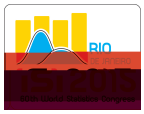


Modelling the kurtosis of spatiotemporal processes

Renata Souza Bueno
Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil - renata@dme.ufrj.br
Escola Nacional de Ciências Estatísticas - IBGE, Rio de Janeiro, Brazil

Thaís C. O. Fonseca
Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil - thais@im.ufrj.br

Alexandra M. Schmidt
Universidade Federal do Rio de Janeiro, Rio de Janeiro, Brazil - alex@im.ufrj.br

Abstract

The goal of this work is to propose a novel class of robust models for phenomena that vary continuously in space and time, in which the variance of the process may depend on spatial covariates. Thus, it is expected that the covariates help to explain spatial heterogeneity. The resulting model is non-stationary as the covariance function depends on spatial locations. In geostatistical data, it is usually assumed Gaussianity for the spatial process of interest. However, in practice, this might not be appropriate when the real data present outliers or heterogeneity. The non-Gaussian process presented by Fonseca and Steel (2011) models the variance of the phenomenon as a spatial process. Our proposal is to allow this variance process to depend on spatial covariates. The idea is to improve the explanation of variability of the process by allowing the kurtosis to vary with spatial location. The inference procedure is performed under the Bayesian framework. We present an illustrative example in the modeling of the maximum temperatures in the Spanish Basque country. The resultant posterior predictive interval for unobserved locations of the proposed model is narrower than the ones obtained under the usual Gaussian process or the process presented by Fonseca and Steel (2011). Further, the parameters associated with the kurtosis process have interesting interpretation.

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

1. Introduction

The development of methods for the analysis of spatiotemporal processes has increased considerably due to computational advances. The models that describe these processes incorporate spatial and temporal dependencies between observations in order to understand the response variable and also to make predictions for future times or unsampled sites. There are many practical applications in environmental, hydrological, and ecological studies. Usually, the models used to describe spatiotemporal processes are based on Gaussian processes. However, real data distributions often deviate from Gaussianity, presenting heavy tails or skewness. The objective of this work is to propose a model for spatiotemporal process that aim to accommodate non-Gaussian tail behaviour.

The approach considered was introduced by Palacios and Steel (2006) and extended to the context of spatiotemporal process by Fonseca and Steel (2011). In this approach a non-Gaussian process is defined through a scale mixture. The mixing variables in this mixture are latent variables which model the variance process in study. This fact allows spatial heteroscedasticity and heavier tails than the Gaussian process. Thus, this proposal is able to accommodate possible outliers. The idea in this work is to incorporate spatially referenced covariates in the model for the latent mixing variable, which models the variance of the process. It is expected that the use of these covariates help to bring information about the variance of the process and the behavior of the data in study. The proposed approach allows the kurtosis of the process to vary with location, making the model able to accommodate different distributions across the observed locations. Also, as the covariance structure depends on the location, the resultant process is non-stationary.

The paper is divided as follows. Section 2 presents a brief review of spatiotemporal models. Section 3 introduces the proposed kurtosis model for spatiotemporal geostatistical data and Section 4 presents an illustrative

example concerning the maximum temperatures in the Spanish Basque country. Finally, Section 5 presents some discussion and conclusions.

2. Spatiotemporal modelling

Consider a stochastic spatiotemporal process defined by $\{Z(s,t) : s \in D; t \in T\}$ where (s,t) are space-time coordinates that vary continuously in $D \times T$, $D \subseteq \mathbb{R}^d, T \subseteq \mathbb{R}$, typically $d = 1, 2$, or 3. The process is Gaussian when is assumed that the finite-dimensional distributions implied by this process are all Gaussian. Thus, assume that $Z(s_1,t_1), Z(s_2,t_2), \ldots, Z(s_I,t_J)$ are observations of this process at locations s_i $(i = 1, \ldots, I)$ and times t_j $(j = 1, \ldots, J)$. Then $Z(s_1,t_1), Z(s_2,t_2), \ldots, Z(s_I,t_J)$ has a multivariate normal distribution with mean vector

$$\boldsymbol{m} = (m(s_1,t_1), \ldots, m(s_I,t_J))' \text{ and } \Sigma_{ij} = Cov[Z(s_i,t_i), Z(s_j,t_j)]$$

is a covariance matrix, where $(s_1,t_1), \ldots, (s_I,t_J) \in D \times T$.

Building adequate models for spatiotemporal processes is not trivial as the chosen covariance function has to be valid, that is, the covariance function must be positive definite. One possibility to ensure that the covariance function is positive definite is the assumption of separability, i.e., the function can be written as

$$\Sigma_{ij} = Cov[Z(s_i,t_i), Z(s_j,t_j)] = C_1(s_i,s_j)C_2(t_i,t_j),$$

where C_1 and C_2 are valid spatial and temporal covariance functions, respectively. Separability is a convenient property for the covariance matrix of the spatiotemporal process, because it can be expressed as a Kronecker product of matrices with smaller dimensions.

The class of Gaussian models is mathematically very convenient because all conditional and marginal distributions are known and prediction is easily obtained. On the other hand, gaussianity is a very restrictive assumption since fit and prediction might not be satisfactory if the dataset presents non-Gaussian characteristics as the presence of outliers. Therefore, it would be useful to consider distributions with heavier tails than the Gaussian distribution to accommodate those values. Fonseca and Steel (2011) extended the ideas by Palacios and Steel (2006) to processes in space and time. Fonseca and Steel (2011) define a spatiotemporal Non-Gaussian process based on a scale mixing of Gaussian processes, which results in a process with heavier tails than the Gaussian one. Consider a particular case of this process defined by

$$Z(s,t) = w(s,t)'\boldsymbol{\delta} + \sigma \frac{\epsilon(s,t)}{\sqrt{\lambda(s)}}, \tag{1}$$

where $\epsilon(s,t)$ follow a Gaussian process defined in $(s,t) \in D \times T$ with a zero mean vector and covariance function $\boldsymbol{C}(\cdot,\cdot)$ defined by $\boldsymbol{C}(d_s,d_t) = \boldsymbol{C}_1(d_s)\boldsymbol{C}_2(d_t)$, where $\boldsymbol{C}_1(d_s)$ is a valid spatial correlation function that only depends on the Euclidean distance, d_s, among locations, and $\boldsymbol{C}_2(d_t)$ is a valid temporal correlation function that only depends on the Euclidean distance, d_t, between instants in time, $\sigma > 0$ is a scale parameter. The mean surface is assumed to be a linear function of $w(s,t)$, a vector of p covariates, with unknown coefficient vector $\boldsymbol{\delta} \in \mathbb{R}^p$.

This process is an alternative to the usual Gaussian spatial process and provides flexibility in the modelling of spatiotemporal data. Possible heterocedasticity in datasets can be incorporated in this model, since it allows the variance to be different in each location. The component $\lambda(s)$ is a latent variable and is responsible for inflating the variance of $Z(.,.)$ when it is close to zero and allows the model to accommodate outliers.

Palacios and Steel (2006) consider that $\lambda(s)$ follows a log-normal distribution with mean 1 and variance $\exp(\nu) - 1$. For large values of ν the variance of the process $Z(.,.)$ decreases. When ν tends to zero, the process becomes Gaussian. In the next section we propose a more flexible model for non-Gaussian processes.

3. Proposed model

The main aim is to account for covariate information in the resultant covariance structure of the proposed model. This proposal allows both to model the variance and the kurtosis of the process as a function of locations. Consider the model proposed by Fonseca and Steel (2011) described in equation (1) where it is assumed that the covariance function of the component $\epsilon(s,t)$ is isotropic, separable and stationary. The

latent variable $\lambda(s)$ is responsible for inflating the variance of the process of interest $Z(.,.)$. The component $\boldsymbol{\lambda}$ in Palacios and Steel (2006) and Fonseca and Steel (2011) is described as

$$ln(\boldsymbol{\lambda})|\nu, \boldsymbol{C}_1 = \{ln(\lambda(s_1)), \ldots, ln(\lambda(s_I))\}' \sim N_I\left(-\frac{\nu}{2}\mathbf{1}\,, \nu\,\boldsymbol{C}_1\right),$$

where $N_I(\cdot,\cdot)$ represents a multivariate normal distribution with dimension I. The parameter $\nu > 0$ controls the behavior of the component $\boldsymbol{\lambda}$ and influences the covariance structure and the tail of the process of interest $Z(.,.)$. We propose to write the parameter ν as a function of spatial covariates allowing it to vary with location. As a result the proposed model allows the tail behavior of the sampling distribution to vary across space. Therefore, for a given location s, the parameter ν will be defined by:

$$\nu(s) = \exp\{\beta_0 + \beta_1 x_1(s) + \beta_2 x_2(s) + \cdots + \beta_{p-1} x_{p-1}(s)\}, \tag{2}$$

where x_i is the i-th spatial covariate and β_i are regression coefficients. The exponential function is a link function used to ensure that in all location s the parameter $\nu(s)$ is positive. Rewriting in matrix form, we obtain that $\boldsymbol{\nu} = \exp\{\boldsymbol{X\beta}\}$, where $\boldsymbol{X}$ is a matrix of dimension $I \times p$. If an intercept is present, the first column is defined by a vector containing 1's, $\mathbf{1}$, and the other $p-1$ columns contain the spatial covariates that are believed to influence $\nu(.)$. The parameter $\boldsymbol{\beta}$ is a $p-$dimensional column vector, containing of the coefficients of the regression variables. Now, the vector $\boldsymbol{\nu}$ is introduced in the modelling of the latent variable $\boldsymbol{\lambda}$ as:

$$ln(\boldsymbol{\lambda}) \mid \boldsymbol{\nu}, \boldsymbol{C}_1 \sim N_I\left(-\frac{1}{2}\,\boldsymbol{\nu}, diag(\sqrt{\boldsymbol{\nu}})\,\boldsymbol{C}_1\, diag(\sqrt{\boldsymbol{\nu}})\right). \tag{3}$$

Thus, the marginal mean and the variance of $\lambda(s)$ are given, respectively, by: $E[\lambda(s)|\nu(s)] = 1$ and $Var[\lambda(s)|\nu(s)] = \exp\{\nu(s)\} - 1$. Note that the variance of $\lambda(s)$ depends on spatial covariates, and so does the variance of the process of interest, $Z(.,.)$. Clearly, $\nu(s)$ controls the variance of $\lambda(s)$. When $\nu(s) \to 0$, the variance of $\lambda(s)$ tends to zero and the $Z(.,.)$ process tends to a Gaussian process. When $\nu(s)$ increases, the distribution of $\lambda(s)$ becomes more dispersed, allowing the variance of $Z(.,.)$ to be inflated at location s and, consequently, accommodating outliers.
The proposed model includes, as a special case, the model described by Fonseca and Steel (2011), in which the parameter $\boldsymbol{\nu}$ is constant across space. Such case occurs when the spatial covariate does not influence the modelling of $\boldsymbol{\nu}$. In our proposal, the covariance function of the process of interest $Z(.,.)$ depends on spatial covariates. This fact makes the proposed process non-stationary. Furthermore, the kurtosis of the process varies with locations, and it can be shown that the kurtosis is given by $Kurt[Z(s,t)] = 3\exp\{\nu(s)\}$.

3.1 Inference Procedure

The inference procedure follows the Bayesian paradigm. Therefore, to complete model specification, a prior distribution must be assigned to the model parameter vector. Following equations (1), (2), and (3), the correlation functions $\boldsymbol{C}_1(d_s)$ and $\boldsymbol{C}_2(d_t)$ are defined by $C_i(d) = \left(1 + \left(\frac{||d||}{a_i}\right)^{\alpha_i}\right)^{-1}$, $i = 1, 2.$, where a_1 and a_2 are decay parameters, and α_1 and α_2 are smoothness parameters. The parameter σ^2 is a scale parameter. In particular, we assume $\nu(s) = \exp\{\beta_0 + \beta_1 x_1^*(s)\}$ where $x_1^*(s)$ is a standardized covariate. Therefore, the parametric vector is defined by: $\boldsymbol{\Psi} = (a_1, \alpha_1, a_2, \alpha_2, \sigma^2, \beta_0, \beta_1, \boldsymbol{\delta})$.
Following the Bayes' theorem, the posterior distribution is proportional to

$$\begin{aligned} p(\boldsymbol{\lambda}, \boldsymbol{\Psi}|\boldsymbol{Z}) &\propto p(\boldsymbol{Z}|\boldsymbol{\lambda}, \boldsymbol{\Psi}) p(\boldsymbol{\lambda}, \boldsymbol{\Psi}) \\ &\propto |\boldsymbol{\Sigma}|^{-1/2} \exp\left\{-\frac{1}{2}(Vec(\boldsymbol{Z}) - \boldsymbol{\omega\delta})'\, \boldsymbol{\Sigma}^{-1}\, (Vec(\boldsymbol{Z}) - \boldsymbol{\omega\delta})\right\} \times p(\boldsymbol{\lambda}|a_1, \alpha_1, \beta_0, \beta_1) \times p(\boldsymbol{\Psi}), \end{aligned}$$

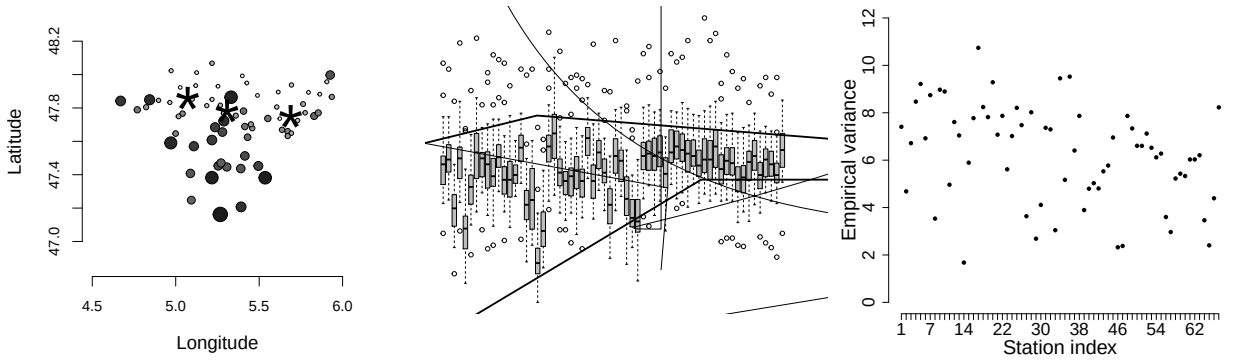
where $\boldsymbol{Z}$ is the matrix of observed data, of dimension $I \times J$; $Vec(\boldsymbol{Z})$ is the matrix $\boldsymbol{Z}$ vectorized by its columns; $\boldsymbol{\omega}$ is the matrix of covariates in the mean function of the process. The covariance matrix $\boldsymbol{\Sigma}$ is given by:

$$\boldsymbol{\Sigma} = C_{\boldsymbol{\theta}_2} \otimes \sigma^2 \left[\Lambda^{-1/2}\, C_{\boldsymbol{\theta}_1}\, \Lambda^{-1/2}\right],$$

where $\Lambda = diag(\lambda(1), \ldots, \lambda(I))$, and the symbol $\otimes$ denotes the Kronecker product.
The resultant posterior distribution does not have a closed form. We make use of Markov chain Monte Carlo algorithms to obtain samples from it. We assume prior independence among the parameters in the higher levels of hierarchy, such that, we assume $a_i \sim Gama(\alpha_{a_i}, \beta_{\alpha_i})$, $i = 1, 2$. For α_1 and α_2, the smoothness parameters in the correlation functions which are defined in the range of $[0, 2]$ we assing $\alpha_i \sim U(0, 2)$, $i = 1, 2$. For the scale parameter, $\sigma^2 > 0$, we assume $\sigma^2 \sim GI(\alpha_{\sigma^2}, \beta_{\sigma^2})$, a priori. Care must be taken when assigning the prior distributions for the coefficients β_0 and β_1, as they might lead to high, therefore unreasonable, values of the resultant kurtosis of the process. Here we suggest $\beta_0 \sim NT^-(0, 3)$, where NT^- stands for a truncated normal distribution with mean -2.4 and variance 3.27 and values smaller than 0; for β_1 we suggest an uniform prior with limits depending on the range of the observed covariate $\beta_1 \sim U(-c, c)$. Finally, $\boldsymbol{\delta}$ is the vector containing the regression coefficients in the mean function and we assume $\boldsymbol{\delta} \sim N_p(0, diag(\sigma_\delta))$. As an illustration of the proposed model, in the sequel we analyze maximum temperature observed in the Spanish Basque country.

4. Application to maximum temperature observed in the Spanish Basque Country

The data studied in this application were analyzed in Fonseca and Steel (2011). It was considered the maximum temperatures recorded daily in July 2006 at 70 locations within the Spanish Basque country. It was used 67 of these locations for parameter estimation and three locations were left out of the estimation procedure for predictive comparison. Panels of Figure 1 show some summaries of the observed data. In particular, panels (b) and (c) provide evidence that the variability of the observed maximum temperature changes over space. Therefore it seems reasonable to consider models that are able to accommodate this behavior.



(a) Position of the 70 stations (b) Temperature over space (c) Empirical variance for each location

Figure 1: Summaries of the observed temperature: (a) Observed Locations (circles are locations considered in the inference procedure, where the size is proportional to altitude and * represents those left out from the inference procedure); (b) Box-plot of the temperatures across spatial locations; (c) Empirical variances of the time series observed at each spatial location.

The altitude in the region of interest varies considerably and might influence the variability maximum temperature. And this is the covariate we consider when modelling $\nu(s)$. To obtain reasonable estimates for β_1, altitude was standardized. In order to compare our proposed model with standard ones in the literature, we also fitted two other models which are obtained as particular cases of the proposed one. The fitted models are the following:

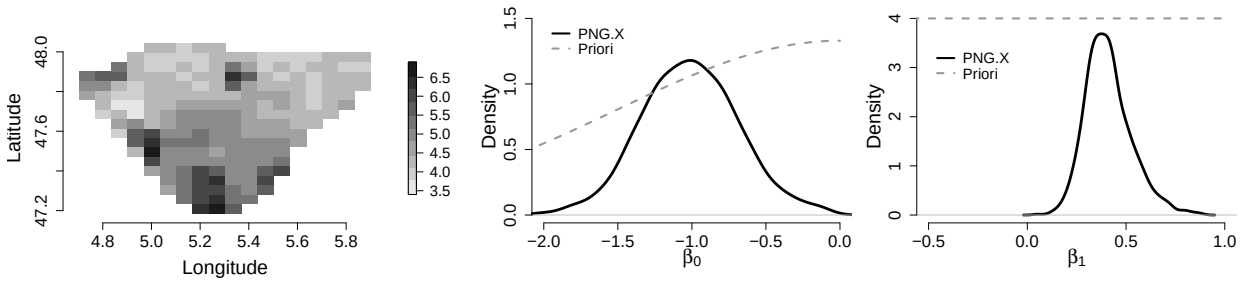
- **Model PNG.X**: $\nu(s)$ as described in equation (2);
- **Model PNG**: $\nu(s) = \nu$ for all $s \in D$;

• **Model PG**: $\nu(s) = 1$ for all $s \in D$.

In the PNG.X model the latent variable $\boldsymbol{\lambda}$, which influences the variance of the process, is modeled as $\nu(s) = \beta_0 + \beta_1 x_1(s)$, where $x_1(s)$ is the standardized values of altitude. Following Fonseca and Steel (2011), the mean function, for all models was modelled as

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

The MCMC algorithm was implemented in the language R (R Development Core Team (2013)). It was considered a burn-in of 20,000 iterations, and after that every 40^{th} draw was stored, resulting in a sample of size 1,000 from the posterior distribution.
Panel (a) of Figure 2 shows the posterior median map of kurtosis under the PNG.X model. It is clear that the kurtosis vary with locations. The south has higher values of altitude, and the estimated kurtosis seem to follow this same pattern, what seems reasonable for maximum temperature. Panels (b) and (c) of this same figure, show posterior densities of β_0 and β_1, which are related with the kurtosis of the process. Clearly, there is information gain when the resultant posteriors are compared to the respective prior distributions. Moreover, β_1 is estimated as positive, suggesting that higher values of altitude lead to higher values of the kurtosis.



(a) (b) (c)

Figure 2: Panel (a) Map of posterior median of kurtosis. Panels (b) and (c) Prior (dashed lines) and posterior (solid lines) densities of the parameters related to the kurtosis, β_0 and β_1.

Figure 3 presents the posterior predictive median and 95% credible intervals for stations 1, 2 and 3, across the temporal sampling period (see Panel (a) of Figure 1), under each of the fitted models. All models provide predictions which follow the pattern of the observed data. For stations 1 and 3, model PNG.X results in the smallest ranges of the 95% predictive intervals, when compared to the other two fitted models. On the other hand, fort station 2, model PNG.X leads to the biggest ranges of the 95% predictive intervals.
Model comparison was performed using the Bayes' factor and the interval score as proposed by Jeffreys (1961). Table 1 shows that model PNG.X resulted in the highest value of the Bayes factor, suggesting it provides the best fit, among the fitted models, under this criterion. As for prediction, under the interval score criterion the best model was also the proposed one, as it resulted in the lowest value of this criterion.

Table 1: Twice the natural logarithm of the Bayes factor in favour of the model in the row versus Gaussian model and the average interval score.

Model	2 Ln(Bayes Factor)	Interval score
PG	-	6,508
PNG	602,45	5,748
PNG.X	874,32	5,497

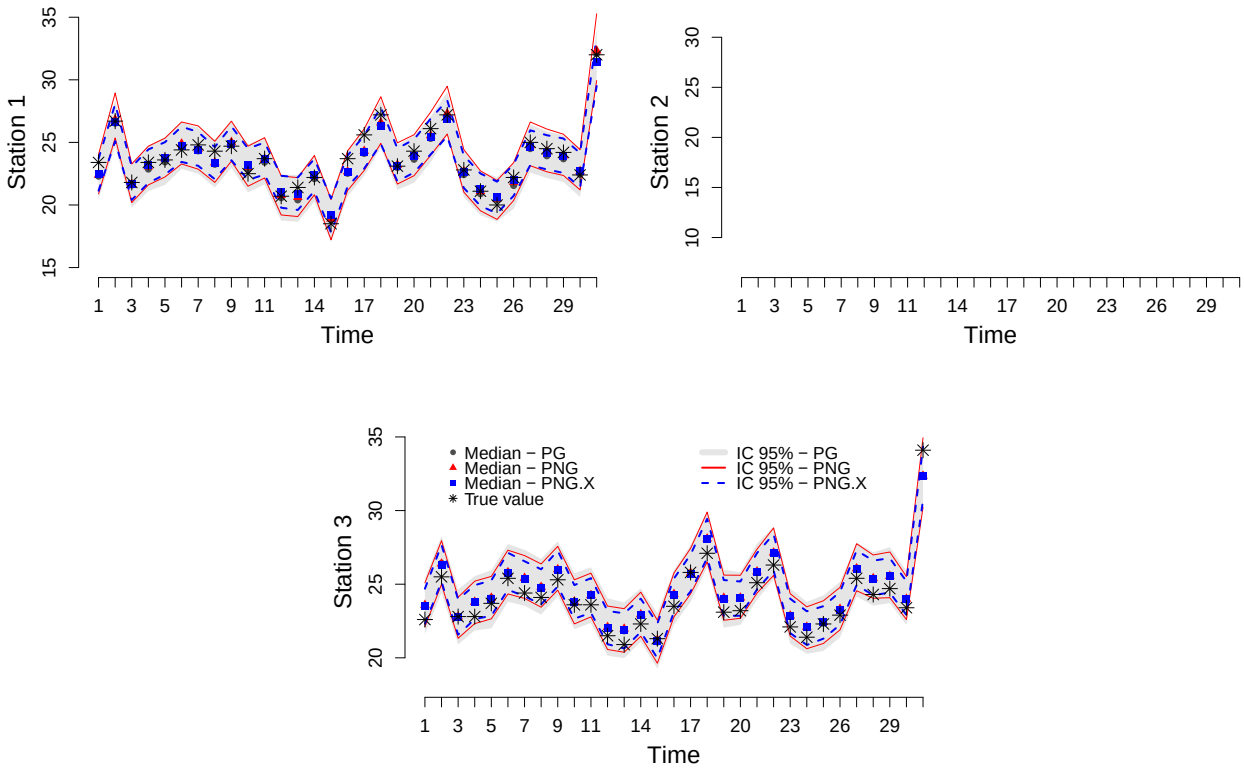


Figure 3: Posterior predictive medians and 95% credible predictive intervals for the differents fitted models and for the three stations across the temporal sampling period. True values are represented by asterisks.

5. Conclusions

This paper proposes the use of covariate information in the mixing latent component of the spatiotemporal model proposed by Fonseca and Steel (2011). This allows the kurtosis of the process to vary with location. The parameters that model the variance of the process have interesting interpretations, and influence the kurtosis of the distribution of the phenomenon under study.

The proposed model was fitted to maximum temperature data osberved at the Spanish Basque country and the mean and the variance of the process were modelled as functions of altitude. The results suggest that altitude influence not only the mean function, but also the variance of the process. And the estimated kurtosis surface suggest that higher altitudes are related to greater variability of the maximum temperatures. Also, when compared to particular cases, the proposed model resulted in the best predictive performance for the 3 locations left out from the inference procedure.

References

Jeffreys, H. (1961). Theory of Probability. Oxford University Press, USA, Third edition.

Palacios, M. B. & Steel, M. F. J. (2006). Non-gaussian bayesian geostatistical modelling. Journal of the American Statistical Association, 101, 604-618.

Fonseca, T. C. O. & Steel, M. F. J. (2011). Non-gaussian spatiotemporal modelling through scale mixing. Biometrika, 98, 761-774.

R Development Core Team (2013) R: A language and environment for statistical computing. URLhttp://www.R-project.org. ISBN 3-900051-07-0.