Robust Multivariate Time Series Analysis in Nonlinear Models with Autoregressive and t-Distributed Errors

Hamza Alkhatib, Boris Kargoll, and Jens-André Paffenholz

Leibniz Universität Hannover, Geodätisches Institut
Nienburger Str. 1, 30167 Hannover, Germany
{alkhatib,kargoll,paffenholz}@gih.uni-hannover.de
https://www.gih.uni-hannover.de

Abstract. We study a time series model which can generally be described as the additive combination of a multivariate, nonlinear deterministic model with multiple univariate, covariance-stationary autoregressive (AR) processes whose white noise components follow independent scaled t-distributions. These distributions allow for the stochastic modeling of heavy tails or multiple outliers and provide the framework for a partially adaptive, robust maximum likelihood (ML) estimation of the deterministic model parameters, of the AR coefficients, of the scale parameters, and of the degrees of freedom of the underlying t-distributions. To obtain the ML estimator, we derive a generalized expectation maximization (GEM) algorithm, which takes the form of linearized, iteratively reweighted least squares. The performance of this estimator is evaluated by means of a Monte Carlo simulation for the observations of a circle in three dimensions, involving different noise models encountered typically in the analysis of global navigation satellite system (GNSS) time series.

Keywords: multivariate time series, nonlinear regression model, AR process, scaled t-distribution, partially adaptive estimation, robust parameter estimation, GEM algorithm, GNSS time series

1 Introduction

Robust estimation is important in many fields of application where the probability density function (pdf) of the random deviations is expected to be heavy-tailed (e.g., as a consequence of multiple outliers). [5] was an early exposition demonstrating the use and usefulness of the scaled (Student) t-distribution in robust maximum likelihood (ML) estimation for regression models. As already indicated by [1], this kind of ML estimation can be expressed in a computationally convenient form as iteratively reweighted least squares, where the weights are used to rescale the variances of the random deviations according to their locations under the pdf. It is possible with this approach to estimate the degree of freedom of the underlying t-distribution, alongside the regression parameters and the scale parameter, turning it into a so-called (partially) adaptive estimator.

In a multivariate regression model, each observable is modeled as a random vector which is explained by a vector-valued (possibly non-linear) deterministic regression function and a vector of random deviations. [6] assumed a multivariate t-distribution with unknown scale factor and unknown degree of freedom for each vector of random deviations and investigated different forms of the expectation maximization (EM) algorithm for the purpose of estimating the unknown model parameters. It was shown earlier in [7] and [10] that the expectation conditional maximization (ECM) and the expectation conditional maximization either (ECME) variants can speed up the convergence of the EM algorithm considerably. To handle models that do not allow for closed form solutions by EM, the optimization principle of generalized expectation maximization (GEM) was proposed by [1]. The idea is to approach the maximum expectation within each EM step rather than trying to reach it fully. GEM algorithms employing Newton-Raphson steps have been applied frequently [9]. A GEM algorithm can in particular be used to handle non-linear regression models. In this situation, an iteratively reweighted least squares algorithm with Gauss-Newton steps was found to be a suitable form of GEM [14, 5].

Besides heavy tails, multivariateness and non-linearity, a further aspect that complicates (partially adaptive) parameter estimation consists in the frequently encountered autocorrelatedness of the random deviations. For instance, many types of sensor data such as inertial sensor data, satellite gravity gradiometry data and GNSS data give measurement results where the random deviations exhibit pronounced colored noise characteristics (see, e.g., [17, 13, 16, 8]). Typically, such datasets contain numerous outliers, so that robust estimation approach is generally desirable. To deal with situation, the aforementioned partially adaptive estimator for regression models based on the scaled t-distribution was extended in [4] to include autoregressive (AR) random deviations, where the white noise components of the AR process are independently and identically t-distributed. A limitation of that method is however that the observables describe only a univariate time series involving a linear regression model.

The purpose of this contribution is to extend the existing univariate, linear model to a multivariate and nonlinear (differentiable) regression model. Concerning the setup of the AR model, we currently limit ourselves to the case where each time series component is associated with a univariate AR process of individual order, independently of the AR processes of the other components. We thus exclude the modeling of cross-correlations, a task which would require the use of vector AR (VAR) processes and which is beyond the scope of the present contribution. The paper is organized as follows.

First, the time series model is described in detail in Sect. 2, and the derivation of a corresponding GEM algorithm is outlined in Sect. 3. Here, it is shown on the one hand how the scaled t-distributions are taken into account within the E step. On the other hand, the linearization of the nonlinear deterministic model is demonstrated in connection with the M step, which is broken up into conditional maximization steps with respect to the different groups of estimated model parameters. In Sect. 4, a time series model for GNSS observations of a circle in 3D is proposed, and the results of a Monte Carlo simulation as well as real world application based on this observation model are discussed. These findings are used to evaluate the performance of the implemented GEM algorithm in this scenario.

2 The Observation Model

We consider q-dimensional observables $\mathbf{Y}_t = [Y_{1,t} \cdots Y_{N,t}]^T$ measured at equidistant time instances $t = 1, \ldots, n$. The task is to approximate the corresponding measurement results $\mathbf{y}_1, \ldots, \mathbf{y}_n$ by a (vector-valued) nonlinear function $\mathbf{h}_t(\boldsymbol{\xi}) = [h_{1,t}(\boldsymbol{\xi}) \cdots h_{N,t}(\boldsymbol{\xi})]^T$ of unknown parameters $\boldsymbol{\xi} = [\xi_1, \ldots, \xi_m]^T$. We model the uncertainties of the measurement process by means of random deviations $\mathbf{E}_t = [E_{1,t} \cdots E_{N,t}]^T$ between the observables and the functional model, so that the observation equations take the form

$$\mathbf{Y}_t = \mathbf{h}_t(\boldsymbol{\xi}) + \mathbf{E}_t \qquad (t = 1, \ldots, n). \tag{1}$$

Here, we assume that each of the N components of the random deviations is subject to autocorrelations in the form of a covariance-stationary autoregressive (AR) model

$$E_{k,t} = \alpha_{k,1} E_{k,t-1} + \ldots + \alpha_{k,p_k} E_{k,t-p_k} + U_{k,t} \qquad (k = 1, \ldots, N; t = 1, \ldots, n), \tag{2}$$

in which the random variables $U_{k,1}, \ldots, U_{k,n}$ are, for every $k = 1, \ldots, N$, independently and identically t-distributed according to

$$U_{k,t} \sim t_{\nu_k}(0, \sigma_k^2) \qquad (k = 1, \ldots, N; t = 1, \ldots, n). \tag{3}$$

Thus, we allow each white noise series $U_{k,1}, \ldots, U_{k,n}$ to have an individual fluctuation and tail behavior, as determined by the component-dependent scale parameter σ_k^2 and degree of freedom ν_k. These quantities, alongside the AR coefficients, are considered as additional unknowns to be estimated jointly with the functional parameters $\boldsymbol{\xi}$. The probability density function (pdf) of the scaled t-distributed white noise components $U_{k,t}$ is thus defined by

$$f(u_{k,t}) = \frac{\Gamma\left(\frac{\nu_k+1}{2}\right)}{\sqrt{\nu_k \pi \sigma_k^2}\,\Gamma\left(\frac{\nu_k}{2}\right)} \left[1 + \left(\frac{u_{k,t}}{\sigma_k}\right)^2 / \nu_k\right]^{-\frac{\nu_k+1}{2}} \tag{4}$$

(where Γ represents the gamma function). The preceding assumption of stochastic independence of the white noise components $\mathbf{u}_k = [u_{k,1} \cdots u_{k,n}]^T$ for each $k = 1, \ldots, N$ implies that their joint pdf is given by

$$f(\mathbf{u}_k) = \prod_{t=1}^{n} f(u_{k,t}) = \prod_{t=1}^{n} \frac{\Gamma\left(\frac{\nu_k+1}{2}\right)}{\sqrt{\nu_k \pi \sigma_k^2}\,\Gamma\left(\frac{\nu_k}{2}\right)} \left[1 + \left(\frac{u_{k,t}}{\sigma_k}\right)^2 / \nu_k\right]^{-\frac{\nu_k+1}{2}}. \tag{5}$$

We assume that no stochastic dependencies between the N white noise series exist, so that the joint pdf of the white noise components throughout all series can be written in the factorized form $f(\mathbf{u}) = f(\mathbf{u}_1)\cdots f(\mathbf{u}_N)$. This implies also that the N colored noise processes (2) can be treated separately. Note that we generally allow these AR processes to have different orders $p_1,\ldots,p_N$. Since we intend to apply the preceding model to rather large time series (with n being at least 100), we deal with the initialization problem of the AR processes in a practical manner, by setting all quantities occurring at time instances $t = 0,-1,\ldots$ equal to 0. Moreover, we assume all AR processes to be invertible, so that we can rewrite them in the form

$$U_{k,t} = E_{k,t} - \alpha_{k,1}E_{k,t-1} - \ldots - \alpha_{k,p_k}E_{k,t-p_k} = \boldsymbol{\alpha}_k(L)E_{k,t}, \tag{6}$$

using the lag operator $L^jE_t := E_{t-j}$ and the lag polynomial $\boldsymbol{\alpha}_k(L) := 1-\alpha_{k,1}L-\ldots-\alpha_{k,p_k}L^{p_k}$. We can interpret the latter as a digital filter, which decorrelates the colored noise series $e_{k,1},\ldots,e_{k,n}$ (into the white noise series $u_{k,1},\ldots,u_{k,n}$).

A maximum likelihood estimation of the unknown model parameters $\boldsymbol{\xi}$, $\boldsymbol{\alpha}_1$, …, $\boldsymbol{\alpha}_N$, $\sigma_1^2,\ldots,\sigma_N^2$ and $\nu_1,\ldots,\nu_N$ based on the pdf $f(\mathbf{u})$ or its natural logarithm

$$\log\mathcal{L}(\boldsymbol{\theta};\mathbf{y}) = \log f(\mathbf{u}) = \log\left[f(\mathbf{u}_1)\cdots f(\mathbf{u}_N)\right] = \sum_{k=1}^{N}\left(n\log\left[\frac{\Gamma\left(\frac{\nu_k+1}{2}\right)}{\sqrt{\nu_k\pi\sigma_k^2}\Gamma\left(\frac{\nu_k}{2}\right)}\right]\right.$$
$$\left.-\frac{\nu_k+1}{2}\sum_{t=1}^{n}\log\left[1+\left(\frac{\boldsymbol{\alpha}_k(L)(y_{k,t}-h_{k,t}(\boldsymbol{\xi}))}{\sigma_k}\right)^2/\nu_k\right]\right) \tag{7}$$

and given measurement results $\mathbf{y}$ requires numerical optimization since a closed-form expression of the estimator is unavailable. Following the ideas of [1] and [5], we transform the preceding t-distribution observation model into an easier-to-manage form by introducing latent variables

$$W_{k,t}\,|\,\boldsymbol{\xi},\sigma_k^2,\boldsymbol{\alpha}_k,\nu_k \sim \frac{\chi^2_{\nu_k}}{\nu_k} = G\left(\frac{\nu_k}{2},\frac{\nu_k}{2}\right) \qquad (k=1,\ldots,N;t=1,\ldots,n), \tag{8}$$

where the gamma distribution is defined by the pdf

$$f(w_{k,t}|\boldsymbol{\theta}) = \begin{cases} \frac{\left(\frac{\nu_k}{2}\right)^{\nu_k/2}}{\Gamma\left(\frac{\nu_k}{2}\right)}\cdot(w_{k,t})^{\nu_k/2-1}\cdot e^{-\nu_k/2\cdot w_{k,t}} & \text{if } \ w_{k,t}>0, \\ 0 & \text{if } \ w_{k,t}\leq 0 \end{cases} \tag{9}$$

(using $\boldsymbol{\theta}$ for convenience as the vector consisting of all the unknown model parameters). These variables are assumed to be stochastically independent within each series, resulting in the factorization $f(\mathbf{w}_k|\boldsymbol{\theta}) = \prod_{t=1}^{n} f(w_{k,t}|\boldsymbol{\theta})$. The idea is now to define further stochastic properties of the white noise $\mathbf{U}_k = [U_{k,1}\cdots U_{k,n}]^T$ and the latent variables $\mathbf{W}_k = [W_{k,1}\cdots W_{k,n}]^T$ in such a way that the Student pdf (5) is obtained as the marginal distribution from the joint pdf $f(\mathbf{u}_k,\mathbf{w}_k|\boldsymbol{\theta})$ (cf. [9]). This is achieved on the one hand by employing the conditional Gaussian

$$f(u_{k,t}|w_{k,t},\boldsymbol{\theta}) = \frac{1}{\sqrt{2\pi(\sigma_k/\sqrt{w_{k,t}})^2}}\exp\left\{-\frac{u_{k,t}^2}{2(\sigma_k/\sqrt{w_{k,t}})^2}\right\}. \tag{10}$$

On the other hand, $\mathbf{U}_{k,t}$ is assumed to be independent of the white noise components and latent variables occurring within the series k at the other time instants $1,\ldots,t-1,t+1,\ldots,n$ and within the other series $1,\ldots,k-1,k+1,\ldots,N$ at all time instances, conditional on the values $w_{k,t}$ and $\boldsymbol{\theta}$. This conditional independence assumption allows us to apply for instance the simplification

$$\begin{aligned} &f(u_{k,t}|u_{k,1},w_{k,1}\ldots,u_{k,t-1},w_{k,t-1},u_{k,t+1},w_{k,t+1},\ldots,u_{k,n},w_{k,n},w_{k,t},\boldsymbol{\theta}) \\ &\quad = f(u_{k,t}|w_{k,t},\boldsymbol{\theta}) \end{aligned} \tag{11}$$

in the derivation of the desired joint pdf (similarly to the proof in [2])

$$\begin{aligned} f(\mathbf{u},\mathbf{w}|\boldsymbol{\theta}) &= \prod_{t=1}^{n} f(u_{1,t},w_{1,t}|\boldsymbol{\theta})\cdots\prod_{t=1}^{n} f(u_{N,t},w_{N,t}|\boldsymbol{\theta}) \\ &= \prod_{t=1}^{n} f(w_{1,t}|\boldsymbol{\theta})\, f(u_{1,t}|w_{1,t},\boldsymbol{\theta})\cdots\prod_{t=1}^{n} f(w_{N,t}|\boldsymbol{\theta})\, f(u_{N,t}|w_{N,t},\boldsymbol{\theta}). \end{aligned} \tag{12}$$

We define this be the likelihood function $\mathcal{L}(\boldsymbol{\theta};\mathbf{y},\mathbf{w})$ of the extended observation model. Before proceeding with the corresponding ML estimation, we note that the second factor in $f(w_{k,t},u_{k,t}|\boldsymbol{\theta}) = f(u_{k,t}|\boldsymbol{\theta})\, f(w_{k,t}|u_{k,t}|\boldsymbol{\theta})$ defines the conditional gamma distribution $G(a,b)$ with parameters $a = (\nu_k+1)/2$ and $b = (\nu_k + u_{k,t}^2/\sigma_k^2)/2$, given the value $u_{k,t}$ (applying a proof in analogy to [3]).

3 The Generalized EM Algorithm

In view of (12), (9) and (10), the log-likelihood function takes the form

$$\begin{aligned} \log\mathcal{L}(\boldsymbol{\theta};\mathbf{y},\mathbf{w}) &= \text{const.} - \frac{n}{2}\sum_{k=1}^{N}\log(\sigma_k^2) + \frac{n}{2}\sum_{k=1}^{N}\nu_k\log\left(\frac{\nu_k}{2}\right) - n\sum_{k=1}^{N}\log\Gamma\left(\frac{\nu_k}{2}\right) \\ &- \sum_{k=1}^{N}\sum_{t=1}^{n}\frac{1}{2}\left[\nu_k + \left(\frac{\boldsymbol{\alpha}_k(L)(y_{k,t}-h_k(\boldsymbol{\xi}))}{\sigma_k}\right)^2\right] w_{k,t} + \sum_{k=1}^{N}\sum_{t=1}^{n}\frac{1}{2}(\nu_k-1)\log w_{k,t}. \end{aligned} \tag{13}$$

To set up the generalized EM (GEM) algorithm, we define the Q-function as the conditional expectation of the preceding log-likelihood function (treated now as a random function), given measurement results $\mathbf{y}$ and trial parameter values $\boldsymbol{\theta}^{(i)}$, in the sense of

$$Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(i)}) = E_{\mathbf{W}|\mathbf{y};\boldsymbol{\theta}^{(i)}}\left\{\log\mathcal{L}\left(\boldsymbol{\theta};\mathbf{y},\mathbf{W}\right)\right\}. \tag{14}$$

3.1 The E Step

Recalling that the likelihood function was defined by (12), we condition directly on the white noise outcome $\mathbf{u}$ and on $\boldsymbol{\theta}^{(i)}$ (which values give $\mathbf{y}$ through the equations (1) and (2)). Then, (13) yields

$$\begin{aligned}
Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(i)}) &= \text{const.} - \frac{n}{2}\sum_{k=1}^{N}\log(\sigma_k^2) + \frac{n}{2}\sum_{k=1}^{N}\nu_k\log\left(\frac{\nu_k}{2}\right) - n\sum_{k=1}^{N}\log\Gamma\left(\frac{\nu_k}{2}\right)\\
&- \sum_{k=1}^{N}\sum_{t=1}^{n}\frac{1}{2}\left[\nu_k + \left(\frac{\boldsymbol{\alpha}_k(L)(y_{k,t}-h_{k,t}(\boldsymbol{\xi}))}{\sigma_k}\right)^2\right]E_{\mathbf{W}|\mathbf{u};\boldsymbol{\theta}^{(i)}}\{W_{k,t}\}\\
&+ \sum_{k=1}^{N}\sum_{t=1}^{n}\frac{1}{2}(\nu_k-1)E_{\mathbf{W}|\mathbf{u};\boldsymbol{\theta}^{(i)}}\{\log W_{k,t}\}. \qquad (15)
\end{aligned}$$

Here, we observe in light of [2] that the two conditional expectations simplify to

$$\begin{aligned}
E_{\mathbf{W}|\mathbf{u};\boldsymbol{\theta}^{(i)}}\{W_{k,t}\} &= E_{W_{k,t}|u_{k,t};\boldsymbol{\theta}^{(i)}}\{W_{k,t}\},\\
E_{\mathbf{W}|\mathbf{u};\boldsymbol{\theta}^{(i)}}\{\log W_{k,t}\} &= E_{W_{k,t}|u_{k,t};\boldsymbol{\theta}^{(i)}}\{\log W_{k,t}\}.
\end{aligned}$$

Since the latent variable $W_{k,t}$ given the value $u_{k,t}$ follows the gamma distribution $G(a,b)$, the previous two expectations are, respectively, a/b and $\psi(a)-\log(b)$ (where ψ is the digamma function), so that we obtain (cf. [3] for details)

$$w_{k,t}^{(i)} := E_{W_{k,t}|u_{k,t};\boldsymbol{\theta}^{(i)}}\{W_{k,t}\} = \frac{\nu_k^{(i)}+1}{\nu_k^{(i)} + \left(\frac{\boldsymbol{\alpha}_k^{(i)}(L)(y_{k,t}-h_{k,t}(\boldsymbol{\xi}^{(i)}))}{\sigma^{(i)}}\right)^2}, \qquad (16)$$

$$E_{W_{k,t}|u_{k,t};\boldsymbol{\theta}^{(i)}}\{\log W_{k,t}\} = \log w_{k,t}^{(i)} + \psi\left(\frac{\nu_k^{(i)}+1}{2}\right) - \log\left(\frac{\nu_k^{(i)}+1}{2}\right). \qquad (17)$$

Consequently, we may rewrite (15) as

$$\begin{aligned}
Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(i)}) &= \text{const.} - \frac{n}{2}\sum_{k=1}^{N}\log(\sigma_k^2) - \sum_{k=1}^{N}\frac{1}{2\sigma_k^2}\sum_{t=1}^{n}w_{k,t}^{(i)}\left[\boldsymbol{\alpha}_k(L)(y_{k,t}-h_{k,t}(\boldsymbol{\xi}))\right]^2\\
&+\frac{n}{2}\sum_{k=1}^{N}\nu_k\log\nu_k - n\sum_{k=1}^{N}\log\Gamma\left(\frac{\nu_k}{2}\right)\\
&+\frac{n}{2}\sum_{k=1}^{N}\nu_k\left[\psi\left(\frac{\nu_k^{(i)}+1}{2}\right) - \log\left(\nu_k^{(i)}+1\right) + \frac{1}{n}\sum_{t=1}^{n}\left(\log w_{k,t}^{(i)} - w_{k,t}^{(i)}\right)\right]. \qquad (18)
\end{aligned}$$

We see in light of (16) that the computation of initial weights requires initial parameter values. In cases where these are not given, we choose unit weights $w_{k,t}^{(0)} = 1$ for all $k = 1,\ldots,N$ and all $t = 1,\ldots,n$ for the subsequent M step.

3.2 The M Step

We break up the M step into four conditional maximization (CM) steps (see [10]), one for each of the parameter groups, and substituting the most recent available estimates whenever needed. Since the regression function $h_{k,t}$ were assumed to be nonlinear functions of $\boldsymbol{\xi}$, it is linearized within the first CM-Step with respect to that parameter group. Choosing for the Taylor point the estimate $\boldsymbol{\xi}^{(i)}$ of the preceding iteration step, we obtain for the partial derivative of the Q-function with respect to ξ_j

$$\begin{aligned}
0 = \frac{\partial}{\partial \xi_j} Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(i)}) &= -\sum_{k=1}^{N} \frac{1}{2\sigma_k^2} \sum_{t=1}^{n} w_{k,t}^{(i)} \frac{\partial}{\partial \xi_j} \left[\boldsymbol{\alpha}_k(L)(y_{k,t} - h_{k,t}(\boldsymbol{\xi}))\right]^2 \\
&= -\sum_{k=1}^{N} \frac{1}{2\sigma_k^2} \sum_{t=1}^{n} w_{k,t}^{(i)} \frac{\partial}{\partial \xi_j} \left[\boldsymbol{\alpha}_k(L) \left(y_{k,t} - \left[h_{k,t}(\boldsymbol{\xi}^{(i)}) + \frac{\partial h_{k,t}(\boldsymbol{\xi}^{(i)})}{\partial \boldsymbol{\xi}} (\boldsymbol{\xi} - \boldsymbol{\xi}^{(i)}) \right] \right) \right]^2 \\
&= -\sum_{k=1}^{N} \frac{1}{2\sigma_k^2} \sum_{t=1}^{n} w_{k,t}^{(i)} \frac{\partial}{\partial \xi_j} \left[\boldsymbol{\alpha}_k(L) \left(\Delta y_{k,t} - \mathbf{A}_{k,t}^{(i)} \boldsymbol{\Delta\xi} \right) \right]^2,
\end{aligned}$$

where $\Delta y_{k,t} = y_{k,t} - h_{k,t}(\boldsymbol{\xi}^{(i)})$, $\boldsymbol{\Delta\xi} = \boldsymbol{\xi} - \boldsymbol{\xi}^{(i)}$, and $\mathbf{A}_{k,t}^{(i)} = \frac{\partial h_{k,t}(\boldsymbol{\xi}^{(i)})}{\partial \boldsymbol{\xi}}$. Denoting in addition $\mathbf{A}_{k,t,j}^{(i)} = \frac{\partial h_{k,t}(\boldsymbol{\xi}^{(i)})}{\partial \xi_j}$ and forming also the diagonal matrix $\mathbf{W}^{(i)}$ from the values $w_1^{(i)}, \ldots, w_n^{(i)}$, we can derive the system of m equations

$$\mathbf{0} = \sum_{k=1}^{N} \frac{1}{\sigma_k^2} \begin{bmatrix} \boldsymbol{\alpha}_k(L) A_{k,1,1}^{(i)} & \cdots & \boldsymbol{\alpha}_k(L) A_{k,n,1}^{(i)} \\ \vdots & & \vdots \\ \boldsymbol{\alpha}_k(L) A_{k,1,m}^{(i)} & \cdots & \boldsymbol{\alpha}_k(L) A_{k,n,m}^{(i)} \end{bmatrix} \mathbf{W}_k^{(i)} \begin{bmatrix} \boldsymbol{\alpha}_k(L)(\Delta y_{k,1} - \mathbf{A}_{k,1}^{(i)} \boldsymbol{\Delta\xi}) \\ \vdots \\ \boldsymbol{\alpha}_k(L)(\Delta y_{k,n} - \mathbf{A}_{k,n}^{(i)} \boldsymbol{\Delta\xi}) \end{bmatrix}.$$

Fixing now the values of the unknown scale parameters and AR coefficients by taking the estimates from the preceding M step i, we can filter the reduced observations and the Jacobi matrices (for every $k = 1, \ldots, N$ and every $t = 1, \ldots, n$) according to

$$\overline{\Delta y}_{k,t}^{(i)} := \boldsymbol{\alpha}_k^{(i)}(L) \Delta y_{k,t}, \qquad \overline{A}_{k,t,j}^{(i)} := \boldsymbol{\alpha}_k^{(i)}(L) A_{k,t,j}, \qquad \overline{\mathbf{A}}_{k,t}^{(i)} := \boldsymbol{\alpha}_k^{(i)}(L) \mathbf{A}_{k,t} \quad (19)$$

and restate the preceding normal equation system as

$$\begin{aligned}
\mathbf{0} &= \sum_{k=1}^{N} \frac{1}{(\sigma_k^2)^{(i)}} \begin{bmatrix} \overline{A}_{k,1,1}^{(i)} & \cdots & \overline{A}_{k,n,1}^{(i)} \\ \vdots & & \vdots \\ \overline{A}_{k,1,m}^{(i)} & \cdots & \overline{A}_{k,n,m}^{(i)} \end{bmatrix} \mathbf{W}_k^{(i)} \begin{bmatrix} \overline{\Delta y}_{k,1} - \overline{\mathbf{A}}_{k,1} \boldsymbol{\Delta\xi} \\ \vdots \\ \overline{\Delta y}_{k,n} - \overline{\mathbf{A}}_{k,n} \boldsymbol{\Delta\xi} \end{bmatrix} \\
&= \sum_{k=1}^{N} \frac{1}{(\sigma_k^2)^{(i)}} \overline{\mathbf{A}}_k^{(i)} \mathbf{W}_k^{(i)} \left(\overline{\boldsymbol{\Delta}\mathbf{y}}_k - \overline{\mathbf{A}}_k^{(i)} \boldsymbol{\Delta\xi} \right).
\end{aligned}$$

Consequently, the estimate of the update $\boldsymbol{\Delta\xi}$ is given by

$$\boldsymbol{\Delta\xi}^{(i+1)} = \left(\sum_{k=1}^{N} \frac{1}{(\sigma_k^2)^{(i)}} (\overline{\mathbf{A}}_k^{(i)})^T \mathbf{W}_k^{(i)} \overline{\mathbf{A}}_k^{(i)} \right)^{-1} \sum_{k=1}^{N} \frac{1}{(\sigma_k^2)^{(i)}} (\overline{\mathbf{A}}_k^{(i)})^T \mathbf{W}_k^{(i)} \overline{\boldsymbol{\Delta}\mathbf{y}}^{(i)}. \quad (20)$$

This update is added entirely or partially to the trial solution (in the sense of a Gauss-Newton step with step size $\gamma \in (0, 1]$), resulting in

$$\boldsymbol{\xi}^{(i+1)} = \boldsymbol{\xi}^{(i)} + \gamma \boldsymbol{\Delta\xi}^{(i+1)}. \tag{21}$$

In the first iteration step, we would typically use unit weight matrices $\mathbf{W}_k^{(0)} = \mathbf{I}_n$, neutral filters $\boldsymbol{\alpha}_k^{(0)}(L) = 1$ and identity scale factors $(\sigma_k^2)^{(i)} = 1$, corresponding to the initial assumption of normally distributed, uncorrelated and homoskedastic white noise components throughout all time series. For the subsequent CM step with respect to the autoregressive coefficients, the colored noise residuals

$$e_{k,t}^{(i+1)} := y_{k,t} - h_{k,t}(\boldsymbol{\xi}^{(i+1)}) \qquad (k = 1, \ldots, N; t = 1, \ldots, n). \tag{22}$$

will play a central role. We assemble for this purpose the matrices

$$\mathbf{E}_k^{(i+1)} := \begin{bmatrix} e_{k,0}^{(i+1)} & \cdots & e_{k,1-p_k}^{(i+1)} \\ \vdots & & \vdots \\ e_{k,n-1}^{(i+1)} & \cdots & e_{k,n-p_k}^{(i+1)} \end{bmatrix} \qquad (k = 1, \ldots, N), \tag{23}$$

in which we substitute the initial values $e_{k,0}^{(i+1)} = \ldots = e_{k,1-p_k}^{(i+1)} = 0$. Setting now the first partial derivative of (18) with respect to the jth AR coefficient within the Kth time series equal to zero, we have

$$0 = \frac{\partial}{\partial \alpha_{K,j}} Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(i)}) = -\frac{\partial}{\partial \alpha_{K,j}} \sum_{k=1}^{N} \frac{1}{2\sigma_k^2} \sum_{t=1}^{n} w_{k,t}^{(i)} \left[\boldsymbol{\alpha}_k(L)(y_{k,t} - h_{k,t}(\boldsymbol{\xi}))\right]^2$$
$$= -\frac{1}{2\sigma_K^2} \sum_{t=1}^{n} w_{K,t}^{(i)} \frac{\partial}{\partial \alpha_{K,j}} \left[\boldsymbol{\alpha}_K(L) e_{K,t}\right]^2 .$$

Substituting for the unknowns $\boldsymbol{\xi}$ within the residual $e_{K,t}$ the already available estimates $\boldsymbol{\xi}^{(i+1)}$ (according to the principle of conditional maximization) and collecting all j partial derivative with respect to the Kth time series in a single equation system, we obtain then for every $K = 1, \ldots, N$ the iteratively reweighted least squares scheme for the estimation of the AR coefficients $\boldsymbol{\alpha}_K$

$$\boldsymbol{\alpha}_K^{(i+1)} = \left((\mathbf{E}_K^{(i+1)})^T \mathbf{W}_K^{(i)} \mathbf{E}_K^{(i+1)}\right)^{-1} (\mathbf{E}_K^{(i+1)})^T \mathbf{W}_K^{(i)} \mathbf{e}_K^{(i+1)}. \tag{24}$$

Since we aim for covariance-stationary and invertible AR processes, it is necessary to determine whether all roots of $\boldsymbol{\alpha}_K^{(i+1)}(z) = 0$ are located within the unit circle. In case this is not true, we stabilize the preceding polynomial by mirroring all roots with magnitude exceeding 1 into the unit circle (cf. [15]), using MATLAB's `polystab` routine. We see from (24) that the individual AR processes can be determined independently, and we use them to filter the colored noise residuals according to (6) through

$$u_{k,t}^{(i+1)} = \boldsymbol{\alpha}_k^{(i+1)}(L) e_{k,t}^{(i+1)} \qquad (k = 1, \ldots, N; t = 1, \ldots, n) \tag{25}$$

in order to obtain the estimated white noise residuals. We are now in a position to estimate within the third CM step each scale factor σ_K^2 via the N independent conditions

$$
\begin{aligned}
0 &= \frac{\partial}{\partial \sigma_K^2} Q(\boldsymbol{\theta}|\boldsymbol{\theta}^{(i)}) \\
&= -\frac{n}{2}\frac{\partial}{\partial \sigma_K^2}\log(\sigma_K^2) - \frac{\partial}{\partial \sigma_K^2}\frac{1}{2\sigma_K^2}\sum_{t=1}^{n} w_{K,t}^{(i)}\left[\boldsymbol{\alpha}_K(L)(y_{K,t} - h_{K,t}(\boldsymbol{\xi}))\right]^2 ,
\end{aligned}
$$

in which we substitute the current estimates $\boldsymbol{\xi}^{(i+1)}$ and $\boldsymbol{\alpha}_K^{(i+1)}$. Making use of (25), we therefore arrive at the solutions

$$
(\sigma_K^2)^{(i+1)} = \frac{1}{n}\sum_{t=1}^{n} w_{K,t}^{(i)}\left(u_{K,t}^{(i+1)}\right)^2 = \frac{(\mathbf{u}_K^{(i+1)})^T \mathbf{W}_K^{(i)} \mathbf{u}_K^{(i+1)}}{n}. \tag{26}
$$

It remains for us to compute the solutions for the degrees of freedom of the t-distributions underlying the N time series. Instead of using the Q-function for this purpose, we follow the recommendation of [7] and maximize the log-likelihood function (7) with respect to these parameters (which turns the current ECM algorithm into an ECME algorithm). Using the digamma function ψ, it can be shown that

$$
\begin{aligned}
0 = \frac{\partial}{\partial \nu_K}\log \mathcal{L}(\boldsymbol{\theta};\mathbf{y}) &= \frac{n}{2}\psi\left(\frac{\nu_K+1}{2}\right) - \frac{n}{2}\psi\left(\frac{\nu_K}{2}\right) + \frac{n}{2}(\log \nu_K + 1) \\
&-\frac{1}{2}\sum_{t=1}^{n}\log\left[\nu_K + \left(\frac{\boldsymbol{\alpha}_K(L)(y_{K,t} - h_{K,t}(\boldsymbol{\xi}))}{\sigma_K}\right)^2\right] \\
&-\frac{1}{2}(\nu_K+1)\sum_{t=1}^{n}\left[\nu_K + \left(\frac{\boldsymbol{\alpha}_K(L)(y_{K,t} - h_{K,t}(\boldsymbol{\xi}))}{\sigma_K}\right)^2\right]^{-1}
\end{aligned}
$$

As with the previous three CM steps, we utilize the most up-to-date parameter estimates, now for $\boldsymbol{\xi}$, $\boldsymbol{\alpha}_1$, …, $\boldsymbol{\alpha}_N$, σ_1^2, …, σ_N^2. Denoting furthermore the desired solution by $\nu_K^{(i+1)}$ for every $K = 1, \ldots, N$, where we define $w_{K,t}^{(i+1)}$ in analogy to (16), we can derive the N equations

$$
\begin{aligned}
0 = \log \nu_K^{(i+1)} &+ 1 - \psi\left(\frac{\nu_K^{(i+1)}}{2}\right) + \psi\left(\frac{\nu_K^{(i+1)}+1}{2}\right) - \log\left(\nu_K^{(i+1)}+1\right) \\
&+\frac{1}{n}\sum_{t=1}^{n}\left(\log w_{K,t}^{(i+1)} - w_{K,t}^{(i+1)}\right).
\end{aligned} \tag{27}
$$

Thus, the estimates $\nu_1^{(i+1)}$, …, $\nu_N^{(i+1)}$ constitute the zeros of these equations, which are to be found numerically (using for instance MATLAB's `fzero` routine). Note for normally distributed white noise components that these degrees of freedom tend to infinity, in which case the function on the right-hand side of (27) does not change its sign. This numerical problem with the zero search is circumvented by testing for the existence sign change over a sufficiently large interval, say, over $[10^{-8}, 10^{8}]$; if this does not happen, the estimated degree of freedom should be set to a large value (say, to 10000). We stopped the reiteration in case the maximum number of iteration steps (500) was reached or in case the parameter values of the preceding step i did not change significantly within step $i+1$. We specified two thresholds with respect to largest maximum parameter changes: 10^{-4} for the degrees of freedom, and 10^{-8} for all other parameters.

4 Monte Carlo (MC) Results and Real World Application

4.1 The Framework of the Simulation

We consider in this section a multivariate, non-linear regression model in terms of a circle in $N = 3$ dimensions, having the following six parameters: two for the orientation (azimuth angle $\Phi \in [-\pi, \pi]$ and zenith angle $\theta \in [0, \pi]$) of its unit normal vector, one for the radius (r), and three for the circle center (C_x, C_y, C_z) (see pp. 24-27 in [11]). The observable 3D circle points are described by

$$\begin{pmatrix} X_t \\ Y_t \\ Z_t \end{pmatrix} = \begin{pmatrix} -r\cos(T_t)\sin(\Phi) + r\sin(T_t)\cos(\theta)\cos(\Phi) + C_x \\ r\cos(T_t)\cos(\Phi) + r\sin(T_t)\cos(\theta)\sin(\Phi) + C_y \\ -r\sin(T_t)\sin(\theta) + C_z \end{pmatrix} + \begin{pmatrix} E_{1,t} \\ E_{2,t} \\ E_{3,t} \end{pmatrix} \quad (28)$$

with $t = 1, \ldots, n$. In our current simulation study, $n = 100,000$ time instances in (28) are sampled equidistantly between $T_1 = 0$ and $T_n = 2\pi$ (corresponding to the time interval $[1, 10000]$ sec), and the circle parameters $\boldsymbol{\xi}$ were assumed to take the true values: $r = 0.487$ m, $\Phi = 0$ rad, $\theta = -\pi$ rad, $C_x = -2487.211$ m $C_y = -6053.041$ m and $C_z = -26.293$ m. according to a realistic scenario within the aforementioned application. Concerning the random deviations $\mathbf{E}_t$, we generated three different kinds of time series: (1) a pure white noise process, which may be viewed as an AR(0) process, (2) the AR(1) process

$$E_{k,t} = -0.9E_{k,t-1} + U_{k,t} \qquad (k = 1, \ldots, 3; t = 1, \ldots, n), \quad (29)$$

and (3) the ARMA(3,2) process (used for all $k = 1, \ldots, 3$)

$$E_{k,t} = -0.73E_{k,t-1} - 0.38E_{k,t-2} + 0.14E_{k,t-3} + U_{k,t} - 0.33U_{k,t-1} - 0.35U_{k,t-2}. \quad (30)$$

These models were investigated in the extensive study [8] (see pp. 230) on the stochastic modeling of GNSS data, where the white noise processes $U_{k,1}, \ldots, U_{k,n}$ were assumed to be Gaussian. Besides generating the white noise components $U_{k,1}, \ldots, U_{k,n}$ with the Gaussian sampling distributions

$$U_{1,t}, U_{2,t} \overset{\text{ind}}{\sim} N(0, 0.001^2), \;\; U_{3,t} \overset{\text{ind}}{\sim} N(0, 0.002^2), \quad (31)$$

we sampled also from the scaled t-distributions

$$U_{1,t}, U_{2,t} \overset{\text{ind}}{\sim} t_{2.5}(0, 0.001^2), \;\; U_{3,t} \overset{\text{ind}}{\sim} t_2(0, 0.002^2) \quad (32)$$

and from the contaminated normal distributions

$$U_{1,t}, U_{2,t} \overset{\text{ind}}{\sim} 0.6 \cdot N(0, 0.001^2) + 0.4 \cdot N(0, 0.008^2) \tag{33}$$

$$U_{3,t} \overset{\text{ind}}{\sim} 0.6 \cdot N(0, 0.002^2) + 0.4 \cdot N(0, 0.008^2) \tag{34}$$

to induce heavy tails or outliers. As the Z coordinates measured by GNSS are known to have much larger random fluctuations than the other coordinate components, the true variances in (31), true scale factors in (32) and true variances of the first Gaussian mixture component of (33) – (34) were chosen differently for the X/Y components (corresponding to $k = 1/k = 2$) and the Z component (associated with $k = 3$). Fluctuations due to systematic effects can also be expected to be largest for the Z components, so that the degree of freedom with respect to the variables $U_{3,t}$ in (32) is assumed to be less than for the components $U_{1,t}$ and $U_{2,t}$; thus, we assume the Student white noise in the Z coordinates (vertical coordinates) to be more heavy-tailed than the noise in the other components.

We generated 1000 random samples for the multivariate white noise series $\mathbf{U}_1, \ldots, \mathbf{U}_n$ from each of the distributions, from which we subsequently computed the corresponding noise series $\mathbf{E}_1, \ldots, \mathbf{E}_n$ and then via (28) the simulated observation time series $\mathbf{Y}_1, \ldots, \mathbf{Y}_n$. The proposed GEM algorithm was applied to each of these observation samples in order to estimate the six circle parameters, the coefficients of AR processes (having a suitable, identical order for each coordinate component), and the scale factors as well as degrees of freedom of the three underlying t-distributions. Note that neither the ARMA(3,2) model (30) nor the contaminated normal distributions (33) – (34) constitute special cases of the stochastic model (2) and (4) underlying the applied GEM algorithm.

4.2 Results of the Simulation and Real Data Application

Concerning the functional parameter $\boldsymbol{\xi}$, the Table 1 gives the means of the estimates of the first and third parameter ($\hat{r}$ and $\hat{\theta}$), computed from the 1000 MC runs. The approximation of the true parameter values by these means leads to bias free estimates for the AR(0) and AR(1) model. Only an insignificant bias in $\hat{\theta}$ in case of the ARMA(3,2) model can be detected.

To assess the goodness-of-fit of the AR models, a periodogram-based white noise test (WNT) is applied within each MC run to each of the three decorrelation-filtered residual series $\hat{u}_{k,1}$, $\hat{u}_{k,2}$ and $\hat{u}_{k,3}$. The test statistic determines the maximum cumulated periodogram excess over a cumulated, theoretical white noise periodogram (see [4] for detailed information concerning the computation of the test value). The white noise hypothesis is rejected if this maximum excess is larger than the critical value at a 95% significance level. More specifically, the critical value of the test is determined individually for each sample size n and each probability distribution in such a way that the acceptance rate, throughout all MC runs with generated random deviations e_1, e_2 and e_3 following the AR(0)-white noise model, is identical with the desired significance level 0.95. This critical value is then employed for the current sample size and probability distribution to determine the acceptance rates with respect to the estimated

AR(1) and ARMA(3,2) models. To approximate the ARMA(3,2) model in the applied GEM algorithm we increased the order an AR-processes gradually until the white noise test has been accepted. This results in appropriate model order 30. The WNT results are given in Table 1. Generally, the WNT acceptance rates increase with the AR model and reach 95,0% for AR(0). Apparently, the ARMA(3,2) models (approximated by an AR(30)) are estimated already reasonably well for this large sample size.

The performance of the estimation of the scale factor σ with respect to the $t_\nu(0, \sigma^2)$-distribution underlying the algorithm in Sect. 3 can be assessed only in the two cases that the white noise sampling distribution is (32) or (31), because the latter distributions are special cases of the family of scaled t-distributions. For the AR(0) and AR(1) models the mean value of the MC estimates $\hat{\sigma}$ coincides with the true value 0.001 for X/Y and 0.002 for Z (see Table 1). In contrast, for the ARMA(3,2) model the estimated scale factor is underestimated. In case of sampling by means of the contaminated normal (CN), the estimated scale factor can evidently not capture the effect of the two different variances in the data.

The evaluation of the algorithm's performance in estimating the degree of freedom of the underlying t-distribution is based on the mode of the MC estimates $\hat{\nu}$. As for the scale factor, the sampling distributions (31 and 32) allow for direct comparisons of the mode of the $\hat{\nu}$ with the corresponding true values $\nu = 2/2/2.5$ (with respect to X/Y/Z) and $\nu \to \infty$. The maximum value of an estimated $\hat{\nu}$ is 10000 for numerical reasons, which we therefore take as a sufficient approximation of $\nu \to \infty$. Table 1 shows that the degree of freedom tends to be overestimated for the ARMA(3,2) model.

Finally, the root mean square error (RMSE) measures the estimator's ability to predict the true observations. Since the predicted or adjusted observations are a consequence of the estimation of all four groups parameter groups $\boldsymbol{\xi}$, $\boldsymbol{\alpha}$, σ^2 and ν, the RMSE expresses the overall performance of the proposed GEM algorithm. This error measure includes both the variance and the bias of the estimator, and should therefore approach 0 for different AR models. The RMSE is computed for each MC run, and the resulting mean value is given in the Table 1. It can be seen that the mean of RMSE is substantially reduced with each increase in the AR model orders and for all error models. Only in case of t-distributed errors for the ARMA(3,2) model, one sample from the tail of the distribution occurred, which lead to an extreme estimation result and therefore to an unusually high RMSE value. To accommodate for this sampling effect we computed also the median of the RMSE values, as a robust measure of goodness of fit. As could be expected, the model reproductions based on the t- and the normal sampling distributions are much superior to the contaminated normal.

We also applied the GEM algorithm to approximate a measured and preprocessed 3D GNSS time series (see [12]) by the circle given in (28). One application of this model serves the geo-referencing of terrestrial laser scanner data where the 3D circle describes the circular, horizontal motion of two global navigation satellite system (GNSS) antenna reference points. Dual frequency receivers with individually and absolutely calibrated GNSS antennas were used. The origin of the coordinates lies in the nearby reference station with a baseline length of approximately 14 m. For further information on the measurement setup (see [12], p. 69). A full rotation consists of 7609 points (acquired with a data rate of 1 Hz) with respect to one antenna. We employed an AR model of order 12 for each time series component. Figure 1 shows the adjusted circle and the observed 3D points. Having obtained an estimated degree of freedom of 10,000 for each component we conclude that given GNSS series are normally distributed.

Table 1. Estimation results based on 1000 MC runs from the generated Student (t), normal (N) and contaminated normal (CN) error models according to (31) - (34). For WNT acceptance rates, Mean($\hat{\sigma}$) and Mode($\hat{\nu}$) results are listed one below the other for the three time series components (X/Y/Z).

	AR(0)			AR(1)			ARMA(3,2)		
Error model	t	N	CN	t	N	CN	t	N	CN
Mean($\hat{r}$)	0.4874	0.4874	0.4874	0.4874	0.4874	0.4874	0.4874	0.4874	0.4874
Mean($\hat{\theta}$)	-3.141593	-3.141592	-3.141585	-3.141592	-3.141592	-3.141589	-3.141593	-3.141592	-3.141573
WNT	0.95 0.95 0.95	0.95 0.95 0.95	0.95 0.95 0.95	0.962 0.962 0.962	0.972 0.972 0.972	0.955 0.955 0.955	0.999 0.999 0.999	1 1 1	1 1 1
Mean($\hat{\sigma}$)	0.0010 0.0010 0.0020	0.0010 0.0010 0.0020	0.0616 0.0616 0.0663	0.0010 0.0010 0.0020	0.0010 0.0010 0.0020	0.0616 0.0616 0.0663	0.0006 0.0006 0.0012	0.0005 0.0005 0.0010	0.0296 0.0296 0.0318
Mode($\hat{\nu}$)	2.50 2.50 2.00	10000 10000 10000	10000 10000 10000	2.50 2.50 2.00	10000 10000 10000	10000 10000 10000	3.00 3.01 2.40	10000 10000 10000	10000 10000 10000
Mean(RMSE) $\times 10^{-6}$	8	6	271	4	3	142	7961	4	191
Median(RMSE) $\times 10^{-6}$	8	6	267	4	3	140	6	4	187

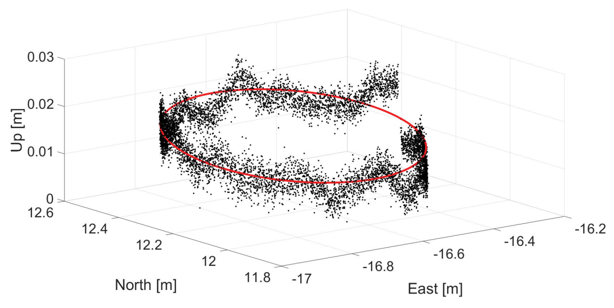

Fig. 1. 3D view of observed (black points) and adjusted circle (red line) for $n = 7827$ real three-dimensional GNSS measurements taken from [12], displayed in a North East Up (NEU) coordinate system.

5 Conclusions

To achieve an adaptive robust adjustment of a multivariate regression time series with outlier-afflicted/heavy-tailed and autocorrelated errors, we described the theory and implementation of a generalized expectation maximization algorithm. Monte Carlo simulations based on different error sampling distributions showed that the bias of the parameter estimates is insignificant when a sufficiently large number of observations (here 100,000) is adjusted. The presented algorithm was also tested in a real-data experiment using GNSS measurements.

References

1. Dempster, A.P., Laird, N.M., Rubin, D.B.: Maximum Likelihood from Incomplete Data via the EM Algorithm. J R Stat Soc (Series B) 39, 1-38 (1977)
2. Gupta, M.R., Chen, Y.: Theory and Use of the EM Algorithm. Foundations and Trends in Signal Processing 4, 223-296 (2011)
3. Koch, K.R., Kargoll, B.: Expectation Maximization Algorithm for the Variance-Inflation Model by Applying the t-Distribution. J Appl Geod 7, 217-225 (2013)
4. Kargoll, B., Omidalizarandi, M., Loth, I., Paffenholz, J.A., Alkhatib, H.: An Iteratively Reweighted Least-Squares Approach to Adaptive Robust Adjustment of Parameters in Linear Regression Models with Autoregressive and t-Distributed Deviations. Journal of Geodesy, https://doi.org/10.1007/s00190-017-1062-6 (2017)
5. Lange, K.L., Little, R.J.A., Taylor, J.M.G.: Robust Statistical Modeling Using the t-Distribution. J Am Stat Ass 84, 881-896 (1989)
6. Liu, C.H.: ML Estimation of the Multivariate t Distribution and the EM Algorithm. J Multivar Anal 63, 296-312 (1997)
7. Liu, C.H., Rubin, D.B.: The ECME Algorithm: A Simple Extension of EM and ECM with Faster Monotone Convergence. Biometrika 81, 633-648 (1994)
8. Luo, X.: GPS Stochastic Modelling: Signal Quality Measures and ARMA Processes. Springer, Berlin, Heidelberg (2013)
9. McLachlan, G.J., Krishnan, T.: The EM Algorithm and Extensions. John Wiley & Sons, Hoboken, New Jersey (2008)
10. Meng, X., Rubin, D.B.: Maximum Likelihood Estimation via the ECM Algorithm: A General Framework. Biometrika 80, 267-278 (1993).
11. Moon, P., Spencer, D.E.: Field Theory Handbook – Including Coordinate Systems, Differential Equations, and Their Solutions. 3rd edition. Springer, New York (1988)
12. Paffenholz, J.A.: Direct Geo-Referencing of 3D Point Clouds with 3D Positioning Sensors. PhD thesis. Deutsche Geodätische Kommission (Series C, 689), Munich, Online available under http://www.dgk.badw.de/fileadmin/docs/c-689.pdf (2013)
13. Park, M., Gao, Y.: Error and Performance Analysis of MEMS-based Inertial Sensors with a Low-cost GPS Receiver. Sensors 8, 2240-2261 (2008)
14. Phillips, R.F.: Least Absolute Deviations Estimation via the EM Algorithm. Statistics and Computing 12:281-285 (2002)
15. Porat, B.: Digital Processing of Random Signals. Dover, Mineola/New York (1994)
16. Schuh, W.D.: The Processing of Band-Limited Measurements; Filtering Techniques in the Least Squares Context and in the Presence of Data Gaps. Space Science Reviews 108, 67-78 (2003)
17. Wang, K., Xiong, S., Li, Y.: Modeling with Noises for Inertial Sensors. In: Position Location and Navigation Symposium (PLANS) 2012, IEEE/ION, 625-632 (2012)