SORT 35 (2) July-December 2011, 165-178

Generalized spatio-temporal models

Edilberto Cepeda Cuervo*

Abstract

An important problem in statistics is the study of spatio-temporal data taking into account the effect of explanatory variables such as latitude, longitude and time. In this paper, a new Bayesian approach for analyzing spatial longitudinal data is proposed. It takes into account linear time regression structures on the mean and linear regression structures on the variance-covariance matrix of normal observations. The spatial structure is included in the time regression parameters and also in the regression structure of the variance covariance matrix. Initially, we present a summary of the spatial models and the Bayesian methodology used to fit the models, as a extension of the longitudinal data analysis. Next, the general spatial temporal model is proposed. Finally, this proposal is used to study rainfall data.

MSC: 62F15

Keywords: Bayesian, Fisher scoring, mean-covariance modelling, antedependence, spatial statistics, spatial temporal models, spatial longitudinal data.

1. Introduction

In the context of the parametric multivariate regression model for longitudinal data and under normality, the response variable for each of the m units under study, each having n observations over time, is denoted by $\mathbf{Y}_i = (Y_{i1},\ldots,Y_{in})'$, $i = 1,\ldots,m$. In this case, it is usually assumed that $\mathbf{Y}_i \sim N(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$, with $\boldsymbol{\mu}_i = \mathbf{X}_i\boldsymbol{\beta}$, where $\mathbf{X}_i$ is a matrix of explanatory variables. Thus, if $nm = n \times m$, it is assumed that the nm-response vector $\mathbf{Y} = (\mathbf{Y}_1,\ldots,\mathbf{Y}_m)'$ follows the model

$$\mathbf{Y} = \boldsymbol{\mu} + \boldsymbol{\epsilon}, \text{ with } \boldsymbol{\epsilon} \sim N(\mathbf{0}, \boldsymbol{\Sigma} = \text{diag}(\boldsymbol{\Sigma}_i)), \tag{1.1}$$

where $\boldsymbol{\mu} = (\boldsymbol{\mu}_1,\ldots,\boldsymbol{\mu}_m)'$, $\boldsymbol{\epsilon} = (\boldsymbol{\epsilon}_1,\ldots,\boldsymbol{\epsilon}_m)'$ and $\boldsymbol{\Sigma}_i = \text{Var}(\boldsymbol{\epsilon}_i)$.

* *Address for correspondence:* Departamento de Estadística. Universidad Nacional de Colombia. ecepedac@unal.edu.co
Received: December 2010
Accepted: September 2011

In these models, as it is well known, ϵ_{ij} and ϵ_{ik}, $j \neq k$, $i = 1, \ldots, m$, are not independent. Thus, $\text{Var}(\boldsymbol{\epsilon}_i) = \boldsymbol{\Sigma}_i$ is no longer a diagonal matrix and it would be necessary to model and estimate the off-diagonal elements of the covariance matrix. This modelling approach usually requires to impose some constraints on the elements of Σ_i to guarantee its positive definiteness. For example, in stationary Gaussian processes, such as the ones used in Geostatistics, the covariance between two observations is explicitly determined by their correlation function. More specifically, it is modelled as a function of the (Euclidean) distance between these two observations. Moreover, and given that some of the properties of this function are imposed by its spatial structure, only correlation functions belonging to the families where these requirements hold can be considered (see, e.g., Diggle and Verbyla, 1998, or Stein, 1999).

Spatial data consist of several measurements taken on the space in each of the experimental coordinates in the sample. This falls into the framework of correlated observations and requires the specification and estimation of both the mean and the covariance structures. A central idea to be able to efficiently estimate the covariance matrix was first introduced by Macchiavelli and Arnold (1994) and Macchiavelli and Moser (1997) and it is based on its Cholesky decomposition. This approach has been used for several joint modelling proposals for the mean and covariance structures in the context of longitudinal data (see, e.g., Pourahmadi, 1999 and 2000, or Pan and MacKenzie, 2006).

In our work, we apply the modified Cholesky decomposition of the precision matrix proposed in Macchiavelli and Arnold (1994), since it offers a simple unconstrained and statistically meaningful reparametrization of the covariance matrix. It has a statistical interpretation in longitudinal data through consideration of antedependence models (Gabriel, 1962; Macchiavelli and Arnold, 1994). With this reparametrization, the dependence between the components of $\mathbf{Y}$ can be modelled as functions of explanatory variables. In this case the covariance matrix structure does not depend on the ordering of observations, so we can apply this models for the variance-covariance matrix in the analysis of spatial data. The parameters do not have a practical interpretation anymore, but estimation of the covariance matrix may lead to better estimates of the parameters of the mean model. When there are many observational units, this parametrization can be useful to alleviate problems associated with the high dimensionality of the variance-covariance matrix. In this case, other variables beyond distance between observational units may be included in the model for the correlations. A simulation study is presented in section 5.

In this paper, we apply the Bayesian methodology proposed by Cepeda and Gamerman (2004) for the analysis of spatial data. In the regression models of joint regressions for the mean and covariance matrix, we include a spatial structure in the regression mean parameters through the spatial dependence of them and in the regression models of the variance covariance matrix, including spatial variables. We also extend the longitudinal models proposed by Pouramadi (1999) and the Bayesian methodology proposed by Cepeda and Gamerman (2004) for modelling spatio-temporal data sets. A spatio-temporal structure is included in the mean, that now is a function of the temporal and spatial variables. The spatial-temporal association is captured through $\mathbf{T}$ in the triangular decomposition $\mathbf{\Sigma}^{-1} = \mathbf{T}'\mathbf{D}^{-1}\mathbf{T}$, that now it is not defined as in equation (1.1), since in the spatio-temporal analysis $\mathbf{\Sigma}$ is not a block-diagonal matrix. Initially, it can have all entries different from zero and, thus, $\mathbf{T}$ is a $nm \times nm$ triangular matrix with 1's in the diagonal.

After this introduction, in section 2, a general spatial model is presented. In section 3 a general Bayesian methodology proposed to fit spatial and spatial temporal models is presented. In section 4 the results of a spatial simulation study are presented. In section 5 a general spatial temporal model is proposed. Finally, in section 6, the results of the analysis of rainfall data are presented.

2. The spatial data analysis

In this section, a single observation of each of the n observational units is assumed. Observations are correspondingly arranged in a n dimensional vector $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_n)$, assumed to follow a multivariate normal distribution, so that $\mathbf{Y} \sim (\boldsymbol{\mu}, \mathbf{\Sigma})$, where $\mathbf{\Sigma}$ is a non-negative definite matrix. A crucial requirement for the analysis is that the inverse of the covariance matrix can be efficiently computed and that, in addition, it should also be allowed to have a very general and flexible specification, so that its functional specification is not too restrictive. For this reasons, we adopt the models suggested by Pourahmadi (1999) following the general model setting presented in Cepeda (2001) and Cepeda and Gamerman (2004) and considering the general ante-dependence model (Gabriel, 1962, Zimmerman and Núñez-Antón, 1997), where for a given individual having n observations we have that

$$Y_i - \mu_i = \sum_{j=1}^{i-1} \phi_{ij}(Y_j - \mu_j) + \nu_i, \nu_i \sim N(0, \sigma_i^2), \quad i = 1, \ldots, n, \tag{2.1}$$

where $E(Y_i) = \mu_i$, with $\mu_i = f(\mathbf{x}_i, \boldsymbol{\beta})$ a linear (or nonlinear) function of the vector of parameter $\boldsymbol{\beta}$, $\nu_i \sim N(0, \sigma_i^2)$ are assumed as mutually independent and by convection $\sum_{j=1}^{0} \phi_{ij}(y_i - \mu_j) = 0$. Although i is typically indexed over time (Gabriel, 1962), when a single series of observations is assumed, the covariance matrix structure does not depend on the ordering of observations, and thus, we can apply this model for the variance-covariance matrix in the analysis of spatial data.

Writing (2.1) in matrix form we obtain

$$\boldsymbol{\nu} = \mathbf{T}(\mathbf{Y} - \boldsymbol{\mu}),\ \boldsymbol{\nu} \sim N(\mathbf{0}, \mathbf{D}) \text{ and } \mathbf{D} = \text{diag}(\sigma_i^2) \tag{2.2}$$

where $\boldsymbol{\nu}' = (\nu_1, \ldots, \nu_n)$, $\boldsymbol{\mu}' = (\mu_1, \ldots, \mu_n)$ and $\mathbf{T} = (\tau_{ij})$, with

$$\tau_{ij} = \begin{cases} 1 & \text{if } j = i \\ -\phi_{ij} & \text{if } j < i \\ 0 & \text{elsewhere} \end{cases}$$

and

$$\text{Var}(\boldsymbol{\nu}) = \mathbf{D} = \mathbf{T}\,\text{Var}(\mathbf{Y} - \boldsymbol{\mu})\mathbf{T}^{'} = \mathbf{T}\boldsymbol{\Sigma}\mathbf{T}^{'} \tag{2.3}$$

Thus, as a consequence of (2.3), $\boldsymbol{\Sigma}$ is obtained indirectly by obtaining $\mathbf{D}$ and $\mathbf{T}$. In addition, we should point out that the triangular decomposition in equation (4) is unique. Moreover, given that $\boldsymbol{\Sigma}$ is a symmetric matrix if and only if there exists a unique lower triangular matrix $\mathbf{T}$, with ones in the diagonal, and a unique diagonal matrix $\mathbf{D}$ with positive diagonal entries such that $\mathbf{T}\boldsymbol{\Sigma}\mathbf{T}^{'} = \mathbf{D}$, we also have that $\boldsymbol{\Sigma}$ is positive definite (Pourahmadi, 1999).

From (2.2),

$$\tilde{\mathbf{Y}} = (\mathbf{I}_n - \mathbf{T})\tilde{\mathbf{Y}} + \boldsymbol{\nu}, \tag{2.4}$$

where $\tilde{\mathbf{Y}} = \mathbf{Y} - \boldsymbol{\mu}$ and $\mathbf{I}_n$ is the $n \times n$ identity matrix. Assuming now that there is a vector of (covariance) explanatory variables $w_{ij} = (w_{ij,1}, \ldots, w_{ij,r})^{'}$, we can write

$$\phi_{ij} = \mathbf{w}_{ij}^{\prime}\boldsymbol{\lambda}, 1 \leq j < i \leq n \tag{2.5}$$

where λ is a vector of parameters $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_r)^{'}$. Since $\phi_{ij} = \sum_{l=1}^{r} w_{ij,l}\,\lambda_l$, the matrix $\mathbf{I}_n - \mathbf{T}$ can be expressed as the linear combination

$$\mathbf{I}_n - \mathbf{T} = \lambda_1 \mathbf{W}_1 + \cdots + \lambda_r \mathbf{W}_r \tag{2.6}$$

where $\mathbf{W}_l = (w_{ij,l})$, $l = 1, \ldots, r$ are $n \times n$ matrices such that $w_{ij,l} = 0$, if $i \leq j$, and $w_{ij,l} = 1$, if $i > j$ and $l = 1$ to allow for a constant intercept in the covariance model. Note that, given the unrestricted and flexible specification of the ϕ_{ij}'s, there are no particular restrictions imposed on $\boldsymbol{\Sigma}$, indirectly specified in $\mathbf{T}$. Therefore, the covariance matrix is allowed to have any dependence form. Particular structures may be imposed on $\mathbf{T}$, for example by setting some of the non-zero ϕ_{ij}'s to 0. This can be handled by choosing the matrices $\mathbf{W}_l$, $l = 1, \ldots, r$ appropriately. In general, in longitudinal data set analysis the explanatory variables, $w_{ij,k} = (t_i - t_j)^k$ are associated with differences in time measurements. In the case of the spatial data analysis $w_{ij,k}$ can be defined as a function of the spatial variables such as longitude and latitude.

As a consequence of (2.4) and (2.6), model (2.1) can be expressed in the form

$$\begin{aligned}\tilde{\mathbf{Y}} &= \lambda_1 \mathbf{W}_1 \tilde{\mathbf{Y}} + \cdots + \lambda_r \mathbf{W}_r \tilde{\mathbf{Y}} + \boldsymbol{\nu} \\ &= \lambda_1 \mathbf{V}_1 + \cdots + \lambda_r \mathbf{V}_r + \boldsymbol{\nu} \\ &= \mathbf{V}\boldsymbol{\lambda} + \boldsymbol{\nu} \end{aligned} \tag{2.7}$$

where $\boldsymbol{\nu} \sim N(\mathbf{0}, \mathbf{D})$ and $\mathbf{V} = (\mathbf{V}_1, \ldots, \mathbf{V}_r)$ with $\mathbf{V}_l = \mathbf{W}_l \tilde{\mathbf{Y}}$, for $l = 1, \ldots, r$. Note that for a fixed value of $\boldsymbol{\beta}$, the model $\tilde{\mathbf{Y}} \sim N(\mathbf{V}\boldsymbol{\lambda}, \mathbf{D})$ is obtained.

Given that ϕ_{ij} can be modelled as in (2.5) and that σ_i^2, $i = 1, 2, \ldots, n$, can be modelled in terms of covariates as $g(\sigma_i^2) = \mathbf{z}_i' \boldsymbol{\Upsilon}$, where g is a real positive function, we summarize the full model for the mean $\boldsymbol{\mu}$ and for the matrices $\mathbf{T}$ and $\mathbf{D}$ by

$$\mu_i = \mathbf{x}_i' \boldsymbol{\beta}, \; g(\sigma_i^2) = \mathbf{z}_i' \boldsymbol{\Upsilon}, \; h(\phi_{ij}) = \mathbf{w}_{ij}' \boldsymbol{\lambda} \tag{2.8}$$

for some appropriate functions g and h. In (2.8), $\mathbf{x}_i$, $\mathbf{z}_i$, $\mathbf{w}_{ij}$ are $k \times 1$, $s \times 1$ and $r \times 1$ vectors of explanatory variables. $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_k)'$, $\boldsymbol{\Upsilon} = (\gamma_1, \ldots, \gamma_s)'$ and $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_r)'$ are the parameters vectors corresponding to the mean, variance and covariance, respectively.

We can apply these models for the variance-covariance matrix in spatial data analysis. We suppose that there are many observational units with a spatial distribution and that we have a variable of interest and the explanatory variables for each observational unit. Then we can propose the mean and covariance models in (2.8) to analyze spatial data. In this situation, we have a random vector, where each component is associated with an observational unit. We consider random vectors $\mathbf{Y} = (Y_1, \ldots, Y_N)' \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, with mean $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_N)'$ and concentration matrix $\boldsymbol{\Sigma}^{-1} = \mathbf{T}'\mathbf{D}^{-1}\mathbf{T}$, since the observations Y_i's are not independent.

3. General spatio-temporal model

In section (2) a single observation of each of the n observational units is assumed but, in general, we have several short series, each one associated with one of the observational units, for example, with a meteorological station, where the variable of interest is measured m times through time. This is, we are considering n nonindependent random vectors $\mathbf{Y}_i = (Y_{i1}, \ldots, Y_{im})'$, $i = 1, 2, \ldots, n$, with mean $\boldsymbol{\mu}_i = (\mu_{i1}, \ldots, \mu_{im})'$ and variance covariance matrix $\boldsymbol{\Sigma}_i^{-1} = \mathbf{T}_i'\mathbf{D}_i^{-1}\mathbf{T}_i$. Thus, if we assume normal distribution and if $\mathbf{Y} = (\mathbf{Y}_1, \ldots, \mathbf{Y}_n) = (Y_{11}, \ldots, Y_{1m}, \ldots, Y_{n1}, \ldots, Y_{nm})'$, $\mathbf{Y} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where the variance-covariance matrix $\boldsymbol{\Sigma}$ is not a block diagonal matrix. In this case, the variance-covariance matrix $\boldsymbol{\Sigma}$ is $(nm) \times (nm)$ and there is an $(nm) \times (nm)$ triangular matrix $\mathbf{T}$ and an $(nm) \times (nm)$ diagonal matrix $\mathbf{D}$, such that $\boldsymbol{\Sigma}^{-1} = \mathbf{T}'\mathbf{D}^{-1}\mathbf{T}$. Thus, rewriting the vector $\mathbf{Y}$ as $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_{nm})'$, we can consider the model given by

$$Y_i - \mu_i = \sum_{j=1}^{i-1} \phi_{ij}(Y_j - \mu_j) + \nu_i, \nu_i \sim N(0, \sigma_i^2), \quad i = 1, \ldots, nm. \tag{3.1}$$

where $E(\mathbf{Y}) = (\mu_1, \mu_2, \ldots, \mu_{nm})'$, ϕ_{ij}s are the entries of $\mathbf{T}$, defined as in (2.2), and $\boldsymbol{\nu} = (\nu_1, \nu_2, \ldots, \nu_{nm})$ is the vector of independent random innovation variables. In this case, the entries ϕ_{ij} of $\mathbf{T}$ can be modelled as a function of the time intervals and spatial differences between the n observational units, as for example, differences between their coordinates or mean temperature, among others. The mean can also be modelled as a spatial temporal function.

In a second approximation, the observation of the interest variable is rearranged as $\mathbf{Y} = (Y_{11}, Y_{12}, \ldots, Y_{1m}, \ldots, Y_{k1}, Y_{k2}, \ldots, Y_{km}, \ldots, Y_{n1}, Y_{n2}, \ldots, Y_{nm})'$ and the spatial-temporal dependence through the models is assumed

$$Y_{ki} - \mu_{ki} = \sum_{k=1}^{n} \sum_{j=1}^{i-1} \phi_{ij}{}^{(k)}(Y_{kj} - \mu_{kj}) + \nu_{ki}, \nu_{ki} \sim N(0, \sigma_i^2), \quad i = 1, \ldots, m, \quad k = 1, \ldots, n, \tag{3.2}$$

where $E(Y_{ki}) = \mu_{ki}$, with $\mu_{ki} = \mathbf{x}'_{ki}\boldsymbol{\beta}$, a linear function of parameter $\boldsymbol{\beta}$, $\nu_{ki} \sim N(0, \sigma_{ki}^2)$ are mutually independent and $\sum_{j=1}^{0} \phi_{ij}{}^{(k)}(Y_{kj} - \mu_{kj}) = 0$ is used, and i indexes over time. Writing (2.1) in matrix form, we obtain

$$\boldsymbol{\nu} = \mathbf{T}(\mathbf{Y} - \boldsymbol{\mu}), \ \boldsymbol{\nu} \sim N(\mathbf{0}, \mathbf{D}) \text{ and } \mathbf{D} = \text{diag}(\sigma_{ki}^2) \tag{3.3}$$

where $\boldsymbol{\mu} = (\mu_{11}, \mu_{12}, \ldots, \mu_{1m}, \ldots, \mu_{k1}, \mu_{k2}, \ldots, \mu_{km}, \ldots, \mu_{n1}, \mu_{11}, \ldots, \mu_{nm})'$ and $\mathbf{T} = (\tau_{ij})$, is an nm lower triangular matrix, with

$$\tau_{ij} = \begin{cases} 1 & \text{if } i = j, \ i = nk_1 + l, \ j = nk_1 + r \\ -\phi_{ij}^{(k)} & \text{if } r < l, \ i = nk_1 + l, \ j = nk_2 + r \\ 0 & \text{elsewhere} \end{cases} \tag{3.4}$$

where $1 \leq l, r \leq m$, $k_1, k_2 = 0, 1, 2 \ldots, n-1$, and $k_1 \leq k_2$. Finally, specifying the mean and the variance-covariance models $\mu_i = f(\mathbf{x}_i, \boldsymbol{\beta})$, $\sigma_i^2 = g(\mathbf{z}_i, \boldsymbol{\Upsilon})$, $\phi_{ij} = h(\mathbf{w}_{ij}, \boldsymbol{\lambda})$, respectively, for some appropriately selected functions h, g, and f, the spatial temporal model is completely defined. These functions can be the same as in equation (2.8), however they can be appropriate nonlinear functions. In the last case it is necessary to build a kernel transition function that can be obtained by defining a normal working variable as proposed in Cepeda and Gamerman (2005). Finally, to fit these spatio-temporal models we propose the Bayesian methodology defined in section (4).

4. Bayesian methodology

In this section we present the Bayesian methodology used to fit spatial and spatial temporal models, following the Bayesian methodology proposed by Cepeda and Gamerman (2004) to fit longitudinal data. Assuming the observational model $\mathbf{Y} = (Y_1, \ldots, Y_n)' \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, where $\boldsymbol{\mu}$ depends on $\boldsymbol{\beta}$ through $\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta}$, where $\mathbf{X}$ is the matrix of explanatory variables, and $\boldsymbol{\Sigma}$ depends on $\boldsymbol{\gamma}$ and $\boldsymbol{\lambda}$ through (2.8), the likelihood function is given by

$$L(\boldsymbol{\beta},\ \boldsymbol{\gamma},\ \boldsymbol{\lambda}|\mathbf{Y}) \propto |\mathbf{D}|^{-1/2} \exp\left\{-\frac{1}{2}(\mathbf{Y}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{Y}-\boldsymbol{\mu})\right\},$$

since $|\boldsymbol{\Sigma}| = |\mathbf{T}'|\ |\mathbf{D}|\ |\mathbf{T}| = |\mathbf{D}|$.

Now, a prior distribution $p(\boldsymbol{\theta})$ for $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\lambda})'$ must be assigned to obtain the posterior distribution. For simplicity we assume $\boldsymbol{\theta} \sim N(\boldsymbol{\theta}_0, \boldsymbol{\Sigma}_0)$ where $\boldsymbol{\theta}_0 = (\mathbf{b}_0, \mathbf{g}_0, \mathbf{l}_0)'$ as prior distribution. One possible model for $\boldsymbol{\Sigma}_0$ is the diagonal form, implying prior independence between $\boldsymbol{\beta}, \boldsymbol{\gamma}$ and $\boldsymbol{\lambda}$. In this case, the full conditional prior distributions for $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$ and $\boldsymbol{\lambda}$ are given by normal distributions, denoted by $N(\mathbf{b},\mathbf{B})$, $N(\mathbf{g},\mathbf{G})$ and $N(\mathbf{l},\mathbf{L})$, respectively. The values of $(\mathbf{b},\mathbf{g},\mathbf{l})$ and $(\mathbf{B},\mathbf{G},\mathbf{L})$ are easily evaluated from $\boldsymbol{\theta}_0$ and $\boldsymbol{\Sigma}_0$.

From the Bayes theorem, the posterior distribution for θ is given by

$$\pi(\boldsymbol{\beta},\boldsymbol{\gamma},\boldsymbol{\lambda}|\mathbf{Y}) \propto |\mathbf{D}|^{-1/2} \exp\left\{-\frac{1}{2}(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta})'\Sigma^{-1}(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta}) - \frac{1}{2}(\boldsymbol{\theta}-\boldsymbol{\theta}_0)'\Sigma_0^{-1}(\boldsymbol{\theta}-\boldsymbol{\theta}_0)\right\} \tag{4.1}$$

The posterior distribution (4.1) is intractable analytically and not easily generated from. However, the posterior full conditional distribution $\pi_{\boldsymbol{\beta}} = \pi(\boldsymbol{\beta}\,|\,\boldsymbol{\gamma},\boldsymbol{\lambda})$ is given by

$$\pi(\boldsymbol{\beta}\,|\,\boldsymbol{\gamma},\boldsymbol{\lambda},\mathbf{Y}) \quad \propto \quad \exp\left\{-\frac{1}{2}(\boldsymbol{\beta}-\mathbf{b}^*)'\mathbf{B}^{*-1}(\boldsymbol{\beta}-\mathbf{b}^*)\right\},$$

where $\mathbf{b}^* = \mathbf{B}^*(\mathbf{B}^{-1}\mathbf{b} + \mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{Y})$ and $\mathbf{B}^* = (\mathbf{B}^{-1} + \mathbf{X}'\Sigma^{-1}\mathbf{X})^{-1}$. Therefore,

$$(\boldsymbol{\beta}\,|\,\boldsymbol{\gamma},\boldsymbol{\lambda},\mathbf{Y}) \sim N(\mathbf{b}^*,\mathbf{B}^*). \tag{4.2}$$

Thus, it is possible to sample $\boldsymbol{\beta}$ directly from $\pi_{\boldsymbol{\beta}}$. Values of $\boldsymbol{\beta}$ can be proposed directly from $\pi_{\boldsymbol{\beta}}$ and accepted with probability 1. This is the Gibbs sampler (Geman and Geman, 1984).

From (2.4) and (2.7), the quadratic form $Q(\mathbf{Y}) = (\mathbf{Y}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{Y}-\boldsymbol{\mu})$ appearing in the likelihood can be rewritten as

$$Q(\mathbf{Y}) = \tilde{\mathbf{Y}}'\mathbf{T}'\mathbf{D}^{-1}\mathbf{T}\tilde{\mathbf{Y}} = (\tilde{\mathbf{Y}}-\mathbf{V}\boldsymbol{\lambda})'\mathbf{D}^{-1}(\tilde{\mathbf{Y}}-\mathbf{V}\boldsymbol{\lambda})$$

Therefore, the full conditional distribution $\pi_{\boldsymbol{\lambda}}$ is given by

$$\begin{aligned}\pi(\boldsymbol{\lambda}|\boldsymbol{\beta},\boldsymbol{\Upsilon}) &\propto \exp\left\{-\frac{1}{2}(\tilde{\mathbf{Y}}-\mathbf{V}\boldsymbol{\lambda})'\mathbf{D}^{-1}(\tilde{\mathbf{Y}}-\mathbf{V}\boldsymbol{\lambda})-\frac{1}{2}(\boldsymbol{\lambda}-\mathbf{l})'\mathbf{L}^{-1}(\boldsymbol{\lambda}-\mathbf{l})\right\}, \\ &\propto \exp\left\{-\frac{1}{2}(\boldsymbol{\lambda}-\mathbf{l}^*)'\mathbf{L}^{*-1}(\boldsymbol{\lambda}-\mathbf{l}^*)\right\}\end{aligned}$$

where $\mathbf{l}^* = \mathbf{L}^*(\mathbf{L}^{-1}\mathbf{l}+\mathbf{V}'\mathbf{D}^{-1}\tilde{\mathbf{Y}})$ and $\mathbf{L}^* = (\mathbf{L}^{-1}+\mathbf{V}'\mathbf{D}^{-1}\mathbf{V})^{-1}$. This is,

$$(\boldsymbol{\lambda}|\boldsymbol{\beta},\boldsymbol{\Upsilon},\mathbf{Y}) \sim N(\mathbf{l}^*,\mathbf{L}^*). \tag{4.3}$$

Thus, values of $\boldsymbol{\lambda}$ can be proposed directly from $\pi_{\boldsymbol{\lambda}}$ and accepted with probability 1.

Unless full conditional distributions of $\boldsymbol{\beta}$ and $\boldsymbol{\lambda}$ are known, the full conditional distribution of $\boldsymbol{\Upsilon}$, given by

$$\pi(\boldsymbol{\Upsilon}\,|\,\boldsymbol{\beta},\boldsymbol{\lambda}) \propto |\mathbf{D}|^{-1/2}\exp\left\{-\frac{1}{2}(\tilde{\mathbf{Y}}-\mathbf{V}\boldsymbol{\lambda})'\mathbf{D}^{-1}(\tilde{\mathbf{Y}}-\mathbf{V}\boldsymbol{\lambda})-\frac{1}{2}(\boldsymbol{\Upsilon}-\mathbf{g})'\mathbf{G}^{-1}(\boldsymbol{\Upsilon}-\mathbf{g})\right\}, \tag{4.4}$$

is intractable analytically and not easily generated from. In this case we have to construct suitable proposals for a Metropolis-Hastings step (Hastings, 1970; Gamerman, 1997a).

We used the methodology proposed by Gamerman (1997b) as applied in Cepeda and Gamerman (2001) for modelling heterogeneity in independent normal regression models. The algorithm requires working variables to approximate transformation of the observations around the current parameter estimates. At the $\boldsymbol{\Upsilon}$ iteration, $\boldsymbol{\beta}$ and $\boldsymbol{\lambda}$ are fixed at their current values $\boldsymbol{\beta}^{(c)}$ and $\boldsymbol{\lambda}^{(c)}$ and, given (2.7), the working observation variables are obtained by Fisher scoring process or by Taylor approximation (Cepeda and Gamerman, 2005).

When $g = \log$, the working observation obtained using Fisher scoring process is

$$\tilde{t}_i = \mathbf{z}_i'\boldsymbol{\Upsilon}^{(c)} + \frac{(\tilde{\mathbf{Y}}_i^{(c)} - \mathbf{v}_i^{(c)'}\boldsymbol{\lambda}^{(c)})^2}{\exp(\mathbf{z}_i'\boldsymbol{\Upsilon}^{(c)})} - 1, \quad i = 1,\ldots,n.$$

It has $E(\tilde{t}_i) = \mathbf{z}_i'\boldsymbol{\Upsilon}^{(c)}$ and associated working variances equal to 2. With the process given in Cepeda and Gamerman (2004), the normal transition kernel $q_{\boldsymbol{\Upsilon}}$ based on Fisher scoring methods is obtained as

$$q_{\boldsymbol{\Upsilon}}(\boldsymbol{\Upsilon}^{(c)},\boldsymbol{\Upsilon}^{(n)}) = N(\mathbf{g}^*,\mathbf{G}^*) \tag{4.5}$$

where

$$\begin{aligned}\mathbf{g}^* &= \mathbf{G}^*(\mathbf{G}^{-1}\mathbf{g}+2^{-1}\mathbf{Z}'\tilde{\mathbf{Y}}) \\ \mathbf{G}^* &= (\mathbf{G}^{-1}+2^{-1}\mathbf{Z}'\mathbf{Z})^{-1}.\end{aligned}$$

Given the characteristic of the conditional distributions we will not sample all the components of vector $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\lambda})^{'}$ simultaneously. Explicitly, we sample $\boldsymbol{\beta}$ and $\boldsymbol{\lambda}$ directly from their full conditionals and $\boldsymbol{\gamma}$ from the proposal given in (4.5), appling Metropolis Hastings algorithm.

5. Spatial data analysis: a simulation study

A simulation study was performed to compare parameter estimates and true values, in a spatial model, assuming that the interest variable $\mathbf{Y}$ has normal distribution. Initially, $n = 50$ values of 5 explanatory variables X_i, $i = 1,2,3$, and W_i, $i = 1,2$, were simulated. Values of X_1, X_2 and X_3 were generated from uniform distributions $U[0,50]$, $U[5,15]$ and $U[0,20]$, respectively, and values of W_i, $i = 1,2$, were generated from uniform distributions $U[0,20]$ and $U[5,15]$, respectively. The values of Y were simulated from a multivariate normal distribution with mean $\mu_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i}$ and variance-covariance matrix $\boldsymbol{\Sigma} = \mathbf{T}^{-1}\mathbf{D}(\mathbf{T}^{'})^{-1}$, where $\mathbf{D} = \text{diag}(\sigma_i^2)$, $\mathbf{T} = (-\phi_{ij})$, $\sigma_i^2 = \exp(\gamma_0 + \gamma_1 x_{1i} + \gamma_2 x_{3i})$ and $\phi_{ij} = \lambda_0 + \lambda_1 w_{ij,1} + \lambda_2 w_{ij,2}$, with $\boldsymbol{\beta} = (20, 3, -1.5)'$, $\boldsymbol{\gamma} = (-6, 0.05, -0.25)'$ and $\boldsymbol{\lambda} = (-0.5, 0.04, -0.02)'$. To apply Bayesian methodology, for simplicity, independent normal prior distributions, $\theta \sim N(0, 10^3\mathbf{I}_9)$ were considered for all the parameters.

The posterior parameter estimates and their respective standard deviation are: $\hat{\beta}_0 = 20.003(6.722 \times 10^{-3})$, $\hat{\beta}_1 = 2.999(1.841 \times 10^{-4})$, $\hat{\beta}_2 = -1.500(6.228 \times 10^{-4})$, $\hat{\lambda}_0 = -0.501(0.004)$, $\hat{\lambda}_1 = 0.040(0.001)$, $\hat{\lambda}_2 = -0.020(4.231 \times 10^{-5})$, $\hat{\gamma}_0 = -5.166(0.619)$, $\hat{\gamma}_1 = 0.025(0.014)$ and $\hat{\gamma}_2 = -0.254(0.041)$. From the comparisons between true values and the corresponding estimates, we conclude that the proposed methodology has excellent performance. The estimates are very close to the true values and, in all cases, they have small standard deviation.

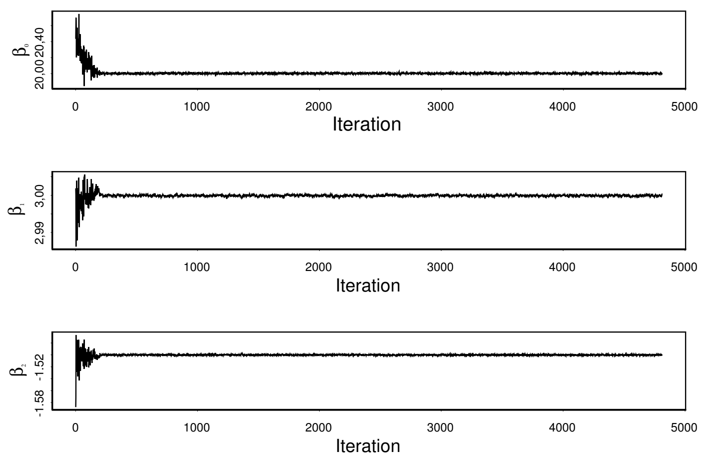

Figure 1: *Posterior chains for the mean parameters:* β_0, β_1, β_2.

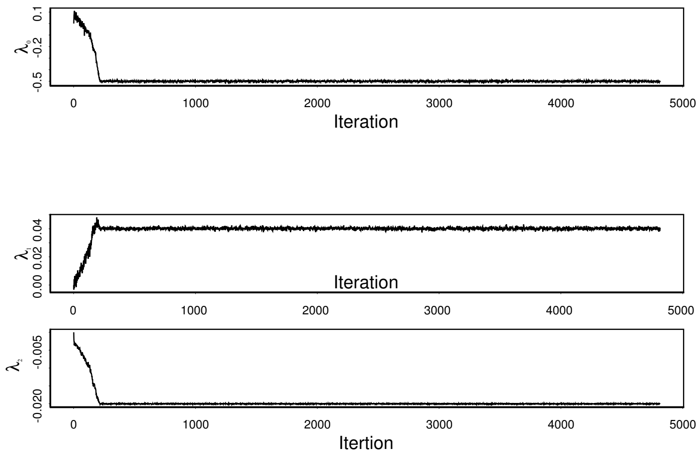

Figure 2: *Posterior samples of the antedependence parameters:* λ_0, λ_1, λ_2.

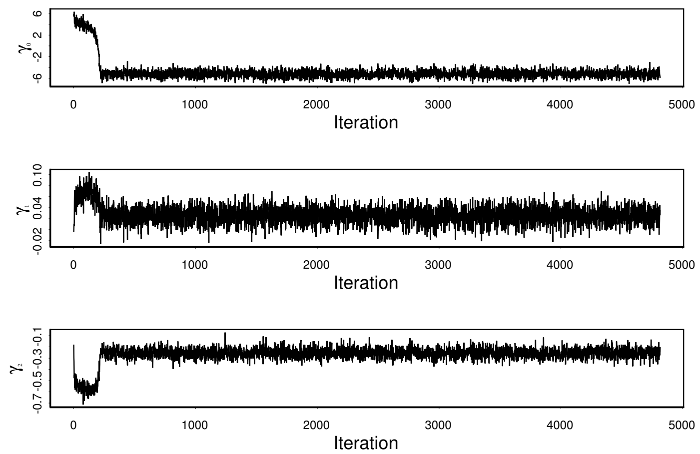

Figure 3: *Posterior samples of the covariance parameters.* γ_1, γ_1, γ_2.

Figures 1, 2 and 3 show the behavior of the chain for the sample simulated for each parameter, where each one has a small transient stage, indicating the speed convergence of simulation for the algorithm. The chain samples are given for the first 4800 iterations. The other results reported in this section are based on a sample of 4000 draws after a burn-in of 800 draws to eliminate the effect of initial values.

The posterior marginal distributions for all the parameters are approximately normal. The p-values of the Kolmogorov-Smirnov test are all larger than 0.05. The posterior sample shows large correlation between parameters of mean models, large correlations between parameters of variance-covariance models, and small but non-negligible correlation between parameters of mean models and parameters of variance models.

6. Application

In this application we consider the precipitation index in the Guajira, a department of Colombia. This department is located in the northeast region of Colombia in a peninsula bounded on the west and north by the Caribbean Sea, on the east by the Gulf of Venezuela and on the south by the Sierra Nevada de Santa Marta. There are six measurement centers distributed along the region. Although it is a flat region, there are differences between the levels of precipitation in each of its points and between seasons and years. For example, in January, February, March and July the level of precipitation is small but in April and May or in October and November all measurements centers report the highest levels of precipitation. In this application we analyze the mean sample of the cumulative monthly precipitation level for 10 years, from 1995 to 2005. The general behavior of the cumulative mean precipitation can be seen in figure 4. From the figure we can infer that the hydrological stations are located in regions with two different regimes of rainfall, characterized by very low levels of precipitation in the first three months of the year. As of April, it is clear that there are three stations located in a region whith high rates of precipitation and three stations where precipitation levels are much lower.

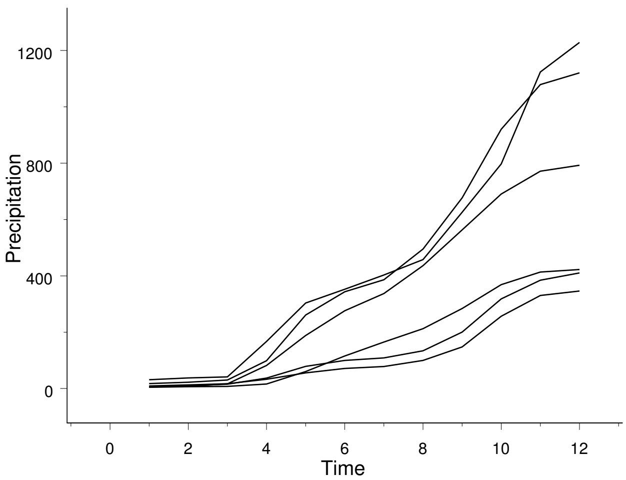

Figure 4: *Annual mean of precipitation.*

In the first analysis of this data we consider the spatio-temporal time model given by

$$\begin{aligned}
\mu_{trs} &= \beta_0 + \beta_1 t + \beta_2 t^2 + \beta_3 t^3 + \beta_4 (st)^3 + \beta_5 (rt)^3 \\
\log\left(\sigma_t^2\right) &= \gamma_0 + \gamma_1 t + \gamma_2 t^2 + \gamma_3 t^3 \\
\phi_{ij} &= \lambda_0 + \lambda_1 (i-j) + \lambda_2 (i-j)^2 + \lambda_3 (i-j)^3 + \lambda_4 (s_i - s_j)^3 + \lambda_5 (r_i - r_j)^3
\end{aligned}$$

where t is the time, s is the latitude with respect to the mean value of the latitude of the observational units, r is the longitude with respect to the mean value of the longitude of the observational units, st is the random variable given by the product of the time by spatial latitude, and rt is the explanatory variable given by the product of the time and spatial longitude. Thus, assuming normal flat prior distribution for the mean, innovation variance and covariance parameters, the posterior estimates and the correspondent standard deviation for the parameters of the model are given by the following values.

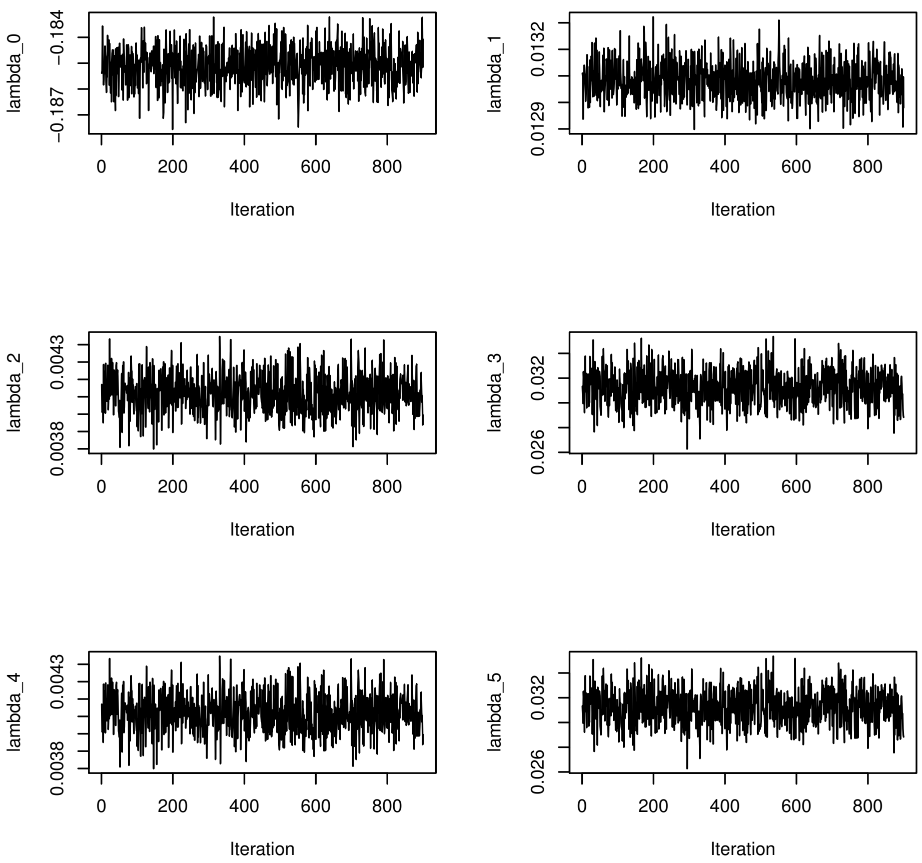

Figure 5: *Posterior samples of the covariance parameters.*

1. For the mean parameters:

$$\hat{\beta}_0 = -32.635(3.877),\ \hat{\beta}_1 = 26.213(2.788),\ \hat{\beta}_2 = -0.548(0.094),$$
$$\hat{\beta}_3 = 0.434(0.001),\ \hat{\beta}_4 = -17.461(0.330),\ \hat{\beta}_5 = 0.184(0.000).$$

2. For the innovation variance:

$$\hat{\gamma}_0 = -9.722(0.357),\ \hat{\gamma}_1 = 0.566(0.000),$$
$$\hat{\gamma}_2 = -0.185(0.000),\ \hat{\gamma}_3 = 0.013(0.000).$$

3. For the covariance parameters:

$$\hat{\lambda}_0 = 0.004(0.000),\ \hat{\lambda}_1 = 1.031(0.000),\ \hat{\lambda}_2 = 53.379(6.611),$$
$$\hat{\lambda}_3 = -45.967(4.604),\ \hat{\lambda}_4 = 9.723(0.201),\ \hat{\lambda}_5 = -0.573(0.001).$$

In all cases, for all the parameters in the model, the chains have a small transient period showing the performance of the proposed algorithm. To illustrate this behavior the chain of the posterior samples for the covariance parameters are included in figure 5. As in the simulation study, in this case, the histograms show that the posterior marginal distributions for all the parameters are approximately normal. The p-values of the Kolmogorov-Smirnov test are all larger than 0.05.

7. Conclusions

In this paper, a new Bayesian flexible and unrestricted approach for analyzing spatial longitudinal data is proposed. We illustrate the usefulness of our proposed methodology by including a Monte Carlo analysis, a simulation study and an application to a real data set. Our proposal puts forward a new way to explore the QR decomposition of the covariance matrix in a longitudinal data setting, by generalizing its application within this context.

Natural extensions of the research presented in this paper are also possible. Classical methodologies applying the Newton Raphson or the Fisher scoring algorithms to fit the proposed models can be easily introduced in the same way as in Pourahmadi (1999) or Cepeda and Gamerman (2004). Nonlinear spatio-temporal models also can be defined by assuming a nonlinear regression models in the mean and covariance models, so that it would be a generalization of the nonlinear longitudinal models proposed in Cepeda-Cuervo and Núñez-Antón (2009), which included a classic and Bayesian generalization that allowed to fit these new models.

Acknowledgements

Cepeda's work was supported by a grant from The Research Division of the National University of Colombia (Universidad Nacional de Colombia). The author wishes to thank the anonymous referees for their careful reading of this manuscript and also for their suggestions that led to very important improvements.

References

Cepeda, E. C. (2001). Variability modelling in generalized linear models. Unpublished Ph.D. thesis, Mathematics Institute, Universidade Federal do Rio de Janeiro.

Cepeda, E. C. and Gamerman D. (2004). Bayesian modelling of joint regression for the mean and covariance matrix. *Biometrical Journal*, 14, 430–440.

Cepeda-Cuervo, E. and Núñez-Antón, V. (2009). Bayesian modelling of the mean and covariance matrix in normal nonlinear models. *Journal of Statistical Computation and Simulation*, 79(6), 837–853.

Cepeda-Cuervo, E. (2005). Bayesian methodology for modelling parameters in the two parameter exponential family. *Revista Estadística*, 57(168-169).

Diggle, P. and Verbyla, A. (1999). Nonparametric estimation of covariance structure in longitudinal date. *Biometrics*, 54, 401–415.

Gabriel, K. R. (1962). Ante-dependence analysis of an ordered set of variables. *Annals of Mathematical Statistics*, 33, 201–212.

Gamerman, D. (1997a). *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*. Chapman & Hall: London.

Gamerman, D. (1997b). Sampling from the posterior distribution in generalized linear mixed models. *Statistics and Computing*, 7, 57–68.

Geman, S. and Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and Bayesian restoration of images. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 6, 721–741.

Hastings, W. K. (1970). Monte Carlo sampling methods using Markov Chains and their applications. *Biometrika*, 57, 97–109.

Macchiavelli, R. E. and Arnold, S. F. (1994). Variable order antedependence models. *Communications in Statistics*, 23, 2683–2699.

Macchiavelli, R. and E. B. Moser (1997). Analysis of repeated measures with ante-dependence models. *Biometrical Journal*, 39(3), 339–350.

Pan, J. and MacKenzie G (2006). Regression models for covariance structures in longitudinal studies. *Statistical Modelling*, 6(1), 43–57.

Pourahmadi, M. (1999). Joint mean-covariance models with applications to longitudinal data: Unconstrained parameterizations. *Biometrika*, 86, 667–690.

Pourahmadi, M. (2000). Maximum likelihood estimation of generalized linear models for multivariate normal covariance matrix. *Biometrika*, 87, 425–435.

Stein, M. (1999). *Interpolation of Spatial Data*. Springer: New York.

Zimmerman, D. and Núñez-Antón, V. (1997). Structured antedependence models for longitudinal data, in *Modelling Longitudinal and Spatially Correlated Data. Methods, Applications, and Future Directions*, T. G. Gregoire, D. R. Brillinger, P. J. Diggle, E. Russek-Cohen, W. G. Warren, and R. Wolnger (eds.), 63–76. Lecture Notes in Statistics No. 122. Springer: New York.