Accounting for covariate information in the scale component of spatio-temporal mixing models

Renata S. Bueno[1], Thaís C. O. Fonseca[2*], and Alexandra M. Schmidt[3]

[1] *Escola Nacional de Ciências Estatísticas, Brazil*

[2] *Instituto de Matemática,*

Universidade Federal do Rio de Janeiro, Brazil

[3] *Department of Epidemiology, Biostatistics and Occupational Health,*

McGill University, Canada

July 2017

Abstract

Spatio-temporal processes in the environmental science are usually assumed to follow a Gaussian process, possibly after some transformation. Gaussian processes might not be appropriate to handle the presence of outlying observations. Our proposal is based on the idea of modelling the process as a scale mixture between a Gaussian and log-Gaussian process. And the novelty is to allow the scale process to vary as a function of covariates. The resultant model has a nonstationary covariance structure in space. Moreover, the resultant kurtosis varies with location, allowing the time series at each location to have different distributions with different tail behaviour. Inference procedure is performed under the Bayesian framework. The analysis of an artificial dataset

[*] *Corresponding author*: Thaís C. O. Fonseca, Departamento de Métodos Estatísticos, Instituto de Matemática, Universidade Federal do Rio de Janeiro, Av. Athos da Silveira Ramos, Centro de Tecnologia, Bloco C, CEP 21941-909.

E-mail: thais@im.ufrj.br. *Homepage*: https://sites.google.com/site/thaisf/

illustrates how this proposal is able to capture heterogeneity in space caused by dependence on some spatial covariate or by a transformation of the process of interest. Furthermore, an application to maximum temperature data observed in the Spanish Basque country illustrates the effects of altitude in the variability of the process and how our proposed model identifies this dependence through parameters which can be interpreted as regression coefficcients in the variance model.

Keywords: Bayesian inference; Heavy-tailed; Non-stationarity; Non-Gaussian process.

1 Introduction

The development of methods for the analysis of spatio-temporal processes has increased considerably in the recent years due to computational advances which enable analysis of possibly high-resolution data observed across space and time. The models that describe these processes incorporate spatial and temporal dependencies among observations in order to better understand the behaviour of the response variable and improve predictions for future times or unsampled sites. Usually, the models used to describe spatio-temporal processes are based on Gaussian processes. However, real data distributions often deviate from Gaussianity, presenting heavy tails or skewness. There are many practical applications in environmental, hydrological, and ecological studies in which Gaussianity is an unrealistic assumption. For data sets with non-Gaussian characteristics, a widely used approach is to find some nonlinear transformation for the data so that the assumption of normality for the transformed data holds. This approach is commonly known as trans-Gaussian Kriging (Cressie, 1993) and common transformations include the logarithm and the square-root.

A different approach is used by Higdon (2002) to construct Gaussian processes by convolving white noise processes with a spatially varying kernel aiming to accommodate non-stationarity. Although this proposal could be used to allow for non-Gaussian behaviour in spatial data, it is computationally intensive and lacks interpretability. Bolin (2014), on the other hand, proposes a non-Gaussian model with Matérn covariance functions formulated as a stochastic partial differential equation driven by a non-Gaussian noise. The estimation procedure proposed by Bolin (2014) is based on the use of the expectation-maximization

(EM) algorithm. Wallin and Bolin (2015) extend this idea using a Monte Carlo EM algorithm which is useful for practical applications. In this paper we pursue a different direction by considering scale mixture of Gaussian processes.

We focus on non-Gaussian processes defined through a scale mixture that results in a Gaussian-log-Gaussian model (GLG). The mixing process in this model formulation is defined by latent variables which describe the variance of the process under study. The model is able to accommodate spatial heteroskedasticity and heavier tails than the usual Gaussian process. The approach considered in this work was initially introduced by Palacios and Steel (2006), and extended to the context of spatio-temporal processes by Fonseca and Steel (2011). Our aim is to propose a flexible model for spatio-temporal processes which is able to accommodate non-Gaussian tail behaviour through the inclusion of covariates in the mixing distribution specification. It is expected that the use of covariates brings more information about the variance and the tail behaviour of the process. Different from the Gaussian, Student-t and Gaussian-log-Gaussian processes, the kurtosis of our proposed model varies with location, allowing the model to accommodate different distributions across space.

1.1 Including covariates in the covariance function of a process

Recently there has been the discussion of including covariate information in the covariance function of spatio-temporal processes. The goal is usually to consider more flexible models for spatial processes when usual setups fail to accommodate heterogeneity. Ver Hoef et al. (2006) propose a spatial model whose covariance structure incorporates covariates through spatial moving averages. Cooley et al. (2007) capture non-stationarity by modeling extreme precipitation as a function of geographical and climatological covariates. Calder (2008) considers wind direction information in the convolution approach for wind modeling. Schmidt et al. (2011) propose a model that allows for both spatial coordinates and covariates to define the latent domain in the deformation approach of Sampson and Guttorp (1992). Reich et al. (2011) consider a spatial model which is a linear combination of stationary fields with different covariance functions, such that the weights in the combination depend on covariates. Viana Neto et al. (2014) present a convolution model which includes wind direction in the covariance function for the process of interest. Ingebrigtsen et al. (2014)

incorporates covariates in stochastic differential equations in a spatial model for precipitation.

The inclusion of covariates in the covariance function may be done implicitly. For instance, the process may be transformed using some Box-Cox transformation (De Oliveira et al., 1997) and the implied covariance function for the process will depend on the mean function. For some discussion about this topic see Wallin and Bolin (2015). However, for transformed fields the modeller has no control on the nonlinear relationship being created and this relationship between the covariance and mean structure might not capture well the actual structure of the process.

Different from previous approaches, our proposal models the mixing component of the process as a function of spatial covariates which results on an anisotropic covariance function, and marginal kurtorsis at each location that varies across apce. We propose a fully Bayesian model which incorporates the inherent characteristic of some spatial data analysis, that even after fitting a mean function depending on covariates, the same covariates might still help in the explanation of the variance. It is argued in this work that this residual heterogeneity could be well modeled through the inclusion of covariates in the scale mixing component of Palacios and Steel (2006) and Fonseca and Steel (2011). We now briefly review the scale mixing spatio-temporal model as proposed by Fonseca and Steel (2011).

1.2 Brief review of a spatio-temporal scale mixing models

Consider a spatio-temporal process defined by $\{Z(s,t) : s \in D; t \in T\}$, where (s,t) are spatio-temporal coordinates varying continuously in $D \times T$, $D \subseteq \mathbb{R}^d, T \subseteq \mathbb{R}$, $d = 1$, 2 or 3. In this context, Gaussianity is an usual assumption for the finite dimensional distribution of observations. Let $Z(s_i, t_j)$ be the observations of the process in locations s_i $(i = 1, \ldots, I)$ and times t_j $(j = 1, \ldots, J)$, thus under Gaussianity, $\mathbf{Z} = (Z(s_1,t_1), \ldots, Z(s_I,t_1), \ldots, Z(s_1,t_J), \ldots, Z(s_I,t_J))'$ follows a multivariate Gaussian distribution with covariance matrix Σ with elements $\Sigma_{kl} = Cov(Z_k, Z_l)$, $(k, l = 1, 2, \ldots, IJ)$ and mean vector $\mathbf{m} = (m(s_1,t_1), \ldots, m(s_I,t_J))'$.

Commonly, in spatial statistics, it is assumed that the process of interest follows a Gaussian process, usually after some suitable transformation. This implies that all finite dimensional distributions defining the spatio-temporal process are Gaussian. The class of Gaussian processes is mathematically convenient because it is defined only through its mean and covariance functions, all conditional and marginal distributions are known, and predictions are easily obtained through the properties of the multivariate normal distribution. However, this assumption might be very restrictive, as the resultant fit might be highly affected by aberrant observations, or regions with larger variability in space and/or time. In this context, distributions with heavier tails than the Gaussian distribution could provide better fit and possibly better predictions. Palacios and Steel (2006) define a spatial process through scale mixing which has heavier tails than the Gaussian process and is able to identify regions in space with larger variability. Fonseca and Steel (2011) extend this idea for the spatio-temporal setup and present an extra mixture to accommodate outliers both in space and time.

Consider a spatio-temporal process defined as a scale mixture as

$$Z(s_i, t_j) = \mathbf{w}(s_i, t_j)'\boldsymbol{\delta} + \sigma \frac{\epsilon(s_i, t_j)}{\sqrt{\lambda(s_i)}}, \tag{1}$$

where $\epsilon(s_i, t_j)$ is a Gaussian process in $(s_i, t_j) \in D \times T$, $i = 1, \ldots, I$, $j = 1, \ldots, J$. The process $\epsilon(\cdot, \cdot)$ has zero mean and a separable covariance function, such that, $C(d_s, d_t) = C_1(d_s)C_2(d_t)$, where $C_1(d_s)$ is a purely spatial correlation function depending on the Euclidean distance among locations, d_s, and $C_2(d_t)$ is a purely temporal correlation function depending on the temporal lag, d_t, $\sigma > 0$ is a scale parameter, $\mathbf{w}(s_i, t_j)$ is a vector of p covariates in location s_i and time t_j and $\boldsymbol{\delta}$ is a vector of regression coefficients. Thus, the resultant covariance matrix for the observed data is $\Sigma = \Sigma_2 \otimes \ \sigma^2 \left[\Lambda^{-1/2}\, \Sigma_1\, \Lambda^{-1/2}\right]$ with $\Sigma_{2,kl} = C_2(|t_k - t_l|)$, $k, l = 1, \ldots, J$, $\Sigma_{1,mn} = C_1(||s_m - s_n||)$, $m, n = 1, \ldots, I$ and $\Lambda = diag(\boldsymbol{\lambda})$, $\boldsymbol{\lambda} = (\lambda(s_1), \cdots, \lambda(s_I))'$.

The parameter $\lambda(\cdot)$ is a latent process, and is responsible for capturing the variance inflation in the process $\epsilon(\cdot, \cdot)$ across different locations, allowing for spatial heterogeneity. Conditional on the latent variable $\lambda(\cdot)$ the process $Z(\cdot, \cdot)$ is Gaussian. If the distribution of $Z(\cdot, \cdot)$ is integrated out with respect to $\lambda(\cdot)$ the resultant process is non-Gaussian.

Although $\lambda(s)$ does not need to be a process to capture dependence in space in $Z(s, t)$, it is required that

$$E[\lambda(s)^{-1/2}\lambda(s')^{-1/2}] \to E[\lambda(s)^{-1}] \ \text{ as } \ s \to s',$$

to obtain a process $Z(s, t)$ (unconditional on $\lambda(s)$) which is mean squared continuous, just as

$\epsilon(s,t)$. This result is proved in Palacios and Steel (2006). This is achieved either by defining $\lambda(s)$ as constant across space, or as a spatially structured process. Note that the former is not flexible enough to capture local heteroskedasticity.

Palacios and Steel (2006) and Fonseca and Steel (2011), model the latent variable $\lambda(s)$, $s \in D \subseteq \mathbb{R}^d$ as a stationary log Gaussian process such that $\ln(\lambda(\cdot))$ is a Gaussian process with mean function $-\nu/2$ and covariance function $\nu C_1(d_s)$. Thus, for locations $s_1, \ldots, s_I$, $\ln(\boldsymbol{\lambda}) \sim N_I\left(-\frac{\nu}{2}\mathbf{1}_I, \nu\Sigma_1\right)$. Note that in their proposal $\Sigma_{1,kl} = C_1(||s_k - s_l||)$, $k, l = 1, \ldots, I$ is assumed to be the same spatial correlation considered for the process $\epsilon(\cdot,\cdot)$ in (1). To make our results comparable to Fonseca and Steel (2011), and for parsimony, we choose the same covariance structure for the variance process $\lambda(s)$ and $\epsilon(s,t)$.

Note that the mean function for $\ln(\boldsymbol{\lambda})$ is constant across space and is given by $-\frac{\nu}{2}$. As a result, $E[\lambda(s)] = 1$ and $Var[\lambda(s)] = \exp(\nu) - 1$. Therefore, the parameter ν is responsible for the inflation in the variance of the process $Z(s,t)$, $t \in T \subseteq \mathbb{R}$. Small values of ν indicate that $\lambda(s)$ has a distribution concentrated around one, while larger values of ν indicate that $\lambda(s)$ tends to zero, inflating the variance of the process $Z(s,t)$. When ν tends to zero the resulting process tends to a Gaussian process.

We propose a model which allows $\ln(\boldsymbol{\lambda})$ to have prior mean depending on spatially varying covariates. As a result, the marginal distributions of $\lambda(s)$ vary with locations, allowing the kurtosis of the resultant spatio-temporal process, $Z(s,t)$, to also vary with location. In this work separability of the space-time covariance structure is assumed for convenience because the main focus of our paper is on the spatial domain, and for this reason non-separable models are not explored here. In our applications time is considered so that replicates are available at each location making it possible to identify different tail behaviours across space.

1.3 Motivation: the non-constant spatial variance problem

In this Section, we motivate the proposed model with an illustrative application to temperature data in which altitude is influential in spatial heterogeneity not only in the mean but also in the variability of the process. Consider a data set of maximum temperatures recorded daily in July 2006 at 70 locations within the Spanish Basque country.

The approach proposed in this paper is an extension of Palacios and Steel (2006) and

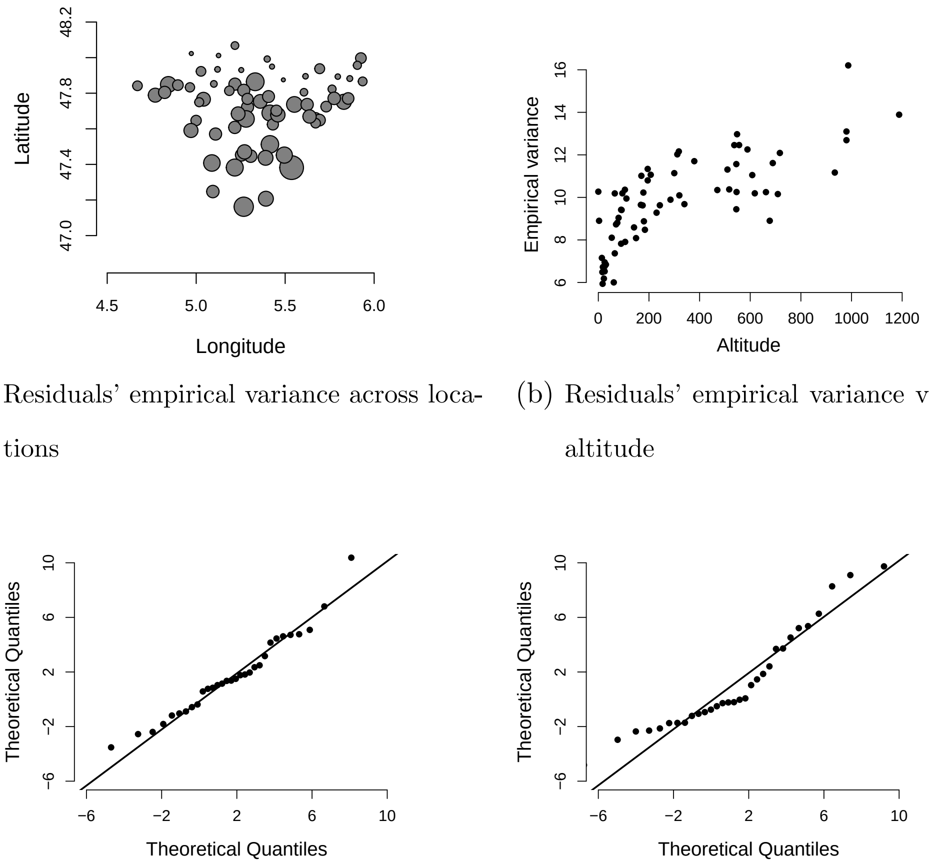
Fonseca and Steel (2011) which also analysed this temperature dataset. Palacios and Steel (2006) considered spatial data without replicates while Fonseca and Steel (2011) considered space-time data, however, none of them included information of altitude in the scaling mixture process. In that context, in order to allow for model comparison, and to illustrate the gains of our proposal, we have chosen to illustrate our model with the same dataset. In their analysis altitude is significant in the mean function. We consider a similar mean structure given by

$$\mathbf{w}(s,t)'\boldsymbol{\delta} = \delta_0 + latitude(s)\,\delta_1 + longitude(s)\,\delta_2 + altitude(s)\,\delta_3 + t\,\delta_4 + t^2\,\delta_5. \tag{2}$$

This is a mountainous region with altitudes varying from 0 to 1188 meters. Exploratory data analysis suggests that this variation in altitude affects not only the mean, but also the variance of the maximum temperatures. Altitude is usually included in the mean function of the process to capture the spatial variability present in the data, with smaller expected means present in higher altitudes. However, even after considering a mean function that depends on altitude, some extra variability might be noticeable in the residuals and the variance of these residuals might be well modeled by a positive process varying in space (Fonseca and Steel, 2011).

Initially we fit a Gaussian model, that is $\lambda(s) = 1$, $\forall s$, in equation (1), with the components of the mean given by (2). Panel (a) of Figure 1 shows the residual empirical variance observed at each location. Clearly, the estimated variance of the process shows a spatial pattern that seems to depend on altitude, even after considering altitude in the mean of the fitted model. Notice that the larger residual variances are observed mainly in the west and southwest portions of the region. Panel (b) of Figure 1 shows a scatter plot of altitude versus the estimated variances which suggests a non-linear relationship between the residual variance and the altitude, with larger residual variances being observed at higher altitudes. A model for the variance should accommodate this behaviour.Panels (c) and (d) of Figure 1 show the theoretical quantiles for the Gaussian distribution versus the empirical quantiles based on the residuals from the Gaussian fit for two different sites. The behaviour of the tails are different for the different sites. These two sites are located at quite different altitudes, site 11 is at sea level, while site 19 has altitude of 1188m.

These features often present in spatial data analysis motivate our proposal to include covariate information in the scaling mixture process proposed by Fonseca and Steel (2011). The proposed model attempts to capture non-stationarity features not only in the mean but also in the resultant covariance function.



(a) Residuals' empirical variance across locations

(b) Residuals' empirical variance vs. altitude

(c) Residuals' qqplot

(d) Residuals' qqplot

Figure 1: Residual analysis of the maximum temperature data. Panel (a): empirical variance across space (the diameter of the circle is proportional to the variance in each location). Panel (b): empirical variance across space versus altitude.

This paper is organized as follows. Section 2 proposes a model which allows the spatial process defined for $\lambda(s)$ in equation (1) to have a mean that depends on spatially varying covariates. Section 3 describes the inference procedure and discusses prior specification for the parameters in the model. Next Section describes an analysis of synthetic data, and Section 5 analyzes the maximum temperature observed in the Spanish Basque country introduced in Section 1.3. Finally, Section 6 presents some discussion and possible avenues for future research.

2 Scale mixture process depending on covariates

Following the spatio-temporal mixture model described in (1), it can be shown that the resultant kurtosis of the process is equal to 3 $\exp\{\nu\}$ (Fonseca and Steel, 2011), indicating that ν controls the tail behaviour of the process $Z(s,t)$. Our proposal focuses on modelling ν as a function of spatially varying covariates. More specifically, let $\ln(\lambda(s))$ be a Gaussian process with mean function $-\nu(s)/2$ and covariance function $Cov[\ln(\lambda(s_k)), \ln(\lambda(s_l))] = \nu(s_k)^{1/2}\nu(s_l)^{1/2}C_1(||s_k - s_l||), s_k, s_l \in D \subseteq \mathbb{R}^d$, such that for $\boldsymbol{\lambda} = (\lambda(s_1), \cdots, \lambda(s_I))'$ at observed locations $s_1, \ldots, s_I$, we have

$$\ln(\boldsymbol{\lambda}) \mid \boldsymbol{\nu}, \Sigma_1 \sim N_I\left(-\frac{1}{2}\,\boldsymbol{\nu}, diag(\sqrt{\boldsymbol{\nu}})\;\Sigma_1\; diag(\sqrt{\boldsymbol{\nu}})\right), \tag{3}$$

where $\Sigma_{1,kl} = C_1(||s_k - s_l||), k, l = 1, \ldots, I$ and $\boldsymbol{\nu} = (\nu(s_1), \ldots, \nu(s_I))'$.

As the parameter $\nu(s)$ has to be positive, we propose to model it as a linear function of spatially varying covariates in the logarithm scale, that is,

$$\ln \nu(s) = \beta_0 + \beta_1 x_1(s) + \beta_2 x_2(s) + \cdots + \beta_{q-1} x_{q-1}(s), \tag{4}$$

$\mathbf{x}(s) = (x_1(s), x_2(s), \ldots, x_{q-1}(s))'$ is the vector containing the covariates that are believed to influence the variance of the process $Z(s,t)$, and $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_{q-1})'$ is the vector of the associated regression coefficients.

The mean and variance of $\lambda(s)$, conditioned on $\nu(s)$ are, respectively, given by $E[\lambda(s)|\nu(s)] = 1$, and $Var[\lambda(s)|\nu(s)] = \exp\{\nu(s)\} - 1$. Notice further that the variance of $\lambda(s)$ varies with s and, therefore, the variance of the resulting process $Z(s,t)$ also varies with spatial locations. Clearly, our proposed model is able to accommodate spatial heterogeneity present in the process $Z(s,t)$ when the heterogeneity is caused by some spatial effect either implicitly or explicitly. It is worth looking at the behaviour of $\lambda(s)$ for different values of $\nu(s)$. If $\nu(s) \to 0$, the variance of $\lambda(s)$ tends to zero and, the process $Z(s,t)$ tends to the usual

Gaussian process. When $\nu(s)$ increases, the marginal distribution of $\lambda(s)$ becomes flatter, inflating the variance of $Z(s,t)$ at location s, and naturally accommodating aberrant or atypical observations at location s. If the covariates $\mathbf{x}(s)$ do not influence $\nu(s)$ then our proposed model has the same structure as that of Fonseca and Steel (2011). Next, we obtain the resultant covariance and kurtosis of the proposed model.

2.1 Properties of the proposed model

Proposition 2.1 *When integrating the distribution of $Z(s,t)$ with respect to $\lambda(s)$, the resultant covariance function of the process $\{Z(s,t) : s \in D; t \in T\}$ defined in (1), with the mixing latent process as defined in (3), is given by:*

$$Cov[Z(s_1,t_1), Z(s_2,t_2)] = \sigma^2 C_1(d_s) C_2(d_t) \exp\left\{\frac{3}{8}\left[\nu(s_1)+\nu(s_2)\right] + \frac{\sqrt{\nu(s_1)\nu(s_2)}}{4} C_1(d_s)\right\}, \quad (5)$$

$(s_1,t_1),(s_2,t_2) \in D \times T$, $d_s = ||s_1 - s_2||$, $d_t = |t_1 - t_2|$.

Proof: See Appendix A.

Clearly, the resultant marginal covariance function in (5) depends on the values of the covariates in $\nu(s)$. From equation (5), it follows that the marginal variance of the process is given by $Var[Z(s,t)] = \sigma^2 \exp\{\nu(s)\}$, and the marginal correlation function by $Cor[Z(s_1,t_1), Z(s_2,t_2)] = C_1(d_s) C_2(d_t) \exp\left\{-\left[\nu(s_1)+\nu(s_2)\right]/8 + \sqrt{\nu(s_1)\nu(s_2)}/4 \; C_1(d_s)\right\}$. Therefore, the proposed model is nonstationary in space, being able to accommodate spatio-temporal processes whose variances change with spatial location.

Proposition 2.2 *The marginal kurtosis (with respect to $\lambda(s)$) in each location for the process $\{Z(s,t) : s \in D; t \in T\}$ defined in (1), with mixing latent process as defined in (3) is given by:*

$$K_Z(\mathbf{x}(s), \boldsymbol{\beta}) = 3\exp\{\exp\{\beta_0 + \beta_1 x_1(s) + \ldots + \beta_{q-1} x_{q-1}(s)\}\}, \quad (6)$$

$(s,t) \in D \times T$, $\mathbf{x}(s) = (x_1(s), x_2(s), \ldots, x_{q-1}(s))'$ and $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_{q-1})'$.

Proof: See Appendix A.

The resulting kurtosis changes with spatial location. An important issue is how the covariates and coefficients in equation (6) influence the kurtosis of $Z(s,t)$. Initially, assume there are no covariates and we model $\nu(s) = \exp(\beta_0)$. Table 1 shows the marginal kurtosis in equation (6) under this scenario. We also present the corresponding degrees of freedom of a standardized Student-t model with ν_0 degrees of freedom. This is to compare the kurtosis of the scale mixture model with a well known model, which allows for fatter tails through an easily interpretable parameter ν_0. When β_0 is positive the kurtosis gets very large, with values that do not represent realistic spatial processes. For this reason, we suggest that the prior specification for β_0 is constrained to be negative.

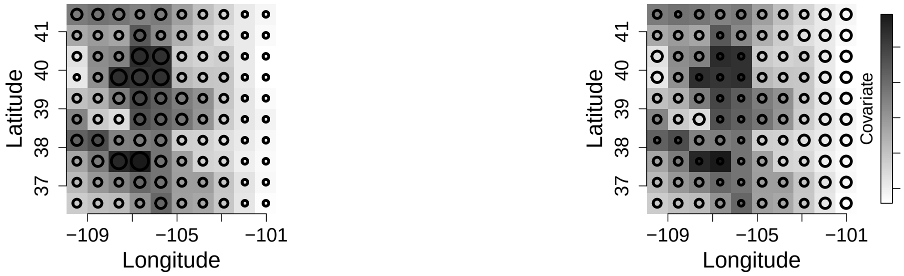
Table 1: Kurtosis for different values of ν_0 (Student-t model) and different values of β_0 assuming no effect of covariates $\beta_1 = 0$.

Kurtosis	200	100	50	20	10	8	5	4	3
ν_0	4.03	4.06	4.12	4.35	4.86	5.2	7	10	200
β_0	1.43	1.25	1.03	0.64	0.18	-0.02	-0.67	-1.24	-16

As an illustration of the covariate effect in the kurtosis we consider an artificial field with 100 observed locations, which is shown in panels of Figure 2. In the panels we explore positive and negative values of β_1. In each panel, the gray scale represents the values of the covariate, whereas the diameter of the open circles at each location is proportional to the resultant marginal kurtosis. For positive β_1 (Panel (a)), larger values of the kurtosis (circles) occur for larger values of the covariate (darker gray squares); whereas for a negative value of β_1 (Panel (b)), larger values of the kurtosis (circles) occur for smaller values of the covariate (lighter gray squares).

As outlined by Viana Neto et al. (2014), introducing covariates in the covariance structure of spatial processes seems to provide reasonably flexible models. However, care must be taken when including covariate information in the covariance structure of spatial processes. It is important to understand well the process under study such that the inclusion of covariates in $\lambda(s)$ are helpful to better explain the second order properties of the process.

In this context, some exploratory data analysis with residuals from fitting a usual Gaus-



(a) $\beta_0 = -0.2$ and $\beta_1 = 0.8$

(b) $\beta_0 = -0.2$ and $\beta_1 = -0.8$

Figure 2: Illustrative example: artificial values of a covariate $x_1(s)$, and resultant marginal kurtosis ($3\exp\{\nu(s)\}$, with $\nu(s) = \exp\{\beta_0 + \beta_1 x_1(s)\}$) under two scenarios: (a) $\beta_0 = -0.2$ and $\beta_1 = 0.8$, and (b) $\beta_0 = -0.2$ and $\beta_1 = -0.8$. The gray scale depicts the values of the covariate across the lattice, and circles are proportional to the marginal kurtosis.

sian process would indicate possible relations with covariates which might be already in the mean of the fitted model but could also help in the explanation of the second order spatial structure. Histograms from residuals at each location could indicate different tail behavior across space and plot of empirical variances versus covariate could indicate a relationship between variance and spatially varying covariates. For instance, this was done with our temperature data example and figure 1 (b) illustrates this relationship between variance and covariate.

3 Prior specification and inference procedure

The inferential approach adopted here follows the Bayesian paradigm; thus, all inference, predictions and model comparison are obtained from the posterior distribution of the parameters of interest. And the posterior distribution results from updating the prior information with the data information that comes from the likelihood function. The model specification follows from equations (1), (3) and (4) and, from the Bayesian point of view, is complete after assigning the prior distribution for the parameter vector of the model.

For the correlation functions we assume a Cauchy correlation function. This function allows for smoother processes than induced by the exponential function and adds some flexibility by allowing for the modeling of long-memory dependence and also correlations at short and intermediate lags. Different from the usual Matérn correlation function, the Cauchy function does not require estimation of a smoothness parameter which might be difficult to estimate (Zhang, 2004). Gneiting (2000) and Gneiting and Schlather (2004) give more details on the properties and power-law behaviours generated by this class of correlation functions. The covariance function in each dimension is given by $C_i(d) = \left(1 + \left(\frac{||d||}{a_i}\right)^{\alpha_i}\right)^{-1}$, $i = 1, 2$. The parameters a_1 and a_2 are, respectively, the range parameters and α_1 and α_2 are shape parameters. The parameter σ^2 in equation (1) is a scale parameter. Therefore, the parameter vector is given by $\mathbf{\Psi} = (a_1, \alpha_1, a_2, \alpha_2, \sigma^2, \boldsymbol{\beta}, \boldsymbol{\delta})$. Following Bayes' theorem, the posterior distribution is proportional to

$$p(\boldsymbol{\lambda}, \mathbf{\Psi}|\mathbf{Z}) \;\; \propto \;\; |\Sigma|^{-1/2} \exp\left\{-\frac{1}{2}(\mathbf{Z} - \mathbf{w}\boldsymbol{\delta})' \, \Sigma^{-1} \; (\mathbf{Z} - \mathbf{w}\boldsymbol{\delta})\right\} \; p(\boldsymbol{\lambda}|a_1, \alpha_1, \boldsymbol{\beta}) \; p(\mathbf{\Psi}),$$

where $\mathbf{Z} = (Z(s_1, t_1), \cdots, Z(s_I, t_1), \cdots, Z(s_1, t_J), \cdots, Z(s_I, t_J))'$ is the IJ-dimensional vector of the observations; $\mathbf{w}$ is the design matrix containing the covariates for the mean process, and Σ, the covariance matrix, is given by $\Sigma = \Sigma_2 \otimes \; \sigma^2 \left[\Lambda^{-1/2} \, \Sigma_1 \, \Lambda^{-1/2}\right]$, where $\Sigma_{2,kl} = C_2(|t_k - t_l|)$, $k, l = 1, 2, \ldots, J$, $\Sigma_{1,ij} = C_1(||s_i - s_j||)$, $i, j = 1, 2, \ldots, I$, $\Lambda = diag(\lambda(s_1), \ldots, \lambda(s_I))$ and $\otimes$ denotes the Kronecker product. *A priori*, the parameters are assumed independent. The covariate coefficients ($\boldsymbol{\delta}$) in the mean process are assumed to follow independent, zero mean normal prior distributions with some large, fixed, variance which result in a vague prior information for $\boldsymbol{\delta}$. For the range parameters we consider $a_i \sim gamma(a_{a_i}, b_{a_i})$, whereas for the smoothness parameters we assume an uniform prior such that $\alpha_i \sim Unif[0, 2]$, $i = 1, 2$. Finally, for the scale parameter σ^2 we assign an inverse gamma prior such that $\sigma^2 \sim IG(a_{\sigma^2}, b_{\sigma^2})$. Care must be taken when specifying the prior distribution for the coefficients in $\nu(s)$. This is because the prior range of $\boldsymbol{\beta}$ affects the range of the kurtosis of the process. Thus, in order to accommodate realistic values for the kurtosis, we propose two different prior specifications for $\boldsymbol{\beta}$. We discuss these proposals in Subsection 3.1.

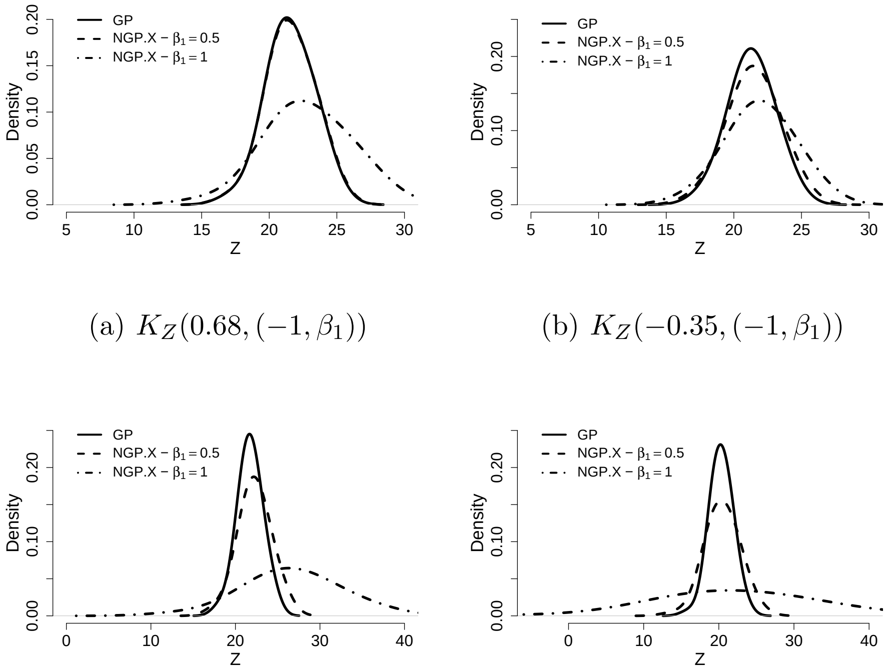
Regardless of the prior specification for $\boldsymbol{\beta}$, the resulting posterior distribution for $(\boldsymbol{\lambda}, \mathbf{\Psi})$ does not have a closed form and inference is based on Markov Chain Monte Carlo (MCMC) methods. The full posterior conditionals are shown in Appendix B. All the algorithms in this paper were implemented using the software `R` (R Core Team, 2015) and are available upon request.

3.1 Prior distribution for $\boldsymbol{\beta}$

Before discussing the prior specification for $\boldsymbol{\beta}$ we suggest to standardize the covariates considered in the model for $\nu(s)$. Then, the kurtosis will be less affected by changes in the scale of the covariates. We suggest the use of medians ($med(x)$) and interquartile distances ($IQ(x)$) in order to avoid high influence of extreme values of the covariate in the standardization. Let $x_j^*(s) = (x_j(s) - med(x_j))/IQ(x_j), j = 1, \ldots, q-1$ be the standardized covariates. The prior specification for the coefficients $\boldsymbol{\beta}$ are assigned after this standardization.

Table 1 suggests we should consider negative values of β_0 in order to ensure realistic values of the marginal kurtosis of the process. Furthermore, the values of $|\beta_j|$ should not be very large. For this reason, the prior specification for $\boldsymbol{\beta}$ is built in terms of the resultant variation of the kurtosis of $Z(s,t)$. Without loss of generality, in what follows we discuss the prior distribution for the case of one covariate in $\nu(s)$, that is, we focus on $\nu(s) = \exp\{\beta_0 + \beta_1 x_1(s)\}$.

To depict the effect of modelling $\nu(s)$ as a function of covariates in equation (3) we have performed simulated studies (not detailed here) to understand the different posterior inference obtained for different specifications of the maximum kurtosis allowed, *a priori*. We have found that if the maximum kurtosis is allowed to be too big then the fit is too sensitive to outliers in the data and the posterior predictive distributions tend to overfit to adapt to very large observed values. In this context, as follows we present realizations from our proposed model for some maximum kurtosis which give an indication of how the prior could be used to truncate the maximum kurtosis and still reflect most of the realistic behaviour one expects to find in real applications. Panels of Figure 3 show the distribution of partial realizations of different GP with mean equals 20, and some covariance structure, based on the same locations of the real data in Section 1.3. We compare the realizations of a GP with those of NGP.X under different values of $x(s)$ and β_1. It is clear that the higher the value of β_1 and $x(s)$ the heavier the tails of the resultant marginal distribution at a particular location. Also, the values of β_1 should not be too high, as depending on the values of the covariate and β_1 we can obtain densities that are quite heavy tailed (see panel (d)). We aim at proposing a prior distribution for β_0 and β_1 that provide realistic realizations of the process under study. Therefore, some kind of prior constrain should be assumed for β_0 and β_1.



(a) $K_Z(0.68, (-1, \beta_1))$

(b) $K_Z(-0.35, (-1, \beta_1))$

(c) $K_Z(1.68, (-1, \beta_1))$

(d) $K_Z(2.12, (-1, \beta_1))$

Figure 3: Partial realizations of the process for four different locations and a fixed point in time, considering $\log \nu(s) = \beta_0 + \beta_1 x(s)$, with different values of $x(s)$, and β_1.

Independent prior for β_0 and β_1 In this specification we assume prior independence between β_0 and β_1. In particular, for β_0 we assign a truncated normal distribution defined on $\mathbb{R}^-$, whose associated normal has mean a and variance b, that is $\beta_0 \sim TN_-(a, b)$. After performing a sensitivity study to investigate the values of a and b, we suggest $a = 0$ and $b = 3$ as benchmark values for the prior specification of β_0. This distribution is concentrated around -2.4, leading the prior kurtosis to be concentrated around 3, which is the kurtosis of a Gaussian process. For β_1 we assign an uniform prior over the interval (c, d). The choice of c and d affects the tails of the kurtosis distribution. Thus, to avoid values of the kurtosis that are too large we assign $c = -1$ and $d = 1$ in this work. We denote this model as NGP.X (I).

Conditional prior specification for β_0 and β_1 In order to avoid unrealistic high values for the prior distribution of the kurtosis we also propose a conditional prior specification for (β_0, β_1), such that $\pi(\beta_0, \beta_1) = \pi(\beta_1|\beta_0)\pi(\beta_0)$. Following the values in table 1, we again constrain β_0 to be negative but we now assign an uniform prior distribution over the interval $(a_0, 0)$, such that $\beta_0 \sim U(a_0, 0)$, and $a_0 < 0$. The value of $a_0 = -4$ is chosen from table 1 to allow for values of the kurtosis that are close to the Gaussian case. Next, we define an upper limit, l_{kurt}, for the kurtosis of the process such that $3 \leq K_Z(\cdot, \cdot) \leq l_{kurt}$. Then we specify a prior distribution for β_1 conditioned on the value of β_0. In particular, we propose an uniform prior such that $\beta_1|\beta_0 \sim U(-L + \beta_0, L - \beta_0)$ with $L = \ln(\ln(l_{kurt}/3))$, a function of l_{kurt}. We denote this model as NGP.X (D).

Section C of the Appendix describes how to perform spatial interpolation and temporal prediction under the proposed model.

4 Analysis of a synthetic dataset

In practice, when analyzing spatio-temporal data, it is common to use some nonlinear transformation, e.g. the square root or the log transformations, to attain approximate normality of the data. Wallin and Bolin (2015) call attention to the fact that in the original scale, the resultant process is nonstationary if the mean has spatially varying covariates, and there is a relationship between the mean and the covariance functions, possibly turning interpretation of the parameters more challenging when nonlinear mean or nonstationary covariance structures are assumed for the transformed field.

Here we focus on the case of a log-Gaussian process. Assume $\ln\, Z(s,t) = \mathbf{w}(s)'\boldsymbol{\delta} + \epsilon(s,t)$, where $\mathbf{w}(s)$ is a vector of covariates that vary smoothly across space, and $\epsilon(s,t)$ is a zero mean Gaussian process with covariance function $C(d_s, d_t) = \sigma^2\rho(d_s, d_t)$, where d_s, d_t represent, respectively, spatial and temporal Euclidian distances, and $\rho(\cdot, \cdot)$ is a valid correlation function. In the original scale, the mean of the process is given by $E[Z(s,t)] = \exp\{\mathbf{w}(s)'\boldsymbol{\delta} + 0.5\sigma^2\}$

and the resultant covariance between any two spatiotemporal coordinates (s,t) and (s',t') is

$$Cov[Z(s,t),Z(s',t')] = \left[\exp\{(\mathbf{w}(s)'\boldsymbol{\delta} + \mathbf{w}(s')'\boldsymbol{\delta}) + \sigma^2\}\right]\left[\exp\{C(d_s,d_t)\} - 1\right],$$

with $d_s = ||s - s'||$, $d_t = |t - t'|$, resulting in a nonstationary covariance structure for $Z(s,t)$, as we assume that $\mathbf{w}(s)$ contains spatially varying covariates.

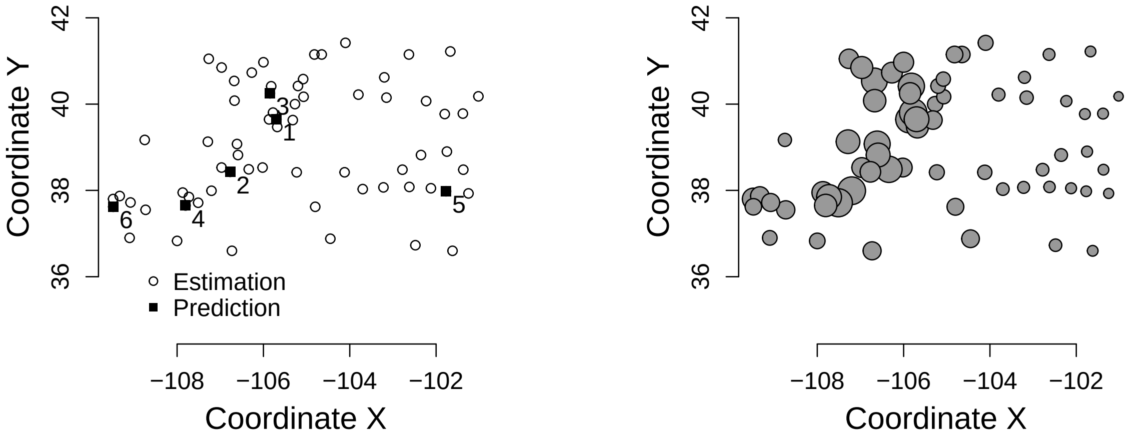
Generation of artificial data We generate data from a log-Gaussian spatiotemporal process as follows. We start by defining $I = 66$ locations over a region, and $J = 30$ instants in time. To depict a realistic region we consider a portion of the Colorado state in the USA, as altitude shows a strong spatial pattern in the east-west direction. Therefore we consider altitude in the modelling of $\nu(s)$. Panel (a) of Figure 4 shows the spatial coordinates together with $p = 6$ locations that are left out from the inference procedure for predictive purposes. Panel (b) of Figure 4 shows the behaviour of the covariate across the region. The diameter of the circles are proportional to the value of the covariate at the respective location. The mean function is given by $\mathbf{w}(s)'\boldsymbol{\delta} = \delta_0 + \delta_1\ w(s)$, with $\delta_0 = 0.33$ and $\delta_1 = 1$.

The covariance structure is assumed to be stationary and separable in the log scale, with $C(d_s,d_t) = \sigma^2 C_1(d_s)C_2(d_t)$ where $C_1(d_s)$ is the spatial correlation structure and $C_2(d_t)$ the temporal one, with d_s representing the Euclidean distance in space and d_t representing the difference in time. We assumed a Cauchy correlation function with smoothness parameters fixed at $\alpha_1 = \alpha_2 = 1.5$, and decay parameters equal to $a_1 = 2.57$ and $a_2 = 1$. The scale parameter was fixed at $\sigma^2 = 0.8$. These values were fixed to provide reasonable values of the artificial data. As our proposed model accounts for covariate effect in the covariance structure of the process, we expect it to better capture the structure in the data when compared to the NGP.

We fit the following three models, assuming the same mean structure that was used to generate the dataset,

NGP.X (I) $\ln\nu(s) = \beta_0 + \beta_1\ x(s)$, where $x(s)$ is the artificial covariate, with the following prior specification $\beta_0 \sim NT_{-}(0,3)$ and $\beta_1 \sim U(-1.5,1.5)$;

NGP.X (D) $\ln\nu(s) = \beta_0 + \beta_1\ x(s)$ with $x(s)$ as above, and the following prior specification $\beta_0 \sim U(-4,0)$ e $\beta_1|\beta_0 \sim U(-L+\beta_0, L-\beta_0)$ with $L = \ln(\ln(30/3))$;



(a) Monitoring locations

(b) Covariate

Figure 4: (a) Monitoring locations (gray circles) for the artificial data generated under a log-Gaussian process. The numbered locations (black squares) are the ones left out from the inference procedure to check the predictive ability of the different fitted models. (b) Monitoring locations with diameter of the circles proportional to the covariate value, $x_1(s)$, at each location.

NGP $\nu(s) = \nu$ for all $s \in D$.

We assume prior independence among the other parameters of the model, and assign the following prior distributions: $a_1 \sim gamma(0.01, 0.01/m)$, m is the median of the Euclidean distance among locations; $a_2 \sim gamma(0.01, 0.01)$; $\alpha_i \sim U(0, 2)$, $i = 1, 2$; $\sigma^2 \sim IG(2.1, 1)$ and $\boldsymbol{\delta} \sim N_2(0, diag(100))$.

We run two chains starting from very different initial values and let the MCMC algorithm run for 50,000 iterations, used 20,000 as burn in and kept every other 30th iteration to avoid autocorrelation among the sampled values. Convergence of the chains was checked using the $\hat{R}$ test of Gelman and Rubin (1992).

Panels of Figure 5 show the posterior distributions of β_0 and β_1 under models NGP.X (I) and NGP.X (D), together with their respective prior distributions. There is clear gain of information when we compare the posterior of β_0 and β_1 with their respective prior distributions. This suggests that there is information in the data to learn about these parameters. Also, the different prior specifications for β_0 and β_1 do not lead to very different posterior distributions for each of the parameters. The coefficient β_1 is estimated as strictly positive suggesting that, as covariate increases, so does the kurtosis of the process.

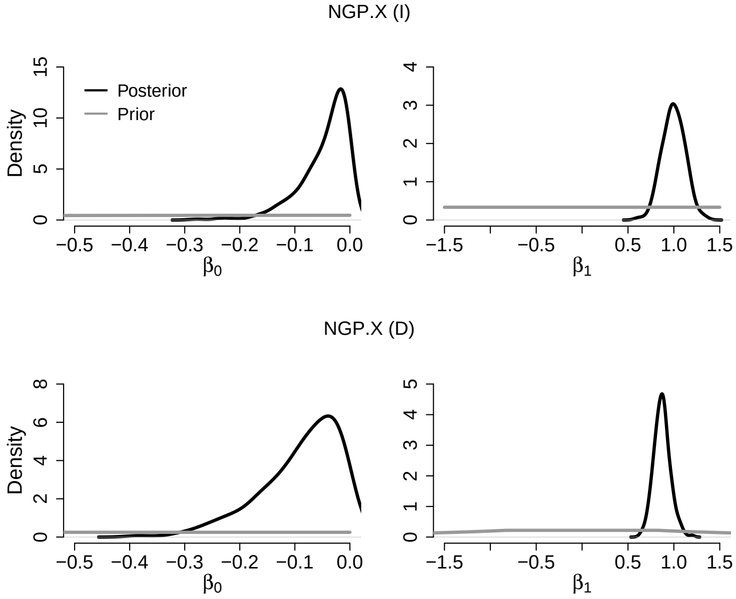


Figure 5: Prior (grey lines) and posterior (black lines) distributions of β_0 (first column) and β_1 (second column) under models NGP.X (I) (top row) and NGP.X (D) (bottom row) for the artificial dataset.

Panels of Figure 6 show the summary of the posterior predictive distribution obtained for locations 4 (first row) and 5 (second row) that were left out from the inference procedure for predictive purposes (see Figure 4). In general, model NGP provides the widest ranges of the 95% credible intervals. Moreover, for locations 4 and 5 model NGP yielded point estimates quite far from the actual observations.

In the interest of model comparison we compute the predictive performance using the interval score criterion (IS)(Gneiting et al., 2007), which compares the predicted value with the true one, and considers the uncertainty in the predictions such that the model is penalized if an interval is too narrow and misses the true value. We also compute the log predictive score (LPS) (Gneiting et al., 2007) which is based on the logarithm of the predictive distribution.

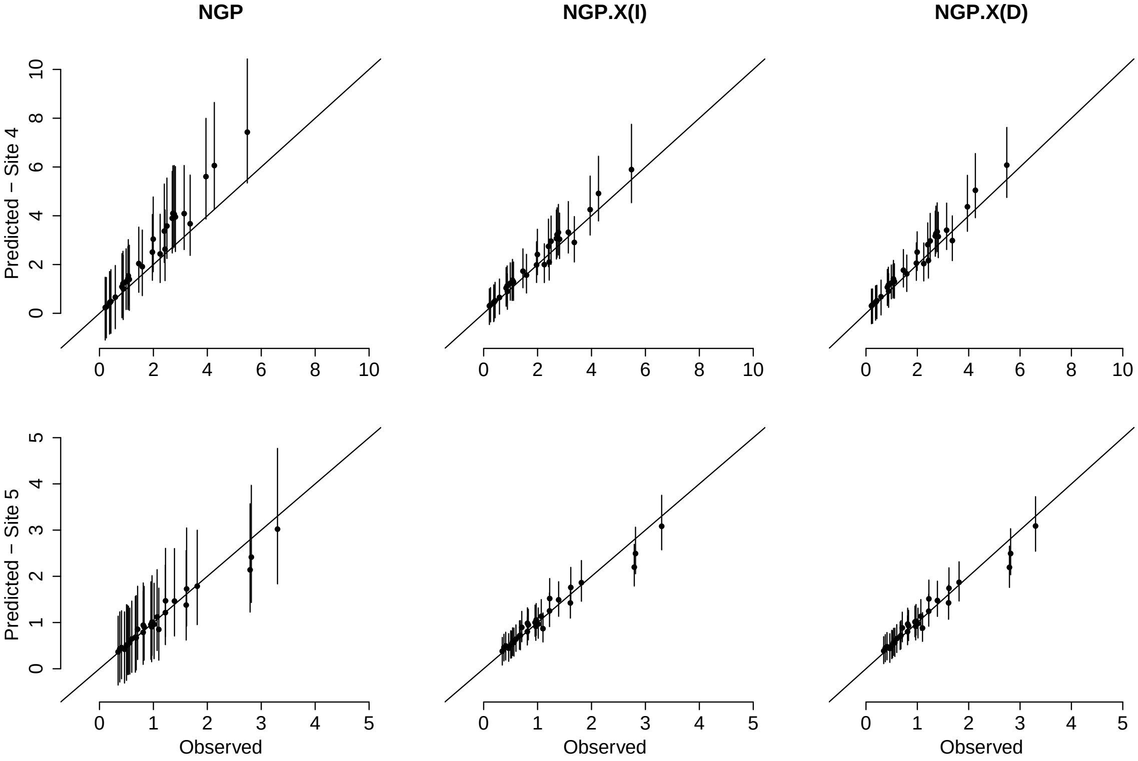


Figure 6: Posterior predictive distribution (median and 95% credible intervals) versus the observed values for locations 4 (first row) and 5 (second row) that were left out from the inference procedure, under the different fitted models (columns) for the artificial dataset.

These comparison criteria are described in Section D of the Appendix. In this example in particular, our aim is to check which model among NGP.X (I), NGP.X (D), and NGP leads to values of IS and LPS that are smaller and closer to the one obtained under the log-Gaussian model, which is the model used to generate the data. Table 2 summarizes the results. The numbers therein should be compared to the results for IS and LPS obtained under the log-Gaussian model which are, respectively 1.18, and -245. Clearly, model NGP.X (D) is the one that leads to closer values of the criteria obtained under the model used to generate the data.

Table 2: Interval score (IS) and log predictive score (LPS) under the predicted observations at the out-of-sample sites under models NGP, NGP.X (I), and NGP.X (D). The log-Gaussian model used to generate the data results in an IS equal to 1.18 and a LPS equal to -245.

Model	NGP	NGP.X (I)	NGP.X (D)
IS	2.33	1.44	1.32
LPS	16.79	-101.87	-97.89

5 Analysis of maximum temperature at the Spanish Basque Country

Now, we fit our proposed models to the temperature data presented in Section 1.3. This data was also analysed in Fonseca and Steel (2011). The maximum temperature data was observed in 70 locations in the Spanish Basque Country in July 2006 with 67 locations used for estimation and 3 left out of the analysis for predictive performance assessment. The coordinates of the locations are considered in utm such that distances in space are in kilometers.

For the temperature data, altitude seems to influence the variability of the process as shown in Figure 1 and will be considered in the mean of the latent process $\ln \lambda(s)$. We compare five model specifications:

GP Gaussian with $\lambda(s) = 1$, for all s.

GP.Var.X Gaussian with $\sigma^2(s) = \exp(\beta_0 + \beta_1 x(s))$, where $x(s)$ is the standardized altitude, with the following prior specification $\beta_0 \sim N(0, 10)$ and $\beta_1 \sim N(0, 10)$;

NGP $\nu(s) = \nu$ for all $s \in D$;

NGP.X (I) $\ln \nu(s) = \beta_0 + \beta_1 \; x(s)$, where $x(s)$ is the standardized altitude, with the following prior specification $\beta_0 \sim NT_{-}(0, 3)$ and $\beta_1 \sim U(-1, 1)$;

NGP.X (D) $\ln \nu(s) = \beta_0 + \beta_1\, x(s)$ with $x(s)$ as above, and the following prior specification $\beta_0 \sim U(-4, 0)$ e $\beta_1 | \beta_0 \sim U(-L + \beta_0, L - \beta_0)$ with $L = \ln(\ln(30/3))$;

Note that the fitted models grow in complexity with the covariance structure. The first model, GP, is the one fitted in Section 1. Model GP.Var.X allows the variance of the Gaussian process to change with location in a deterministic fashion, as a log-linear function of the standardized altitude of the location. Model NGP assumes a stochastic process for $\lambda(s)$ as in Fonseca and Steel (2011), allowing for heavier tails at different locations, whereas models NGP.X(I) and NGP.X(D) are the ones proposed here. We allow for the prior mean of $\lambda(s)$ to be a function of altitude. Therefore, this results in a more flexible structure than the deterministic structure proposed by model GP.Var.X. The mean function depends on spatiotemporal covariates and is detailed in Section 1.3.

The prior distributions considered were $a_1 \sim gamma(0.01, 0.01/m)$, with m equals the median distance among the observed locations; $a_2 \sim gamma(0.01, 0.01)$; $\alpha_i \sim U(0, 2)$, $i = 1, 2$; $\sigma^2 \sim IG(2.1, 1)$ and $\boldsymbol{\delta} \sim N_6(0, diag(100))$. For the NGP.X (I) model, the prior distributions are $\beta_0 \sim NT_{-}(0, 3)$ and $\beta_1 \sim U(-1, 1)$. We run two chains starting from different initial values and let the MCMC algorithm run for 60,000 iterations, used 20,000 as burn in and kept every other 40th iteration to avoid autocorrelation among the sampled values. Convergence of the chains was checked using R test of Gelman and Rubin (1992).

Figure 7 shows the posterior marginal densities for the mean parameters δ_0, δ_1, δ_2, δ_3, δ_4, δ_5. The three models give very similar posterior distributions for these parameters. The latitude and altitude are significant in all models. The estimated values of δ_1 indicate that latitude positively influences the mean, that is, the higher the latitude the higher the temperature. The altitude coefficient negatively influences the mean indicating that the higher the altitude the lower the temperature. On the other hand, the estimated association between altitude and the mean of $\ln \lambda(s)$ is positive (See Figure 8 for details). That is, the higher the altitude the higher the variability of the process. Note that the posterior uncertainty of the coefficients under the mixture models is much smaller than in the Gaussian case.

Figure 8 shows the prior and posterior distributions for the coefficients in the mean of $ln\, \lambda(s)$, β_0 (first column) and β_1 (second column) under models NGP.X (I) (top row) and

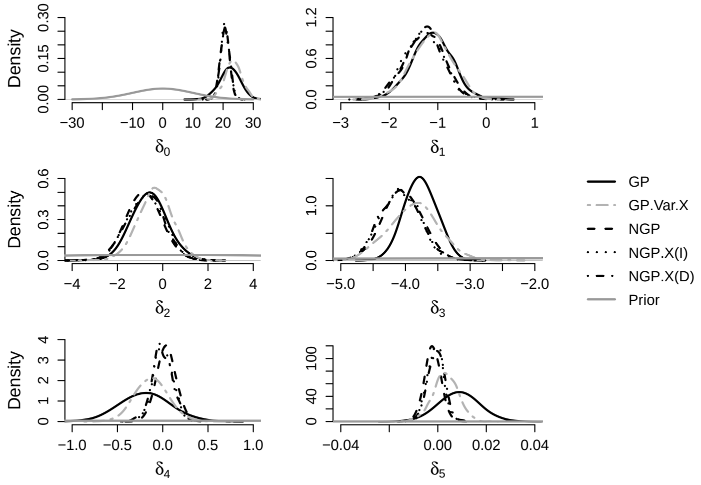


Figure 7: Prior (grey lines) and posterior distributions for the parameters (δ_0, δ_1, δ_2, δ_3, δ_4 and δ_5) in the mean function of the maximum temperature dataset, under each of the five fitted models.

NGP.X (D) (bottom row). These parameters influence directly the marginal kurtosis for each location. The parameter β_1 was significantly different from 0 in both models indicating that higher altitudes lead to larger variability. In other words, the marginal distribution of the process has fatter tails at sites located in high altitudes.

We now compare the predictive performance of the five competing models. Data for three locations were left out of the estimation procedure and predictions were obtained based on the predictive distribution as described in Section C of the Appendix. Figure 9 shows the 95% credible intervals for out-of-sample observations. As previously mentioned, our goal is to improve predictions with our proposed model by better modeling the uncertainty in the variance process. We notice that the credible intervals are narrower under our proposed model for locations 1 and 3, while it provides wider ranges of the predictive credible interval for location 2. Notice that for location 2, the other models are not able to accommodate the larger uncertainty for some extreme observations which presented larger temperature values. Furthermore, the two prior distributions considered under our proposed models led to similar predictive distributions.

Like in the previous section we use the Interval Score (IS) and the Log Predictive Score (LPS) (Gneiting et al., 2007) to compare the different fitted models. The values of the

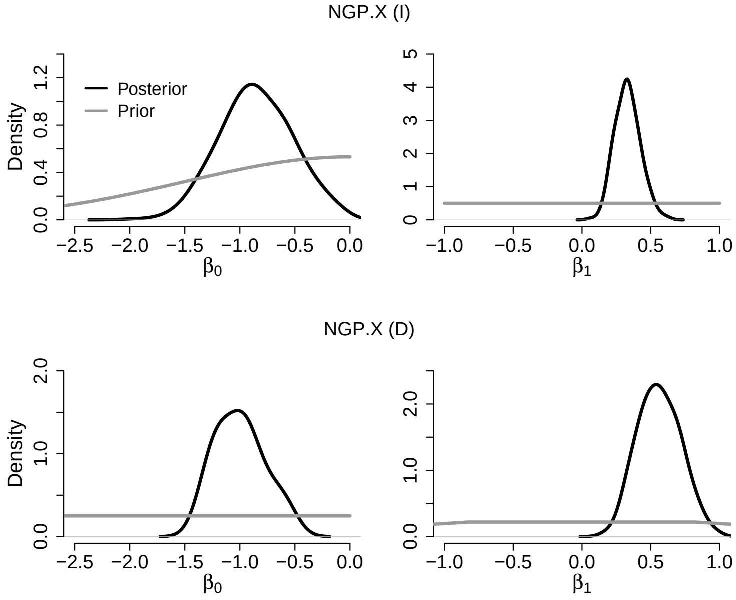


Figure 8: Prior (grey lines) and posterior (black lines) distributions of β_0 (first column) and β_1 (second column) under models NGP.X (I) (top row) and NGP.X (D) (bottom row) for the maximum temperature dataset.

criteria are shown in Table 3. Under criterion IS model NGP.X (I) performs best, among the fitted ones. Under LPS the best model is NGP.X (I) followed by NGP.X(D).

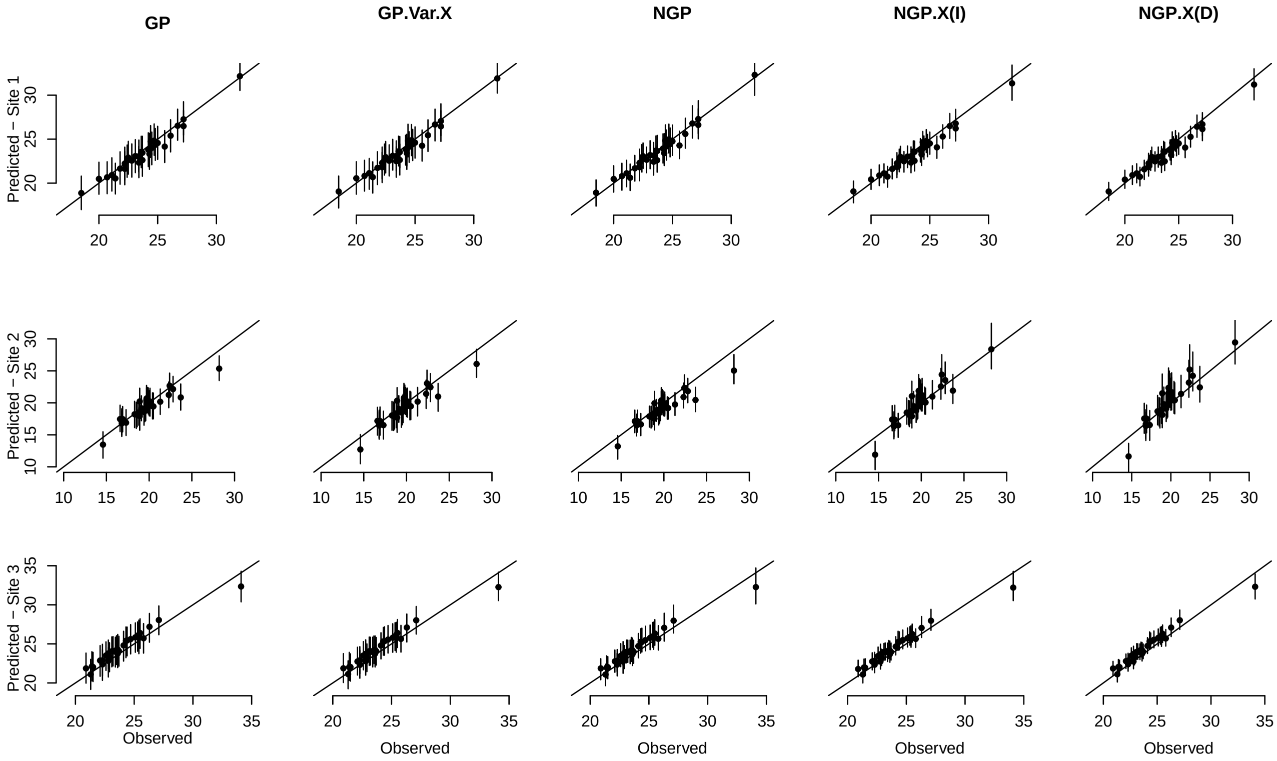


Figure 9: Posterior summary (median and 95% credible intervals) of the predictive distribution versus the actual observed values of the maximum temperature observed at three different locations left out from the inference procedure.

Table 3: Model comparison based on IS and LPS criteria for the predicted observations at the out-of-sample sites under each of the fitted models for the maximum temperature dataset.

Model	GP	GP.Var.X	NGP	NGP.X (I)	NGP.X (D)
IS	4.51	4.12	4.12	3.59	4.36
LPS	89.90	87.08	61.24	38.31	59.79

6 Discussion

We introduce a new class of non-stationary spatio-temporal geostatistical models by allowing spatially varying covariates that influence the tail behaviour of the process across space. Understanding the tail behaviour of spatio-temporal processes is crucial for efficient prediction. Unexplained large variances may result in large prediction intervals while well modelled variances will tend to capture the correct amount of uncertainty in the predictive distribution of the process.

One important aspect of our proposed framework is that the variance process is allowed to depend on covariates, providing some interpretation about the behaviour of the tail of the process as a function of a known covariate. For instance, in our synthetic application to the log-transformed field, it is known that a relationship between the mean and the covariance is imposed through the log transformation of a Gaussian field. However, our model is able to identify the dependence on the covariate without requiring any transformation of the data. Of course the unknown transformation could be estimated from the data, however, the induced mean-covariance relationship would be nonlinear. The effect of considering the response in the original scale and estimating the covariance-covariate dependence with our model led to narrower predictive intervals when compared to the ones obtained by a model which does not allow for covariance-covariate dependence.

In order to allow for flexible representation of the scale mixture process some caution is needed in the prior specification of the coefficients of the covariates in the mean of the log-scale process. We suggest two different prior specifications as benchmarks for practitioners and evaluate the impact of these different prior distributions in the resulting inference for the kurtosis of the process.

Notice that the inclusion of covariates in the scale mixture depends on the choice of a link function connecting the scale, which is positive, to the covariates. We have chosen the log link as it is the most often used to transform from positive to real line. However, other choices of link functions could be considered depending on the application and model comparison criteria could be used to select the best link function.

In the real data analysis, although altitude is considered in the mean structure of the process, the inclusion of altitude in the mean of the scale mixing process led to improved predictions when compared to the NGP model. As already observed in different studies (e.g. Schmidt et al. (2011), Viana Neto et al. (2014)), the mean of the predicted values under the different fitted models do not differ much. However, our proposed model seems to perform better in terms of the uncertainty of the predictions, providing better accommodation of the outlying observations (see Figure 9).

Overall, the proposed model added flexibility to the class of spatial mixture models often considered in the literature as an alternative to the Gaussian assumption. A natural extension of this work is to investigate how this covariance-covariate dependence may be changing in time. This is a direction we intend to investigate further in future research. One possibility is to follow the specification in Fonseca and Steel (2011) for a spatiotemporal process $Z(s_i, t_j) = \mathbf{w}(s_i, t_j)'\boldsymbol{\delta} + \sigma \frac{\epsilon(s_i,t_j)}{\sqrt{\lambda_1(s_i)\lambda_2(t_j)}}$ and consider $\ln(\boldsymbol{\lambda}_1) \sim N_I\left(-\frac{\nu_1}{2}\mathbf{1}_I, \nu_1\Sigma_1\right)$ and $\ln(\boldsymbol{\lambda}_2) \sim N_J\left(-\frac{\nu_2}{2}\mathbf{1}_J, \nu_2\Sigma_2\right)$. As in equation (4) the parameter ν_1 might depend on spatial covariates while ν_2 might depend on temporal covariates. We believe this extension might add flexibility in applied data analysis.

Acknowledgements

This work was part of the Ph.D. thesis of R. S. Bueno in the Graduate Program in Statistics of the Federal University of Rio de Janeiro, under the supervision of A. M. Schmidt and T. C.

Fonseca. Bueno benefited from a scholarship from *Conselho de Aperfeiçoamento de Pessoal de Nível Superior* (CAPES), Brazil. T. C. Fonseca was partially supported by *Conselho Nacional de Desenvolvimento Científico e Tecnológico* (CNPq). A. M. Schmidt was partially supported by CNPq and *Fundação de Amparo à Pesquisa do Estado do Rio de Janeiro* (FAPERJ), Brazil, and the Natural Sciences and Engineering Research Council (NSERC) of Canada.

References

Bolin, D. (2014). "Spatial Matérn fields driven by non-Gaussian noise." *Scandinavian Journal of Statistics*, 41, 557–579.

Calder, C. A. (2008). "A dynamic process convolution approach to modeling ambient particular matter concentrations." *Environmetrics*, 19, 39–48.

Cooley, D., Nychka, D., and Naveau, P. (2007). "Bayesian spatial modelling of extreme precipitation return levels." *Journal of the American Statistical Association*, 102, 824–840.

Cressie, N. (1993). *Statistics for Spatial Data*. John Wiley & Sons, New York.

De Oliveira, V., Kedem, B., and Short, D. A. (1997). "Bayesian prediction of transformed Gaussian random fields." *Journal of the American Statistical Association*, 92, 1422–1433.

Fonseca, T. C. O. and Steel, M. F. J. (2011). "Non-Gaussian spatiotemporal modelling through scale mixing." *Biometrika*, 98, 761–774.

Gelman, A. and Rubin, D. B. (1992). "Inference from iterative simulation using multiple sequences." *Statistical Science*, 7, 457–511.

Gneiting, T. (2000). "Power-law correlations, related models for long-range dependence and their simulation." *Journal of Applied Probability*, 37, 4, 1104–1109.

Gneiting, T., Balabdaoui, F., and Raftery, A. E. (2007). "Probabilistic forecasts, calibration and sharpness." *Journal of Royal Statistical Society: Series B*, 69, 243–268.

Gneiting, T. and Schlather, M. (2004). "Stochastic Models That Separate Fractal Dimension and the Hurst Effect." *SIAM review*, 46, 269–282.

Higdon, D. (2002). "Space and space-time modeling using process convolutions." In *Quantitative Methods for Current Environmental Issues*, eds. C. W. Anderson, V. Barnett, P. C. Chatwin, and A. H. El-Shaarawi, 37–56. London: Springer.

Ingebrigtsen, R., Lindgren, F., and Steinsland, I. (2014). "Spatial models with explanatory variables in the dependence structure." *Spatial Statistics*, 8, 20–38.

Palacios, M. B. and Steel, M. F. J. (2006). "Non-Gaussian Bayesian geostatistical modelling." *Journal of the American Statistical Association*, 101, 604–618.

R Core Team (2015). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Reich, B., Eidvisk, J., Guindani, M., Nail, A. J., and Schmidt, A. M. (2011). "A class of covariate-dependent spatiotemporal covariance functions." *The Annals of Applied Statistics*, 5, 2425–2447.

Sampson, P. and Guttorp, P. (1992). "Nonparametric estimation of nonstationary spatial covariance structure." *Journal of the American Statistical Association*, 87, 108–119.

Schmidt, A. M., Guttorp, P., and O'Hagan, A. (2011). "Considering covariates in the covariance structure of spatial processes." *Environmetrics*, 22, 487–500.

Ver Hoef, J. M., Peterson, E., and Theobald, D. (2006). "Spatial statistical models that use flow and stream distance." *Environmental and Ecological Statistics*, 13, 449–464.

Viana Neto, J. H., Schmidt, A. M., and Guttorp, P. (2014). "Accounting for spatially varying directional effects in spatial covariance structures." *Journal of the Royal Statistical Society: Series C*, 63, 103–122.

Wallin, J. and Bolin, D. (2015). "Geostatistical modelling using non-Gaussian Matérn fields." *Scandinavian Journal of Statistics*, 42, 872–890.

Zhang, H. (2004). “Inconsistent Estimation and Asymptotically Equal Interpolations in Model-Based Geostatistics.” *Journal of the American Statistical Association*, 99, 250–261.

A Proofs to the theorems of Section 2.1

In this Section we prove the results shown in Propositions 3.1 and 3.2. Consider the spatio-temporal model in equation (1), the mixture process in equation (3) and the parameter $\nu(s)$ as specified in equation (4).

Proof of Proposition 2.1

$$
\begin{aligned}
Cov[Z(s_1,t_1),Z(s_2,t_2)] &= Cov\left[\mathbf{w}(s_1,t_1)'\boldsymbol{\delta}+\sigma\frac{\epsilon(s_1,t_1)}{\sqrt{\lambda(s_1)}}\ ,\ \mathbf{w}(s_2,t_2)'\boldsymbol{\delta}+\sigma\frac{\epsilon(s_2,t_2)}{\sqrt{\lambda(s_2)}}\right] \\
&= Cov\left[\sigma\frac{\epsilon(s_1,t_1)}{\sqrt{\lambda(s_1)}}\ ,\ \sigma\frac{\epsilon(s_2,t_2)}{\sqrt{\lambda(s_2)}}\right] = \sigma^2\ E\left[\frac{\epsilon(s_1,t_1)}{\sqrt{\lambda(s_1)}}\ \frac{\epsilon(s_2,t_2)}{\sqrt{\lambda(s_2)}}\right] \\
&= \sigma^2\ E\left[\epsilon(s_1,t_1)\ \epsilon(s_2,t_2)\right]\ E\left[\lambda(s_1)^{-1/2}\ \lambda(s_2)^{-1/2}\right] \\
&= \sigma^2\ C(d_s,d_t)\ E\left[\lambda(s_1)^{-1/2}\ \lambda(s_2)^{-1/2}\right] \\
&= \sigma^2\ C_1(d_s)C_2(d_t)\ E\left[\exp\left\{-\frac{1}{2}ln[\lambda(s_1)]-\frac{1}{2}ln[\lambda(s_2)]\right\}\right] \\
&= \sigma^2C_1(d_s)C_2(d_t)\exp\left\{\frac{3}{8}\left[\nu(s_1)+\nu(s_2)\right]+\frac{\sqrt{\nu(s_1)\nu(s_2)}}{4}C_1(d_s)\right\}.
\end{aligned}
$$

Proof of Proposition 2.2

Let $m(s,t)=\mathbf{w}(s,t)'\boldsymbol{\delta}$. As the kurtosis is computed through the fourth central moment of $Z(s,t)$ scaled by its squared variance, we have that

$$
\begin{aligned}
K_Z(\mathbf{x}(s),\boldsymbol{\beta}) = Kurt[Z(s,t)] &= \frac{E\{[Z(s,t)-E[Z(s,t)]]^4\}}{[Var[Z(s,t)]]^2} \\
&= \frac{E[Z(s,t)^4]-4m(s,t)E[Z(s,t)^3]+6m(s,t)^2E[Z(s,t)^2]-3m(s,t)^4}{[\sigma^2\exp\{\nu(s)\}]^2}
\end{aligned}
$$

Below we compute each of the expected values in the equation above separately.

$$
\begin{aligned}
E[Z(s,t)^4] &= E\left\{m(s,t)^4+4m(s,t)^3\sigma\frac{\epsilon(s,t)}{\lambda(s)^{1/2}}+4m(s,t)\sigma^3\frac{\epsilon(s,t)^3}{\lambda(s)^{3/2}}+6m(s,t)^2\sigma^2\frac{\epsilon(s,t)^2}{\lambda(s)}+\sigma^4\frac{\epsilon(s,t)^4}{\lambda(s)^2}\right\} \\
&= m(s,t)^4+6m(s,t)^2\sigma^2E\left\{\frac{\epsilon(s,t)^2}{\lambda(s)}\right\}+\sigma^4E\left\{\frac{\epsilon(s,t)^4}{\lambda(s)^2}\right\} \\
&= m(s,t)^4+6m(s,t)^2\sigma^2E[\epsilon(s,t)^2]E[\lambda(s)^{-1}]+\sigma^4E[\epsilon(s,t)^4]E[\lambda(s)^{-2}] \\
&= m(s,t)^4+6m(s,t)^2\sigma^2E[\exp\{-ln(\lambda(s))\}]+3\sigma^4E[\exp\{-2ln(\lambda(s))\}] \\
&= m(s,t)^4+6m(s,t)^2\sigma^2e^{\nu(s)}+3\sigma^4e^{3\nu(s)}.
\end{aligned}
$$

$$\begin{aligned}
E[Z(s,t)^3] &= E\left\{m(s,t)^3 + 3m(s,t)^2\sigma\frac{\epsilon(s,t)}{\lambda(s)^{1/2}} + 3m(s,t)\sigma^2\frac{\epsilon(s,t)^2}{\lambda(s)} + \sigma^3\frac{\epsilon(s,t)^3}{\lambda(s)^{3/2}}\right\} \\
&= m(s,t)^3 + 3m(s,t)\sigma^2 E\left\{\frac{\epsilon(s,t)^2}{\lambda(s)}\right\} = m(s,t)^3 + 3m(s,t)\sigma^2 e^{\nu(s)}.
\end{aligned}$$

$$\begin{aligned}
E[Z(s,t)^2] &= E\left\{m(s,t)^2 + 2m(s,t)\sigma\frac{\epsilon(s,t)}{\lambda(s)^{1/2}} + \sigma^2\frac{\epsilon(s,t)^2}{\lambda(s)}\right\} \\
&= m(s,t)^2 + \sigma^2 E\left\{\frac{\epsilon(s,t)^2}{\lambda(s)}\right\} = m(s,t)^2 + \sigma^2 e^{\nu(s)}.
\end{aligned}$$

Substituing the equalities above into the expression of $Kurt[Z(s,t)]$, we finally get that $Kurt[Z(s,t)] = 3\exp\{\nu(s)\}$.

B Posterior full conditionals

In this section we show the resultant full conditional posterior distributions for the parameters of the model. Again, assume the spatio-temporal model defined in equations (1), (3) and (4) then the full conditional posterior distributions are, respectively, given by:

- Full conditional posterior for σ^2

  $$p(\sigma^2|\cdot) = IG\left(\frac{IJ}{2} + a_{\sigma^2}; \left[\frac{1}{2}(\mathbf{Z} - \mathbf{w}\boldsymbol{\delta})'\ R^{-1}\ (\mathbf{Z} - \mathbf{w}\boldsymbol{\delta})\right] + b_{\sigma^2}\right),$$

  where $R = \Sigma_2 \otimes \left[\Lambda^{-1/2}\Sigma_1\Lambda^{-1/2}\right]$.

- Full conditional posterior for α_1, a_1

  $$\begin{aligned}
  p(\alpha_1, a_1|\cdot) &\propto |\Sigma|^{-1/2}\exp\left\{-\frac{1}{2}(\mathbf{Z} - \mathbf{w}\boldsymbol{\delta})'\ \Sigma^{-1}\ (\mathbf{Z} - \mathbf{w}\boldsymbol{\delta})\right\}|W|^{-1/2} \\
  &\times \exp\left\{-\frac{1}{2}\left[\left(ln(\boldsymbol{\lambda}) + \frac{1}{2}\boldsymbol{\nu}\right)' W^{-1}\left(ln(\boldsymbol{\lambda}) + \frac{1}{2}\boldsymbol{\nu}\right)\right]\right\} I_{\alpha_1}(0,2)\ a_1^{a_{a_1}-1}\exp\{-a_1 b_{a_1}\},
  \end{aligned}$$

  where $W = \text{diag}(\sqrt{\boldsymbol{\nu}})\ \Sigma_1\ \text{diag}(\sqrt{\boldsymbol{\nu}})$. As this kernel does not belong to an known distribution, we use a Metropolis-Hastings step, with random-walk proposals based on transformations of the parameters. In particular, the proposed value for a_1 was sampled in the log scale, and for α_1 we used the transformation $\ln(\alpha_1/(2-\alpha_1))$, such that the back transformation falls within the interval $(0,2)$.

- Full conditional posterior for α_2, a_2

  $$p(\alpha_2, a_2|\cdot) \propto |\Sigma|^{-1/2} \exp\left\{ -\frac{1}{2}(\mathbf{Z} - \mathbf{w}\boldsymbol{\delta})' \, \Sigma^{-1} \, (\mathbf{Z} - \mathbf{w}\boldsymbol{\delta}) \right\} I_{\alpha_2}(0,2) \; a_2^{a_{a_2}-1} \exp\{-a_2 b_{a_2}\}.$$

  The steps to sample from this full conditional are the same as those for α_1 and a_1 described above.

- Full conditional posterior for $\boldsymbol{\beta}$

  $$p(\boldsymbol{\beta}|\cdot) \propto |W|^{-1/2} \exp\left\{ -\frac{1}{2} \left[\left(ln(\boldsymbol{\lambda}) + \frac{1}{2}\boldsymbol{\nu} \right)' W^{-1} \left(ln(\boldsymbol{\lambda}) + \frac{1}{2}\boldsymbol{\nu} \right) \right] \right\} \; p(\boldsymbol{\beta}).$$

  As this is an unknown distribution we use a Metropolis-Hastings step, with proposal based on random walk proposals for transformations of the parameters. In particular, for β_0 we made proposals for $\ln(-\beta_0)$, whereas for β_1 the proposal was based on the transformation $\ln\left((\beta_1 - a.beta1)/(b.beta1 - \beta_1)\right)$, where $a.beta1$ and $b.beta1$ are the limits of the uniform conditional prior assigned to β_1.

- Full conditional posterior for $\boldsymbol{\delta}$

  $$p(\boldsymbol{\delta}|\cdot) = N_p(\mathbf{Z}\Sigma^{-1}\mathbf{w}D; D),$$

  where $D = [\mathbf{w}'\Sigma^{-1}\mathbf{w} + [\; diag(\sigma^2_\delta) \;]^{-1}]^{-1}$.

- Full conditional posterior for $\boldsymbol{\lambda}$

  $$\begin{aligned} p(\boldsymbol{\lambda}|\cdot) \;\; &\propto \;\; |\Sigma|^{-1/2} \exp\left\{ -\frac{1}{2}(\mathbf{Z} - \mathbf{w}\boldsymbol{\delta})' \, \Sigma^{-1} \, (\mathbf{Z} - \mathbf{w}\boldsymbol{\delta}) \right\} \\ &\qquad \times \exp\left\{ -\frac{1}{2} \left[\left(ln(\boldsymbol{\lambda}) + \frac{1}{2}\boldsymbol{\nu} \right)' W^{-1} \left(ln(\boldsymbol{\lambda}) + \frac{1}{2}\boldsymbol{\nu} \right) \right] \right\} \end{aligned}$$

  In this step, we use random walk proposals to generate values of $\boldsymbol{\lambda}$ considering groups in space in order to block the sampler for $\boldsymbol{\lambda}$ (Palacios and Steel, 2006).

C Predictive distribution

Usually, the main aims in spatiotemporal modelling are spatial interpolation and temporal predictions. Under our proposed model this is easily achieved due to the conditional nature of the model specification. Given the mixing latent variables $\lambda(s)$, the data follows a multivariate Gaussian distribution and predictions are obtained through the properties of the multivariate normal distribution.

Specifically, consider an unobserved part of the spatiotemporal process in arbitrary space-time coordinates $(s_{p1}, t_{p1}), \ldots (s_{pn}, t_{p1}), \ldots, (s_{p1}, t_{pm}), \ldots (s_{pn}, t_{pm}) \in D \times T$. Our goal is to make conditional inference for $\mathbf{Z}_p = (Z(s_{p1}, t_{p1}), \ldots, Z(s_{pn}, t_{p1}), \ldots, Z(s_{p1}, t_{pm}), \ldots, Z(s_{pn}, t_{pm}))'$, based on the observed data $\mathbf{Z} = (Z(s_1, t_1), \ldots, Z(s_I, t_J))'$. Let $\boldsymbol{\theta} = (\boldsymbol{\lambda}, \boldsymbol{\Psi}) \in \Theta$ be the unknown parameters for the proposed model (3), thus the predictive distribution is given by $p(\mathbf{Z}_p|\mathbf{Z}) = \int_{\boldsymbol{\theta}} p(\mathbf{Z}_p|\boldsymbol{\lambda}, \boldsymbol{\Psi}, \mathbf{Z}) p(\boldsymbol{\lambda}|\boldsymbol{\Psi}, \mathbf{Z}) p(\boldsymbol{\Psi}|\mathbf{Z}) \; d\boldsymbol{\theta}$. The latent variable vector is partitioned according to the respective set of observed and unobserved locations in space, $(\boldsymbol{\lambda}, \boldsymbol{\lambda}_p)$, and the predictive distribution may be rewritten as

$$p(\mathbf{Z}_p|\mathbf{Z}) = \int_{\boldsymbol{\theta}} p(\mathbf{Z}_p|\boldsymbol{\lambda}, \boldsymbol{\Psi}, \mathbf{Z}) p(\boldsymbol{\lambda}_p|\boldsymbol{\lambda}, \boldsymbol{\Psi}, \mathbf{Z}) p(\boldsymbol{\lambda}|\boldsymbol{\Psi}, \mathbf{Z}) p(\boldsymbol{\Psi}|\mathbf{Z}) \; d\boldsymbol{\theta}. \tag{7}$$

The predictive distribution is then obtained by composition sampling using the parameter values sampled from the posterior distributions $p(\boldsymbol{\Psi} \mid \mathbf{Z})$ and $p(\boldsymbol{\lambda} \mid \boldsymbol{\Psi}, \mathbf{Z})$ in the MCMC algorithm. The densities $p(\mathbf{Z}_p|\boldsymbol{\lambda}, \boldsymbol{\Psi}, \mathbf{Z})$ and $p(\boldsymbol{\lambda}_p|\boldsymbol{\lambda}, \boldsymbol{\Psi}, \mathbf{Z})$ are sampled for each $(\boldsymbol{\lambda}, \boldsymbol{\Psi})$ obtained from the posterior distribution. The covariance matrix for $(\ln(\boldsymbol{\lambda}), \ln(\boldsymbol{\lambda}_p))$ and $(\mathbf{Z}', \mathbf{Z}_p')'$, respectively are partitioned according to

$$\tilde{S} = \begin{pmatrix} S & S_{op} \\ S_{po} & S_{pp} \end{pmatrix}, \tilde{\Sigma} = \begin{pmatrix} \Sigma & \Sigma_{op} \\ \Sigma_{po} & \Sigma_{pp} \end{pmatrix}.$$

Define $\boldsymbol{\nu} = (\nu(s_1), \ldots, \nu(s_I))'$ and $\boldsymbol{\nu}_p = (\nu(s_{p1}), \ldots, \nu(s_{pn}))'$ thus $\tilde{S}$ is computed according to the model in (3), so that $S_{kl} = \sqrt{\nu(s_k)\nu(s_l)} \; C_1(||s_k - s_l||), k, l = 1, \ldots, I$ and $S_{pp,kl} = \sqrt{\nu(s_{pk})\nu(s_{pl})} \; C_1(||s_{pk} - s_{pl}||)$, $k, l = 1, \ldots, n$ and $S_{op,kl} = \sqrt{\nu(s_k)\nu(s_{pl})} \; C_1(||s_k - s_{pl}||)$, $k = 1, \ldots, I, l = 1, \ldots, n$. Notice that the covariate in the covariance model has to be defined for all spatial locations in the domain of interest. Furthermore, it is desired that the covariate varies smoothly across space in order to preserve the smoothness properties of the process $Z(\cdot, \cdot)$. In addition, $\tilde{\Sigma}$ is defined from (1) such that $\Sigma = \Sigma_2 \otimes \; \sigma^2 \left[\Lambda^{-1/2} \; \Sigma_1 \; \Lambda^{-1/2}\right]$, and $\Sigma_{pp} = \Sigma_{2p} \otimes \; \sigma^2 \left[\Lambda_p^{-1/2} \; \Sigma_{1p} \; \Lambda_p^{-1/2}\right]$, with $\Sigma_{1p,kl} = C_1(||s_{pk} - s_{pl}||)$,$k, l = 1, \ldots, n$, $\Sigma_{2p,kl} = C_2(|t_{pk} - t_{pl}|)$, $k, l = 1, \ldots, m$ and $\Sigma_{op} = \Sigma_{2op} \otimes \; \sigma^2 \left[\Lambda^{-1/2} \; \Sigma_{1op} \; \Lambda_p^{-1/2}\right]$, with $\Sigma_{1op,kl} = C_1(||s_k - s_{pl}||)$,$k = 1, \ldots, I, l = 1, \ldots, n$, $\Sigma_{2op,kl} = C_2(|t_k - t_{pl}|)$, $k = 1, \ldots, J, l = 1, \ldots, m$.

Thus, the predictive distribution for $\ln(\boldsymbol{\lambda}_p)$ is

$$p(\ln(\boldsymbol{\lambda}_p)|\boldsymbol{\lambda}, \boldsymbol{\Psi}) = f_m\left(\ln(\boldsymbol{\lambda}_p); -\frac{\boldsymbol{\nu}_p}{2} S_{po} S^{-1}\left(\ln(\boldsymbol{\lambda}) + \frac{\boldsymbol{\nu}}{2}\right), S_{pp} - S_{po} S^{-1} S_{op}\right).$$

And given the sampled values of $(\boldsymbol{\lambda}, \boldsymbol{\lambda}_p)$, and the desired predictions $\mathbf{Z}_p$ are obtained from

$$p(\mathbf{Z}_p|\mathbf{Z}, \boldsymbol{\lambda}, \boldsymbol{\lambda}_p, \boldsymbol{\Psi}) = f_N\left(\mathbf{Z}_p; \mathbf{m}_p + \Sigma_{po}\Sigma^{-1}(\mathbf{Z} - \mathbf{m}), \Sigma_{pp} - \Sigma_{po}\Sigma^{-1}\Sigma_{op}\right),$$

with $f_K(\mathbf{x}; \tilde{m}, \tilde{\Sigma})$ denoting a K-variate Gaussian distribution with mean function $\tilde{m}$ and covariance matrix $\tilde{\Sigma}$, where $\mathbf{m} = \mathbf{w}\boldsymbol{\delta}$ and $\mathbf{m}_p = \mathbf{w}_p\boldsymbol{\delta}$ are the mean vectors as defined in Section 3.

D Model comparison criteria

Scoring rules provide summaries for the evaluation of probabilistic forecasts by comparing the predictive distribution with the actual value which is observed for the process (Gneiting et al., 2007). We will use scoring rules in a Bayesian context as measures for comparing models based on their posterior predictive distribution, in particular, we consider the logarithmic predictive score and the interval score.

Interval score (IS)

Define the $(1-\alpha)100\%$ prediction interval by (q_1, q_2). The interval score is given by

$$IS(q_1, q_2; z_o) = (q_2 - q_1) + \frac{2}{\alpha}(q_1 - z_o)I(z_o < q_1) + \frac{2}{\alpha}(z_o - q_2)I(z_o > q_2),$$

where z_o is the observed value. The first term refers to the range of prediction interval and the other terms increments the IS when the interval does not contain the true value. In general, it is used $\alpha = 0,05$ resulting in a range of 95% of credibility.

Logarithmic predictive score (LPS)

Here we use the log predictive score based on the predictive density value at the observed $\mathbf{z}$,

$$LPS(\mathbf{z}) = -\ln\{p(\mathbf{z}|\mathbf{z}_o)\},$$

where $p(\mathbf{z}|\mathbf{z}_o)$ denotes the posterior predictive density at $\mathbf{z}$ of the model under consideration and $\mathbf{z}_o$ is the observed value. Consider $\boldsymbol{\theta}^j$, $j = 1, \ldots, M$, a sample from the posterior distribution of $\boldsymbol{\theta}$. Then, an approximation of $\ln\{p(\mathbf{z}|\mathbf{z}_o)\}$ is

$$\ln\widehat{\{p(\mathbf{z}|\mathbf{z}_o)\}} = \frac{1}{M}\sum_{j=1}^{M} \ln\{p(\mathbf{z}|\mathbf{z}_o, \boldsymbol{\theta}^j)\},$$

where $p(\mathbf{z}|\mathbf{z}_o, \boldsymbol{\theta}^j)$ denotes the predictive density at the observed value $\mathbf{z}$ based on the j-th sampled value from the posterior distribution of $\boldsymbol{\theta}$.

E Analysis of synthetic data generated from the proposed model

Following Section 2, we generated synthetic data from the proposed spatio-temporal model defined in equations (1), (3) and (4). Our aim is to check if the procedure of inference is correct and if all the model parameters are identifiable. Consider the model given by:

$$Z(s_i, t_j) = \delta_0 + \delta_1 \, w(s_i) + \sigma\frac{\epsilon(s_i, t_j)}{\sqrt{\lambda(s_i)}}, \; i = 1, \ldots, 65, \; j = 1, \ldots, 30.$$

To depict a realistic region we considered the coordinates of the Colorado state in the USA for data generation. The process $\epsilon(\cdot, \cdot)$ has zero mean and covariance function $C(d_s, d_t) = C_1(d_s)C_2(d_t)$, with d_s representing the Euclidean distances in space and d_t representing the difference in time. We assume $C_i(d) = \left(1 + \left(\frac{||d||}{a_i}\right)^{\alpha_i}\right)^{-1}$, $i = 1, 2$; $w(s_i)$ is a coordinate Y in location s_i. We consider $I = 65$ locations and $J = 30$ instants in time. We assume $\ln \nu(s_i) = \beta_0 + \beta_1 x_1(s_i)$ where $x_1(s_i)$ is a standardized altitude in location s_i. Table 4 shows the fixed values of the parameters for data simulation. These values were fixed to provide realistic values of the artificial data. Figure 10 shows the spatial coordinates together with 5 locations that are left out from the inference procedure for predictive purposes.

Table 4: Values of the parameters used create the synthetic data.

Parameter	δ_0	δ_1	σ^2	a_1	α_1	a_2	α_2	β_0	β_1
True value	0.6	1.2	0.2	2.57	1.5	1.0	1.5	-0.8	0.5

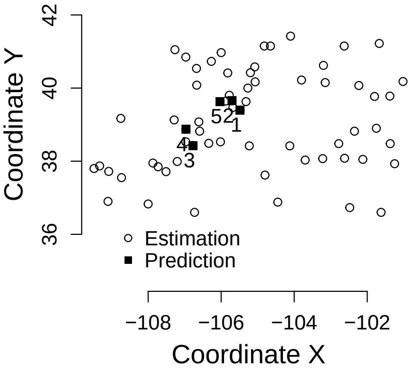


Figure 10: Monitoring locations (circles) of the artificial data generated under the proposed process. The numbered locations (black squares) are the ones left out from the inference procedure to check the predictive ability of the model.

Next we fitted the following two models:

NGP.X (I) with the following prior specification $\beta_0 \sim NT_{-}(0, 3)$ and $\beta_1 \sim U(-1, 1)$ (Prior 1 in Subsection 3.1);

NGP.X (D) with the following prior specification $\beta_0 \sim U(-4, 0)$ e $\beta_1|\beta_0 \sim U(-L + \beta_0, L - \beta_0)$ with $L = \ln(\ln(30/3))$ (Prior 2 in Subsection 3.1);

We assume prior independence among the other parameters of the model, and assign the following prior distributions: $a_1 \sim gamma(0.01, 0.01/m)$, m is the median of the Euclidean distance among locations; $a_2 \sim gamma(0.01, 0.01)$; $\alpha_i \sim U(0, 2)$, $i = 1, 2$; $\sigma^2 \sim IG(2.1, 1)$ and $\boldsymbol{\delta} \sim N_2(0, diag(100))$. We run two chains starting from very different initial values and let the MCMC algorithm run for 30,000 iterations, used 10,000 as burn in and kept every other 20th iteration to avoid autocorrelation among the sampled values. Convergence of the chains was checked using $\hat{R}$ test of Gelman and Rubin (1992).

The panels of Figure 11 show the posterior summary (median and 95% credible intervals) for all parameters under models NGP.X(I) (black lines) and NGP.X(D) (gray lines). The symbol '*' represents the respective true value of the parameter. Clearly, all posterior credible intervals contain the true values of the parameters. Moreover, the point estimates are very close to the true values.

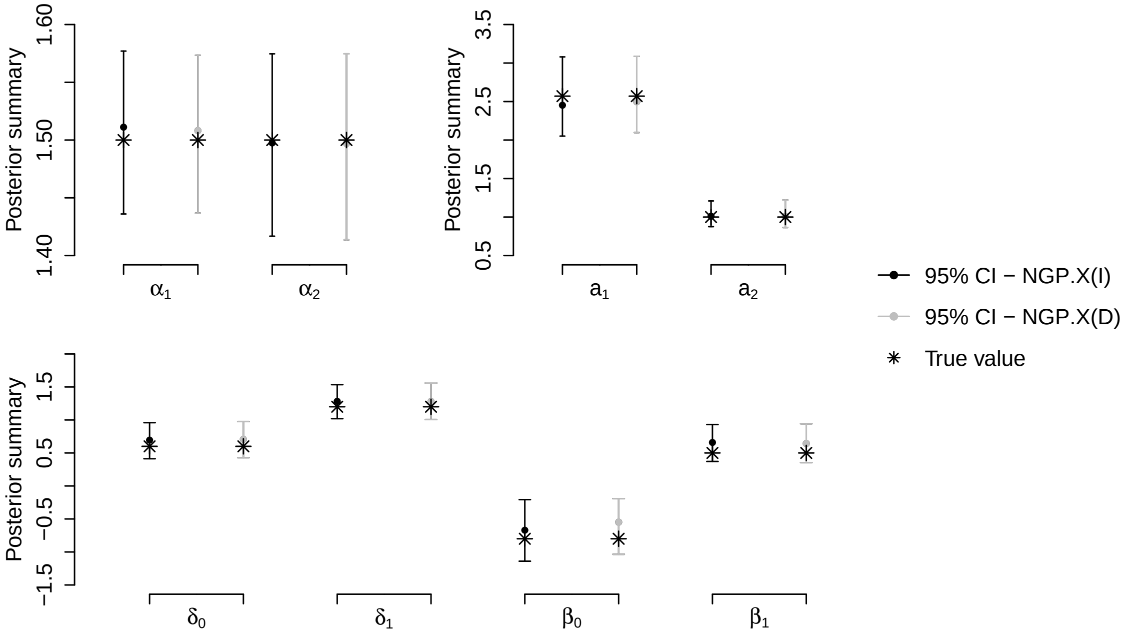


Figure 11: Posterior summary (median and 95% credible intervals) for all parameters from model NGP.X(I) (black line) and NGP.X(D) (gray line). The asterisks represents true value.

The panels of Figure 12 show the summary of the posterior predictive distribution obtained for the five locations that were left out from the inference procedure for predictive purposes. The proposed models recover quite well these observations.

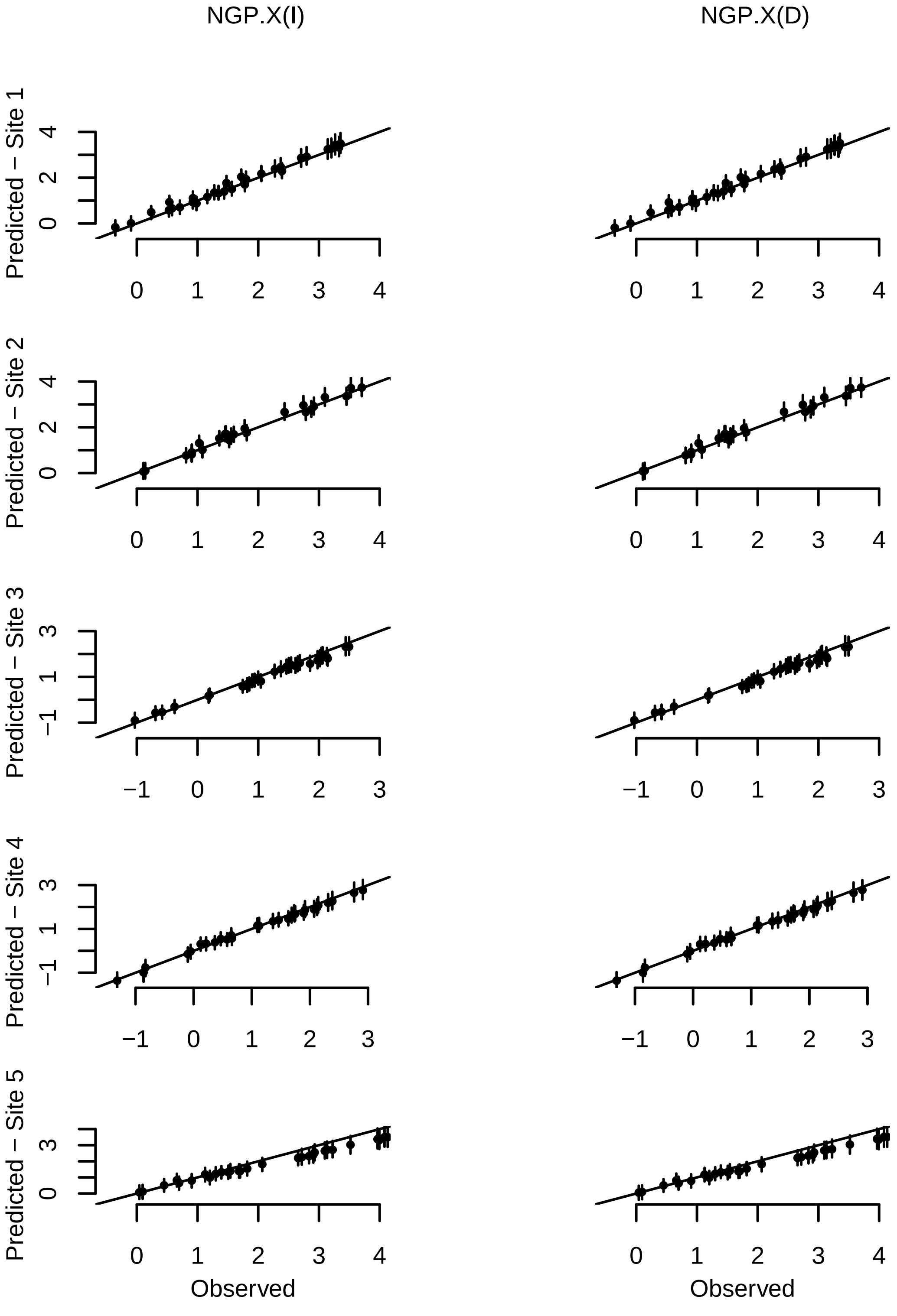


Figure 12: Posterior summary (median and 95% credible intervals) for five locations left out from the inference procedure for the 30 time points.