Hierarchical Dynamic Models

Marina Silva Paez and Dani Gamerman

Universidade Federal do Rio de Janeiro, Brazil

marina,dani@im.ufrj.br

1 Introduction

Consider a population of students divided into groups (schools or classes), for which we believe there are similarities (about a certain aspect of interest) within students in the same group. More specifically, suppose we are interested in analyzing the proficiency in maths of students attending to the same level at different schools, and all of them are submitted to the same test. It is reasonable to expect that the average score obtained by the students is not the same at every school, as characteristics of the school could possibly be affecting this average. For this specific example, the teaching skills of the maths teacher could be an important aspect. Other variables could also be important, such as the average social-economic background of the parents. This problem can be expressed in a hierarchical framework written in two hierarchical levels: firstly, a regression equation is specified at the student level, possibly including explanatory variables such as gender or age. The intercept and possibly some regression coefficients of this regression are allowed to vary between schools, according to equations specified in the second level of the hierarchy. This equations can be specified, for example, as regressions at the school level, taking into account explanatory variables at this level. By allowing the regression coefficients of certain explanatory variables to vary between schools, the model is taking into account the fact that the effects of these variables in the maths score of the students

can vary by school.

Suppose now that the scores are obtained yearly for a certain period of time. It is reasonable to expect that the conditions of each school could be changing through time. Therefore, a modification can be made in the model described above to allow the regression intercept to vary between schools and also through time. In a more general framework, all the regression coefficients can be allowed to vary within these dimensions. Hierarchical models that consider a time variation for the parameters through dynamic linear models (Harrison and Stevens, 1979), are denominated Dynamic Hierarchical Models. These models are well documented in Gamerman and Migon (1993).

Dynamic Hierarchical Models can be applied to data in many fields of interest. Specifically, many applications are made to environmental data due to the usual need of modeling their variation in time, and many times in space as well. Some motivating examples and applications to real data-sets are presented throughout this chapter. The remainder of the chapter is organized as follows: in Section 2, the Dynamic Hierarchical Models are presented, and inference for these models is presented in Section 3. Section 4 presents extensions of these models for observations in time and space. Finally Section 5 presents some concluding remarks.

2 Dynamic Hierarchical Models for Univariate Observations

In this section, Dynamic Hierarchical Models are introduced through *Example 1* below, that illustrates the problem presented in the introduction of this chapter. After the methodology is presented and its specific notation is introduced, two other examples will

be given.

Example 1: maths scores of pupils from different schools over time

This example illustrates the problem presented in the introduction of this chapter, where the interest lies in modeling the scores obtained in maths tests by students of different schools.

Firstly, suppose we have a total J schools, with n_j students being tested in the j^{th} school, $j = 1, \cdots, J$. Let y_{ij} be the maths score for the i^{th} student of the j^{th} school, and suppose his/her age (x_{1ij}) and gender (x_{2ij}) were also observed. A linear regression model could be specified to explain the maths scores using age and gender as explanatory variables. A possibly better specification, however, would take into consideration that the students coming from the same school tend to have similarities. This means that students with the same characteristics (gender and age) should not be expected to achieve the same scores if they study at different schools. To accommodate this, one could allow the intercept of the original regression to vary by school. Variables observed at the school level could help explaining this variation. In this example, we suppose that the following school related variables were also observed: years of experience of the maths teacher (x_{3j}), number of maths classes per week (x_{4j}), and an indicator of whether the students received any support from a tutor (x_{5j}), $j = 1, \cdots, J$. It is reasonable to expect that the regression coefficients related to these explanatory variables will be all positive - meaning that the scores of the students of a certain school tend to be higher if the maths teacher is more experienced, they have higher number of maths classes and received the support of a tutor. A possible hierarchical model for this data would be:

$$\begin{aligned} y_{ij} &= \beta_{0,j} + \beta_1 x_{1ij} + \beta_2 x_{2ij} + \epsilon_{ij}, \quad \epsilon_{ij} \sim N(0, \sigma^2), \\ \beta_{0,j} &= \beta_0 + \beta_3 x_{3j} + \beta_4 x_{4j} + \beta_5 x_{5j} + u_j, \quad u_j \sim N(0, \sigma_u^2). \end{aligned}$$

Now suppose tests are applied yearly to students of a certain level over a period of T years. For simplification purposes, we suppose that the number of students tested in each school does not vary through the years, under the hypothesis that no student will change schools over this period, or fail an year. It is now reasonable to assume that the intercept of the regression at the student level is changing over time, as well as by school, if you take into account that characteristics of the schools could be changing over time. We will also suppose that the intercept of the regression at the school level is changing over time through an autoregressive structure. That reflects that other characteristics of the school, which are not specified in the model, can also be affecting the scores of the students - and these characteristics may vary smoothly in time. The proposed model becomes:

$$
\begin{aligned}
y_{ijt} &= \beta_{0,jt} + \beta_1 x_{1ij} + \beta_2 x_{2ij} + \epsilon_{ijt}, \quad \epsilon_{ijt} \sim N(0, \sigma^2), && (1)\\
\beta_{0,jt} &= \beta_{0,t} + \beta_3 x_{3jt} + \beta_4 x_{4jt} + \beta_5 x_{5jt} + u_{jt}, \qquad u_{jt} \sim N(0, \sigma_u^2), && (2)\\
\beta_{0,t} &= \beta_{0,t-1} + w_t, \qquad w_t \sim N(0, \sigma_w^2), && (3)
\end{aligned}
$$

where y_{ijt} denotes the score obtained by the i^{th} student of the j^{th} school at the t^{th} year. Note also that the variables measured at the school level are now indexed by time. The age of the students can be considered fixed if we work, for example, with the age given at the beginning of the experiment.

For this example, the coefficients related to gender and age are fixed. Note, however, that they could also be varying by school, time or both. Similarly, the slope coefficients at the second level of the hierarchy: β_3, β_4 and β_5, could also be varying over time.

In matrix notation, the above model can be written as

$$\boldsymbol{y}_t = \boldsymbol{X}_{1t}\boldsymbol{\beta}_{1,t} + \boldsymbol{v}_{1t}, \quad \boldsymbol{v}_{1t} \sim N_n(\boldsymbol{0}, \boldsymbol{V}_1), \tag{4}$$

$$\boldsymbol{\beta}_{1,t} = \boldsymbol{X}_{2t}\boldsymbol{\beta}_{2,t} + \boldsymbol{v}_{2t}, \quad \boldsymbol{v}_{2t} \sim N_{3J}(\boldsymbol{0}, \boldsymbol{V}_2), \tag{5}$$

$$\boldsymbol{\beta}_{2,t} = \boldsymbol{I}_4\boldsymbol{\beta}_{2,t-1} + \boldsymbol{w}_t, \qquad \boldsymbol{w}_t \sim N_4(\boldsymbol{0}, \boldsymbol{W}), \quad t \geq 1, \tag{6}$$

where $\boldsymbol{I}_k$ denotes a $k \times k$ identity matrix. $\boldsymbol{y}_t$ is a vector of size n where the first n_1 elements represent the scores of the students of the 1^{st} school, the following n_2 elements represent the scores of the students of the 2^{nd} school, and so on, for a fixed time t. That way, $n = n_1 + n_2 + \cdots + n_J$. $\boldsymbol{X}_{1t}$ is a matrix of dimension $n \times 3J$, obtained by the direct sum of the matrices $\boldsymbol{\theta}_{1j}, j = 1, \cdots, J$, where $\boldsymbol{\theta}_{1j}$ is a $n_j \times 3$ matrix with all the elements in the firt column being equal to one, and the i^{th} elements of the second and third columns being the gender and the age of the i^{th} student of the j^{th} school. $\boldsymbol{\beta}_{1,t}$ is a vector of order $3J \times 1$ of coefficients, where the first n_1 sequences of three elements are given by $\beta_{0,1t}$, β_1 and β_2 (the coefficents corresponding to the first school); the following n_2 sequences of three elements are given by $\beta_{0,2t}$, β_1 and β_2 (the coefficents corresponding to the second school), and so forth. $\boldsymbol{v}_{1t}$ is a vector of errors $n \times 1$ and $\boldsymbol{V}_1 = \boldsymbol{I}_n\sigma^2$. Analogously, $\boldsymbol{X}_{2t}$ is a matrix $3J \times 4$, obtained by the direct sum of the matrices $\boldsymbol{\theta}_{2j}$, where $\boldsymbol{\theta}_{2j}$ is a 1×4 matrix with unitary elements in the first column and the other three columns being the values of the covariates at the second level of the hierarchy for school j; $\boldsymbol{\beta}_{2,t}$ and $\boldsymbol{v}_{2t}$ are vectors of coefficients and errors, respectively, of dimension 4×1 and $\boldsymbol{V}_2 = \boldsymbol{I}_J\sigma_u^2$. $\boldsymbol{w}_t$ is a vector of size 4, with the first element representing the error term in equation (3) and the other elements being constant and equal to zero (as β_3, β_4 and β_5 do not vary in time). That way, $\boldsymbol{W}$ is a 4×4 matrix with element $[1, 1]$ equal to σ_w^2 and the other elements are equal to zero.

Now we are ready to present a general formulation for the Dynamic Hierarchical

Models, as introduced by Gamerman and Migon (1993). The model is composed of three parts: an observation equation, structural equations and a system equation. The observation equation describes the distribution of the observations through a regression model (as in (4)); the structural equation describes the structure of hierarchy of the regression parameters (as in (5)); and the system equation describes the evolution of the parameters through time (as in (6)). For a linear hierarchical model of three levels, these equations can be written, respectively, as:

observation equation:

$$\boldsymbol{y}_t = \boldsymbol{X}_{1t}\boldsymbol{\beta}_{1,t} + \boldsymbol{v}_{1t}, \quad \boldsymbol{v}_{1t} \sim N_n(\boldsymbol{0}, \boldsymbol{V}_{1t}), \tag{7}$$

structural equation:

$$\boldsymbol{\beta}_{1,t} = \boldsymbol{X}_{2t}\boldsymbol{\beta}_{2,t} + \boldsymbol{v}_{2t}, \quad \boldsymbol{v}_{2t} \sim N_{r_1}(\boldsymbol{0}, \boldsymbol{V}_{2t}), \tag{8}$$

$$\boldsymbol{\beta}_{2,t} = \boldsymbol{X}_{3t}\boldsymbol{\beta}_{3,t} + \boldsymbol{v}_{3t}, \quad \boldsymbol{v}_{3t} \sim N_{r_2}(\boldsymbol{0}, \boldsymbol{V}_{3t}), \tag{9}$$

system equation:

$$\boldsymbol{\beta}_{3,t} = \boldsymbol{G}_t\boldsymbol{\beta}_{3,t-1} + \boldsymbol{w}_t, \quad \boldsymbol{w}_t \sim N_{r_3}(\boldsymbol{0}, \boldsymbol{W}_t), \tag{10}$$

where n is the total number of observations, $\boldsymbol{v}_{1t}$, $\boldsymbol{v}_{2t}$, $\boldsymbol{v}_{3t}$ and $\boldsymbol{w}_t$ are disturbance terms which are independent; $\boldsymbol{X}_{1t}, \boldsymbol{X}_{2t}, \boldsymbol{X}_{3t}$ and $\boldsymbol{G}_t$ are known matrices possibly incorporating explanatory variables; $\boldsymbol{V}_{1t}, \boldsymbol{V}_{2t}, \boldsymbol{V}_{3t}$ and $\boldsymbol{W}_t$ are variance-covariance matrices that can be allowed to vary over time; $\boldsymbol{\beta}_i$ is a vector of coefficients of size $r_i, i = 1, \cdots, 3$, with $r_1 > r_2 > r_3$.

Hierarchical models of higher levels can be easily obtained adding extra levels in the structural equations. The two levels model can be obtained setting $\boldsymbol{X}_{3t}$ to be the identity matrix, and $\boldsymbol{V}_{3t}$ to be a matrix of null elements.

Other simple hypothetical examples are presented bellow to illustrate the use of these

models.

Example 2: weight measurements in a population of patients under treatment

Suppose that the variation of weight in a population of patients under the same kind of experimental treatment is being investigated. Since the beginning of the experiment, a different sample of patients is selected every week from the population and weighted, for a total of T weeks. Suppose that at the t^{th} week, a sample of size n_t is selected. To model the variation of weight through time, we can use a simple two-stage model, given by:

observation equation:

$$y_{it} = \beta_{i,t} + \epsilon_{it}, \quad \epsilon_{it} \sim N(0, \sigma^2), \quad i = 1, \cdots, n_t,$$

structural equations:

$$\beta_{i,t} = \mu_t + u_{it}, \quad u_{it} \sim N(0, \sigma_u^2), \quad i = 1, \cdots, n_t,$$

system equations:

$$\mu_t = \mu_{t-1} + w_t, \quad w_t \sim N(0, \sigma_w^2).$$

This model is a collection of the steady models of West and Harrison (1997), which are related through similar mean levels. The $\beta_{i,t}s$ are the observation levels assumed to form an exchangeable (with respect to index i) sample of means with common mean μ_t. Note that the mean μ_t is allowed to vary over time - the treatment can cause average weight loss or weight gain through time. This model can be written in the matrix notation as in (7)-(10):

$$\boldsymbol{y}_t = \boldsymbol{X}_{1t}\boldsymbol{\beta}_{1,t} + \boldsymbol{v}_{1t}, \quad \boldsymbol{v}_{1t} \sim N_{n_t}(\boldsymbol{0}, \boldsymbol{V}_{1t}), \tag{11}$$

$$\boldsymbol{\beta}_{1,t} = \boldsymbol{X}_{2t}\beta_{2,t} + \boldsymbol{v}_{2t}, \quad \boldsymbol{v}_{2t} \sim N_{n_t}(\boldsymbol{0}, \boldsymbol{V}_{2t}), \tag{12}$$

$$\beta_{2,t} = G_t\beta_{2,t-1} + w_t, \quad w_t \sim N(0, W_t), \tag{13}$$

where $\boldsymbol{y}_t = (y_{1t}, \cdots, y_{n_t t})^T$, $\boldsymbol{\beta}_{1,t} = (\beta_{1,t}, \cdots, \beta_{n_t,t})^T$, $\boldsymbol{X}_{1t} = \boldsymbol{I}_{n_t}$, $\boldsymbol{V}_{1t} = \boldsymbol{I}_{n_t}\sigma^2, \boldsymbol{X}_{2t} = \mathbf{1}_{n_t}, \beta_{2,t} = \mu_t, \boldsymbol{V}_{2t} = \boldsymbol{I}_{n_t}\sigma_u^2$, $G_t = 1$ and $W_t = \sigma_w^2$. $\mathbf{1}_n$ represents a n-dimensional vector of ones.

Example 3: weight measurements in a population of children with malnutrition

The second example can be illustrated by the following experiment: A population of children with malnutrition is being treated with a caloric diet. As in *example 2*, a different sample (of size n_t) of children is selected from the population at the t^{th} week of the experiment and weighted. Differently from the previous example, however,the children are expected to gain weight through time. The model proposed in *example 1* can therefore be modified to accommodate this expected growth in the average weight. The proposed model for this example can be written as:

observation equation:

$$y_{it} = \beta_{i,t} + \epsilon_t, \quad \epsilon_t \sim N(0, \sigma^2), \quad i = 1, \cdots, n_t,$$

structural equations:

$$\beta_{i,t} = \mu_t + u_{it}, \quad u_{it} \sim N(0, \sigma_u^2), \quad i = 1, \cdots, n_t,$$

system equations:

$$\mu_t = \rho_1(\mu_{t-1} + \delta_{t-1}) + w_{1t}, \quad w_{1t} \sim N(0, \sigma_{w_1}^2),$$

$$\delta_t = \rho_2\delta_{t-1} + w_{2t}, \quad w_{2t} \sim N(0, \sigma_{w_2}^2).$$

Note that the expected growth of the mean μ_t is specified by the system equations. This model can be represented in the matrix notation as in (11)-(13), where $\boldsymbol{y}_t$ and $\boldsymbol{\beta}_{1,t}$ are defined as before and $\boldsymbol{X}_{1t} = \boldsymbol{I}_{n_t}, \boldsymbol{V}_{1t} = \boldsymbol{I}_{n_t}\sigma^2, \boldsymbol{\beta}_{2,t} = (\mu_t, \delta_t)^T, \boldsymbol{X}_{2t} = (\mathbf{1}_{n_t}, \mathbf{0}_{n_t}), V_{2t} = \boldsymbol{I}_{n_t}\sigma_u^2$, $\boldsymbol{G}_t = \begin{pmatrix} \rho_1 & \rho_1 \\ 0 & \rho_2 \end{pmatrix}$ and $\boldsymbol{W}_t = \begin{pmatrix} \sigma_{w_1}^2 & 0 \\ 0 & \sigma_{w_2}^2 \end{pmatrix}$. $\mathbf{0}_n$ denotes a n-dimensional vector of zeros.

3 Inference for Dynamic Hierarchical Models

Inference will be presented here from the Bayesian point of view. To do so, it will be necessary to specify prior distributions for all the unknown parameters of the model. More details about the Bayesian approach can be seen in Chapter 4. Our aim for this kind of modeling is usually to obtain the posterior distribution of these parameters and perform forecasting for future observations. In this section, some basic results which were presented in Gamerman and Migon (1993) will be reviewed.

Let us define D_t as all the information obtained up to time t, including the observations $\boldsymbol{y} = \{\boldsymbol{y}_1, \cdots, \boldsymbol{y}_t\}$ and the prior information, represented by D_0. We assume therefore that $D_t = \{\boldsymbol{y}_t, D_{t-1}\}$. We can represent the (k-stage) dynamic hierarchical model as:

$$\boldsymbol{y}_t|\boldsymbol{\beta}_{1,t} \sim N_n(\boldsymbol{X}_{1t}\boldsymbol{\beta}_{1,t}, \boldsymbol{V}_{1t}), \tag{14}$$

$$\boldsymbol{\beta}_{i,t}|\boldsymbol{\beta}_{i+1,t} \sim N_{r_i}(\boldsymbol{X}_{i+1,t}\boldsymbol{\beta}_{i+1}, \boldsymbol{V}_{i+1,t}), \quad i = 1, \cdots, k-1, \tag{15}$$

$$\boldsymbol{\beta}_{k,t}|\boldsymbol{\beta}_{k,t-1} \sim N_{r_k}(\boldsymbol{G}_t\boldsymbol{\beta}_{k,t-1}, \boldsymbol{W}_t), \tag{16}$$

with initial prior $\boldsymbol{\beta}_{k,0}|D_0 \sim N(\boldsymbol{m}_{k,0}, \boldsymbol{C}_{k,0})$. Note that the matrices $\boldsymbol{X}_{it}, i = 1, \cdots, k-1$ and $\boldsymbol{G}_t$ are assumed known.

3.1 Updating and forecasting: known variances case

At this point, let us also assume that the variances $\boldsymbol{V}_{it}$ are known. Even though it is in most cases an unrealistic assumption, it enables us to obtain the basic updating and forecasting operations. Under that assumption, Gamerman and Migon (1993) showed that the model specified by equations (14)-(16), considering the prior $\boldsymbol{\beta}_{k,0}|D_0 \sim N(\boldsymbol{m}_{k,0}, \boldsymbol{C}_{k,0})$, gives the following prior, posterior and predictive distributions:

- the prior distribution at time t is given by

$$\boldsymbol{\beta}_{i,t}|D_{t-1} \sim N(\boldsymbol{a}_{it}, \boldsymbol{R}_{it}), \quad i = 1, \cdots, k, \tag{17}$$

where $\boldsymbol{a}_{it} = \boldsymbol{X}_{i+1,t}\boldsymbol{a}_{i+1,t}, \boldsymbol{R}_{it} = \boldsymbol{X}_{i+1,t}\boldsymbol{R}_{i+1,t}\boldsymbol{X}_{i+1,t}^T + \boldsymbol{V}_{i+1,t}, i = 1, \cdots, k+1$, and $\boldsymbol{a}_{kt} = \boldsymbol{G}_t\boldsymbol{m}_{k,t-1}$ and $\boldsymbol{R}_{kt} = \boldsymbol{G}_t\boldsymbol{C}_{k,t-1}\boldsymbol{G}_t^T + \boldsymbol{W}_t$;

- the predictive distribution one step-ahead is given by

$$\boldsymbol{y}_t|D_{t-1} \sim N(\boldsymbol{f}_t, \boldsymbol{Q}_t), \tag{18}$$

where $\boldsymbol{f}_t = \boldsymbol{X}_{1t}\boldsymbol{a}_{1t}$ and $\boldsymbol{Q}_t = \boldsymbol{X}_{1t}\boldsymbol{R}_{1t}\boldsymbol{X}_{1t}^T + \boldsymbol{V}_{1t}$;

- the posterior distribution at time t is given by

$$\boldsymbol{\beta}_{i,t}|D_t \sim N(\boldsymbol{m}_{it}, \boldsymbol{C}_{it}), \quad i = 1, \cdots, k, \tag{19}$$

where $\boldsymbol{m}_{it} = \boldsymbol{a}_{it} + \boldsymbol{S}_{it}\boldsymbol{Q}_t^{-1}(\boldsymbol{y}_t - \boldsymbol{f}_t), \boldsymbol{C}_{it} = \boldsymbol{R}_{it} - \boldsymbol{S}_{it}\boldsymbol{Q}_t^{-1}\boldsymbol{S}_{it}^T$, $\boldsymbol{S}_{it} = \boldsymbol{R}_{it}\boldsymbol{E}_{0it}^T$, $\boldsymbol{E}_{ijt} = \prod_{l=i+1}^{j}\boldsymbol{X}_{lt}, 0 \leq i < j \leq k$, and $\boldsymbol{E}_{iit} = \boldsymbol{I}_{r_i}$ for $1 \leq i < k$.

Gamerman and Migon (1993) also show that h-steps-ahead forecasts, $h > 1$, can be easily obtained from this result. Suppose we are interested in predicting $\boldsymbol{y}_{t+h}$ given D_t, or in other words, we want to obtain the distribution of $\boldsymbol{y}_{t+h}|D_t$. Note that

$$\begin{aligned} \boldsymbol{y}_{t+h}|\boldsymbol{\beta}_{1,t+h} &\sim N(\boldsymbol{X}_{1,t+h}\boldsymbol{\beta}_{1,t+h}, \boldsymbol{V}_{1,t+h}), \text{ and} \\ \boldsymbol{\beta}_{k,t}|D_t &\sim N(\boldsymbol{m}_{kt}, \boldsymbol{C}_{kt}). \end{aligned}$$

Then, the distribution of $\boldsymbol{\beta}_{k,t+h}|D_t \sim N(\boldsymbol{a}_{kt}(h), \boldsymbol{R}_{kt}(h))$ can be recursively obtained (West and Harrison, 1997) with $\boldsymbol{a}_{kt}(h) = \boldsymbol{G}_{t+h}\boldsymbol{a}_{kt}(h-1)$ and $\boldsymbol{R}_{kt}(h) = \boldsymbol{G}_{t+h}\boldsymbol{R}_{kt}(h-1)\boldsymbol{G}_{t+h}^T + \boldsymbol{W}_{t+h}$ with starting values $\boldsymbol{a}_{kt}(0) = \boldsymbol{m}_{kt}$ and $\boldsymbol{R}_{kt}(0) = \boldsymbol{C}_{kt}$. Successive integrations give:

$$\begin{aligned} \boldsymbol{\beta}_{i,t+1}|D_t &\sim N(\boldsymbol{a}_{it}(h), \boldsymbol{R}_{it}(h)), \qquad i = 1, \cdots, k-1 \\ \boldsymbol{y}_{t+h}|D_t &\sim N(\boldsymbol{f}_t(h), \boldsymbol{Q}_t(h)), \end{aligned}$$

where $\boldsymbol{f}_t(h)$ and $\boldsymbol{Q}_t(h)$ are analogous to $\boldsymbol{f}_t$ and $\boldsymbol{Q}_t$ as defined previously, substituting $\boldsymbol{a}_{it}$ and $\boldsymbol{R}_{it}$ by $\boldsymbol{a}_{it}(h)$ and $\boldsymbol{R}_{it}(h)$.

When a more accurate description of the model parameters is required, or if events in the past are still of interest, a procedure called *smoothing* or *filtering* can be applied at any given time point. This procedure filters back the information smoothly via a recursive algorithm, allowing for parametric estimation of the system at previous times given all available information at a given period of time. Thus, the smoothed distributions of state parameters are denoted by $[\boldsymbol{\beta}_{i,t}|D_n]$, as opposed to their online posterior distributions, denoted by $[\boldsymbol{\beta}_{i,t}|D_t]$. Gamerman and Migon (1993) show that the smoothed distribution for a dynamic hierarchical model is $\boldsymbol{\beta}_{i,t}|D_n \sim N(\boldsymbol{m}_{it}^n, \boldsymbol{C}_{it}^n), t = 1, \cdots, n, i = 1, \cdots, k$, with moments recursively defined as

$$\boldsymbol{m}_{it}^n = \boldsymbol{m}_{it} + \boldsymbol{A}_{it}^T \boldsymbol{R}_{i,t+1}^{-1}(\boldsymbol{m}_{i,t+1}^n - \boldsymbol{a}_{i,t+1}),$$

$$\boldsymbol{C}_{it}^n = \boldsymbol{C}_{it} - \boldsymbol{A}_{it}^T \boldsymbol{R}_{i,t+1}^{-1}(\boldsymbol{R}_{i,t+1} - \boldsymbol{C}_{i,t+1}^n)\boldsymbol{R}_{i,t+1}^{-1}\boldsymbol{A}_{it},$$

and initialized at $t = n$ with $\boldsymbol{m}_{in}^n = \boldsymbol{m}_{in}$ and $\boldsymbol{C}_{in}^n = \boldsymbol{C}_{in}$, where $\boldsymbol{A}_{it} = \boldsymbol{E}_{ik,t+1}\boldsymbol{G}_{t+1}\boldsymbol{C}_{kt}\{(\boldsymbol{I}_{r_i} - \boldsymbol{V}_{ikt}^* \boldsymbol{E}_{0it}^T \boldsymbol{V}_{0kt}^{*-1} \boldsymbol{E}_{0it})\boldsymbol{E}_{ikt}\}^T$, and $\boldsymbol{V}_{ijt}^* = \sum_{l=i+1}^{j} \boldsymbol{E}_{i,l-1,t}\boldsymbol{V}_{lt}\boldsymbol{E}_{i,l-1,t}^T$.

3.2 Updating and forecasting: unknown variances case

As stated before, the hypothesis that the variances in the model are known is hardly ever realistic in a real application. In many applications, however, is it reasonable to suppose independence of the errors in the observation equation, such that $\boldsymbol{V}_{it} = \sigma_{it}^2 \boldsymbol{I}_N$, where N is the dimension of $\boldsymbol{y}_t$. A conjugate analysis is possible when all the variances $\boldsymbol{V}_{it}$, $\boldsymbol{W}_t$ and $\boldsymbol{C}_{k0}$ are scaled by σ^2, an unknown factor, with an Inverted Gamma prior distribution: $\sigma^2|D_0 \sim IG(n_0/2, d_0/2)$. The model can be written as in (14)-(16) with the variances multiplied by the factor σ^2.

The distribution of σ^2 can be updated at time t by $IG(n_t/2, d_t/2)$ where $n_t = n_{t-1}+n$ and $d_t = d_{t-1} + (\boldsymbol{y}_t - \boldsymbol{f}_t)^T \boldsymbol{Q}_t^{-1}(\boldsymbol{y}_t - \boldsymbol{f}_t)$. The results presented by equations (13), (14) and (15) remain valid except that all variances should be multiplied by σ^2. The predictive distribution (one-step ahead) in this case (after integrating σ^2 out) is obtained by replacing the Normal distributions by a Student T distributions with n_t degrees of freedom and substituting σ^2 by its estimate d_t/n_t. The posterior distribution is also obtained by substituting the Normal by a Student T distribution, but in this case with n_{t-1} degrees of freedom, and the estimate for σ^2 is given by d_{t-1}/n_{t-1}. We denote the Student T distribution with mean $\boldsymbol{m}$, ν degrees of freedom and variance-covariance matrix $\boldsymbol{\Sigma}$ by $T(\boldsymbol{m}, \nu, \boldsymbol{\Sigma})$.

In a more realistic approach, however, just one unknown factor σ^2 is not enough to handle all of the uncertainty at every level of the hierarchy. In general, a more realistic representation considers at least one unknown parameter for the variance of the errors of each hierarchical level - with the single parameter being the constant variance under the assumption of independent errors. Models of this kind are not analytically tractable and therefore it is necessary to make use of computational methods to simulate from the distributions of interest (posterior and/or predictive), or make use of deterministic approximation methods.

Numerical methods are a very important tool for applied Bayesian analysis. When the posterior (or predictive) distribution of the unknown parameters is not analytically tractable, samples can hardly ever be obtained directly from this distribution, and as an alternative, simulation methods that sample indirectly from the posterior distribution can be applied. One of the most used methods to obtain these samples is the Markov Chain Monte Carlo (MCMC), a computer-intensive method where the idea is to sample from

a Markov chain whose equilibrium distribution is the posterior/predictive distribution of interest. That way, after convergence is reached, the simulated values can be seen as a sample of the distribution of interest, and inference can be performed based on this sample. Different algorithms can be proposed in MCMC, with Gibbs Sampling and Metropolis-Hastings being the most used. A broad discussion about these and others algorithms can be found in Gamerman and Lopes (2006).

More recently, a promising alternative to inference via MCMC in latent Gaussian models - the Integrated Nested Laplace Approximations (INLA) - was proposed by Rue et. al. (2009). INLA is an approximation method for the posterior/predictive distributions of interest, that unlike the empirical Bayes approaches (Fahrmeir *et al.* 2004), incorporates posterior uncertainty with respect to hyperparameters. Other alternative methods that have been successfully used for inference in problems of this kind are the Particle Filters, which are sequential Monte Carlo methods based on point masses (or "particles") representations of probability densities, and are very attractive computationally. Examples of the use of these methods can be found in Liu and West (2001), Johannes and Polson (2007) and Carvalho *et. al.* (2010), amongst others. Both these methods (INLA and Particle Filters) can be applied to the Hierarchical Dynamic Models presented here and extensions of these models that will be presented in chapter 5.

4 Model Extensions

In this section, extensions to the Dynamic Hierarchical Models will be presented, including examples illustrating applications to real data-sets. In Section 4.1, the Matrix-variate Dynamic Hierarchical Models will be presented as a multivariate extension to the univariate case already introduced. In Section 4.2, a particular case of this model will be

considered, imposing a parametric structure to account for the spatial correlation between observations made in different locations in space. In Section 4.3, an extension to observations in the exponential family will be presented.

4.1 Matrix-variate Dynamic Hierarchical Models

Suppose q observations are made through time in r different locations in space such that $\boldsymbol{y}_{1t}, \cdots, \boldsymbol{y}_{qt}$ represent q vectors of observations of dimension $r \times 1$. Data with this structure can be easily found in the literature for the study of environmental processes. As an example, in many locations around the world, monitoring stations are implemented to monitor the air quality, measuring concentrations of all kinds of pollutants at a certain periodicity. q observations of certain pollutants made through time, and registered at r different locations, can be seen as q vectors of observations of dimension $r \times 1$. In this section, the Hierarchical Dynamic Models for univariate responses will be extended to responses of this kind, leading to the Matrix-variate Hierarchical Dynamic Models. To introduce the notation, Section 4.1.1 presents the Matrix-variate Dynamic models. The Matrix-variate Hierarchical Dynamic Models are then presented in Section 4.1.2 as a simple extension. Inference to these models is presented in Section 4.1.3, and finally Section 4.1.4 presents an application.

4.1.1 Matrix-variate Dynamic Models

Let $\boldsymbol{y}_{1t}, \cdots, \boldsymbol{y}_{qt}$ represent q vectors of observations of dimension $r \times 1$, with the i^{th} element of each vector representing an observation made at location i. If we denote

$$\boldsymbol{y}_t = [\boldsymbol{y}_{1t}, \cdots, \boldsymbol{y}_{qt}],$$

we have a matrix of observations through time. West and Harrison (1989) define a Matrix-variate Dynamic Model by the equations:

$$\begin{aligned} \boldsymbol{y}_t &= \boldsymbol{X}_t\boldsymbol{\beta}_t + \boldsymbol{v}_t, \\ \boldsymbol{\beta}_t &= \boldsymbol{G}_t\boldsymbol{\beta}_{t-1} + \boldsymbol{w}_t, \end{aligned}$$

where

- $\boldsymbol{y}_t$ is the $r \times q$ observations matrix in time t;
- $\boldsymbol{X}_t$ is a known matrix of regressors $r \times p$;
- $\boldsymbol{\beta}_t$ is the matrix of unknown parameters;
- v_t is a matrix of errors;
- $\boldsymbol{G}_t$ is a known evolution matrix $p \times p$;
- $\boldsymbol{w}_t$ is the matrix of evolutions of errors $p \times q$;

It is assumed that $\boldsymbol{\beta}_t$ and $\boldsymbol{v}_t$ are independent, $\boldsymbol{\beta}_{t-1}$ and $\boldsymbol{w}_t$ are independent, and $\boldsymbol{v}_t$ and $\boldsymbol{w}_t$ are independent. Under the hypotheses of normality, the errors $\boldsymbol{v}_t$ and $\boldsymbol{w}_t$, which are matrices, will follow a matrix-variate normal distribution. If $\boldsymbol{Z}$ is a random matrix of dimension $r \times q$, we say that $\boldsymbol{Z}$ has a matrix-variate normal distribution with right covariance matrix $\boldsymbol{\Sigma}$ and left covariance matrix $\boldsymbol{C}$, denoted by $\boldsymbol{Z} \sim N(\boldsymbol{M}, \boldsymbol{C}, \boldsymbol{\Sigma})$, if $vec(\boldsymbol{Z}) \sim N_{rq}(vec(\boldsymbol{M}), \boldsymbol{\Sigma} \otimes \boldsymbol{C})$, where $vec(\boldsymbol{Z})$ represents the vectorization of the matrix $\boldsymbol{Z}$ in columns.

As an example, suppose that q types of pollutants are observed in r different locations in space, resulting in a $r \times q$ matrix of observations, for a fixed period of time t. If this matrix follows a $N(\boldsymbol{M}, \boldsymbol{C}, \boldsymbol{\Sigma})$ distribution, then $\boldsymbol{M}$ is a $r \times q$ matrix of means, the left

matrix $\boldsymbol{C}$ is $r \times r$, and represents the correlation of the observations in space, and the right matrix $\boldsymbol{\Sigma}$ is $q \times q$ and represents the correlation between the different kinds of pollutants. Note that in this model, $\boldsymbol{C}$ and $\boldsymbol{\Sigma}$ are not identifiable, and $N(\boldsymbol{M}, \boldsymbol{C}, \boldsymbol{\Sigma})$ is equivalent to $N(\boldsymbol{M}, \boldsymbol{C}/k, k\boldsymbol{\Sigma}), k \in \Re$. If the observations of each variable of interest are not correlated in space at a given period of time t, then $\boldsymbol{C} = \boldsymbol{I}_r$. If the observations are independent at a fixed period of time and fixed location in space, then $\boldsymbol{\Sigma} = \boldsymbol{I}_q$.

4.1.2 Hierarchical Formulation of the Model

The Matrix-variate Hierarchical Dynamic Models are an extension of the model described above, and it is an extension of the Hierarchical Dynamic Models, where for each period of time, we observe a matrix of observations instead of a vector. This model can be written, as defined by Landim and Gamerman (2000), as follows:

$$\begin{aligned}
\boldsymbol{y}_t &= \boldsymbol{X}_{1t}\boldsymbol{\beta}_{1,t} + \boldsymbol{v}_{1t}, & \boldsymbol{v}_{1t}|\boldsymbol{\Sigma} \sim N(\boldsymbol{0}, \boldsymbol{V}_1, \boldsymbol{\Sigma}), & (20)\\
\boldsymbol{\beta}_{1,t} &= \boldsymbol{X}_{2t}\boldsymbol{\beta}_{2,t} + \boldsymbol{v}_{2t}, & \boldsymbol{v}_{2t}|\boldsymbol{\Sigma} \sim N(\boldsymbol{0}, \boldsymbol{V}_2, \boldsymbol{\Sigma}), & (21)\\
&\vdots & & \\
\boldsymbol{\beta}_{k-1,t} &= \boldsymbol{X}_{kt}\boldsymbol{\beta}_{k,t} + \boldsymbol{v}_{kt}, & \boldsymbol{v}_{kt}|\boldsymbol{\Sigma} \sim N(\boldsymbol{0}, \boldsymbol{V}_k, \boldsymbol{\Sigma}), & (22)\\
\boldsymbol{\beta}_{k,t} &= \boldsymbol{G}_t\boldsymbol{\beta}_{k,t-1} + \boldsymbol{w}_t, & \boldsymbol{w}_t|\boldsymbol{\Sigma} \sim N(\boldsymbol{0}, \boldsymbol{W}, \boldsymbol{\Sigma}), & (23)
\end{aligned}$$

where, for a fixed period of time t,

- $\boldsymbol{y}_t$ is the $r \times q$ observations matrix;
- $\boldsymbol{\beta}_{i,t}, i = 1, \cdots, k$ are regression coefficients matrixes of order $r_i \times q$;
- $\boldsymbol{X}_{it}$ are known matrices of regressors of order $r_{i-1} \times r_i$, with $r_0 = r$, $i = 1, \cdots, k$;
- $\boldsymbol{G}_t$ is the known evolution matrix $r_k \times r_k$;

- $\boldsymbol{w}_t$ is the matrix of evolutions of errors $p \times q$.

We assume that $\boldsymbol{\beta}_{i,t}$ and $\boldsymbol{v}_{it}$ are independent, $i = 1, \cdots, k$, $\boldsymbol{\beta}_{k,t-1}$ and $\boldsymbol{w}_t$ are independent, and $\boldsymbol{v}_{1t}, \boldsymbol{v}_{2t}, \cdots, \boldsymbol{v}_{kt}$ and $\boldsymbol{w}_t$ are independent. Matrices $\boldsymbol{V}_1, \boldsymbol{V}_2, \cdots, \boldsymbol{V}_k$ and $\boldsymbol{W}$ are considered to be known, and for simplicity, are considered to be fixed over time.

Note that a restriction is made, stating that part of the conditional covariance between elements of $\boldsymbol{y}_t$, which account for the different responses, is specified in $\boldsymbol{\Sigma}$. The same restriction is imposed to the conditional covariances of $\boldsymbol{\beta}_{1,t}, \boldsymbol{\beta}_{2,t}, \cdots, \boldsymbol{\beta}_{k,t}$, with all sharing the same structure $\boldsymbol{\Sigma}$ to account for the covariance between elements coming from different responses. This assumption is usually made in multivariate dynamic models (see, for example, West and Harrison (1997, ch. 16) and Quintana (1987)), and it is reasonable since the magnitude of the covariance between elements at different hierarchical levels are not imposed to be the same. The models above could be specified and operated without this restriction but that would substantially increase parameter dimensionality.

To complete the specification of the model presented above, it is necessary to specify prior distributions for the unknown parameters of the model. Usual choices of priors, used by West and Harrison (1997), Quintana (1987) and Landim and Gamerman (2000), are the following:

$$\begin{aligned} \boldsymbol{\beta}_{k,0}|\boldsymbol{\Sigma}, D_0 &\sim N(\boldsymbol{M}_{k,0}, \boldsymbol{C}_{k,0}, \boldsymbol{\Sigma}), \text{ and} \\ \boldsymbol{\Sigma}|D_0 &\sim WI(n_0, \boldsymbol{S}_0). \end{aligned}$$

4.1.3 Inference

As in the beginning of Section 3, we will start this section presenting results of inference based on the assumption that $\boldsymbol{V}_1, \boldsymbol{V}_2, \cdots, \boldsymbol{V}_k$ and $\boldsymbol{W}$ are known. Conditionally on the value of $\boldsymbol{\Sigma}$, it is possible to obtain the distributions of $\boldsymbol{\beta}_{i,t}|D_{t-1}$, $\boldsymbol{y}_t|D_{t-1}$ and $\boldsymbol{\beta}_{i,t}|D_t$

analytically. These results are given below, and the demonstration can be found in Landim and Gamerman (2000). For $t = 1, 2, \cdots$ and $i = 1, \cdots, k$,

a) $\boldsymbol{\beta}_{i,t}|D_{t-1} \sim N(\boldsymbol{a}_{it}, \boldsymbol{R}_{it}, \boldsymbol{\Sigma})$,

$\boldsymbol{a}_{it} = \boldsymbol{X}_{i+1}\boldsymbol{a}_{i+1,t}, \quad i = 1, \cdots, k-1, \quad \boldsymbol{a}_{kt} = \boldsymbol{G}_t\boldsymbol{M}_{k,t-1},$

$\boldsymbol{R}_{it} = \boldsymbol{X}_{i+1,t}\boldsymbol{R}_{i+1,t}\boldsymbol{X}_{i+1,t}^T + \boldsymbol{V}_{i+1,t}, \quad 1 = 1, \cdots, k, \quad \boldsymbol{R}_{kt} = \boldsymbol{G}_t\boldsymbol{C}_{k,t-1}\boldsymbol{G}_t^T + \boldsymbol{W}_t;$

b) $\boldsymbol{y}_t|D_{t-1} \sim N(\boldsymbol{f}_t, \boldsymbol{Q}_t, \boldsymbol{\Sigma})$,

$\boldsymbol{f}_t = \boldsymbol{X}_{1t}\boldsymbol{a}_{1t}, \quad \boldsymbol{Q}_t = \boldsymbol{X}_{1t}\boldsymbol{R}_{1t}\boldsymbol{F}_{1t}^T + \boldsymbol{V}_{1t};$

c) $\boldsymbol{\theta}_{i,t}|D_t \sim N(\boldsymbol{M}_{it}, \boldsymbol{C}_{it}, \boldsymbol{\Sigma})$,

$\boldsymbol{M}_{it} = \boldsymbol{a}_{it} + \boldsymbol{S}_{it}\boldsymbol{Q}_t^{-1}\boldsymbol{E}_t, \quad \boldsymbol{S}_{it} = \boldsymbol{R}_{it}\boldsymbol{E}_0\boldsymbol{R}_{it}^T,$

$\boldsymbol{E}_t = \boldsymbol{y}_t - \boldsymbol{f}_t,$

$\boldsymbol{C}_{it} = \boldsymbol{R}_{it} - \boldsymbol{S}_{it}\boldsymbol{Q}_t^{-1}\boldsymbol{S}_{it}^T.$

The posterior distribution of $\boldsymbol{\Sigma}$ can also be obtained analytically, and it is given by:

$$\boldsymbol{\Sigma}|D_t \sim WI(n_t, \boldsymbol{S}_t),$$

with $n_t = n_{t-1} + N$ and $\boldsymbol{S}_t = \boldsymbol{S}_{t-1} + \boldsymbol{E}_t^T\boldsymbol{Q}_t^{-1}\boldsymbol{E}_t$. Unconditionally on $\boldsymbol{\Sigma}$, the distributions of $\boldsymbol{\Theta}_{it}|D_{t-1}$, $\boldsymbol{y}_t|D_{t-1}$ and $\boldsymbol{\beta}_{i,t}|D_t$ become Student T, given by:

a) $\boldsymbol{\beta}_{i,t}|D_{t-1} \sim T(\boldsymbol{a}_{it}, \boldsymbol{R}_{it}, n_t, \boldsymbol{S}_t)$,

b) $\boldsymbol{y}_t|D_{t-1} \sim T(\boldsymbol{f}_t, \boldsymbol{Q}_t, n_t, \boldsymbol{S}_t)$,

c) $\boldsymbol{\beta}_{i,t}|D_t \sim T(\boldsymbol{M}_{it}, \boldsymbol{C}_{it}, n_t, \boldsymbol{S}_t)$,

where $T(\boldsymbol{a}, \boldsymbol{R}, n, \boldsymbol{S})$ denote the Matrix Variate Student T distribution with mean $\boldsymbol{a}$, left side covariance matrix $\boldsymbol{R}$, n degrees of freedom and right side covariance matrix $\boldsymbol{S}$. If $\boldsymbol{Z}$ is a random matrix such that $\boldsymbol{Z} \sim T(\boldsymbol{a}, \boldsymbol{R}, n, \boldsymbol{S})$ then $vec(\boldsymbol{X}) \sim T(vec(\boldsymbol{a}), n, \boldsymbol{R} \otimes \boldsymbol{S})$.

Equivalently to the case of Univariate Hierarchical Dynamic Linear Models, forecasts h-steps-ahead, or the distribution of $(\boldsymbol{y}_{t+h}|D_t, \boldsymbol{\Sigma})$, can be obtained.

When the variance matrices $\boldsymbol{V}_1, \boldsymbol{V}_2, \cdots, \boldsymbol{V}_k$, $\boldsymbol{W}$ and $\boldsymbol{\Sigma}$ are unknown, it is not possible to obtain the distributions of $\boldsymbol{\beta}_{i,t}|D_{t-1}, \boldsymbol{y}_t|D_{t-1}$ and $\boldsymbol{\beta}_{i,t}|D_t$ analytically. In that case, numerical methods can help obtaining approximations for these distributions.

Under the Bayesian point of view, prior distributions must be specified for these quantities. Landim and Gamerman (2000) considered independent Inverted Wishart prior distributions for all of the variance-covariance matrices, and a Normal prior distribution for $\boldsymbol{\beta}_{k,0}|\boldsymbol{\Sigma}, D_0$, given by $\boldsymbol{\beta}_{k,0}|\boldsymbol{\Sigma}, D_0 \sim N(\boldsymbol{M}_{k0}, \boldsymbol{C}_{k0}, \boldsymbol{\Sigma})$.

The joint posterior distribution of $(\{\boldsymbol{\beta}_1\}, \cdots, \{\boldsymbol{\beta}_k\}, \boldsymbol{V}_1, \cdots, \boldsymbol{V}_k, \boldsymbol{W}, \boldsymbol{\Sigma})$ is given by

$$
\begin{aligned}
& p(\{\boldsymbol{\beta}_1\}, \cdots, \{\boldsymbol{\beta}_k\}, \boldsymbol{V}_1, \cdots, \boldsymbol{V}_k, \boldsymbol{W}, \boldsymbol{\Sigma}|\boldsymbol{y}) \propto \prod_{t=1}^{T} p(\boldsymbol{y}_t|\boldsymbol{\beta}_{1,t}, \boldsymbol{V}_1, \boldsymbol{\Sigma}) \\
\times \quad & p(\boldsymbol{W})p(\boldsymbol{\Sigma})p(\boldsymbol{\beta}_{k,0})|\boldsymbol{\Sigma}) \prod_{i=1}^{k} p(\boldsymbol{V}_i) \prod_{t=1}^{T} \prod_{i=1}^{k-1} p(\boldsymbol{\beta}_{i,t}|\boldsymbol{\beta}_{i+1,t}, \boldsymbol{V}_{i+1}, \boldsymbol{\Sigma}) \prod_{t=2}^{T} p(\boldsymbol{\beta}_{k,t}|\boldsymbol{\beta}_{k,t-1}, \boldsymbol{W}, \boldsymbol{\Sigma}).
\end{aligned}
$$

Even though it is not possible to obtain the posterior distribution of the model parameters analytically, it is possible to obtain their full conditional distributions. Landim and Gamerman (2000) showed that the full conditional distribution for each of the parameters $\boldsymbol{\beta}_{i,t}, t = 1, \cdots, n$, $i = 1, \cdots, k$ is Normal and the full conditional distribution for the variance parameters $\boldsymbol{V}_1, \boldsymbol{V}_2, \cdots, \boldsymbol{V}_k, \boldsymbol{W}$ and $\boldsymbol{\Sigma}$ is Inverted Wishart. That is, it is possible to obtain samples from the posterior distribution of these parameters using Gibbs Sampling, which is an MCMC algorithm. Details of how to obtain the full conditional distributions and the proposed algorithm can be seen in Landim and Gamerman

(2000).

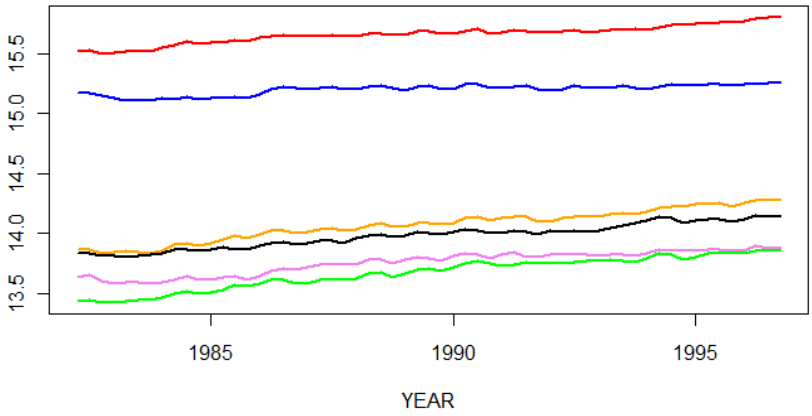
4.1.4 Application to the occupied population and average salary in Brazil

Landim and Gamerman (2000) presented an application to jointly model the population at work and the average salary for the biggest Metropolitan regions in Brazil: Rio de Janeiro (RJ), Salvador (SAL), São Paulo (SP), Belo Horizonte (BH), Porto Alegre (PA) and Recife (RE). The data was made available by Instituto de Pesquisa Econômica Apliacada (IPEA), in Brazil.

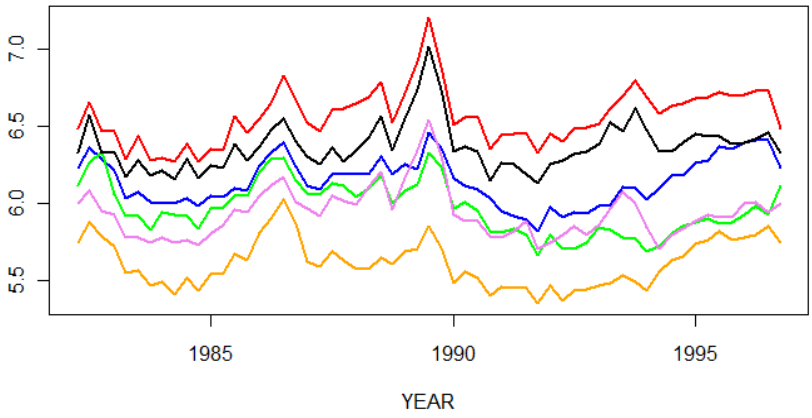
Observations were monthly made from May, 1982 to December, 1996. Preliminary analysis showed a seasonal pattern for both variables in time. Landim and Gamerman (2000) worked with the series aggregated at every three months, having a total of $T = 56$ periods of time of observation.

The original data was transformed via a logarithm transformation which made the observations closer to those coming from a Normal distribution. Also, the series of salaries was transformed using the INPC index, which is a National Index of prices for Consumer, using December 1996 as a base. That way, the salaries were all in the same scale of reais ($R\$$) of December 1996. Figure 1 shows the aggregated series of the logarithm of people at work and average salary for the six Metropolitan regions, showing that both variables change significantly and smoothly in time. Note that this example concerns two response variables which are varying not only in time but also in space. Preliminary analysis showed that it is reasonable to consider that the observations follow an autoregressive model in time and that the coefficients of this model vary in time and space.

Let $\boldsymbol{y}_t$ be a 6×2 matrix with element $[i, 1]$ representing the logarithm of the population size in the i^{th} region, and $[i, 2]$ representing the logarithm of the average salary in the i^{th}



(a)



(b)

Figure 1: (a) logarithm of people at work; and (b) logarithm of average salary, for each one of the six regions through time.

region for a fixed period of time t. That way, the response variable in the model can be represented by

$$\boldsymbol{y}_t = \begin{bmatrix} y_{11t} & y_{12t} \\ \vdots & \\ y_{61t} & y_{62t} \end{bmatrix}.$$

Landim and Gamerman (2000) suggest a dynamic autoregressive model of order 1 to model this response matrix, where the observation equation for the element $[i, j]$ of $\boldsymbol{y}_t$ is given by:

$$y_{ijt} = \beta_{1,ijt} + \beta_{2,ijt} y_{i1,t-1} + \beta_{3,ijt} y_{i2,t-1} + \epsilon_{ijt},$$

$i = 1, \cdots, 6, j = 1, 2$ and $t = 1, \cdots, 56$. In matrix notation, we can write:

$$\boldsymbol{y}_t = \boldsymbol{X}_{1t}\boldsymbol{\beta}_{1,t} + \boldsymbol{v}_{1t}, \qquad \boldsymbol{v}_{1t}|\boldsymbol{\Sigma} \sim N(\boldsymbol{0}, \boldsymbol{V}_1, \boldsymbol{\Sigma}),$$

where $\boldsymbol{X}_{1t}$ is given by:

$$\boldsymbol{X}_{1t} = \begin{bmatrix} 1 & y_{11,t-1} & y_{12,t-1} & 0 & \cdots & 0 & 0 & 0 \\ & & & \vdots & & & & \\ 0 & 0 & 0 & 0 & \cdots & 1 & y_{61,t-1} & y_{62,t-1} \end{bmatrix},$$

and the matrix of autoregressive coefficients is given by:

$$\boldsymbol{\beta}_{1,t}^T = \begin{bmatrix} \beta_{1,11t} & \beta_{2,11t} & \beta_{3,11t} & \cdots & \beta_{1,61t} & \beta_{2,61t} & \beta_{3,61t} \\ \beta_{1,12t} & \beta_{2,12t} & \beta_{3,12t} & \cdots & \beta_{1,62t} & \beta_{2,62t} & \beta_{3,62t} \end{bmatrix}.$$

For each $j = 1, 2$, and fixed period of time t, the parameters β_1, β_2 and β_3 are supposed to have the same mean for every location i, as specified by the equations bellow:

$$\begin{aligned} \beta_{1,ijt} &= \beta_{1,jt} + v_{21,ijt}, \\ \beta_{2,ijt} &= \beta_{2,jt} + v_{22,ijt}, \\ \beta_{3,ijt} &= \beta_{3,jt} + v_{23,ijt}, \end{aligned}$$

where $i = 1, \cdots, 6$. That way, we can write the matrix $\boldsymbol{\beta}_{1,t}$ in the form:

$$\boldsymbol{\beta}_{1,t} = \boldsymbol{X}_{2t}\boldsymbol{\beta}_{2,t} + \boldsymbol{v}_{2t}, \qquad \boldsymbol{v}_{2t}|\boldsymbol{\Sigma} \sim N(\boldsymbol{0}, \boldsymbol{V}_2, \boldsymbol{\Sigma}),$$

where $\boldsymbol{X}_{2t}$ is 18×3 and it is given by

$$\boldsymbol{X}_{2t}^T = \begin{bmatrix} 1 & 0 & 0 & \cdots & 1 & 0 & 0 \\ 0 & 1 & 0 & \cdots & 0 & 1 & 0 \\ 0 & 0 & 1 & \cdots & 0 & 0 & 1 \end{bmatrix} = 1_6^T \otimes I_3,$$

the hyperparameters matrix is given by

$$\boldsymbol{\beta}_{2,t} = \begin{bmatrix} \beta_{1,1t} & \beta_{1,1t} \\ \beta_{2,1t} & \beta_{2,2t} \\ \beta_{3,1t} & \beta_{3,2t} \end{bmatrix},$$

and

$$\boldsymbol{v}_{2t}^T = \begin{bmatrix} v_{21,11t} & v_{22,11t} & v_{23,11t} & \cdots & v_{21,61t} & v_{22,61t} & v_{23,61t} \\ v_{21,12t} & v_{22,12t} & v_{23,12t} & \cdots & v_{21,62t} & v_{22,62t} & v_{23,62t} \end{bmatrix}.$$

The evolution matrix is given by:

$$\boldsymbol{\beta}_{2,t} = \boldsymbol{\beta}_{2,t-1} + \boldsymbol{w}_t, \qquad \boldsymbol{w}_t \sim N(\boldsymbol{0}, \boldsymbol{W}, \boldsymbol{\Sigma}).$$

To complete the model, Inverted Wishart distributions were set as the prior distributions for the parameters $\boldsymbol{V}_1$, $\boldsymbol{V}_2$, $\boldsymbol{W}$ and $\boldsymbol{\Sigma}$, and a Normal distribution was set as the prior distribution for $\boldsymbol{\beta}_{2,0}|\boldsymbol{\Sigma}$.

Landim and Gamerman (2000) used an MCMC algorithm to obtain samples from the posterior distribution of the unknown model parameters. After 59500 iterations, convergence was achieved and the last 5000 iterations were used as samples of the posterior distribution. They obtained credibility intervals for each unknown parameter of the model. One important finding was that the correlations in matrix $\boldsymbol{V}_1$ were not significantly different from zero, which indicates independence between observations made in different regions. Besides that, there is a strong indication that matrix $\boldsymbol{V}_2$ could be

written as a block diagonal matrix. These hypothesis were incorporated into the model, and it was re-estimated. Convergence for this case was faster. The posterior distributions of coefficients $\boldsymbol{\beta}_2$ showed that past observations of population size are a good explanatory variable for the present population size, but do not explain much of the average salary. The same way, past observations of average salary can explain part of the variability of the present average salary, but do not explain much of the population size. Matrix $\boldsymbol{\Sigma}$ showed small but positive correlation between the response variables.

4.2 Spatially Structured Matrix-variate Dynamic Hierarchical Models

In this section we describe a class of models proposed by Paez *et al.* (2008) which are a particular case of the Matrix-variate Dynamic Hierarchical Models of Landim and Gamerman (2000), but impose a parametric structure to account for the spatial correlation between observations made in different locations in space. With that restriction, the spatial correlation can be captured without the need of estimating completely unknown covariance matrices. A spatially structured matrix can be specified with a small number of parameters, and as a consequence, interpolation of the response variable can be made for any fixed period of time, to any set of non-observed locations in space.

Gelfand *et al.* (2005) use a similar approach to deal with this same kind of problem, where the temporal variance is described by dynamic components which capture the association between measurements made for a fixed location in space and period of time, as well as the dependence between space and time themselves. The multivariate space processes described in that paper were developed through the use of corregionalization techniques and non-separable space-time covariance structures. They applied

the proposed methodology to model the variation of monthly maximum temperature in the State of Colorado, U.S.A., using the monthly mean precipitation as an explanatory variable.

Other applications of multilevel modeling taking spatial structures into consideration can be seen in Chapter 33.

4.2.1 General model Framework

The model presented in this section is a special case of the model described by equations (20) to (23). Without loss of generality, the model will be described for a two-level hierarchical dynamic model. A simple extension can be made to obtain models of higher hierarchical levels. Suppose we observe q ($q > 1$) response variables in a discrete set of periods of time $t = 1, ..., T$ and set of locations $\{s_1, ..., s_r\}$ in a continuous space S. Analogously to equations (20) to (23), the matrix-variate hierarchical dynamic space-time model can be written as:

$$\boldsymbol{y}_t = \boldsymbol{X}_{1t}\boldsymbol{\beta}_{1,t} + \boldsymbol{v}_{1t}, \qquad \boldsymbol{v}_{1t} \sim N(\boldsymbol{0}, \boldsymbol{V}_1, \boldsymbol{\Sigma}), \tag{24}$$

$$\boldsymbol{\beta}_{1,t} = \boldsymbol{X}_{2t}\boldsymbol{\beta}_{2,t} + \boldsymbol{v}_{2t}, \qquad \boldsymbol{v}_{2t} \sim N(\boldsymbol{0}, \boldsymbol{V}_2, \boldsymbol{\Sigma}), \tag{25}$$

$$\boldsymbol{\beta}_{2,t} = \boldsymbol{G}_t\boldsymbol{\beta}_{2,t-1} + \boldsymbol{w}_t, \qquad \boldsymbol{w}_t \sim N(\boldsymbol{0}, \boldsymbol{W}, \boldsymbol{\Sigma}), \tag{26}$$

for $t = 1, ..., T$. We assume that the matrices $\boldsymbol{X}_{1t}$, $\boldsymbol{X}_{2t}$ and $\boldsymbol{G}_t$ are known, with $\boldsymbol{X}_{1t}$ and $\boldsymbol{X}_{2t}$ possibly incorporating the values of explanatory variables. The dimensions of the matrices in equation (24)-(26) are the same as specified previously for the equations (20)-(23).

A spatial structure can be incorporated in the variance matrices $\boldsymbol{V}_1$ and $\boldsymbol{V}_2$, being specified through parametric structures describing spatial dependency. In particular, one can define $\boldsymbol{V}_i = \boldsymbol{C} \otimes \boldsymbol{V}$, which corresponds to $\boldsymbol{v}_{it} \sim N(\boldsymbol{0}, \boldsymbol{V}_i\rho_i(\lambda_i, \cdot), \boldsymbol{\Sigma}_i)$, where $\boldsymbol{C}$ is the

matrix specified through the correlation function $\rho_i(\lambda_i, \cdot)$, $i = 1, 2$. For a fixed period of time t and given the parameter λ_i, the correlation function $\rho_i(\lambda_i, \cdot)$ gives the correlation between the elements of $\boldsymbol{v}_{it}, i = 1, 2$. This function can be appropriately specified in order to model spatial dependency, giving higher correlation between elements which correspond to observations made closer in space.

Paez *et al.* (2008) suggest independent Inverted Wishart prior distributions for the covariance matrices and matrix-variate Normal prior distribution for $(\boldsymbol{\beta}_{2,0}|\boldsymbol{\Sigma})$. A prior distribution must also be specified for the parameters $\lambda_i, i = 1, 2$, which define the spatial correlation matrices. Inference for this model, including the estimation of the model parameters, time forecasting and spatial interpolation, can be performed using MCMC methods as described in Paez *et al.* (2008).

4.2.2 Pollution in the Northeast of the United States

This example was presented in Paez *et al.* (2008), and refers to a data-set made available by the Clean Air Status and Trends Network (CASTNet) at the web site *www.epa.gov/castnet*. CASTNet is the primary source for data on dry acidic deposition and rural ground-level ozone in the United States. It consists of over 70 sites across the eastern and western United States and is cooperatively operated and funded by the United States Environmental Protection Agency (EPA) with the National Park Service.

When dealing with pollutants (and most environmental data-sets), the main interest is usually to be able to explain some possible causes of variation as well as to be able to perform predictions. In this particular example, the interest is to model and make predictions of levels of two pollutants: SO_2 and NO_3, which were measured in micrograms per cubic meter of air ($\mu g/m^3$). In this example, Paez *et al.* (2008) work with observations

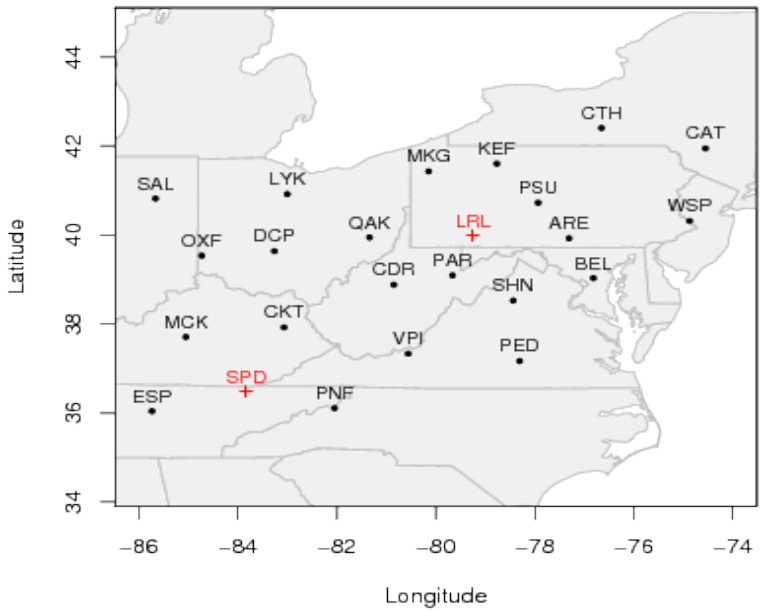


Figure 2: Map of the 24 monitoring stations in the northeast United States. The coordinates Latitude and Longitude are expressed in decimal degrees.

made in 24 monitoring stations from the 1^{st} of January 1998 to the 30^{th} of December 2004, in a total of 342 periods of time. The data, however, is not completely available, and about 4% of it is missing. Under the Bayesian approach, handling missing observations is straightforward, as they can be considered as unknown parameters of the model and can be estimated if needed.

Figure 2 shows a map of the 24 monitoring stations in the northeast United States and Figure 3 shows the trajectory of the time series observed at five randomly selected locations in space, after a logarithmic transformation. No explanatory variables are available to explain the variation of these pollutants. However, it is clear by Figure 3 that part of the time variation can be explained by seasonal components, such as sine and co-sine waves.] Exploratory analysis clearly show that the levels of the time series vary with the location where they were measured. Also, it can be noticed that the amplitude of their oscillation varies as well. The mean level of the series is also varying in time

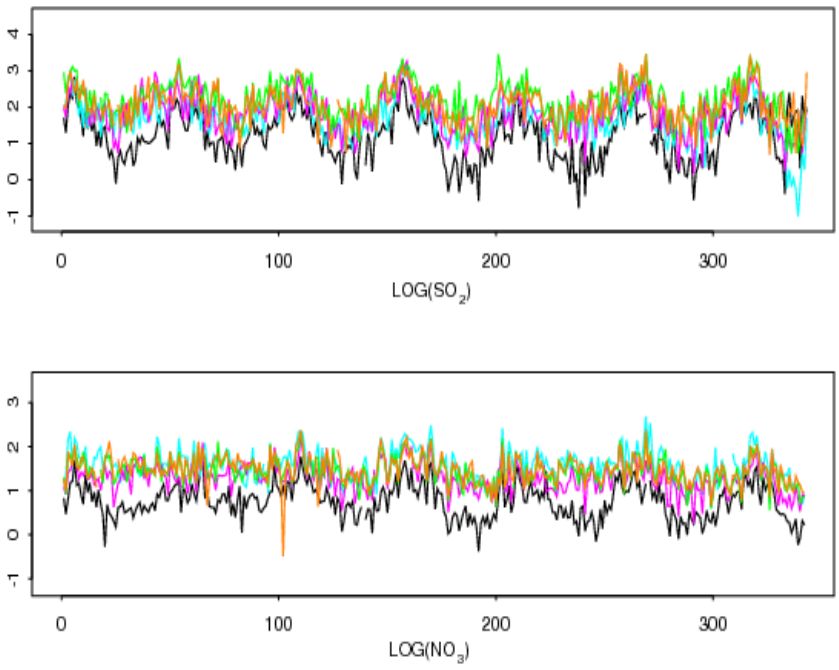


Figure 3: Series of $log(SO_2)$ and $log(NO_3)$ through time for five randomly selected monitoring stations.

(it is slowly decreasing), as well as their amplitude (which is slowly increasing). These variations suggest the use of a hierarchical model with varying intercept and also varying coefficients of sine and co-sine waves. It is intuitive, however, that these coefficients are not independent for each period of time or location in space, but rather vary smoothly in these dimensions. This should also be considered in the model spacification.

A logarithmic transformation was applied to the observed levels of both pollutants. Preliminary analysis showed strong correlation between the transformed response variables $log(SO_2)$ and $log(NO_3)$. Paez *et al.* (2008) worked with observations of 22 monitoring stations and left the other 2 to validate the models (stations LRL and SPD, which can be seen in Figure 2).

The authors compared the interpolation performance between two univariate models

(one for each response variable) and one bivariate model (where the two response variables are jointly modeled). They showed an advantage of working under the multivariate model, which was described by the authors through the following equations:

$$\begin{aligned} \boldsymbol{y}_t &= \boldsymbol{X}_{1t}\boldsymbol{\beta}_{1,t} + \boldsymbol{v}_{1t}, & \boldsymbol{v}_{1t} &\sim N(\boldsymbol{0}, \boldsymbol{I}_r, \boldsymbol{\Sigma}), \\ \boldsymbol{\beta}_{1,t} &= \boldsymbol{\beta}_{2,t} + \boldsymbol{v}_{2t}, & \boldsymbol{v}_{2t} &\sim N(\boldsymbol{0}, \boldsymbol{C} \otimes \boldsymbol{V}, \boldsymbol{\Sigma}), \\ \boldsymbol{\beta}_{2,t} &= \boldsymbol{\beta}_{2,t-1} + \boldsymbol{w}_t, & \boldsymbol{w}_t &\sim N(\boldsymbol{0}, \boldsymbol{W}, \boldsymbol{\Sigma}), \end{aligned}$$

where $\boldsymbol{y}_t = (\boldsymbol{y}_{1t}, \boldsymbol{y}_{2t})$, with $\boldsymbol{y}_{1t} = log(SO_2)_t$ and $\boldsymbol{y}_{2t} = log(NO_3)_t$. Sinuses and cosinuses waves were used as explanatory variables to explain the seasonality present in the observations, so for each location s_i, $\boldsymbol{X}_{1t}(s_i) = (1, sin(2\pi t/52), cos(2\pi t/52)), i = 1, \cdots, 22$, with $\boldsymbol{X}_{1t} = (\boldsymbol{X}_{1t}(s_1), \cdots, \boldsymbol{X}_{1t}(s_r))^T$. Note that this model is a special case of the model specified by equations (24)-(26), where $\boldsymbol{V}_1 = \boldsymbol{I}_r$, $\boldsymbol{V}_2 = \boldsymbol{C} \otimes \boldsymbol{V}$, $\boldsymbol{X}_{2t} = \boldsymbol{I}_3$ and $\boldsymbol{G}_t = \boldsymbol{I}_3$. The spatial structure is specified through a spatial correlation function which defines matrix $\boldsymbol{C}$.

In this application, the spatial correlation function was specified in the Matérn family (Matérn, 1986; Handcock and Stein, 1993), which is given by

$$C(i,j) = \frac{1}{2^{\kappa-1}\Gamma(\kappa)} \left(\frac{d_{ij}}{\phi}\right)^{\kappa} K_{\kappa}\left(\frac{d_{ij}}{\phi}\right), \qquad \phi, \kappa > 0,$$

where K_κ is the modified Bessel function of order κ. This is a flexible family, containing the exponential function ($\phi = 0.5$), and the squared exponential function, which is the limiting case when $\kappa \to \infty$. κ is a range parameter that controls how fast the correlation decays with distance, and ϕ is a smoothness (or roughness) parameter that controls geometrical properties of the random field. d_{ij} is a measure of distance between the locations of observation.

As the scale and range parameters in the Matérn family cannot be estimated consis-

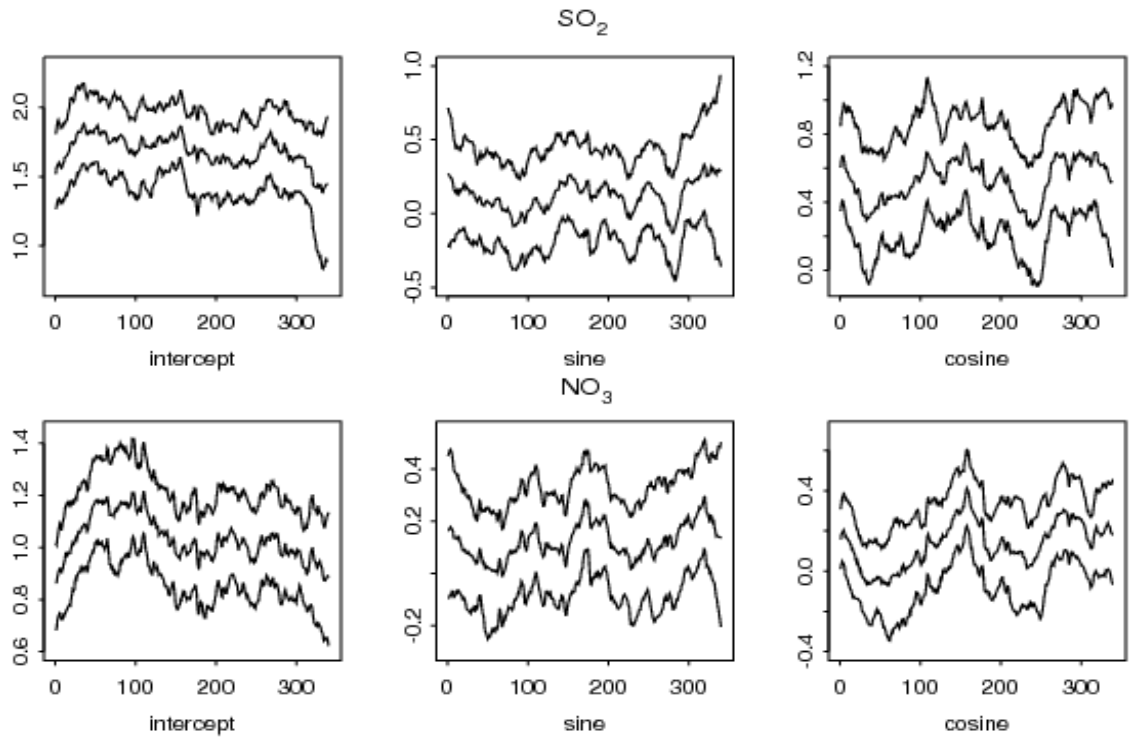


Figure 4: Posterior median and 95% credibility intervals for the elements of matrix $\boldsymbol{\beta}_2$ through time.

tently (Zhang, 2004), Paez *et al.* (2008) worked with a fixed value of $\kappa = 2.5$, and gave to ϕ a prior distribution based on the reference prior of Berger, de Oliveira and Sansó (2001). The choice of $\kappa = 2.5$ was made based on forecast performances comparing a set of different values for this parameter. Non-informative priors were specified for the other unknown parameters of the model.

Samples from the posterior distributions of the parameters were obtained through an MCMC algorithm. The elements of matrix $\boldsymbol{\beta}_2$ were shown to change significantly in time, as can be seen in Figure 4, for both $log(SO_2)$ and $log(NO_3)$. Thus, the use of temporally varying coefficients seems to be justified in this application. An interpolation of $\boldsymbol{\beta}_1$ was performed at an equally spaced grid of points, for a fixed period of time $t = 342$. Figure 5 shows a significant spatial variation of the elements of $\boldsymbol{\beta}_1$, which supports the importance

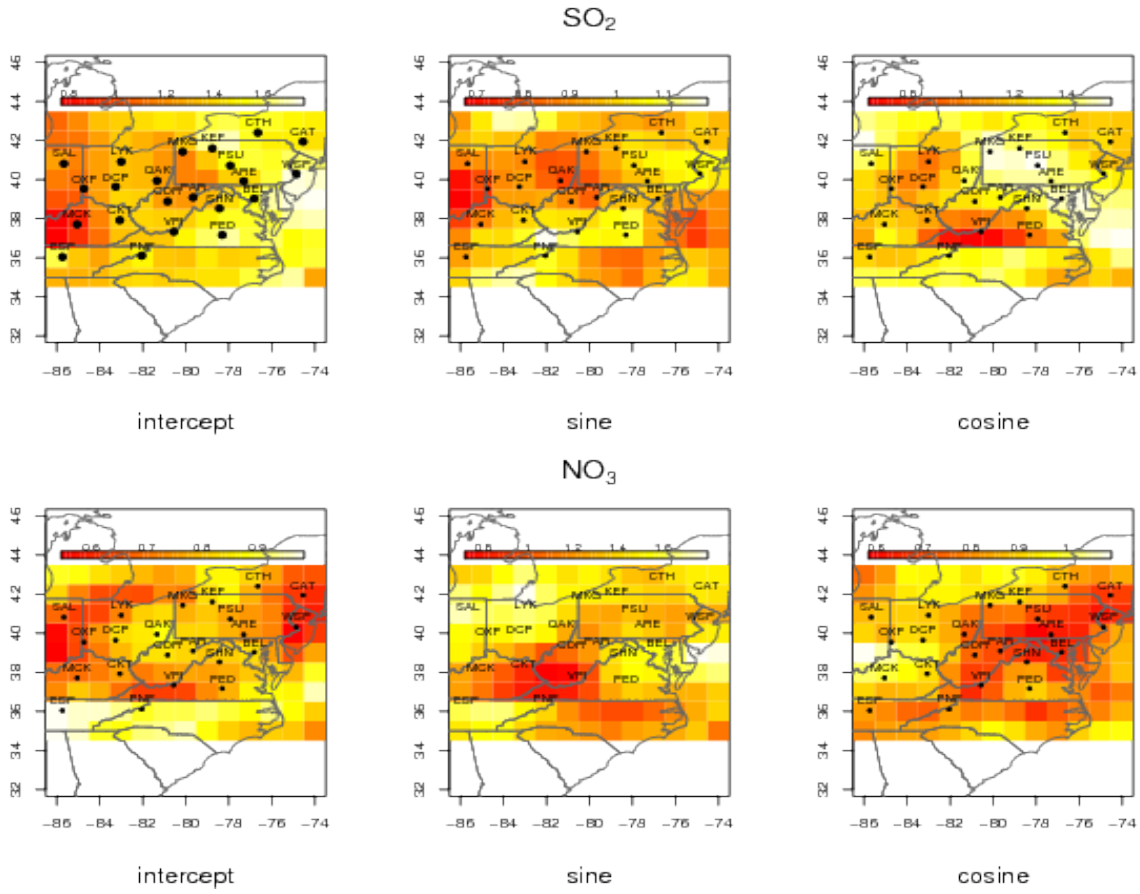


Figure 5: Posterior median of the elements of matrix $\boldsymbol{\beta}_1$ through space.

of allowing these parameters to vary in this dimension. Thus, the use of spatially varying coefficients also seems to be justified in this application. The variation of these elements is smooth in space, specially for the coefficients of sine and co-sine of $log(NO_3)$. Based on the interpolated coefficients $\boldsymbol{\beta}_1$, the response process $\boldsymbol{y}_t$ is also interpolated at $t = 342$. Each value sampled from $\boldsymbol{y}_{1,342}$ and $\boldsymbol{y}_{2,342}$ received an exponential transformation, so that a sample of values in the original scale of the pollutants SO_2 and NO_3 were obtained. The spatial variation of the posterior median of these pollutants is shown, in the original scale, in Figure 6. Finally, Figures 7 and 8 show the predictions (95% credibility intervals and median of the posterior distribution) made for SO_2 (Figure 7) and NO_3 (figure 8), for the two stations (SPD and LRL) which were used for validation. The true observed

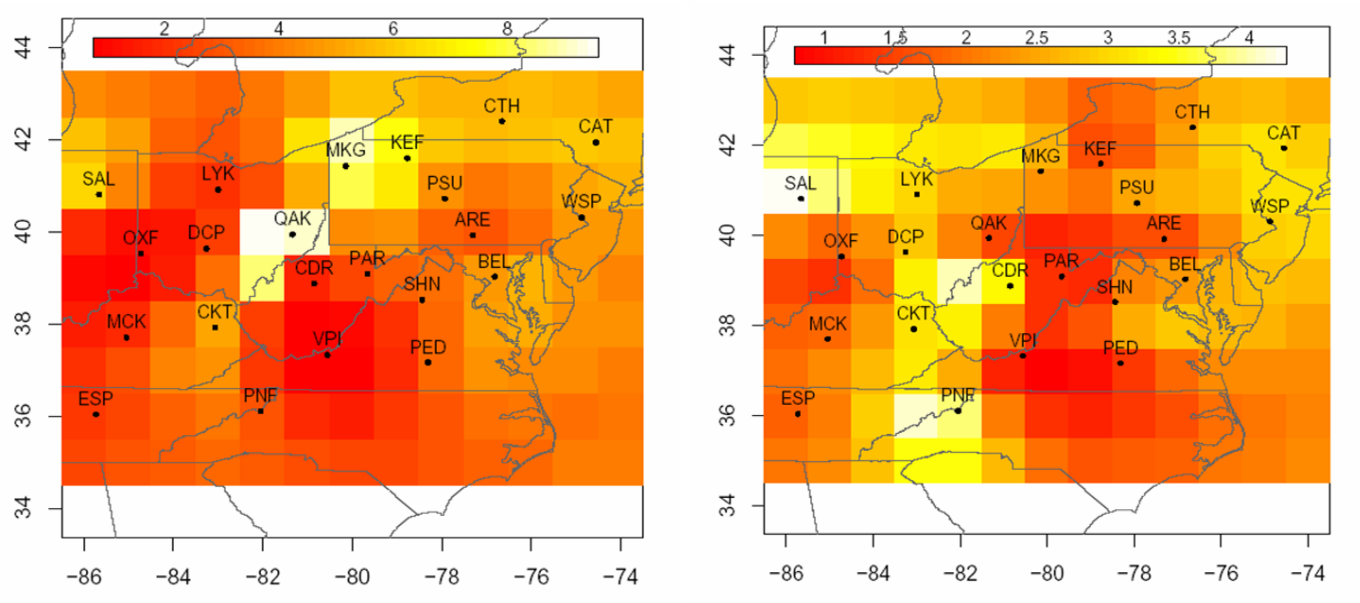


Figure 6: Predicted surface of SO_2 and NO_3 levels (μ/m^3) for $t = 342$.

trajectories are also shown, for comparison. It can be seen that the obtained predictions are good, with most of the real observations following inside the posterior credibility intervals.

4.3 Dynamic Hierarchical Models: Exponential Family Observations

The methodology which will be introduced here was presented by Hansen (2009) and it is an extension of the previous models to observations in the exponential family.

The motivation for this extension is that, sometimes, it is not possible to work under the hypothesis of normality of the observations that otherwise could be modeled through the Hierarchical Dynamic Models with structure in space proposed by Paez *et al.* (2008). Measurements of rainfall are an example of observations that could benefit from this new approach, as they cannot be assumed to be normally distributed. It is reasonable, however, to assume that they come from another distribution belonging to the exponential

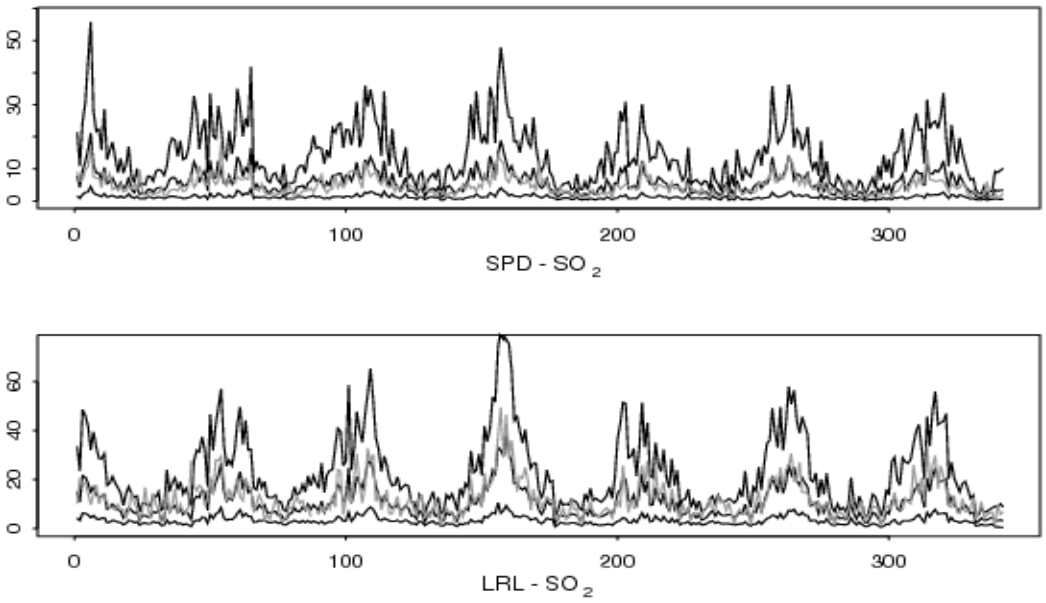


Figure 7: SO_2 levels (μ/m^3) observed at stations SPD and LRL, for $t = 1, \cdots, 342$, and the posterior median and 95% credibility intervals obtained for SO_2. The true trajectory is in gray.

family.

4.3.1 General Model Framework

Consider a set of discrete periods of time: $t = 1, \cdots, T$, where for every t, q variables are observed in r different locations in space $s_1, \cdots, s_r$. Let $\boldsymbol{y}_t$ be the $r \times q$ observation matrix. Suppose that the distribution of $\boldsymbol{y}_t$ belongs to the exponential family. A family of distributions is said to belong to the exponential family if the probability density function (or probability mass function, for discrete distributions) can be written as:

$$f(x|\lambda) = h(x)exp\left(\sum_{i=1}^{s} \eta_i(\lambda)T_i(x) - A(\lambda)\right),$$

where $h(x)$, $A(\lambda)$, $T_i(x)$ and $\eta_i(\lambda)$, $i = 1, \cdots, s$ are known functions. We will use the notation $X \sim EF(\mu)$ to denote that the random variable (either a scalar, vector or

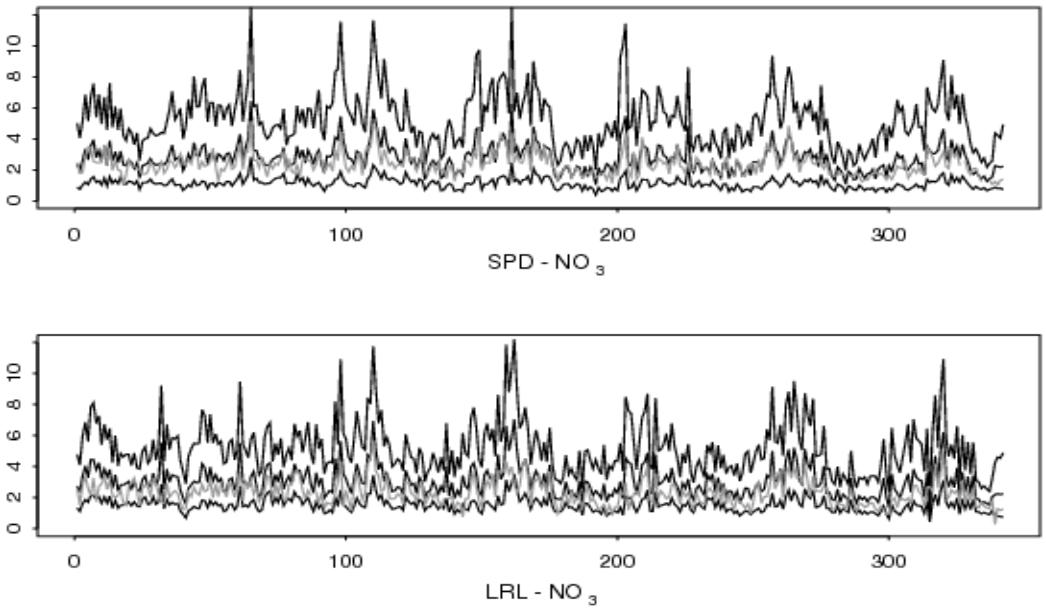


Figure 8: NO_3 levels (μ/m^3) observed at stations SPD and LRL, for $t = 1, \cdots, 342$, and the posterior median and 95% credibility intervals obtained for NO_3. The true trajectory is in gray.

matrix) X follows a distribution coming from the exponential family with mean μ.

Suppose that $\boldsymbol{y}_t$ has mean $\boldsymbol{\phi}_t$ and that $\boldsymbol{\phi}_t$ can be modeled through a function of a regression equation in which the covariate effects vary smoothly through time and space. This function is called link function and it links the linear predictor to the mean of the distribution function as in the Generalized Linear Models (Nelder and Weddenburn, 1972). The model can be specified as bellow:

$$\boldsymbol{y}_t \sim EF(\boldsymbol{\mu}_t), \tag{27}$$

$$g(\boldsymbol{\mu}_t) = \boldsymbol{X}_{1t}\boldsymbol{\beta}_{1,t} + \boldsymbol{v}_{1t}, \qquad \boldsymbol{v}_{1t} \sim N(\boldsymbol{0}, \boldsymbol{V}_{1t}, \boldsymbol{\Sigma}), \tag{28}$$

$$\boldsymbol{\beta}_{1,t} = \boldsymbol{X}_{2t}\boldsymbol{\beta}_{2,t} + \boldsymbol{v}_{2t}, \qquad \boldsymbol{v}_{2t} \sim N(\boldsymbol{0}, \boldsymbol{V}_{2t}, \boldsymbol{\Sigma}), \tag{29}$$

$$\boldsymbol{\beta}_{2,t} = \boldsymbol{G}_t\boldsymbol{\Theta}_{2,t-1} + \boldsymbol{w}_t, \qquad \boldsymbol{w}_t \sim N(\boldsymbol{0}, \boldsymbol{W}, \boldsymbol{\Sigma}), \tag{30}$$

where $g(\boldsymbol{\mu}_t)$ is a known link function and the other quantities are defined as in (24)-(26).

4.3.2 Application: Rainfall data-set in Australia

In this section we present an application of the models presented above to a single response variable consisting of measurements of rainfall in $r = 15$ monitoring stations in Australia (Hansen, 2009). Here the observations can be assumed as coming from a Gamma distribution, which belongs to the exponential family. We denote by $y_t(s_i)$ the amount of rainfall observed in time t and location s_i. We assume that $y_t(s_i)$ follows a Gamma distribution with mean $\phi_t(s_i)$ and coefficient of variation η denoted by $y_t(s_i) \sim G(\phi_t(s_i), \eta)$.

Exploratory analysis showed that the observations present a clear seasonal cycle of one year. Hansen (2009) worked with rainfall amounts aggregated by year to eliminate the seasonal effect. No explanatory variables were considered in the model. Another possibility would be to consider seasonal covariates. To model the rainfall observations, capturing time and space dependencies, Hansen (2009) proposed the use of a particular case of the model in (27)-(30) for a univariate response, where $\boldsymbol{X}_{1t} = \boldsymbol{I}_r$, $\boldsymbol{X}_{2t} = \boldsymbol{1}_r$, $G_t = 1$, $\boldsymbol{V}_{1t} = 0 \times \boldsymbol{I}_r$, $\boldsymbol{V}_{2t} = \boldsymbol{C}$ and $W = \sigma_w^2$. A spatial structure is incorporated through the specification of the elements of matrix $\boldsymbol{C}$. In this application, these elements are specified by the exponential family, where $C[i, j] = \rho \exp\{-\lambda d_{i,j}\}$. Note that this is an isotropic covariance function, which depends on the locations of observation s_i and s_j only through the distance between them ($d_{i,j}$), and depends on two unknown parameters ρ and λ. Hansen (2009) also works with a logarithmic link function. The model can be

written as follows:

$$y_t(s_i) \sim G(\phi_t(s_i), \eta), \tag{31}$$

$$log(\phi_t(s_i)) = \beta_{1,t}(s_i), \tag{32}$$

$$\boldsymbol{\beta}_{1,t} = \beta_{2,t} + \boldsymbol{v}_{2t}, \qquad \boldsymbol{v}_{2t} \sim N(\mathbf{0}, \boldsymbol{C}), \tag{33}$$

$$\beta_{2,t} = \beta_{2,t-1} + w_t, \qquad w_t \sim N(0, \sigma_w^2), \tag{34}$$

where $\boldsymbol{\beta}_{1,t} = (\beta_{1,t}(s_1), \cdots, \beta_{1,t}(s_r))$. To complete the model specifications, prior distributions for the unknown model parameters must be specified. Hansen (2009) works with vaguely informative priors. As in the previous applications, $\beta_{2,0}$ follows a Normal distribution and W follows an Inverted Gamma distribution. The prior distribution for ρ, which is also a variance parameter, was set to be an Inverted Gamma as well. Gamma priors were specified for η and λ. Samples of the posterior distribution of the unknown parameters were obtained through MCMC, using Gibbs Sampling and Metropolis-Hastings. Table 1 shows statistics of the posterior distribution of the fixed parameters of the model.

parameter	mean	2.5%	median	97.5%
η	0.	0.09	0.18	0.22
ρ	0.80	0.67	0.79	0.94
λ	0.08	0.07	0.08	0.10
σ_w^2	0.08	0.05	0.08	0.14

Table 1: Statistics from the posterior samples obtained for the fixed parameters.

It can be noticed that λ is centered around 0.08 and ρ is centered around 0.8. Taking into account the distances between the monitoring stations, these estimated values define a covariance matrix $\boldsymbol{C}$ with values that vary from 0.16 to 0.70. That way we conclude that the model captured some high correlations between observations made in different locations in space. The posterior samples obtained for σ_w^2 show high probability density

around 0.08, meaning that the parameters $\beta_{2,t}$ do not vary much in time. The spatial variation of these parameters, however, is higher, with the estimated value for the spatial variance parameter ρ being around ten times larger than the estimated value for σ_w^2.

5 Concluding Remarks

In this chapter we presented a methodological review on hierarchical dynamic models and generalizations of these models to accommodate multivariate responses, spatial variation of regression coefficients and observations in the exponential family.

The models presented here are very flexible, permitting the smooth variation of regression coefficients in time and/or space, and they can be applied to model data in many areas of interest. In this chapter we presented examples and applications made with real data-sets focusing mainly on environmental problems. The interest in this kind of application is usually to do time forecasting and interpolation in space, which can be easily done under the proposed methodology.

Inference is made under the Bayesian point of view. Usually for models like the ones presented in this chapter, the posterior distributions of the unknown parameters are not analytically tractable, and numerical methods must be used to approximate these distributions. In the applications presented here, MCMC methods were used.

Acknowledgments

The authors would like to thank Dr. Aline Maia for providing us with the Australian rainfall data-set and Dr. Flávia Landim for providing us with the occupied population and mean salary data-set for the Metropolitan regions in Brazil. They would also like to thank CNPq and FAPERJ for the financial support.

References

[1] Berger, J.O., De Oliveira, V. and Sanso, B. (2001). Objective Bayesian analysis of spatially correlated data. *Journal of the American Statistical Association*, **96**, 1361-1374.

[2] Fahrmeir, L., Kneib, T. and Lang, S. (2004). Penilized structured additive regression for space-time data: a Bayesian perspective. *Statistica Sinica*, **14(3)**, 731-761.

[3] Gamerman, D. and Lopes, H. F. (2006). *Monte Carlo Markov Chain: Stochastic Simulation for Bayesian Inference*, 2nd. edition. Chapman and Hall, London.

[4] Gamerman, D. and Migon, H.S. (1993). Dynamic hierarchical models. *Journal of the Royal Statistical Society: Series B*, **55**, 3, 629-642.

[5] Gelfand, A. E., Banerjee, S. and Gamerman, D. (2005). Spatial process modelling for univariate and multivariate dynamic spatial data. *Environmetrics*, **16**, 465-479.

[6] Handcock, M. S. and Stein, M. L. (1993). A Bayesian analysis of Krigging. *Technometrics*, **35**, 403-410.

[7] Hansen, N. (2009). Models with dynamic coefficients varying in space for data in the exponential family. Unpublished Msc dissertation, IM, Universidade Federal do Rio de Janeiro (in Portuguese).

[8] Harrison, P. J. and Stevens, C. F. (1979). Bayesian Forecasting. *Journal of the Royal Statistical Society Series B*, **38**, 205-247.

[9] Johannes, M., Polson, N.G. (2007) Exact Particle Filtering and Learning. *Working Paper.* The University of Chicago Booth School of Business.

[10] Landim, F. M. P. F. and Gamerman, D. (2000). Dynamic hierarchical models - an extension to matrix variate observations. *Computational Statistics and Data Analysis,* **35**, 11-42.

[11] Liu, J. and West, M. (2001). Combined parameters and state estimation in simulation-based Filtering. *Sequential Monte Carlo Methods in Practice.* Springer-Verlag, New York.

[12] Carvalho, C. M., Johannes, M. S., Lopes, H. F. and Polson, N. G. (2010). Particle Learning and Smoothing. *Statistical Science*, **25(1)**, 88-106.

[13] Matérn, B. (1986). *Spatial Variation*, 2nd. edition. Berlin: Springer-Verlag.

[14] Nelder, J. A. and Wedderburn, R.W.M. (1972). Generalized linear models. *Journal of the Royal Statistical Society: Series A*, **135**, 370-384.

[15] Paez, M. S., Gamerman, D., Landim, F. M. P. F. and Salazar, E. (2008). Spatially varying dynamic coefficient models. *Journal of Statistical Planning and Inference*, **138**, 1038-1058.

[16] Quintana, J. (1987). Multivariate Bayesian forcasting models, unpublished Ph.D. Thesis, University of Warwick.

[17] Rue, H., Martino, S., and Chopin, N. 2009. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations (with discussion). *Journal of the Royal Statistical Society Series B*, **71(2)**, 319-392.

[18] West, M. and Harrison, P. J. (1989). *Bayesian Forecasting and Dynamic Model*, New York: Springer.

[19] West, M. and Harrison, P. J. (1997). *Bayesian Forecasting and Dynamic Model*, 2nd. edition. London: Springer.

[20] Zhang (2004). Inconsistent estimation and asymptotically equal interpolations in model-based geostatistics. *Journal of the American Association*, **99**, 250-261.