Bayesian outlier detection in Geostatistical models

Viviana G. R. Lobo[a] and Thaís C. O. Fonseca[b]

[a]Fluminense Federal University, Rio de Janeiro, Brazil; [b]Federal University of Rio de Janeiro, Rio de Janeiro, Brazil

ARTICLE HISTORY
Compiled October 4, 2016

ABSTRACT

This work considers residual analysis and predictive techniques in the identification of individual and multiple outliers in geostatistical data. The standardized Bayesian spatial residual is proposed and computed for three competing models: the Gaussian, Student-t and Gaussian-log-Gaussian spatial processes. In this context, the spatial models are investigated regarding their plausibility for datasets contaminated with outliers. The posterior probability of an outlying observation is computed based on the standardized residuals and different thresholds for outlier discrimination are tested. From a predictive point of view, methods such as the conditional predictive ordinate, the predictive concordance and the Savage-Dickey density ratio for hypothesis testing are investigated in the identification of outliers in the spatial setting. For illustration, contaminated datasets are considered to access the performance of the three spatial models in the identification of outliers in spatial data.

KEYWORDS
Residual analysis; Spatial statistics; Outlier detection; Predictive performance; Contaminated datasets.

1. Introduction

From a theoretical point of view, statistical inference goes beyond parameter estimation and prediction [see 20, page 343]. Often, tests are performed regarding model parameters which are based on models that are not adequate to the data under study. That is, model adequacy checking should not be based on model parameter testing. Some verification of model goodness of fit is then called for. From a Bayesian perspective, the issue is the same, statements are done regarding the posterior distribution which is also based on the chosen sampling distribution for the data. The usual model criticism is done throught model comparison and prediction for few out of sample observations. Often, these model checkings are not able to access whether the assumed model is plausible for the data. This work is motivated by the idea that model determination or checking should be based on residual analysis, predictive performance and outlier detection. [14] suggest a general framework based on goodness of fitting checking from a Bayesian point of view. This is the kind of approach this paper aims

CONTACT Thais C. O. Fonseca, Av. Athos da Silveira Ramos, Centro de Tecnologia, Bloco C Sala C114 D, IM-UFRJ, CEP 21941-909, Rio de Janeiro, Brazil, Email: thais@im.ufrj.br

to pursue for spatially correlated data modelling.

An stylized fact of statistical applications in general is that if the data contain aberrant observations, the estimated model might not be a good representation of the phenomenon under study, leading to poor predictions for out of sample observations. An important tool in the identification of atypical observations is the residual analysis. In the classical linear regression model, the residuals are usually defined as the difference between observed and fitted values. In the Bayesian context, [4] defines an outlier as an observation with large random error generated by the sampling distribution of the data. In this case, the discrepant observation might be detected through the posterior distribution of these random errors. On the other hand, [11] considers an outlier any observation which was not generated by the mechanism generating the majority of the data. In the independent data setting several papers discussed the detection of outliers such as [23] which considers heavy tailed distributions defined throught a Gaussian mixture model to accommodate and detect outliers in a regression setup.

The focus of this work is the analysis and detection of regional outliers which might be neightboors in space. Indeed, in the spatial statistics context the issue of outlier detection or modeling is even more important than in the independent case. Prediction at new locations is usually based on krigging ideas and krigging predictors are well known to be affected by outliers as they are obtained as linear combination of observations. In geostatistics, an outlier may have a strong effect in the prediction of its neighboors when the observed value for the process at this location is much higher or lower than expected for that region in space. [5] comment that, in applied settings, even small changes in some regions in space might cause large differences between the predicted and observed process. Observations in these regions should not be discarted as this might cause bias in the estimation of parameters and predictions [5, page 221].

Several papers have proposed robust alternatives or modifications of usual kriging predictor. [9] proposed a model to robustify the kriging predictor by defining the model for geostatistical data as a mixture of a spatial process and a contamination process. In this proposal each site has a corresponding contamination variable which indicates whether the site is contaminated or not. The optimal predictor in this case depends on weights which will be affected by the contamination variables. However, the predictor is unfiasible in practice and an approximation is considered. [18] proposed the Gaussian-Log-Gaussian process which is able to capture heterogeneity in space through a mixing process used to increase the Gaussian process variability. This proposal is an alternative to the usual Gaussian process which is very sensitive to outliers. The mixture approach is able to both accommodate and detect outliers. The detection step is done throught hypothesis testing. In particular, [18] considered Bayes factors for that purpose. Notice however that hypothesis testing based on Bayes Factors will depend on the loss function considered to reach a conclusion regarding the outlying observation. Thus, it might be useful to consider other identification techniques jointly with the hypothesis testing.

A potentially robust alternative model for spatial processes is the Student-t process discussed in [21]. However, the Student-t process inflates the variance of the whole process in the presence of outliers in the data and does not allow for individual or regional outlier detection as it does not allow for different kurtosis behaviours across space.

In the literature few proposals deal with model checking or validation for correlated data. In particular, few papers discuss model checking for random functions. [15] describes a deletion scheme for models based on correlated observations. [10] and [16] proposed graphical diagnostics for time series models. [16] propose a rotated residual for independent and time series data which has good asymptotic properties. And [2] propose Bayesian diagnostics for computer models through Cholesky decompositions of the covariance matrix. The proposal results in numerical and graphical tools for model checking in the context of Gaussian processes.

This work proposes to extend the Bayesian residual approach of [4] to accomodate spatially correlated observations. In particular, the data is assumed to vary continuously in a spatial domain of interest D and residual analysis and predictive tecniques are investigated aiming to identify potential outliers or regions of larger variability in the data. The chosen model is crucial in the definition of residuals, thus this paper considers a flexible mixture model for geostatistical data.

The paper is organized as follows. Section 2 describes three competing models for geostatistical data analysis: the Gaussian process, the Student-t process and the Gaussian-Log-Gaussian process. Section 3 presents the proposed spatial residual for outlier identification, discusses the predictive approach and defines a new measure for outlier detection which is based on cross-validation ideas. In addition, the hypothesis test for outlying observations in the context of Gaussian-Log-Gaussian processes is presented for comparison with the other proposed techniques. Section 4 illustrates the methods for outlier detection with contaminated datasets and section 5 concludes.

2. Mixture modelling for outlier detection

As follows we consider a mixture model which mimics a mechanism for outlier generation in a geostatistical context. Consider the spatial process defined in $s \in D$ such that

$$Z(s) = \mathbf{x}^T(s)\boldsymbol{\beta} + \sigma \frac{\tilde{Z}(s)}{\lambda(s)^{1/2}} + \epsilon(s), \tag{1}$$

where $\tilde{Z}(s)$ is a Gaussian process defined in $s \in D$ with zero mean and correlation function $\rho(s, s')$, $s, s' \in D$. The process $\tilde{Z}(s)$ is independent of $\epsilon(s) \sim N(0, \tau^2)$ which models the measurement error parametrized by τ^2, the nugget effect. The mean function depend on covariates $\mathbf{x}^T(s) = (x_1(s), \ldots, x_k(s))$ and $\boldsymbol{\beta}$, a vector of regression coefficcients. The process $\lambda(s)$ is the mixing process allowing for spatial heterogeneity. If $\lambda(s) \neq 1$ the process $Z(s)$ is non-Gaussian. In the absence of nugget effect, the process $\lambda(s)$ must be correlated to induce mean squared continuity of $Z(s)$ [see 18, for details]. Consider $s_1, \ldots, s_n$ spatial locations in D and $\mathbf{Z} = (Z(s_1), \ldots, Z(s_n))$ the observed data at these locations. The models investigated in this work are detailed as follows.

(A) **Gaussian model:** we set $\lambda(s) = 1$, $\forall s \in D$ as a benchmark. The distribution of $\mathbf{Z}$ is

$$\mathbf{Z} \mid \beta, \sigma^2, \tau^2, \theta \sim N_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2\Sigma_\theta + \tau^2 I_n), \tag{2}$$

$\Sigma_{\theta(i,j)} = \rho(s_i, s_j) = \rho(s_i, s_j; \theta)$, i,j=1,. . . ,n.

(B) **Student-t model:** define $\lambda(s) = \lambda$, $\forall s \in D$ such that $\lambda \mid \nu \sim Ga(\nu/2, \nu/2)$.

Then, by marginalization, the distribution of $\mathbf{Z}$ is

$$\mathbf{Z} \mid \beta, \nu, \sigma^2, \tau^2, \theta \sim \text{Student-t}(\nu, \mathbf{X}\boldsymbol{\beta}, \sigma^2\Sigma_\theta + \tau^2 I_n). \tag{3}$$

Similar to the Gaussian process, the Student-t process has the advantage of depending on the mean and covariance functions only for its definition. Details about the Student-t process in a non-Bayesian context may be seen in [21]. Appendix A presents the likelihood function resulting from this model. Appendix B.1 presents the posterior inference considered which is based on Jeffreys independent prior distribution for the unknown degree of freedom parameter as proposed by [7].

(C) **Gaussian-Log-Gaussian model:** consider $ln(\boldsymbol{\lambda}) \mid \nu, \theta \sim N_n\left(-\frac{\nu}{2}1_n, \nu\Sigma_\theta\right)$ with $\boldsymbol{\lambda} = (\lambda(s_1), \ldots, \lambda(s_n))$. Then, the distribution of $\mathbf{Z}$ is

$$\mathbf{Z} \mid \beta, \sigma^2, \tau^2, \mathbf{\Lambda}, \theta \sim N_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2(\mathbf{\Lambda}^{-1/2}\Sigma_\theta\mathbf{\Lambda}^{-1/2}) + \tau^2 I_n), \tag{4}$$

with $\mathbf{\Lambda} = diag(\boldsymbol{\lambda})$. Properties, estimation and predition for the GLG model are introduced in [18] and extendend to the space-time case in [8]. Appendix B.2 presents the posterior inference for the model parameters.

Although the Student-t model allows for variance inflation, it increases the kurtosis for the process in every location and does not allow for individual changes in variability. For the GLG model, if $\lambda(s_k)$ is close to one then the observation is not considered an outlier. However values of $\lambda(s_k)$ close to zero indicate outlying observation. The marginal kurtosis for the process $Z(s)$ is given by $\kappa = 3\exp\{\nu\}$ implying that $\nu \to 0$ results in the Gaussian case with kurtosis 3 and large values of ν indicate fatter tails than the Gaussian model.

In this paper, we investigate the three models in the detection of outliers in spatial data. For that purpose, we compare the performance of methods for outlier detection in the context of correlated data. Furthermore, we propose a new measure based on cross validation ideas and extend the Bayesian residual proposed by [4] to the spatial context.

3. Outlier detection in spatial modelling

In this section we describe three approaches to outlier detection in spatial modelling: the posterior probability computation of a large residual, predictive techniques such as the predictive concordance and the hypothesis testing for the latent mixing variables.

3.1. *Bayesian residual analysis*

Definition 3.1. Consider $\mathbf{Z} = (Z(s_1), \ldots, Z(s_n))$ observations at n spatial locations of the the spatial process $\{Z(s), s \in D\}$ as defined in (1) such that $\mathbf{Z} \mid \beta, \sigma^2, \mathbf{\Lambda} \sim N_n(\mathbf{X}\boldsymbol{\beta}, \sigma^2(\mathbf{\Lambda}^{-1/2}\Sigma_\theta\mathbf{\Lambda}^{-1/2}))$, with $\mathbf{\Lambda} = diag(\boldsymbol{\lambda})$. Then the standardized Bayesian spatial residual for the mixture model without nugget effect is

$$\mathbf{r} = \sigma^{-1}\mathbf{\Lambda}^{1/2}\Sigma_\theta^{-1/2}(Z - X\beta) \tag{5}$$

If the errors are Gaussian distributed then approximately 95% of the individual residuals are expected to be in the interval $[-2, 2]$. If an observation is out of this interval there is some evidence that this observation could be an outlier. In order to detect outlying observations, [4] define the posterior probability that an observation is an outlier as $p_i = P(|r_i| > t \mid \mathbf{z})$. According to [4] the value of t can be chosen so that the prior probability of no outliers is large, say 0.95. The constant t is chosen to be $\Phi^{-1}(0.5 + 0.5(0.95^{1/n}))$. Any observation with posterior probability of being an outlier larger than the prior probability $2\Phi(-t)$ would be suspect. In a context of binary regression [1] and [22] consider $t = 0.75$.

In the geostatistical setting, we will set the value of t to different constants and verify by simulation for several scenarios how the mixture process is sensitive to this choice. It is expected that not all values used for regression model will have good performance in the correlated data context.

Furthermore, we investigate the joint posterior probability of two observations being outliers. This is a phenomena which is expected in spatial applications. In particular, due to spatial correlation of observations and smoothness of the spatial process, two observations which are close together are expected to have large errors if there is a mechanism causing outliers in the spatial region where these two observations are located. Thus, the joint posterior probability that the pair (r_i, r_j) is a regional or multiple outlier is

$$p_{ij} = P(|r_i| > t, |r_j| > t \mid \mathbf{z}). \tag{6}$$

In particular, the variance process $1/\lambda(s)$ in the GLG model is considered to be correlated with $ln(\boldsymbol{\lambda}) \mid \nu \sim N_n\left(-\frac{\nu}{2}1_n, \nu\Sigma_\theta\right)$. Thus, if an individual outlier is detected then this indicates that observations in the close neighborhood are potential outliers. This could be verified computing p_{ij}. [8] extends this proposal by allowing for independent outlying observations through individual nugget effects for each location. This approach is not discussed in this work as replicates in time would be required for parameter estimation.

3.2. *Predictive approach*

An alternative definition of an outlier was given by [13]. An observation is said aberrant or discrepant if it is in the tails of the predictive posterior distribution. The author define the Predictive Concordance for observed value z_i as

$$pc_i = P(z^{rep} > z_i) = \int_{z_i}^{\infty} p(z^{rep}) dz^{rep}, \tag{7}$$

with z^{rep} an imaginary observation and $p(z^{rep})$ the predictive distribution of z^{rep}. This measure is similar to the Bayesian p-value. [13] says that any observation which is in the 2.5% tail of $p(z^{rep})$ should be considered an outlier. The percentage of outliers in the data should be smaller than $(100 - C)\%$ where $C\%$ is the Predictive Concordance. [13] suggests the achievement of 95% predictive concordance for model adequacy.

Notice that the pc_i is computed based on the full predictive distribution, however, to check whether z_i is an outlier it actually uses z_i to obatin the predictive distribution. Thus, the model might predict this observation better than it should if z_i was not in the data. The leave-one-out predictive distribution obtained by removing z_i from the data might give better information about model performance to predict z_i. [13]

proposed the Conditional Predictive Ordinate or CPO

$$CPO_i = p(z_i|\mathbf{z}_{(i)}) = \int p(z_i|\boldsymbol{\theta})p(\boldsymbol{\theta}|\mathbf{z}_{(i)})d\boldsymbol{\theta} \quad (8)$$

where z_i represents a observed value from $\mathbf{z}$ and $\mathbf{z}_{(i)}$ represents the vector $\mathbf{z}$ without z_i. Notice that $p(z^{rep}|\mathbf{z}_{(i)})$ represents the predictive density of a new observation given the dataset which does not include z_i. Values of CPO_i close to zero suggest that observation i is a potential outlier. [19] comments that althought the CPO_i could be used as a surprise index it might return similar values for all observations failling in identifying outlying observations. In these situations [19] suggest a new measure called *Ratio ordinate measure* (ROM) which is the CPO standardized by $max\{p(z^{rep}|z_{(i)})\}$. This measure aims at given more realistic indications of outliers in a dataset.

Following the ideas based on predictive distributions and Predictive Concordance this work proposes a measure for spatial data which is based on cross-validation ideas.

Definition 3.2. The p-value from CPO is defined as

$$CPOp_i = P(z^{rep} > z_i \,|\, \mathbf{z}_{(i)}) \quad (9)$$

The proposed measure is similar to the predictive concordance however it leaves z_i out of the dataset used for parameter estimation. This proposal checks if the observed value z_i is in accordance with the predictive distribution which was obtained excluding z_i from the dataset. [12] argues that z_i should not be used to determine the predictive distribution for model checking. For this measure, values of z_i in either tails of the predictive distribution will indicate that z_i is an outlier.

3.3. *Savage-Dickey ratio test*

A different approach to outlier detection considers inference directly in the mixing process $\lambda(s)$, $s \in D$. Thus, $\lambda(s_k)$ close to one indicates that observation at location s_k is not an outlier. Thus, the model that considers $\lambda(s_k) = 1$ could be compared to the model which considers free $\lambda(s_k)$. This model comparison could be done through Bayes Factors after fitting both models to the data. An alternative to fitting both models and then computing the Bayes Factor [17] is to consider only the model with free $\lambda(s_k)$ and the Savage-Dickey density ratio to approximate the Bayes Factor for the hypothesis that $\lambda(s_k) = 1$ versus $\lambda(s_k) \neq 1$. The Savage-Dickey density ratio was proposed by [6] and can be used when the restriction in the parameter being tested in the null hypothesis is not in the boundary of the parameter domain.

According to [18] the hypothesis testing is useful to indicate outliers in the data or regions with larger variability in space.

The resulting approximation for the Bayes Factor for each location s_k is given by

$$R_k = \frac{p(\lambda(s_k)\,|\,\mathbf{z})}{p(\lambda(s_k))}_{\,|\lambda(s_k)=1} \quad (10)$$

with the ratio R_k being favorable to the model with $\lambda(s_k) = 1$ and all the other λ free against the model with free $\lambda(s_k)$. Thus, small values of R_k (much smaller than 1) will indicate outliers. [17] give some guidelines for interpretation of Bayes Factors. According to the authors, values of R_k smaller than 0.10 give strong evidence that

$\lambda(s_k) \neq 1$. Values between 0.1 and 0.3 give substancial evidence that $\lambda(s_k) \neq 1$, while values between 0.3 and 1 give some evidence but are not very conclusive.

4. Application

4.1. *Simulated dataset*

In this subsection the three models presented in section 2, Gaussian model (GM), Student-t model (STM) and Gaussian log Gaussian model (GLGM) are considered in the identification of outliers in contaminated datasets. In the context of dynamical models [23] comments that contaminated datasets, simulated originally from Gaussian distribution and then contaminated to characterize aberrant observations, are a useful tool to evaluate the performance of robust models. In this simulated study we consider the ideia of Gaussian data contamination in the context of spatial data. Three scenarios of contamination are presented: no contamination, weak contamination and moderate contamination as presented in table 1. Observations in the weak scenario were contaminated summing a random increment $u\sigma$ such that σ is the observational standard deviation and $u \sim Uniform(1, 3.5)$ for observations 1 and 20 and $u \sim Uniform(1, 2.5)$ for observation 6. While in the moderate scenario the random increment was generated from $u \sim Uniform(1, 3.5)$ for observations 1, 13, 15, 16, 20, 30, from $u \sim Uniform(1, 2.5)$ for observations 6 and from $u \sim Uniform(1, 6.5)$ for observations 29. The locations considered for data simulation and contamination are presented in figure 1. Notice that some of the contaminated locations are neighboors in space.

Assume that $Z(s)$ is a spatial process in D. Consider data observed in $n = 30$ spatial locations $\mathbf{z} = (z(s_1), \ldots, z(s_n))'$. Then, $\mathbf{z}$ is simulated from $f_n(\boldsymbol{\mu}, \sigma^2\Sigma_\theta)$, such that $\mu_i = \mu(s_i) = \beta_0 + \beta_1 lat_i + \beta_2 long_i$ and covariance matrix Σ_θ with components $\Sigma_{\theta(ij)} = \exp\{-||s_i - s_j||/\phi\}$, with lat_i and $long_i$ the latitude and longitude for location i, respectively. The parameter values considered for simulation were $\beta_0 = 6,716, \beta_1 = 2,7, \beta_2 = -1,808$ for the mean vector, and $\sigma = 1, \phi = 0,61$ for the covariance matrix parameters.

Samples from the posterior distribution for the parameters of interest are obtained by Markov Chain Monte Carlo simulation. Some details about posterior simulation are presented in the Appendix B. For more details of sampling from parameters from these models see [18].

In this simulated example, it is expected that contaminated observations have large posterior residual probabilities while non-contaminated observations should have small residual probabilities. For the limiar t three choices were considered: $t_1 = 0.75, t_2 = 2, t_3 = 3.1$ with t_1 from [1] and [22], t_2 an arbitrary intermediate constant and t_3 based on [4]. Figure 2 presents the posterior distribution for the residuals for each model and scenario of contamination. Note that in scenario 1 (first row), which has no contamination, the three models have a desirable behavior and do not indicate outlying observations in the data. In scenario 2, only the GLG model identifies the observation 6 as an outlier and the three models identify observations 1 and 20. In the scenario 3, with moderate outliers the GLG model is the only model which has realistic posterior probabilities of large residuals for all the contaminated observations while the Gaussian and Student-t models fails to identify most of these observations.

For the actual identification of outlying observations the limiar t_1, t_2 and t_3 are considered. Table 2 presents the posterior probabilities of atypical observations for

t_3, $p_i(|r_i| > t_3 \mid \mathbf{z})$. In the scenario 1 (no contamination) it is expected that the probabilities of outliers are small. However, if limiars t_1 and t_2 are considered these probabilities are unexpectedly high. Thus, this suggests that only limiar t_3 should be considered for outlier identification in the spatial data scenario simulated in this work.

Contaminated observations in scenarios 2 and 3 are presented in tables 3 and 4. The GLG model give large probabilities of outliers for the contaminated observations while gives small probabilities for observations which were not contaminated. The Gaussian and Student-t models fail to identify the contaminated observation 6 as an outlier. In scenario 2 with weak contamination, the GLG model gave a probability of 0.228 for observation 6 being an outlier while the Gaussian and Student-t models had 0.000 and 0.001, respectively. For instance, in scenario 3, the models gave probabilities for the contaminated observation 30 of being an outlier 0.000 (GM),0.042 (STM) and 0.795 (GLGM). Thus, the GLG was the only model to correctly identify all the outlying observations in the data. Notice that in this simulated example no observation was classified as an outlier when it was not contaminated.

In addition, scenarios 2 and 3 are investigated for multiple outliers. In the context of geostatistical data, it is expected that outliers are not observed isolated but in a region in space. In this direction, the joint probability of multiple outliers was computed p_{ij} as in equation (6). Table 5 presents the probabilities for a set of pairs in the sampled locations and the corresponding posterior correlation.

Once again the GLG model is the only model which correctly give reasonably large posterior probabilities for all pairs of outliers in scenario 2. The Gaussian and Student-t models give reasonably large probabilities for the pair $(1, 20)$ however give close to zero posterior probability for the pair $(1, 6)$ and $(6, 20)$. This is explained by the fact that the Gaussian and Student-t model do not model heterogeneity in space.

Table 6 presents the results for scenario 3. Only the pair $(1, 15)$ have reasonably large posterior probabilities of being multiple outliers in the Gaussian and Student-t models althought all the other pairs in the table have very large spatial correlation. The GLG model correctly indicates all the contaminated pairs as multiple outliers.

As follows, the predictive measures were computed for the contaminated scenarios. In particular, tables 7 to 9 present the pc_i as presented in (7) for the scenarios: no outliers, weak outliers and moderate outliers, respectively. If the observation is not in the 5% tail it is not classified as outlier. Observation in the tail are indicated as potential outliers.

The CPO is misleading in scenario 1, resulting in cpo_i of 0.006 (GM), 0.000 (STM) and 0.137 (GLG) for observation 27 which would indicate an outlier, however, no observation was contaminated in this scenario. The predictive concordance pc_i and the proposed p-value based on CPO behave as expect and do not indicate any observation as an outlier in this scenario. For scenario 3, the advantages of the proposed p-value is even more evident. In scenario 2, observation 27 was not contaminated however, cpo_i was 0.039 (GM), 0.000 (STM) and 0.000 (GLGM) indicating an outlier. While the proposed CPO_p were 0.769 (GM), 0.898 (STM) and 0.385 (GLGM) which correctly do not indicate outlying observation. The predictive concordance also give reasonable results for observation 27: 0.742 (GM), 0.651 (STM) and 0.486 (GLGM). In scenario 3, while the cpo_i is smaller than 10^{-4} for observation 27 (non-contaminated) indicating an outlier, the CPO_p is 0.281 and the pc_i is 0.402. An analogous behaviour is obtained for observation 3 which was not contaminated.

Table 10 indicate strong evidence that $\lambda_i \neq 1$ for all contaminated observations in the hypothesis testing for scenario 2. For instance, for observation 1, the Bayes Factor in favour of free λ_i is 0.011 giving strong evidence of free $\lambda(s)$ in this location. The results for scenario 3 with moderate contamination are presented in table 11. The ratio R_i correctly indicates all contaminated observation as outliers such as observation 6 which has $R_i = 0.05$ indicating strong evidence of an outlying observation.

Table 1. Contamination scenarios.

Scenario 1	no outliers in the data
Scenario 2	weak outliers: 3 points were contaminated
Scenario 3	moderate outliers: 8 points were contaminated

Table 2. Standardized residuals and posterior probabilities of outliers, $p_i(|r_i| > t|\mathbf{z})$ for scenario 1 (no outlier).

Scenario 1 (no outlier)												
	Gaussian				*Student-t*				*GLG*			
i	r_i	$p(t_1)_i$	$p(t_2)_i$	$p(t_3)_i$	r_i	$p(t_1)_i$	$p(t_2)_i$	$p(t_3)_i$	r_i	$p(t_1)_i$	$p(t_2)_i$	$p(t_3)_i$
1	0.730	0.489	0.033		0.738	0.530	0.044		1.170	0.742	0.298	0.062
3	-0.575	0.403	0.020		-0.547	0.420	0.021		-0.467	0.519	0.084	0.011
6	-0.100	0.296	0.002		-0.033	0.269	0.006		-0.181	0.474	0.073	0.008
15	0.759	0.536	0.039		0.794	0.551	0.058	0.003	1.372	0.772	0.301	0.072
20	0.495	0.391	0.020		0.495	0.392	0.024		0.928	0.663	0.193	0.030
27	-0.724	0.517	0.082	0.003	-0.580	0.480	0.058	0.002	-0.691	0.645	0.208	0.038
30	0.074	0.309	0.007		0.143	0.317	0.007		0.443	0.519	0.111	0.009

[a]Posterior probabilities smaller than 10^{-3} are omitted from the table.

Table 3. Standardized residuals and posterior probabilities of outliers, $p_i(|r_i| > t|\mathbf{z})$ for scenario 2 (weak outliers). bold-faced observations represent contaminated observations.

Scenario 2 (weak outlier)												
	Gaussian				*Student-t*				*GLG*			
i	r_i	$p(t_1)_i$	$p(t_2)_i$	$p(t_3)_i$	r_i	$p(t_1)_i$	$p(t_2)_i$	$p(t_3)_i$	r_i	$p(t_1)_i$	$p(t_2)_i$	$p(t_3)_i$
1	3.711	1.000	0.998	**0.797**	3.731	1.000	0.995	**0.839**	5.426	1.000	1.000	**0.997**
3	-0.796	0.524	0.007		-0.639	0.455	0.016		0.222	0.432	0.053	0.005
6	0.920	0.618	0.010		1.050	0.698	0.092	**0.001**	2.364	0.965	0.664	**0.228**
15	0.680	0.473	0.004		0.756	0.522	0.044		1.883	0.877	0.436	0.112
20	2.622	0.999	0.848	**0.114**	2.668	0.999	0.863	**0.299**	4.300	0.999	0.995	**0.904**
27	-0.871	0.548	0.060		-0.592	0.505	0.053	0.001	0.137	0.578	0.143	0.033
30	-0.056	0.152			0.102	0.273	0.005		1.058	0.667	0.209	0.045

[a]Posterior probabilities smaller than 10^{-3} are omitted from the table.

Table 4. Standardized residuals and posterior probabilities of outliers, $p_i(|r_i| > t|\mathbf{z})$ for scenario 3 (moderate outliers). Bold-faced observations represent contaminated observations.

	Scenario 3 (moderate outlier)											
	Gaussian				*Student-t*				*GLG*			
i	r_i	$p(t_1)_i$	$p(t_2)_i$	$p(t_3)_i$	r_i	$p(t_1)_i$	$p(t_2)_i$	$p(t_3)_i$	r_i	$p(t_1)_i$	$p(t_2)_i$	$p(t_3)_i$
1	3.71	1.000	0.979	**0.344**	3.080	0.999	0.925	**0.456**	5.338	1.000	1.000	**0.986**
3	-0.796	0.756	0.008		-0.958	0.630	0.044	0.001	0.507	0.478	0.072	0.007
6	0.920	0.114			0.472	0.345	0.017		2.628	0.969	0.733	**0.310**
15	0.680	1.000	0.993	**0.723**	3.817	1.000	0.994	**0.805**	5.734	1.000	1.000	**0.996**
20	2.622	0.997	0.563	**0.002**	2.135	0.971	0.574	**0.066**	4.273	1.000	0.989	**0.896**
27	-0.871	0.959	0.335	0.023	-1.459	0.808	0.254	0.017	0.702	0.624	0.244	0.067
30	-0.056	0.965	0.192		1.937	0.957	0.443	**0.042**	3.988	1.000	0.975	**0.795**

[a]Posterior probabilities smaller than 10^{-3} are omitted from the table.

Table 5. Posterior probabilities of multiple outliers $p_{ij} = p(|r_i| > t_3, |r_j| > t_3|\mathbf{z})$ and posterior correlation ρ_{ij} for r_i and r_j, for each model in scenario 2.

i,j	**Gaussian**	**Student-t**	**GLG**	**Correlation** ρ_{ij}
1,6		0.001	0.228	0.869
1,20	**0.114**	**0.299**	**0.904**	0.950
6,20		0.001	0.227	0.854

[a]Posterior probabilities smaller than 10^{-3} are omitted from the table.

Table 6. Posterior probabilities of multiple outliers $p_{ij} = p(|r_i| > t_3, |r_j| > t_3|\mathbf{z})$ and posterior correlation ρ_{ij} for r_i and r_j, for each model in scenario 3.

i,j	**Gaussian**	**Student-t**	**GLG**	**Correlation** ρ_{ij}
1,15	**0.307**	**0.433**	**0.982**	0.834
1,20	0.002	0.066	**0.896**	0.958
1,29		0.006	0.603	0.850
1,30		0.042	0.794	0.847
6,20			0.307	0.875
15,20	0.002	0.066	**0.893**	0.933

[a]Posterior probabilities smaller than 10^{-3} are omitted from the table.

Table 7. pc_i,cpo_i, $CPOp_i$ for some observations in the sample. Bold-faced observations represent the contaminated observations.

	Scenario 1								
	Gaussian			**Student-T**			**GLG**		
Obs.	pc_i	cpo_i	$CPOp_i$	pc_i	cpo_i	$CPOp_i$	pc_i	cpo_i	$CPOp_i$
1	0.271	0.145	0.257	0.297	0.039	0.201	0.252	0.001	0.261
3	0.685	0.200	0.712	0.680	0.016	0.769	0.611	0.259	0.261
15	0.278	0.112	0.262	0.267	0.011	0.197	0.226	0.101	0.235
20	0.337	0.227	0.341	0.361	0.017	0.286	0.307	0.174	0.291
27	0.678	0.006	0.700	0.647	0.000	0.873	0.643	0.137	0.627

[a]Values smaller than 10^{-4} are omitted from the table.

Table 8. pc_i,cpo_i, $CPOp_i$ for some observations in the sample. Bold-faced observations represent the contaminated observations.

	Scenario 2								
	Gaussian			**Student-T**			**GLG**		
Obs.	pc_i	cpo_i	$CPOp_i$	pc_i	cpo_i	$CPOp_i$	pc_i	cpo_i	$CPOp_i$
1	**0.001**			**0.001**			**0.013**		**0.019**
3	0.767	0.2395	0.780	0.700	0.0577	0.886	0.457	0.239	0.507
15	0.264	0.281	0.243	0.269	0.173	0.199	0.177	0.232	0.277
20	**0.009**	0.003	0.001	**0.016**			**0.031**	0.004	**0.110**
27	0.742	0.039	0.769	0.651		0.898	0.486		0.385

[a]Values smaller than 10^{-4} are omitted from the table.

Table 9. pc_i,cpo_i, $CPOp_i$ for some observations in the sample. Bold-faced observations represent the contaminated observations.

	Scenario 3								
	Gaussian			Student-T			GLG		
Obs.	pc_i	cpo_i	$CPOp_i$	pc_i	cpo_i	$CPOp_i$	pc_i	cpo_i	$CPOp_i$
1	**0.002**	0.000	0.004	**0.006**	0.004	**0.037**	**0.031**		**0.070**
3	0.827	0.192	0.835	0.801	0.014	0.964	0.402		0.257
15	**0.004**				0.031	**0.046**	**0.094**		**0.044**
20	**0.025**	0.022	0.035	0.001	**0.017**	**0.017**	**0.025**		**0.115**
27	0.945	0.006	0.968	0.851		1	0.402		0.281

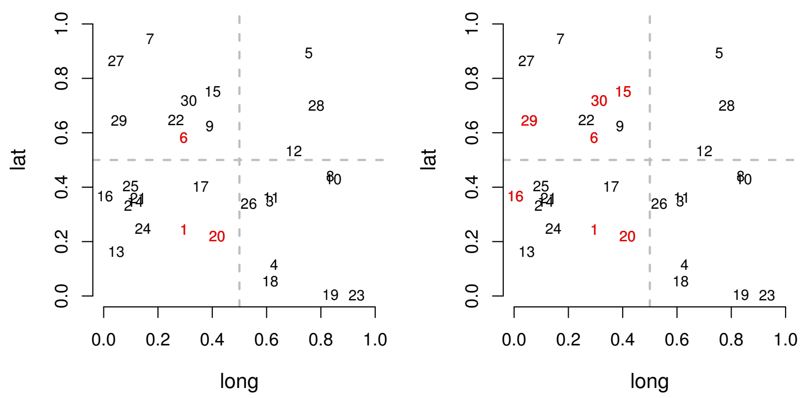
[a]Values smaller than 10^{-4} are omitted from the table.

Table 10. Savage-Dickey density ratio R_i for hypotesis testing in favour of $\lambda_i = 1$ in scenario 2. Bold-faced observations represent the contaminated observations.

obs.	$E(\lambda_i\|\mathbf{z})$	$SD(\lambda_i\|\mathbf{z})$	R_i
[1]	0.271	0.119	**0.011**
[3]	0.601	0.321	0.454
[6]	0.516	0.193	**0.149**
[20]	0.309	0.141	**0.016**
[27]	0.643	0.280	0.596

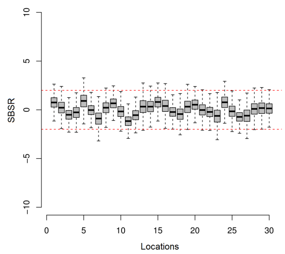
Table 11. Savage-Dickey density ratio R_i for hypotesis testing in favour of $\lambda_i = 1$ in scenario 3. Bold-faced observations represent the contaminated observations.

obs.	$E(\lambda_i\|\mathbf{z})$	$SD(\lambda_i\|\mathbf{z})$	R_i
[1]	0.199	0.097	**0.0002**
[3]	0.573	0.352	0.381
[6]	0.353	0.175	**0.050**
[15]	0.188	0.089	**0.001**
[20]	0.242	0.120	**0.017**
[27]	0.434	0.246	0.596

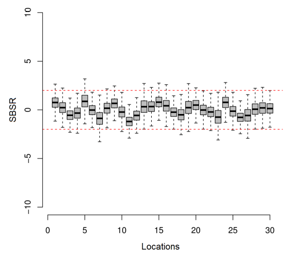


(a) Weak outliers. (b) Moderate outliers.

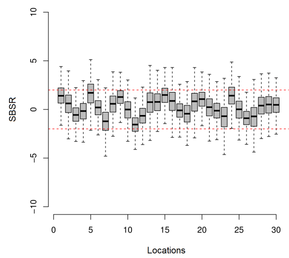
Figure 1. Spatial location considered for data simulation for the contaminated data scenarios. Locations in red represent the contaminated locations in each scenario. Plot (a) represents scenario 2 with weak outliers and (b) represents scenario 3 with moderate outliers.



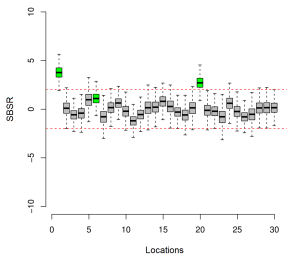
(a) GM.



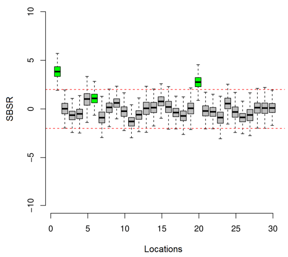
(b) STM.



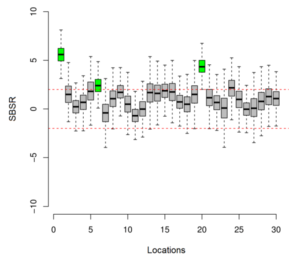
(c) GLGM.



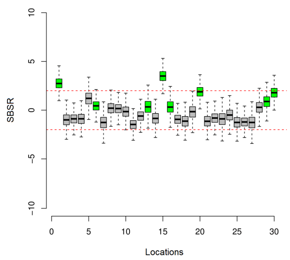
(d) GM.



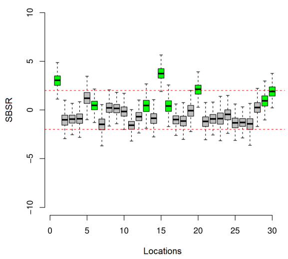
(e) STM.



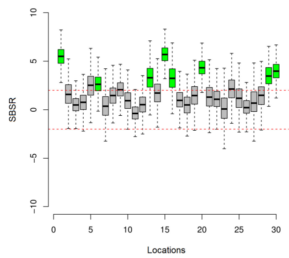
(f) GLGM.



(g) GM.



(h) STM.



(i) GLGM.

Figure 2. Posterior distribution for the residuals in each scenario: no contamination (first row), weak contamination (second row) and moderate contamination (tird row).

5. Conclusions

The idea that the Student-t process would result in robustness to outliers is misleading as the degree of freedom is the same in all spatial locations leading to inflation in the global variance. However, the model does not allow for actual detection of regions with aberrant observation in spatial data. These ideas were discussed in [3] and [18]. This work adds in this direction by presenting simulation for spatial contaminated datasets and obtaining no outlier detection for most of outliers in the data using the Student-t model. The Student-t process for spatial data is not able to detect outliers either by marginal probabilities or multiple detection procedures. The GLG mixture process, on the other hand, indicates the outliers in all simulated scenarios for all detection tools presented in this work. The residual analyses presented here is purely spatial in the sense that the mixture process is considered in space only. [8] considered a mixture model in space-time which is not exploit in this work.

The spatial Bayesian residuals, the proposed cross-validation p-value and the Savage-Dickey density indicate correctly the outliers and multiple outliers under the GLG model assumption. The CPO fail in identifying outliers for all models, giving misleading indications for the contaminated scenarios. As an alternative, this work proposed the p-value based on the CPO, which is able to correctly detect aberrant observations in the spatial data.

Acknowledgements

This work was part of the Master's dissertation of Viviana G. R. Lobo under the supervision of T. C. O. Fonseca. Lobo benefited from a scholarship from Conselho de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Brazil. T. C. O. Fonseca was partially supported by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq).

References

[1] Albert, J. and Chib, S. *Bayesian residual analysis for binary response regression models.* Biometrika, 82 (1995), pp. 747-759.

[2] Bastos, L. S. and OHagan, A. *Diagnostics for Gaussian process emulators.* Technometrics, 51 (2008), pp. 425-438.

[3] Breusch, T. S., Robertson, J. C., and Welsh, A. H. *The emperors new clothes: a critique of the multivariate t regression model.* Statistica Neerlandica, 1997.

[4] Chaloner, K. and Brant, R. *A Bayesian approach to outlier detection and residual analysis.* Biometrika, 75, 4 (1988), pp. 651-659.

[5] Chilès, J.-P. and Delfiner, P. *Modeling Spatial Uncertainty.* New York: Wiley, 1999.

[6] Dickey, J. *The Weighted Likelihood Ratio, Linear Hypotheses on Normal Location Parameters.* The Annals of Statistics, 42, 1 (1971), pp. 204-223.

[7] Fonseca, T. C. O. and Ferreira, M. A. R. and Migon, H. S. *Objective Bayesian analysis for the Student-t regression model* Biometrika, 95, 2 (2008), pp. 325–333.

[8] Fonseca, T. C. O. and Steel, M. F. J. *Non- Gaussian Spatiotemporal Modelling through Scale Mixing.* Biometrika, 98, 4 (2011), pp. 761-774.

[9] Fournier, B. and Furrer, R. *Automatic mapping in the presence of substutive errors: a robust kriging approach.* Applied GIS, 1 (2005), pp. 121-1216.

[10] Fraccaro, R., Hyndman, R. J., and Veevers, A. *Residual Diagnostic Plots for Checking*

for Model Misspecification in Time Series Regression. Australian & New Zealand Journal of Statistics, 42, 4 (2000), pp. 463-477.
[11] Freeman, P. R. *On the number of outliers in data from a linear model*, 349365. Valencia: University Press, 1980.
[12] Gelfand, A. *Model determination using predictive distributions with implementation via sampling-based methods.* Technical report, Department of Statistics, Stanford University 1992.
[13] Gelfand, A. *Model Determination Using Samplings Based Methods.* Chapman & Hall, Boca Raton, FL, 1996.
[14] Gelman, A., Meng, X.-L., and Stern, H. *Posterior predictive assessment of model fitness via realized discrepancies.* Statistica Sinica, 6 (1996), pp. 733-807.
[15] Hasllet, J. *A Simple Derivation of Deletion Diagnostic Results for the General Linear Model with Correlated Errors.* Journal of the Royal Statistical Society: Series B (Statistical Methodology), 61, 3 (1999), pp. 603-609.
[16] Houseman, E. A., Ryan, L., and Coull, B. *Cholesky Residuals for Assessing Normal Errors in a Linear Model with Correlated Outcomes.* Journal of the American Statistical Association, 99 (2004), pp. 383-394.
[17] Kass, R. E. and Raftery, A. E. *Bayes Factors.* Journal of the American Statistical Association, 90, 430 (1995), pp. 773-795.
[18] Palacios, M. B. and Steel, M. F. J. *Non-Gaussian Bayesian Geostatistical Modeling.* Journal of the American Statistical Association, 101, 474 (2006), pp. 604-618.
[19] Petit, L. *The conditional predictive ordinate for the normal distribution.* Journal of the Royal Statistical Society: Series B (Statistical Methodology), 52, 21 (1990), pp. 175-184.
[20] Robert, C. P. *The Bayesian Choice.* Second edition ed. Springer, New York, 2007.
[21] Roislien, J. and Omre, H. *T-distributed Random Fields: A Parametric Model for Heavy-tailed Well-log Data.* Mathematical Geology, 38, 7 (2006), pp. 821849.
[22] Souza, A. D. P. and Migon, H. S. *Bayesian outlier analysis in binary regression.* Journal of Applied Statistics, 37, 8 (2010), pp. 1355-1368.
[23] West, M. *Outlier models and prior distributions in Bayesian linear regression.* Journal of the Royal Statistical Society: Series B (Statistical Methodology), 48, 3 (1984), pp. 431-439.

Appendix A. Student-t Process

The joint density function for the spatial data $\mathbf{z} = (z(s_1), \dots, z(s_n))$ in the Student-t model is given by

$$p(\mathbf{z} \mid \boldsymbol{\mu}, \sigma^2, \nu, \theta) = \frac{\Gamma(\frac{\nu+n}{2})}{\Gamma(\frac{\nu}{2})(\sigma^2\nu\pi)^{n/2}} |\Sigma_\theta|^{-1/2} \left[1 + \frac{(\mathbf{z}-\boldsymbol{\mu})^T \Sigma_\theta^{-1} (\mathbf{z}-\boldsymbol{\mu})}{\sigma^2 \nu}\right]^{-\nu+n/2} \tag{A1}$$

$\mathbf{z} \in \Re^n$, $\nu, \sigma^2 > 0$.

Appendix B. Markov Chain Monte Carlo Sampler

The prior distributions considered for the parameters, the complete conditional distributions and proposal densities used in the MCMC algorithm are detailed as follows.

B.1. ***Student-t Bayesian model***

Consider the marginalized model (A1) obtained by integrating the mixing variable λ.

$$\mathbf{z}|\boldsymbol{\beta},\sigma^2,\phi,\nu \sim Student-t_n(\boldsymbol{\mu},\nu,\Sigma_{\boldsymbol{\theta}}), \tag{B1}$$

with $\Sigma_{\boldsymbol{\theta}(ij)} = \sigma^2(1+(||s_i - s_j||/\phi)^{1.5})^{-1}$.

(1) Prior distribution: $\sigma^2 \sim GI(a,b)$, $a,b>0$. Thus,

$$p(\sigma^2\,|\,\mathbf{z},\boldsymbol{\beta},\phi,\nu) \quad \propto \quad \underbrace{p(\mathbf{z}\,|\,\boldsymbol{\beta},\phi,\sigma^2,\nu)\pi(\sigma^2)}_{\text{Metropolis-Hastings Step}}$$

The proposal density in the MCMC sampler is:

$$ln(\sigma^2) \sim Normal(ln(\sigma^{2(k-1)}),\sigma^2_{(\sigma^2)}).$$

(2) Prior distribution: $\boldsymbol{\beta} \sim Normal_n(\mathbf{0},\tau^2_\beta I_n)$, $\tau^2_\beta>0$. Thus,

$$p(\boldsymbol{\beta}\,|\,\mathbf{z},\sigma^2,\phi,\nu) \quad \propto \quad \underbrace{p(\mathbf{z}\,|\,\boldsymbol{\beta},\sigma^2,\phi,\nu)\pi(\boldsymbol{\beta})}_{\text{Metropolis-Hastings Step}}$$

The proposal density in the MCMC sampler is:

$$\beta \sim Normal(\beta^{(k-1)},\sigma^2_{(\beta)}).$$

(3) Prior distribution: $\phi \sim Gama(1,c/med(u_s))$, with $c>0$ and $med(u_s)$ the median distance in the observed data. Thus,

$$p(\phi|\mathbf{z},\boldsymbol{\beta},\sigma^2,\nu) \quad \propto \quad \underbrace{p(\mathbf{z}|\boldsymbol{\beta},\sigma^2,\phi,\nu)\pi(\phi)}_{\text{Metropolis-Hastings Step}}$$

The proposal density in the MCMC sampler is:

$$ln(\phi) \sim Normal(ln(\phi^{(k-1)}),\sigma^2_{(\phi)}).$$

(4) Jeffreys independent prior distribution [7]:

$$p(\nu) \propto \left(\frac{\nu}{\nu+3}\right)^{1/2}\left\{\psi'\left(\frac{\nu}{2}\right)-\psi'\left(\frac{\nu+1}{2}\right)-\frac{2(\nu+3)}{\nu(\nu+1)^2}\right\}^{1/2},$$

with $\psi'(a)=\frac{d\{\psi(a)\}}{da}$ the trigamma function. In the context of regression models, this prior distribution garantees that the posterior distribution for ν is proper. Thus,

$$p(\nu|\mathbf{z},\boldsymbol{\beta},\sigma^2,\phi) \quad \propto \quad \underbrace{p(\mathbf{z}|\boldsymbol{\beta},\sigma^2,\phi,\nu)\pi(\nu)}_{\text{Metropolis-Hastings step}}$$

The proposal density in the MCMC sampler is:

$$ln(\nu) \sim Normal(ln(\nu^{(k-1)}), \sigma^2_{(\nu)}).$$

B.2. *GLG Bayesian model*

We follow [18] in the simulation from the posterior distribution for parameters in the GLG model. The vector $\mathbf{z}$ has conditional distribution given by

$$f(\mathbf{z} \,|\, \boldsymbol{\beta}, \boldsymbol{\theta}, \sigma^2, \Lambda) \sim Normal_n(\mu, \sigma^2 \Lambda^{-1} \Sigma_{\boldsymbol{\theta}} \Lambda^{-1})$$

with $\Lambda = diag(\lambda_1, \ldots, \lambda_n)$ and $\boldsymbol{\theta} = \phi$ the spatial range parameter. Define $\Sigma^*_{\boldsymbol{\theta}} = \Lambda^{-1}\Sigma(\boldsymbol{\theta})\Lambda^{-1}$.

(1) Prior distribution: $\sigma^2 \sim GI(a, b)$, $a, b > 0$. Thus,

$$\begin{aligned} p(\sigma^2 \mid \mathbf{z}, \boldsymbol{\beta}, \boldsymbol{\theta}, \boldsymbol{\lambda}, \nu) &\propto p(\mathbf{z} \,|\, \boldsymbol{\beta}, \boldsymbol{\theta}, \sigma^2, \boldsymbol{\lambda}, \nu)\pi(\sigma^2) \\ &\propto (\sigma^2)^{-(a+n/2+1)} exp\left\{ -\frac{1}{\sigma^2}\left[\frac{1}{2}(\mathbf{z}-\boldsymbol{\mu})'\Sigma^{*(-1)}_{\boldsymbol{\theta}}(\mathbf{z}-\boldsymbol{\mu})\right] + b \right\} \end{aligned}$$

Thus, $\sigma^2 \mid \Phi \sim IGamma\left(a + \frac{n}{2}, \frac{1}{2}(\mathbf{z}-\boldsymbol{\mu})'\Sigma^{*(-1)}_{\boldsymbol{\theta}}(\mathbf{z}-\boldsymbol{\mu}) + b\right)$.

(2) Prior distribution: $\boldsymbol{\beta} \sim Normal_n(\mathbf{0}, \tau^2_\beta I_n)$, $\tau^2_\beta > 0$. Thus,

$$\begin{aligned} p(\boldsymbol{\beta} \mid \mathbf{z}, \nu, \sigma^2, \phi, \boldsymbol{\lambda}) &\propto p(\mathbf{z} \mid \boldsymbol{\beta}, \sigma^2, \phi, \boldsymbol{\lambda}, \nu)\pi(\boldsymbol{\beta}) \\ &\propto \exp\left\{ -\frac{1}{2}\left[(\mathbf{z}-\boldsymbol{\mu})'\sigma^{-2}\Sigma^{*(-1)}_{\boldsymbol{\theta}}(\mathbf{z}-\boldsymbol{\mu}) + \tau_\beta^{-2}\boldsymbol{\beta}'\boldsymbol{\beta}\right]\right\} \end{aligned}$$

As a result $\beta \mid \Phi \sim Normal_n\,(m, M)$ with

$$M = \left(\frac{I_p}{\tau^2_\beta} + \frac{\mathbf{X}'\Sigma^{*(-1)}_{\boldsymbol{\theta}}}{\sigma^2}\right)^{-1} \qquad \text{e} \qquad m = M \times \left(\frac{1}{\tau^2_\beta} + \frac{\mathbf{X}'\mathbf{z}}{\sigma^2}\right)$$

(3) Prior distribution: $\nu \sim GIG(\zeta, \delta, \iota)$

$$\begin{aligned} p(\nu \mid \mathbf{z}, \boldsymbol{\beta}, \boldsymbol{\theta}, \boldsymbol{\lambda}, \sigma^2) &\propto p(\boldsymbol{\lambda} \mid \nu)\pi(\nu) \\ &\propto \nu^{\zeta-n/2-1} exp\left\{ -\frac{1}{2\nu}\left[\left(ln\boldsymbol{\lambda} + \frac{\nu}{2}\right)^T \Sigma^{*(-1)}_{\boldsymbol{\theta}} \left(ln\boldsymbol{\lambda} + \frac{\nu}{2}\right) + \delta^2\right] - \frac{1}{2}\iota^2\nu \right\} \end{aligned}$$

Thus, $\nu \mid \Phi \sim GIG\left(\zeta - \frac{n}{2}, \delta^2 + \iota^2\right)$ and n represents the dimension of $\Sigma^*_{\boldsymbol{\theta}}$.

(4) Prior distribution: $\phi \sim Gama(1, c/med(u_s))$, with $c > 0$ and $med(u_s)$ the median distance in the observed data. Thus,

$$p(\phi \,|\, \mathbf{z}, \boldsymbol{\beta}, \nu, \boldsymbol{\lambda}, \sigma^2) \quad \propto \quad \underbrace{p(\mathbf{z} \,|\, \boldsymbol{\beta}, \sigma^2, \boldsymbol{\lambda}, \nu)\pi(\phi)}_{\text{Metropolis-Hastings step}}$$

The proposal density in the MCMC sampler is:

$$ln(\phi) \sim Normal(ln(\phi^{(k-1)}), \sigma^2_{(\phi)}).$$

(5) Prior distribution: $\boldsymbol{\lambda} \mid \nu, \phi \sim Log - Normal\left(-\frac{\nu}{2}\mathbf{1}, \nu\Sigma_{\boldsymbol{\theta}}\right)$

$$p(\boldsymbol{\lambda} \,|\, \phi, \nu, \mathbf{z}, \boldsymbol{\beta}, \sigma^2) \quad \propto \quad \underbrace{p(\mathbf{z} \,|\, \boldsymbol{\lambda}, \phi, \nu, \boldsymbol{\beta}, \sigma^2)\pi(\boldsymbol{\lambda} \,|\, \nu)}_{\text{Metropolis-Hastings step}}$$

The spatial region is divided in subregions and a random walk proposal density is used for each subregion. [18] propose a independent sampler which might be more efficient than random walk proposals in the case of large datasets.