

Proceeding Paper

Bayesian Robust Multivariate Time Series Analysis in Nonlinear Models with Autoregressive and t-Distributed Errors [†]

Alexander Dorndorf ¹, Boris Kargoll ², Jens-André Paffenholtz ³ and Hamza Alkhatib ^{1,*}

¹ Geodätisches Institut, Leibniz Universität Hannover, Nienburger Str. 1, D-30167 Hannover, Germany; dorndorf@gih.uni-hannover.de

² Institute of Geoinformation and Surveying, Anhalt University of Applied Sciences, Seminarplatz 2a, D-06846 Dessau-Rosslau, Germany; boris.kargoll@hs-anhalt.de

³ Institute of Geo-Engineering, Clausthal University of Technology, Erzstraße 18, D-38678 Clausthal-Zellerfeld, Germany; jens-andre.paffenholtz@tu-clausthal.de

* Correspondence: alkhatib@gih.uni-hannover.de; Tel.: +49-5117622464

[†] Current address: Presented at the 7th International conference on Time Series and Forecasting, Gran Canaria, Spain, 19–21 July 2021.

Abstract: Many geodetic measurement data can be modelled as a multivariate time series consisting of a deterministic (“functional”) model describing the trend, and a stochastic model of the correlated noise. These data are also often affected by outliers and their stochastic properties can vary significantly. The functional model of the time series is usually nonlinear regarding the trend parameters. To deal with these characteristics, a time series model, which can generally be explained as the additive combination of a multivariate, nonlinear regression model with multiple univariate, covariance-stationary autoregressive (AR) processes the white noise components of which obey independent, scaled t-distributions, was proposed by the authors in previous research papers. In this paper, we extend the aforementioned model to include prior knowledge regarding various model parameters, the information about which is often available in practical situations. We develop an algorithm based on Bayesian inference that provides a robust and reliable estimation of the functional parameters, the coefficients of the AR process and the parameters of the underlying t-distribution. We approximate the resulting posterior density using Markov chain Monte Carlo (MCMC) techniques consisting of a Metropolis-within-Gibbs algorithm.

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



check for
updates

Citation: Dorndorf, A.; Kargoll, B.; Paffenholtz, J.-A.; Alkhatib, H. Bayesian Robust Multivariate Time Series Analysis in Nonlinear Models with Autoregressive and t-Distributed Errors. *Eng. Proc.* **2021**, *5*, 20. <https://doi.org/10.3390/engproc2021005020>

Academic Editors: Ignacio Rojas, Fernando Rojas, Luis Javier Herrera and Hector Pomare

Published: 28 June 2021

Publisher’s Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Adjustment calculus offers a rich toolbox of statistical models and procedures for parameter estimation and hypothesis testing based on given numerical observations (cf. [1]). Such models usually consist of a deterministic functional model (e.g., a linear model describing some trend function), a correlation model (e.g., in the form of a variance-covariance matrix or an autoregressive (AR) error process), and a stochastic model (i.e., a probability distribution of the observation errors or the innovations of the AR error process). The stochastic model is often taken to be some multivariate normal distribution, which, however, easily leads to erroneous estimation results if the observations are afflicted by outliers. To take outliers into account, the normal distribution can be replaced by some outlier distribution, for example, a heavy-tailed t-distribution (cf. [2]). A multivariate time series model, including a nonlinear functional model and an autoregressive observation error model with t-distributed innovations, was suggested and investigated in [3] and [4]. A shortcoming of that model is that it does not include prior knowledge about the parameters of the functional, correlation or stochastic model, the information about which may readily

be available. Therefore, the current paper describes a Bayesian extension of that time series model, which can be expected to result in more robust and more accurate parameter estimates (cf. [5]).

A general Bayesian estimation approach in the specific context of models based on the t-distribution was introduced by [6]. Due to the complexity of such a model, the posterior density function must be approximated numerically or analytically. For numerical approximation, Monte-Carlo (MC) simulation and, in particular, Markov-Chain Monte-Carlo (MCMC) methods, which are suitable also for multivariate distributions, have been applied routinely (cf. [7]). In particular, the Gibbs sampler and the Metropolis–Hastings algorithm have been employed for (non-robust) Bayesian estimation of the parameters of a linear functional model with autoregressive moving-average (ARMA) and normally distributed errors [8]. MCMC methods have also been applied in the context of the robust Bayesian estimation of ARMA models [9] and AR models [10], with one additional (directly observed) mean parameter in the functional model. In both studies, outliers within the auto-correlated errors and within the uncorrelated innovations were modeled as normally distributed random variables with variances inflated by unknown multipliers. Thus, the stochastic error model was based on a discrete mixture of normal distributions, not on the t-distribution. To incorporate an automatic model selection procedure regarding the AR/ARMA model into the adjustment, the preceding studies also included unknown index parameters, taking the value 0 in case the corresponding AR (or MA) coefficient is 0 (or not significant) and taking the value 1 otherwise. Prior distributions for all of the parameter groups and the likelihood function for the data were fixed, and sampling distributions were then derived in order to obtain a numerical approximation of the posterior distribution for all the unknowns. In [11], an MCMC-based computational algorithm was proposed, to facilitate Bayesian analysis of real data when the error structure can be expressed as a p-order AR model.

The paper is organized as follows: First, the Bayesian multivariate time series model with AR and t-distributed errors is described in detail in Section 2. It is shown how the generic deterministic functional model, the AR process and the t-distribution model are first combined to a likelihood function and how prior information about the model parameters to be estimated is taken into account by means of a specified prior density. Here, we denote unknown parameters with Greek letters, random variables with calligraphic letters, and constants with Roman letters. Furthermore, we distinguish between a random variable (e.g., \mathcal{L}_t) and its realization (l_t). Matrices are shown, as usual, as bold capital letters and vectors as bold small letters. Section 3 outlines an MCMC algorithm for determining the posterior density of the unknown parameters of the functional model, the coefficients of the AR process and the scale parameter, as well as the degrees of freedom of the t-distribution. In Section 4, a time series model for GNSS observations of a circle in 3D is proposed, and the results of a Monte Carlo simulation are discussed. These findings are used to evaluate the performance of the implemented Metropolis–Hastings-within Gibbs algorithm in this scenario.

2. The Bayesian Time Series Model

We assume that an N -dimensional time series (\mathcal{L}_t) = ($[\mathcal{L}_{1,t} \ \dots \ \mathcal{L}_{N,t}]^T$) is observed at equi-spaced time instances t without data gaps. The observation model consists of the three interconnected model components,

$$\mathcal{L}_{k,t} = h_{k,t}(\beta_1, \dots, \beta_m) + \mathcal{E}_{k,t}, \quad (1)$$

$$\mathcal{E}_{k,t} = \alpha_{k,1}\mathcal{E}_{k,t-1} + \dots + \alpha_{k,p_k}\mathcal{E}_{k,t-p_k} + \mathcal{U}_{k,t}, \quad (2)$$

$$\mathcal{U}_{k,t} \mid \psi_k^2, \nu_k \stackrel{\text{ind.}}{\sim} t_{\nu_k}(0, \psi_k^2, \nu_k), \quad (3)$$

where (1) defines the “observation equations”, (2) the “error equations” and (3) defines the probability distribution of the innovations. The parameters of this observation model are combined within the vector,

$$\theta = [\beta^T \quad \alpha^T \quad \psi^T \quad \nu^T]^T, \tag{4}$$

with

$$\beta = [\beta_1 \quad \dots \quad \beta_m]^T \tag{5}$$

$$\alpha = [\alpha_1^T \quad \dots \quad \alpha_N^T]^T = [\alpha_{1,p_1} \quad \dots \quad \alpha_{1,p_1} \quad \dots \quad \alpha_{N,1} \quad \dots \quad \alpha_{N,p_N}]^T \tag{6}$$

$$\psi = [\psi_1 \quad \dots \quad \psi_N]^T \tag{7}$$

$$\nu = [\nu_1 \quad \dots \quad \nu_N]^T. \tag{8}$$

On the one hand, the parameters θ are treated as variables of the likelihood function $f_{\mathcal{L}|\Theta}(\mathbf{L}|\theta)$, defined by the observation model (1)–(3). On the other hand, the parameters θ are viewed as a realization of a random vector Θ , having a specified pdf independent of the observables. According to the Bayes theorem, this prior density $f_{\Theta}(\theta)$ and the likelihood function $f_{\mathcal{L}|\Theta}(\mathbf{L}|\theta)$ are connected to the posterior density $f_{\Theta|\mathcal{L}}(\theta|\mathbf{L})$ via proportionality relationship

$$f_{\Theta|\mathcal{L}}(\theta|\mathbf{L}) \propto f_{\Theta}(\theta) \cdot f_{\mathcal{L}|\Theta}(\mathbf{L}|\theta), \tag{9}$$

which serves as the foundation of the inference of the parameters and the adjustment of the observations. The details of this model are described in the following.

The observation equations: Equation (1) reflects the idea that geodetic measurements $\mathcal{L}_{k,t}$ are approximated by a “deterministic” model using mathematical functions $h_{k,t}(\beta)$, which are assumed to be partially differentiable. The index k refers to the time series surveyed by the k th sensor or sensor component, and the time instances $t = 1, \dots, n$ are the same for all sensors. In some applications, the functional model $h_{k,t}(\beta)$ takes the form

$$h_{k,t}(\beta) = \mathbf{X}_{k,t}\beta, \tag{10}$$

of a “linear model”, where \mathbf{X} denotes the design matrix and has a full rank. Since geodetic observables can generally not be modeled using a deterministic model alone, random deviations $\mathcal{E}_{k,t}$ are added to absorb the remaining effects. It is assumed that the instruments used to survey the observables are calibrated, so that no systematic errors occur. Thus, the expected values of the random deviations are assumed to be 0.

The error equations: Equation (2) is included to take account of auto-correlations within each of the N time series. Since the different sensors or sensor components may have different noise characteristics, AR processes, with individual orders p_1, \dots, p_N and sets of coefficients $\alpha_1, \dots, \alpha_N$, are selected. The noise characteristics are assumed to be constant throughout the measurement period. For practical purposes, the AR processes considered are therefore required to be (asymptotically) covariance-stationary. The random variables $\mathcal{U}_{k,t}$ are referred to as “innovations”. Since the observation window is finite, ranging from $t = 1, \dots, n$, the error equations involve errors at times $t = 0, -1, \dots$. To ensure asymptotic covariance-stationarity and the computability of the recursive equations, these quantities are set as equal to 0 (cf. [12]). This initial distortion of the AR process fades out as the process advances in time.

The stochastic model: The innovations of an AR process are usually assumed to be Gaussian white noise. Since the assumption of normal distributions is unrealistic in some

geodetic applications, for example, due to outliers, the heavy-tailed t-distributions are employed here. These are defined by the probability density function (pdf),

$$f(x) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\sqrt{\nu\pi}\psi} \left[1 + \frac{(x-\mu)^2}{\nu\psi^2}\right]^{-\frac{\nu+1}{2}}, \tag{11}$$

where Γ is the gamma function. Since the expected values of the random deviations $\mathcal{E}_{k,t}$ should be 0, we can also restrict the location parameter μ to 0. Since the noise of different sensors or sensor components may exhibit different levels of variance and outliers, each time series involves a t-distribution with individual scale parameter ψ_k^2 and degree of freedom (df) ν_k . It should be mentioned that the alternative usage of a multivariate t-distribution (as defined in [2]) involves a single df and would therefore not allow for the modeling of distinct outlier characteristics within the different time series.

The likelihood function: A likelihood function $f_{\mathcal{L}|\Theta}(\mathbf{L}|\theta)$ can be obtained by combining the observation Equation (1), the error Equation (2) and the stochastic model of the innovations (3). To do so, the well-known method of conditional likelihoods in connection with AR processes with various forms of non-Gaussian innovations is applied (cf. [13–15]). Assuming the AR processes to be invertible, the error Equation (2), in terms of their numerical realizations, can be rewritten as “innovation equations,”

$$u_{k,t} = e_{k,t} - \alpha_{k,1}e_{k,t-1} - \dots - \alpha_{k,p_k}e_{k,t-p_k}. \tag{12}$$

As the errors $e_{k,t}$ contained in the observation Equation (1) can be expressed as

$$e_{k,t} = \ell_{k,t} - h_{k,t}(\beta), \tag{13}$$

the innovation Equation (12) become

$$u_{k,t} = e_{k,t} - \alpha_{k,1}(\ell_{k,t-1} - h_{k,t-1}(\beta)) - \dots - \alpha_{k,p_k}(\ell_{k,t-p_k} - h_{k,t-p_k}(\beta)). \tag{14}$$

The conditional likelihood function is then obtained as the product of the univariate pdf (11), evaluated at all the stochastically independent innovations $u_{k,t}$ with location $\mu = 0$, associated scale factor ψ_k^2 and df ν_k , that is,

$$\mathbf{L}(\theta|\mathbf{L}) = f_{\mathcal{L}|\Theta}(\mathbf{L}|\theta) = \prod_{k=1}^N \prod_{t=1}^n \frac{\Gamma\left(\frac{\nu_k+1}{2}\right)}{\Gamma\left(\frac{\nu_k}{2}\right)\sqrt{\nu_k\pi}\psi_k} \left[1 + \frac{u_{k,t}^2}{\nu_k\psi_k^2}\right]^{-\frac{\nu_k+1}{2}}. \tag{15}$$

For the purpose of maximum likelihood (ML) estimation, the logarithm of this likelihood function is easier to handle (see [3]). In that contribution, a computationally convenient ML estimation of the model parameters was achieved by rewriting the t-distributions as conditional normal distributions with latent variables; these variables play the role of weights in an iteratively reweighted least-squares algorithm. As the main innovation of the current contribution, the likelihood function (15) is incorporated into a Bayesian model instead, which is described in the following.

The Bayesian model: In this contribution, both informative and non-informative prior information is considered. The result of using a fully non-informative prior is that the posterior density follows directly from the likelihood function. In the case of an informative prior, a joint pdf must be specified for the random vector Θ . This task is simplified by the assumption of stochastic independence of the parameter groups β , α , ψ and ν , so that the factorization property,

$$f_{\Theta}(\theta) = f_{\Theta}(\beta, \alpha, \psi, \nu) = f_{\Theta}(\beta)f_{\Theta}(\alpha)f_{\Theta}(\psi)f_{\Theta}(\nu), \tag{16}$$

holds. Consequently, individual prior densities can be specified for these parameter groups. As the prior density of β depends on the selected functional model $h(\beta)$, its specification is fixed after the introduction of the application example in Section 4. In contrast, the prior densities for α , ψ and ν do not depend on the choice of the functional model but mainly on the precision of the sensors or instruments employed. In the case that no prior information is available for the employed sensors or instruments, it is still possible to define prior densities for these three groups of parameters. Due to the assumption, in connection with the error Equation (2) and the stochastic model (3), that the N time series is stochastically independent, the prior density can be further simplified to

$$f_{\Theta}(\alpha) = f_{\Theta}(\alpha_1)f_{\Theta}(\alpha_2) \cdots f_{\Theta}(\alpha_N), \quad (17)$$

$$f_{\Theta}(\psi) = f_{\Theta}(\psi_1)f_{\Theta}(\psi_2) \cdots f_{\Theta}(\psi_N), \quad (18)$$

$$f_{\Theta}(\nu) = f_{\Theta}(\nu_1)f_{\Theta}(\nu_2) \cdots f_{\Theta}(\nu_N). \quad (19)$$

Consequently, as far as the parameters ψ and ν are concerned, only univariate prior densities $f_{\Theta}(\psi_k)$ and $f_{\Theta}(\nu_k)$ need to be specified. When it is known that the scale factor ψ is between ψ_{\min} and ψ_{\max} , the prior density defining the continuous uniform distribution $U(\psi_{\min}, \psi_{\max})$ can be used as a weak form of prior information. The specification of the prior for the df ν_k can be based, on the one hand, on the requirement $\nu_k > 2$, so that the variance of the t-distributed random variables is defined. On the other hand, the t-distribution is practically indistinguishable from a normal distribution for dfs greater than 120 (cf. [16]), so that the upper limit $\nu_k \leq 120$ can be fixed. In the absence of further information about the dfs, the prior density defining the continuous uniform distribution $U(2, 120)$ is reasonable. The auto-correlations of a time series may, for instance, be induced by calibration corrections within the measurement device, by movements of the measured object, or by a combination of the two effects. Therefore, a general definition of the prior density of the AR coefficients is not trivial. For this reason, a non-informative prior density is specified under the additional assumption that all of the AR coefficients are stochastically independent.

3. The Developed MCMC Algorithm

Because of the use of the Student distribution for the white measurement noise (Equation (15)), an analytical solution of the posterior density based on the Bayes theorem (Equation (9)) is not possible, so it can only be solved numerically. The general solution approach is based on generating a so-called Markov Chain for the unknown posterior density using the MCMC method. MCMC algorithms are commonly used in all fields of statistics because of their versatility and generality. When an MCMC method is applied to solve the posterior density function given in (9), it is usually realized with a Gibbs sampler. An implementation of a Gibbs sampler relies on the availability of the complete conditional pdfs of all parameters of interest in our particular problem (cf. Equation (4)). However, the complete conditional pdfs of all parameters of interest are not readily available. In such cases, a Metropolis–Hastings (MH) method can be incorporated within a Gibbs sampler to draw samples from the parameters, the full conditional pdf of which cannot be analytically determined. In this paper, we demonstrate the development of such an algorithm, known as Metropolis–Hastings-within Gibbs. In this algorithm, the Gibbs sampler is used to generate the Markov Chain for $\theta|L$, and within the Gibbs sampler the MH algorithm is used to generate random numbers. For a clearer presentation, the solution algorithm

is encapsulated by two separate functions. For the Gibbs sampler, the outer function is given by,

$$\begin{aligned}
 & \text{for: } j = 1, \dots, \mathbb{M} & (20) \\
 & \text{Draw } \boldsymbol{\psi}^j | \boldsymbol{\beta}^{j-1}, \boldsymbol{\nu}^{j-1}, \boldsymbol{\alpha}^{j-1}, \mathbf{L} & \sim f_{\Theta|\mathcal{L}}(\boldsymbol{\psi} | \boldsymbol{\beta}, \boldsymbol{\nu}, \boldsymbol{\alpha}, \mathbf{L}) \\
 & \text{Draw } \boldsymbol{\beta}^j | \boldsymbol{\psi}^j, \boldsymbol{\nu}^{j-1}, \boldsymbol{\alpha}^{j-1}, \mathbf{L} & \sim f_{\Theta|\mathcal{L}}(\boldsymbol{\beta} | \boldsymbol{\psi}, \boldsymbol{\nu}, \boldsymbol{\alpha}, \mathbf{L}) \\
 & \text{Draw } \boldsymbol{\nu}^j | \boldsymbol{\beta}^j, \boldsymbol{\psi}^j, \boldsymbol{\alpha}^{j-1}, \mathbf{L} & \sim f_{\Theta|\mathcal{L}}(\boldsymbol{\nu} | \boldsymbol{\beta}, \boldsymbol{\psi}, \boldsymbol{\alpha}, \mathbf{L}) \\
 & \text{Draw } \boldsymbol{\alpha}^j | \boldsymbol{\beta}^j, \boldsymbol{\psi}^j, \boldsymbol{\nu}^j, \mathbf{L} & \sim f_{\Theta|\mathcal{L}}(\boldsymbol{\alpha} | \boldsymbol{\beta}, \boldsymbol{\psi}, \boldsymbol{\nu}, \mathbf{L})
 \end{aligned}$$

where \mathbb{M} is the length of the MCMC. Such a Markov chain ensures the convergence of the distribution of the samples to the target distribution after a few burn-in periods \mathbb{V} [17]. We observe that the full conditional pdf shown in (20) does not fit to any known pdf and, therefore, we cannot directly draw samples from it.

However, there are two challenges to calculating the conditional posterior densities. The first challenge is that it results from the likelihood function and the prior density. Consequently, changing the distributional assumption for the prior density results in a new conditional posterior density. While this challenge can be overcome with a small amount of effort, the second challenge is much more fundamental. For the calculation of the conditional posterior densities, several integrals have to be solved and this may not be analytically possible. To overcome these challenges, the MH algorithm is used to draw the required random numbers. The general algorithm for drawing a random number θ_i^j from the posterior density $f_{\Theta|\mathcal{L}}(\boldsymbol{\theta}|\mathbf{L})$ follows from the following steps:

1. Generate: $\theta_i^{\text{new}} \sim N(\theta_i^{j-1}, \lambda_{\theta_i})$ (21)
2. Set: $\boldsymbol{\theta}^{\text{new}} = [\theta_1^j, \theta_2^j, \dots, \theta_i^{\text{new}}, \dots, \theta_{m-1}^{j-1}, \theta_m^{j-1}]^T$
 $\boldsymbol{\theta}^{\text{old}} = [\theta_1^j, \theta_2^j, \dots, \theta_i^{j-1}, \dots, \theta_{m-1}^{j-1}, \theta_m^{j-1}]^T$
3. Calculate: $\zeta = \min \left[1, \frac{f_{\Theta}(\boldsymbol{\theta}^{\text{new}})\mathbf{L}(\boldsymbol{\theta}^{\text{new}}|\mathbf{L})}{f_{\Theta}(\boldsymbol{\theta}^{\text{old}})\mathbf{L}(\boldsymbol{\theta}^{\text{old}}|\mathbf{L})} \right]$
4. Accept or Reject: $\tau \sim U(0, 1)$
 if $\tau \leq \zeta$: $\theta_i^j = \theta_i^{\text{new}}$
 else: $\theta_i^j = \theta_i^{j-1}$,

where m in Equation (22) denotes the length of the parameter vector. The results of the Metropolis–Hastings-within Gibbs are \mathbb{M} random realizations of the unknown parameters $\boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\psi}$ and $\boldsymbol{\nu}$ from the posterior density $f_{\Theta|\mathcal{L}}(\boldsymbol{\theta}|\mathbf{L})$. The estimated values $\hat{\boldsymbol{\theta}}$ for the unknown parameters with their variance–covariance matrix (VCM) result from (cf. [5]):

$$\hat{\theta}_i = \frac{1}{\mathbb{M} - \mathbb{V}} \sum_{j=\mathbb{V}+1}^{\mathbb{M}} \theta_i^j \quad ; \quad \hat{\boldsymbol{\Sigma}}_{\theta_i,s} = \frac{1}{\mathbb{M} - \mathbb{V}} \sum_{j=\mathbb{V}+1}^{\mathbb{M}} (\theta_i^j - \hat{\theta}_i)(\theta_s^j - \hat{\theta}_s). \quad (22)$$

4. Closed Loop Monte Carlo Simulation

4.1. The Framework of the Simulation

In our Closed Loop Monte Carlo simulation (CLS), we rely on the real-world application demonstrated in [3], in which we used a multi-sensor-system (MSS) composed of a laser scanner and two firmly attached pieces of GNSS equipment (see [18] for details). We consider a multivariate, non-linear regression model in terms of a circle in $N = 3$ dimensions that has the following six parameters: two for the orientation (φ and ω) of

its unit normal vector; one for the radius (r); and three for the circle center (c_x, c_y, c_z). The observable 3D circle points are described by

$$l_{x,t} := l_{1,t} = h_{1,t}(\boldsymbol{\beta}) = r \cos(\kappa_t) \cos(\varphi) + c_x \tag{23}$$

$$l_{y,t} := l_{2,t} = h_{2,t}(\boldsymbol{\beta}) = r \cos(\kappa_t) \sin(\varphi) \sin(\omega) + r \sin(\kappa_t) \cos(\omega) + c_y \tag{24}$$

$$l_{z,t} := l_{3,t} = h_{3,t}(\boldsymbol{\beta}) = -r \cos(\kappa_t) \sin(\varphi) \cos(\omega) + r \sin(\kappa_t) \sin(\omega) + c_z, \tag{25}$$

with $t = 1, \dots, 2000$ and where $\kappa_t = \frac{2\pi}{2000} \cdot t$ represents fixed rotation angles around the z-axis. In this simulation, the functional parameters are the circle parameters $\boldsymbol{\beta}$, which were assumed to take the true values

$$\boldsymbol{\beta} = [c_x \ c_y \ c_z \ r \ \omega \ \varphi]^T = [1716.0 \ 3012.0 \ 1064.0 \ 30.0 \ 0.0019 \text{ rad} \ -0.0013 \text{ rad}]^T. \tag{26}$$

The random deviations \mathcal{E}_t were generated by the AR(2) processes

$$e_{k,t} = \sum_{j=1}^2 \alpha_{k,j} e_{k,t-1} + u_{k,t}, \text{ for } k = 1, 2, 3, \tag{27}$$

with true coefficients

$$\alpha_{1,1} = 0.57, \alpha_{1,2} = 0.11, \alpha_{2,1} = 0.67, \alpha_{2,2} = 0.22, \alpha_{3,1} = 0.35, \alpha_{3,2} = 0.55.$$

The innovations of these processes are sampled from the scaled t-distributions

$$u_{k,t} \stackrel{\text{ind.}}{\sim} t_{\nu_k}(0, \psi_k^2), \tag{28}$$

with true scale parameters

$$\psi_x := \psi_1 = 0.2, \quad \psi_y := \psi_2 = 0.2, \quad \psi_z := \psi_3 = 0.4$$

and true dfs

$$\nu_x := \nu_1 = 8, \quad \nu_y := \nu_2 = 10, \quad \nu_z := \nu_3 = 4.$$

In Equation (16), the prior density has been introduced for the general Bayes model. In this simulation, only prior densities for the functional parameters $\boldsymbol{\Theta}$ (see Equation (26)) are assumed to be known:

$$f_{\boldsymbol{\Theta}}(\boldsymbol{\beta}) = f_{\boldsymbol{\Theta}}(c_x, c_y, c_z) f_{\boldsymbol{\Theta}}(r) f_{\boldsymbol{\Theta}}(\omega) f_{\boldsymbol{\Theta}}(\varphi). \tag{29}$$

In Equation (29), we assume that the prior information for the center of the circle, the radius and the angles are independent of each other. The reason for this assumption is that this information is obtained from data sheets or from calibrations. These are specified as follows.

The prior density of the circle center: The prior for the center of the circle is the knowledge that it must be located approximately in the middle between the observations constituting a circle. The location parameter μ_c for the definition of the prior density of the circle center is thus dependent on the observations. Therefore, we consider the prior of the circle center as weak prior information. Hence, we use the identity matrix as a VCM for $\boldsymbol{\Sigma}_c$. The uncertainty of a coordinate component of the prior information of the circle center is thus significantly larger than the assumed measurement precision of a single observation. As a pdf for the prior $f_{\boldsymbol{\Theta}}(c_x, c_y, c_z)$, the multivariate normal distribution is assumed with the expected value μ_c and VCM $\boldsymbol{\Sigma}_c$.

The prior density for the radius: The prior information for the radius of the circle is the result of a calibration measurement using a laser tracker. The manufacturer specifies the

accuracy of a single point measurement with this instrument as the maximum permissible error of $MPE_{x,y,z} = \pm 15 \mu m + 6 \frac{\mu m}{m}$. During the measurement, the diameter of the wing, on which the GNSS antenna was mounted, was determined. For this purpose, the 3D coordinates were measured on the left and right hand side and the Euclidean distance was calculated, which corresponds to the diameter. In total, the radius was estimated four times in this way. By averaging these results, $\mu_r = 30.0026$ was obtained as the location parameter for the prior density of the radius. The scale parameter for the prior is the result of the standard deviation $\sigma_r = 0.0043$ of the four determined radii of the laser tracker measurement. The measurement results of the laser tracker are assumed to be normally distributed. Hence, the normal distribution is assumed for the prior density of the radius of the circle.

The prior density for the rotation angles: The prior information for the rotation angles is derived from the instrument's levelling. The manufacturer specifies the accuracy of the bubble level as ± 0.0047 rad for a 99.9% confidence interval. It is assumed that this information refers separately to one vial axis, so that $f_{\Theta}(\omega)$ and $f_{\Theta}(\varphi)$ are independent of each other. Due to the specification of the accuracy by means of the confidence interval, the uniform distribution is used as the prior family for the angles. The limit of the confidence interval is used as the limit for the uniform distribution, from which follows $a_{\omega} = -0.0047$ rad and $b_{\omega} = 0.0047$ rad. This results in the prior density:

$$f_{\Theta}(\omega) = \begin{cases} \frac{1}{b_{\omega}-a_{\omega}} & a_{\omega} \leq \omega \leq b_{\omega} \\ 0 & \text{else} \end{cases} \quad (30)$$

For φ , the density of the prior is identical to that of the previous formula.

4.2. Results of the Simulation

The results described in this section were achieved with $\mathbb{M} = 10,000$ iterations and a burn-in period with $\mathbb{V} = 3000$ for the MCMC algorithm. We compare in the following three estimation procedures: Bayesian informative (Bayes Inf) using prior knowledge about the functional parameters β , Bayesian non-informative (Bayes Non-Inf) without using prior knowledge about β and the EM (expectation–maximization) algorithm developed in [3]. Before the results of the three estimation methods from the entire CLS are compared, the result of the Bayes Non-Inf solution is considered in more detail. For this purpose, the result of the Markov chains from a single simulation solution is considered in Figure 1. In total, Markov chains were generated for 18 unknown parameters. Figure 1 only shows a representative selection of the results and is limited to the results of the z-component and only one rotation angle. For the other components, the generated chains correspond to the behavior of the z-component. The green dashed line in Figure 1 shows the true value of the parameters used to generate the simulated observations. The red dashed line, or the red cross on the secondary diagonal, show the estimated parameters resulting from the Markov chains. On the diagonal, the distribution of the generated random numbers of the chain is shown as a histogram. The histograms show how the generated random numbers scatter around the estimated parameter and that the true value is close to the estimated parameter. The secondary diagonal of the figure represents the scatter of the generated random numbers depending on two parameters. From this distribution, the correlation of the generated chain between two parameters can be calculated. A dependency can be seen, especially for the scaling parameter and the df of the z-components. This also holds for the coefficients of the AR(2) process of the z-component. The same can be noticed for the x- and y-components. For all other parameters, the major axes of the ellipse are rather parallel to the axes of the expected values, which corresponds to a correlation around zero. The fact that, for example, the scale parameter and the df show stronger correlation behavior is also to be expected. The reason for this is that, with a smaller scale factor, more observations lie in the tails of the distribution, which leads to a smaller df. In contrast, if the scaling

factor is large, there are fewer observations in the tails of the distribution, so a larger df can be selected.

To compare the results of the three approaches, the adjusted observations \hat{l} are used instead of the estimated parameters $\hat{\beta}$. The reason for this is that, in all estimation procedures, the estimated parameters are closer to the nominal value of the simulation parameters and, therefore, it is difficult to judge which approach produces the best results. On the other hand, the predicted observations \hat{l} include the cumulative estimation uncertainties of all parameters $\hat{\beta}$, allowing for an easier comparison. Furthermore, we restrict ourselves here to the representation of the z-component because \hat{l}_z is most sensitive to inaccuracies in the estimated angles $\hat{\omega}$ and $\hat{\varphi}$. The results for the x- and y-components show a result similar to that of the model shown in Figure 2 for the z-component.

Before discussing the results of the three approaches, Figure 2 is first explained. The \hat{l} were calculated with the estimated parameters $\hat{\beta}$ using Equation (25). Subsequently, the \hat{l}_z were reduced by the true value for the observations $E(I_z)$, which is why the predicted observations scatter around 0. The dashed line is the mean value of observation $l_{z,t}$, which results from the 10,000 results of the CLS for one observation t . The dashed blue line of the mean value of the EM algorithm cannot be seen in the figure because it is overlaid by the dashed red line. The colored area shows, for observation t , the 95% confidence interval that results from the 10,000 predicted observations $\hat{l}_{z,t}$. The colored lines >0 show the maximum deviation from the true value for the observation $\hat{l}_{z,t}$ that appeared in the 10,000 simulations. The colored lines <0 show the minimum deviation from the true value for the corresponding approaches.

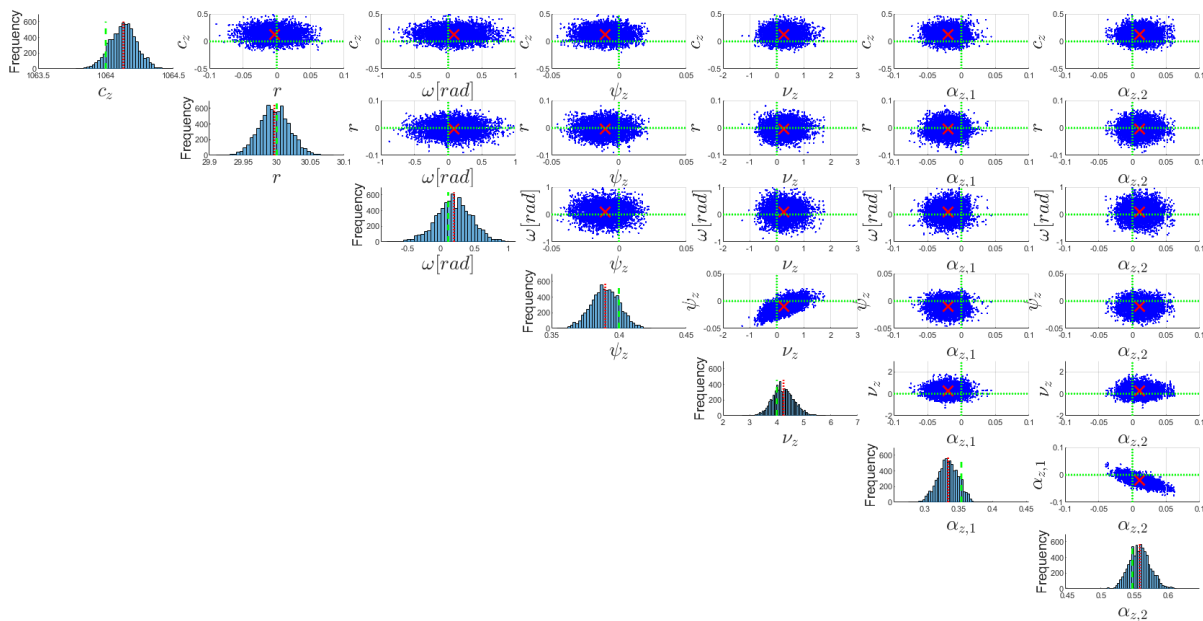


Figure 1. Result of the generated Markov chain after burn-in for Bayes Non-Inf: The main diagonal shows the distribution of the generated random numbers of a parameter. The secondary diagonal shows the correlation behavior of the generated random numbers at time j between two parameters.

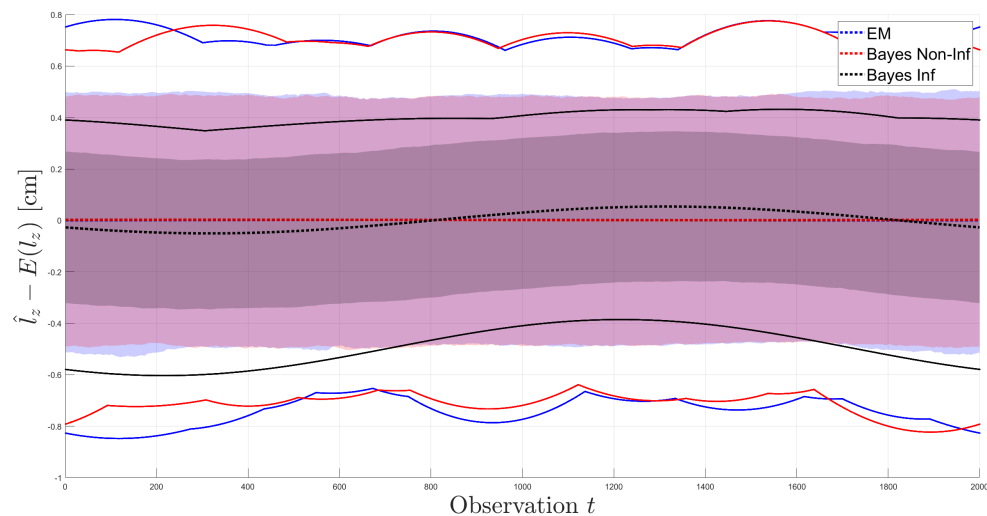


Figure 2. Result of the 10,000 CLS for the 2000 predicted observations of the z-component \hat{l}_z of the EM algorithm, Bayes non-informative (Bayes Non-Inf) and Bayes informative (Bayes Inf).

It is expected that the result of the EM algorithm is identical to the Bayes Non-Inf solution. This can be seen clearly in the result of the mean, where the two dashed lines (blue and red) overlap almost completely and only deviate from each other by a maximum of ≈ 0.002 cm. Furthermore, the mean values of the two estimation procedures are identical to the nominal value of 0 for all predicted observations within 10^{-2} cm. This deviation can be explained by the fact that the CLS was only performed 10,000 times and not infinitely often. For the confidence interval of the EM and Bayes Non-Inf solution, a similar behavior can be seen as for the mean value. For most observations t , the two confidence intervals overlap almost perfectly and, only for a few observations, a difference of at most ≈ 0.05 cm can be identified. However, this identical behavior is not seen in the maximum and minimum deviations of the blue and red lines, where the Bayes Non-Inf solution more often has smaller minimum deviations than the EM results. The reason for this has not yet been analysed in more detail and will be addressed in future work.

In the result of the mean value of Bayes Inf, a deviation from the mean value of EM and Bayes Non-Inf can be seen. The solution of the mean value of Bayes Inf oscillates cyclically around the nominal value of 0 with a maximum deviation of ± 0.05 cm, whereby this deviation is smaller by a factor of 10 than the scale factor ψ_z of the measurement noise. The mean values of the EM and Bayes also oscillate around 0, but this is smaller by a factor of 100 and therefore cannot be seen visually in Figure 2. The influence of the prior information $f_{\Theta}(\beta)$ on the observations $\hat{l}_{z,t}$ can be seen well in the confidence interval and the lines marked in black. These are clearly closer to the true value than in the EM algorithm and Bayes Non-Inf. It is of particular interest that the maximum deviation lies in the 95% confidence interval of EM and Bayes Non-Inf. This is not always the case for the minimum deviation of Bayes Inf, but the black line is still closer to zero than the blue and red lines. The reason for this is mainly due to the prior information for the angles $f_{\Theta}(\omega)$ and $f_{\Theta}(\varphi)$, which improves the estimation of ω and φ .

To compare the results \hat{l}_x , \hat{l}_y and \hat{l}_z of the three approaches, the RMSE of the true observations is calculated for each CLS result. From these RMSE values, the statistical measures in Table 1 were calculated for each approach. All results in the table show the same behavior as the results previously displayed in Figure 2. The EM and Bayes Non-Inf results are almost identical, with the small deviation explained by the different methods used to estimate the parameters. In the Bayes Inf solution, however, all statistical measures are smaller. For the mean, the RMSE of Bayes Inf is about 37% smaller than the RMSE of the EM algorithm, and for the median the difference is slightly larger, at about 41%. Only for the maximum RMSE value from the 10,000 simulations is the difference between EM and

Bayes Informative significantly smaller at about 13%, but the maximum RMSE of Bayes Inf is still significantly smaller than the result from EM.

Table 1. Root-Mean-Square-Error (RMSE) for the predicted observations to the true observations.

| Method | Mean [cm] | Median [cm] | Min [cm] | Max [cm] | σ [cm] |
|---------------|-----------|-------------|----------|----------|---------------|
| EM | 0.107 | 0.103 | 0.013 | 0.337 | 0.041 |
| Bayes Non-Inf | 0.104 | 0.100 | 0.012 | 0.303 | 0.040 |
| Bayes Inf | 0.067 | 0.060 | 0.004 | 0.290 | 0.032 |

5. Conclusions

To achieve an Bayesian adaptive robust adjustment of a multivariate regression time series with outlier-afflicted/heavy-tailed and autocorrelated errors, we described the theory and implementation of an MCMC based approach consisting of a Metropolis-within-Gibbs algorithm. In particular, the Gibbs sampler and the Metropolis–Hastings algorithm have been employed for robust Bayesian estimation of the parameters of a non-linear functional model with AR and t-distributed errors. An advantage of this procedure compared to the EM algorithm, besides the capability to process additional prior knowledge, is that the approximation of the posterior model parameters is feasible without linearization of the functional model. Furthermore, the approximation of the VCM $\hat{\Sigma}_\theta$ of the estimated parameters can be derived directly from the generated chains. CLS showed that the bias of the parameter estimates and the adjusted observations, as well as the RMSE, are significantly reduced when an informative Bayesian approach is used, but only under the condition that good prior information, that is, information that contains the true value, is assumed to be known. Otherwise, the prior can also have the opposite effect and may cause the estimated parameters to deteriorate.

Author Contributions: Conceptualization, H.A., A.D., B.K., J.-A.P.; methodology, A.D., H.A. and B.K.; software, A.D., B.K. and H.A.; validation, A.D. and H.A.; formal analysis, A.D. and H.A.; investigation, A.D., B.K., and H.A.; writing—original draft preparation, A.D., H.A. and B.K.; writing—review and editing, H.A. and B.K.; visualization, H.A. and A.D.; supervision, H.A., B.K. and J.-A.P.; project administration, H.A., B.K. and J.-A.P.; funding acquisition, H.A., B.K. and J.-A.P. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation)—project number 386369985.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Koch, K.R. *Parameter Estimation and Hypothesis Testing in Linear Models*; Springer: Berlin, Germany, 1999.
- Lange, K.L.; Little, R.J.A.; Taylor, J.M.G. Robust Statistical modeling using the t-distribution. *J. Am. Stat. Assoc.* **1989**, *84*, 881–896. [[CrossRef](#)]
- Alkhatib, H.; Kargoll, B.; Paffenholz, J.-A. Further results on robust multivariate time series analysis in nonlinear models with autoregressive and t-distributed errors. In *Time Series Analysis and Forecasting*; Valenzuela, O., Rojas, F., Pomares, H., Rojas, I., Eds.; ITISE 2017; Contributions to Statistics; Springer: Cham, Switzerland, 2018; pp. 25–38. [[CrossRef](#)]
- Kargoll, B.; Omidalizarandi, M.; Loth, I.; Paffenholz, J.A.; Alkhatib, H. An iteratively reweighted least-squares approach to adaptive robust adjustment of parameters in linear regression models with autoregressive and t-distributed deviations. *J. Geod.* **2018**, *92*, 271–297. [[CrossRef](#)]
- Koch, K.R. *Introduction to Bayesian Statistics*, 2nd ed.; Springer: Berlin/Heidelberg, Germany, 2007. [[CrossRef](#)]
- Geweke, J. Bayesian treatment of the independent Student-t linear model. *J. Appl. Econom.* **1993**, *8*, 19–40. [[CrossRef](#)]
- Koch, K.R. Monte Carlo methods. In *Mathematische Geodäsie/Mathematical Geodesy*; Freeden, W., Ed.; Springer Spektrum: Berlin/Heidelberg, Germany, 2020; pp. 445–475. [[CrossRef](#)]

8. Chib, S.; Greenberg, E. Bayes inference in regression models with ARMA(p,q) errors. *J. Econom.* **1994**, *64*, 183–206. [[CrossRef](#)]
9. Barnett, G.; Kohn, R.; Sheather, S. Robust Bayesian estimation of autoregressive-moving-average models. *J. Time Ser. Anal.* **1997**, *18*, 11–28. [[CrossRef](#)]
10. Barnett, G.; Kohn, R.; Sheather, S. Bayesian estimation of an autoregressive model using Markov chain Monte Carlo. *J. Econom.* **1996**, *74*, 237–254. [[CrossRef](#)]
11. Alkhamisi, M.A.; Shukur, G. Bayesian analysis of a linear mixed model with AR(p) errors via MCMC. *J. Appl. Stat.* **2011**, *32*, 741–755. [[CrossRef](#)]
12. Priestley, M.B. *Spectral Analysis and Time Series*; Academic Press: Cambridge, UK, 1981.
13. Bera, A.K.; Jarque, C.M. Model specification tests: A simultaneous approach. *J. Econom.* **1982**, *20*, 59–82. [[CrossRef](#)]
14. McDonald, J.B. Partially adaptive estimation of ARMA time series models. *Int. J. Forecast.* **1989**, *5*, 217–230. [[CrossRef](#)]
15. Tiku, M.L.; Wong, W.-K.; Vaughan, D.C.; Bian, G. Time series models in non-normal situations: symmetric innovations. *J. Time Ser. Anal.* **2000**, *21*, 571–596. [[CrossRef](#)]
16. Koch, K.R. Expectation maximization algorithm and its minimal detectable outliers. *Stud. Geophys. Geod.* **2017**, *61*, 1–18. [[CrossRef](#)]
17. Robert, C.P.; Casella, G. *Monte Carlo Statistical Methods*; Springer: New York, NY, USA, 2005.
18. Paffenholz, J.-A. Direct Geo-Referencing of 3D Point Clouds with 3D Positioning Sensors. Ph.D. Thesis, Deutsche Geodätische Kommission, Munich, Germany, 2012; Series C, No. 689.