ORIGINAL ARTICLE

Self-tuning robust adjustment within multivariate regression time series models with vector-autoregressive random errors

Boris Kargoll[1] · Gaël Kermarrec[2] · Johannes Korte[3] · Hamza Alkhatib[2]

Abstract

The iteratively reweighted least-squares approach to self-tuning robust adjustment of parameters in linear regression models with autoregressive (AR) and t-distributed random errors, previously established in Kargoll et al. (in J Geod 92(3):271–297, 2018. https://doi.org/10.1007/s00190-017-1062-6), is extended to multivariate approaches. Multivariate models are used to describe the behavior of multiple observables measured contemporaneously. The proposed approaches allow for the modeling of both auto- and cross-correlations through a vector-autoregressive (VAR) process, where the components of the white-noise input vector are modeled at every time instance either as stochastically independent t-distributed (herein called "stochastic model A") or as multivariate t-distributed random variables (herein called "stochastic model B"). Both stochastic models are complementary in the sense that the former allows for group-specific degrees of freedom (df) of the t-distributions (thus, sensor-component-specific tail or outlier characteristics) but not for correlations within each white-noise vector, whereas the latter allows for such correlations but not for different dfs. Within the observation equations, nonlinear (differentiable) regression models are generally allowed for. Two different generalized expectation maximization (GEM) algorithms are derived to estimate the regression model parameters jointly with the VAR coefficients, the variance components (in case of stochastic model A) or the cofactor matrix (for stochastic model B), and the df(s). To enable the validation of the fitted VAR model and the selection of the best model order, the multivariate portmanteau test and Akaike's information criterion are applied. The performance of the algorithms and of the white noise test is evaluated by means of Monte Carlo simulations. Furthermore, the suitability of one of the proposed models and the corresponding GEM algorithm is investigated within a case study involving the multivariate modeling and adjustment of time-series data at four GPS stations in the EUREF Permanent Network (EPN).

✉ Boris Kargoll
boris.kargoll@hs-anhalt.de

[1] Institute of Geoinformation and Surveying, Anhalt University of Applied Sciences, Seminarplatz 2a, 06846 Dessau-Roßlau, Germany

[2] Geodetic Institute, Gottfried Wilhelm Leibniz Universität Hannover, Nienburger Str. 1, 30167 Hannover, Germany

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

1 Introduction

As Tienstra (1947) noted, "in our observations correlation is present, always and everywhere." The importance of modeling auto-correlations in the context of regression models has been pointed out, for instance, by Kuhlmann (2001). In the context of time series measured by a sensor, auto-correlations can be expected, e.g., as a consequence of calibration corrections being applied to all of the measurements (cf. Lira and Wöger 2006). Due to their efficiency, flexibility and straightforward relationships with auto-covariance and spectral density functions, autoregressive (AR) models have been

used in diverse fields of geodetic science[1] (e.g., Xu 1988; Nassar et al. 2004; Amiri-Simkooei et al. 2007; Park and Gao 2008; Luo et al. 2011; Schuh et al. 2014; Didova et al. 2016; Zeng et al. 2018). In previous geodetic studies involving AR models, stochastic dependencies between different, contemporaneously observed times series, have usually been neglected. Thus, the common practice of multivariate geodetic time series analysis consists of a fusion of time series at the level of a functional (usually spatial regression) model, with separate modeling of the correlation pattern, say, through multiple independent AR models (cf. Alkhatib et al. 2018).

However, it is known for certain multivariate geodetic observables that stochastic dependencies between them indeed occur. For instance, spatial correlations between GPS position time series at different sites have been reported by Williams et al. (2004). Amiri-Simkooei (2009) suggested to model such correlations by means of a cross-covariance matrix involving adjustable variance–covariance components. Cross-covariance matrices have been estimated also by Koch et al. (2010) to analyze the correlations between coordinates triples of grid points measured by a terrestrial laser scanner. In analogy to the univariate case, multivariate ("vector") AR (VAR) models may be an efficient and flexible alternative to cross-covariance matrices or functions. However, to the best knowledge of the authors, VAR models have not been considered to describe the colored noise of multivariate geodetic observables that include a functional, spatial regression model. Instead, VAR processes have been applied in two main geodetic studies to model pure time series, which are not fused with a spatial model. In the first one, Niedzielski and Kosek (2012) used a VAR process for predictions of Universal Time (UT1-UTC) based on geodetic and geophysical data. In the second one, Zhang and Gui (2015) modeled double-differenced carrier phase observations by a VAR process to identify outliers. In climate research, VAR processes have been used frequently to investigate relationships between different datasets. Mosedale et al. (2006) fitted daily wintertime sea surface temperatures and North Atlantic Oscillation (NAO) time series by means of a bivariate VAR model. Strong et al. (2009) employed a VAR process to quantify feedback between the NAO and winter sea ice variability. Matthewman and Magnusdottir (2011) fitted weekly averaged observations of Bering Sea ice concentrations and the Western Pacific pattern (defined in Wallace and Gutzler 1981). Furthermore, Matthewman and Magnusdottir (2012) applied a VAR-based test of causality to relationships

between geopotential height anomalies at different locations. Recently, Papagiannopoulou et al. (2017) explored climate-vegetation dynamics in the framework of VAR processes.

To extend the applicability of VAR models to time series consisting more generally of a deterministic regression model and a VAR (correlation) model, it is necessary to more systematically study both the underlying theory and computationally convenient inferential procedures. In doing so, it is prudent to incorporate the possibility of outliers, which geodetic measurements are often susceptible to. Outlier-resistant ("robust") parameter estimation is one of the main strategies to accounting for outliers. In this category, iteratively least squares (IRLS) procedures have been used by geodesists for decades due to their conceptual simplicity and proven ability to reduce the effect of outliers by down-weighting outlier-afflicted observations. It should be mentioned that the robustness of these procedures is usually limited to adjustment problems which do not contain outliers that occur in leverage points (i.e., in poorly controlled observations characterized by very small partial redundancies) or as clusters. Such problems, which are not considered in the present paper, can be dealt with by employing a robust estimator with high breakdown point, e.g., the sign-constrained robust least squares method introduced by Xu (2005). Particularly popular robust IRLS procedures have been the L1-norm- and M-estimators (cf. Huber and Ronchetti 2009). The robustness of these estimators can be explained by the fact that random errors are modeled by means of a probability density function (pdf) $f(x)$ that decreases rather slowly (often as a power of x, i.e., $f(x) \sim c \cdot x^{-\alpha}$) as $|x|$ increases. Such a pdf is thus characterized by "heavy tails" (cf. Koch and Kargoll 2013; Alkhatib et al. 2017), which are reflected by high kurtosis (cf. Westfall 2014). Thus, outliers occur more likely with a heavy-tailed pdf than with a "light-tailed" pdf, such as a Gaussian pdf. Therefore, the former may be expected to define a more realistic stochastic model than the latter when numerous outliers are present. If outliers are thus modeled stochastically by means of a heavy-tailed pdf, then robust maximum likelihood (ML) estimators can be employed (cf. Wiśniewski 2014). Particularly attractive appear to be procedures that include the widely used method of least squares alongside more robust procedures. Since the true nature of the random error, including the stochastic outlier characteristics, is usually unknown, it makes sense to adjust the actual probability distribution of the random errors from the given measurements themselves, as already pointed out by Helmert (1907). One possibility for doing so is to adopt the family of scaled t-distributions (cf. Lange et al. 1989). The shape characteristics of the pdfs defining the members of this family are controlled by two parameters: The scale factor

---

[1] It should be mentioned that the more flexible autoregressive moving average (ARMA) models have also been employed frequently. However, the estimation of such models is more involved than the estimation of AR models (cf. Siemes 2013). As we found algorithms of the type Footnote1 continued

elaborated in the following papers to be unsuitable for this task due to lack of convergence, ARMA models are not considered here.

determines the spread, and the so-called degree of freedom[2] (*df*) is related to the tail behavior. These pdfs have a number of beneficial mathematical properties such as heavy tails (high kurtosis), so that t-distributions can be used for the purpose of robust estimation (cf. Schön et al. 2018; Alkhatib et al. 2017; Briegel and Tresp 1999; Lange et al. 1989). As this family contains the Gaussian distributions as limit cases (as the *df* tends to infinity), the former is merely a generalization of the latter. Thus, the often-made assumption of a Gaussian random error law is not really abandoned, but becomes a testable hypothesis in light of the fact that the *df* of the t-distribution can be estimated by means of ML estimation (as demonstrated in Sect. 3). Such an estimator that adapts the *df* to the data itself was called a *self-tuning robust estimator* by Parzen (1979). The usage of Student distributions with a relatively low *df* $\nu \in [3, 4]$ has also been recommended by advocates of the Guide to the Expression of Uncertainty in Measurement (ISO/IEC 2008) in situations, where input quantities with statistically determined ("type-A") standard uncertainties affect the output quantities (i.e., the observables to be adjusted) (cf. Sommer and Siebert 2004).

Recently, Liu et al. (2017) used the t-distribution in a VAR process added to a deterministic model consisting solely of the unknown observation offsets ("intercepts"). While the focus of that study was on influence diagnostics for model perturbations, one purpose of our current contribution is to demonstrate effective estimation algorithms in the more general situation where the deterministic model is described by either a linear or nonlinear regression. This is achieved by setting up an expectation maximization (EM) algorithm in the linear case and a generalized expectation maximization (GEM) algorithm in the nonlinear case. These are suitable approaches in view of their generally stable convergence and their capacity to reduce an ML estimation with random errors (based on the t-distribution) to a computationally simple IRLS method (cf. McLachlan and Krishnan 2008; Koch and Kargoll 2013). The expectation step ("E-step") corresponds to the computation of the weights in the current iteration, which are used to downweight outliers in the maximization step ("M-step") to solve for the unknown model parameters. With an EM algorithm in the context of a linear deterministic model, the likelihood function is maximized globally within the M-step. However, when the functional model is nonlinear and therefore linearized, this cannot be achieved. In this situation, a GEM algorithm may be applied instead to compute, after the usual E-step, a solution that increases the likelihood function, but which does not necessarily maximize it (cf. Phillips 2002). Our proposed algorithm extends the GEM algorithm for adjusting nonlinear, multivariate regres-

sion time series where the random errors of each component follow a univariate AR process with univariate t-distributed white-noise components (cf. Alkhatib et al. 2018). Algorithms for the simpler case of linear, univariate time series with AR and t-distributed random errors were independently investigated by Kargoll et al. (2018a), Tuac et al. (2018) and Nduka et al. (2018).

The new GEM algorithm dealing more generally with a VAR process (derived in Sect. 3) will be established in two variants for two different stochastic models (defined in Sect. 2): The first stochastic model ("A") is based on the assumption that the different white-noise series of the VAR process have independent, univariate t-distributions whose scale factors and *df*s are time-constant but possibly different for all series. The second stochastic model ("B") is defined by a multivariate t-distribution, which has only a single *df* applying to all series jointly, besides a cofactor matrix. Since these GEM algorithms maximize the original (log-)likelihood function, information criteria for model selection are readily available. In particular, a bias-corrected version of the Akaike information criterion (AIC) is defined. As a tool specifically for selecting an adequate VAR model order, the multivariate portmanteau test by Hosking (1980) is adapted to the current observation models. The convergence behavior of the suggested GEM algorithms is investigated through Monte Carlo (MC) simulations, the results of which are presented and analyzed in Sect. 4. In Sect. 5, various Global Positioning System (GPS) time series from the EUREF Permanent Network (EPN) are modeled and adjusted by means of the proposed methodology. In this numerical example, the issue of model selection is also discussed.

2 The observation models

Our goal is to adjust N groups $\mathcal{L}_{1:}, \ldots, \mathcal{L}_{N:}$ of observables, where each group $k \in \{1, \ldots, N\}$ consists of n observables $\mathcal{L}_{k:} = [\mathcal{L}_{k,1} \ \ldots \ \mathcal{L}_{k,n}]^T$, defined to be a regression time series

$$\mathcal{L}_{k,t} = h_{k,t}(\boldsymbol{\xi}) + \mathcal{E}_{k,t} \qquad (t = 1, \ldots, n). \tag{1}$$

These equations specify then a vector time series[3] $(\mathcal{L}_t \mid t \in \{1, \ldots, n\})$ with $\mathcal{L}_t = [\mathcal{L}_{1,t} \ \ldots \ \mathcal{L}_{N,t}]^T$. The measurements $\ell_{1,1}, \ldots, \ell_{N,n}$ are used to form the total observation

---

[2] The degree of freedom as a parameter of t-distributions is not to be confused with the notion of degree of freedom as the redundancy of an adjustment model.

[3] Colons occurring, e.g., in the subscript of $\mathcal{L}_{k:}$, are used to distinguish quantities collected throughout time within a single group k from quantities, such as $\mathcal{L}_t$, collected within the various groups at a single time instance t. Unknown parameters are mostly denoted by Greek letters, random variables by calligraphic letters, and constants by Roman letters. Thus, a random variable (e.g., $\mathcal{E}_t$) and its realization (e_t) are distinguished. Furthermore, matrices and vectors are represented by bold letters.

vector $\boldsymbol{\ell}$. The observation model on the right-hand side of (1) is defined by a deterministic model, through possibly nonlinear (differentiable) functions, $h_{k,t}$ of $m < n$ unknown parameters $\boldsymbol{\xi} = [\xi_1, \ldots, \xi_m]^T$, with random errors $\boldsymbol{\mathcal{E}}_{k:} = [\mathcal{E}_{k,1} \ \ldots \ \mathcal{E}_{k,n}]^T$ forming a vector time series with terms $\boldsymbol{\mathcal{E}}_t = [\mathcal{E}_{1,t} \ \ldots \ \mathcal{E}_{N,t}]^T$. Similarly, the vector-valued function $\mathbf{h}_t$ can be defined from the scalar functions $h_{1,t}, \ldots, h_{N,t}$, so that one can write

$$\boldsymbol{\mathcal{L}}_t = \mathbf{h}_t(\boldsymbol{\xi}) + \boldsymbol{\mathcal{E}}_t. \tag{2}$$

Clearly, observation equations (1) can also be stated in the form of "random error equations"

$$\mathcal{E}_{k,t} = \mathcal{L}_{k,t} - h_{k,t}(\boldsymbol{\xi}) \quad (t = 1, \ldots, n) \tag{3}$$
or in vector form

$$\boldsymbol{\mathcal{E}}_t = \boldsymbol{\mathcal{L}}_t - \mathbf{h}_t(\boldsymbol{\xi}). \tag{4}$$

Numerical realizations $e_{k,t}$ of the random errors $\mathcal{E}_{k,t}$ are also referred to as "residuals" below. To allow for both auto-correlations within each time series and cross-correlations between the time series, the random errors $\boldsymbol{\mathcal{E}}_t$ are modeled jointly as a p-th order VAR process $\boldsymbol{\mathcal{E}}_t = \mathbf{A}^{(1)}\boldsymbol{\mathcal{E}}_{t-1} + \cdots + \mathbf{A}^{(p)}\boldsymbol{\mathcal{E}}_{t-p} + \boldsymbol{\mathcal{U}}_t$ with components

$$\begin{bmatrix} \mathcal{E}_{1,t} \\ \vdots \\ \mathcal{E}_{N,t} \end{bmatrix} = \begin{bmatrix} \alpha_{1;1,1} & \cdots & \alpha_{1;1,N} \\ \vdots & \ddots & \vdots \\ \alpha_{1;N,1} & \cdots & \alpha_{1;N,N} \end{bmatrix} \begin{bmatrix} \mathcal{E}_{1,t-1} \\ \vdots \\ \mathcal{E}_{N,t-1} \end{bmatrix} + \cdots$$
$$+ \begin{bmatrix} \alpha_{p;1,1} & \cdots & \alpha_{p;1,N} \\ \vdots & \ddots & \vdots \\ \alpha_{p;N,1} & \cdots & \alpha_{p;N,N} \end{bmatrix} \begin{bmatrix} \mathcal{E}_{1,t-p} \\ \vdots \\ \mathcal{E}_{N,t-p} \end{bmatrix} + \begin{bmatrix} \mathcal{U}_{1,t} \\ \vdots \\ \mathcal{U}_{N,t} \end{bmatrix} \tag{5}$$

Here, the entries $\alpha_{1;1,1}, \ldots, \alpha_{p;N,N}$ of the N-by-N matrices $\mathbf{A}^{(1)}, \ldots, \mathbf{A}^{(p)}$ are unknown parameters and $\mathcal{U}_{1,1}, \ldots, \mathcal{U}_{N,n}$ are, at least within every group, stochastically independent random variables, having zero mean and group-specific variances $\sigma_{k,0}^2$. Since a computationally convenient method of parameter estimation is desirable, the random errors $\boldsymbol{\mathcal{E}}_0, \ldots, \boldsymbol{\mathcal{E}}_{1-p}$ are all assumed to take the value 0. The conditions for asymptotic covariance-stationarity of VAR(p) model (5) can be found, e.g., in Hamilton (1994, p. 259). We now rewrite process equations (5) as

$$\boldsymbol{\mathcal{U}}_t = \boldsymbol{\mathcal{E}}_t - \mathbf{A}^{(1)}\boldsymbol{\mathcal{E}}_{t-1} - \cdots - \mathbf{A}^{(p)}\boldsymbol{\mathcal{E}}_{t-p} \quad (t = 1, \ldots, n). \tag{6}$$

Denoting by $\mathbf{A}_k^{(j)}$ the kth row of the matrix $\mathbf{A}^{(j)}$ with $j \in \{1, \ldots, p\}$ and by $\mathbf{i}_k^T$ the kth row of the N-dimensional identity matrix $\mathbf{I}$, one may write for the kth component of $\boldsymbol{\mathcal{U}}_t$

$$\mathcal{U}_{k,t} = \mathbf{i}_k^T \boldsymbol{\mathcal{E}}_t - \mathbf{A}_k^{(1)}\boldsymbol{\mathcal{E}}_{t-1} - \cdots - \mathbf{A}_k^{(p)}\boldsymbol{\mathcal{E}}_{t-p}. \tag{7}$$

It is also assumed that the random variables $\mathcal{L}_0, \ldots, \mathcal{L}_{1-p}$ and the functions $h_{k,0}, \ldots, h_{k,1-p}$ occurring after substitution of lagged versions of (3) into (7) take values of 0. It will be convenient to collect all VAR coefficients contained in the kth rows $\mathbf{A}_k^{(1)}, \ldots, \mathbf{A}_k^{(p)}$ within the single ($pN \times 1$)-vector

$$\boldsymbol{\alpha}_{k:} = [\mathbf{A}_k^{(1)} \ \ldots \ \mathbf{A}_k^{(p)}]^T, \tag{8}$$

and to stack then all of these column vectors within the single ($pN^2 \times 1$)-vector

$$\boldsymbol{\alpha} = [\boldsymbol{\alpha}_{1:}^T \ \ldots \ \boldsymbol{\alpha}_{N:}^T]^T. \tag{9}$$

To set up computationally efficient filter schemes, it is useful to introduce the notation $L\mathbf{Z}_t := \mathbf{Z}_{t-1}$ and more generally $L^j\mathbf{Z}_t := \mathbf{Z}_{t-j}$, where L is the so-called lag operator that shifts the time index t by $j \in \{1, 2, \ldots\}$ into the past (see Chap. 2 in Hamilton 1994, for details). As shown in Brockwell and Davis (2016, p. 243) this lag operator can be used to define so-called lag polynomials

$$\mathbf{A}(L) = \mathbf{I} - \mathbf{A}^{(1)}L - \cdots - \mathbf{A}^{(p)}L^p, \tag{10}$$
$$\mathbf{A}_k(L) = \mathbf{i}_k^T - \mathbf{A}_k^{(1)}L - \cdots - \mathbf{A}_k^{(p)}L^p, \tag{11}$$

which may then be multiplied with some element $\mathbf{Z}_t$ of a given time series to obtain filtered quantities

$$\overline{\mathbf{Z}}_t = \mathbf{A}(L)\mathbf{Z}_t \tag{12}$$
$$= \mathbf{I}\mathbf{Z}_t - \mathbf{A}^{(1)}L\mathbf{Z}_t - \cdots - \mathbf{A}^{(p)}L^p\mathbf{Z}_t$$
$$= \mathbf{Z}_t - \mathbf{A}^{(1)}\mathbf{Z}_{t-1} - \cdots - \mathbf{A}^{(p)}\mathbf{Z}_{t-p} \tag{13}$$

and

$$\overline{Z}_{k,t} = \mathbf{A}_k(L)\mathbf{Z}_t \tag{14}$$
$$= \mathbf{i}_k^T \mathbf{Z}_t - \mathbf{A}_k^{(1)}L\mathbf{Z}_t - \cdots - \mathbf{A}_k^{(p)}L^p\mathbf{Z}_t$$
$$= \mathbf{i}_k^T \mathbf{Z}_t - \mathbf{A}_k^{(1)}\mathbf{Z}_{t-1} - \cdots - \mathbf{A}_k^{(p)}\mathbf{Z}_{t-p}. \tag{15}$$

Thus, (6) and (7) become with (4)

$$\boldsymbol{\mathcal{U}}_t = \overline{\boldsymbol{\mathcal{E}}}_t = \mathbf{A}(L)\boldsymbol{\mathcal{E}}_t = \mathbf{A}(L)(\boldsymbol{\mathcal{L}}_t - \mathbf{h}_t(\boldsymbol{\xi})), \tag{16}$$
$$\mathcal{U}_{k,t} = \overline{\mathcal{E}}_{k,t} = \mathbf{A}_k(L)\boldsymbol{\mathcal{E}}_t = \mathbf{A}_k(L)(\boldsymbol{\mathcal{L}}_t - \mathbf{h}_t(\boldsymbol{\xi})). \tag{17}$$

Here, $\mathbf{A}(L)$ and the collection of $\mathbf{A}_1(L), \ldots, \mathbf{A}_N(L)$ may be viewed as digital filters, which transform the auto- and cross-correlated ("colored-noise") time series ($\boldsymbol{\mathcal{E}}_t \,|\, t \in \{1, \ldots, n\}$) into the "white-noise" time series ($\boldsymbol{\mathcal{U}}_t \,|\, t \in \{1, \ldots, n\}$). The white noise components $\boldsymbol{\mathcal{U}}_t$ and $\mathcal{U}_{k,t}$ in (16)–(17) simplify to the right-hand sides of (3)–(4) if the VAR(p) process is partially "switched off" by setting $\mathbf{A}(L) = \mathbf{I}$ and $\mathbf{A}_k(L) = \mathbf{i}_k^T$, respectively. Numerical values $u_{k,t}$ of the white-noise variables $\mathcal{U}_{k,t}$ will also be called "white-noise residuals" in

the sequel. Having defined the functional and the correlation model, the stochastic model is specified next, for which at least two structurally different alternatives are conceivable.

2.1 Stochastic model A: group-specific univariate t-distributions

To allow for flexible modeling of random errors, which may involve heavy tailedness or multiple outliers of unknown forms, as well as different measurement precisions across the various observation groups, it may be assumed that the white-noise variables ($\mathcal{U}_{k,t}, k = 1, \ldots, N; t = 1, \ldots, n$) are all independently t-distributed with mean 0, group-specific (generally unknown) df ν_k and group-specific (generally unknown) variance component σ_k^2, symbolically,

$$\mathcal{U}_{k,t} \mid \boldsymbol{\theta} \overset{\text{ind.}}{\sim} t_{\nu_k}(0, \sigma_k^2) \quad (t = 1, \ldots, n). \quad (18)$$

Here, $\boldsymbol{\theta} = [\boldsymbol{\xi}^T \ \boldsymbol{\alpha}^T \ (\boldsymbol{\sigma^2})^T \ \boldsymbol{\nu}^T]^T$ is the complete parameter vector consisting of the ($m \times 1$)-vector $\boldsymbol{\xi}$ of functional parameters, the ($pN^2 \times 1$)-vector $\boldsymbol{\alpha}$ of VAR coefficients (8)–(9), the ($N \times 1$)-vector $\boldsymbol{\sigma^2} = [\sigma_1^2 \ \ldots \ \sigma_N^2]^T$ of variance components and the ($N \times 1$)-vector $\boldsymbol{\nu} = [\nu_1 \ \ldots \ \nu_N]^T$ of dfs. The estimation of the model parameters $\boldsymbol{\theta}$ is explained in Sect. 3. Note that the widely used assumption of normally distributed random errors is approximated by this model as $\nu_k \to \infty$. Practically, a close approximation of a normal distribution is reached already for $\nu_k = 120$ (cf. Koch 2017). Note also that the white-noise variance $\sigma_{k,0}^2$ is linked to the variance component σ_k^2 via the relationship $\sigma_{k,0}^2 = \frac{\nu_k}{\nu_k - 2} \cdot \sigma_k^2$, which requires that $\nu_k > 2$.

Due to the assumed stochastic independence of the white-noise variables, the joint pdf of $\mathcal{U}$ factors into

$$f_A(\mathbf{u}|\boldsymbol{\theta}) = f(u_{1,1}, \ldots, u_{N,n}|\boldsymbol{\theta})$$
$$= \prod_{k=1}^{N} \prod_{t=1}^{n} \frac{\Gamma\left(\frac{\nu_k+1}{2}\right)}{\sqrt{\nu_k \pi} \, \sigma_k \, \Gamma\left(\frac{\nu_k}{2}\right)} \left[1 + \left(\frac{u_{k,t}}{\sigma_k}\right)^2 / \nu_k\right]^{-\frac{\nu_k+1}{2}}, \quad (19)$$

where Γ is the gamma function and where the realizations $u_{k,t}$ of $\mathcal{U}_{k,t}$ carry the complete information of the functional and the correlation model via (17). This extends, in a natural way, the pdf considered in Alkhatib et al. (2018) from a multivariate regression time series with independent, univariate autoregressive random errors to one with random error following a fully multivariate VAR process. The log-likelihood function based on this pdf, as our basic observation model, is then given by

$$\log L_A(\boldsymbol{\theta}; \boldsymbol{\ell}) = \log f(\mathbf{u}|\boldsymbol{\theta}) = \sum_{k=1}^{N} \left(n \log\left[\frac{\Gamma\left(\frac{\nu_k+1}{2}\right)}{\sqrt{\nu_k \pi} \, \sigma_k \, \Gamma\left(\frac{\nu_k}{2}\right)}\right]\right.$$

$$\left. -\frac{\nu_k+1}{2} \cdot \sum_{t=1}^{n} \log\left[1 + \left(\frac{\mathbf{A}_k(L)(\boldsymbol{\ell}_t - \mathbf{h}_t(\boldsymbol{\xi}))}{\sigma_k}\right)^2 / \nu_k\right]\right). \quad (20)$$

Note that the vector-filtering operation $\mathbf{A}_k(L)(\boldsymbol{\ell}_t - \mathbf{h}_t(\boldsymbol{\xi}))$ in (20) simplifies to a scalar-filtering operation $\boldsymbol{\alpha}_k(L)(\ell_{k,t} - h_{k,t}(\boldsymbol{\xi}))$ used in the log-likelihood function involving multiple independent AR processes (see Alkhatib et al. 2018, Eq. (7)). As mentioned before, these filters make use of fixed boundary conditions for the unobservable past $\mathcal{L}_0, \ldots, \mathcal{L}_{-p}$, so that (20) constitutes a *conditional* log-likelihood function (see also Hamilton 1994, p. 291).

Before deriving the expectation maximization (EM) algorithm in analogy to the algorithms in Kargoll et al. (2018a) and Alkhatib et al. (2018), the Student distribution model is recast as a conditionally Gaussian variance inflation model. For this purpose, stochastically independent, gamma-distributed random variables

$$\mathcal{W}_{k,t} \mid \boldsymbol{\theta} \overset{\text{ind.}}{\sim} G\left(\frac{\nu_k}{2}, \frac{\nu_k}{2}\right) \quad (k = 1, \ldots, N; t = 1, \ldots, n) \quad (21)$$

are introduced, which are viewed as unobservable ("latent") variables or "missing data." Under the conditions that the random variable $\mathcal{W}_{k,t}$ takes the value $w_{k,t}$ and that the parameter values $\boldsymbol{\theta}$ are also given, the white-noise components are assumed to be Gaussian random variables

$$\mathcal{U}_{k,t} \mid w_{k,t}, \boldsymbol{\theta} \overset{\text{ind.}}{\sim} N(0, \sigma_k^2/w_{k,t}). \quad (22)$$

Thus, the value $w_{k,t}$ that the distribution of the white noise component $\mathcal{U}_{k,t}$ is conditional on plays the role of an additional weight, a decrease of which causes an increase ("inflation") of the variance of $\mathcal{U}_{k,t}$. Since these weights cannot be observed directly, it will be the essential idea of the GEM algorithms developed in Sect. 3 to determine their values as conditional expectations of the latent variables $\mathcal{W}_{k,t}$ given the observations $\ell_{k,t}$ and an estimate of the parameters $\boldsymbol{\theta}$. The conditional independence in (22) means that each random variable $\mathcal{U}_{k,t}$ is independent of the white-noise components $\mathcal{U}_{\kappa,\tau}$ and latent variables $\mathcal{W}_{\kappa,\tau}$ occurring within the series k at the other time instances (i.e., for $\kappa = k$ and $\tau = 1, \ldots, t-1, t+1, \ldots, n$) and within the other series at all time instances (i.e., for $\kappa = 1, \ldots, k-1, k+1, \ldots, N$ and $\tau = 1, \ldots, n$), conditional on the value $w_{k,t}$. This crucial assumption enables the simplification

$$f(u_{k,t}|u_{k,1}, w_{k,1} \ldots, u_{k,t-1}, w_{k,t-1}, u_{k,t+1}, w_{k,t+1}, \ldots, u_{k,n}, w_{k,n}$$
$$, w_{k,t}, \boldsymbol{\theta}) = f(u_{k,t}|w_{k,t}, \boldsymbol{\theta}). \quad (23)$$

Since assumptions (21)–(22) are exactly the same as the ones applied by Alkhatib et al. (2018) in connection with group-

specific (univariate) AR processes, their derivation of the log-likelihood function can be applied to the current model involving a VAR process with only one modification: $u_{k,t}$ is not replaced by the quantity $\boldsymbol{\alpha}_k(L)(\ell_{k,t} - \mathbf{h}_{k,t}(\boldsymbol{\xi}))$ using decorrelation filters $\boldsymbol{\alpha}_k(L)$ based on group-specific AR processes, but by the quantity $\mathbf{A}_k(L)(\ell_t - \mathbf{h}_t(\boldsymbol{\xi}))$ where the decorrelation filter $\mathbf{A}_k(L)$ is defined by a VAR process. With this modification, the log-likelihood function in (20) becomes

$$
\begin{aligned}
\log L_A(\boldsymbol{\theta}; \boldsymbol{\ell}, \mathbf{w}) = {} & \frac{n}{2} \log(2\pi) - \frac{n}{2} \sum_{k=1}^{N} \log(\sigma_k^2) \\
& + \frac{n}{2} \sum_{k=1}^{N} \nu_k \log\left(\frac{\nu_k}{2}\right) - n \sum_{k=1}^{N} \log\left(\Gamma\left(\frac{\nu_k}{2}\right)\right) \\
& - \sum_{k=1}^{N} \sum_{t=1}^{n} \frac{w_{k,t}}{2}\left[\nu_k + \left(\frac{\mathbf{A}_k(L)(\ell_t - \mathbf{h}_t(\boldsymbol{\xi}))}{\sigma_k}\right)^2\right] \\
& + \sum_{k=1}^{N} \sum_{t=1}^{n} \frac{1}{2}(\nu_k - 1) \log(w_{k,t}),
\end{aligned}
\tag{24}
$$

A certain disadvantage of this t-distribution model is that it does not allow for the modeling of stochastic dependence between the white-noise variables $\mathcal{U}_{1,t}, \ldots, \mathcal{U}_{N,t}$ at a given time instance t, so that the stochastic dependence between the colored-noise variables $\mathcal{E}_{1,t}, \ldots, \mathcal{E}_{N,t}$ arises merely from the past time instances $t-1, t-2, \ldots$, as shown by (5). The following type of stochastic model overcomes this limitation.

2.2 Stochastic model B: multivariate t-distribution

The second stochastic model uses a multivariate scaled t-distribution (cf. Lange et al. 1989) in order to model correlations between the white-noise variables $\mathcal{U}_{1,t}, \ldots, \mathcal{U}_{N,t}$. Instead of (18), the stochastic model assumption now is

$$
\mathcal{U}_t \mid \boldsymbol{\theta} \overset{\text{ind.}}{\sim} t_\nu(\mathbf{0}, \boldsymbol{\Sigma}) \qquad (t = 1, \ldots, n),
\tag{25}
$$

where $\boldsymbol{\Sigma}$ denotes an N-by-N cofactor matrix, being a factor of the covariance matrix $\boldsymbol{\Sigma}_0 = \frac{\nu}{\nu-2} \cdot \boldsymbol{\Sigma}$ of each random vector $\mathcal{U}_t$ for $\nu > 2$. All unknown entries of the inverse matrix $\boldsymbol{\Sigma}^{-1}$ are treated as parameters to be estimated. Instead of this variance–covariance model, a variance-component model as described in Koch (2014) could also be here. When the covariance matrix $\boldsymbol{\Sigma}_0$ of the random errors is given *a priori*, then the cofactor matrix $\boldsymbol{\Sigma}$ is computable via the equation $\boldsymbol{\Sigma} = \frac{\nu-2}{\nu} \cdot \boldsymbol{\Sigma}_0$. In certain applications, such a covariance matrix might depend on time (e.g., analysis of GNSS time series preprocessed by means of standard software); one would then simply add the time index to the corresponding cofactor matrix ($\boldsymbol{\Sigma}_t$) in (25) and in the sequel.

The total parameter vector reads $\boldsymbol{\theta} = [\boldsymbol{\xi}^T \ \boldsymbol{\alpha}^T \ (\boldsymbol{\sigma}^{-1})^T \ \nu]^T$, where $\boldsymbol{\sigma}^{-1}$ symbolizes the vectorized matrix $\boldsymbol{\Sigma}^{-1}$, that is,

the $(N^2 \times 1)$-vector obtained by stacking the columns of the matrix $\boldsymbol{\Sigma}^{-1}$. A certain drawback of the current model is that all observation groups share the same *df* ν and thus the same tail or outlier characteristics, which might be an unrealistic assumption in some geodetic applications.

Due to the assumed stochastic independence of the random vectors $\mathcal{U}_1, \ldots, \mathcal{U}_n$, their joint pdf factors into

$$
\begin{aligned}
f_B(\mathbf{u}|\boldsymbol{\theta}) &= \prod_{t=1}^{n} f_B(\mathbf{u}_t|\boldsymbol{\theta}) \\
&= \prod_{t=1}^{n} \frac{\Gamma\left(\frac{\nu+N}{2}\right)}{(\sqrt{\nu\pi})^N \sqrt{\det \boldsymbol{\Sigma}} \, \Gamma\left(\frac{\nu}{2}\right)} \left(1 + \frac{\mathbf{u}_t^T \boldsymbol{\Sigma}^{-1} \mathbf{u}_t}{\nu}\right)^{-\frac{\nu+N}{2}}
\end{aligned}
\tag{26}
$$

and leads to the log-likelihood function

$$
\begin{aligned}
\log L_B(\boldsymbol{\theta}; \boldsymbol{\ell}) = {} & n \log\left(\frac{\Gamma\left(\frac{\nu+N}{2}\right)}{(\sqrt{\nu\pi})^N \sqrt{\det \boldsymbol{\Sigma}} \, \Gamma\left(\frac{\nu}{2}\right)}\right) \\
& - \frac{\nu+N}{2} \cdot \sum_{t=1}^{n} \log\left(1 + \frac{\mathbf{u}_t^T \boldsymbol{\Sigma}^{-1} \mathbf{u}_t}{\nu}\right),
\end{aligned}
\tag{27}
$$

with $\mathbf{u}_t = \mathbf{A}(L)(\ell_t - \mathbf{h}_t(\boldsymbol{\xi}))$ according to (16). The variance inflation model corresponding to this multivariate Student random error model is now defined by

$$
\mathcal{W}_t \mid \boldsymbol{\theta} \overset{\text{ind.}}{\sim} G\left(\frac{\nu}{2}, \frac{\nu}{2}\right) \qquad (t = 1, \ldots, n)
\tag{28}
$$

and

$$
\mathcal{U}_t \mid w_t, \boldsymbol{\theta} \overset{\text{ind.}}{\sim} N\left(\mathbf{0}, \frac{1}{w_t} \cdot \boldsymbol{\Sigma}\right),
\tag{29}
$$

where the preceding conditional independence means that

$$
\begin{aligned}
& f(\mathbf{u}_t|\mathbf{u}_1, w_1 \ldots, \mathbf{u}_{t-1}, w_{t-1}, \mathbf{u}_{t+1}, w_{t+1}, \ldots, \mathbf{u}_n, w_n, w_t, \boldsymbol{\theta}) \\
& = f(\mathbf{u}_t|w_t, \boldsymbol{\theta}).
\end{aligned}
\tag{30}
$$

In contrast to model (18), a single weight w_t is assigned to the entire time series term $\mathcal{U}_t$; the larger the weight, the more extreme is the location of the vector $\mathbf{u}_t$ under the multidimensional tail of the multivariate distribution. The pdfs defining the preceding distributions are given by

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

and

$$
f_B(\mathbf{u}_t|w_t, \boldsymbol{\theta}) = \frac{1}{\sqrt{(2\pi)^N \det(\boldsymbol{\Sigma}/w_t)}} \exp\left(-\frac{w_t}{2} \mathbf{u}_t^T \boldsymbol{\Sigma}^{-1} \mathbf{u}_t\right).
$$

Then, the joint distribution of all the white-noise variables $\mathcal{U}$ and all latent variables $\mathcal{W}$ can be written as

$$
\begin{aligned}
f_B(\mathbf{u}, \mathbf{w}|\boldsymbol{\theta}) &= \prod_{t=1}^{n} f_B(\mathbf{u}_t, w_t|\boldsymbol{\theta}) \\
&= \prod_{t=1}^{n} f(w_t|\boldsymbol{\theta}) f_B(\mathbf{u}_t|w_t, \boldsymbol{\theta}) \\
&= \prod_{t=1}^{n} \frac{\left(\frac{v}{2}\right)^{\frac{v}{2}}}{\Gamma\left(\frac{v}{2}\right)} w_t^{\frac{v}{2}-1} \exp\left(-\frac{v}{2}w_t\right) \\
&\quad \times \frac{1}{\sqrt{(2\pi)^N \det(\boldsymbol{\Sigma}/w_t)}} \exp\left(-\frac{w_t}{2}\mathbf{u}_t^T \boldsymbol{\Sigma}^{-1}\mathbf{u}_t\right)
\end{aligned}
\tag{31}
$$

Since the unknown values $\mathbf{u}$ of the white-noise variables depend on the given observations $\boldsymbol{\ell}$ as well as the unknown parameters $\boldsymbol{\xi}$ and $\boldsymbol{\alpha}$, this pdf may be used as a likelihood function, for which one obtains

$$
\begin{aligned}
\log L_B(\boldsymbol{\theta}; \boldsymbol{\ell}, \mathbf{w}) &= \log f_B(\mathbf{u}, \mathbf{w}|\boldsymbol{\theta}) \\
&= -\frac{Nn}{2}\log(2\pi) - \frac{n}{2}\log\det\boldsymbol{\Sigma} + \frac{nv}{2}\log\left(\frac{v}{2}\right) \\
&\quad - n\log\Gamma\left(\frac{v}{2}\right) + \left(\frac{N}{2}-1\right)\sum_{t=1}^{n}\log w_t \\
&\quad - \frac{1}{2}\sum_{t=1}^{n} w_t[\mathbf{A}(L)(\boldsymbol{\ell}_t - \mathbf{h}_t(\boldsymbol{\xi}))]^T\boldsymbol{\Sigma}^{-1} \\
&\quad \times [\mathbf{A}(L)(\boldsymbol{\ell}_t - \mathbf{h}_t(\boldsymbol{\xi}))] + \frac{v}{2}\sum_{t=1}^{n}(\log w_t - w_t).
\end{aligned}
\tag{32}
$$

3 Adjustment of the observation models

In this section, ML estimates of the model parameters $\boldsymbol{\theta}$ under the stochastic models A and B are made. For this purpose, the fact that the structure of log-likelihood functions (24) and (27) allows for the setup of GEM algorithms is exploited. These algorithms iteratively perform an expectation ("E-step") and a maximization step ("M-step"). Let the number s denote the iteration step and $\widehat{\boldsymbol{\theta}}^{(s)}$ the parameter estimates obtained in that step. Then, the $(s+1)$th iteration of the E-step consists of the computation of the conditional expectation

$$
Q_{A/B}(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(s)}) = E_{\boldsymbol{\mathcal{W}}|\boldsymbol{\ell};\widehat{\boldsymbol{\theta}}^{(s)}}\left\{\log L_{A/B}(\boldsymbol{\theta}; \boldsymbol{\ell}, \boldsymbol{\mathcal{W}})\right\},
\tag{33}
$$

which is called "Q-function" below; "A/B" indicates that either log-likelihood function (24) for stochastic model A or log-likelihood function (32) for stochastic model B is sub-

stituted. The E-step will be seen to provide values $\mathbf{w}$ of the latent variables $\mathcal{W}$. Next, one would solve within the M-step the maximization problem

$$
\widehat{\boldsymbol{\theta}}^{(s+1)} = \arg\max_{\boldsymbol{\theta}} Q_{A/B}(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(s)}).
\tag{34}
$$

3.1 Generalized expectation maximization algorithm for stochastic model A

Replacing the decorrelation filter $\boldsymbol{\alpha}_k(L)(\ell_{k,t} - h_{k,t}(\boldsymbol{\xi}))$ by $\mathbf{A}_k(L)(\boldsymbol{\ell}_t - \mathbf{h}_t(\boldsymbol{\xi}))$ as in Alkhatib et al. (2018), the Q-function of the current stochastic model A can similarly be derived, that is,

$$
\begin{aligned}
Q_A(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(s)}) &= \text{const.} - \frac{n}{2}\sum_{k=1}^{N}\log(\sigma_k^2) \\
&\quad - \sum_{k=1}^{N}\frac{1}{2\sigma_k^2}\sum_{t=1}^{n}\widetilde{w}_{k,t}^{(s)}\left[\mathbf{A}_k(L)(\boldsymbol{\ell}_t - \mathbf{h}_t(\boldsymbol{\xi}))\right]^2 \\
&\quad + \frac{n}{2}\sum_{k=1}^{N} v_k \log v_k - n\sum_{k=1}^{N}\log\Gamma\left(\frac{v_k}{2}\right) \\
&\quad + \frac{n}{2}\sum_{k=1}^{N} v_k\left[\psi\left(\frac{\widehat{v}_k^{(s)}+1}{2}\right) - \log\left(\widehat{v}_k^{(s)}+1\right)\right. \\
&\quad \left. + \frac{1}{n}\sum_{t=1}^{n}\left(\log\widetilde{w}_{k,t}^{(s)} - \widetilde{w}_{k,t}^{(s)}\right)\right],
\end{aligned}
\tag{35}
$$

where

$$
\begin{aligned}
\widetilde{w}_{k,t}^{(s)} &= E_{\mathcal{W}_{k,t}|u_{k,t};\widehat{\boldsymbol{\theta}}^{(s)}}\{\mathcal{W}_{k,t}\} \\
&= \frac{\widehat{v}_k^{(s)}+1}{\widehat{v}_k^{(s)} + \left(\frac{\widehat{\mathbf{A}}_k^{(s)}(L)(\boldsymbol{\ell}_t - \mathbf{h}_t(\widehat{\boldsymbol{\xi}}^{(s)}))}{\widehat{\sigma}^{(s)}}\right)^2}.
\end{aligned}
\tag{36}
$$

The tilde symbol is used with these conditional expectations to distinguish these quantities more clearly from parameter estimates (which obtain a "hat"). The M-step is carried out in four sequential conditional maximization (CM) steps (in the sense of Meng and Rubin 1993) that correspond to the parameter groups $\boldsymbol{\xi}, \boldsymbol{\alpha}, \boldsymbol{\sigma}$ and $\boldsymbol{v}$ forming $\boldsymbol{\theta}$. The essential idea behind conditioning is to substitute the most recent available estimates for parameters belonging to groups other than the one to be determined by the current CM-step.

In the first CM-step the functional parameters $\boldsymbol{\xi}$, which occur as variables of the nonlinear regression functions $h_{k,t}$, are solved for. To resolve the nonlinearity, the functions are linearized by Taylor series expansion, which is possible within the framework of a GEM algorithm, as shown for instance in Phillips (2002). A natural choice for the Taylor

point is the solution $\widehat{\boldsymbol{\xi}}^{(s)}$ from the preceding iteration step s. GEM algorithm (Algorithm 1) is started with approximate, initial parameter values $\widehat{\boldsymbol{\xi}}^{(0)}$, which may be known from a previous adjustment or determined as part of a preprocessing step. Setting the partial derivative of the previous Q-function with respect to ξ_j equals to zero, as shown in Appendix B.2, one arrives at the normal equations

$$
\mathbf{0} = \sum_{k=1}^{N} \frac{1}{\sigma_k^2} \begin{bmatrix} \mathbf{A}_k(L)\mathbf{X}_{1,:1} & \cdots & \mathbf{A}_k(L)\mathbf{X}_{n,:1} \\ \vdots & & \vdots \\ \mathbf{A}_k(L)\mathbf{X}_{1,:m} & \cdots & \mathbf{A}_k(L)\mathbf{X}_{n,:m} \end{bmatrix} \widetilde{\mathbf{W}}_k^{(s)}
$$
$$
\times \begin{bmatrix} \mathbf{A}_k(L)\left(\boldsymbol{\Delta\ell}_1 - \mathbf{X}_1\boldsymbol{\Delta\xi}\right) \\ \vdots \\ \mathbf{A}_k(L)\left(\boldsymbol{\Delta\ell}_n - \mathbf{X}_n\boldsymbol{\Delta\xi}\right) \end{bmatrix} \tag{37}
$$

with the incremental observations

$$
\boldsymbol{\Delta\ell}_t = \boldsymbol{\ell}_t - \mathbf{h}_t(\widehat{\boldsymbol{\xi}}^{(s)}), \tag{38}
$$

the incremental parameters

$$
\boldsymbol{\Delta\xi} = \boldsymbol{\xi} - \widehat{\boldsymbol{\xi}}^{(s)} \tag{39}
$$

and $\mathbf{X}_{t,:j}$ being the jth column of the Jacobi matrix

$$
\mathbf{X}_t = \frac{\partial \mathbf{h}_t(\boldsymbol{\xi})}{\partial \boldsymbol{\xi}}\bigg|_{\boldsymbol{\xi}=\widehat{\boldsymbol{\xi}}^{(s)}} = \begin{bmatrix} \frac{\partial h_{1,t}(\boldsymbol{\xi})}{\partial \xi_1} & \cdots & \frac{\partial h_{1,t}(\boldsymbol{\xi})}{\partial \xi_m} \\ \vdots & & \vdots \\ \frac{\partial h_{N,t}(\boldsymbol{\xi})}{\partial \xi_1} & \cdots & \frac{\partial h_{N,t}(\boldsymbol{\xi})}{\partial \xi_m} \end{bmatrix}\bigg|_{\boldsymbol{\xi}=\widehat{\boldsymbol{\xi}}^{(s)}}. \tag{40}
$$

The step number s is omitted for brevity of expressions. Conditioning these equations on the estimated variance components $\widehat{\boldsymbol{\sigma}}^{(s)}$ and VAR coefficients $\widehat{\boldsymbol{\alpha}}^{(s)}$ from the preceding iteration step s, and performing also filtering operation (14) on the incremental observations and the Jacobi matrix results in the estimates

$$
\overline{\boldsymbol{\Delta\ell}}_{k,t} = \widehat{\mathbf{A}}_k^{(s)}(L)\boldsymbol{\Delta\ell}_t, \tag{41}
$$
$$
\overline{X}_{t,k,j} = \widehat{\mathbf{A}}_k^{(s)}(L)\mathbf{X}_{t,:j}, \tag{42}
$$
$$
\overline{\mathbf{X}}_{t,k:} = \widehat{\mathbf{A}}_k^{(s)}(L)\mathbf{X}_t, \tag{43}
$$

where $\overline{X}_{t,k,j}$ is the entry in the kth row and the jth column of the filtered Jacobi matrix at time t and $\overline{\mathbf{X}}_{t,k:}$ is accordingly the kth row of that matrix. The solution $\widehat{\boldsymbol{\Delta\xi}}^{(s+1)}$ satisfies

$$
\mathbf{0} = \sum_{k=1}^{N} \frac{1}{(\sigma_k^2)^{(s)}} \begin{bmatrix} \overline{X}_{1,k,1} & \cdots & \overline{X}_{n,k,1} \\ \vdots & & \vdots \\ \overline{X}_{1,k,m} & \cdots & \overline{X}_{n,k,m} \end{bmatrix} \widetilde{\mathbf{W}}_k^{(s)}
$$

$$
\times \begin{bmatrix} \overline{\Delta\ell}_{k,1} - \overline{\mathbf{X}}_{1,k:}\widehat{\boldsymbol{\Delta\xi}}^{(s+1)} \\ \vdots \\ \overline{\Delta\ell}_{k,n} - \overline{\mathbf{X}}_{n,k:}\widehat{\boldsymbol{\Delta\xi}}^{(s+1)} \end{bmatrix}
$$
$$
= \sum_{k=1}^{N} \frac{1}{(\widehat{\sigma}_k^2)^{(s)}} \begin{bmatrix} \overline{X}_{1,k,1} & \cdots & \overline{X}_{n,k,1} \\ \vdots & & \vdots \\ \overline{X}_{1,k,m} & \cdots & \overline{X}_{n,k,m} \end{bmatrix} \widetilde{\mathbf{W}}_k^{(s)}
$$
$$
\times \left(\begin{bmatrix} \overline{\Delta\ell}_{k,1} \\ \vdots \\ \overline{\Delta\ell}_{k,n} \end{bmatrix} - \begin{bmatrix} \overline{X}_{1,k,1} & \cdots & \overline{X}_{1,k,m} \\ \vdots & & \vdots \\ \overline{X}_{n,k,1} & \cdots & \overline{X}_{n,k,m} \end{bmatrix} \widehat{\boldsymbol{\Delta\xi}}^{(s+1)} \right)
$$
$$
=: \sum_{k=1}^{N} \frac{1}{(\widehat{\sigma}_k^2)^{(s)}} \overline{\mathbf{X}}_{k:}^{T}\widetilde{\mathbf{W}}_k^{(s)} \left(\overline{\boldsymbol{\Delta\ell}}_{k:} - \overline{\mathbf{X}}_{k:}\widehat{\boldsymbol{\Delta\xi}}^{(s+1)} \right).
$$

If the matrices $\overline{\mathbf{X}}_{1:}, \ldots, \overline{\mathbf{X}}_{N:}$ have full rank m, estimates of the incremental parameters can be computed by

$$
\widehat{\boldsymbol{\Delta\xi}}^{(s+1)} = \left(\sum_{k=1}^{N} \frac{1}{(\widehat{\sigma}_k^2)^{(s)}} \overline{\mathbf{X}}_{k:}^{T}\widetilde{\mathbf{W}}_k^{(s)}\overline{\mathbf{X}}_{k:} \right)^{-1}
$$
$$
\times \sum_{k=1}^{N} \frac{1}{(\widehat{\sigma}_k^2)^{(s)}} \overline{\mathbf{X}}_{k:}^{T}\widetilde{\mathbf{W}}_k^{(s)}\overline{\boldsymbol{\Delta\ell}}_{k:}, \tag{44}
$$

which yield with (39) the new solution for the full parameters

$$
\widehat{\boldsymbol{\xi}}^{(s+1)} = \widehat{\boldsymbol{\xi}}^{(s)} + \widehat{\boldsymbol{\Delta\xi}}^{(s+1)}. \tag{45}
$$

Notice that (44) involves additions of the normal equation matrices $\overline{\mathbf{X}}_{k:}^{T}\widetilde{\mathbf{W}}_k^{(s)}\overline{\mathbf{X}}_{k:}$ and corresponding "right-hand sides" $\overline{\mathbf{X}}_{k:}^{T}\widetilde{\mathbf{W}}_k^{(s)}\overline{\boldsymbol{\Delta\ell}}_{k:}$ across the N individual time series, because the white-noise time series were modeled independently by univariate (t-)distributions.

Next, estimates of random errors (4), also called "colored-noise residuals" below, are immediately obtained through

$$
\widehat{\mathbf{e}}_t^{(s+1)} = \boldsymbol{\ell}_t - \mathbf{h}_t(\widehat{\boldsymbol{\xi}}^{(s+1)}). \tag{46}
$$

To derive the new solution for the VAR coefficients $\boldsymbol{\alpha}$ within the second CM-step, the matrix

$$
\widehat{\mathbf{E}}^{(s+1)} = \begin{bmatrix} \widehat{\mathbf{e}}_0^{(s+1)T} & \cdots & \widehat{\mathbf{e}}_{1-p}^{(s+1)T} \\ \vdots & & \vdots \\ \widehat{\mathbf{e}}_{n-1}^{(s+1)T} & \cdots & \widehat{\mathbf{e}}_{n-p}^{(s+1)T} \end{bmatrix} \tag{47}
$$

is defined. As the colored-noise residuals can be computed by (46) only for $t = 1, \ldots, n$, the residuals $\widehat{\mathbf{e}}_0^{(s+1)}, \widehat{\mathbf{e}}_{-1}^{(s+1)}$, $\ldots$ with nonpositive time index values occurring in matrix (47) are undefined. To account for these missing values, the initial conditions $\widehat{\mathbf{e}}_0^{(s+1)} = \widehat{\mathbf{e}}_{-1}^{(s+1)} = \cdots = \mathbf{0}$ are used.

Appendix B.3 shows that the first-order conditions for $\boldsymbol{\alpha}_{k:}$ result in

$$\widehat{\boldsymbol{\alpha}}_{k:}^{(s+1)} = \left(\widehat{\mathbf{E}}^{(s+1)T} \widetilde{\mathbf{W}}_k^{(s)} \widehat{\mathbf{E}}^{(s+1)} \right)^{-1} \widehat{\mathbf{E}}^{(s+1)T} \widetilde{\mathbf{W}}_k^{(s)} \widehat{\boldsymbol{e}}_{k:}^{(s+1)}, \tag{48}$$

given that the matrix $\widehat{\mathbf{E}}^{(s+1)}$ has full rank. Note that this matrix is the same for each group $k \in \{1, \ldots, N\}$. Stacking the vectors $\widehat{\boldsymbol{\alpha}}_{1:}^{(s+1)}, \ldots, \widehat{\boldsymbol{\alpha}}_{N:}^{(s+1)}$ yields the vector $\widehat{\boldsymbol{\alpha}}^{(s+1)}$, which defines the updated VAR(p) model. Applying the decorrelation filter based on this model to correlated residuals (46), as indicated by (16), gives

$$\widehat{\mathbf{u}}_t^{(s+1)} = \widehat{\mathbf{A}}^{(s+1)}(L) \widehat{\mathbf{e}}_t^{(s+1)}. \tag{49}$$

Now, the variance components are estimated in the third CM-step group-by-group according to

$$(\widehat{\sigma}_k^2)^{(s+1)} = \frac{1}{n} \sum_{t=1}^{n} \widetilde{w}_{k,t}^{(s)} \widehat{u}_{k,t}^{(s+1)2} \tag{50}$$

$$= \frac{\widehat{\mathbf{u}}_{k:}^{(i+1)T} \widetilde{\mathbf{W}}_k^{(s)} \widehat{\mathbf{u}}_{k:}^{(s+1)}}{n}, \tag{51}$$

as the result of the corresponding first-order condition (see Appendix B.4 for details). To increase the computational efficiency of the remaining estimation of the *df*, the first-order condition with respect to the original log-likelihood function $L_A(\boldsymbol{\theta}; \boldsymbol{\ell})$ is used. This turns the fourth CM-step into a CM *either*-v(CME-) step (as suggested by Liu and Rubin 1994), which consists of the group-wise zero search of the equation (derived in Appendix B.5)

$$0 = \log \widehat{v}_k^{(s+1)} + 1 - \psi \left(\frac{\widehat{v}_k^{(s+1)}}{2} \right) + \psi \left(\frac{\widehat{v}_k^{(s+1)} + 1}{2} \right)$$
$$- \log \left(\widehat{v}_k^{(s+1)} + 1 \right) + \frac{1}{n} \sum_{t=1}^{n} \left(\log \widetilde{w}_{k,t}^{(s+1)} - \widetilde{w}_{k,t}^{(s+1)} \right) \tag{52}$$

with

$$\widetilde{w}_{k,t}^{(s+1)} = \frac{\widehat{v}_k^{(s+1)} + 1}{\widehat{v}_k^{(s+1)} + \left(\frac{\widehat{\mathbf{A}}_k^{(s+1)}(L)(\boldsymbol{\ell}_t - \mathbf{h}_t(\widehat{\boldsymbol{\xi}}^{(s+1)}))}{\widehat{\sigma}^{(s+1)}} \right)^2} \tag{53}$$

The steps of the current GEM algorithm under stochastic model A can be implemented easily as shown in following Algorithm 1, which includes the initialization and stop criterion.

Algorithm 1: GEM algorithm for stoch. model A

Input : $\boldsymbol{\ell}_1, \ldots, \boldsymbol{\ell}_n; \mathbf{h}_1(\boldsymbol{\xi}), \ldots, \mathbf{h}_n(\boldsymbol{\xi}); \widehat{\boldsymbol{\xi}}^{(0)}; p;$ itermax; $\varepsilon, \varepsilon_v$
Output: $\widehat{\boldsymbol{\xi}}; \widehat{\boldsymbol{\alpha}}; \widehat{\sigma}_1^2, \ldots, \widehat{\sigma}_N^2; \widehat{v}_1, \ldots, \widehat{v}_N; \widetilde{\mathbf{w}}_1, \ldots, \widetilde{\mathbf{w}}_n$
Set initial weights $\widetilde{w}_{1,1}^{(0)} = \cdots = \widetilde{w}_{N,n}^{(0)} = 1$
Set initial filter $\widehat{\mathbf{A}}^{(0)}(L) = \mathbf{I}$
Compute incremental observations $\boldsymbol{\Delta\ell}_1, \ldots, \boldsymbol{\Delta\ell}_n$ by (38)
Compute Jacobi matrices $\mathbf{X}_1, \ldots, \mathbf{X}_n$ by (40)
Compute functional parameter update $\widehat{\boldsymbol{\Delta\xi}}^{(1)}$ by (44) using
   $\overline{\boldsymbol{\Delta\ell}}_t = \boldsymbol{\Delta\ell}_t$ and $\overline{\mathbf{X}}_t = \mathbf{X}_t$ $(t = 1, \ldots, n)$
Compute functional parameter solution $\widehat{\boldsymbol{\xi}}^{(1)}$ by (45)
Compute colored-noise residuals $\widehat{\mathbf{e}}_1^{(1)}, \ldots, \widehat{\mathbf{e}}_n^{(1)}$ by (46)
Compute VAR coefficients $\widehat{\boldsymbol{\alpha}}^{(1)}$ via (48)
Compute white-noise residuals $\widehat{\mathbf{u}}_1^{(1)}, \ldots, \widehat{\mathbf{u}}_n^{(1)}$ by (49)
Compute variance components $(\widehat{\sigma}_1^2)^{(1)}, \ldots, (\widehat{\sigma}_N^2)^{(1)}$ by (50)
Set $df \, \widehat{v}^{(1)} = 30$
for $s = 1 \ldots$ itermax **do**
   Compute weights $\widetilde{w}_{1,1}^{(s)}, \ldots, \widetilde{w}_{N,n}^{(s)}$ by (36)
   Compute incremental observations $\boldsymbol{\Delta\ell}_1, \ldots, \boldsymbol{\Delta\ell}_n$ by (38)
   Compute Jacobi matrices $\mathbf{X}_1, \ldots, \mathbf{X}_n$ by (40)
   Filter incremental observations: $\overline{\boldsymbol{\Delta\ell}}_{1,1}, \ldots, \overline{\boldsymbol{\Delta\ell}}_{N,n}$ by (41)
   Filter Jacobi matrices: $\overline{\mathbf{X}}_1, \ldots, \overline{\mathbf{X}}_n$ by (42)
   Compute functional parameter update $\widehat{\boldsymbol{\Delta\xi}}^{(s+1)}$ by (44)
   Compute functional parameter solution $\widehat{\boldsymbol{\xi}}^{(s+1)}$ by (45)
   Compute colored-noise residuals $\widehat{\mathbf{e}}_1^{(s+1)}, \ldots, \widehat{\mathbf{e}}_n^{(s+1)}$ by (46)
   Compute VAR coefficients $\widehat{\boldsymbol{\alpha}}^{(s+1)}$ via (48)
   Compute white-noise residuals $\widehat{\mathbf{u}}_1^{(s+1)}, \ldots, \widehat{\mathbf{u}}_n^{(s+1)}$ by (49)
   Compute variance components $(\widehat{\sigma}_1^2)^{(s+1)}, \ldots, (\widehat{\sigma}_N^2)^{(s+1)}$ by (50)
   Search zeros (*df*s) $\widehat{v}_1^{(s+1)}, \ldots, \widehat{v}_N^{(s+1)}$ in (52)
   Compute maximum absolute parameter changes
   $d = \max(\max |\widehat{\boldsymbol{\xi}}^{(s)} - \widehat{\boldsymbol{\xi}}^{(s+1)}|, \max |\widehat{\boldsymbol{\alpha}}^{(s)} - \widehat{\boldsymbol{\alpha}}^{(s+1)}|,$
      $\max_k |(\widehat{\sigma}_k^2)^{(s)} - (\widehat{\sigma}_k^2)^{(s+1)}|),$
   $d_v = \max_k |\widehat{v}_k^{(s)} - \widehat{v}_k^{(s+1)}|$

   if $d > \varepsilon$ or $d_v > \varepsilon_v$ **then**
      $\lfloor \, s = s + 1$
   else
      $\lfloor \,$ break
Set $\widehat{\boldsymbol{\xi}} = \widehat{\boldsymbol{\xi}}^{(s+1)}, \widehat{\boldsymbol{\alpha}} = \widehat{\boldsymbol{\alpha}}^{(s+1)}, \widehat{\sigma}_1^2 = (\widehat{\sigma}_1^2)^{(s+1)}, \ldots,$
$\widehat{\sigma}_N^2 = (\widehat{\sigma}_N^2)^{(s+1)}, \widetilde{w}_{1,1} = \widetilde{w}_{1,1}^{(s)}, \ldots, \widetilde{w}_{N,n} = \widetilde{w}_{N,n}^{(s)}$

3.2 Generalized expectation maximization algorithm for stochastic model B

The solution of maximization problem (34) for the log-likelihood function $\log L_B(\boldsymbol{\theta}; \boldsymbol{\ell}, \boldsymbol{w})$ involves the multivariate t-distribution. To bring out clearly the modifications that have to be applied to the previously given derivation of EM algorithms in the univariate case (cf. Koch and Kargoll 2013; Kargoll et al. 2018a), the E-step is firstly elaborated in some detail.

3.2.1 The E-step

To derive the E-step analytically, it is convenient to separate within log-likelihood function (32) the terms involving w_t from those involving $\log w_t$, that is,

$$
\begin{aligned}
\log L_B(\boldsymbol{\theta}; \boldsymbol{\ell}, \mathbf{w}) = {}& -\frac{Nn}{2}\log(2\pi) - \frac{n}{2}\log \det \boldsymbol{\Sigma} \\
& + \frac{n\nu}{2}\log\left(\frac{\nu}{2}\right) - n\log\Gamma\left(\frac{\nu}{2}\right) \\
& - \sum_{t=1}^{n}\frac{1}{2}\left(\nu + [\mathbf{A}(L)(\boldsymbol{\ell}_t - \mathbf{h}_t(\boldsymbol{\xi}))]^T \boldsymbol{\Sigma}^{-1}\right. \\
& \qquad\qquad \times[\mathbf{A}(L)(\boldsymbol{\ell}_t - \mathbf{h}_t(\boldsymbol{\xi}))]\big) w_t \\
& + \sum_{t=1}^{n}\frac{1}{2}(\nu + N - 2)\log w_t.
\end{aligned}
$$

As mentioned in Appendix A, the previously considered univariate case is embedded in this multivariate case as the special case $N = 1$. Having defined the probabilistic model of the random variables in $\mathcal{U}$, we condition directly on the values $\mathbf{u}$, so that the Q-function becomes

$$
\begin{aligned}
Q_B(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(s)}) = {}& E_{\boldsymbol{\mathcal{W}}|\mathbf{u};\widehat{\boldsymbol{\theta}}^{(s)}}\{\log L_B(\boldsymbol{\theta}; \boldsymbol{\ell}, \boldsymbol{\mathcal{W}})\} = -\frac{Nn}{2}\log(2\pi) \\
& - \frac{n}{2}\log\det\boldsymbol{\Sigma} + \frac{n\nu}{2}\log\left(\frac{\nu}{2}\right) - n\log\Gamma\left(\frac{\nu}{2}\right) \\
& - \sum_{t=1}^{n}\frac{1}{2}\left(\nu + [\mathbf{A}(L)(\boldsymbol{\ell}_t - \mathbf{h}_t(\boldsymbol{\xi}))]^T \boldsymbol{\Sigma}^{-1}\right. \\
& \qquad\qquad \times[\mathbf{A}(L)(\boldsymbol{\ell}_t - \mathbf{h}_t(\boldsymbol{\xi}))]\big) E_{\boldsymbol{\mathcal{W}}|\boldsymbol{u};\widehat{\boldsymbol{\theta}}^{(s)}}\{\mathcal{W}_t\} \\
& + \sum_{t=1}^{n}\frac{1}{2}(\nu - 1)E_{\boldsymbol{\mathcal{W}}|\mathbf{u};\widehat{\boldsymbol{\theta}}^{(s)}}\{\log \mathcal{W}_t\}. \qquad (54)
\end{aligned}
$$

The conditional expectations are determined in analogy to those occurring in the Q-function for the univariate t-distribution model (see Kargoll et al. 2018a, Eqs. (39)–(41)). In the present case, the required conditional distribution of $\mathcal{W}_t|\mathbf{u}_t; \widehat{\boldsymbol{\theta}}^{(s)}$ can be shown to be a univariate gamma distribution with parameter values $a = \frac{\widehat{\nu}^{(s)}+N}{2}$ and $b = \frac{1}{2}(\widehat{\nu}^{(s)} + \mathbf{u}_t^T(\widehat{\boldsymbol{\Sigma}}^{(s)})^{-1}\mathbf{u}_t)$, according to Appendix B.1. Using the property that the expected value of a gamma-distributed random variable is given by a/b and that of its natural logarithm by $\psi(a) - \log b$ (see Appendix A.2 in Kargoll et al. 2018a) one finds

$$
\begin{aligned}
\widetilde{w}_t^{(s)} = E_{\boldsymbol{\mathcal{W}}|\mathbf{u};\widehat{\boldsymbol{\theta}}^{(s)}}\{\mathcal{W}_t\} &= E_{\mathcal{W}_t|\mathbf{u}_t;\widehat{\boldsymbol{\theta}}^{(s)}}\{\mathcal{W}_t\} \\
&= \frac{\widehat{\nu}^{(s)} + N}{\widehat{\nu}^{(s)} + \mathbf{u}_t^T(\widehat{\boldsymbol{\Sigma}}^{(s)})^{-1}\mathbf{u}_t} \qquad (55)
\end{aligned}
$$

as well as

$$
\begin{aligned}
E_{\boldsymbol{\mathcal{W}}|\mathbf{u};\widehat{\boldsymbol{\theta}}^{(s)}}\{\log \mathcal{W}_t\} &= E_{\mathcal{W}_t|\mathbf{u}_t;\widehat{\boldsymbol{\theta}}^{(s)}}\{\log \mathcal{W}_t\} = \psi\left(\frac{\widehat{\nu}^{(s)} + N}{2}\right) \\
&\quad - \log\left(\frac{1}{2}\left[\widehat{\nu}^{(s)} + \mathbf{u}_t^T(\widehat{\boldsymbol{\Sigma}}^{(s)})^{-1}\mathbf{u}_t\right]\right) \\
&= \log \widetilde{w}_t^{(s)} + \psi\left(\frac{\widehat{\nu}^{(s)} + N}{2}\right) \\
&\quad - \log\left(\frac{\widehat{\nu}^{(s)} + N}{2}\right). \qquad (56)
\end{aligned}
$$

Since (55) is conditioned on $\mathbf{u}_t$ and $\widehat{\boldsymbol{\theta}}^{(s)}$ in (55), (16) becomes $\mathbf{u}_t = \widehat{\mathbf{A}}^{(s)}(L)(\boldsymbol{\ell}_t - \mathbf{h}_t(\widehat{\boldsymbol{\xi}}^{(s)}))$, which is substituted into (55). Substituting these into (54) and omitting all terms that do not involve any parameter in $\boldsymbol{\theta}$, the Q-function can be written as

$$
\begin{aligned}
Q_B(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(s)}) = {}& -\frac{n}{2}\log\det\boldsymbol{\Sigma} - \frac{1}{2}\sum_{t=1}^{n}\widetilde{w}_t^{(s)}[\mathbf{A}(L)(\boldsymbol{\ell}_t - \mathbf{h}_t(\boldsymbol{\xi}))]^T \\
& \times \boldsymbol{\Sigma}^{-1}[\mathbf{A}(L)(\boldsymbol{\ell}_t - \mathbf{h}_t(\boldsymbol{\xi}))] + \frac{n\nu}{2}\log\nu \\
& - n\log\Gamma\left(\frac{\nu}{2}\right) + \frac{n\nu}{2}\left[\psi\left(\frac{\widehat{\nu}^{(s)} + N}{2}\right)\right. \\
& - \log\left(\widehat{\nu}^{(s)} + N\right) + \frac{1}{n}\sum_{t=1}^{n}\left(\log\widetilde{w}_t^{(s)} - \widetilde{w}_t^{(s)}\right)\Big].
\end{aligned}
$$

$$(57)$$

The omission of parameter-independent constants will not affect maximization with respect to the parameters $\boldsymbol{\theta}$, which is described next.

3.2.2 The M-step

To obtain estimates of the regression parameters in the first CM-step, the regression function is linearized exactly as before (see Sect. 3.1). The first-order conditions with respect to $\boldsymbol{\xi}$ can then be shown to lead to the normal equations

$$
\sum_{t=1}^{n}[\mathbf{A}(L)\mathbf{X}_t]^T\left[\widetilde{w}_t^{(s)}\boldsymbol{\Sigma}^{-1}\right][\mathbf{A}(L)(\boldsymbol{\Delta\ell}_t - \mathbf{X}_t\boldsymbol{\Delta\xi})] \qquad (58)
$$

with the incremental observations $\boldsymbol{\Delta\ell}_t$ computed by (38), incremental parameters $\boldsymbol{\Delta\xi}$ by (39), and Jacobi matrix $\mathbf{X}_t$ by (40). If these equations depend on the latest available estimates $\widehat{\boldsymbol{\Sigma}}^{(s)}$ and $\widehat{\boldsymbol{\alpha}}^{(s)}$, and the incremental observations and the Jacobi matrix are filtered as in (12) to obtain the vectors

$$
\overline{\boldsymbol{\Delta\ell}}_t = \widehat{\mathbf{A}}^{(s)}(L)\boldsymbol{\Delta\ell}_t, \qquad (59)
$$

$$
\overline{\mathbf{X}}_t = \widehat{\mathbf{A}}^{(s)}(L)\mathbf{X}_t, \qquad (60)
$$

then one can write for the normal equations

$$0 = \sum_{t=1}^{n} \overline{\mathbf{X}}_t^T \left[\widetilde{w}_t^{(s)} (\widehat{\boldsymbol{\Sigma}}^{(s)})^{-1} \right] \left(\overline{\boldsymbol{\Delta\ell}}_t - \overline{\mathbf{X}}_t \widehat{\boldsymbol{\Delta\xi}}^{(s+1)} \right).$$

Given that the filtered Jacobi matrices have full rank, the solution of the normal equations is given by

$$\widehat{\boldsymbol{\Delta\xi}}^{(s+1)} = \left(\sum_{t=1}^{n} \overline{\mathbf{X}}_t^T \left[\widetilde{w}_t^{(s)} (\widehat{\boldsymbol{\Sigma}}^{(s)})^{-1} \right] \overline{\mathbf{X}}_t \right)^{-1}$$
$$\times \sum_{t=1}^{n} \overline{\mathbf{X}}_t^T \left[\widetilde{w}_t^{(s)} (\widehat{\boldsymbol{\Sigma}}^{(s)})^{-1} \right] \overline{\boldsymbol{\Delta\ell}}_t, \quad (61)$$

resulting in the new solution $\widehat{\boldsymbol{\xi}}^{(s+1)}$ via (45). Since the white-noise time series are now generally correlated across the N groups through the multivariate t-distribution, the normal equation matrices $\overline{\mathbf{X}}_t^T [\widetilde{w}_t^{(s)} (\widehat{\boldsymbol{\Sigma}}^{(s)})^{-1}] \overline{\mathbf{X}}_t$ and right-hand sides $\overline{\mathbf{X}}_t^T [\widetilde{w}_t^{(s)} (\widehat{\boldsymbol{\Sigma}}^{(s)})^{-1}] \overline{\boldsymbol{\Delta\ell}}_t$ are formed group-wise. As the white-noise vectors are stochastically independent throughout time, these normal equations can be added up accordingly, as shown in (61). Next, the residuals $\widehat{\mathbf{e}}_t^{(s+1)}$ and the associated matrix $\widehat{\mathbf{E}}^{(s+1)}$ are computed as in (46) and (47), respectively.

Regarding the second CM-step, one obtains from the first-order conditions with respect to the VAR coefficient matrices (see Appendix B.3)

$$\left[\widehat{\mathbf{A}}^{(1)(s+1)} \cdots \widehat{\mathbf{A}}^{(p)(s+1)} \right] = \left[\widehat{\mathbf{e}}_1^{(s+1)} \cdots \widehat{\mathbf{e}}_n^{(s+1)} \right] \widetilde{\mathbf{W}}^{(s)} \widehat{\mathbf{E}}^{(s+1)}$$
$$\times \left(\widehat{\mathbf{E}}^{(s+1)T} \widetilde{\mathbf{W}}^{(s)} \widehat{\mathbf{E}}^{(s+1)} \right)^{-1}. \quad (62)$$

The new filter $\widehat{\mathbf{A}}(L)^{(s+1)}$ enables the estimation of the white-noise errors $\widehat{\mathbf{u}}_t^{(s+1)}$ through (49). These enter, alongside the current weights, the third CM-step, in which the cofactor matrix $\boldsymbol{\Sigma}$ is estimated. Observe that the relevant first two terms of Q-function (57) are similar to the log-likelihood function for the multivariate Gauss–Markov model based on normally distributed observations, so that the first-order conditions with respect to $\boldsymbol{\Sigma}^{-1}$ can be shown to lead to (see Appendix B.4)

$$\widehat{\boldsymbol{\Sigma}}^{(s+1)} = \frac{1}{n} \sum_{t=1}^{n} \widetilde{w}_t^{(s)} \widehat{\mathbf{u}}_t^{(s+1)} \widehat{\mathbf{u}}_t^{(s+1)T}. \quad (63)$$

Finally, the first-order condition with respect to the log-likelihood function and ν yields the fourth CM-step (in the form of CME, similarly as shown for stochastic model A in

Appendix B.5), consisting of the zero search

$$0 = \log \widehat{\nu}^{(s+1)} + 1 - \psi \left(\frac{\widehat{\nu}^{(s+1)}}{2} \right) + \psi \left(\frac{\widehat{\nu}^{(s+1)} + N}{2} \right)$$
$$- \log \left(\widehat{\nu}^{(s+1)} + N \right) + \frac{1}{n} \sum_{t=1}^{n} \left(\log \widetilde{w}_t^{(s+1)} - \widetilde{w}_t^{(s+1)} \right)$$
$$\quad (64)$$

where

$$\widetilde{w}_t^{(s+1)} = \frac{\widehat{\nu}^{(s+1)} + N}{\widehat{\nu}^{(s+1)} + \mathbf{u}_t^T (\widehat{\boldsymbol{\Sigma}}^{(s+1)})^{-1} \mathbf{u}_t} \quad (65)$$

and $\mathbf{u}_t = \widehat{\mathbf{A}}^{(s+1)}(L)(\boldsymbol{\ell}_t - \mathbf{h}_t(\widehat{\boldsymbol{\xi}}^{(s+1)}))$. Following Algorithm 2 summarizes the computational steps of the E- and M-step under the current stochastic model B.

3.3 Model selection

In practical situations, it might not be clear at the outset which VAR model order p or which stochastic model (A or B) is most adequate for the given observations. Therefore, this section provides some techniques for identifying the model that is best in a statistical sense. First, a statistical white noise test is elaborated, which enables one to assess the capability of an estimated VAR model to fully capture the auto- and cross-correlation pattern. Subsequently, a standard information criterion adapted to the generic observation models of Sect. 2 is defined.

3.3.1 Multivariate portmanteau test

The portmanteau test by Box and Pierce (1970) may be used to test whether the residuals of an autoregressive moving average (ARMA) are uncorrelated. Hosking (1980) suggested an extension of the portmanteau statistic to vector ARMA (VARMA) models, of which the VAR models considered in our current paper constitute special cases. This statistic has an approximate Chi-square distribution with *df* $N^2(h - p)$ and may be written as

$$P = n \sum_{l=1}^{h} \text{vec}(\widehat{\boldsymbol{\Sigma}}_l)^T \left(\widehat{\boldsymbol{\Sigma}}_0^{-1} \otimes \widehat{\boldsymbol{\Sigma}}_0^{-1} \right) \text{vec}(\widehat{\boldsymbol{\Sigma}}_l), \quad (66)$$

where h is the maximum lag to be considered and $\widehat{\boldsymbol{\Sigma}}_l$ the empirical covariance matrix at lag l with respect to the decorrelation-filtered residuals $\widehat{\mathbf{u}}_t$—estimated by

$$\widehat{\boldsymbol{\Sigma}}_l = \frac{1}{n} \sum_{t=1}^{n} \widehat{\mathbf{u}}_t \widehat{\mathbf{u}}_{t+l}^T \quad (l = 0, \dots, h). \quad (67)$$

Algorithm 2: GEM algorithm for stoch. model B

Input   : $\boldsymbol{\ell}_1, \ldots, \boldsymbol{\ell}_n; \mathbf{h}_1(\boldsymbol{\xi}), \ldots, \mathbf{h}_n(\boldsymbol{\xi}); \widehat{\boldsymbol{\xi}}^{(0)}; p$; itermax; $\varepsilon, \varepsilon_\nu$
Output: $\widehat{\boldsymbol{\xi}}; \widehat{\boldsymbol{\alpha}}; \widehat{\boldsymbol{\Sigma}}; \widehat{\nu}; \widetilde{w}_1, \ldots, \widetilde{w}_n$
Set initial weights $\widetilde{w}_1^{(0)} = \cdots = \widetilde{w}_n^{(0)} = 1$
Set initial filter $\widehat{\mathbf{A}}^{(0)}(L) = \mathbf{I}$
Compute incremental observations $\boldsymbol{\Delta\ell}_1, \ldots, \boldsymbol{\Delta\ell}_n$ by (38)
Compute Jacobi matrices $\mathbf{X}_1, \ldots, \mathbf{X}_n$ by (40)
Compute functional parameter update $\widehat{\boldsymbol{\Delta\xi}}^{(1)}$ by (61) using
    $\overline{\boldsymbol{\Delta\ell}}_t = \boldsymbol{\Delta\ell}_t$ and $\overline{\mathbf{X}}_t = \mathbf{X}_t$ $(t = 1, \ldots, n)$
Compute functional parameter solution $\widehat{\boldsymbol{\xi}}^{(1)}$ by (45)
Compute colored-noise residuals $\widehat{\mathbf{e}}_1^{(1)}, \ldots, \widehat{\mathbf{e}}_n^{(1)}$ by (46)
Compute VAR coefficients $\widehat{\boldsymbol{\alpha}}^{(1)}$ via (62)
Compute white-noise residuals $\widehat{\mathbf{u}}_1^{(1)}, \ldots, \widehat{\mathbf{u}}_n^{(1)}$ by (49)
Compute cofactor matrix $\widehat{\boldsymbol{\Sigma}}^{(1)}$ by (63)
Set $df\,\widehat{\nu}^{(1)} = 30$
for $s = 1 \ldots$ itermax **do**
    Compute weights $\widetilde{w}_1^{(s)}, \ldots, \widetilde{w}_n^{(s)}$ by (55)
    Compute incremental observations $\boldsymbol{\Delta\ell}_1, \ldots, \boldsymbol{\Delta\ell}_n$ by (38)
    Compute Jacobi matrices $\mathbf{X}_1, \ldots, \mathbf{X}_n$ by (40)
    Filter incremental observations: $\overline{\boldsymbol{\Delta\ell}}_1, \ldots, \overline{\boldsymbol{\Delta\ell}}_n$ by (59)
    Filter Jacobi matrices: $\overline{\mathbf{X}}_1, \ldots, \overline{\mathbf{X}}_n$ by (60)
    Compute functional parameter update $\widehat{\boldsymbol{\Delta\xi}}^{(s+1)}$ by (61)
    Compute functional parameter solution $\widehat{\boldsymbol{\xi}}^{(s+1)}$ by (45)
    Compute colored-noise residuals $\widehat{\mathbf{e}}_1^{(s+1)}, \ldots, \widehat{\mathbf{e}}_n^{(s+1)}$ by (46)
    Compute VAR coefficients $\widehat{\boldsymbol{\alpha}}^{(s+1)}$ via (62)
    Compute white-noise residuals $\widehat{\mathbf{u}}_1^{(s+1)}, \ldots, \widehat{\mathbf{u}}_n^{(s+1)}$ by (49)
    Compute cofactor matrix $\widehat{\boldsymbol{\Sigma}}^{(s+1)}$ by (63)
    Search zero $(df)\,\widehat{\nu}^{(s+1)}$ in (64)
    Compute maximum absolute parameter changes
    $d = \max(\max|\widehat{\boldsymbol{\xi}}^{(s)} - \widehat{\boldsymbol{\xi}}^{(s+1)}|, \max|\widehat{\boldsymbol{\alpha}}^{(s)} - \widehat{\boldsymbol{\alpha}}^{(s+1)}|,$
        $\max|\widehat{\boldsymbol{\Sigma}}^{(s)} - \widehat{\boldsymbol{\Sigma}}^{(s+1)}|),$
    $d_\nu = \max|\widehat{\nu}^{(s)} - \widehat{\nu}^{(s+1)}|$
    if $d > \varepsilon$ or $d_\nu > \varepsilon_\nu$ **then**
        $\llcorner$ $s = s + 1$
    else
        $\llcorner$ break
Set $\widehat{\boldsymbol{\xi}} = \widehat{\boldsymbol{\xi}}^{(s+1)}, \widehat{\boldsymbol{\alpha}} = \widehat{\boldsymbol{\alpha}}^{(s+1)}, \widehat{\boldsymbol{\Sigma}} = \widehat{\boldsymbol{\Sigma}}^{(s+1)},$
$\widetilde{w}_1 = \widetilde{w}_1^{(s)}, \ldots, \widetilde{w}_n = \widetilde{w}_n^{(s)}$

In particular, the matrix $\widehat{\boldsymbol{\Sigma}}_0$ occurring in (66) represents the empirical covariance matrix at lag $l = 0$. Furthermore, "vec" is the vectorization operator that stacks all columns of a matrix within a single column vector, and $\otimes$ is the Kronecker product. In view of the similarity of the estimation of the lag-dependent covariance matrix with third CM step (63), we propose to replace estimates (67) by the reweighted estimates

$$\widehat{\boldsymbol{\Sigma}}_l = \frac{1}{n} \sum_{t=1}^{n} \widetilde{w}_t \widehat{\mathbf{u}}_t \widehat{\mathbf{u}}_{t+l}^T \quad (l = 0, \ldots, h). \tag{68}$$

in connection with the multivariate t-distribution of stochastic model B. Similarly, the weights regarding the univariate

t-distributions of stochastic model A can be taken into account by

$$\widehat{\boldsymbol{\Sigma}}_l = \frac{1}{n} \sum_{t=1}^{n} \sqrt{\widetilde{w}_t \widetilde{w}_{t+l}} \, \widehat{\mathbf{u}}_t \widehat{\mathbf{u}}_{t+l}^T \quad (l = 0, \ldots, h). \tag{69}$$

In Sect. 4.2 the effects of these modifications regarding the portmanteau test's type-I error rate are quantified in a Monte Carlo simulation.

3.3.2 Akaike information criterion

Various information criteria for selecting an adequate AR model order p have been proposed (cf. Brockwell and Davis 2016, Sect. 5.5). Although our focus is not on forecasting, we follow these authors' recommendation to adopt a criterion that strongly penalizes and thus avoids overly large model orders. Since the assumption that the auto-and cross-correlation pattern of a measured geodetic time series is always generated by an AR process appears to be unrealistic (see Sect. 5), the Bayesian information criterion (BIC) is not considered in the sequel. Instead, the Akaike information criterion (AIC) and a bias-corrected version of the AIC, commonly termed AIC *corrected* (AICC), will be used for VAR model selection (cf. Brockwell and Davis 2016, Sects. 5.5.2 and 8.6). Since the usual assumption of normally distributed white noise is in general not justified for the stochastic models A and B, the general definitions of the AIC and AICC (cf. Burnham and Anderson 2002, Sects. 2.2 and 2.4) are considered here. Using log-likelihood functions (20) and (27) and letting

$$K_A = m + N^2 \cdot p + 2N, \tag{70}$$
$$K_B = m + N^2 \cdot p + N^2 + 1 \tag{71}$$

be the corresponding total numbers of unknown model parameters (where N is the number of groups of observations, m the number of parameters of the deterministic model, and p the order of the VAR process specified by the user) for stochastic model A and B, the corresponding AIC and AICC are then given by

$$\mathrm{AIC}_A = -2\log L_A(\widehat{\boldsymbol{\theta}}; \boldsymbol{\ell}) + 2K_A, \tag{72}$$

$$\mathrm{AIC}_B = -2\log L_B(\widehat{\boldsymbol{\theta}}; \boldsymbol{\ell}) + 2K_B, \tag{73}$$

$$\mathrm{AICC}_A = -2\log L_A(\widehat{\boldsymbol{\theta}}; \boldsymbol{\ell}) + 2K_A + \frac{2K_A(K_A + 1)}{N \cdot n - K_A - 1}, \tag{74}$$

$$\mathrm{AICC}_B = -2\log L_B(\widehat{\boldsymbol{\theta}}; \boldsymbol{\ell}) + 2K_B + \frac{2K_B(K_B + 1)}{N \cdot n - K_B - 1}. \tag{75}$$

The required log-likelihood values are computable using the outputs of GEM Algorithms 1 and 2, whereas the scalar quantities K_A and K_B determined by (70)–(71) are computed in the setup of the adjustment problem.

4 Monte Carlo results

4.1 Parameter estimation

In the first part of this section, the performance of Algorithms 1 and 2 in producing correct estimates within a Monte Carlo simulation is studied. For this purpose, generated measurements are approximated by a 3D circle defined by the (nonlinear) equations

$$
\begin{aligned}
x_t = \ell_{1,t} &= h_{1,t}(r, c_x, c_y, c_z, r, \varphi, \omega) \\
&= -r \cos(T_t) \cos(\varphi) + c_x
\end{aligned}
\tag{76}
$$

$$
\begin{aligned}
y_t = \ell_{2,t} &= h_{2,t}(r, c_x, c_y, c_z, r, \varphi, \omega) \\
&= r \cos(T_t) \sin(\varphi) \sin(\omega) + r \sin(T_t) \cos(\omega) + c_y
\end{aligned}
\tag{77}
$$

$$
\begin{aligned}
z_t = \ell_{3,t} &= h_{3,t}(r, c_x, c_y, c_z, r, \varphi, \omega) \\
&= -r \cos(T_t) \sin(\varphi) \cos(\omega) + r \sin(T_t) \sin(\omega) + c_z,
\end{aligned}
\tag{78}
$$

where the center coordinates c_x, c_y, c_z, the radius r and the rotation angles φ, ω about the x- and y-axis, respectively, are treated as unknown parameters to be estimated. The (simulated) true values are specified by

$$
\boldsymbol{\xi} = \begin{bmatrix} c_x \\ c_y \\ c_z \\ r \\ \varphi \\ \omega \end{bmatrix} = \begin{bmatrix} -1663.1 \\ 1223.4 \\ 1.6 \\ 29.7 \\ 0 \\ 0 \end{bmatrix}.
$$

This model approximately describes the trajectory of two GNSS receivers mounted on a terrestrial laser scanner (TLS), which rotates about the z-axis. The GNSS data are observed for the purpose of geo-referencing the TLS point clouds by using 3D coordinates in a coordinate frame of superordinate precision (see Paffenholz 2012, for details of this multi-sensor system and the geo-referencing methodology). In our current simulation study, the time instances are simply determined by $T_t = (t-1) \cdot \frac{2\pi}{n}$ s for $t \in \{1, \ldots, n\}$, where $n = 1000$, $n = 10{,}000$ or $n = 100{,}000$ is the total number of modeled 3D points $\boldsymbol{\ell}_t = [x_t, y_t, z_t]^T$. Thus, the measurements are equidistantly distributed along the circle. The numerical realizations $\mathbf{e}_t = [e_{1,t}, e_{2,t}, e_{3,t}]^T$ of the corresponding random errors are assumed to be determined by

the VAR process

$$
\begin{bmatrix} e_{1,t} \\ e_{2,t} \\ e_{3,t} \end{bmatrix} = \begin{bmatrix} 0.5653 & -0.0066 & -0.0197 \\ 0.0150 & 0.6657 & 0.0102 \\ -0.0431 & 0.0207 & 0.7577 \end{bmatrix} \begin{bmatrix} e_{1,t-1} \\ e_{2,t-1} \\ e_{3,t-1} \end{bmatrix} + \begin{bmatrix} u_{1,t} \\ u_{2,t} \\ u_{3,t} \end{bmatrix},
$$

where random numbers $\mathbf{u}_t = [u_{1,t}, u_{2,t}, u_{3,t}]^T$ for the white-noise components are generated based on the following instances of stochastic model A and B:

(A1) stochastic model A with an approximate normal distribution ($v = 120$) and different variance components:

$$
\mathcal{U}_{1,t} \overset{\text{ind.}}{\sim} t_{120}(0, \sigma_0^2),
$$
$$
\mathcal{U}_{2,t} \overset{\text{ind.}}{\sim} t_{120}(0, 2\sigma_0^2),
$$
$$
\mathcal{U}_{3,t} \overset{\text{ind.}}{\sim} t_{120}(0, 4\sigma_0^2).
$$

with $\sigma_0^2 = 0.001^2$.

(B1) stochastic model B with an approximate normal distribution ($v = 120$) and a fully populated cofactor matrix:

$$
\begin{bmatrix} \mathcal{U}_{1,t} \\ \mathcal{U}_{2,t} \\ \mathcal{U}_{3,t} \end{bmatrix} \overset{\text{ind.}}{\sim} t_{120}\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \sigma_0^2 \cdot \begin{bmatrix} 1 & 0.98 & 1.4 \\ 0.98 & 2 & 1.96 \\ 1.4 & 1.96 & 4 \end{bmatrix} \right),
$$

where the given cofactor matrix reflects the assumption that the entries of $\mathcal{U}_t$ are pairwise correlated by correlation coefficient $\rho = 0.7$.

(A2) stochastic model A with variance components as in model A1 and different small dfs:

$$
\mathcal{U}_{1,t} \overset{\text{ind.}}{\sim} t_3(0, \sigma_0^2),
$$
$$
\mathcal{U}_{2,t} \overset{\text{ind.}}{\sim} t_4(0, 2\sigma_0^2),
$$
$$
\mathcal{U}_{3,t} \overset{\text{ind.}}{\sim} t_5(0, 4\sigma_0^2).
$$

(B2) stochastic model B with cofactor matrix as in model B1 and small df:

$$
\begin{bmatrix} \mathcal{U}_{1,t} \\ \mathcal{U}_{2,t} \\ \mathcal{U}_{3,t} \end{bmatrix} \overset{\text{ind.}}{\sim} t_3\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \sigma_0^2 \cdot \begin{bmatrix} 1 & 0.98 & 1.4 \\ 0.98 & 2 & 1.96 \\ 1.4 & 1.96 & 4 \end{bmatrix} \right),
$$

In total 1000 Monte Carlo (MC) samples of the white-noise components for the stochastic models A1–B2 were randomly generated using MATLAB's trnd and mvtrnd routines. The resulting simulated observations in models A1 and A2 were then adjusted by means of GEM Algorithm 1, whereas the samples in models B1 and B2 were adjusted using Algorithm 2. The maximum number of iterations was set to itermax = 100, and the convergence thresholds to $\varepsilon = 10^{-8}$

and $\varepsilon_v = 0.0001$. For each MC run i the root-mean-square errors (RMSE) was computed for different parameters. For example, the RMSE for the center of the circle was obtained by

$$
\text{RMSE}_c^{(i)} = \left[\left(c_x - \hat{c}_x^{(i)} \right)^2 + \left(c_y - \hat{c}_y^{(i)} \right)^2 + \left(c_z - \hat{c}_z^{(i)} \right)^2 \right]^{1/2},
$$

where c_x, c_y, c_z and $\hat{c}_x^{(i)}$, $\hat{c}_y^{(i)}$, $\hat{c}_z^{(i)}$, respectively, are the true values and the estimates. The RMSE for the other parameters shown in Table 1 is defined similarly (the RMSE with respect to the VAR coefficients was also determined collectively). Only the results for the Student random error models A2 and B2 are given since the RMSE values for the Gaussian random error models A1 and B1 were similar. To compute the interval estimates, the empirical 2.5 and 97.5 percentiles were determined from the sorted MC estimates; e.g., $\hat{r}_{0.025}$ and $\hat{r}_{0.975}$ for the radius estimates $\hat{r}^{(i)}$, so that

$$
P(\hat{r}_{0.025} < r < \hat{r}_{0.975}) = 0.95.
$$

As it is intended to analyze the degree to which the bias is reduced by increasing the number of observations, decimal floating point numbers are listed with only one digit before the decimal point of the mantissa.

In the majority of cases the mean, minimum and maximum RMSE decreases by one order of magnitude when the number observations per time series are increased from 1000 to 100,000 for the multivariate Student random error model B2. With model A2 the improvement of the RMSE is notable but usually less than one order of magnitude. The largest RMSE values occur for the df and the VAR coefficients. It may therefore be concluded that these parameters are most difficult to estimate. This problem is most obvious for the model involving univariate t-distributions. A similar finding was obtained through the MC study in Kargoll et al. (2018a) for the df and the (univariate) AR coefficients. We therefore conclude that the estimation of the df of the multivariate Student random error model is much more accurate than the df of the univariate Student random error models. The improvement of the accuracy of the estimated df obtained for the former model when the number of observations per time series is increased from 1000 to 100,000 is clearly visible in histogram plots (Fig. 1). On the one hand, the histogram for $n = 100,000$ is much narrower and more centered about the true value $v = 3$ than the histogram for $n = 1000$ and $n = 10,000$. The fact that these histogram are skewed is a phenomenon well known in the context of estimating the df of a Student distribution (e.g., Koop 2003, Sect. 6.4.3).

4.2 White noise test

Reweighted and weighted multivariate portmanteau statistics (66) based on (67)–(69) were evaluated for 1000 MC runs with respect to all four simulation models A1–B2 and by generating either $n = 1000$, $n = 10,000$ or $n = 100,000$ observations per time series. For each of these scenarios the number of times that the white-noise hypothesis was rejected due to the test with significance level 0.05 was counted, and the obtained numbers were divided by the number of MC samples. Ideally, the resulting rejection rates (see Table 2) are equal to the significance level since the actual frequency of falsely rejecting a true null hypothesis corresponds to the prescribed probability of a type-I error in connection with the test distribution used. As the MC simulation is a closed-loop simulation in the sense that the same model is used both for the generation of the observations and their subsequent adjustment, large deviations of the rejection rates from the significance level indicate that the model is not estimated well by the EM algorithm or that the test distribution does not apply exactly. To differentiate these systematic deviations from random fluctuations, the rejection rates were validated by an additional MC simulation using newly generated samples. The resulting differences between the rejection rates of the two MC simulations turned out to be a few per mille. Based on these results, it can be concluded that differences between the four rejection rates for model A1 are not significant. Furthermore, for model B1 there is no substantial difference in rate between reweighted and unweighted application of the portmanteau test, whereas increasing the number n of observations per time series from 1000 to 100,000 reduces the deviations from the significance level notably (giving even the exact value 0.05 for the reweighted test). With the Student models A2 and B2 the reweighted test performs slightly better than the unweighted test, whereas an increase of the number of observations does not bring about a general improvement. The reweighted test produced rejection rates between 0.036 and 0.076, which do not grossly deviate from the specified significance level of 0.05. We therefore conclude that both the usage of the EM algorithm and of the approximate test distribution yield satisfactory estimation and testing results.

5 Analysis of Global Positioning System time series from the EUREF Permanent Network

5.1 Introduction

The monitoring of geophysical phenomena such as tectonic velocity can be achieved by analyzing the site velocities of Global Positioning System (GPS) permanent stations. For the proper interpretation of the results and deduction of

Table 1 Monte Carlo RMSE statistics for the selected parameter estimates $\widehat{c}_x$, $\widehat{c}_y$, $\widehat{c}_z$ (circle center coordinates), $\widehat{r}$ (circle radius), $\widehat{\varphi}$ (one of the rotation angles), $\widehat{\boldsymbol{\alpha}}$ (all VAR coefficients), $\widehat{\sigma}_1^2$ (scale factor for the first time series), $\widehat{\nu}_1$ (df for the first time series with model A2) or $\widehat{\nu}_1$ (df with model B2)

Model	AT			BT		
n	1000	10,000	100,000	1000	10,000	100,000
$\widehat{c}_{x,y,z}$						
Mean RMSE	4×10^{-4}	1×10^{-4}	4×10^{-5}	3×10^{-4}	9×10^{-5}	3×10^{-5}
Minimum RMSE	1×10^{-5}	3×10^{-5}	3×10^{-6}	1×10^{-5}	6×10^{-6}	1×10^{-6}
maximum RMSE	1×10^{-3}	4×10^{-4}	1×10^{-4}	1×10^{-3}	3×10^{-4}	1×10^{-4}
95% interval	$[9 \times 10^{-5}, 8 \times 10^{-4}]$	$[3 \times 10^{-5}, 2 \times 10^{-4}]$	$[1 \times 10^{-5}, 8 \times 10^{-5}]$	$[6 \times 10^{-5}, 7 \times 10^{-4}]$	$[1 \times 10^{-5}, 2 \times 10^{-4}]$	$[6 \times 10^{-6}, 8 \times 10^{-5}]$
$\widehat{r}$						
Mean RMSE	1×10^{-4}	4×10^{-5}	1×10^{-5}	6×10^{-4}	2×10^{-5}	6×10^{-6}
Minimum RMSE	7×10^{-8}	9×10^{-8}	5×10^{-8}	2×10^{-9}	1×10^{-8}	2×10^{-8}
Maximum RMSE	6×10^{-4}	2×10^{-4}	6×10^{-5}	3×10^{-4}	7×10^{-5}	3×10^{-5}
95% interval	$[5 \times 10^{-6}, 4 \times 10^{-4}]$	$[1 \times 10^{-6}, 1 \times 10^{-4}]$	$[5 \times 10^{-7}, 4 \times 10^{-5}]$	$[2 \times 10^{-6}, 2 \times 10^{-4}]$	$[6 \times 10^{-7}, 5 \times 10^{-5}]$	$[2 \times 10^{-7}, 2 \times 10^{-5}]$
$\widehat{\varphi}$						
Mean RMSE	7×10^{-6}	3×10^{-6}	9×10^{-7}	8×10^{-6}	2×10^{-6}	8×10^{-7}
Minimum RMSE	3×10^{-8}	9×10^{-9}	7×10^{-10}	4×10^{-9}	8×10^{-9}	3×10^{-9}
Maximum RMSE	8×10^{-5}	5×10^{-5}	5×10^{-6}	4×10^{-5}	1×10^{-5}	4×10^{-6}
95% interval	$[4 \times 10^{-7}, 6 \times 10^{-5}]$	$[2 \times 10^{-7}, 3 \times 10^{-6}]$	$[4 \times 10^{-8}, 2 \times 10^{-6}]$	$[3 \times 10^{-7}, 2 \times 10^{-5}]$	$[1 \times 10^{-7}, 7 \times 10^{-6}]$	$[3 \times 10^{-8}, 2 \times 10^{-6}]$
$\widehat{\alpha}_{k;i,j}$						
Mean RMSE	7×10^{-2}	2×10^{-2}	7×10^{-3}	8×10^{-2}	2×10^{-2}	7×10^{-3}
Minimum RMSE	2×10^{-2}	1×10^{-2}	1×10^{-3}	2×10^{-2}	5×10^{-3}	1×10^{-3}
Maximum RMSE	2×10^{-1}	5×10^{-2}	2×10^{-2}	2×10^{-1}	8×10^{-2}	2×10^{-2}
95% interval	$[4 \times 10^{-2}, 1 \times 10^{-1}]$	$[1 \times 10^{-2}, 4 \times 10^{-2}]$	$[4 \times 10^{-3}, 1 \times 10^{-2}]$	$[3 \times 10^{-2}, 2 \times 10^{-1}]$	$[9 \times 10^{-2}, 1 \times 10^{-1}]$	$[3 \times 10^{-3}, 1 \times 10^{-2}]$
$\widehat{\sigma}_1^2$						
Mean RMSE	5×10^{-12}	5×10^{-12}	4×10^{-12}	4×10^{-15}	4×10^{-16}	4×10^{-17}
Minimum RMSE	8×10^{-13}	2×10^{-12}	3×10^{-12}	2×10^{-21}	2×10^{-21}	6×10^{-25}
Maximum RMSE	6×10^{-10}	2×10^{-9}	2×10^{-11}	4×10^{-14}	7×10^{-15}	5×10^{-16}
95% interval	$[1 \times 10^{-12}, 1 \times 10^{-11}]$	$[3 \times 10^{-12}, 9 \times 10^{-12}]$	$[3 \times 10^{-12}, 6 \times 10^{-12}]$	$[4 \times 10^{-18}, 2 \times 10^{-14}]$	$[5 \times 10^{-19}, 2 \times 10^{-15}]$	$[2 \times 10^{-20}, 2 \times 10^{-16}]$
$\widehat{\nu}_1$ or $\widehat{\nu}$						
Mean RMSE	26	26	26	2×10^{-1}	5×10^{-2}	1×10^{-2}
Minimum RMSE	25	25	25	2×10^{-5}	0	1×10^{-7}
Maximum RMSE	27	27	27	8×10^{-1}	2×10^{-1}	7×10^{-2}
95% interval	[25, 26]	[25, 26]	[25, 26]	$[6 \times 10^{-3}, 5 \times 10^{-1}]$	$[3 \times 10^{-3}, 2 \times 10^{-1}]$	$[5 \times 10^{-4}, 4 \times 10^{-2}]$

The boundaries of the 95% interval estimates are defined by the empirical 2.5 and 97.5 percentiles
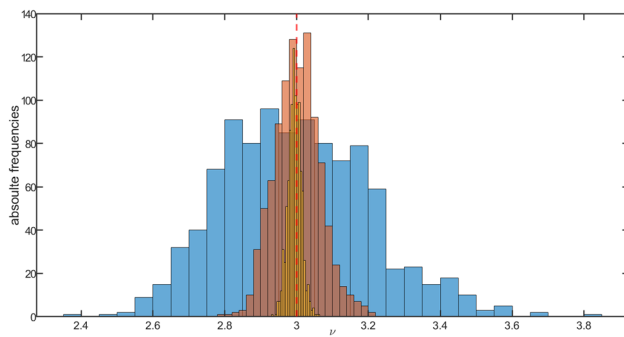
Fig. 1 Histograms of the estimated *df* of the model B2 for $n = 1000$ (blue), for $n = 10,000$ (orange) and for $n = 100,000$ (yellow)

Table 2 Rejection rate of the null hypothesis that the decorrelation-filtered residuals $\hat{\mathbf{u}}_1, \ldots, \hat{\mathbf{u}}_n$ are uncorrelated both throughout time and across the three time series

	n	AN	BN	AT	BT
Reweighted	1000	0.036	0.039	0.042	0.073
	10,000	0.039	0.050	0.060	0.076
	100,000	0.046	0.048	0.063	0.078
Unweighted	1000	0.037	0.040	0.034	0.081
	10,000	0.039	0.049	0.051	0.087
	100,000	0.046	0.048	0.047	0.075

physically relevant processes, knowledge of the noise characteristics of the GPS coordinate time series is required. This includes a description of their temporal correlations, the omission of which affects the uncertainty of the estimated quantities, leading to an unrealistic estimation of precision (Mao et al. 1999; Altamimi et al. 2002; Dong et al. 2006; Santamaria-Gomez et al. 2011; Klos et al. 2018). Cross-correlations between time series, sometimes called spatial correlations, should also be considered, particularly for regional networks (Williams et al. 2004). This kind of correlation is due, e.g., to atmospheric effects and mismodeling of satellite orbits or Earth Orientation Parameters (Klos et al. 2015a). A reduction of their impact can be achieved by the use of stacking filtering techniques (Wdowinski et al. 1997). Another approach is to model the temporal and spatial correlations within the framework of a multivariate regression model. Amiri-Simkooei (2009) made use of the least-squares variance component estimation (LS-VCE) to assess the noise characteristics of daily GPS time series of permanent stations, restricting attention to the estimation of a combination of white and flicker noise. Whereas many empirical studies confirmed this modeling approach (e.g. Amiri-Simkooei et al. 2007; Teferle et al. 2008; Kenyeres and Bruyninx 2009), other studies found the noise as being a combination of white and different power-law noises: white, flicker and/or random-walk or fractional Brownian motion (see, e.g., Mao et al. 1999; Klos et al. 2015b). Spectral indices estimated by MLE assuming normally distributed noise (Langbein and Johnson 1997; Williams et al. 2004) were shown to depend on the stations or the component under consideration (North, East, Up, abbreviated in the following by N, E and U, respectively).

Nearly all previous studies intending to fit a deterministic AR model to the GPS coordinate time series neglected the potential cross-correlations (Didova et al. 2016) and used a simplified AR model. The strength of our methodology, as developed in the previous sections, is to allow specifically for their estimation. We follow thus the results of previous noise analysis and model the noise of GPS coordinate time series as being a combination of a white and a correlated

noise. To take potential cross-correlations into account, a VAR model is used, whose coefficients are estimated alongside functional parameters such as the intercept, trend and the amplitudes of harmonic functions of the GPS position time series. As the outlier characteristics are expected to be heterogeneous, stochastic model A is applied, which allows for individual variance components and *df*s of the univariate t-distributions associated with the different time series. As we do not estimate a fractal dimension by setting up an autoregressive fractionally integrated moving average (ARFIMA) model (see Hosking 1981), which is beyond the scope of the present paper, a direct comparison between the results obtained with the VAR and power law approaches to describe the noise structure can hardly be done. We believe, however, that our method allows for an alternative, complementary determination of the noise structure with respect to the usual method based on spectral analysis, by achieving simultaneously a robust estimation of the parameters of the deterministic observation model. To achieve a robust adjustment of the GPS data, t-distributions are used for two reasons: Firstly, the t-distributions are used as outlier distributions to deal with outliers in the GPS time series. To the best knowledge of the authors, there currently is no other modeling framework that incorporates a general deterministic model, multiple outliers as well as a VAR (auto- and cross-correlation) model. Secondly, there exists an intimate connection of t-distributions with the well-known Mátern covariance model, which has been successfully employed in the stochastic modeling of GPS time series (e.g., Kermarrec and Schön 2014). Using the concept of scale invariance, Schön et al. (2018) show the relationship between the t-distribution model and processes having a power spectral density, such as correlated processes from a Mátern model. Both models share the property of being heavy tailed. Moreover, whereas most of the previous studies concentrate on the analysis of the noise of GPS coordinate time series of one station (univariate model), we make use of a multivariate regression. Auto-correlations as well as cross-correlations between GPS coordinate time at different stations can be easily estimated in one step and if needed, accounted for in

Fig. 2 Location of the four stations under consideration with their mutual distances

further calculations. Specific statistical tests for multivariate observations are further used to judge the goodness of the decorrelation.

5.2 Description of the dataset

The multivariate regression is applied to the GPS daily coordinates time series provided by the EUREF Permanent Network[4] (EPN) (cf. Bruyninx et al. 2012). For the sake of simplicity and without loss of generality, we restrict our analysis to the N component. The aim of this case study is not to provide a detailed and general description of the noise structure of GPS position time series, but to highlight a possible usage of the proposed algorithm. Further contributions will analyze more systematically how the correlated noise of the N, E and U components or the velocity can be modeled by means of a VAR process. Four stations located in the Czech Republic are considered in the following. Their respective locations are shown in Fig. 2. We intentionally selected three stations (CPAR, GOPE and CTAB, station A) close to each other, thereby creating a small network with a maximum distance between stations of approximately 100 km. The station KRAW (station B of the EPN network) is 400 km away from the other stations and allows us to analyze medium-distance cross-correlations. A total of 171.85 weeks (i.e., approximately 3 years) of solution were randomly downloaded, starting at GPS second of week 1824 and ending at 1995.857. We followed thus the recommendation of Bos et al. (2010) to avoid an absorbed correlated noise content in estimating trends of time series and unreliable seasonal signals (cf. Blewitt and Lavallée 2002). In order to test the feasibility of the proposed adjustment technique, neither outliers were removed nor data gaps were interpolated. Instead, missing solutions were filled with random numbers, which we generated using the MATLAB func-

4 http://www.epncb.oma.be/_networkdata/stationlist.php.

tion trnd as following a t-distribution with df $\nu = 5$ and scaling $\sigma = 1$ mm. The latter value is thus larger than the estimated standard deviation (0.57 mm) of the white noise components of the time series. Figure 2 shows the four stations under consideration. Although it may seem crude to insert—usually undesired—outliers instead of interpolating the missing values, our robust multivariate regression approach will downweight the outlying observations within the IRLS algorithm and thus yield parameter estimates that are affected little by irregular observations. It could be further shown that outliers with higher or lower standard deviation (10 mm and 0.5 mm were tested) did not change the VAR coefficients by more than 1 percent, indicating a high robustness of the algorithm. The corresponding results are not presented for the sake of brevity.

5.3 Functional model

We describe the deterministic part of the observations as a combination of a mean, a linear trend and periodic effects. The vector for the $N = 4$ groups of observations is expressed as $\ell_t = [\ell_{\mathrm{GOPE},t}, \ell_{\mathrm{KRAW},t}, \ell_{\mathrm{CPAR},t}, \ell_{\mathrm{CTAB},t}]^T$. Following Klos et al. (2015a), we estimate an annual and a semiannual period, i.e., 365 and 365/2 days. Since the estimation of lower periodicities (cf. Amiri-Simkooei et al. 2007) did not influence the coefficients of the VAR process more than 2 percent, they will not be considered in this contribution. We acknowledge that further investigations with more observations are needed to decide whether additional frequency parameters should be estimated and to adapt their determination to station specificities. This is, however, beyond the scope of the present paper on robust multivariate regression. Interested readers should see Klos et al. (2018) for an analysis of the impact of misspecifying the functional model on velocity estimates, as well as Santamaria-Gomez et al. (2011) for a detailed description of the frequencies present in GPS coordinate time series. Consequently, we simplify the model chosen by Amiri-Simkooei (2009) and express the functional model for each time series as

$$h_{k,t}(\boldsymbol{\xi}) = a_0 + a_1 t + \sum_{j=1}^{2} \left[A_j \cos \omega_j t + B_j \sin \omega_j t \right] \quad (79)$$

where a_0 is the intercept and a_1 the slope or trend, which can be interpreted as the permanent GPS station velocity (Klos et al. 2015b). A_j and B_j are the coefficients with respect to frequency $f_j = 2\pi \omega_j$, where ω_j is the angular velocity. As mentioned previously, two periods were considered as having a significant amplitude, which correspond to $\omega_1 = \frac{2\pi}{365}$ and $\omega_2 = \frac{2\pi}{365/2}$. Thus, the chosen functional model contains six parameters to estimate. Here, we apply stochastic model A (Sect. 2.1), where white noise components are considered
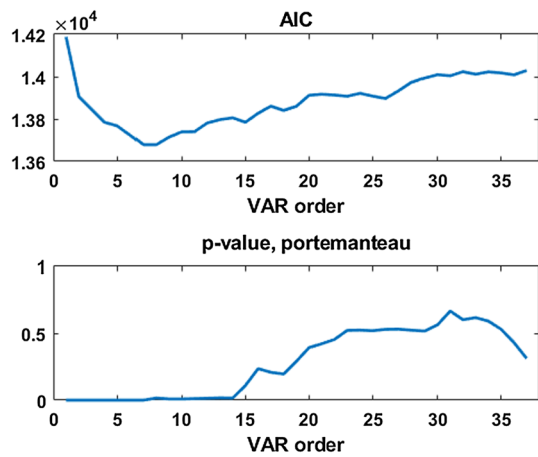
Fig. 3 Results of the AIC and portmanteau test (p value) obtained for VAR model orders $p = 1, \ldots, 38$
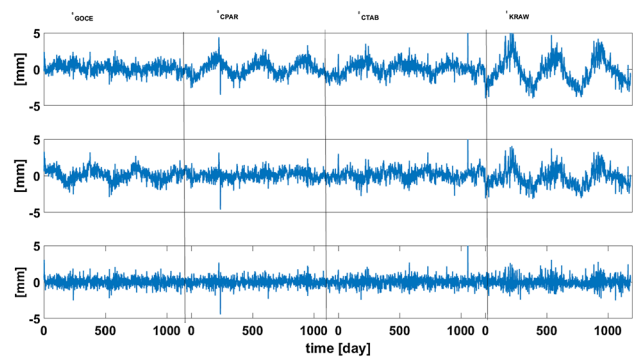

Fig. 4 Original observations (top), colored noise residuals (middle) and white noise residuals (bottom); the four segments of these time series are sorted per stations versus time in day

to be independently t-distributed, which is justified by the assumption that the noise is station specific.

5.4 Determination of the vector-autoregressive model order

5.4.1 Application of Akaike's Information Criterion

The order p of the VAR model to be estimated was determined by means of the Akaike information criteria based on the log-likelihood function L_A with respect to the univariate t-distributions, as presented in Sect. 3.3.2. For this purpose, the order was varied from $p = 1$ up to $p = 38$ by steps of 1, and the AIC as well as AICC value was computed for each model. The results are presented in Fig. 3 (top) only for the AIC since the AICC gave very similar values. Using proposed functional model (79), the order $p = 8$ is clearly identified as the optimal one since the AIC is a minimum there. It should be mentioned that the identified order may vary depending on the modeling of the periodic components, in particular on the number and values of the modeled frequencies. A similar interaction between a Fourier model and the characteristics of a fitted univariate AR model was observed in Kargoll et al. (2018a, Sect. 5). The reason for this is the AR model's capacity to capture certain systematic effects not included in the deterministic model.

5.4.2 Application of the portmanteau test

The multivariate portmanteau test as presented in Sect. 3.3.1 was applied to the adjustment results obtained for the $N = 4$ stations and VAR model orders $p = 1, \ldots, 38$, as was done for the AIC. The significance level was defined as $\alpha = 0.05$, and the maximum lag to be considered was set to $h = 20$. Due to the lack of practical rec-

ommendations in the literature, these somewhat arbitrary choices were adopted for use in MATLAB's Ljung–Box test, which corresponds to the univariate portmanteau test. Then, to obtain more detailed information, the p values resulting from the applications of the multivariate portmanteau test were computed. Since the test is one-sided in the sense that the null hypothesis of white noise is rejected if the value of test statistic (66) exceeds the critical value of the Chi-square distribution with $df\ N^2(h - p)$, the p value (denoted by pv below) is defined as the probability that the portmanteau test statistic exceeds the computed value of the test statistic. This probability can be computed by

$$pv = 1 - F_{\chi^2(4^2 \cdot (20-p))}(P). \tag{80}$$

where P is the value of test statistic (66) and $F_{\chi^2(4^2 \cdot (20-p))}$ the cumulative distribution function of the Chi-square distribution with $df\ 4^2 \cdot (20 - p)$. A p value below α indicates that this test statistic exceeds the critical value (and vice versa). Thus, the p value pv indicates the strength of evidence in favor of the white noise hypothesis, which hypothesis is rejected in case of $pv < \alpha\ (= 0.05)$. Figure 3 (bottom) shows that the p value pv increases sharply between the VAR model orders $p = 14$ and $p = 15$, taking values $pv < 0.05$ for $p < 15$ and $pv > 0.05$ for $p \geq 15$. Thus, $p = 15$ constitutes the least VAR model order for which the portmanteau test accepts the white noise hypothesis. A p value $pv < 0.015$ is obtained for the optimal model order of 8 found by means of the AIC, so that both model selection procedures prefer quite different orders. This difference could be linked with an inaccuracy of the functional model. Indeed, modeling the greater number of coefficients may be an overfitting of the VAR model and an indication of the presence of nonperiodic signals or/and amplitude-varying harmonic components, as displayed by the estimated white noise time series in Fig. 4. As low-order models should be preferred for the description of VAR processes, we consider a VAR
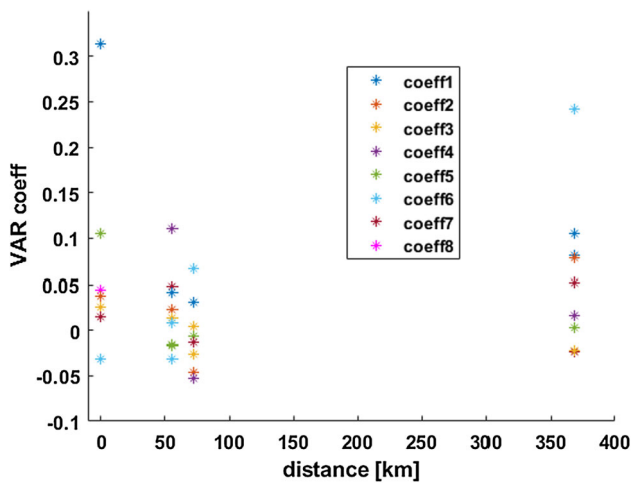
Fig. 5 Estimated VAR coefficients are sorted per distance for the time series of station GOPE and represented by stars with different colors
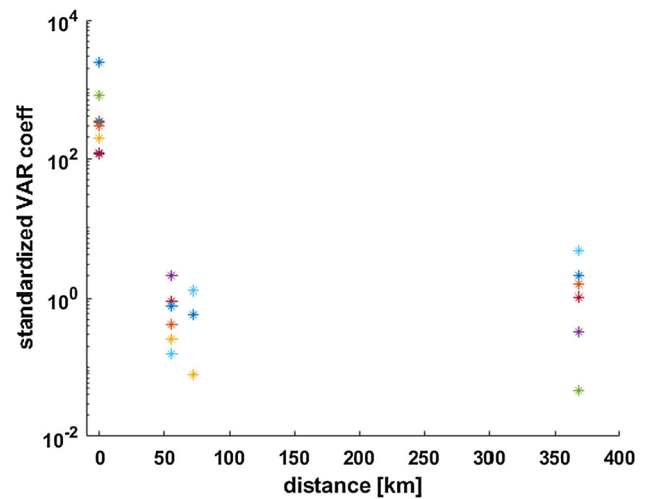

Fig. 6 Normalized VAR coefficients (log plot)

model order of $p = 8$ as more adequate to describe the noise structure.

5.5 Results

5.5.1 Vector-autoregressive coefficients

Figure 5 depicts the 32 coefficients of the fitted VAR(8) model, which describe the correlation structure of the time series of the station GOPE as well the cross-correlations between GOPE and the three other time series (stations CPAR, CTAB and KRAW, respectively). The plotted coefficients, which are sorted by distance, generally decrease in value with increasing lag $j = 1, \ldots, 8$. However, no clear physically explainable dependency on distance can be deduced from Fig. 5. The coefficients become smaller as the distance between stations grows, although the sixth coefficient for GOPE-KRAW is surprisingly larger than the other one. This pattern was confirmed for the other stations (not presented here for the sake of brevity). This large coefficient may be the reason why an optimal VAR model order of $p = 8$ was found with the AIC. The first VAR coefficient is greater than the other one (distance 0), i.e., current values depend more strongly on the previous one than on higher time-lagged values. This dependency does not hold true for the cross-correlation coefficients. Figure 6 shows the standardized coefficients, i.e., the coefficients divided by their estimated standard deviations. The latter have been extracted after execution of the final EM iteration step from the inverted normal equation matrix in (48) multiplied by the corresponding variance component estimate. They clearly have greater values for the correlation of the GOPE time series (distance 0), which are determined with a lower variance than the cross-correlation coefficients.
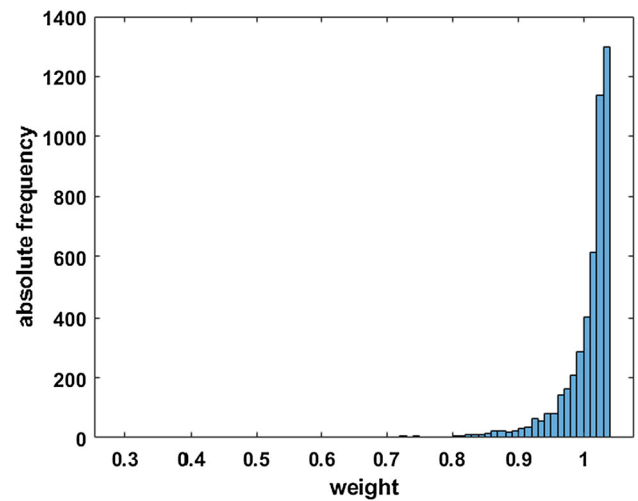

Fig. 7 Histogram plot of the weights

5.5.2 Degree of freedom and weights

The *df* of the t-distribution was found to reach 30 for all time series, indicating the presence of a low level of heavy-tailedness or outlier contamination. The histogram of the adjusted weights is presented in Fig. 7. As found in Koch and Kargoll (2013) and Kargoll et al. (2018a), the weights increase rather smoothly from smaller to larger values. The white-noise components associated with small weights are located in the tails of the t-distribution, so that the corresponding observations could be labeled as outliers.

5.5.3 Interpretation of the residuals

The colored noise residuals $\widehat{\mathbf{e}}_1, \ldots, \widehat{\mathbf{e}}_n$ and the white noise residuals $\widehat{\mathbf{u}}_1, \ldots, \widehat{\mathbf{u}}_n$ are plotted together with the original observations in Fig. 4. As already mentioned in Sect. 5.4.2,

Table 3 Ratio of standard deviations of the white and correlated noise for the four time series analyzed

station	GOPE	CPAR	CTAB	KRAW
σ_u/σ_e	0.68	0.51	0.52	0.57

some oscillations which do not correspond to physically plausible ones remain in these residuals (see, e.g., Santamaria-Gomez et al. 2011). These harmonics may stem from periodic loading or multipath effects that (i) cannot be modeled with the chosen sine and cosine functions or (ii) have varying amplitudes. The residuals of KRAW, classified as "B station" by the EPN clearly show that the low quality of the observations is transferred into the residuals. In such a case the VAR coefficients take the role of "waste containers" in the sense that they express residual deterministic effects rather than the actual auto- and cross-correlations, so that the coefficients lose their physical meaning as noise. As an example, Stralkowski et al. (1970) show how some periodic signals can be modeled by an AR(2) process. Alternatively, a periodic autoregressive moving average (ARMA) model could be used, or an ARFIMA model to take the fractal dimension additionally into consideration. This problem may be the subject of further research, although such fine differences are mostly interesting for the purpose of forecasting, which is not a goal of our study. Time-varying (V)AR coefficients to account for the time variability of the correlations (cf. Kargoll et al. 2018b) is another option to investigate, particularly for longer coordinate time series.

5.5.4 Ratio white noise to correlated noise

The ratios of the standard deviation of the white-noise residuals to the standard deviation of the correlated noise residuals are given in Table 3. For station GOPE, this ratio is slightly larger than for the other stations. This result matches the conclusions obtained by Amiri-Simkooei (2009) for the ratio of flicker noise to white-noise amplitude. The power law description of the noise and its AR modeling both have advantages and shortcomings: Whereas the power laws may be difficult to estimate accurately, the AR coefficients by themselves are physically hardly interpretable. We believe that the two approaches are complementary and benefit from each other for a better comprehension of the correlation structure. Further analysis should be performed in that direction.

6 Conclusions and outlook

Two structurally different stochastic outlier models for multivariate time series involving generally nonlinear deterministic model functions and a VAR model have been introduced. The first stochastic model is based on univariate-scaled t-distributions for the different time series (groups) and allows for group-specific *df* (i.e., outlier characteristics), but not for correlations between white-noise variables at the same time instance. Thus, it is recommended in situations where time series from independent, heterogeneous sensors are combined. The second model is based on a multivariate-scaled t-distribution and allows for the modeling of correlations between white-noise variables at the same time instance, but not for group-specific *df*. This model in turn may be useful when time series from different, but interconnected sensor components are fused. These two stochastic frameworks are therefore viewed as complementary. Two corresponding GEM algorithms were derived, which show that the estimation of the VAR model is based on normal equations that are structurally similar to those for the estimation of a single or multiple independent AR model(s). The extension of the univariate to the multivariate scaled t-distribution does not complicate the algorithm significantly, either.

The MC closed-loop simulation demonstrated a fast and stable convergence of the proposed GEM algorithms for a nonlinear observation model. Furthermore, it was shown that the bias decreased for the majority of estimated parameters when the number of observations is increased. The accurate estimation of the *df* of the model involving univariate t-distributions remains a challenging task to be investigated in future research. Moreover, it was demonstrated that the empirical size of the standard multivariate portmanteau test using the Chi-square distribution is reasonably close to the specified significance level. As this test allows for an evaluation of the fitted VAR model's capacity to fully de-correlate the colored-noise residuals, it may also serve as a tool for selecting an adequate VAR model order, as an alternative to Akaike's information criterion derived for the two stochastic models considered.

The GEM algorithm derived for the stochastic model based on univariate-scaled t-distributions and group-specific *df* was tested on observations from the EPN Network for a particular case study with four stations. We applied the AIC, which is based directly on the maximum likelihood estimates obtained by means of the GEM algorithm, to identify the relatively most adequate order of the VAR model to model the auto- and cross-correlation structure of the GPS coordinate time series. We also employed the multivariate portmanteau test to evaluate the VAR noise models of different orders absolutely, by testing whether the decorrelation of the estimated colored-noise residuals by means of the inverted VAR models resulted in white noise or not. Thus, a VAR model of order 8 was found to be relatively optimal by the AIC, whereas an order of 15 was preferred by the portmanteau test. These findings highlight

that an AR(1) model, as often assumed, is suboptimal to describe the correlation structure. However, no physically explainable dependency of the coefficient values regarding the distance between stations could be found, a finding we link to the respective quality of the observations. The presented approach is an alternative to the usual power-law noise description. It would, therefore, be very valuable to compare the performance of these different noise models in detail. Such comparisons were beyond the scope of our present contribution since it is focused on the methodology of stochastic Student-VAR modeling, which is general enough to be applicable to different types of geodetic time series.

Further investigations to improve the modeling of the noise structure could include, e.g., ARFIMA or periodic ARMA models. Additionally, the impact of the stochastic model chosen on velocity estimates should be investigated. Since these time series have gaps, it would appear to be a useful further step to extend the proposed GEM algorithms to handle such missing observations as well.

A The multivariate, scaled (Student's) t-distribution

If a random vector $\boldsymbol{\mathcal{X}} = [\mathcal{X}_1 \ \ldots \ \mathcal{X}_N]^T$ follows a multivariate, scaled Student's t-distribution with $\nu > 0$ degrees of freedom, location parameter vector $\boldsymbol{\mu} \in \mathbb{R}^N$ and positive definite (N-by-N) cofactor matrix $\boldsymbol{\Sigma}$, symbolically $\boldsymbol{\mathcal{X}} \sim t_\nu(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, then the density of $\boldsymbol{\mathcal{X}}$ is given by

$$f_{\boldsymbol{\mathcal{X}}}(\mathbf{x}) = \frac{\Gamma\left(\frac{\nu+N}{2}\right)}{(\sqrt{\nu\pi})^N \sqrt{\det \boldsymbol{\Sigma}} \, \Gamma\left(\frac{\nu}{2}\right)} \left(1 + \frac{(\mathbf{x}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})}{\nu}\right)^{-\frac{\nu+N}{2}}, \quad (81)$$

with the gamma function $\Gamma(.)$; here, if $\nu > 1$, then the expectation of $\boldsymbol{\mathcal{X}}$ is defined and equals $\boldsymbol{\mu}$; if $\nu > 2$, then the variance–covariance matrix of $\boldsymbol{\mathcal{X}}$ is defined and equals $\frac{\nu}{\nu-2} \cdot \boldsymbol{\Sigma}$ (cf. Lange et al. 1989). Setting $N = 1$ yields the univariate scaled t-distribution (mentioned in Kargoll et al. 2018a, Appendix A.1).

B Derivations

The equations of this section constitute multivariate extensions of the equations derived in Kargoll et al. (2018a, Appendix C). To demonstrate the nature of these extensions, we maintain the original structure as much as possible.

B.1 A derivation for the E-step

Using the expression for $f_B(\mathbf{u}_t, w_t|\boldsymbol{\theta})$ in (31) and the expression for $f_B(\mathbf{u}_t|\boldsymbol{\theta})$ in (26), one finds

$$\begin{aligned}
f_B(w_t|\mathbf{u}_t; \boldsymbol{\theta}) &= \frac{f_B(\mathbf{u}_t, w_t|\boldsymbol{\theta})}{f_B(\mathbf{u}_t|\boldsymbol{\theta})} \\
&= \frac{\left(\frac{\nu}{2}\right)^{\frac{\nu}{2}}}{\Gamma\left(\frac{\nu}{2}\right)} w_t^{\frac{\nu}{2}-1} \exp\left(-\frac{\nu}{2}w_t\right) \\
&\quad \times \frac{1}{\sqrt{(2\pi)^N \det(\boldsymbol{\Sigma}/w_t)}} \exp\left(-\frac{w_t}{2}\mathbf{u}_t^T \boldsymbol{\Sigma}^{-1}\mathbf{u}_t\right) \\
&\quad \times \left\{\frac{\Gamma\left(\frac{\nu+N}{2}\right)}{(\sqrt{\nu\pi})^N \sqrt{\det \boldsymbol{\Sigma}} \, \Gamma\left(\frac{\nu}{2}\right)} \left(1 + \frac{\mathbf{u}_t^T \boldsymbol{\Sigma}^{-1}\mathbf{u}_t}{\nu}\right)^{-\frac{\nu+N}{2}}\right\}^{-1} \\
&= \frac{[\frac{1}{2}(\nu + \mathbf{u}_t^T \boldsymbol{\Sigma}^{-1}\mathbf{u}_t)]^{\frac{\nu+N}{2}}}{\Gamma\left(\frac{\nu+N}{2}\right)} w_t^{\frac{\nu+N-2}{2}} \\
&\quad \times \exp\left(-\frac{1}{2}(\nu + \mathbf{u}_t^T \boldsymbol{\Sigma}^{-1}\mathbf{u}_t)w_t\right).
\end{aligned} \quad (82)$$

This pdf defines the univariate gamma distribution (cf. Kargoll et al. 2018a, Appendix A.2) with parameter values $a = \frac{\nu+N}{2}$ and $b = \frac{1}{2}(\nu + \mathbf{u}_t^T \boldsymbol{\Sigma}^{-1}\mathbf{u}_t)$.

B.2 CM-step for $\boldsymbol{\xi}$

Concerning stochastic model A, setting the partial derivative of Q-function (35) with respect to ξ_j equal to zero yields

$$
\begin{aligned}
0 &= \frac{\partial}{\partial \xi_j} Q_A(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(s)}) \\
&= -\sum_{k=1}^{N} \frac{1}{2\sigma_k^2} \sum_{t=1}^{n} \widetilde{w}_{k,t}^{(s)} \frac{\partial}{\partial \xi_j} \left[\mathbf{A}_k(L)(\boldsymbol{\ell}_t - \mathbf{h}_t(\boldsymbol{\xi})) \right]^2 \\
&\approx -\sum_{k=1}^{N} \frac{1}{2\sigma_k^2} \sum_{t=1}^{n} \widetilde{w}_{k,t}^{(s)} \frac{\partial}{\partial \xi_j} \left[\mathbf{A}_k(L) \left(\boldsymbol{\ell}_t - \left[\mathbf{h}_t(\widehat{\boldsymbol{\xi}}^{(s)}) \right. \right. \right. \\
&\quad \left. \left. \left. + \frac{\partial \mathbf{h}_t(\boldsymbol{\xi})}{\partial \boldsymbol{\xi}^T} \Big|_{\boldsymbol{\xi}=\widehat{\boldsymbol{\xi}}^{(s)}} \cdot (\boldsymbol{\xi} - \widehat{\boldsymbol{\xi}}^{(s)}) \right] \right) \right]^2 \\
&= -\sum_{k=1}^{N} \frac{1}{2\sigma_k^2} \sum_{t=1}^{n} \widetilde{w}_{k,t}^{(s)} \frac{\partial}{\partial \xi_j} \left[\mathbf{A}_k(L) \left(\boldsymbol{\Delta}\boldsymbol{\ell}_t - \mathbf{X}_t \boldsymbol{\Delta}\boldsymbol{\xi} \right) \right]^2 \\
&= -\sum_{k=1}^{N} \frac{1}{\sigma_k^2} \sum_{t=1}^{n} \widetilde{w}_{k,t}^{(s)} \left[\mathbf{A}_k(L) \left(\boldsymbol{\Delta}\boldsymbol{\ell}_t - \mathbf{X}_t \boldsymbol{\Delta}\boldsymbol{\xi} \right) \right] \\
&\quad \cdot \frac{\partial}{\partial \xi_j} \left[-\mathbf{A}_k(L)(\mathbf{X}_t \boldsymbol{\Delta}\boldsymbol{\xi}) \right]
\end{aligned}
$$

with the incremental observations $\boldsymbol{\Delta}\boldsymbol{\ell}_t$, the incremental parameters $\boldsymbol{\Delta}\boldsymbol{\xi}$ and the Jacobi matrix $\mathbf{X}_t$ given in (38)–(40). Denoting by $\mathbf{X}_{t,:j}$ the jth column of $\mathbf{X}_t$, one then obtains equivalently

$$
\begin{aligned}
0 &= \sum_{k=1}^{N} \frac{1}{\sigma_k^2} \sum_{t=1}^{n} \widetilde{w}_{k,t}^{(s)} \left[\mathbf{A}_k(L)\mathbf{X}_{t,:j} \right] \left[\mathbf{A}_k(L) \left(\boldsymbol{\Delta}\boldsymbol{\ell}_t - \mathbf{X}_t \boldsymbol{\Delta}\boldsymbol{\xi} \right) \right] \\
&= \sum_{k=1}^{N} \frac{1}{\sigma_k^2} \left[\mathbf{A}_k(L)\mathbf{X}_{1,:j} \cdots \mathbf{A}_k(L)\mathbf{X}_{n,:j} \right] \widetilde{\mathbf{W}}_k^{(s)} \begin{bmatrix} \mathbf{A}_k(L) \left(\boldsymbol{\Delta}\boldsymbol{\ell}_1 - \mathbf{X}_1 \boldsymbol{\Delta}\boldsymbol{\xi} \right) \\ \vdots \\ \mathbf{A}_k(L) \left(\boldsymbol{\Delta}\boldsymbol{\ell}_n - \mathbf{X}_n \boldsymbol{\Delta}\boldsymbol{\xi} \right) \end{bmatrix}
\end{aligned}
$$

One can now write all m of these equations as equations system (37).

For stochastic model B, one obtains with the same linearization as in the preceding equations

$$
\begin{aligned}
\mathbf{0} &= \frac{\partial}{\partial \boldsymbol{\xi}} Q_B(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(s)}) \\
&= -\frac{1}{2} \sum_{t=1}^{n} \widetilde{w}_t^{(s)} \frac{\partial}{\partial \boldsymbol{\xi}} [\mathbf{A}(L)(\boldsymbol{\ell}_t - \mathbf{h}_t(\boldsymbol{\xi}))]^T \boldsymbol{\Sigma}^{-1} [\mathbf{A}(L)(\boldsymbol{\ell}_t - \mathbf{h}_t(\boldsymbol{\xi}))] \\
&\approx -\frac{1}{2} \sum_{t=1}^{n} \widetilde{w}_t^{(s)} \frac{\partial}{\partial \boldsymbol{\xi}} [\mathbf{A}(L)(\boldsymbol{\Delta}\boldsymbol{\ell}_t - \mathbf{X}_t \boldsymbol{\Delta}\boldsymbol{\xi})]^T \boldsymbol{\Sigma}^{-1} [\mathbf{A}(L)(\boldsymbol{\Delta}\boldsymbol{\ell}_t - \mathbf{X}_t \boldsymbol{\Delta}\boldsymbol{\xi})] \\
&= -\sum_{t=1}^{n} \widetilde{w}_t^{(s)} [\mathbf{A}(L)(\boldsymbol{\Delta}\boldsymbol{\ell}_t - \mathbf{X}_t \boldsymbol{\Delta}\boldsymbol{\xi})] \boldsymbol{\Sigma}^{-1} \cdot \frac{\partial}{\partial \boldsymbol{\xi}} [-\mathbf{A}(L)(\mathbf{X}_t \boldsymbol{\Delta}\boldsymbol{\xi})] \\
&= \sum_{t=1}^{n} [\mathbf{A}(L)\mathbf{X}_t]^T \left[\widetilde{w}_t^{(s)} \boldsymbol{\Sigma}^{-1} \right] [\mathbf{A}(L)(\boldsymbol{\Delta}\boldsymbol{\ell}_t - \mathbf{X}_t \boldsymbol{\Delta}\boldsymbol{\xi})],
\end{aligned}
$$

which is identical with (58).

B.3 CM-step for $\boldsymbol{\alpha}$

Setting the first partial derivative of the Q-function for stochastic model A with respect to the autoregressive coefficient $\alpha_{h;i,j}$ (with $h \in \{1, \ldots, p\}$ and $i, j \in \{1, \ldots, N\}$) equal to zero gives

$$
\begin{aligned}
0 &= \frac{\partial}{\partial \alpha_{h;i,j}} Q_A(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(s)}) \\
&= -\frac{1}{2\sigma_i^2} \sum_{t=1}^{n} \widetilde{w}_{i,t}^{(s)} \cdot \frac{\partial}{\partial \alpha_{h;i,j}} \left[\mathbf{A}_i(L)(\boldsymbol{\ell}_t - \mathbf{h}_t(\boldsymbol{\xi})) \right]^2 \\
&= -\frac{1}{2\sigma_i^2} \sum_{t=1}^{n} \widetilde{w}_{i,t}^{(s)} \\
&\quad \cdot 2 \left[\mathbf{A}_i(L)(\boldsymbol{\ell}_t - \mathbf{h}_t(\boldsymbol{\xi})) \right] \cdot \frac{\partial}{\partial \alpha_{h;i,j}} \left[\mathbf{A}_i(L)(\boldsymbol{\ell}_t - \mathbf{h}_t(\boldsymbol{\xi})) \right] \\
&= -\frac{1}{\sigma_i^2} \sum_{t=1}^{n} \widetilde{w}_{i,t}^{(s)} \left[\mathbf{i}_i^T \mathbf{e}_t - \mathbf{A}_i^{(1)} \mathbf{e}_{t-1} - \cdots - \mathbf{A}_i^{(p)} \mathbf{e}_{t-p} \right] \\
&\quad \times \frac{\partial}{\partial \alpha_{h;i,j}} \left[\mathbf{i}_i^T \mathbf{e}_t - \mathbf{A}_i^{(1)} \mathbf{e}_{t-1} - \cdots - \mathbf{A}_i^{(p)} \mathbf{e}_{t-p} \right] \\
&= -\frac{1}{\sigma_i^2} \sum_{t=1}^{n} e_{j,t-h} \widetilde{w}_{i,t}^{(s)} (e_{i,t} - \mathbf{A}_i^{(1)} \mathbf{e}_{t-1} - \cdots - \mathbf{A}_i^{(p)} \mathbf{e}_{t-p}) \\
&= -\frac{1}{\sigma_i^2} \left[e_{j,1-h} \cdots e_{j,n-h} \right] \widetilde{\mathbf{W}}_i^{(s)} \begin{bmatrix} e_{i,1} - \mathbf{A}_i^{(1)} \mathbf{e}_0 - \cdots - \mathbf{A}_i^{(p)} \mathbf{e}_{1-p} \\ \vdots \\ e_{i,n} - \mathbf{A}_i^{(1)} \mathbf{e}_{n-1} - \cdots - \mathbf{A}_i^{(p)} \mathbf{e}_{n-p} \end{bmatrix}.
\end{aligned}
$$

Writing these equations for all $j \in \{1, \ldots, N\}$ and all $h \in \{1, \ldots, p\}$ in matrix notation results in the system of $N \cdot p$ equations

$$
\begin{bmatrix} 0 \\ \vdots \\ 0 \\ \vdots \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \begin{bmatrix} e_{1,0} & \cdots & e_{1,n-1} \\ \vdots & & \vdots \\ e_{N,0} & \cdots & e_{N,n-1} \\ \vdots & & \vdots \\ e_{1,1-p} & \cdots & e_{1,n-p} \\ \vdots & & \vdots \\ e_{N,1-p} & \cdots & e_{N,n-p} \end{bmatrix} \widetilde{\mathbf{W}}_i^{(s)} \begin{bmatrix} e_{i,1} - \mathbf{A}_i^{(1)} \mathbf{e}_0 - \cdots - \mathbf{A}_i^{(p)} \mathbf{e}_{1-p} \\ \vdots \\ e_{i,n} - \mathbf{A}_i^{(1)} \mathbf{e}_{n-1} - \cdots - \mathbf{A}_i^{(p)} \mathbf{e}_{n-p} \end{bmatrix}
$$

for every row $i \in \{1, \ldots, N\}$ of the matrices $\mathbf{A}_i^{(1)}, \ldots, \mathbf{A}_i^{(p)}$. These equation systems read, using vector notation,

$$
\begin{bmatrix} \mathbf{0} \\ \vdots \\ \mathbf{0} \end{bmatrix} = \begin{bmatrix} \mathbf{e}_0 & \cdots & \mathbf{e}_{n-1} \\ \vdots & & \vdots \\ \mathbf{e}_{1-p} & \cdots & \mathbf{e}_{n-p} \end{bmatrix} \widetilde{\mathbf{W}}_i^{(s)} \left(\begin{bmatrix} e_{i,1} \\ \vdots \\ e_{i,n} \end{bmatrix} - \begin{bmatrix} \mathbf{e}_0^T & \cdots & \mathbf{e}_{1-p}^T \\ \vdots & & \vdots \\ \mathbf{e}_{n-1}^T & \cdots & \mathbf{e}_{n-p}^T \end{bmatrix} \begin{bmatrix} \mathbf{A}_i^{(1)T} \\ \vdots \\ \mathbf{A}_i^{(p)T} \end{bmatrix} \right).
$$

If the available solution $\boldsymbol{\xi} = \widehat{\boldsymbol{\xi}}^{(s+1)}$ is substituted to compute the residuals, the normal equations

$$
\mathbf{0} = \widehat{\mathbf{E}}^{(s+1)T} \widetilde{\mathbf{W}}_i^{(s)} \left(\widehat{\mathbf{e}}_{i:}^{(s+1)} - \widehat{\mathbf{E}}^{(s+1)} \widehat{\boldsymbol{\alpha}}_{i:}^{(s+1)} \right)
$$

can be found. These are solved for every $i \in \{1, \ldots, N\}$ and result in (48).

For stochastic model B, it is convenient to set

$$
\begin{aligned}
\mathbf{0} &= \frac{\partial}{\partial \mathbf{A}^{(h)}} Q_B(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(s)}) \\
&= -\frac{1}{2} \sum_{t=1}^{n} \widetilde{w}_t^{(s)} \frac{\partial}{\partial \mathbf{A}^{(h)}} [\mathbf{A}(L)\mathbf{e}_t]^T \boldsymbol{\Sigma}^{-1} [\mathbf{A}(L)\mathbf{e}_t] \\
&= -\frac{1}{2} \sum_{t=1}^{n} \widetilde{w}_t^{(s)} \frac{\partial}{\partial \mathbf{A}^{(h)}} [\mathbf{I}\mathbf{e}_t - \mathbf{A}^{(1)}\mathbf{e}_{t-1} - \cdots - \mathbf{A}^{(p)}\mathbf{e}_{t-p}]^T \boldsymbol{\Sigma}^{-1} \\
&\quad \times [\mathbf{I}\mathbf{e}_t - \mathbf{A}^{(1)}\mathbf{e}_{t-1} - \cdots - \mathbf{A}^{(p)}\mathbf{e}_{t-p}] \\
&= -\frac{1}{2} \sum_{t=1}^{n} \widetilde{w}_t^{(s)} \frac{\partial}{\partial \mathbf{A}^{(h)}} \left[\left(\mathbf{e}_t - \sum_{j=1, j \neq h}^{p} \mathbf{A}^{(j)}\mathbf{e}_{t-j} \right) - \mathbf{A}^{(h)}\mathbf{e}_{t-h} \right]^T \boldsymbol{\Sigma}^{-1} \\
&\quad \times \left[\left(\mathbf{e}_t - \sum_{j=1, j \neq h}^{p} \mathbf{A}^{(j)}\mathbf{e}_{t-j} \right) - \mathbf{A}^{(h)}\mathbf{e}_{t-h} \right].
\end{aligned}
$$

Using now Eq. (88) in Petersen and Pedersen (2012) to determine the derivative with respect to the matrix $\mathbf{A}^{(h)}$, one obtains

$$
\begin{aligned}
\mathbf{0} &= -\frac{1}{2} \sum_{t=1}^{n} \widetilde{w}_t^{(s)} \cdot (-2) \\
&\quad \cdot \boldsymbol{\Sigma}^{-1} \left[\left(\mathbf{e}_t - \sum_{j=1, j \neq h}^{p} \mathbf{A}^{(j)}\mathbf{e}_{t-j} \right) - \mathbf{A}^{(h)}\mathbf{e}_{t-h} \right] \mathbf{e}_{t-h}^T \\
&= \sum_{t=1}^{n} \widetilde{w}_t^{(s)} \boldsymbol{\Sigma}^{-1} [\mathbf{A}(L)\mathbf{e}_t] \mathbf{e}_{t-h}^T.
\end{aligned}
$$

After transposition of this equation, one may multiply it from the right with $\boldsymbol{\Sigma}$ and thereby eliminate $\boldsymbol{\Sigma}^{-1}$. Writing now all p of these resulting normal equation systems, one finds

$$
\begin{aligned}
\begin{bmatrix} \mathbf{0} \\ \vdots \\ \mathbf{0} \end{bmatrix} &= \begin{bmatrix} \sum_{t=1}^{n} \widetilde{w}_t^{(s)} \mathbf{e}_{t-1}[\mathbf{A}(L)\mathbf{e}_t]^T \\ \vdots \\ \sum_{t=1}^{n} \widetilde{w}_t^{(s)} \mathbf{e}_{t-p}[\mathbf{A}(L)\mathbf{e}_t]^T \end{bmatrix} \\
&= \begin{bmatrix} \widetilde{w}_1^{(s)}\mathbf{e}_{1-1} \cdots \widetilde{w}_n^{(s)}\mathbf{e}_{n-1} \\ \vdots \qquad \vdots \\ \widetilde{w}_1^{(s)}\mathbf{e}_{1-p} \cdots \widetilde{w}_n^{(s)}\mathbf{e}_{n-p} \end{bmatrix} \begin{bmatrix} [\mathbf{e}_1 - \mathbf{A}^{(1)}\mathbf{e}_{1-1} - \cdots - \mathbf{A}^{(p)}\mathbf{e}_{1-p}]^T \\ \vdots \\ [\mathbf{e}_n - \mathbf{A}^{(1)}\mathbf{e}_{n-1} - \cdots - \mathbf{A}^{(p)}\mathbf{e}_{n-p}]^T \end{bmatrix} \\
&= \begin{bmatrix} \mathbf{e}_0 \cdots \mathbf{e}_{n-1} \\ \vdots \qquad \vdots \\ \mathbf{e}_{1-p} \cdots \mathbf{e}_{n-p} \end{bmatrix} \widetilde{\mathbf{W}}^{(s)} \left(\begin{bmatrix} \mathbf{e}_1^T \\ \vdots \\ \mathbf{e}_n^T \end{bmatrix} - \begin{bmatrix} \mathbf{e}_0^T \cdots \mathbf{e}_{1-p}^T \\ \vdots \qquad \vdots \\ \mathbf{e}_{n-1}^T \cdots \mathbf{e}_{n-p}^T \end{bmatrix} \begin{bmatrix} \mathbf{A}^{(1)T} \\ \vdots \\ \mathbf{A}^{(p)T} \end{bmatrix} \right).
\end{aligned}
$$

Substitution of $\boldsymbol{\xi} = \widehat{\boldsymbol{\xi}}^{(s+1)}$ and the resulting residuals $\widehat{\mathbf{e}}_1^{(s+1)}$, …, $\widehat{\mathbf{e}}_n^{(s+1)}$ yields then after transposition (62).

B.4 CM-step for σ^2 or $\boldsymbol{\Sigma}$

Setting the first partial derivative of the Q-function for stochastic model A with respect to the variance component σ_i^2 for observation group i equal to zero, one finds

$$
\begin{aligned}
0 &= \frac{\partial}{\partial \sigma_i^2} Q_A(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(s)}) \\
&= -\frac{n}{2} \frac{\partial}{\partial \sigma_i^2} \log(\sigma_i^2) - \frac{\partial}{\partial \sigma_i^2} \frac{1}{2\sigma_i^2} \sum_{t=1}^{n} \widetilde{w}_{i,t}^{(s)} \left[\mathbf{A}_i(L)(\ell_{i,t} - h_{i,t}(\boldsymbol{\xi})) \right]^2 \\
&= -\frac{n}{2\sigma_i^2} + \frac{1}{2\sigma_i^4} \sum_{t=1}^{n} \widetilde{w}_{i,t}^{(s)} \left[\mathbf{A}_i(L)(\ell_{i,t} - h_{i,t}(\boldsymbol{\xi})) \right]^2.
\end{aligned}
$$

Conditioning this equation on the previously determined solutions $\widehat{\boldsymbol{\xi}}^{(s+1)}$ and $\widehat{\boldsymbol{\alpha}}^{(s+1)}$, and using the resulting residual estimates $\widehat{\mathbf{e}}_t^{(s+1)}$ as well as $\widehat{\mathbf{u}}_t^{(s+1)}$, we arrive at (50)–(51).

In case of the multivariate t-distribution model, one obtains (using the arguments given in Koch 1999, Sect. 373)

$$
\begin{aligned}
\mathbf{0} &= \frac{\partial}{\partial \boldsymbol{\Sigma}^{-1}} Q_B(\boldsymbol{\theta}|\widehat{\boldsymbol{\theta}}^{(s)}) \\
&= \frac{n}{2} \frac{\partial}{\partial \boldsymbol{\Sigma}^{-1}} (-\log \det \boldsymbol{\Sigma}) - \frac{1}{2} \sum_{t=1}^{n} \widetilde{w}_t^{(s)} \frac{\partial}{\partial \boldsymbol{\Sigma}^{-1}} \mathbf{u}_t^T \boldsymbol{\Sigma}^{-1} \mathbf{u}_t \\
&= \frac{n}{2} \boldsymbol{\Sigma} - \frac{1}{2} \sum_{t=1}^{n} \widetilde{w}_t^{(s)} \mathbf{u}_t \mathbf{u}_t^T.
\end{aligned}
$$

Conditioning on the available residual estimates $\widehat{\mathbf{u}}_t^{(s+1)}$ leads to (63).

B.5 CM-step for ν

Setting the first partial derivative of log-likelihood function (20) under stochastic model A with respect to the degree of freedom ν_i^2 for observation group i equal to zero, we find

$$
\begin{aligned}
0 &= \frac{\partial}{\partial \nu_i} \log L_A(\boldsymbol{\theta}; \boldsymbol{\ell}) \\
&= \frac{n}{2} \psi \left(\frac{\nu_i + 1}{2} \right) - \frac{n}{2} \psi \left(\frac{\nu_i}{2} \right) + \frac{n}{2} (\log \nu_i + 1) \\
&\quad - \frac{1}{2} \sum_{t=1}^{n} \log \left[\nu_i + \left(\frac{\mathbf{A}_i(L)(\ell_{i,t} - h_{i,t}(\boldsymbol{\xi}))}{\sigma_i} \right)^2 \right] \\
&\quad - \frac{1}{2}(\nu_i + 1) \sum_{t=1}^{n} \left[\nu_i + \left(\frac{\mathbf{A}_i(L)(\ell_{i,t} - h_{i,t}(\boldsymbol{\xi}))}{\sigma_i} \right)^2 \right]^{-1}.
\end{aligned}
$$

Conditioning this equation on the solutions $\widehat{\boldsymbol{\xi}}^{(s+1)}$, $\widehat{\boldsymbol{\alpha}}^{(s+1)}$ and $(\widehat{\sigma}_i^2)^{(s+1)}$ (as determined in the previous three CM-steps) and substituting then the estimated residual $\widehat{u}_{i,t}^{(s+1)}$, one obtains (52).

References

Alkhatib H, Kargoll B, Neumann I, Kreinovich V (2017) Normalization-invariant fuzzy logic operations explain empirical success of Student distributions in describing measurement uncertainty. In: Melin P, Castillo O, Kacprzyk J, Reformat M, Melek W (eds) Fuzzy logic in intelligent system design. NAFIPS 2017. Advances in intelligent systems and computing, vol 648. Springer, Cham, pp 300–306. https://doi.org/10.1007/978-3-319-67137-6_34

Alkhatib H, Kargoll B, Paffenholz JA (2018) Further results on robust multivariate time series analysis in nonlinear models with autoregressive and t-distributed errors. In: Valenzuela O, Rojas F, Pomares H, Rojas I (eds) Time series analysis and forecasting. ITISE 2017. Contributions to statistics. Springer, Cham, pp 25–38. https://doi.org/10.1007/978-3-319-96944-2_3

Altamimi Z, Sillard P, Boucher C (2002) ITRF2000: a new release of the International Terrestrial Reference Frame for earth science applications. J Geophys Res 107(B10):2214. https://doi.org/10.1029/2001JB000561

Amiri-Simkooei AR (2009) Noise in multivariate GPS position time-series. J Geod 83(2):175–187. https://doi.org/10.1007/s00190-008-0251-8

Amiri-Simkooei AR, Tiberius CCJM, Teunissen PJG (2007) Assessment of noise in GPS coordinate time series: methodology and results. J Geophys Res. https://doi.org/10.1029/2006JB004913

Blewitt G, Lavallée D (2002) Effect of annual signals on geodetic velocity. J Geophys Res 107(B7):2145. https://doi.org/10.1029/2001JB000570

Bos MS, Bastos L, Fernandes RMS (2010) The influence of seasonal signals on the estimation of the tectonic motion in short continuous GPS time-series. J Geodyn 49:205–209. https://doi.org/10.1016/j.jog.2009.10.005

Box GEP, Pierce DA (1970) Distribution of residual autocorrelations in autoregressive-integrated moving average time series models. J Am Stat Ass 65(332):1509–1526. https://doi.org/10.2307/2284333

Briegel T, Tresp V (1999) Robust neural network regression for offline and online learning. In: NIPS'99 proceedings of the 12th international conference on neural information processing systems, pp 407–413

Brockwell PJ, Davis RA (2016) Introduction to time series and forecasting, 3rd edn. Springer, Berlin

Bruyninx C, Habrich H, Söhne W, Kenyeres A, Stangl G, Völksen C (2012) Enhancement of the EUREF permanent network services and products. In: Kenyon S, Pacino M, Marti U (eds) Geodesy for planet earth. International Association of Geodesy Symposia, vol 136. Springer, Berlin, pp 27–35. https://doi.org/10.1007/978-3-642-20338-1_4

Burnham KP, Anderson DR (2002) Model selection and multimodel inference—a practical information-theoretic approach, 2nd edn. Springer, New York

Didova O, Gunter B, Riva R, Klees R, Roese-Koerner L (2016) An approach for estimating time-variable rates from geodetic time series. J Geod 90:1207–1221. https://doi.org/10.1007/s00190-016-0918-5

Dong D, Fang P, Bock Y, Webb F, Prawirodirdjo L, Kedar S, Jamason P (2006) Spatiotemporal filtering using principal component analysis and Karhunen-Loeve expansion approaches for regional GPS network analysis. J Geophys Res. https://doi.org/10.1029/2005JB003806

Hamilton JD (1994) Time series analysis. Princeton University Press, Princeton

Helmert FR (1907) Die Ausgleichsrechnung nach der Methode der kleinsten Quadrate, 2nd edn. Teubner, Leipzig

Hosking JRM (1980) The multivariate portmanteau statistic. J Am Stat Ass 75(371):602–608. https://doi.org/10.2307/2287656

Hosking JRM (1981) Fractional differencing. Biometrika 68(1):165–176. https://doi.org/10.1093/biomet/68.1.165

Huber PJ, Ronchetti EM (2009) Robust statistics, 2nd edn. Wiley, Hoboken

ISO/IEC (2008) JCGM 100:2008 evaluation of measurement data—guide to the expression of uncertainty in measurement. First edition 2008, corrected version 2010. International Organization for Standardization (ISO), Geneva

Kargoll B, Omidalizarandi M, Loth I, Paffenholz JA, Alkhatib H (2018a) An iteratively reweighted least-squares approach to adaptive robust adjustment of parameters in linear regression models with autoregressive and t-distributed deviations. J Geod 92(3):271–297. https://doi.org/10.1007/s00190-017-1062-6

Kargoll B, Omidalizarandi M, Alkhatib H, Schuh WD (2018b) Further results on a modified EM algorithm for parameter estimation in linear models with time-dependent autoregressive and t-distributed errors. In: Valenzuela O, Rojas F, Pomares H, Rojas I (eds) Time series analysis and forecasting. ITISE 2017, Contributions to statistics. Springer, Cham, pp 323–337. https://doi.org/10.1007/978-3-319-96944-2_22

Kenyeres A, Bruyninx C (2009) Noise and periodic terms in the EPN time series. In: Drewes H (ed) Geodetic reference frames. International Association of Geodesy Symposia, vol 134. Springer, Berlin, pp 143–148. https://doi.org/10.1007/978-3-642-00860-3_22

Kermarrec G, Schön S (2014) On the Mátern covariance family: a proposal for modeling temporal correlations based on turbulence theory. J Geod 88(11):1061–1079. https://doi.org/10.1007/s00190-014-0743-7

Klos A, Bogusz J, Figurski M, Kosek W (2015a) Noise analysis of continuous GPS time series of selected EPN stations to investigate variations in stability of monument types. In: Sneeuw N, Novák P, Crespi M, Sansò F (eds) III Hotine-Marussi symposium on mathematical geodesy. International Association of Geodesy Symposia, vol 142. Springer, Cham, pp 19–26. https://doi.org/10.1007/1345_2015_62

Klos A, Bogusz J, Figurski M, Gruszczynska M, Gruszczynski M (2015b) Investigation of noises in the EPN weekly time series. Acta Geodyn Geomater 12(2):117–126. https://doi.org/10.13168/AGG.2015.0010

Klos A, Olivares G, Teferle FN, Hunegnaw A, Bogusz J (2018a) On the combined effect of periodic signals and colored noise on velocity uncertainties. GPS Solut. https://doi.org/10.1007/s10291-017-0674-x

Koch KR (1999) Parameter estimation and hypothesis testing in linear models. Springer, Berlin

Koch KR (2014) Outlier detection for the nonlinear Gauss Helmert model with variance components by the expectation maximization algorithm. J Appl Geod 8(3):185–194. https://doi.org/10.1515/jag-2014-0004

Koch KR (2017) Expectation maximization algorithm and its minimal detectable outliers. Stud Geophys Geod 61:1–18. https://doi.org/10.1007/s11200-016-0617-y

Koch KR, Kargoll B (2013) Expectation-maximization algorithm for the variance-inflation model by applying the t-distribution. J Appl Geod 7:217–225. https://doi.org/10.1515/jag-2013-0007

Koch KR, Kuhlmann H, Schuh WD (2010) Approximating covariance matrices estimated in multivariate models by estimated auto- and cross-covariances. J Geod 84:383–397. https://doi.org/10.1007/s00190-010-0375-5

Koop G (2003) Bayesian econometrics. Wiley, Hoboken

Kuhlmann H (2001) Importance of autocorrelation for parameter estimation in regression models. In: Proceedings of the 10th FIG international symposium on deformation measurements. FIG, pp 354–361

Langbein J, Johnson H (1997) Correlated errors in geodetic time series: implications for time-dependent deformation. J Geophys Res 102(B1):591–603. https://doi.org/10.1029/96JB02945

Lange KL, Little RJA, Taylor JMG (1989) Robust statistical modeling using the t-distribution. J Am Stat Assoc 84:881–896. https://doi.org/10.2307/2290063

Lira I, Wöger W (2006) Comparison between the conventional and Bayesian approaches to evaluate measurement data. Metrologia 43:S249–S259

Liu CH, Rubin DB (1994) The ECME algorithm: a simple extension of EM and ECM with faster monotone convergence. Biometrika 81:633–648. https://doi.org/10.1093/biomet/81.4.633

Liu Y, Sang R, Liu S (2017) Diagnostic analysis for a vector autoregressive model under Student's t-distributions. Stat Neerl 71(2):86–114. https://doi.org/10.1111/stan.12102

Luo X, Mayer M, Heck B (2011) Verification of ARMA identification for modelling temporal correlations of GNSS observations using the ARMASA toolbox. Stud Geophys Geod 55:537–556. https://doi.org/10.1007/s11200-011-0033-2

Mao A, Harrison CGA, Dixon TH (1999) Noise in GPS coordinate time series. J Geophys Res 104(B2):2797–2816. https://doi.org/10.1029/1998JB900033

Matthewman NJ, Magnusdottir G (2011) Observed interaction between pacific sea ice and the Western Pacific pattern on intraseasonal time scales. J Clim 24:5031–5042. https://doi.org/10.1175/2011JCLI4216.1

Matthewman NJ, Magnusdottir G (2012) Clarifying ambiguity in intraseasonal Southern Hemisphere climate modes during austral winter. J Geophs Res 117:D03105. https://doi.org/10.1029/2011JD016707

McLachlan GJ, Krishnan T (2008) The EM algorithm and extensions, 2nd edn. Wiley, Hoboken

Meng X-L, Rubin DB (1993) Maximum likelihood estimation via the ECM algorithm: a general framework. Biometrika 80:267–278

Mosedale TJ, Stepehnson DB, Collins M, Mills TC (2006) Granger causality of coupled climate processes: Ocean feedback on the North Atlantic oscillation. J Clim 19:1182–1194. https://doi.org/10.1175/JCLI3653.1

Nassar S, Schwarz KP, El-Sheimy N, Noureldin A (2004) Modeling inertial sensor errors using autoregressive (AR) models. Navigation 51:259–268. https://doi.org/10.1002/j.2161-4296.2004.tb00357.x

Nduka UC, Iwueze IS, Nwogu EC (2018) Efficient algorithms for robust estimation in autoregressive regression models using student's t distribution. Commun Stat Simul Comput. https://doi.org/10.1080/03610918.2018.1484481

Niedzielski T, Kosek W (2012) Prediction analysis of UT1-UTC time series by combination of the least-squares and multivariate autoregressive method. In: International Association of Geodesy Symposia, vol 137. Springer, pp 153–157. https://doi.org/10.1007/978-3-642-22078-4_23

Paffenholz JA (2012) Direct geo-referencing of 3D point clouds with 3D positioning sensors. Deutsche Geodätische Kommission, Series C (Dissertations), No 689, Munich

Papagiannopoulou C, Miralles DG, Decubber S, Demuzere M, Verhoest NEC, Dorigo WA, Waegeman W (2017) A non-linear Granger-causality framework to investigate climate-vegetation dynamics. Geosci Model Dev 10:1945–1960. https://doi.org/10.5194/gmd-10-1945-2017

Park M, Gao Y (2008) Error and performance analysis of MEMS-based inertial sensors with a low-cost GPS receiver. Sensors 8(4):2240–2261. https://doi.org/10.3390/s8042240

Parzen E (1979) A density-quantile function perspective on robust estimation. In: Launer L, Wilkinson GN (eds) Robustness in statistics. Academic Press, New York, pp 237–258. https://doi.org/10.1016/B978-0-12-438150-6.50019-4

Petersen KB, Pedersen MS (2012) The matrix cookbook. Version November 15, 2012, Technical University of Denmark. http://www2.imm.dtu.dk/pubdb/views/edoc_download.php/3274/pdf/imm3274.pdf. Accessed 3 April 2020

Phillips RF (2002) Least absolute deviations estimation via the EM Algorithm. Stat Comput 12(3):281–285. https://doi.org/10.1023/A:1020759012226

Santamaria-Gomez A, Bouin MN, Collilieux X, Woppelmann G (2011) Correlated errors in GPS position time series: implications for velocity estimates. J Geophys Res. https://doi.org/10.1029/2010JB007701

Schön S, Kermarrec G, Kargoll B, Neumann I, Kosheleva O, Kreinovich V (2018) Why student distributions? Why Mátern's covariance model? A symmetry-based explanation. In: Ly AH, Le DS, Kreinovich V, Nguyen TN (eds) ECONVN 2018: econometrics for financial applications. Studies in computational intelligence, vol 760. Springer, Cham, pp 266–275. https://doi.org/10.1007/978-3-319-73150-6_21

Schuh W-D, Krasbutter I, Kargoll B (2014) Korrelierte Messungen - Was nun? In: Neuner H (ed) Zeitabhängige Meßgrößen - Ihre Daten haben (Mehr-)Wert DVW Schriftenreihe 74, pp 85–101

Siemes C (2013) Digital filtering algorithms for decorrelation within large least squares problems. Deutsche Geodätische Kommission, Series C (Dissertations), No 707, Munich

Sommer KD, Siebert BRL (2004) Praxisgerechtes Bestimmen der Messunsicherheit nach GUM. Tech Mess 71:52–66

Stralkowski CM, Wu SM, DeVor RE (1970) Charts for the interpretation and estimation of the second order autoregressive model. Technometrics 12(3):669–685. https://doi.org/10.1080/00401706.1970.10488705

Strong C, Magnusdottir G, Stern H (2009) Observed feedback between winter sea ice and the North Atlantic Oscillation. J Clim 22:6021–6032. https://doi.org/10.1175/2009JCLI3100.1

Teferle FN, Williams SDP, Kierulf HP, Bingley RM, Plag HP (2008) A continuous GPS coordinate time series analysis strategy for high-accuracy vertical land movements. Phys Chem Earth 33(3–4):205–216. https://doi.org/10.1016/j.pce.2006.11.002

Tienstra JM (1947) An extension of the technique of the methods of least squares to correlated observations. Bull Geod 6(1):301–335. https://doi.org/10.1007/BF02525959

Tuaç Y, Güney Y, Şenoğlu B, Arslan O (2018) Robust parameter estimation of regression model with AR(p) error terms. Commun Stat Simul Comput. https://doi.org/10.1080/03610918.2017.1343839

Wallace JM, Gutzler DS (1981) Teleconnections in the geopotential height field during the northern hemisphere winter. Mon Weather Rev 109(4):784–812. 10.1175/1520-0493(1981)109<0784:TITGHF>2.0.CO;2

Wdowinski S, Bock Y, Zhang J, Fang P, Genrich J (1997) Southern California permanent GPS geodetic array: spatial filtering of daily positions for estimating coseismic and postseismic displacements induced by the 1992 Landers earthquake. J Geophys Res 102(B8):18,057–18,070. https://doi.org/10.1029/97JB01378

Westfall PH (2014) Kurtosis as peakedness, 1905–2014 R.I.P. Am Stat 68(3):191–195. https://doi.org/10.1080/00031305.2014.917055

Williams SDP, Bock Y, Fang P, Jamason P, Nikolaidis RM, Prawirodirdjo L, Miller M, Johnson DJ (2004) Error analysis of continuous GPS position time series. J Geophys Res 109:B03412. https://doi.org/10.1029/2003JB002741

Wiśniewski Z (2009) *M*-estimation with probabilistic models of geodetic observations. J Geod 88(10):941–957. https://doi.org/10.1007/s00190-014-0735-7

Xu PL (1988) Application of time series methods in prediction of displacements on large dams. J Wuhan Tech Univ Surv Mapp 13(3):23–31

Xu PL (2005) Sign-constrained robust least squares, subjective breakdown point and the effect of weights of observations on robustness. J Geod 79:146–159. https://doi.org/10.1007/s00190-005-0454-1

Zeng W, Fang X, Lin Y, Huang X, Zhou Y (2018) On the total least-squares estimation for autoregressive model. Surv Rev 50(359):186–190. https://doi.org/10.1080/00396265.2017.1281096

Zhang Q, Gui Q (2015) Carrier-phase RAIM algorithm based on a vector autoregressive model. In: Sun J, Liu J, Fan S, Lu X (eds) China Satellite Navigation Conference (CSNC) 2015 Proceedings: volume II. Lecture Notes in Electrical Engineering, vol 341. Springer, Berlin, pp 125–142. https://doi.org/10.1007/978-3-662-46635-3_11