Dordrecht, Kluwer Academic Publishers, 1999, 77–104.

[52] Neitzel F, Niemeier W, Weisbrich S. Investigation of low-cost accelerometer, terrestrial laser scanner and ground-based radar interferometer for vibration monitoring of bridges. In: Proceedings of the 6th European Workshop on Structural Health Monitoring, 2012, 542–551.

[53] Omidalizarandi M, Herrmann R, Marx S, Kargoll B, Paffenholz JA, Neumann I. A validated robust and automatic procedure for vibration analysis of bridge structures using MEMS accelerometers. J Appl Geod 2020, 14, 327–354.

[54] Omidalizarandi, M. Robust deformation monitoring of bridge structures using MEMS accelerometers and image-assisted total stations. München 2020. Deutsche Geodätische Kommission: C (PhD Dissertation): Heft Nr. 859. http://publikationen.badw.de/de/047037691.

[55] Omidalizarandi M, Neumann I, Kemkes E, Kargoll B, Diener D, Rüffer J, Paffenholz JA. MEMS based bridge monitoring supported by image-assisted total station. In: ISPRS –

International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XLII-4/W18, 2019, 833–842.

[56] Alkhatib H, Kargoll B, Paffenholz JA. 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, Cham, Springer, 2018, 25–38.

[57] Hosking JRM. The multivariate portmanteau statistic. J Am Stat Ass 1980, 75, 602–608.

[58] Hargreaves GI. Interval analysis in MATLAB. Numerical Analysis Report No. 416, Manchester Centre for Computational Mathematics, The University of Manchester, ISSN 1360-1725, 2002.

[59] Petersen KB, Pedersen MS. The matrix cookbook. Version November 15, 2012, Technical University of Denmark, https://www2.imm.dtu.dk/pubdb/edoc/imm3274.pdf. Accessed 22 February 2021.

[60] Hexagon Manufacturing Intelligence. Leica Absolute Tracker AT960 product brochure. Version 2017, https://www.hexagonmi.com/en-US/products/laser-tracker-systems/leica-absolute-tracker-at960. Accessed 20 April 2021.