

*Análise de modelos para a estimação e  
previsão de processos espaço-temporais*

Marina Silva Paez

IM/UFRJ

Setembro 2004

# **Livros Grátis**

<http://www.livrosgratis.com.br>

Milhares de livros grátis para download.

*Análise de modelos para a estimação e  
previsão de processos espaço-temporais*

Marina Silva Paez

Prof. Orientador: Dani Gamerman

Universidade Federal do Rio de Janeiro

Instituto de Matemática

Rio de Janeiro, RJ

Setembro 2004

Aos meus queridos avós Guilherme e Carmen.

# Agradecimentos

Agradeço aos meus pais, Diana e Adauto, e a minha irmã, Anita, por estarem sempre ao meu lado me dando felicidade e segurança.

ao Trey, meu namorado, por todo carinho e apoio que me deu na Inglaterra, tornando muito feliz um período que tinha tudo pra ser o mais difícil da minha vida.

ao professor Dani pela orientação, e principalmente por me incentivar constantemente e me fazer acreditar que tudo é possível quando se quer.

ao professor Peter Diggle pela orientação no período que estive em Lancaster.

ao professor Victor de Oliveira e à professora Flávia Landim pela colaboração nos capítulos 2 e 3 dessa tese, respectivamente.

aos demais professores e colegas do departamento de estatística da UFRJ por tudo que vem me ensinando ao longo dos anos.

à CAPES e ao CNPq pelo apoio financeiro.

# Resumo

Nesse trabalho analisamos observações feitas no espaço e/ou tempo que apresentam estrutura de correlação nessas dimensões, usando para isso modelos que levam em consideração essa dependência, e tendo como objetivo fazer previsões para tempos futuros e interpolação espacial.

Para tratar a variação espaço-temporal, propomos modelos hierárquicos que incorporam uma estrutura de correlação temporal e ou espacial na média (condicional aos processos explicativos) da variável resposta, e nos casos onde existe considerável heterogeneidade espacial no efeito de processos explicativos, incorporam também uma estrutura de correlação no tempo/espaço nos coeficientes relativos a esses processos. Tendo como base os modelos dinâmicos (para tratar variação temporal) e os campos aleatórios gaussianos (para tratar variação espacial), e trabalhando sob o ponto de vista Bayesiano, propomos três diferentes modelagens para dados temporais ou espaço-temporais.

A primeira proposta é uma extensão dos modelos usados em Paez e Gamerman (2003), onde permitimos que os coeficientes das covariáveis variem no espaço, e estudamos os efeitos da incorporação de incertezas no modelo. A segunda proposta é uma generalização da primeira, onde os parâmetros associados às variáveis resposta variam no tempo e no espaço. Essa modelagem é um caso particular dos modelos dinâmicos hierárquicos matriz-variados propostos por Landim e Gamerman (2000) para respostas gaussianas multivariadas, onde impomos uma forma estruturada para as matrizes de variância espacial. A terceira e última proposta é um estudo da variação temporal da incidência de eventos através de processos de Cox log-gaussianos, baseado no artigo de Brix e Diggle (2001).

# Abstract

In this work we analyze data sets observed in time and/or space which present correlation structure in these dimensions, and propose models that take into account this dependence, aiming to make predictions for future times and do spatial interpolation.

To explain the spatio-temporal variation we propose hierarchical models which incorporate a temporal and/or spatial correlation structure in the mean (conditionally on explanatory processes) of the response variable, and in cases where there is spatial heterogeneity in the effects of the explanatory processes, also incorporate a time/space correlation structure in the regression coefficients. Based on dynamic models (to explain temporal variation) and gaussian random fields (to explain spatial variation), and working under a Bayesian point of view, we propose three different approaches to model temporal or space-time data.

The first approach is an extension of the models used in Paez and Gamerman (2003), where the regression coefficients are allowed to vary in space, and we study the effects of adding uncertainties to the model. The second approach is a generalization of the first one, where the regression parameters vary in time and space. This model is a particular case of the matrix-variate dynamic hierarchical models proposed by Landim and Gamerman (2000) for multivariate gaussian responses, where we impose a structured form to the covariance matrixes which define the spatial dependence. The third and last approach is a study of the temporal variation of the incidence of events using log-gaussian Cox processes, based on the models proposed in Brix and Diggle (2001).

# Sumário

<b>1</b>	<b>Introdução</b>	<b>1</b>
<b>2</b>	<b>Modelos Espaço-temporais com Coeficientes Variando no Espaço: Uma aplicação a dados de <math>PM_{10}</math> no Rio de Janeiro</b>	<b>8</b>
2.1	Introdução . . . . .	8
2.2	O modelo . . . . .	10
2.2.1	Definição do modelo . . . . .	10
2.2.2	A priori . . . . .	12
2.2.3	Interpolação . . . . .	14
2.3	Trabalhando com dados simulados . . . . .	16
2.3.1	Resultados . . . . .	19
2.4	Análise de dados $PM_{10}$ no Rio de Janeiro . . . . .	30
2.4.1	Inferência sobre os parâmetros do modelo . . . . .	33
2.4.2	Interpolação espacial . . . . .	35



2.5	Conclusões . . . . .	38
<b>3</b>	<b>Modelos Dinâmicos Hierárquicos Espaço-Temporais para Respostas Multivariadas</b>	<b>40</b>
3.1	Introdução . . . . .	40
3.2	Descrição do modelo . . . . .	43
3.2.1	Exemplo 1 . . . . .	44
3.2.2	Exemplo 2 . . . . .	46
3.3	Inferência e métodos computacionais . . . . .	48
3.4	Predição . . . . .	53
3.5	Interpolação . . . . .	54
3.6	Aplicação com dados simulados . . . . .	56
3.6.1	Descrição do modelo . . . . .	56
3.6.2	Estimação do modelo . . . . .	58
3.7	Aplicação com dados de poluição . . . . .	63
3.8	Conclusão . . . . .	71
<b>4</b>	<b>Análise de Processos Temporais de Cox e a Influência da Agregação dos Dados Observados</b>	<b>74</b>
4.1	Introdução . . . . .	74
4.2	Processo de Cox . . . . .	76

4.2.1	Abordagem 1: respostas no tempo contínuo . . . . .	78
4.2.2	Abordagem 2: dados de contagem . . . . .	82
4.3	Estimação por métodos de MCCM . . . . .	86
4.3.1	Estimando parâmetros sob a abordagem 1 . . . . .	86
4.3.2	Estimando parâmetros sob a abordagem 2 . . . . .	90
4.4	Análise de dados simulados . . . . .	91
4.4.1	Modelando $z$ - abordagem 1 . . . . .	94
4.4.2	Modelando $y_1$ - abordagem 2.1 . . . . .	97
4.4.3	Modelando $y_2$ - abordagem 2.2 . . . . .	101
4.5	Análise de casos de doenças gastro-intestinais . . . . .	105
4.5.1	Motivação e análises preliminares . . . . .	105
4.5.2	Modelagem e resultados . . . . .	108
4.6	Conclusão . . . . .	112
4.7	Apêndice - aproximação usada na geração de $\gamma$ no algoritmo de MCCM sob a abordagem 1 . . . . .	115
<b>5</b>	<b>Extensões e conclusões</b>	<b>118</b>
5.1	Extensões . . . . .	118
5.2	Conclusões . . . . .	122
	<b>Bibliografia</b>	<b>127</b>

# Lista de Figuras

2.1	Superfícies Simuladas dos componentes do modelo (1) para $t=30$ . . . . .	17
2.2	Locais de amostragem no espaço para os quatro conjuntos de dados . . .	18
2.3	Histograma da posteriori de $\gamma_0, \gamma_1, \sigma_e^{-2}, \sigma_0^{-2}$ e $\sigma_1^{-2}$ considerando as especificações a priori PNI e PPI. . . . .	23
2.4	Histograma da posteriori de $\lambda_0$ e $\lambda_1$ baseada nas três especificações a priori.	24
2.5	Superfícies interpoladas de $\beta_0(\cdot)$ e $\beta_1(\cdot)$ sob as especificações PPI e PNI. .	25
2.6	Superfície interpolada de $Y(\cdot, 30)$ . . . . .	26
2.7	Superfície verdadeira de $X(\cdot, 30)$ . . . . .	29
2.8	Superfície interpolada de $X(\cdot, 30)$ baseada nos conjuntos de dados 1 e 2 .	29
2.9	Superfície interpolada de $Y(\cdot, 30)$ quando $X(\cdot, \cdot)$ é desconhecido, baseada nos conjuntos de dados 1 e 2. . . . .	30
2.10	Mapa da Região Metropolitana do Rio de Janeiro, mostrando a localização das estações de monitoramento e a área de interpolação. . . . .	31
2.11	Efeito da temperatura por estação. . . . .	32
2.12	Histogramas da amostra a posteriori dos parâmetros $\sigma_e^{-2}, \sigma_0^{-2}, \sigma_1^{-2}, \lambda_0$ e $\lambda_1$ .	35

2.13	Superfície interpolada de $\beta_0(\cdot)$ . . . . .	37
2.14	Superfície interpolada de $\beta_1(\cdot)$ . . . . .	37
2.15	Superfície interpolada das concentrações de $PM_{10}$ em $t=59$ . . . . .	38
3.1	histograma da amostra da distribuição a posteriori dos parâmetros $b$ e $k$	59
3.2	histograma da amostra da distribuição a posteriori dos elementos da matriz $\Sigma$	59
3.3	histograma da amostra da distribuição a posteriori dos elementos da matriz $V$	60
3.4	histograma da amostra da distribuição a posteriori dos elementos da matriz W . . . . .	60
3.5	Quantis de 2.5%, 50% e 97.5% da amostra da distribuição a posteriori dos componentes de $\Theta_{2t}$ , $t = 1, \dots, 70$ , com a trajetória verdadeira em negrito.	61
3.6	Superfícies reais e interpoladas para $\alpha_{(1,1),61}$ e $\beta_{(1,1),61}$ . . . . .	62
3.7	Superfícies reais e interpoladas para $\alpha_{(2,1),61}$ e $\beta_{(2,1),61}$ . . . . .	62
3.8	Superfícies reais e interpoladas para $Y_{1,61}$ e $Y_{2,61}$ . . . . .	63
3.9	Mapa da Alemanha com a localização das estações de monitoramento . .	64
3.10	Variáveis resposta $Y_1(LN(SO_2))$ e $Y_2(LN(NO_2))$ ao longo do tempo, por estação de monitoramento. . . . .	65
3.11	Plot de $Y_1$ contra $Y_2$ . . . . .	65
3.12	Plot de $X$ contra $Y_1$ ; plot de $X$ contra $Y_2$ . . . . .	66
3.13	Coefficientes de regressão linear ao longo do espaço . . . . .	67

3.14	Histograma da amostra da distribuição a posteriori obtida para os elementos da matriz $\Sigma$ . . . . .	69
3.15	Histograma da amostra da distribuição a posteriori obtida para os elementos da matriz $V$ . . . . .	69
3.16	Histograma da amostra da distribuição a posteriori obtida para os elementos da matriz $W$ . . . . .	69
3.17	Histograma da amostra da distribuição a posteriori obtida para os parâmetros $b$ e $k$ . . . . .	70
3.18	Quantis de 2.5%, 50% e 97.5% da amostra da distribuição a posteriori dos elementos de $\Theta_{2t}$ , $t = 1, \dots, 84$ . . . . .	71
4.1	valores simulados dos processos $\exp(\gamma)$ , $\rho$ e $\lambda$ ao longo do tempo. . . . .	93
4.2	Histograma das amostras da posteriori de $\sigma^2$ e $\phi$ . . . . .	95
4.3	Trajetória real de $\gamma$ com intervalos de credibilidade da amostra da posteriori (em vermelho): (A) nos tempos 1 a 9600; (B) nos tempos 9601 a 19200. . . . .	95
4.4	Previsão para o número de casos nas 1008 últimas horas geradas sob as abordagens A e B . . . . .	96
4.5	Histograma das amostras da posteriori de $\sigma^2$ e $\phi$ . . . . .	98
4.6	Trajetória real de $\gamma_1$ com intervalos de credibilidade da amostra da posteriori (em vermelho):(A) nos tempos 1 a 400; (B) nos tempos 401 a 800. . . . .	98
4.7	Previsão para o número de casos nos 42 últimos dias - abordagem 1. . . . .	100
4.8	Previsão para o número de casos nos 42 últimos dias - abordagem 2.1. . . . .	100

4.9	Histograma das amostras da posteriori de $\sigma_1^2$ e $\phi_1$ . . . . .	103
4.10	Trajectoria real de $\gamma_7$ com intervalos de credibilidade da amostra da posteriori (em vermelho). . . . .	103
4.11	comparação das estimativas do processo $\log(\Lambda)$ agregado por semana, pelas três abordagens. . . . .	104
4.12	Previsão para o número de casos nas 13 últimas semanas usando as abordagens 1, 2.1 e 2.2 e cenários A e B. . . . .	104
4.13	Casos reportados ao longo do tempo. . . . .	107
4.14	Histograma dos casos reportados por dia. . . . .	107
4.15	Intervalos de 95% de credibilidade dos coeficientes (obtido no ajuste de um MLG) para os indicadores no dia da semana de segunda a domingo. . . . .	108
4.16	Histogramas das amostras da posteriori de $\sigma^2$ e $\phi$ sob as abordagens 1 e 2. . . . .	110
4.17	Intervalos de 95% de credibilidade dos parâmetros $\beta$ correspondendo aos indicadores de segunda a domingo - abordagem 1. . . . .	110
4.18	Intervalos de 95% de credibilidade da trajetória do processo $\Lambda$ e número de casos observados por dia: (A) do dia 1 a 365; (B) do dia 366 a 730. . . . .	111
4.19	Previsão para o número de casos nos 10 últimos dias usando as abordagens 1 e 2 . . . . .	112

# Lista de Tabelas

2.1	Resumo da inferência a posteriori dos parâmetros baseada no conjunto de dados 1 . . . . .	20
2.2	Resumo da inferência a posteriori dos parâmetros baseada no conjunto de dados 2 . . . . .	20
2.3	Resumo da inferência a posteriori dos parâmetros baseada no conjunto de dados 3 . . . . .	21
2.4	Resumo da inferência a posteriori dos parâmetros baseada no conjunto de dados 4 . . . . .	21
2.5	Estatística $DS$ para $\beta_0(\cdot)$ , $\beta_1(\cdot)$ , e $Y(\cdot, 30)$ (desvio padrão estimado entre parêntesis) . . . . .	27
2.6	Sumário da inferência a posteriori dos parâmetros do processo $X(\cdot, \cdot)$ baseado nos quatro conjuntos de dados . . . . .	27
2.7	Estações de monitoramento de $PM_{10}$ na Região Metropolitana do Rio de Janeiro . . . . .	31
2.8	Estatísticas Descritivas baseadas na amostra das distribuições a posteriori estimadas por MCCM . . . . .	34

# Capítulo 1

## Introdução

A aleatoriedade é um conceito criado pelo homem para justificar aquilo que não pode entender. Perguntas como "será que vai chover amanhã?" ou "será que está chovendo em São Paulo?" (para quem não está em São Paulo) estão presentes no cotidiano de cada um de nós, e a resposta para essas perguntas é demasiadamente complicada para se dar uma resposta exata. Depende de uma combinação de fatores ambientais tão complexos que nem o mais competente meteorologista é capaz de responder com 100% de certeza. Para isso são desenvolvidos modelos que levam em consideração diversos processos associados, como a velocidade dos ventos e a umidade relativa do ar, entre outros. A velocidade dos ventos, por sua vez, é determinada por outros fatores complexos, e assim por diante. Os processos ambientais, que são objeto de estudo nessa tese, são causados por uma infinidade de outros processos desconhecidos e interdependentes, tornando a modelagem dos mesmos uma tarefa difícil. A nosso favor está o fato de que a natureza não dá saltos, como afirma o princípio da continuidade. Processos ambientais são em geral contínuos no tempo e no espaço, variando portanto de forma suave em ambas as dimensões. A distribuição espacial desses processos pode ser atribuída a uma combinação de outros processos humanos e naturais (também contínuos), que operam em diferentes escalas espaciais e temporais.



A suavidade presente no processo de interesse em uma determinada dimensão é relativa, assim como a noção de tempo e espaço. Um processo que varia bruscamente a cada hora pode estar variando suavemente a cada segundo, e um processo que varia bruscamente entre dois locais a 20km de distância pode estar variando suavemente a cada metro. Dessa forma, quando um processo é dito "independente" no espaço (ou tempo), isso simplesmente significa que ele não apresenta dependência significativa para uma dada unidade de espaço (ou tempo). Muitas vezes variáveis explicativas em modelos de regressão ajudam a "remover" parte da correlação existente nesses processos, dando origem, em alguns casos, a resíduos independentes. Novamente, a independência desses resíduos é relativa as unidades de tempo e espaço em questão.

Considere, por exemplo, o estudo da temperatura ambiental ao longo do ano. Suponha que a temperatura em um determinado local seja registrada a cada 15 dias. Sem que haja a incorporação de um componente sazonal que explique a variação de temperatura ao longo do ano, os erros obtidos na modelagem não podem ser considerados independentes. Se, ao contrário, a sazonalidade está presente no modelo, a suposição de independência dos erros é bastante razoável. Suponha agora que os dados são observados a cada dia. A suposição de independência nesse caso parece pouco realista. Provavelmente se a temperatura registrada no dia  $t$  está abaixo do esperado pelo modelo, a temperatura no dia  $t+1$  também estará. Essa dependência é ainda mais evidente se os dados são observados a cada hora, minuto ou segundo.

A percepção da correlação é portanto dependente da frequência de observação dos dados no tempo e espaço. Em problemas reais, entretanto, não é possível observar processos contínuos nessas dimensões, e a análise fica limitada às observações feitas desses processos em determinados períodos de tempo e locais no espaço. O mesmo problema se dá na observação de processos explicativos.

A dependência espaço-temporal não está restrita apenas a processos contínuos. Processos pontuais (onde o tempo e/ou local de observação de eventos são registrados) também

apresentam correlação nessas dimensões. Isso porque são fenômenos geralmente ligados a processos contínuos não observáveis. Como um exemplo, suponha que estamos interessados em estudar o padrão de variação espaço-temporal da incidência da Dengue no Rio de Janeiro. A suposição de que a probabilidade de um indivíduo contrair a doença é contínua no tempo/espaço é bastante razoável. Entretanto, tudo que se pode observar são os locais e tempos de registro de cada caso de Dengue.

Em muitos casos, no entanto, registros não são feitos com tamanha precisão, e tudo que se observa é o número de casos ocorridos a cada dia, semana, mês, etc.. A dependência espacial remanescente nos dados observados, nesse caso, é um compromisso entre o padrão de correlação espaço-temporal e o nível de agregação das observações. Um dos pontos de interesse nessa pesquisa é estudar os efeitos de diferentes níveis de informação proveniente do conjunto de dados analisado. Isso é feito através da comparação da modelagem feita usando diferentes níveis de informação a respeito de processos explicativos (capítulo 2), e através da comparação da modelagem considerando diferentes níveis de agregação da variável resposta (capítulo 4).

Nosso objetivo principal é analisar observações feitas no espaço e/ou tempo que apresentam estrutura de correlação nessas dimensões mesmo após a remoção dos efeitos dos processos explicativos disponíveis, usando para isso modelos que levem em consideração essa dependência para fazer previsões para tempos futuros e interpolação espacial.

Diferentes tipos de modelagem têm sido propostos recentemente na literatura para a análise de dados desse tipo (uma revisão pode ser vista em Kyriakidis e Journel (1999)), sendo que a abordagem Bayesiana vem obtendo um aumento substancial em popularidade para a análise desses modelos. Esse fenômeno se dá principalmente devido a avanços computacionais recentes, tais como o algoritmo de Monte Carlo em Cadeias de Markov (MCCM), mas também pelo fato da abordagem Bayesiana ser particularmente adequada para fazer inferência preditiva. Ela tem a vantagem de contemplar plenamente a incerteza dos parâmetros, o que não ocorre na abordagem clássica. Além disso, inter-

polação e previsão são técnicas "naturais" sob o ponto de vista Bayesiano, onde qualquer quantidade desconhecida pode ser caracterizada pela sua distribuição a posteriori, que pode ser obtida analiticamente ou estimada através de métodos numéricos quando apresentam estruturas muito complicadas. Exemplos de trabalhos recentes da análise Bayesiana de modelos espaço-temporais incluem Le e Zidek (1992), Guttorp, Meiring e Sampson (1994), Handcock e Wallis (1994), Wikle e Cressie (1999), Wikle, Berliner e Cressie (1998), Brown et al. (2001), Tonellato (2001), e Shaddick e Wakefield (2001).

Para tratar a variação espaço-temporal, propomos modelos hierárquicos que incorporam uma estrutura de correlação temporal e ou espacial na média (condicional aos processos explicativos) da variável resposta, e nos casos onde existe considerável heterogeneidade espacial no efeito de processos explicativos, incorporam também uma estrutura de correlação no tempo/espaço nos coeficientes relativos a esses processos. Uma forma flexível de tratar a variação temporal é através de modelos dinâmicos lineares, que foram introduzidos por Harrison e Stevens (1976) e estão bem documentados em West e Harrison (1997). O princípio fundamental dessa abordagem é a definição de modelos sequenciais definidos através de estruturas paramétricas. Esses modelos são caracterizados por uma equação de observação e uma ou mais equações de transição, dadas abaixo:

$$\begin{aligned}
 y_t &= F_t' \Theta_t + \nu_t, & \nu_t &\sim N(0, V_t) \\
 \Theta_t &= G_t \Theta_{t-1} + w_t, & w_t &\sim N(0, W_t)
 \end{aligned}$$

com condição inicial  $(\Theta_0 | D_0) \sim N(m_0, C_0)$ , onde  $t$  é o índice de tempo,  $\nu_t$  e  $w_t$  são erros mutuamente independentes,  $F_t$  e  $G_t$  são matrizes de evolução, e  $N(\mu, \sigma^2)$  denota a distribuição normal com média  $\mu$  e variância  $\sigma^2$ . A primeira equação define a distribuição da variável aleatória  $y_t$  condicionalmente ao vetor de parâmetros de estado  $\Theta_t$ , e a segunda equação define a evolução temporal dos parâmetros  $\Theta$ . Como exemplo podemos citar os modelos autoregressivos de ordem 1 (AR1), obtidos ao especificar  $G_t$  como sendo uma matriz identidade vezes um escalar, utilizados no capítulo 4.

Nesse trabalho adotamos a teoria de campos aleatórios gaussianos (Cressie, 1993) para tratar a variação espacial dos parâmetros que variam nessa dimensão. Esses são modelos probabilísticos úteis que levam em consideração a correlação existente no espaço através de uma estrutura paramétrica. Sob a suposição de normalidade, o modelo é especificado através de funções de média e covariância. Hipóteses simplificadoras geralmente utilizadas são as de estacionariedade e isotropia. Um processo  $w(\cdot)$  na região  $S \in R^d$  é dito estacionário se a esperança e a variância de  $w(s_i)$  são as mesmas para todo  $i=1, \dots, N$ , e a correlação entre  $w(s_i)$  e  $w(s_j)$  depende somente de  $s_i - s_j$  para qualquer par de inteiros  $0 < i, j < n$ . Se, adicionalmente, essa correlação depender somente de  $u_{ij} = \|s_i - s_j\|$ , a distância euclidiana entre  $s_i$  e  $s_j$ , dizemos que o processo além de estacionário é isotrópico. Nesse último caso, temos que  $\rho[w(s_i), w(s_j)] = \rho_\theta(u_{ij})$  para alguma função definida positiva  $\rho$ , possivelmente dependendo de parâmetros  $\theta$  e  $u_{ij}, \forall(i, j)$ .

Assumindo isotropia, a função de covariância do modelo tipicamente depende da distância entre locais no espaço, sendo em geral especificada de acordo com uma família de distribuições com propriedades desejáveis, como a exponencial, exponencial de potência, esférica e Mátern. Nesse caso, fazer interpolação espacial, que é um dos nossos objetivos, torna-se uma tarefa simples, levando em consideração propriedades da teoria normal.

Nesse trabalho, tendo como base os modelos dinâmicos (para tratar variação temporal) e os campos aleatórios gaussianos (para tratar variação espacial), e trabalhando sob o ponto de vista Bayesiano, propomos três diferentes modelagens para dados temporais ou espaço-temporais.

A primeira proposta é uma extensão dos modelos usados em Paez e Gamerman (2003) para processos gaussianos, onde as observações são condicionalmente independentes, com a média especificada pela soma de componentes espaciais e temporais, e uma função linear das variáveis explicativas. Aqui é feita uma extensão desse modelo, onde permitimos que os coeficientes das covariáveis variem no espaço. Trabalhamos com dados simulados avaliando o efeito do padrão espacial da amostra, o efeito das especificações das distribuições

a priori, e a incorporação de interteza no modelo quando uma os mais covariáveis não são observadas em toda a região de interesse, tendo como objetivo fazer interpolação. Também analisamos o mesmo conjunto de dados utilizado em Paez e Gamerman (2003) para efeito de comparação. Esse conjunto de dados consiste em níveis de  $PM_{10}$  coletados em 16 estações de monitoramento para 59 períodos de tempo na Região Metropolitana do Rio de Janeiro, tendo como variáveis explicativas a temperatura máxima observada em um ponto fixo do espaço, e os dias da semana.

A segunda proposta pode ser vista como uma generalização da primeira proposta considerada, onde os parâmetros associados às variáveis resposta podem variar no tempo e no espaço. Essa modelagem é um caso particular dos modelos dinâmicos hierárquicos matriz-variados propostos por Landim e Gamerman (2000) para respostas gaussianas multivariadas. Esses modelos são bastante flexíveis pois além de terem uma estrutura de evolução temporal que possibilita explorar possíveis movimentos dinâmicos, também nos permitem incorporar estrutura espacial em qualquer nível da hierarquia. A partir dessa modelagem podemos desenvolver modelos bem estruturados no espaço e no tempo, contemplando inclusive interações entre essas duas dimensões. Nesse trabalho propusemos uma forma estruturada para as matrizes de variância espacial, alternativamente à estimação da matriz cheia de Landim e Gamerman (2000). Com essa restrição, nosso modelo, além da previsão de dados para tempos futuros, tem a vantagem de permitir a interpolação das variáveis de interesse para qualquer ponto no espaço, levando em consideração a estrutura de correlação espacial presente nas variáveis resposta. Como exemplo consideramos o modelo com dois níveis hierárquicos, trabalhando sob a suposição de independência espacial a nível das variáveis resposta. Uma análise utilizando dados simulados foi feita para a validação do algoritmo de estimação dos parâmetros do modelo, assim como o estudo da sua capacidade preditiva no tempo e no espaço. Por fim a modelagem proposta foi aplicada a dados de poluição ( $SO_2$  e  $NO_2$ ) observados mensalmente em 7 estações de monitoramento na Alemanha, em um período de 84 meses, tendo como variável explicativa o índice pluviométrico.

A terceira e última proposta é um estudo da variação temporal da incidência de eventos através de processos de Cox log-gaussianos, baseado no artigo de Brix e Diggle (2001). Nesse artigo, os autores propõem métodos de momento para fazer inferência sobre esses processos, que são propostos para modelar a variação espaço-temporal em dados de contagem. Nessa pesquisa propomos processos similares para tratar a variação temporal dos dados sob o ponto de vista Bayesiano, considerando não somente dados de contagem, mas também os registros de tempo de ocorrência de eventos como variável resposta. O modelo proposto supõe que o processo em questão é um processo de poisson não homogêneo com função de intensidade contínua  $\Lambda$ , que é o produto da intensidade populacional  $\rho$  e uma função de risco  $\Pi$ , que por sua vez segue um modelo log linear incorporando um processo dinâmico no tempo e efeitos de variáveis explicativas. Trabalhamos com dados simulados para validar o algoritmo, tendo também como objetivo comparar a performance da estimação do modelo e predição para tempos futuros em diferentes níveis de agregação da variável resposta. Uma aplicação foi feita a dados de infecções gastro-intestinais observados sob forma de contagens diárias na Inglaterra nos anos de 2002 e 2003. Essa última pesquisa foi realizada de setembro de 2003 a setembro de 2004, na Universidade de Lancaster, sob a supervisão do professor Peter Diggle.

Inferência para os parâmetros dos modelos propostos foram feitas por simulação através do uso de Métodos de Monte Carlo em Cadeias de Markov (MCCM). Nas três propostas temos como objetivo não apenas estimar os parâmetros do modelo para entender o processo em questão, mas também fazer previsão no tempo e interpolação no espaço, feitos também através de MCCM.

Essa tese está dividida em 5 capítulos, sendo o primeiro a introdução. Nos capítulos 2, 3 e 4 são feitas as descrições dos modelos para cada proposta, respectivamente, com a apresentação de resultados de inferência, análise de dados simulados e aplicação a dados reais. Por fim temos no capítulo 5 possíveis extensões para as modelagens propostas nesse trabalho e as conclusões obtidas.

## Capítulo 2

# Modelos Espaço-temporais com Coeficientes Variando no Espaço: Uma aplicação a dados de $PM_{10}$ no Rio de Janeiro

### 2.1 Introdução

Nos dias de hoje, dados ambientais tais como níveis de poluição, são coletados tanto no tempo quanto no espaço. Geralmente as dinâmicas que governam esses processos dão origem a dados que quando observados próximos, em qualquer das duas dimensões, são mais similares que aqueles observados com maior distância. Os Campos Aleatórios são modelos probabilísticos úteis para a análise desse tipo de dado, especialmente quando o objetivo principal é de fazer predição. Cressie (1993) é uma referência básica para esse tipo de análise.

Muitos modelos espaço-temporais tem como hipótese que as observações podem ser decompostas em uma função de regressão e um processo de erros correlacionados, onde os parâmetros de regressão são supostamente desconhecidos. Essa hipótese, entretanto, pode não ser realista para descrever conjuntos de dados que possuam uma considerável heterogeneidade espacial no efeito de alguns processos explicativos da variável resposta. Nesse caso, permitir que um ou mais parâmetros de regressão variem no espaço, resulta numa representação mais realista da variabilidade espacial. Exemplos de dados com esse tipo de comportamento podem ser encontrados em Assunção, Gamerman e Assunção (1999), onde os autores usam um modelo com parâmetros variando no espaço para descrever a variação da produtividade agrícola brasileira por micro região. Gamerman, Moreira e Rue (2003) propõem diferentes algoritmos de MCCM para fazer inferência sobre esses modelos e suas extensões multivariadas, e Huerta, Sansó e Stroud (2004) e Gelfand, Kim, Sirmans e Banerjee (2003) usam modelos com parâmetros variando no espaço para descrever a variação contínua de, respectivamente, níveis de ozônio na cidade do México e preços de mercado de residências familiares.

Neste capítulo analisamos níveis de poluição do ar medidos em  $PM_{10}$  (partículas inaláveis com diâmetro menor que  $10\mu g/m^3$ ) na Região Metropolitana do Rio de Janeiro, usando um modelo similar ao proposto por Gelfand et al. (2003), e extendendo o modelo proposto por Paez e Gamerman (2003) para análise do mesmo conjunto de dados. A análise exploratória de dados feita em Paez e Gamerman (2003) mostra que o efeito da temperatura máxima diária nos níveis de  $PM_{10}$  apresenta uma variação espacial significativa. O modelo aqui proposto assume que as observações se decompõem em uma função de regressão e um processo de erros não correlacionados, onde os parâmetros de regressão variam no espaço como campos aleatórios gaussianos independentes. O processo resultante é não estacionário tanto na média quanto na função de covariância, onde espaço e tempo interagem.

O objetivo desse capítulo engloba três pontos principais. Primeiramente descrevemos



como fazer interpolação dos parâmetros de regressão e variável resposta para o modelo proposto, levando em consideração dois possíveis cenários para o processos explicativos: quando eles são conhecidos para toda a região de interesse e quando eles são conhecidos apenas nos locais de amostragem. Em segundo lugar, investigamos como a inferência sobre os parâmetros e a performance da interpolação são afetados pela configuração dos locais de amostragem, tamanho da amostra, e diferentes especificações de priori. Finalmente, aplicamos o modelo proposto para descrever a variação espacial dos níveis de  $PM_{10}$  na Região Metropolitana do Rio de Janeiro.

Esse capítulo é organizado da seguinte forma: na seção 2.2 definimos o modelo e as distribuições a priori para os parâmetros desconhecidos, e descrevemos a metodologia para fazer interpolação dos parâmetros de regressão e da variável resposta sob dois diferentes cenários de observação dos procesos explicativos. Na seção 2.3 analisamos um conjunto de dados simulados e investigamos como a inferência sobre os parâmetros e a performance da interpolação são afetados pela configuração dos locais de amostragem, tamanho amostral e diferentes especificações a priori. Na seção 2.4 aplicamos o modelo proposto para a análise dos níveis de  $PM_{10}$ . Finalmente apresentamos as conclusões na seção 2.5.

## 2.2 O modelo

### 2.2.1 Definição do modelo

Seja  $Y(s, t)$  a variável resposta no local  $s \in D \subset R^2$  e tempo  $t \in E \subset Z$ , e suponha que os dados observados consistem em medições feitas em  $N$  regiões de amostragem  $\{s_1, s_2, \dots, s_N\}$ , e  $T$  tempos de registro  $\{t_1, t_2, \dots, t_T\}$ . Propomos aqui que a resposta seja modelada como um processo espaço-temporal  $\{Y(s, t); s \in D, t \in E\}$  definido pelas equações

$$Y(s, t) \mid \beta(s), \phi(t) \sim N(\mu(s, t), \sigma_e^2), \quad \sigma_e^2 > 0,$$

$$\mu(s, t) = \beta_0(s) + \beta_1(s)X_1(s, t) + \dots + \beta_p(s)X_p(s, t) + \phi(t) , \quad (2.1)$$

onde  $N(\mu, \sigma^2)$  denota a distribuição normal com média  $\mu$  e Variância  $\sigma^2$ ,  $X(s, t) = (X_1(s, t), \dots, X_p(s, t))$  são processos explicativos no espaço e no tempo, sendo a princípio supostamente conhecidos na região  $D \times E$ , e  $\beta(s) = (\beta_0(s), \dots, \beta_p(s))$  são processos de regressão espacial que descrevem os efeitos dos processos explicativos na variável resposta. Assume-se que os processos de regressão são independentes, onde para cada  $j = 0, 1, \dots, p$ ,  $\{\beta_j(s), s \in D\}$  é um processo Gaussiano com média  $\gamma_j \in R$  e função de covariância  $\sigma_j^2 \rho_j(s, u; \lambda_j)$ ,  $\sigma_j^2 > 0$ ,  $\lambda_j \in \Lambda_j \subset R^+$  e  $\rho_j(s, u; \lambda_j)$  é uma função de correlação em  $R^2$ ; denotada por

$$\beta_j(\cdot) \stackrel{ind}{\sim} \text{GP}(\gamma_j, \sigma_j^2 \rho_j(\cdot; \lambda_j)), \quad j = 0, 1, \dots, p . \quad (2.2)$$

O componente  $\phi(t)$  descreve uma possível dependência temporal nas observações que os efeitos dos processos explicativos não foram capazes de explicar. Assumimos que  $\phi(t)$  é um processo Gaussiano AR(1), independente do processo de regressão, e que satisfaça

$$\phi(t) = \delta \phi(t - 1) + w(t), \quad w(t) \mid \sigma_\phi^2 \stackrel{ind}{\sim} N(0, \sigma_\phi^2) , \quad (2.3)$$

onde  $\delta \in [0, 1)$ ,  $\sigma_\phi^2 > 0$  e  $t \in Z$ . Nesse modelo os efeitos das variáveis explicativas na variável resposta variam no espaço, mas para cada local do espaço esses efeitos são constantes no tempo. Finalmente, consideramos aqui funções de correlação isotrópicas em  $R^2$ ,  $\rho_j(d; \lambda_j)$ , tais como a esférica, a exponencial ou a exponencial ao quadrado, que dependem apenas de  $d = \|s - u\|$ , a distância Euclideana entre os locais  $s$  e  $u$ , e um único parâmetro que controla a razão de decaimento da correlação; veja Cressie (1993) para outros exemplos.

A estrutura de segunda ordem do processo  $Y(\cdot)$  é dada por

$$\begin{aligned} \text{cov}(Y(s, t), Y(s^*, t^*)) &= \sigma_0^2 \rho_0(d_s; \lambda_0) + \sigma_1^2 X_1(s, t) X_1(s^*, t^*) \rho_1(d_s; \lambda_1) + \dots + \\ &+ \sigma_p^2 X_p(s, t) X_p(s^*, t^*) \rho_p(d_s; \lambda_p) + \frac{\sigma_\phi^2}{1 - \delta^2} \delta^{d_t} , \end{aligned}$$

onde  $d_s = \|s - s^*\|$  e  $d_t = |t - t^*|$ . Ambas as funções de média e covariância são não estacionárias.

O modelo proposto (2.1) é uma extensão do modelo proposto por Paez e Gamerman (2003), onde se permite que ambos o intercepto e os efeitos dos processos explicativos variem no espaço. O modelo (2.1) também inclui o caso onde os efeitos de alguns dos processos explicativos variam espacialmente enquanto outros permanecem constantes. Nesse último caso, assume-se que os parâmetros do modelo de regressão satisfaçam  $\beta_j \sim N(\gamma_j, \sigma_j^2)$ .

Os processos explicativos podem apresentar uma variação substancial na região de interesse para um dado período de tempo. Em alguns casos esses processos são observados apenas nos locais de amostragem e nos tempos de registro. Assim sendo, assumimos que o modelo (2.1) seja válido condicionalmente aos processos explicativos, e, além disso, assumimos que para qualquer período de tempo  $t \in E$ , os processos espaciais  $X_j(\cdot, t)$ ,  $j = 0, 1, \dots, p$ , sejam processos Gaussianos independentes, e também independentes do processo de regressão e de  $\phi(\cdot)$ , isso é

$$X_j(\cdot, t) \stackrel{ind}{\sim} \text{GP}(\gamma_{x_j}, \sigma_{x_j}^2 \rho_{x_j}(\cdot; \lambda_{x_j})), \quad j = 1, \dots, p, \quad (2.4)$$

onde  $\gamma_{x_j} \in R$ ,  $\sigma_{x_j}^2 > 0$ ,  $\lambda_{x_j} \in \Lambda_{x_j} \subset R^+$  e  $\rho_{x_j}(\cdot; \lambda_{x_j})$  é uma função de correlação em  $R^2$ .

## 2.2.2 A priori

Vamos denotar os parâmetros do modelo (2.1) por  $\Psi' = (\boldsymbol{\gamma}', \boldsymbol{\sigma}^{2'}, \sigma_e^2, \boldsymbol{\lambda}', \sigma_\phi^2, \delta)$ , onde  $\boldsymbol{\gamma}' = (\gamma_0, \gamma_1, \dots, \gamma_p)$ ,  $\boldsymbol{\sigma}^{2'} = (\sigma_0^2, \sigma_1^2, \dots, \sigma_p^2)$  e  $\boldsymbol{\lambda}' = (\lambda_0, \lambda_1, \dots, \lambda_p)$ . Escolhemos trabalhar aqui com distribuições a priori próprias, evitando assim o risco de obter posterioris impróprias (Berger, De Oliveira and Sansó, 2001). Especificamos as seguintes prioris independentes:

$$\sigma_e^{-2} \sim G(\alpha_e, \nu_e), \quad \sigma_\phi^{-2} \sim G(\alpha_\phi, \nu_\phi), \quad \sigma_j^{-2} \sim G(\alpha_j, \nu_j), \quad \gamma_j \sim N(\mu_j, \tau_j^2),$$

$$\delta \sim U(0, 1), \quad \lambda_j \sim G(\alpha_{\lambda_j}, \nu_{\lambda_j}), \quad j = 0, 1, \dots, p. \quad (2.5)$$

onde  $G(a, b)$  denota a distribuição gama com média  $a/b$  e variância  $a/b^2$ , e  $U(a, b)$  denota a distribuição uniforme no intervalo  $(a, b)$ . Para os parâmetros de correlação  $\boldsymbol{\lambda}'$ , também consideramos a priori  $\lambda_j \sim U(0, c), j = 1, \dots, p$ , como uma especificação não-informativa. Nesse trabalho consideramos  $c = 1$ .

Os hiperparâmetros de algumas dessas prioris podem ser obtidos como se segue. Tendo fixado cada um dos locais de amostragem  $s_i$ , e supondo que não exista um componente autoregressivo, o modelo proposto determina que os dados variam como um modelo de regressão linear. Nesse caso, estimativas preliminares para  $\beta_0(s_i), \beta_1(s_i), \dots, \beta_p(s_i)$ , denotadas por  $\hat{\beta}_0(s_i), \hat{\beta}_1(s_i), \dots, \hat{\beta}_p(s_i), i = 1, \dots, N$ , podem ser obtidas por mínimos quadrados ordinários. Seja  $\hat{\boldsymbol{\beta}}_j = (\hat{\beta}_j(s_1), \hat{\beta}_j(s_2), \dots, \hat{\beta}_j(s_N))'$ , para  $j = 1, \dots, p$ . Estimativas preliminares para os parâmetros  $\gamma_j, \sigma_j^2$  e  $\lambda_j, j = 1, \dots, p$ , podem ser obtidas tratando  $\hat{\boldsymbol{\beta}}_j$  como se fossem dados observados com distribuição  $N(\gamma_j \mathbf{1}, \sigma_j^2 \boldsymbol{\Sigma}_j)$ , onde  $\mathbf{1} = (1, 1, \dots, 1)'$  e  $\boldsymbol{\Sigma}_j$  é  $N \times N$ ,  $(\boldsymbol{\Sigma}_j)_{lr} = \rho_j(s_r, s_l; \lambda_j)$ . Além disso, uma estimativa preliminar para  $\sigma_e^2$  pode ser obtida como a variância amostral do conjunto de todos os resíduos  $\{Y(s_i, t_j) - \hat{\mu}(s_i, t_j)\}$ , onde  $\hat{\mu}(s_i, t_j) = \hat{\beta}_0(s_i) + \hat{\beta}_1(s_i)X_1(s_i, t_j) + \dots + \hat{\beta}_p(s_i)X_p(s_i, t_j), i = 1, \dots, N, j = 1, \dots, T$ . Essas estimativas preliminares servem como locação para suas respectivas prioris em (2.5), enquanto o parâmetro de escala é determinado subjetivamente. Métodos similares foram aplicados, num contexto não-Bayesiano, por Eynon e Switzer (1983), Oehlert (1993) e Holland et al (2000), para obter estimativas de parâmetros em alguns modelos espaço-temporais. Para o parâmetro de precisão  $\sigma_\phi^{-2}$  usamos a priori não informativa  $\sigma_\phi^{-2} \sim G(10^{-6}, 10^{-6})$ .

Vamos denotar os parâmetros do modelo (2.4) por  $\boldsymbol{\Theta}' = (\boldsymbol{\theta}'_{x_1}, \dots, \boldsymbol{\theta}'_{x_p})$ , onde  $\boldsymbol{\theta}'_{x_j} = (\gamma_{x_j}, \sigma_{x_j}^2, \lambda_{x_j}), j = 1, \dots, p$ . No caso em que os processos explicativos são observados apenas nos locais de amostragem e tempos de registro, especificamos prioris independentes para esses parâmetros, dadas por

$$\gamma_{x_j} \sim N(\mu_{x_j}, \tau_{x_j}^2), \quad \sigma_{x_j}^{-2} \sim G(\alpha_{x_j}, \nu_{x_j}), \quad \lambda_{x_j} \sim U(0, 1), \quad j = 1, \dots, p,$$

onde os hiperparâmetros são escolhidos de forma a definir prioris vagas.

### 2.2.3 Interpolação

Estamos interessados em interpolar o processo resposta  $Y(\cdot, \cdot)$  para locais e períodos de tempo não observados, e também interpolar, para toda a região, os processos de regressão  $\beta_1(\cdot), \dots, \beta_p(\cdot)$ , que representam os efeitos dos processos explicativos.

Primeiramente introduzimos a notação necessária. Seja  $ST^{obs} = \{(s, t)\}$  o conjunto de pontos no espaço-tempo onde observações estão disponíveis,  $ST^{new} = \{(s, t)\}$  os pontos no espaço-tempo onde desejamos fazer interpolação,  $\mathbf{Y}^{obs} = (Y(s, t) : (s, t) \in ST^{obs})$ ,  $\mathbf{Y}^{new} = (Y(s, t) : (s, t) \in ST^{new})$ ,  $\mathbf{X}^{obs} = (X(s, t) : (s, t) \in ST^{obs})$ ,  $\mathbf{X}^{new} = (X(s, t) : (s, t) \in ST^{new})$ ,  $\boldsymbol{\beta}_k^{obs} = (\beta_k(s) : s \text{ é um local de amostragem})$  e  $\boldsymbol{\beta}_k^{new} = (\beta_k(s) : s \text{ é um local de predição})$ ,  $k = 1, \dots, p$ ,  $\boldsymbol{\beta}^{obs} = (\boldsymbol{\beta}_1^{obs}, \dots, \boldsymbol{\beta}_p^{obs})$  e  $\boldsymbol{\beta}^{new} = (\boldsymbol{\beta}_1^{new}, \dots, \boldsymbol{\beta}_p^{new})$ .

Agora descrevemos como interpolar os processos de regressão e de resposta sob dois cenários distintos: (A) quando os processos explicativos são conhecidos em  $ST^{obs} \cup ST^{new}$ , e (B) quando os processos explicativos são conhecidos apenas em  $ST^{obs}$ . Para ambos os cenários usamos a abordagem Bayesiana para fazer predições, que é baseada em amostrar das funções de densidade preditivas apropriadas.

*Cenário A.* A distribuição requerida para se fazer interpolação é  $\pi(\mathbf{Y}^{new} | \mathbf{X}^{obs}, \mathbf{X}^{new}, \mathbf{Y}^{obs})$ . Todas as distribuições que se seguem estão (implicitamente) condicionadas a  $(\mathbf{X}^{obs}, \mathbf{X}^{new})$ , que são ambos conhecidos. Inferência a respeito de  $\boldsymbol{\beta}^{new}$  é feita com base na amostra obtida da sua distribuição a posteriori.

Com essa finalidade, amostramos da distribuição a posteriori conjunta de  $\boldsymbol{\beta}^{new}$ ,  $\boldsymbol{\beta}^{obs}$  e  $\boldsymbol{\Psi}$ , que pode ser escrita como

$$\begin{aligned} \pi(\boldsymbol{\beta}^{new}, \boldsymbol{\beta}^{obs}, \boldsymbol{\Psi} | \mathbf{Y}^{obs}) &= \pi(\boldsymbol{\beta}^{new} | \boldsymbol{\beta}^{obs}, \boldsymbol{\Psi}, \mathbf{Y}^{obs}) \pi(\boldsymbol{\beta}^{obs}, \boldsymbol{\Psi} | \mathbf{Y}^{obs}) \\ &= \pi(\boldsymbol{\beta}^{new} | \boldsymbol{\beta}^{obs}, \boldsymbol{\Psi}) \pi(\boldsymbol{\beta}^{obs}, \boldsymbol{\Psi} | \mathbf{Y}^{obs}), \end{aligned} \quad (2.6)$$

visto que  $\beta^{new}$  e  $\mathbf{Y}^{obs}$  são condicionalmente independentes dados  $(\beta^{obs}, \Psi)$ . Para amostrar de (2.6), primeiramente amostramos de  $\pi(\beta^{obs}, \Psi | \mathbf{Y}^{obs})$  baseado no fato do modelo (2.1) ser um modelo linear Gaussiano hierárquico, e então podemos simular de  $\pi(\beta^{new} | \beta^{obs}, \Psi)$  que é multivariado Gaussiano. Os valores de  $\beta^{new}$  são amostras de  $\pi(\beta^{new} | \mathbf{Y}^{obs})$ , e podem ser usados para fazer inferência sobre os efeitos dos processos explicativos, através de, por exemplo, sua média e variância amostrais.

A interpolação do processo resposta, isso é, inferência sobre  $\mathbf{Y}^{new}$ , é feita de forma similar, baseada em amostras da sua distribuição preditiva. Com esse objetivo, amostramos da distribuição conjunta a posteriori de  $\mathbf{Y}^{new}$ ,  $\beta^{new}$ ,  $\beta^{obs}$  e  $\Psi$ , que pode ser escrita como

$$\begin{aligned} \pi(\mathbf{Y}^{new}, \beta^{new}, \beta^{obs}, \Psi | \mathbf{Y}^{obs}) &= \pi(\mathbf{Y}^{new} | \beta^{new}, \beta^{obs}, \Psi, \mathbf{Y}^{obs})\pi(\beta^{new}, \beta^{obs}, \Psi | \mathbf{Y}^{obs}) \\ &= \pi(\mathbf{Y}^{new} | \beta^{new}, \Psi)\pi(\beta^{new}, \beta^{obs}, \Psi | \mathbf{Y}^{obs}) \end{aligned} \quad (2.7)$$

pois  $\mathbf{Y}^{new}$  e  $(\beta^{new}, \mathbf{Y}^{obs})$  são condicionalmente independentes dados  $\beta^{obs}$  e  $\Psi$ . Para amostrar de (2.7), para cada  $(\beta^{new}, \beta^{obs}, \Psi)$  que foram amostrados anteriormente de (2.6), amostramos de  $\pi(\mathbf{Y}^{new} | \beta^{new}, \Psi)$  que é Normal multivariada. Os valores de  $\mathbf{Y}^{new}$  são uma amostra de  $\pi(\mathbf{Y}^{new} | \mathbf{Y}^{obs})$  e podem ser usados para fazer interpolação dos processos resposta, por exemplo, através de suas médias e variâncias amostrais.

*Scenario B.* Aqui a distribuição requerida para fazer interpolação é  $\pi(\mathbf{Y}^{new} | \mathbf{X}^{obs}, \mathbf{Y}^{obs})$ . Assim sendo, primeiramente geramos uma amostra da distribuição a posteriori de  $\mathbf{X}^{new}$  e  $\Theta$ , o que é necessário para se fazer inferência sobre os processos de regressão e resposta. Usando o modelo (2.4), a distribuição a posteriori conjunta de  $\mathbf{X}^{new}$  and  $\Theta$  é dada por

$$\begin{aligned} \pi(\mathbf{X}^{new}, \Theta | \mathbf{X}^{obs}, \mathbf{Y}^{obs}) &= \pi(\mathbf{X}^{new} | \Theta, \mathbf{X}^{obs}, \mathbf{Y}^{obs})\pi(\Theta | \mathbf{X}^{obs}, \mathbf{Y}^{obs}) \\ &= \pi(\mathbf{X}^{new} | \Theta, \mathbf{X}^{obs})\pi(\Theta | \mathbf{X}^{obs}), \end{aligned} \quad (2.8)$$

pois  $\mathbf{X}^{new}$  e  $\mathbf{Y}^{obs}$  são condicionalmente independentes dado  $\Theta$ , e  $\Theta$  e  $\mathbf{Y}^{obs}$  são condicionalmente independentes dado  $\mathbf{X}^{obs}$ . Para amostrar dessa distribuição primeiramente simulamos de  $\pi(\Theta | \mathbf{X}^{obs})$  usando um algoritmo de MCCM, e então simulamos de  $\pi(\mathbf{X}^{new} | \Theta, \mathbf{X}^{obs})$  que é Normal multivariada.

Tendo uma amostra da distribuição a posteriori de  $(\mathbf{X}^{new}, \Theta)$ , essencialmente seguimos o mesmo procedimento descrito no cenário A. A distribuição a posteriori conjunta de  $\mathbf{Y}^{new}$ ,  $\beta^{new}$ ,  $\beta^{obs}$ ,  $\Psi$ ,  $\mathbf{X}^{new}$  e  $\Theta$  pode ser escrita como

$$\begin{aligned}
& \pi(\mathbf{Y}^{new}, \beta^{new}, \beta^{obs}, \Psi, \mathbf{X}^{new}, \Theta \mid \mathbf{Y}^{obs}, \mathbf{X}^{obs}) = \\
& \pi(\mathbf{Y}^{new} \mid \beta^{new}, \beta^{obs}, \Psi, \mathbf{X}^{new}, \Theta, \mathbf{Y}^{obs}, \mathbf{X}^{obs}) \pi(\beta^{new}, \beta^{obs}, \Psi \mid \mathbf{X}^{new}, \Theta, \mathbf{Y}^{obs}, \mathbf{X}^{obs}) \\
& \quad \times \pi(\mathbf{X}^{new}, \Theta \mid \mathbf{Y}^{obs}, \mathbf{X}^{obs}) = \\
& \pi(\mathbf{Y}^{new} \mid \beta^{new}, \Psi, \mathbf{X}^{new}) \pi(\beta^{new}, \beta^{obs}, \Psi \mid \mathbf{X}^{obs}, \mathbf{Y}^{obs}) \pi(\mathbf{X}^{new}, \Theta \mid \mathbf{X}^{old}, \mathbf{Y}^{obs}) \quad (2.9)
\end{aligned}$$

usando relações de independência condicional similares às usadas no cenário A.

Para amostrar de (2.9), para cada  $(\mathbf{X}^{new}, \Theta)$  que foi previamente amostrado de (2.8), amostramos de  $\pi(\beta^{new}, \beta^{obs}, \Psi \mid \mathbf{X}^{obs}, \mathbf{Y}^{obs})$  usando o mesmo algoritmo que foi usado para amostrar de (2.6), e depois simulamos de  $\pi(\mathbf{Y}^{new} \mid \beta^{new}, \Psi, \mathbf{X}^{new})$  que é uma Normal multivariada. Os valores de  $\beta^{new}$  e  $\mathbf{Y}^{new}$  são amostras, respectivamente, de  $\pi(\beta^{new} \mid \mathbf{X}^{obs}, \mathbf{Y}^{obs})$  e  $\pi(\mathbf{Y}^{new} \mid \mathbf{X}^{obs}, \mathbf{Y}^{obs})$ , que são usadas para fazer inferência sobre os processos de regressão e resposta.

## 2.3 Trabalhando com dados simulados

Nessa seção utilizamos dados simulados para estudar a influência da quantidade de informação disponível nos dados, e de diferentes especificações de priors, na performance da inferência sobre os parâmetros do modelo e interpolação espacial.

Simulamos uma realização dos processos definidos pelas equações (2.1) e (2.2) com  $p = 1$  (apenas um processo explicativo) sob uma grade regular  $51 \times 51$  na região  $D = [0, 1]^2$  ( $N = 2601$ ) e para  $T = 30$  pontos igualmente espaçados no tempo. Assumimos  $\sigma_\epsilon^2 = 1$ , e

$$\beta_0(\cdot) \sim \text{GP}(100, 10\rho^E(\cdot; 0.4)) \quad \text{e} \quad \beta_1(\cdot) \sim \text{GP}(5, 0.1\rho^E(\cdot; 0.8)),$$

com  $\rho^E(d; \lambda)$  sendo a função de correlação exponencial isotrópica

$$\rho^E(d; \lambda) = \exp(-\lambda d). \quad (2.10)$$

Os valores do processo explicativo foram simulados independentemente no tempo, sendo que para cada  $t = 1, \dots, T$ ,

$$X(\cdot, t) \sim \text{GP}(10, 0.15\rho^E(\cdot; 1.5)).$$

A Figura 2.1 mostra os valores simulados dos processos  $\beta_0(\cdot), \beta_1(\cdot)X(\cdot, t), e(\cdot, t) \stackrel{iid}{\sim} N(0, \sigma_e^2)$ , e  $Y(\cdot, t)$  ao longo da região  $D = [0, 1]^2$  para o período de tempo fixo  $t = 30$ .

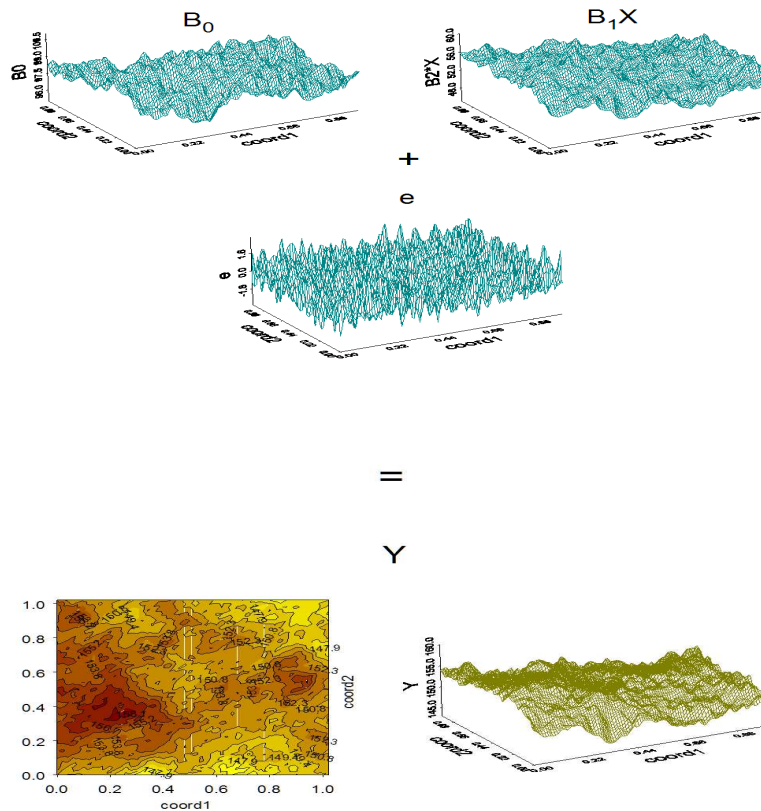


Figura 2.1: Superfícies Simuladas dos componentes do modelo (1) para  $t=30$

A inferência será feita baseada em quatro conjuntos de dados, onde cada um deles



consiste em observações da realização acima em todos os períodos de tempo, e uma grade de regiões amostrais (ver figura 2.2), obtidas como se segue:

*dados 1*: observações selecionadas em 25 locais de uma grade regular em  $[0, 1]^2$ ;

*dados 2*: observações selecionadas aleatoriamente em 25 locais em  $[0, 1]^2$ ;

*dados 3*: observações selecionadas em 81 locais de uma grade regular em  $[0, 1]^2$ ;

*dados 4*: observações selecionadas aleatoriamente em 81 locais em  $[0, 1]^2$ .

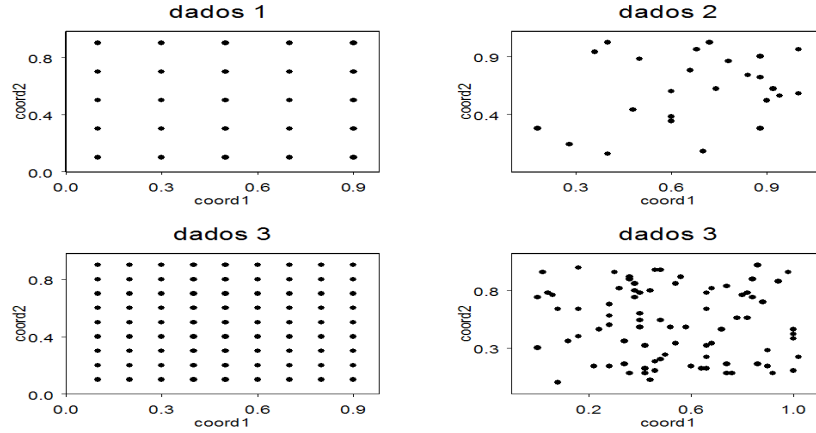


Figura 2.2: Locais de amostragem no espaço para os quatro conjuntos de dados

Consideramos também 3 especificações de distribuição a priori para os parâmetros, levando em conta diferentes graus de informação:

*Priori 1.* Nessa especificação utilizamos a priori em (2.5), escolhendo os hiperparâmetros de forma a obter priors difusas

$$\sigma_e^{-2}, \sigma_0^{-2}, \sigma_1^{-2} \stackrel{ind}{\sim} G(0.001, 0.001), \quad \gamma_0, \gamma_1 \stackrel{ind}{\sim} N(0.0, 0.00001), \quad \lambda_0, \lambda_1 \stackrel{ind}{\sim} U(0, 1). \quad (2.11)$$

Vamos nos referir à essa especificação como priors não informativas (PNI).

*Priori 2.* Aqui utilizamos a mesma especificação da priori 1 para  $\sigma_e^{-2}$ ,  $\gamma_0$  e  $\gamma_1$ , mas adotamos distribuições a priori gamma para  $\sigma_0^{-2}$ ,  $\sigma_1^{-2}$ ,  $\lambda_0$  e  $\lambda_1$ , com parâmetro de forma 1 e (inverso do) parâmetro de escala igual às estimativas preliminares obtidas como descrito

na seção 2.2.2. Assim sendo, essas prioris estão centradas nas estimativas preliminares  $\hat{\sigma}_0^{-2} = 0.07$ ,  $\hat{\sigma}_1^{-2} = 6.25$ ,  $\hat{\lambda}_0 = 0.35$  e  $\hat{\lambda}_1 = 0.70$ . Especificamente temos que,

$$\sigma_0^{-2} \sim G(1, 14.50), \quad \sigma_1^{-2} \sim G(1, 0.16), \quad \lambda_0 \sim G(1, 2.86) \quad \text{and} \quad \lambda_1 \sim G(1, 1.43). \quad (2.12)$$

Vamos nos referir a essa especificação como prioris pouco informativas (PPI).

*Priori 3.* As prioris para  $\sigma_1^{-2}$ ,  $\sigma_2^{-2}$ ,  $\lambda_1$  e  $\lambda_2$  são especificadas de forma similar à da Priori 2, mas aqui essas distribuições estão mais concentradas em torno das suas estimativas preliminares. O parâmetro de forma para essas prioris é igual a 10 e o (inverso do) parâmetro de escala é o mesmo da Priori 2 multiplicado por 10, de forma a manter a mesma média. Especificamente, temos

$$\sigma_1^{-2} \sim G(10, 145), \quad \sigma_2^{-2} \sim G(10, 1.6), \quad \lambda_1 \sim G(10, 28.57) \quad \text{e} \quad \lambda_2 \sim G(10, 14.28) \quad (2.13)$$

Vamos nos referir a essa especificação como prioris muito informativas (PMI).

## 2.3.1 Resultados

### Inferência sobre os Parâmetros do Modelo

As tabelas 2.1, 2.2, 2.3 e 2.4 mostram um resumo da inferência para os parâmetros do modelo, baseada, respectivamente, nos conjuntos de dados 1, 2, 3 e 4, e considerando cada uma das especificações a priori. Essa inferência é baseada em amostras obtidas através do BUGS.<sup>1</sup>

---

<sup>1</sup>BUGS (Bayesian inference Using Gibbs Sampling) é um software computacional para a análise Bayesiana de modelos estatísticos complexos usando métodos de Monte Carlo em Cadeias de Markov (MCCM). Ele pode ser obtido na homepage <http://www.mrc-bsu.cam.ac.uk/bugs>

Tabela 2.1: Resumo da inferência a posteriori dos parâmetros baseada no conjunto de dados 1

parâmetro	valor real	PNI			PPI			PMI		
		média	mediana	d.p.	média	mediana	d.p.	média	mediana	d.p.
$\gamma_0$	100	97.40	97.39	0.79	97.90	98.34	2.35	98.24	97.94	1.74
$\gamma_1$	5	5.56	5.60	0.17	5.50	5.52	0.19	5.41	5.41	0.20
$\sigma_e^{-2}$	1	0.94	0.94	0.05	0.94	0.94	0.94	0.05	0.94	0.05
$\sigma_0^{-2}$	0.1	173.70	23.33	372.90	0.10	0.08	0.07	0.07	0.07	0.02
$\sigma_1^{-2}$	10	4.08	3.90	2.10	8.34	7.12	5.57	6.57	6.39	1.77
$\lambda_0$	0.4	0.49	0.49	0.29	0.33	0.23	0.33	0.36	0.34	0.11
$\lambda_1$	0.8	0.63	0.66	0.26	0.95	0.79	0.71	0.71	0.69	0.20

Tabela 2.2: Resumo da inferência a posteriori dos parâmetros baseada no conjunto de dados 2

parâmetro	valor real	PNI			PPI			PMI		
		média	mediana	d.p.	média	mediana	d.p.	média	mediana	d.p.
$\gamma_0$	100	99.30	99.25	1.10	97.86	97.97	2.50	99.16	99.13	1.26
$\gamma_1$	5	5.83	5.84	0.13	5.48	5.45	0.17	5.42	5.44	0.17
$\sigma_e^{-2}$	1	1.09	1.09	0.06	1.09	1.09	0.06	1.09	1.09	0.06
$\sigma_0^{-2}$	0.1	47.35	0.34	199.40	0.08	0.06	0.06	0.07	0.06	0.02
$\sigma_1^{-2}$	10	3.23	2.80	2.39	6.67	5.43	5.04	6.07	5.91	1.67
$\lambda_0$	0.4	0.52	0.53	0.28	0.39	0.29	0.34	0.38	0.36	0.12
$\lambda_1$	0.8	0.51	0.50	0.27	0.88	0.70	0.71	0.76	0.73	0.21

Tabela 2.3: Resumo da inferência a posteriori dos parâmetros baseada no conjunto de dados 3

parâmetro	valor real	PNI			PPI			PMI		
		média	mediana	d.p.	média	mediana	d.p.	média	mediana	d.p.
$\gamma_0$	100	124.20	124.40	0.94	122.90	122.60	2.31	129.70	129.10	3.18
$\gamma_1$	5	2.79	2.75	0.30	2.55	2.59	0.25	2.49	2.54	0.22
$\sigma_e^{-2}$	1	0.24	0.24	0.01	0.24	0.24	0.01	0.24	0.24	0.01
$\sigma_0^{-2}$	0.1	39.45	3.34	153.30	0.11	0.10	0.07	0.07	0.07	0.02
$\sigma_1^{-2}$	10	6.56	5.48	6.62	7.81	6.71	5.06	7.26	7.10	1.79
$\lambda_0$	0.4	0.55	0.56	0.27	0.25	0.15	0.28	0.31	0.30	0.09
$\lambda_1$	0.8	0.57	0.59	0.26	0.77	0.64	0.54	0.62	0.60	0.17

Tabela 2.4: Resumo da inferência a posteriori dos parâmetros baseada no conjunto de dados 4

parâmetro	valor real	PNI			PPI			PMI		
		média	mediana	d.p.	média	mediana	d.p.	média	mediana	d.p.
$\gamma_0$	100	95.96	95.86	0.96	96.08	95.87	1.66	96.33	96.39	1.17
$\gamma_1$	5	5.65	5.64	0.09	5.57	5.56	0.14	5.73	5.73	0.06
$\sigma_e^{-2}$	1	1.01	1.01	0.03	1.01	1.01	0.03	1.01	1.01	0.03
$\sigma_0^{-2}$	0.1	26.10	0.72	152.20	0.12	0.10	0.08	0.08	0.08	0.02
$\sigma_1^{-2}$	10	7.91	7.32	5.08	8.97	7.77	5.68	7.60	7.44	1.81
$\lambda_0$	0.4	0.53	0.54	0.28	0.24	0.16	0.26	0.30	0.30	0.09
$\lambda_1$	0.8	0.58	0.61	0.26	0.64	0.54	0.45	0.54	0.52	0.14

Analisando as tabelas acima chegamos as seguintes conclusões:

- Os parâmetros  $\gamma_0, \gamma_1,$  e  $\sigma_e^{-2}$  são razoavelmente bem estimados em todos os casos, com exceção do conjunto de dados 3. Para nenhum conjunto de dados a estimativa desses parâmetros foi sensível à priori.

- A inferência sobre  $\sigma_0^{-2}$  considerando a especificação PNI é extremamente pobre (ao ponto de ser inútil), de forma que essa especificação a priori se mostra inconveniente. Apesar de  $\sigma_0^{-2}$  ser subestimado nas especificações PPI e PMI, ambas levam a melhores inferências a respeito de  $\sigma_0^{-2}$  que PNI, sendo que PPI apresenta resultados ligeiramente superiores. Conclusões similares podem ser tiradas na inferência sobre  $\sigma_1^{-2}$ . A figura 2.3 apresenta histogramas da posteriori de  $\sigma_0^{-2}$  sob as três especificações usando o conjunto de dados 1, e também os histogramas da posteriori de  $\sigma_0^2$ . Através dessa figura podemos enxergar com mais clareza o quão pobre é a estimação de  $\sigma_0^2$  sob a especificação PNI, pois mostra que o valor real 10 está longe de ser incluído no intervalo de credibilidade de 95%.
- Os parâmetros de correlação  $\lambda_0$  e  $\lambda_1$  são os mais difíceis de se estimar, e para todos os conjuntos de dados e prioris há uma grande incerteza a respeito deles. As estimativas obtidas considerando as especificações PPI e PMI são um pouco superioris àquelas obtidas sob a especificação PNI. Os histogramas das amostras das distribuições a posteriori marginais de  $\lambda_0$  e  $\lambda_1$ , que podem ser vistos na Figura 2.4, revelam que quando a especificação PNI é usada há uma substancial concentração da densidade da função de probabilidade perto do limite de  $\lambda = 1$ , sugerindo que restringir o espaço paramétrico ao intervalo  $[0, 1]$  não é uma escolha apropriada nesse caso.

Inferência sobre os parâmetros foram obtidas utilizando outras duas realizações do mesmo processo. Obtivemos resultados similares aos apresentados aqui para uma delas. Para a outra realização obtivemos melhores resultados na estimação de  $\sigma_0^{-2}$  sob a especificação PNI. Em contrapartida, os resultados da estimação de  $\sigma_1^{-2}$  foram bastante pobres.

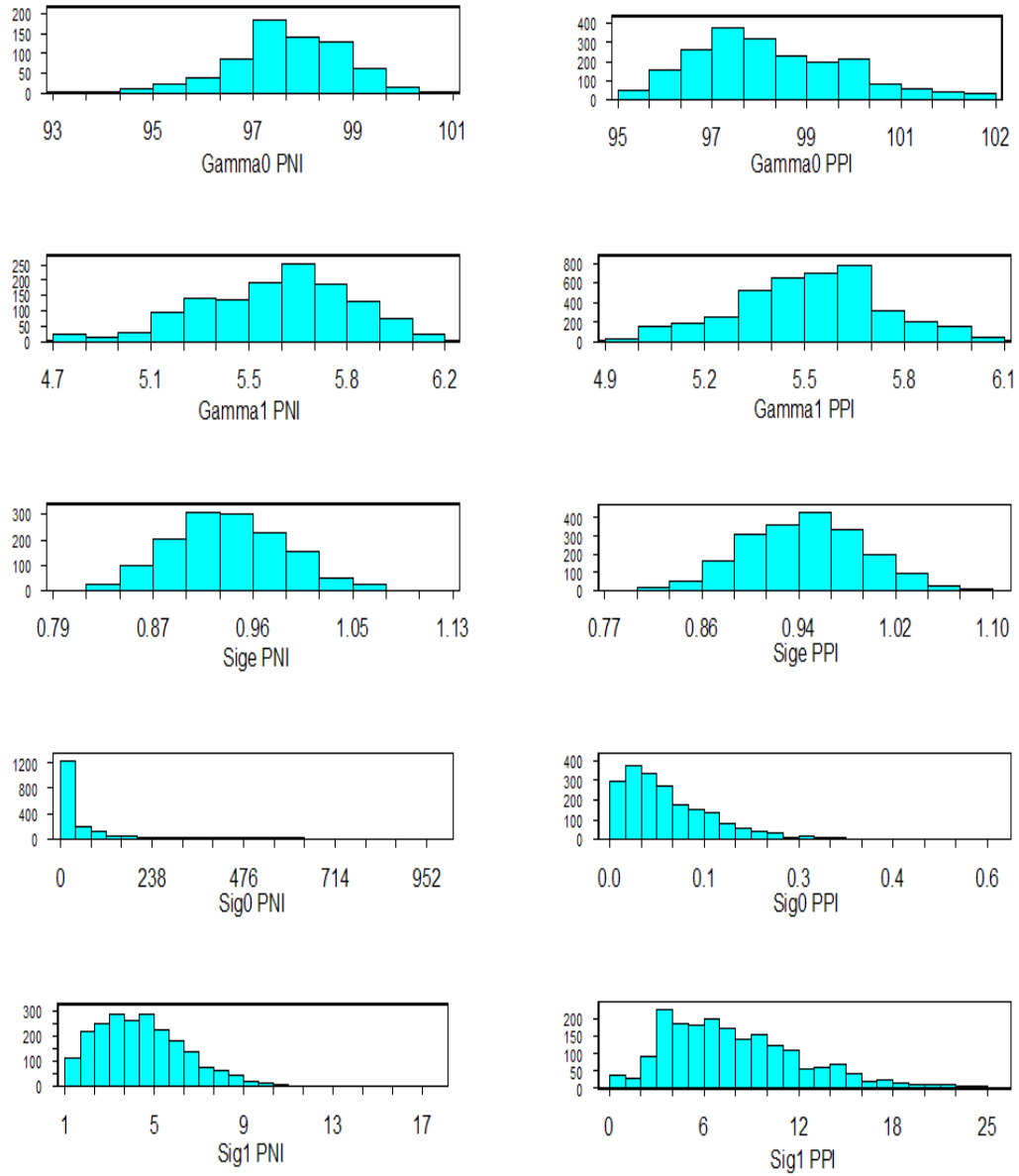


Figura 2.3: Histograma da porteriori de  $\gamma_0, \gamma_1, \sigma_e^{-2}, \sigma_0^{-2}$  e  $\sigma_1^{-2}$  considerando as especificações a priori PNI e PPI.

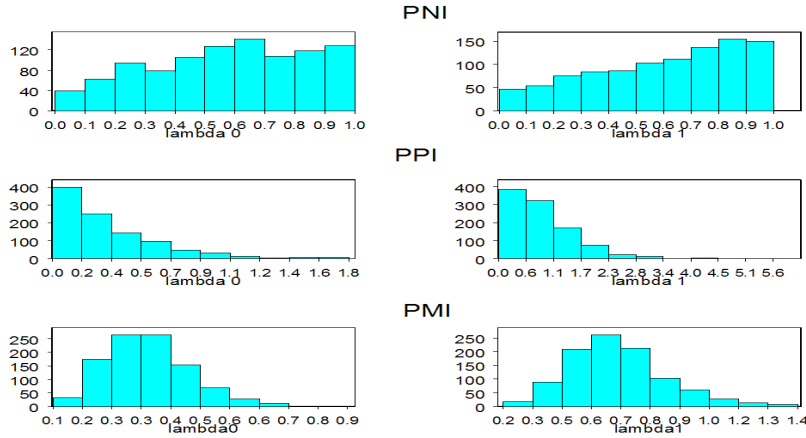


Figura 2.4: Histograma da posteriori de  $\lambda_0$  e  $\lambda_1$  baseada nas três especificações a priori.

## Resultados da Interpolação Espacial

Com base no modelo proposto e na metodologia descrita na seção 2.2.3, investigamos nessa subseção a performance da interpolação espacial para ambos os processos de regressão e resposta, analisando a dependência dos resultados nos diferentes conjuntos de dados e especificações a priori. Essa análise é feita para o período de tempo fixo em  $t = 30$  e todos os pontos localizados na grade original de  $51 \times 51$ , sob ambos os cenários A e B. Por razões computacionais restringimos a maior parte da análise aos conjuntos de dados 1 e 2 (ambos de tamanho 25) e especificações a priori PNI e PPI. Os parâmetros foram estimados pela média de suas respectivas densidades a posteriori, e essas estimativas serão utilizadas nessa seção e na seção 2.4 para construir superfícies de interpolação.

Primeiramente consideramos o cenário A, onde  $X(\cdot, \cdot)$  é conhecido em todos os pontos da grade e períodos de tempo. A Figura 2.5 mostra o resultado obtido pela interpolação dos processos de regressão  $\beta_0(\cdot)$  e  $\beta_1(\cdot)$ , baseado no conjunto de dados 1 e prioris PNI e PPI.

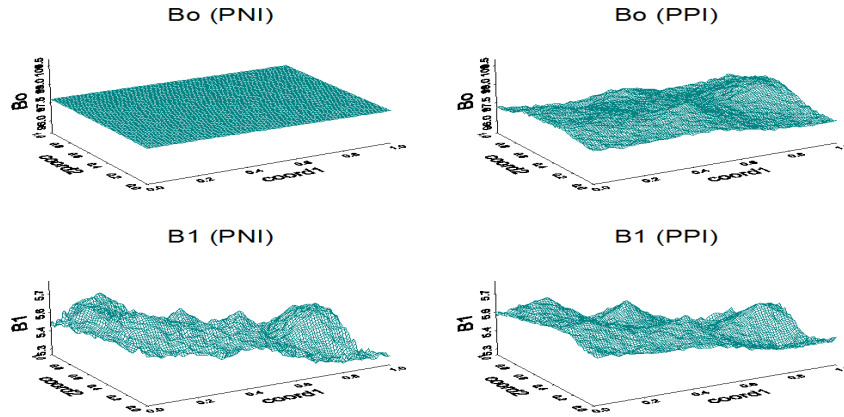


Figura 2.5: Superfícies interpoladas de  $\beta_0(\cdot)$  e  $\beta_1(\cdot)$  sob as especificações PPI e PNI.

Quando utilizamos a especificação PNI, a superfície interpolada de  $\beta_0(\cdot)$  é próxima de uma constante, visto que  $\sigma_0^{-2}$  é severamente superestimado. Esse problema não ocorre quando utilizamos PPI, que dá origem a uma superfície interpolada próxima à verdadeira, como pode ser visto pela comparação das Figuras 2.1 e 2.5. As superfícies interpoladas de  $\beta_1(\cdot)$  considerando as duas especificações de priori estão próximas uma da outra, sendo a que corresponde à priori PNI mais variável, visto que  $\sigma_1^{-2}$  foi subestimado nesse caso.

Baseado no conjunto de dados 2,  $\sigma_0^{-2}$  foi menos superestimado e  $\sigma_1^{-2}$  foi mais subestimado (comparado ao conjunto de dados 1) considerando as prioris PNI, fazendo com que ambas as superfícies estimadas  $\beta_0(\cdot)$  e  $\beta_1(\cdot)$  mostrem mais variabilidade. Quando a priori PPI é utilizada, as superfícies estimadas são similares às obtidas anteriormente, porém com menos variabilidade, ficando mais próximas às superfícies verdadeiras.

A Figura 2.6 mostra as superfícies interpoladas de  $Y(\cdot, 30)$  para ambas as prioris sob o conjunto de dados 1. Podemos verificar que elas estão próximas uma da outra, mostrando que o resultado da interpolação não é muito sensível à especificação a priori. O mesmo ocorre com o conjunto de dados 2.



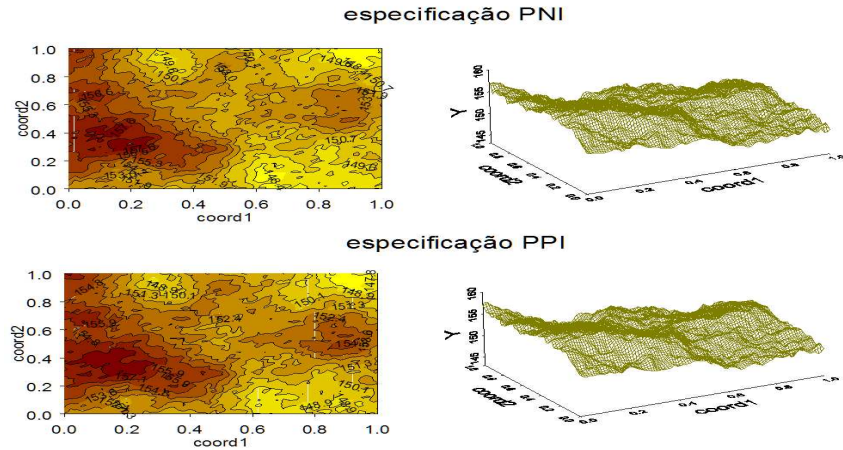


Figura 2.6: Superfície interpolada de  $Y(\cdot, 30)$ .

Para quantificar a "distância" entre a superfície verdadeira e a interpolada de um parâmetro  $M$  em uma grade  $G$  com  $m$  pontos, definimos a estatística  $DS$  (distância em superfície), que nos dá a soma de quadrados média da diferença entre as medidas de ambas as superfícies.

$$DS(M) = \frac{\sum_{i=1}^m (\mu_i - \mu'_i)^2}{m},$$

onde  $\mu'_i$  e  $\mu_i$  são respectivamente os valores verdadeiros e interpolados em  $i$ ,  $i \in G$ .

Na tabela 2.5 temos os valores obtidos pela estatística  $DS$  aplicada a  $\beta_0(\cdot)$ ,  $\beta_1(\cdot)$  e  $Y(\cdot, 30)$ . Quando aplicada a  $\beta_0(\cdot)$  e  $\beta_1(\cdot)$ , a estatística apresenta menores valores quando o conjunto de dados 2 é utilizado, o que pode ser um simples efeito da amostra selecionada. Quando aplicada a  $Y(\cdot, 30)$ , melhores resultados foram obtidos usando o conjunto de dados 1, e considerando as especificações PNI e PMI. Comparando as especificações a priori PNI, PPI e PMI vimos que na maioria das vezes a especificação PPI levou a melhores valores da estatística  $DS$ .

Tabela 2.5: Estatística  $DS$  para  $\beta_0(\cdot)$ ,  $\beta_1(\cdot)$ , e  $Y(\cdot, 30)$  (desvio padrão estimado entre parêntesis)

parâmetro	dados 1-PNI	dados 1-PPI	dados 1-PMI	dados 2-PNI	dados 2-PPI	dados 2-PMI
$\beta_0(\cdot)$	1.248 (0.063)	1.121 (0.060)	2.668 (0.071)	0.866 (0.052)	0.548 (0.043)	1.620 (0.053)
$\beta_1(\cdot)$	0.024 (0.013)	0.023 (0.011)	0.040 (0.014)	0.014 (0.012)	0.012 (0.013)	0.024 (0.010)
$Y(\cdot, 30)$	2.478 (0.067)	2.258 (0.050)	2.144 (0.067)	2.385 (0.068)	2.342 (0.087)	2.864 (0.047)

A critério de comparação, a estatística  $DS(Y(\cdot, 30))$  também foi computada para a interpolação feita sob um modelo com parâmetros fixos  $\beta_0$  e  $\beta_1$ , usando o conjunto de dados 1. Como resultado obtivemos  $DS(Y) = 6.49 (0.2)$ , que é consideravelmente maior que o valor obtido usando o modelo aqui proposto, o que mostra a importância de permitir que os coeficientes variem no espaço.

Considere agora o cenário B onde  $X(\cdot, \cdot)$  é observado apenas nos locais amostrais e tempos de registro. Suponha que esse processo siga o modelo (2.4), e seus parâmetros tenham prioris independentes e não-informativas dadas por

$$\gamma_x \sim N(0, 0.00001), \quad \sigma_x^{-2} \sim G(0.001, 0.001), \quad \lambda_x \sim U(0, 1) .$$

A Tabela 2.6 sumariza a inferência feita para os parâmetros  $\theta'_x = (\gamma_x, \sigma_x^{-2}, \lambda_x)$  baseada nos quatro conjuntos de dados. Os resultados mostram uma melhora significativa na precisão quando comparados com a inferência sobre parâmetros similares no modelo (2.1):  $\sigma_0^2$ ,  $\sigma_1^2$ ,  $\lambda_0$  e  $\lambda_1$ . Isso se explica pelo fato de  $X(\cdot, \cdot)$  ter sido replicado  $T$  vezes, ao contrário do processo  $Y(\cdot, \cdot)$ .

Tabela 2.6: Sumário da inferência a posteriori dos parâmetros do processo  $X(\cdot, \cdot)$  baseado nos quatro conjuntos de dados

		dados 1	dados 2	dados 3	dados 4
parâmetro	valor verdadeiro	média (d.p.)	média (d.p.)	média (d.p.)	mean (s.d.)
$\gamma_x$	10	9.988 (0.003)	9.995 (0.002)	9.985 (0.002)	9.980 (0.003)
$\sigma_x^{-2}$	6.67	6.499 (0.828)	7.021 (0.903)	6.630 (0.779)	6.398 (0.790)
$\lambda_x$	1.5	1.476 (0.211)	1.596 (0.231)	1.507 (0.186)	1.408 (0.182)

O padrão amostral (regular ou aleatório) aparentemente tem pouca influência na precisão das estimativas dos parâmetros, e apenas uma melhora moderada nas estimativas da precisão de  $\sigma_x^{-2}$  e  $\lambda_x$  é obtida com o aumento do tamanho amostral de 25 para 81. Entretanto, não há nenhuma melhora aparente na estimação de  $\gamma_x$  com o aumento do tamanho amostral, o que pode ser causado pelo fato de  $\gamma_x$  estar bem estimado com 30 replicações em 25 locais.

A Figura 2.7 mostra a superfície verdadeira de  $X(\cdot, 30)$ , enquanto na Figura 2.8 temos a superfície interpolada de  $X(\cdot, 30)$  baseada nos conjuntos de dados 1 e 2. Uma comparação visual entre essas superfícies mostra que a interpolação baseada na amostra com padrão espacial regular é bem melhor que a baseada na amostra com padrão espacial aleatório. Essa superioridade é confirmada pela estatística  $DS$  obtida com base nesses dois conjuntos de dados, sendo  $DS(X(\cdot, 30)) = 0.028$  considerando o conjunto de dados 1, e  $DS(X(\cdot, 30)) = 0.035$ , o conjunto de dados 2, sendo o desvio padrão da estatística  $SD$  aproximadamente igual a 0.002 nos dois casos.

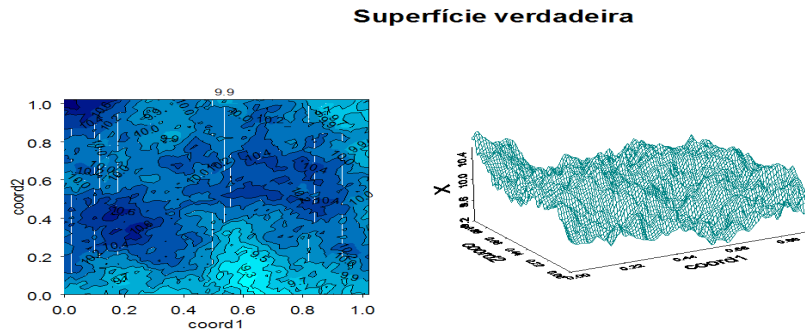


Figura 2.7: Superfície verdadeira de  $X(\cdot, 30)$

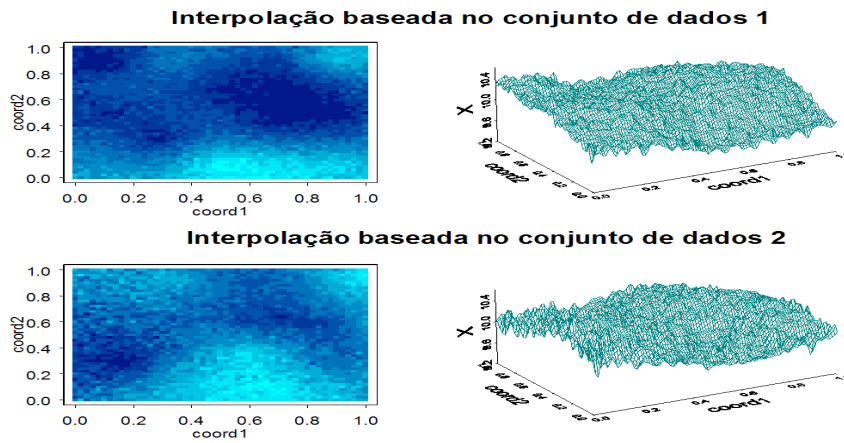


Figura 2.8: Superfície interpolada de  $X(\cdot, 30)$  baseada nos conjuntos de dados 1 e 2

A Figura 2.9 apresenta o resultado obtido na interpolação do processo  $Y(\cdot, 30)$  baseada nos conjuntos de dados 1 e 2, e priori PPI. A superfície estimada com base no conjunto de dados 1 é próxima àquelas apresentadas na Figura 2.6, que correspondem à interpolação feita quando  $X(\cdot, \cdot)$  é conhecido. As superfícies de Figura 2.6 são, entretanto, menos suaves que a da Figura 2.9. Como já era esperado, a estatística  $DS$  obtida para  $Y(\cdot, 30)$  é maior no caso de  $X$  desconhecido, sendo  $DS(Y(\cdot, 30))=3.133$  considerando o conjunto de dados 1, e  $DS(Y(\cdot, 30))=3.226$  considerando o conjunto de dados 2. Os desvios padrões estimados da estatística  $DS$  foram respectivamente 0.07 e 0.11. A superfície estimada

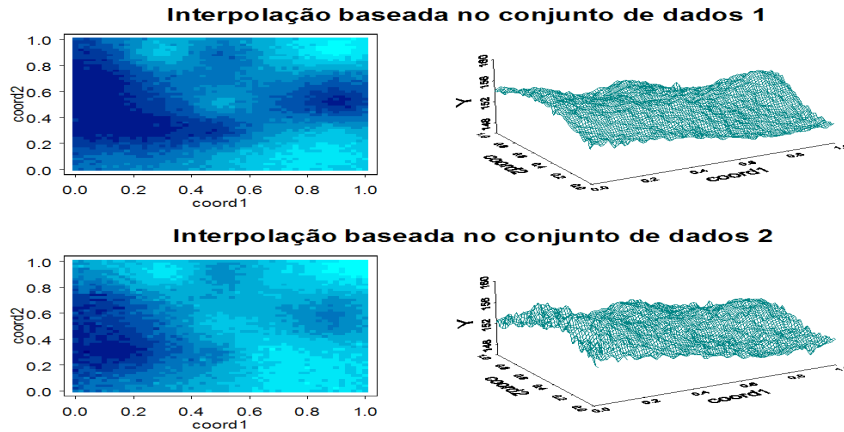


Figura 2.9: Superfície interpolada de  $Y(\cdot, 30)$  quando  $X(\cdot, \cdot)$  é desconhecido, baseada nos conjuntos de dados 1 e 2.

baseada no conjunto de dados 2 é mais distante da superfície apresentada na Figura 2.6 e da superfície verdadeira de  $Y$ , dada na Figura 2.1.

## 2.4 Análise de dados $PM_{10}$ no Rio de Janeiro

O conjunto de dados aqui analisado consiste em medições de  $PM_{10}$  em 16 estações de monitoramento na Região Metropolitana do Rio de Janeiro, coletados a cada 6 dias no ano de 1999, totalizando 59 observações de tempo. Uma descrição da campanha de monitoramento que deu origem a esse conjunto de dados pode ser encontrada em Paez (2000). A Figura 2.10 mostra o mapa da Região Metropolitana do Rio de Janeiro com o local dos postos de monitoramento, e na tabela 2.7 é dada a listagem de nomes desses postos. A região inclui a Baía de Guanabara, que deságua no Oceano Atlântico e separa os postos 13 e 15 dos outros. Para referências posteriores, a Figura 2.10 também mostra a área que será utilizada para interpolação. Mais detalhes sobre a região estudada, o conjunto de dados, e os fatores que influenciam os níveis de  $PM_{10}$  são descritos em Paez

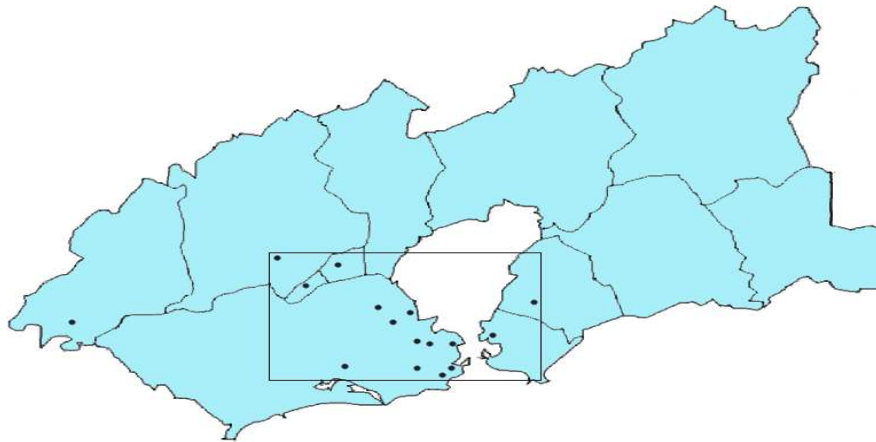


Figura 2.10: Mapa da Região Metropolitana do Rio de Janeiro, mostrando a localização das estações de monitoramento e a área de interpolação.

e Gamerman (2003), e Paez (2000).

Tabela 2.7: Estações de monitoramento de  $PM_{10}$  na Região Metropolitana do Rio de Janeiro

índice	estação	índice	estação
1	Bonsucesso	9	Jacarepaguá
2	Botafogo	10	Maracanã
3	Caxias	11	Nova Iguaçu
4	Centro	12	Nilópolis
5	Sumaré	13	Niterói
6	Copacabana	14	São Cristóvão
7	Inhaúma	15	São Gonçalo
8	Itaguaí	16	São João de Meriti

As investigações feitas em Paez and Gamerman (2003) mostram que a temperatura máxima diária observada na região estudada, e indicadores do dia da semana, são

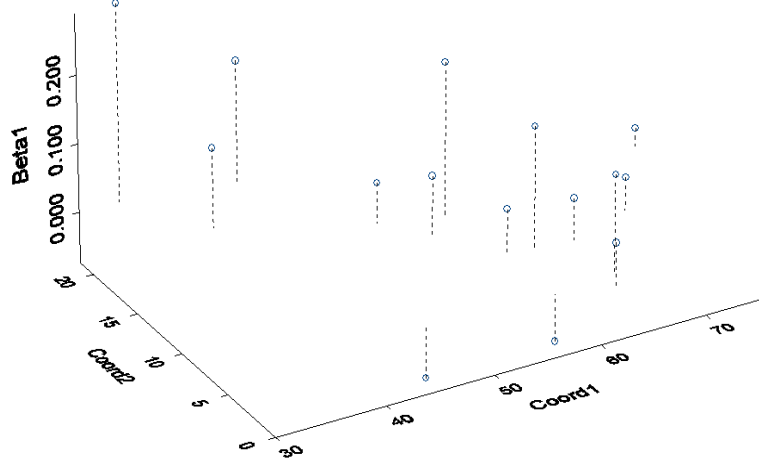


Figura 2.11: Efeito da temperatura por estação.

variáveis úteis para ajudar a explicar a variação de  $PM_{10}$ . A análise exploratória descrita nesse mesmo artigo sugere que os efeitos da temperatura máxima diária variam de forma significativa no espaço, ao contrário dos efeitos do dia da semana, que podem ser considerados fixos. A figura 2.11 mostra a variação dos efeitos da temperatura máxima diária estimados independentemente, estação por estação, através de regressão simples.

Baseado nessas análises, o modelo usado por Paez and Gamerman (2003) é estendido aqui de modo a acomodar efeitos de temperatura que variam no espaço. Especificamente

$$\begin{aligned}
 Y(s, t) &\sim N(\mu(s, t), \sigma_e^2) \\
 \mu(s, t) &= \beta_0(s) + \beta_1(s)TEMP(t) + \boldsymbol{\beta}'\mathbf{X}(t) + \phi(t)
 \end{aligned}
 \tag{2.14}$$

onde  $Y(s, t)$  é a raiz quadrada dos níveis de concentração de  $PM_{10}$  na estação de moni-

toramento  $s$  e período de tempo  $t$ ,  $\text{TEMP}(t)$  é a temperatura máxima diária no período de tempo  $t$  e  $\mathbf{X}(t) = (\text{SEG}(t), \text{TER}(t), \text{QUA}(t), \text{QUI}(t), \text{SEX}(t), \text{SAB}(t))$  são indicadores do dia da semana (por exemplo,  $\text{SEG}(t) = 1$  quando a observação é coletada em uma segunda-feira, e 0 caso contrário). Assim como em Paez and Gamerman (2003), escolhemos trabalhar com a raiz quadrada dos níveis de  $PM_{10}$  como variável resposta, visto que análises exploratórias mostram que essa transformação ajuda a "normalizar" essa variável (uma forma mais estruturada de levar em conta a incerteza na transformação normalizadora pode ser vista em De Oliveira, Kedem e Short, 1997). Os parâmetros de regressão que variam no espaço,  $\beta_0(s)$  e  $\beta_1(s)$ , são definidos em (2.2), ambos com função de correlação dada em (2.10), enquanto os parâmetros de regressão fixos  $\boldsymbol{\beta} = (\beta_2, \dots, \beta_7)$  tem distribuição  $\beta_2, \dots, \beta_7 \stackrel{\text{ind}}{\sim} N(0, 10^{10})$ . Finalmente,  $\phi(t)$  é um componente temporal definido em (2.3).

As priors para os parâmetros do modelo são especificadas em (2.5), e dadas por:

$$\sigma_e^{-2} \sim G(1, 2.74), \sigma_0^{-2} \sim G(1, 8.0), \sigma_1^{-2} \sim G(1, 0.005), \sigma_\phi^{-2} \sim G(10^{-6}, 10^{-6}),$$

$$\gamma_0, \gamma_1 \sim N(0, 10^{10}), \lambda_0 \sim G(1, 4), \lambda_1 \sim G(1, 2.17), \text{ and } \delta \sim U(0, 1),$$

onde a locação para cada uma das priors de  $\sigma_e^{-2}$ ,  $\sigma_0^{-2}$ ,  $\sigma_1^{-2}$ ,  $\lambda_1$  e  $\lambda_0$  foi obtida por análises preliminares, utilizando o método descrito na seção 2.2.2, após remover os efeitos dos dias da semana. Escolhemos trabalhar com priors com 1 grau de liberdade em todos os casos, visto que essa especificação (PPI) forneceu melhores resultados na análise de dados simulados. As priors para  $\sigma_\phi^{-2}$ ,  $\gamma_0$  e  $\gamma_1$ , e  $\delta$  foram especificadas como sendo não informativas.

### 2.4.1 Inferência sobre os parâmetros do modelo

A inferência sobre os parâmetros do modelo é feita via MCCM (Gamerman, 1997a) como descrito em Paez (2000), e esses resultados são sumarizados na tabela 2.8.



Tabela 2.8: Estatísticas Descritivas baseadas na amostra das distribuições a posteriori estimadas por MCCM

parâmetro	2.5%	97.5%	média	d.p.
$\beta_2$	-0.098	2.024	1.012	0.4996
$\beta_3$	-0.8887	1.237	0.228	0.5415
$\beta_4$	-0.4206	2.267	1.025	0.6801
$\beta_5$	0.2025	2.854	1.63	0.673
$\beta_6$	0.6004	3.309	1.933	0.6587
$\beta_7$	-0.5784	1.346	0.3786	0.4847
$\gamma_0$	5.385	9.286	7.509	1.055
$\gamma_1$	-0.040	0.069	0.012	0.033
$\lambda_0$	0.041	0.950	0.315	0.272
$\lambda_1$	0.017	1.620	0.388	0.449
$\sigma_0^{-2}$	0.094	1.221	0.380	0.367
$\sigma_1^{-2}$	129.0	1141.0	485.1	263.5
$\delta$	0.221	0.822	0.530	0.153
$\sigma_\phi^{-2}$	0.445	1.279	0.788	0.215
$\sigma_e^{-2}$	0.451	0.587	0.517	0.035

As estimativas dos parâmetros  $\beta$  confirmam os resultados obtidos por Paez and Gamerman (2003), que mostram que as concentrações de partículas inaláveis tendem a ser mais altas nas quintas e sexta-feiras, visto que as estimativas para  $\beta_4$  e  $\beta_5$  são maiores que as estimativas para os outros  $\beta$ .

A precisão estimada  $\hat{\sigma}_0^{-2}$  do processo correspondente ao intercepto é maior que as estimativas das precisões do erro e componentes temporais,  $\hat{\sigma}_e^{-2}$  e  $\hat{\sigma}_\phi^{-2}$ , assim como encontrado em Paez and Gamerman (2003), mostrando a importância dos coeficientes do intercepto para explicar a variabilidade de  $Y(\cdot, \cdot)$ . A precisão estimada  $\hat{\sigma}_1^{-2}$  é muito mais alta que  $\hat{\sigma}_0^{-2}$ .

Em Paez e Gamerman (2003), o parâmetro de decaimento  $\lambda_0$  foi estimado por  $\hat{E}(\lambda_0|y) =$

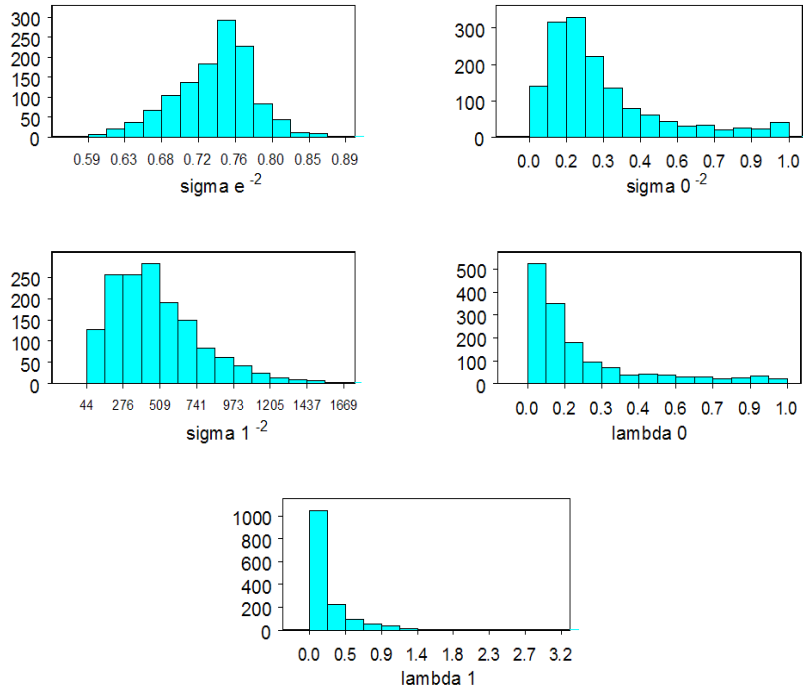


Figura 2.12: Histogramas da amostra a posteriori dos parâmetros  $\sigma_e^{-2}$ ,  $\sigma_0^{-2}$ ,  $\sigma_1^{-2}$ ,  $\lambda_0$  e  $\lambda_1$ .

0.177, enquanto aqui obtivemos  $\hat{E}(\lambda_0|y) = 0.315$ . Apesar dessas esperanças terem sido obtidas por dois modelos diferentes e não comparáveis, isso sugere que com a incorporação do parâmetro  $\beta_1(\cdot)$  variando no espaço, a correlação entre o intercepto de duas diferentes estações de monitoramento diminui. Parte dessa correlação foi absorvida pela correlação entre os  $\beta_1(\cdot)$ 's dessas duas estações.

A Figura 2.12 mostra os histogramas das posterioris dos parâmetros  $\sigma_e^{-2}$ ,  $\sigma_0^{-2}$  e  $\sigma_1^{-2}$ ,  $\lambda_0$  e  $\lambda_1$ , estando todos concentrados em pequenos intervalos com exceção de  $\sigma_e^{-2}$ .

## 2.4.2 Interpolação espacial

Com objetivo de fazer interpolação espacial, trabalhamos com a mesma grade regular  $50 \times 50$  usada em Paez e Gamerman (2003), cobrindo a Região Metropolitana do Rio

de Janeiro, que pode ser vista na Figura 2.10; Essa Figura contém todas as regiões de monitoramento exceto uma. Uma vez obtida uma amostra da posteriori de  $Y(\cdot, t)$ , uma amostra da posteriori na escala original de  $PM_{10}$  pode ser obtida aplicando o a transformação quadrática a cada valor amostrado. As Figuras 2.13, 2.14 e 2.15 mostram as superfícies estimadas para, respectivamente, os processos de regressão  $\beta_0(\cdot)$  e  $\beta_1(\cdot)$ , e o processo resposta  $(Y(\cdot, t))^2$  para o período de tempo  $t = 59$ . Note que ambas as superfícies de regressão estimadas têm uma variação espacial substancial, o que apoia o uso de modelos com coeficientes variando no espaço.

Quando analisamos a interpolação de  $(Y(\cdot, t))^2$  em  $t=59$ , chegamos à mesma conclusão que Paez e Gamerman (2003), ou seja, os níveis de  $PM_{10}$  preditos para o sul e leste da região são baixos, visto que essas são em sua maioria áreas residenciais. A medida que nos movemos para o norte e oeste, essas concentrações aumentam por causa da maior quantidade de atividades industriais, e ventos mais fracos que provocam menos circulação de ar.

As superfícies estimadas de  $\beta_0(\cdot)$ ,  $\beta_1(\cdot)$  e  $(Y(\cdot, t))^2$  mostram um padrão similar de variação espacial. Isso significa que os efeitos de temperatura tendem a ser baixos e altos nos mesmo locais que a concentração de  $PM_{10}$ , o que está de certa forma relacionado com o fato da dispersão de emissões industriais na região variarem com a temperatura. Entretanto, a razão científica por trás desse resultado está além dos propósitos desse estudo, e deveria ser analisada com a ajuda de outras variáveis explicativas.

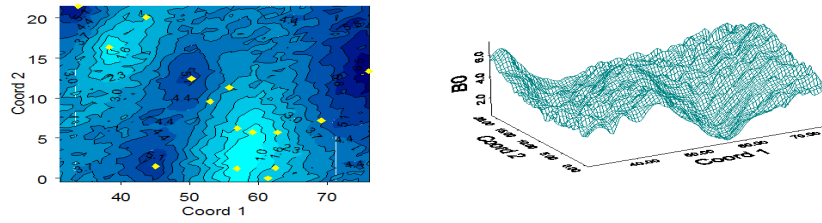


Figura 2.13: Superfície interpolada de  $\beta_0(\cdot)$ .

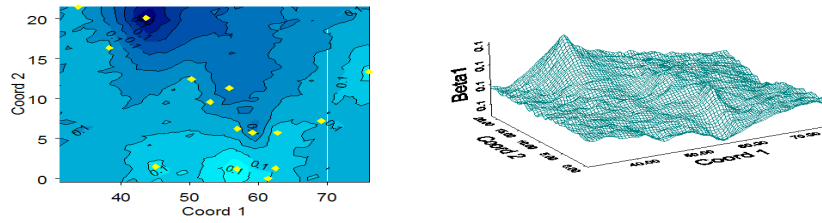


Figura 2.14: Superfície interpolada de  $\beta_1(\cdot)$ .

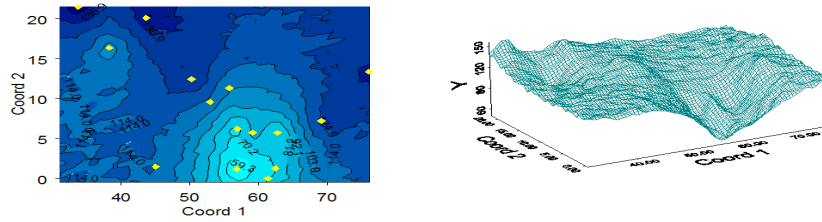


Figura 2.15: Superfície interpolada das concentrações de  $PM_{10}$  em  $t=59$ .

## 2.5 Conclusões

Nosso objetivo nesse capítulo foi de estudar os efeitos obtidos ao adicionar incerteza em modelos espaço-temporais, com base nos modelos propostos em Paez and Gamerman (2003). Trabalhamos com um conjunto de dados simulados com propósitos investigativos. Uma forma de incorporar incerteza foi permitir a variação espacial dos coeficientes de regressão. Também comparamos o caso onde as covariáveis eram conhecidas para todo o espaço de interesse, com o caso onde eram conhecidas em apenas um conjunto de pontos. Os dois casos deram origem a superfícies de interpolação similares, sendo que o primeiro levou a melhores resultados no ajuste dos valores interpolados à variável resposta. O segundo caso, baseado em valores interpolados de  $X(.,.)$ , deu origem a uma superfície mais suave.

Trabalhando sob o ponto de vista Bayesiano, analisamos diferentes especificações a priori para os parâmetros dos campos aleatórios Gaussianos. Isso é necessário para acomodar variações espaciais, devido a dificuldades muitas vezes presentes na estimação

desses parâmetros. Chegamos à conclusão que usar prioris centradas em valores previamente estimados pode ser benéfico para a estimação de certos parâmetros num campo aleatório Gaussiano, quando esse campo não é replicado. Entretanto, o uso de prioris com graus de liberdade maiores que 10 não deu origem a nenhuma melhoria nas estimativas dos parâmetros. Não chegamos a nenhuma conclusão quanto à vantagem de se trabalhar com amostras ordenadas, mas pelo menos em alguns casos esse benefício é evidente. Aumentar o tamanho amostral também é benéfico na maioria das vezes.

Como uma aplicação, trabalhamos com o conjunto de dados de poluição analisado em Paez and Gamerman (2003), generalizando os modelos propostos nesse artigo de forma a permitir a variação dos coeficientes de temperatura no espaço. As superfícies interpoladas obtidas apresentaram um padrão similar às obtidas anteriormente, mostrando que os níveis de  $PM_{10}$  são mais baixos no sudeste da região estudada. Agora também podemos dizer que os efeitos da temperatura são menores nessa área.

# Capítulo 3

## Modelos Dinâmicos Hierárquicos Espaço-Temporais para Respostas Multivariadas

### 3.1 Introdução

Nesse capítulo trabalhamos com modelos dinâmicos hierárquicos para modelar séries temporais multivariadas medidas em diferentes locais no espaço.

Modelos dinâmicos para dados espaço-temporais foram propostos em vários artigos como Huang e Cressie (1996), Wikle e Cressie (1999) e Mardia et al (1998) sob uma abordagem não Bayesiana, e Sansó e Guenni (1999), Stroud et al (2001) e Huerta et al (2003) sob abordagem Bayesiana.

Quintana (1987) propõe uma extensão aos modelos dinâmicos lineares para o caso de observações matriz-variadas, onde as matrizes de covariância observacional e evolucionar são definidas por fatores de escala comuns. Gamerman e Migon (1993) propõem um mo-

delo dinâmico hierárquico para respostas univariadas, também restringindo as variâncias do modelo a um fator de escala comum. Combinando as duas abordagens, Landim e Gamerman (2000) apresentam uma classe de modelos dinâmicos hierárquicos mais geral, onde as observações são matriz-variadas.

Seja  $Y_t$  uma matriz de dimensão  $N \times q$  formada por  $q$  variáveis resposta obtidas em  $N$  estações de monitoramento, para um período de tempo  $t$ ,  $t \in T \subset Z^+$ . Suponha que os dados são medidos em  $T$  períodos de tempo  $t = 1, 2, \dots, T$ . Landim e Gamerman (2000) modelam essa resposta multivariada como um processo hierárquico com  $k$  níveis, explorando as similaridades entre as estações de monitoramento, e especificando uma estrutura de evolução dinâmica para explicar a variação temporal dos dados. O modelo para  $Y_t$  é definido por

$$\begin{aligned}
Y_t &= F_{1t}\Theta_{1t} + v_{1t}, & v_{1t}|\Sigma, V_1 &\sim N(0, V_1, \Sigma) \\
\Theta_{1t} &= F_{2t}\Theta_{2t} + v_{2t}, & v_{2t}|\Sigma, V_2 &\sim N(0, V_2, \Sigma) \\
&& \vdots & \\
\Theta_{k+1,t} &= F_{kt}\Theta_{kt} + v_{kt}, & v_{kt}|\Sigma, V_k &\sim N(0, V_k, \Sigma) \\
\Theta_{k,t} &= G_t\Theta_{k,t-1} + w_t, & w_t|\Sigma, W &\sim N(0, W, \Sigma)
\end{aligned} \tag{3.1}$$

para  $t = 1, \dots, T$ , onde  $N(M, C, W)$  denota a distribuição Normal matriz-variada, com média  $M$ , variância a esquerda  $C$  e variância a direita  $\Sigma$  (veja Landim (1998) para mais detalhes). As matrizes  $\Theta_{it}$ ,  $i = 1, \dots, k$  são  $N_i \times q$ , os  $F_{it}$ s são  $N_{i-1} \times N_i$ , com  $i = 1, \dots, k$ , e  $G_t$  é  $N_k \times N_k$ . Assume-se que as matrizes  $F_{it}$  e  $G_t$  são conhecidas, com  $F_{it}$  possivelmente acomodando valores de variáveis explicativas.  $\Theta_{it}$  e  $v_{it}$  são independentes por hipótese para  $i = 1, \dots, k$ , assim como  $\Theta_{k(t-1)}$  e  $w_t$ , e as matrizes de variância  $v_{1t}, v_{2t}, \dots, v_{kt}$  e  $w_t$ .

$V_1, V_2, \dots, V_k$  e  $W$  são matrizes positivas definidas. Landim e Gamerman (2000) abordam



o caso onde elas são supostamente conhecidas, e o caso em que são desconhecidas, sob o ponto de vista Bayesiano. Nesse último caso, assume-se que as matrizes de variância têm distribuição Wishart Invertida a priori, e são estimadas usando o amostrador de Gibbs em MCMC. Note que nesse capítulo a referência a um período de tempo  $t$  é denotada pelo índice  $t$  (por exemplo,  $Y_t$ ), enquanto nos capítulos 2 e 4 o tempo aparece como argumento de função (por exemplo,  $Y(t)$ ).

Nessa pesquisa exploramos a mesma idéia de Landim e Gamerman (2000), mas impondo aqui uma estrutura paramétrica para a matriz de correlações entre parâmetros de regiões distintas. Com essa restrição, a correlação espacial pode ser capturada sem haver a necessidade de estimar a matriz completa. Nosso modelo tem a vantagem de permitir a interpolação de  $Y_t$  para qualquer conjunto de pontos não observados, levando em consideração a variação espacial presente no processo. Nossa proposta é a de fazer previsão e interpolação para esses modelos.

Adotamos aqui uma abordagem completamente Bayesiana para fazer inferência, e usamos métodos de MCMC para amostrar da distribuição a posteriori dos parâmetros.

Para tratar esse mesmo tipo de problema, Gelfand, Banerjee e Gamerman (2003) usam uma abordagem similar que descreve a variância temporal através de componentes dinâmicos, capturando associações entre medidas feitas em um dado local no espaço e período de tempo, assim como a dependência envolvendo espaço e tempo. Os processos espaciais multivariados descritos nesse artigo foram desenvolvidos através de técnicas de correionalização. Outras pesquisas nessa área levando em conta apenas a variação temporal foram desenvolvidas por Cargnoni, Muller e West (1997) e Aguilar e West (1998).

Nesse capítulo apresentamos a descrição do modelo na seção 3.2. Na seção 3.3 apresentamos alguns resultados de inferência para os parâmetros desconhecidos do modelo, e o algoritmo utilizado para gerar amostras das suas respectivas posteriores. As seções 3.4

e 3.5 mostram como fazer previsão e interpolação, respectivamente, para esses modelos. Na seção 3.6 apresentamos um exemplo simulado, e finalmente na seção 3.7 temos uma aplicação a dados de poluição.

## 3.2 Descrição do modelo

Uma forma de incorporar estrutura espacial no modelo definido em (3.1) é através da especificação de estruturas paramétricas para as matrizes de variância  $V_1, V_2, \dots, V_k$ . Aqui vamos considerar o caso onde a correlação espacial presente nas variáveis resposta se dá exclusivamente através dos parâmetros de regressão  $\Theta_{1t}$ . Dessa forma, definimos  $V_1$  como sendo uma matriz identidade, e  $V_2$  como sendo o produto de kronecker entre duas matrizes  $C$  e  $V$ . A matriz  $V$  define a correlação existente entre coeficientes que correspondem, por exemplo, a diferentes variáveis explicativas, enquanto  $C$  define a estrutura de correlação espacial que é igual para todos esses coeficientes.

Por simplicidade vamos considerar  $k = 2$  níveis hierárquicos. O Modelo aqui proposto é um caso particular do modelo (3.1) e é definido pelas equações abaixo:

$$\begin{aligned}
 Y_t &= F_{1t}\Theta_{1t} + v_{1t}, & v_{1t}|\Sigma &\sim N(0, I_N, \Sigma) \\
 \Theta_{1t} &= F_{2t}\Theta_{2t} + v_{2t}, & v_{2t}|C, V, \Sigma &\sim N(0, C \otimes V, \Sigma) \\
 \Theta_{2t} &= G_t\Theta_{2,t-1} + w_t, & w_t|\Sigma, W &\sim N(0, W, \Sigma)
 \end{aligned} \tag{3.2}$$

onde  $I_N$  denota uma matriz identidade de ordem  $N$ .

As matrizes  $F_{1t}(N_0 \times N_1)$ ,  $F_{2t}(N_1 \times N_2)$  e  $G_t(N_2 \times N_2)$  são supostamente conhecidas. A matriz  $C(N \times N)$ , que define a estrutura de correlação espacial presente no modelo, pode ser definida por qualquer função capaz de capturar essa estrutura. Se trabalhar-

mos sob a hipótese de isotropia, ela tipicamente depende de  $d_{ij} = ||i - j||$ , a distância Euclideana entre locais  $i$  e  $j$ , e parâmetros  $\lambda$ .

A matriz de parâmetros de regressão  $\Theta_{1t}$  ( $N_1 \times q$ ) pode ser escrita como uma coleção de blocos  $\Theta_{1t} = [\Theta_{1,1t}, \dots, \Theta_{N,1t}]'$ , onde  $\Theta_{j,1t}$  é  $N_2 \times q$  tal que  $N_1 = N \times N_2$ . O modelo é completado com prioris Wishart Invertidas independentes para as matrizes de covariância, e distribuição Normal matriz-variada para  $(\Theta_{2,0}|\Sigma)$ , além de prioris para os parâmetros  $\lambda$  que definem a matriz de correlação espacial  $C$ .

A nível de ilustração, vamos apresentar a seguir dois exemplos. No primeiro exemplo definimos um modelo univariado, e no segundo, um modelo com duas variáveis resposta.

### 3.2.1 Exemplo 1

Nesse exemplo consideramos apenas uma variável resposta  $Y$  ( $q=1$ ), observada em  $N = 3$  locais no espaço e  $T = 10$  períodos de tempo. Suponha que para cada local  $i = 1, \dots, 3$  e tempo  $t = 1, \dots, 10$ , também observamos uma variável explicativa  $X_{i,t}$ . A equação de regressão que define a relação entre  $Y$  e  $X$  é dada por

$$Y_{i,t} = \alpha_{i,t} + \beta_{i,t}X_{i,t} + v_{1,i,t}, \quad i = 1, \dots, 3,$$

onde  $\alpha_{i,t}$  e  $\beta_{i,t}$  são parâmetros de regressão que variam tanto no tempo quanto no espaço. Supondo que os erros  $v_{1,1,t}$ ,  $v_{1,2,t}$  e  $v_{1,3,t}$  sejam independentes para  $t$  fixo, temos que  $V_1 = I_3$ .

De acordo com o modelo (3.2), a matriz  $F_{1t}$  ( $N_0 = 3 \times N_1 = 6$ ) é definida por

$$F_{1t} = \begin{bmatrix} 1 & X_{1,t} & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & X_{2,t} & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & X_{3,t} \end{bmatrix},$$

e  $\Theta_{1t}$ , que é a coleção de parâmetros de regressão no primeiro nível da hierarquia, tem dimensão  $N_1 = 6 \times q = 1$ , e pode ser escrito como

$$\Theta_{1t} = \begin{bmatrix} \alpha_{1,t} \\ \beta_{1,t} \\ \alpha_{2,t} \\ \beta_{2,t} \\ \alpha_{3,t} \\ \beta_{3,t} \end{bmatrix}.$$

Suponha que para um período de tempo fixo  $t$  os parâmetros  $\alpha_{1,t}$ ,  $\alpha_{2,t}$  e  $\alpha_{3,t}$  tenham uma média comum  $\delta_t$ , e do mesmo modo, os parâmetros  $\beta_{1,t}$ ,  $\beta_{2,t}$  e  $\beta_{3,t}$  tenham a mesma média  $\psi_t$ . Podemos definir então as matrizes  $\Theta_{2t}$  ( $N_2 = 2 \times q = 1$ ) e  $F_{2t}$  ( $N_1 = 6 \times N_2 = 2$ ) como

$$\Theta_{2t} = \begin{bmatrix} \delta_{1,t} \\ \psi_{1,t} \end{bmatrix}, \quad F_{2t} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & 1 \end{bmatrix},$$

e as equações

$$\begin{aligned} \alpha_{i,t} &= \delta_t + v_{2,1+2(i-1),t} \\ \beta_{i,t} &= \psi_t + v_{2,2i,t} \quad i = 1, 2, 3. \end{aligned}$$

Vamos supor além disso que  $(\alpha_{1,t}, \alpha_{2,t}, \alpha_{3,t})$  sejam espacialmente correlacionados, assim como  $(\beta_{1,t}, \beta_{2,t}, \beta_{3,t})$ , e ambos os conjuntos de parâmetros  $\alpha$  e  $\beta$  compartilham da mesma matriz de correlação espacial  $C$ . Suponha também que a correlação entre os parâmetros  $\alpha$  e  $\beta$  seja dada por  $V$ , e não dependa dos locais de observação dos dados. Dessa forma, a matriz de correlação de  $(\Theta_{1t}|\Theta_{2t})$  é separável e pode ser escrita como  $V_2 = C \otimes V$ .

Para definir o último nível da hierarquia, supomos que  $\Theta_{2t}$  seja um passeio aleatório, tal que  $\Theta_{2t} = \Theta_{2,t-1} + w_t$ , sendo  $W$  a matriz de variâncias de  $w_t$ . Sendo assim, temos que  $G_t = I_{N_2}$ .

### 3.2.2 Exemplo 2

Considere agora um modelo com  $q = 2$  variáveis resposta  $Y$ , e uma variável explicativa  $X$ , observadas em  $N = 3$  estações de monitoramento e  $T = 10$  períodos de tempo. Suponha também que uma segunda variável  $Z_{i,t}$  foi observada para cada  $i = 1, \dots, 3$  e  $t = 1, \dots, 10$ , ajudando a explicar a variabilidade dos parâmetros de regressão  $\Theta_{1t}$  do nível 1 da hierarquia.

Para um dado período de tempo, definimos cada elemento da matriz  $Y_{ij,t}$  pela equação de regressão  $Y_{ij,t} = \alpha_{ij,t} + \beta_{ij,t}X_{i,t} + v_{1,ij,t}$ ,  $i = 1, \dots, 3; j = 1, 2$ , onde  $\alpha_{ij,t}$  e  $\beta_{ij,t}$  são os coeficientes de regressão que variam no tempo e no espaço, para cada variável resposta. Dessa forma, para cada resposta temos dois parâmetros de regressão ( $N_2 = 2$ ).

Vamos supor, como no exemplo 1, que para um período de tempo fixo  $t$ , e variável resposta  $j$ , os erros  $v_{1,ij,t}$ ,  $i = 1, \dots, 3$  são independentes. Temos então  $V_1 = I_N$ . Para  $t$  e  $i$  fixos, a covariância entre  $v_{1,i1,t}$  e  $v_{1,i2,t}$  é dada por  $\Sigma$ . Dessa forma, definimos o primeiro nível hierárquico do modelo como em (3.2), onde as matrizes  $F_{1t}$  e  $\Theta_{1t}$  são dadas por:

$$F_{1t} = \begin{bmatrix} 1 & X_{1,t} & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & X_{2,t} & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & X_{3,t} \end{bmatrix}, \quad \Theta_{1t} = \begin{bmatrix} \alpha_{11,t} & \alpha_{12,t} \\ \beta_{11,t} & \beta_{12,t} \\ \alpha_{21,t} & \alpha_{22,t} \\ \beta_{21,t} & \beta_{22,t} \\ \alpha_{31,t} & \alpha_{32,t} \\ \beta_{31,t} & \beta_{32,t} \end{bmatrix}.$$

Para o segundo nível da hierarquia, supomos que para um dado  $t$ , ambas as coleções de parâmetros  $\alpha$  e  $\beta$ , que formam a matriz  $\Theta_{1t}$ , podem ser explicadas pela variável  $Z_{i,t}$ ,  $i = 1, \dots, 3$ , de forma que:

$$\begin{aligned} \alpha_{ij,t} &= \delta_{j,t} + \delta_{j,t}^* Z_{i,t} + v_{2,1+2(i-1),t} \\ \beta_{ij,t} &= \psi_{j,t} + \psi_{j,t}^* Z_{i,t} + v_{2,2i,t}, \quad i = 1, \dots, 3; j = 1, 2 \end{aligned}$$

Sendo assim, define-se as matrizes  $F_{2t}$  e  $\Theta_{2t}$  por:

$$F_{2t} = \begin{bmatrix} 1 & 0 & Z_{1,t} & 0 \\ 0 & 1 & 0 & Z_{1,t} \\ 1 & 0 & Z_{2,t} & 0 \\ 0 & 1 & 0 & Z_{2,t} \\ 1 & 0 & Z_{3,t} & 0 \\ 0 & 1 & 0 & Z_{3,t} \end{bmatrix}, \quad \Theta_{2t} = \begin{bmatrix} \delta_{1,t} & \delta_{2,t} \\ \psi_{1,t} & \psi_{2,t} \\ \delta_{1,t}^* & \delta_{2,t}^* \\ \psi_{1,t}^* & \psi_{2,t}^* \end{bmatrix}.$$

Assim como no exemplo 1, vamos supor que a matriz de variâncias da distribuição de  $(\Theta_{1t}|\Theta_{2t})$  é separável e pode ser escrita como  $V_2 = C \otimes V$ . A matriz  $C$  define a estrutura de correlação espacial do modelo, e a matriz  $V$  define a estrutura de correlação entre as

coleções de parâmetros  $\alpha$  e  $\beta$ . Para  $t$  e  $i$  fixos, a matriz de correlações entre os parâmetros  $\alpha$ , assim como entre os parâmetros  $\beta$ , é dada por  $\Sigma$  ( $q = 2 \times q = 2$ ).

Para a evolução de  $\Theta_{2t}$  especificamos  $G_t = I_2$ , assim como no exemplo 1, definindo a equação  $\Theta_{2t} = \Theta_{2,t-1} + w_t$ . Temos que para um dado  $t$ ,  $w_t | \Sigma, W \sim N(0, W, \Sigma)$ . Fixando  $j$ ,  $W$  corresponde à matriz de correlações entre os parâmetros  $\delta_{j,t}$ ,  $\psi_{j,t}$ ,  $\delta_{j,t}^*$ ,  $\psi_{j,t}^*$ .  $\Sigma$  corresponde à matriz de correlações entre elementos de um mesmo parâmetro, ou seja, entre  $\delta_{1,t}$  e  $\delta_{2,t}$ ; entre  $\psi_{1,t}$  e  $\psi_{2,t}$ ; entre  $\delta_{1,t}^*$  e  $\delta_{2,t}^*$ ; e entre  $\psi_{1,t}^*$  e  $\psi_{2,t}^*$ . Note que a matriz  $\Sigma$  permanece constante, ou seja, a correlação entre os elementos dos parâmetros é sempre a mesma.

É importante notar também que  $\Sigma$  define a correlação entre as duas variáveis resposta no primeiro nível hierárquico do modelo (3.2), assim como a correlação entre os elementos dos parâmetros que são relacionados às diferentes variáveis resposta nos níveis hierárquicos que se seguem.

### 3.3 Inferência e métodos computacionais

No modelo proposto as quantidades desconhecidas que devem ser estimadas são  $\Sigma$ ,  $V$ ,  $W$ ,  $\Theta_{1t}$ ,  $\Theta_{2t}$ , para  $t = 1, \dots, T$ , além do conjunto de parâmetros  $\lambda$  da função de correlação espacial que define a matriz  $C$ . Consideramos aqui as mesmas priors utilizadas em Landim e Gamerman (2000) para os parâmetros, exceto para  $\lambda$ , cuja priori vai depender de cada definição particular de  $C$ , não sendo portanto especificada nessa seção. As priors são dadas por

$$\begin{aligned} \Theta_{2,0} | \Sigma &\sim N(M_2, C_2, \Sigma), & \Sigma &\sim WI(T_0, S_0), \\ W &\sim WI(T_{W0}, S_{W0}), & V &\sim WI(T_{V0}, S_{V0}), \end{aligned} \tag{3.3}$$

onde  $WI(T_0, S_0)$  denota a distribuição Wishart Invertida (Dawid, 1981) com média  $S_0/(T_0 - 2)$ ,  $T_0 > 2$ .

Vamos definir  $Y = \{Y_1, \dots, Y_T\}$ ,  $\{\Theta_1\} = \{\Theta_{11}, \dots, \Theta_{1T}\}$ ,  $\{\Theta_2\} = \{\Theta_{21}, \dots, \Theta_{2T}\}$ , e  $\Phi$  a coleção de todos os parâmetros, tal que  $\Phi = \{\{\Theta_1\}, \{\Theta_2\}, \Theta_{20}, \lambda, V, W, \Sigma\}$ . A distribuição a posteriori conjunta dos parâmetros desconhecidos do modelo (3.2) é dada por:

$$P(\Phi|Y) \propto P(\Theta_{20}|\Sigma)P(\Sigma)P(V)P(W)P(\lambda) \times \prod_{t=1}^T P(Y_t|\Theta_{1t}, \Sigma)P(\Theta_{1t}|\Theta_{2t}, \Sigma, V, \lambda)P(\Theta_{2t}|\Theta_{2,t-1}, \Sigma, W)$$

Como essa distribuição a posteriori não é analiticamente tratável, os parâmetros  $\Phi$  são amostrados por um algoritmos eficiente usando o amostrador de Gibbs, proposto por Landim e Gamerman (2000). Amostramos de  $\{\Theta_1\}, \Sigma, V$  e  $W$  através de suas condicionais completas, e amostramos de  $(\Theta_{20}, \{\Theta_2\})$  por FFBS (Forward Filtering Backward Sampling - Frühwirth-Schnatter, 1995). Passos adicionais de Metropolis-Hastings (Metropolis, 1953; Hastings, 1970) foram incluídos para gerar amostras de  $\lambda$ .

Vamos denotar por  $\Phi_{-\Sigma}$  a coleção de todos os parâmetros exceto  $\Sigma$ .  $\Phi_{-\{\Theta_1\}}$ ,  $\Phi_{-\{\Theta_2\}}$ ,  $\Phi_{-\Theta_{20}}$ ,  $\Phi_{-\lambda}$ ,  $\Phi_{-V}$  e  $\Phi_{-W}$  são definidos analogamente. Vamos também definir  $Y_t^* = Y_t - F_{1t}\Theta_{1t}$ ,  $\Theta_{1t}^* = \Theta_{1t} - F_{2t}\Theta_{2t}$ ,  $\Theta_{2t}^* = \Theta_{2t} - G_t\Theta_{2,t-1}$ ,  $t = 1, \dots, T$  e  $\Theta_{20}^* = \Theta_{20} - M_2$ . A seguir temos alguns resultados probabilísticos necessários para a implementação do algoritmo proposto para a geração de amostras das distribuições a posteriori dos parâmetros desconhecidos do modelo.

- Distribuição condicional completa de  $\Sigma$

$$P(\Sigma|\Phi_{-\Sigma}, Y) \propto P(\Theta_{20}|\Sigma)P(\Sigma) \prod_{t=1}^T P(Y_t|\Theta_{1t}, \Sigma)P(\Theta_{1t}|\Theta_{t-1}, \Theta_{2t}, \Sigma, V, \lambda)P(\Theta_{2t}|\Theta_{2,t-1}, \Sigma, W)$$



$$\begin{aligned}
&\propto |\Sigma|^{-\frac{N_2}{2}} \exp\left(-\frac{1}{2}tr\left(\Theta_{20}^{*'}C_2^{-1}\Theta_{20}^*\Sigma^{-1}\right)\right) |\Sigma|^{-\frac{T_0}{2}+q} \exp\left(-\frac{1}{2}tr[S_0\Sigma^{-1}]\right) \\
&\times \prod_{t=1}^T |\Sigma|^{-\left(\frac{N+N_1+N_2}{2}\right)} \times \\
&\exp\left(-\frac{1}{2}tr\left(\sum_{t=1}^T Y_t^{*'}Y_t^*\Sigma^{-1} + \sum_{t=1}^T \Theta_{1t}^{*'}V_2^{-1}\Theta_{1t}^* + \sum_{t=1}^T \Theta_{2t}^{*'}W^{-1}\Theta_{2t}^*\Sigma^{-1}\right)\right) \\
&= |\Sigma|^{-\left(\frac{T_0+N_2+T(N+N_1+N_2)}{2}+q\right)} \times \\
&\exp\left(-\frac{1}{2}tr\left(S_0 + \Theta_{20}^{*'}C_2^{-1}\Theta_{20}^* + \sum_{t=1}^T (Y_t^{*'}Y_t^* + \Theta_{1t}^{*'}V_2^{-1}\Theta_{1t}^* + \Theta_{2t}^{*'}W^{-1}\Theta_{2t}^*)\right)\right) \Sigma^{-1} \\
&\Rightarrow P(\Sigma|\Phi_{-\Sigma}) \sim WI(T_0 + N_2 + T(N_1 + N_2 + N), S_0), \quad \text{onde} \\
&S = S_0 + \left(\Theta_{20}^{*'}C_2^{-1}\Theta_{20}^* + \sum_{t=1}^T (y_t^{*'}y_t^* + \Theta_{1t}^{*'}V_2^{-1}\Theta_{1t}^* + \Theta_{2t}^{*'}W^{-1}\Theta_{2t}^*)\right)
\end{aligned}$$

e  $tr(A)$  denota o traço da matriz  $A$ .

- Distribuição condicional completa de  $W$

$$\begin{aligned}
P(W|\Phi_{-W}, Y) &\propto P(W) \prod_{t=1}^T P(\Theta_{2t}|W, \Sigma, \Theta_{2,t-1}) \\
&\propto |W|^{-\left(\frac{T_{W0}}{2}+N_2\right)} \exp\left(-\frac{1}{2}tr(S_{W0}W^{-1})\right) |W|^{-\frac{Tq}{2}} \exp\left(-\frac{1}{2}tr\sum_{t=1}^T \Theta_{2t}^{*'}\Sigma^{-1}\Theta_{2t}^*W^{-1}\right) \\
&\Rightarrow (W|\Phi_{-W}) \sim WI\left(T_{W0} + Tq; S_{W0} + \sum_{t=1}^T \Theta_{2t}^{*'}\Sigma^{-1}\Theta_{2t}^*\right).
\end{aligned}$$

- Distribuição condicional completa de  $V$

$$\begin{aligned}
P(V|\Phi_{-V}, Y) &\propto P(V) \prod_{t=1}^T P(\Theta_{1t}|\Theta_{2t}, C, V, \Sigma) \\
&\propto |V|^{-\left(\frac{T_{V0}}{2}+N_2\right)} \exp\left(-\frac{1}{2}tr(S_{V0}V^{-1})\right) |V|^{-\frac{Tq}{2}} \exp\left(-\frac{1}{2}tr\sum_{t=1}^T \Theta_{1t}^{*'}V_2^{-1}\Theta_{1t}^*\Sigma^{-1}\right) \\
&\propto |V|^{-\left(\frac{T_{V0}+TNq}{2}+N_2\right)} \exp\left(-\frac{1}{2}tr(S_{V0}V^{-1} + \Theta_{1t}^{*'}(C \otimes V)^{-1}\Theta_{1t}^*\Sigma^{-1})\right).
\end{aligned}$$

Usando a propriedade  $(C \otimes V)^{-1} = C^{-1} \otimes V^{-1}$ , e definindo  $K_{ij} = (C^{-1})_{ij}$ , temos o seguinte resultado:

$$\begin{aligned}
P(V|\Phi_{-V}, Y) &\propto |V|^{-\left(\frac{TV_0+TNq}{2}+N_2\right)} \times \\
&\exp\left(-\frac{1}{2}tr\left(S_{V_0} + \sum_{t=1}^T \sum_{i=1}^N \sum_{j=1}^N K_{ji}(\Theta_{i,1t} - \Theta_{2t})' \Sigma^{-1} (\Theta_{j,1t} - \Theta_{2t}) V^{-1}\right)\right) \\
\Rightarrow (V|\Phi_{-V}) &\sim WI\left(T_{V_0} + TNq; S_{V_0} + \sum_{t=1}^T \sum_{i=1}^N \sum_{j=1}^N K_{ji}(\Theta_{i,1t} - \Theta_{2t})' \Sigma^{-1} (\Theta_{j,1t} - \Theta_{2t})\right).
\end{aligned}$$

- Distribuição condicional completa de  $\{\Theta_1\}$

$$\begin{aligned}
P(\{\Theta_1\}|\Phi_{-\{\Theta_1\}}, Y) &\propto \prod_{t=1}^T P(Y_t|\Theta_{1t}, \Sigma) P(\Theta_{1t}|\Theta_{2t}, \lambda, V, \Sigma) \\
&\propto \exp\left(-\frac{1}{2}tr \sum_{t=1}^T (Y_t^* Y_t^* + \Theta_{1t}^* V_2^{-1} \Theta_{1t}^*) \Sigma^{-1}\right) \\
&= \exp\left[-\frac{1}{2}tr \sum_{t=1}^T (Y_t' Y_t - \Theta_{1t}' F_{1t}' Y_t - Y_t' F_{1t} \Theta_{1t} + \Theta_{1t}' F_{1t}' F_{1t} \Theta_{1t} \right. \\
&\quad \left. + \Theta_{2t}' F_2' V_2^{-1} F_{2t} \Theta_{2t} - \Theta_{2t}' F_2' V_2^{-1} \Theta_{1t} - \Theta_{1t}' V_2^{-1} F_{2t} \Theta_{2t} + \Theta_{1t}' V_2^{-1} \Theta_{1t}) \Sigma^{-1}\right] \\
&\propto \exp\left(-\frac{1}{2}tr \sum_{t=1}^T ((\Theta_{1t} - B_t^{-1} M_t)' B_t (\Theta_{1t} - B_t^{-1} M_t)) \Sigma^{-1}\right),
\end{aligned}$$

onde  $B_t = F_{1t}' F_{1t} + V_2^{-1}$  e  $M_t = F_{1t}' Y_t + V_2^{-1} F_{2t} \Theta_{2t}$ .

$$\Rightarrow (\Theta_{1t}|\Phi_{-\Theta_{1t}}) \sim N(B_t^{-1} M_t; B_t^{-1}; \Sigma).$$

- Distribuição condicional completa de  $\{\Theta_2\}$

$$P(\{\Theta_2\}|\Phi_{-\Theta_2}, Y) = P(\Theta_{2T}|\Theta_1, W, \Sigma, \lambda, V) \prod_{t=0}^{T-1} P(\Theta_{2t}|\Theta_{2,t+1}, \Theta_{2T}, \Theta_1, W, \Sigma, \lambda, V)$$

Propomos aqui o uso do algoritmo FFBS para amostrar de  $\{\Theta_2\}$ . Esse algoritmo é baseado nos seguintes resultados:

$$P(\Theta_{2t}|\Theta_{2,t+1}, \dots, \Theta_{2T}, \Theta_1, W, \Sigma, \lambda, V) = P(\Theta_{2t}|\Theta_{2,t+1}, \Theta_{11}, \dots, \Theta_{1t}, W, \Sigma, \lambda, V)$$

e

$$P(\Theta_{2t}|\Theta_{2,t+1}, \Theta_{11}, \dots, \Theta_{1t}, W, \Sigma, \lambda, V) \propto P(\Theta_{2,t+1}|\Theta_{2t}, W, \Sigma)P(\Theta_{2t}|\Theta_{11}, \dots, \Theta_{1t}, W, \Sigma, \lambda, V),$$

para  $t = 1, \dots, T - 1$ . Pode ser mostrado que

$$(\Theta_{2t}|\Theta_{2,t+1}, W, \Sigma) \sim N(h_t, H_t, \Sigma)$$

onde  $H_t = (C^{-1} + G'_t W^{-1} G_t)^{-1}$  e  $h_t = H_t(C_t^{-1} M_t + G'_t W^{-1} \Theta_{2,t+1})$  e

$$M_t = G_t M_{t-1} + R_t F'_{2t} Q_t^{-1} (\Theta_{1t} - G_t M_{t-1})$$

$$M_0 = M_{20} = M_2$$

$$C_t = R_t - R_t F'_{2t} Q_t^{-1} F_{2t} R_t$$

$$Q_t = V_2 + F_{2t} R_t F'_{2t}$$

$$R_t = W + G_t C_{t-1} G'_t$$

$$C_0 = C_{20} = C_2,$$

onde as quantidades acima são obtidas pela atualização dos parâmetros (West e Harrison, 1997), tomando  $\Theta_1$  como sendo os dados. Após sucessivas atualizações obtemos  $\Theta_{2T}|\Phi_{-\Theta_{2T}} \sim N(M_T, C_T, \Sigma)$ .

Com base nesses resultados, propomos o seguinte algoritmo para gerar amostras das distribuições a posteriori dos parâmetros  $\Phi$ :

1. Dê valores iniciais para todos os parâmetros e faça  $j=1$ .

2. Gere  $\{\Theta_1\}$ ,  $\Sigma$ ,  $V$  e  $W$  das suas distribuições condicionais completas.
3. Amostre  $(\Theta_{20}, \{\Theta_2\})$  por FFBS (Forward Filtering Backward Sampling - Frühwirth-Schnatter, 1995). Para  $t$  variando de 1 a  $T$  calcule  $M_t, C_t, Q_t$  e  $R_t$ . Depois gere  $\Theta_{2T}$  de  $N(M_T, C_T, \Sigma)$ , e para  $t$  variando de  $(T-1)$  a zero gere  $\Theta_{2t}$  de  $N(h_t, H_t, \Sigma)$ .
4. Gere  $\lambda$  por Metropolis-Hastings (Metropolis et al, 1953 e Hasting, 1970).
5. Faça  $j=j+1$  e volte para o passo 1 até atingir convergência.

Depois que a convergência é atingida, usamos o algoritmo descrito acima para obter amostras das distribuições a posteriori das quantidades desconhecidas.

### 3.4 Predição

Suponha que estamos interessados na distribuição da variável resposta  $h$  passos a frente, dadas as observações passadas. Dessa forma, estamos interessados em obter amostras da distribuição  $(Y_{T+h}|Y)$ .

A distribuição conjunta a posteriori de  $Y_{T+h}, \Theta_{1,T+h}, \Theta_{2,T+h}$  e  $\Phi$  pode ser escrita como

$$\begin{aligned}
P(Y_{T+h}, \Theta_{1,T+h}, \Theta_{2,T+h}, \Phi|Y) &= P(Y_{T+h}, \Theta_{1,T+h}, \Theta_{2,T+h}|Y, \Phi)P(\Phi|Y) \\
&= P(Y_{T+h}|\Theta_{1,T+h}, \Theta_{2,T+h}, \Phi, Y)P(\Theta_{1,T+h}|\Theta_{2,T+h}, \Phi, Y)P(\Theta_{2,T+h}|\Phi, Y)P(\Phi|Y) \\
&= P(Y_{T+h}|\Theta_{1,T+h}, \Sigma)P(\Theta_{1,T+h}|\Theta_{2,T+h}, \Sigma, V, \lambda)P(\Theta_{2,T+h}|\Theta_{2T}, \Sigma, W)P(\Phi|Y)
\end{aligned}$$

$P(\Phi|Y)$  pode ser amostrado através do algoritmo descrito na seção anterior. As distribuições  $P(Y_{T+h}|\Theta_{1,T+h}, \Sigma)$  e  $P(\Theta_{1,T+h}|\Theta_{2,T+h}, \Sigma, V, \lambda)$  são conhecidas, e  $P(\Theta_{2,T+h}|\Theta_{2T}, \Sigma, W)$

pode ser facilmente encontrada através de repetidos usos da equação  $\Theta_{2T} = G_t \Theta_{2,T-1} + w_T$ ,  $w_T|W, \Sigma \sim N(0, W, \Sigma)$ . Usando propriedades da distribuição Normal matriz-variada temos que

$$P(\Theta_{2,T+h}|\Theta_{2T}, W, \Sigma) \sim N \left( \prod_{t=T+1}^{T+h} G_t \Theta_{2,T}, W + \sum_{i=T+1}^{T+h} \left( \prod_{t=i}^{T+h} G'_t \right) W \left( \prod_{t=i}^{T+h} G_t \right), \Sigma \right).$$

Dessa forma, para gerar uma amostra de  $Y_{T+h}$  basta adicionar um passo no algoritmo apresentado na seção anterior onde geramos  $\Theta_{2,T+h}$  de  $P(\Theta_{2,T+h}|\Theta_{2,T}, \Sigma, W)$ , depois geramos  $\Theta_{1,T+h}$  de  $P(\Theta_{1,T+h}|\Theta_{2,T+h}, \Sigma, V, \lambda)$ , e finalmente geramos  $Y_{T+h}$  de  $P(Y_{T+h}|\Theta_{1,T+h}, \Sigma)$ .

### 3.5 Interpolação

Suponha que os dados são observados em um conjunto de locais  $S = \{s_1, \dots, s_N\} \subset R^2$ , e estamos interessados em interpolar a variável resposta para outras  $N_n$  localidades  $S_n = \{s_{N+1}, \dots, s_{N+N_n}\}$ , para um dado período de tempo  $t$ .

Vamos denotar por  $Y_t^o$  a matriz de valores observados  $N \times q$ , e por  $Y_t^n$  a matriz de valores a serem interpolados  $N_n \times q$ , para um período de tempo fixo  $t$ . Da mesma forma definimos  $F_{t1}^o, F_{t1}^n, F_{t2}^o, F_{t2}^n, \Theta_{1t}^o, \Theta_{1t}^n, v_{1t}^o, v_{1t}^n, v_{2t}^o$  e  $v_{2t}^n$ . Vamos também definir  $\psi = \{\Sigma, V, W, \lambda\}$ . O modelo extendido é dado por:

$$\begin{aligned} \begin{pmatrix} Y_t^o \\ Y_t^n \end{pmatrix} &= \begin{pmatrix} F_{1t}^o \\ F_{1t}^n \end{pmatrix} \begin{pmatrix} \Theta_{1t}^o \\ \Theta_{1t}^n \end{pmatrix} + \begin{pmatrix} v_{1t}^o \\ v_{1t}^n \end{pmatrix}, \quad \begin{pmatrix} v_{1t}^o \\ v_{1t}^n \end{pmatrix} | \Sigma \sim N(0, I_{N+N_n}, \Sigma) \\ \begin{pmatrix} \Theta_{1t}^o \\ \Theta_{1t}^n \end{pmatrix} &= \begin{pmatrix} F_{2t}^o \\ F_{2t}^n \end{pmatrix} \Theta_{2t} + \begin{pmatrix} v_{2t}^o \\ v_{2t}^n \end{pmatrix}, \quad \begin{pmatrix} v_{2t}^o \\ v_{2t}^n \end{pmatrix} | \lambda, V, \Sigma \sim N \left( 0, \begin{pmatrix} C^o & C^{on} \\ C^{no} & C^n \end{pmatrix} \otimes V, \Sigma \right) \\ \Theta_{2t} &= G_t \Theta_{2,t-1} + w_t, \quad w_t | \Sigma, W \sim N(0, W, \Sigma) \end{aligned}$$

sendo  $C^o$ ,  $C^{no}$ ,  $C^{on}$  e  $C^n$  as partições de  $C$  obtidas de acordo com a função que descreve a correlação espacial.

Sob a hipótese de que  $F_{1t}^o$ ,  $F_{1t}^n$ ,  $F_{2t}^o$ ,  $F_{2t}^n$  e  $G_t$  são conhecidos, a distribuição conjunta a posteriori de  $Y_t^n$ ,  $\Theta_{1t}^n$ ,  $\Theta_{1t}^o$ ,  $\Theta_{2t}$  e  $\psi$  é dada por

$$\begin{aligned}
& P(Y_t^n, \Theta_{1t}^n, \Theta_{1t}^o, \Theta_{2t}, \psi | Y_t^o) \\
& = P(Y_t^n, \Theta_{1t}^n | \Theta_{1t}^o, \Theta_{2t}, \psi, Y_t^o) P(\Theta_{1t}^o, \Theta_{2t}, \psi | Y_t^o) \\
& = P(Y_t^n | \Theta_{1t}^n, \psi, Y_t^o) P(\Theta_{1t}^n | \Theta_{1t}^o, \Theta_{2t}, \psi) P(\Theta_{1t}^o, \Theta_{2t}, \psi | Y_t^o), \tag{3.4}
\end{aligned}$$

visto que dados  $\Theta_{1t}^n$ ,  $\psi$  e  $Y_t^o$ ,  $Y_t^n$  não depende de  $\Theta_{1t}^o$  e  $\Theta_{2t}$ , e dados  $\Theta_{1t}^o$ ,  $\Theta_{2t}$  e  $\psi$ ,  $\Theta_{1t}^n$  não depende de  $Y_t^o$ .

Para amostrar dessa distribuição basta amostrar de cada parcela de (3.4). Primeiramente,  $(\Theta_{1t}^o, \Theta_{2t}, \psi | Y_t^o)$  pode ser amostrado pelo algoritmo descrito na seção anterior, notando que  $\Theta_{1t}^o, \Theta_{2t}, \psi$  são um subconjunto de  $\Phi$ . Depois podemos amostrar da distribuição  $(\Theta_{1t}^n | \Theta_{1t}^o, \Theta_{2t}, \psi)$ , usando o seguinte resultado obtido usando propriedades da distribuição Normal:

$$\begin{aligned}
& (\Theta_{1t}^n | \Theta_{1t}^o, \Theta_{2t}, \psi) \sim N(M, H, \Sigma) \text{ onde } M = F_{2t}^n \Theta_{2t} + (C^{no} \otimes V)(C^o \otimes V)^{-1}(\Theta_{1t}^o - F_{2t}^o \Theta_{2t}) \\
& \text{e } H = (C^n - C^{no}(C^o)^{-1}C^{on}) \otimes V.
\end{aligned}$$

Finalmente, podemos amostrar de  $(Y_t^n | \Theta_{1t}^n, \psi, Y_t^o)$ , usando um resultado análogo ao anterior:  $(Y_t^n | \Theta_{1t}^n, \psi, Y_t^o) \sim N(M^*, H^*, \Sigma)$ , onde  $M^* = F_{1t}^n \Theta_{1t}^n$  e  $H^* = I_{N_n}$

## 3.6 Aplicação com dados simulados

### 3.6.1 Descrição do modelo

Nessa seção apresentamos um exemplo com um conjunto de dados simulado, segundo o modelo descrito pela equação (3.2). Observações foram geradas para  $T = 70$  períodos de tempo e  $N = 100$  locais igualmente espaçados, com as coordenadas "latitude" e "longitude" variando de 1 a 10. Consideramos aqui duas variáveis resposta ( $q = 2$ ), tal que a matriz  $Y_t$  tem dimensão  $100 \times 2$ . Uma variável explicativa  $X_{i,t}$  foi gerada de uma distribuição  $U(0, 1)$ , para  $i = 1, \dots, 100$  locais no espaço, e  $t = 1, \dots, 70$  períodos de tempo, independentemente. Essa variável foi incluída na especificação de  $F_{1t}$ , que é uma matriz  $100 \times 200$ , definida como

$$F_{1t} = \begin{bmatrix} 1 & X_{1,t} & 0 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 1 & X_{2,t} & 0 & 0 & \cdots & 0 \\ & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & X_{100,t} \end{bmatrix}$$

para todo  $t = 1, \dots, 70$ .

$F_{2t}$  tem dimensão  $N_1 = 200 \times N_2 = 2$ , e é definida por  $F_{2t} = [1, 1, \dots, 1] \otimes I_2$ . Dessa forma, para cada particular elemento da matriz  $Y_t$ , temos a seguinte equação:

$$Y_{(i,j),t} = \alpha_{(i,j),t} + \beta_{(i,j),t} X_{it} + v_{1,(i,j),t}, \quad (3.5)$$

para  $i = 1, \dots, 100$  e  $j = 1, 2$ , onde  $Y_{(i,j),t}$  é o elemento da  $i$ -ésima linha e  $j$ -ésima coluna

de  $Y_t$ . A matriz de parâmetros  $\Theta_{1t}$  pode então ser escrita como:

$$\Theta_{1t} = \begin{bmatrix} \alpha_{(1,1),t} & \alpha_{(1,2),t} \\ \beta_{(1,1),t} & \beta_{(1,2),t} \\ \alpha_{(2,1),t} & \alpha_{(2,2),t} \\ \beta_{(2,1),t} & \beta_{(2,2),t} \\ \vdots & \vdots \\ \alpha_{(100,1),t} & \alpha_{(100,2),t} \\ \beta_{(100,1),t} & \beta_{(100,2),t} \end{bmatrix}$$

Para os elementos em  $\Theta_1$ , temos as seguintes equações:

$$\begin{aligned} \alpha_{(i,j),t} &= \delta_{j,t} + v_{2,(1+2(i-1),j),t} \\ \beta_{(i,j),t} &= \psi_{j,t} + v_{2,(2i,j),t} \quad i = 1, \dots, 100; \quad j = 1, 2 \end{aligned} \quad (3.6)$$

tal que a matriz de hiperparâmetros  $\Theta_{2t}$  é dada por

$$\Theta_{2t} = \begin{bmatrix} \delta_{1,t} & \delta_{2,t} \\ \psi_{1,t} & \psi_{2,t} \end{bmatrix} \quad (3.7)$$

$G_t$  é especificado por  $G_t = I_2$ , tal que  $\Theta_{2t}$  segue um passeio aleatório, definido por

$$\Theta_{2t} = \Theta_{2,t-1} + w_t \quad (3.8)$$

A função de correlação espacial que define a matriz  $C$ , foi especificada como sendo uma função exponencial de potência (De Oliveira et al., 1997), dada por

$$C(i, j) = \exp\{-(bd_{ij})^k\}, \quad \text{com } b > 0 \quad \text{e} \quad 0 < k < 2. \quad (3.9)$$



tal que  $\lambda = (b, k)$ , com  $b = 0, 1$  e  $k = 1, 2$ .

Finalmente, as demais matrizes de covariância são especificadas por

$$\Sigma = \begin{bmatrix} 1 & 0.2 \\ 0.2 & 1 \end{bmatrix}, V = \begin{bmatrix} 0.7 & 0 \\ 0 & 0.7 \end{bmatrix}, W = \begin{bmatrix} 0.5 & 0 \\ 0 & 0.5 \end{bmatrix}.$$

### 3.6.2 Estimação do modelo

Nesse exercício temos como objetivo comparar os valores verdadeiros dos parâmetros com os estimados através do algoritmo proposto, e também investigar a predição no tempo e interpolação no espaço para  $Y_t$ . Para isso escolhemos um conjunto de 16 estações de monitoramento, e supomos observados os 60 primeiros períodos de tempo nessas estações. Esse conjunto de dados foi usado para prever os últimos 10 períodos de tempo e interpolar nas 84 estações restantes.

As distribuições a priori, especificadas em (3.3), foram escolhidas de forma a não serem muito informativas, com

$$\begin{aligned} M_2 &= 0, & C_2 &= I_2, & T_0 &= T_{W_0} = T_{V_0} = 1 \\ S_0 &= I_2 & S_{W_0} &= I_2, & S_{V_0} &= I_2 \end{aligned} \tag{3.10}$$

As distribuições a priori para  $b$  e  $k$  também não são informativas, sendo especificadas por  $b \sim U[0, 2]$  e  $k \sim G(0.01, 0.01)$ . As distribuições propostas para gerar desses parâmetros no algoritmo de Metropolis-Hastings foram Normais truncadas com variância escolhida de forma a controlar a razão de aceitação. A cada passo do algoritmo, os valores gerados para  $b$  e  $k$  foram testados conjuntamente.

Após a convergência das cadeias, obtemos os resultados abaixo. As Figuras 3.1, 3.2,

3.3 e 3.4 mostram o histograma dos parâmetros  $b$  e  $k$ , e os elementos das matrizes  $\Sigma$ ,  $V$  e  $W$ , respectivamente. Podemos observar que todas as quantidades estão dentro do seu intervalo de credibilidade de 95%, com exceção de  $V(1, 2)$  e  $W(1, 2)$ , e talvez  $\Sigma(1, 1)$ . Os elementos nulos das matrizes  $V$  e  $W$  foram superestimados.

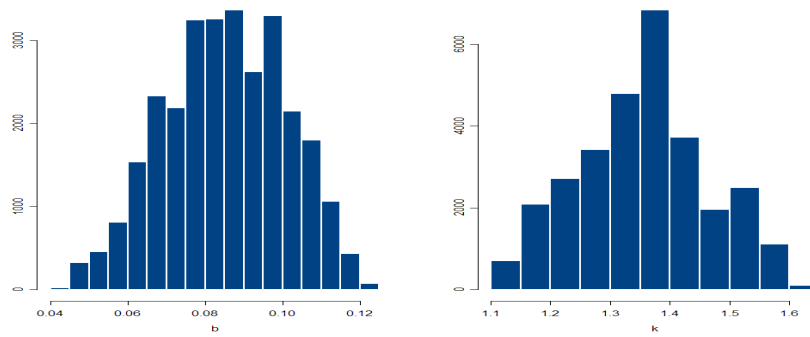


Figura 3.1: histograma da amostra da distribuição a posteriori dos parâmetros  $b$  e  $k$

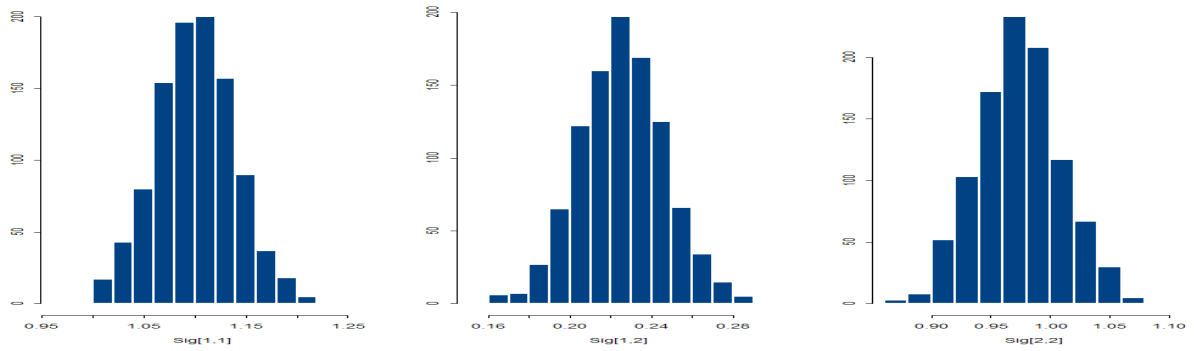


Figura 3.2: histograma da amostra da distribuição a posteriori dos elementos da matriz  $\Sigma$

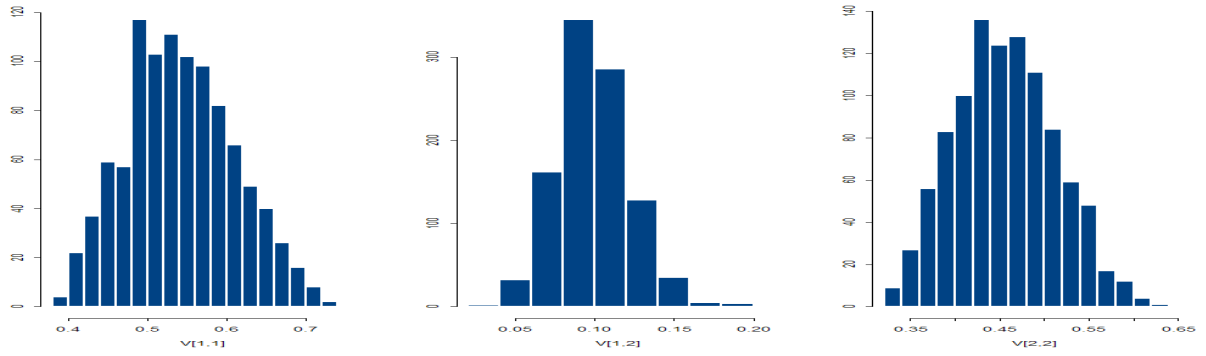


Figura 3.3: histograma da amostra da distribuição a posteriori dos elementos da matriz  $V$

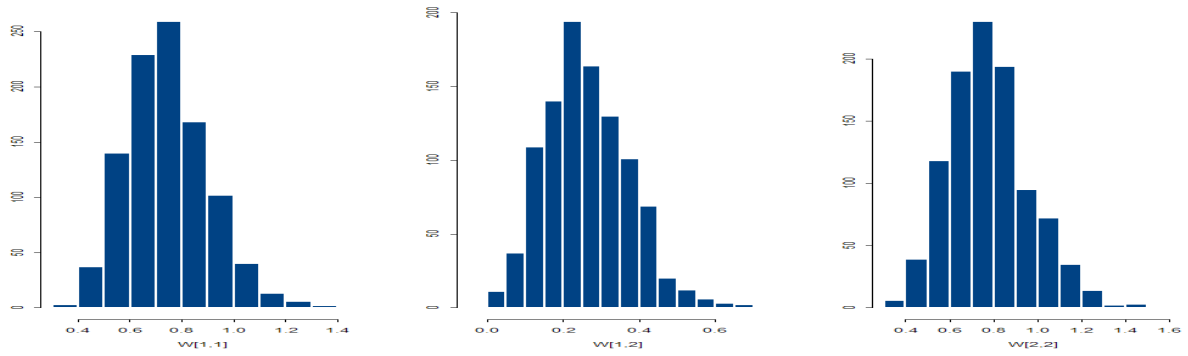


Figura 3.4: histograma da amostra da distribuição a posteriori dos elementos da matriz  $W$

A Figura 3.5 mostra a trajetória dos elementos verdadeiros e estimados de  $\Theta_{2t}$ , para  $t = 1, \dots, 70$ . Para  $t = 1, \dots, 60$ , os períodos de tempo "observados", os valores reais estão em sua maioria contidos nos intervalos de credibilidade de 95%. O modelo, entretanto, não faz boas previsões para  $t = 61, \dots, 70$ , visto que este é um modelo não estacionário de primeira ordem.

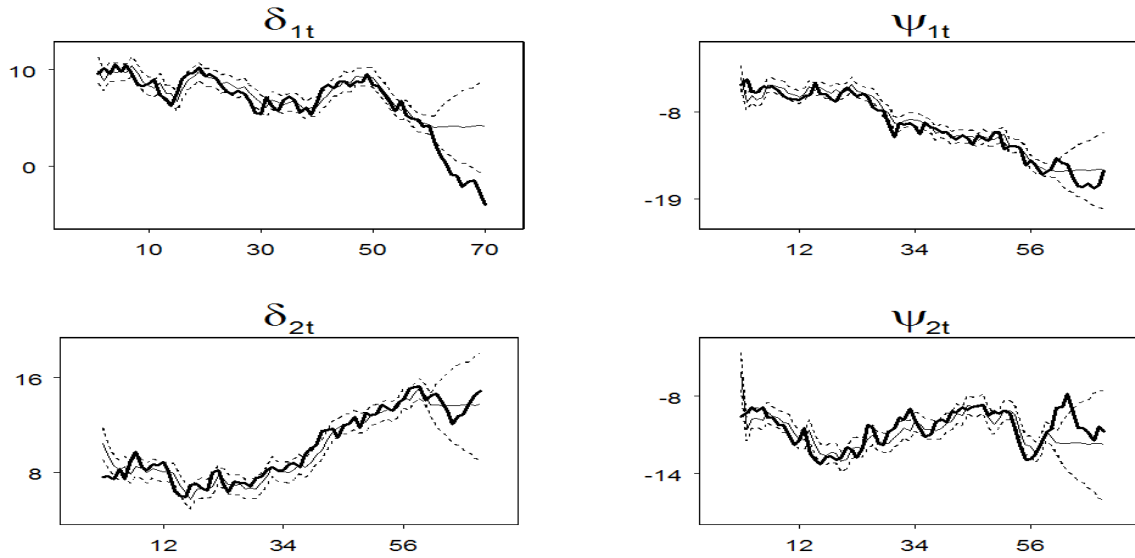


Figura 3.5: Quantis de 2.5%, 50% e 97.5% da amostra da distribuição a posteriori dos componentes de  $\Theta_{2t}$ ,  $t = 1, \dots, 70$ , com a trajetória verdadeira em negrito.

As figuras 3.6 e 3.7 mostram as superfícies reais e interpoladas dos parâmetros  $\alpha$  e  $\beta$  para  $t=61$ , e a Figura 3.8 mostra as superfícies reais e interpoladas para  $Y_{1,61}$  e  $Y_{2,61}$ . Apesar do padrão dos parâmetros interpolados não ser o mesmo dos parâmetros reais, o padrão das superfícies interpoladas de  $Y_{1,61}$  e  $Y_{2,61}$  é muito próximo ao verdadeiro. Isso mostra que o modelo tem dificuldade de diferenciar a variabilidade espacial trazida por  $\alpha$  e  $\beta$ , mas consegue capturar a variação espacial de  $Y$ .

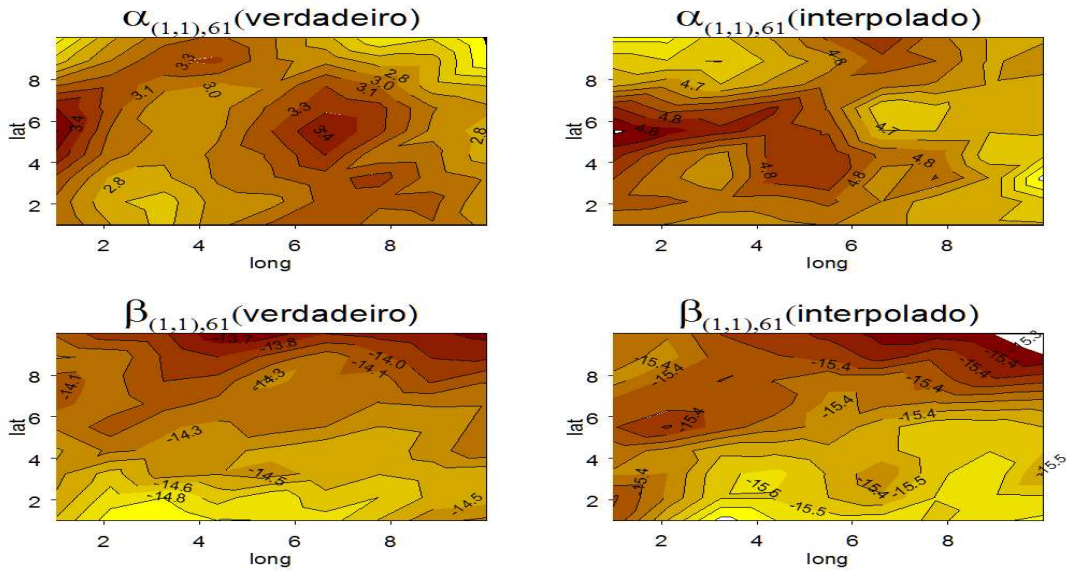


Figura 3.6: Superfícies reais e interpoladas para  $\alpha_{(1,1),61}$  e  $\beta_{(1,1),61}$

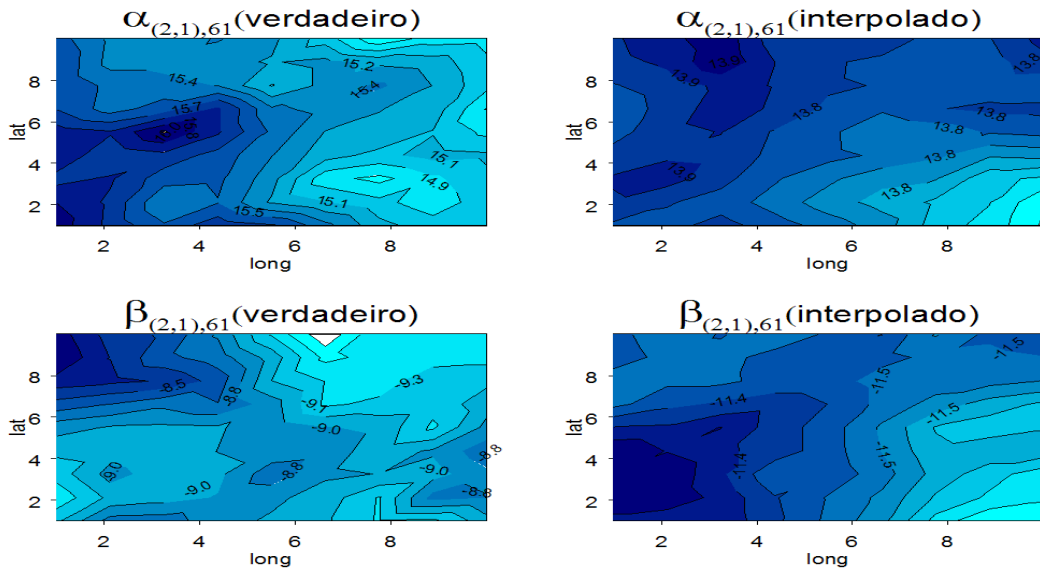


Figura 3.7: Superfícies reais e interpoladas para  $\alpha_{(2,1),61}$  e  $\beta_{(2,1),61}$

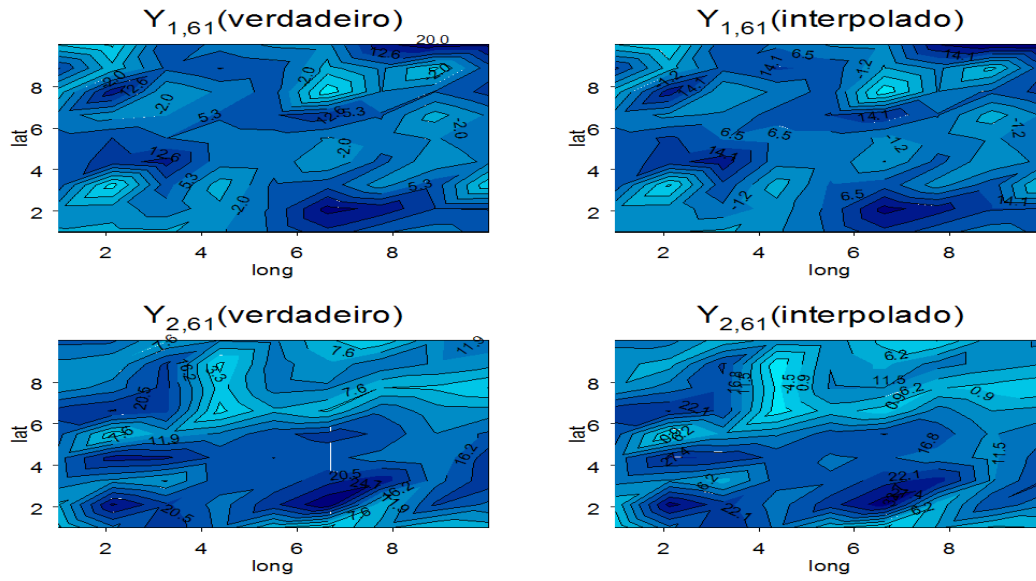


Figura 3.8: Superfícies reais e interpoladas para  $Y_{1,61}$  e  $Y_{2,61}$

### 3.7 Aplicação com dados de poluição

Nosso segundo exercício foi a aplicação do modelo (3.2) a dados de poluição na Alemanha, que foram medidos pela EMEP, um programa cooperativo para a monitoração e avaliação de transmissões de poluentes do ar na Europa. O conjunto de dados foi obtido através do site "www.emep.int", e consiste em duas variáveis resposta ( $SO_2$  e  $NO_2$ ), medidas mensalmente em 7 estações de monitoramento do ano de 1993 a 1999, totalizando 84 períodos de tempo, com distâncias medidas em unidade de 10kms. Também foi medido um índice de pluviosidade (em mm) para  $i = 1, \dots, 7$  e  $t = 1, \dots, 84$ , que servirá como variável explicativa para os níveis de  $SO_2$  e  $NO_2$ . A figura 3.9 mostra o mapa da Alemanha com a localização das estações, que se apresentam razoavelmente bem espalhadas na região de estudo.

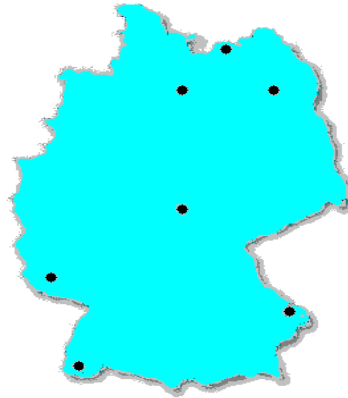


Figura 3.9: Mapa da Alemanha com a localização das estações de monitoramento

Trabalhamos com a transformação logarítmica das três variáveis observadas, pois análises preliminares mostraram que os dados transformados estão mais próximos de satisfazer a hipótese de normalidade. Sendo assim, a variável resposta  $Y$  é o logaritmo das medições de  $SO_2$  e  $NO_2$  ( $Y_1 = LN(SO_2)$  e  $Y_2 = LN(NO_2)$ ), e a variável explicativa  $X$  é o logaritmo do índice pluviométrico. Na figura 3.10 temos o gráfico de  $LN(SO_2)$  e  $LN(NO_2)$  ao longo do tempo, por estação de monitoramento, mostrando clara existência de um componente sazonal nos dados. O modelo proposto, entretanto, não trata essa sazonalidade de forma explícita, de forma que esta será explicada unicamente pelo componente diâmico no tempo. A figure 3.11 mostra o gráfico de  $Y_1$  contra  $Y_2$ , considerando todos os períodos de tempo e estações de monitoramento juntos. A correlação entre as variáveis resposta é alta, como pode ser visto na figura 3.10, sendo estimado em 0.7.

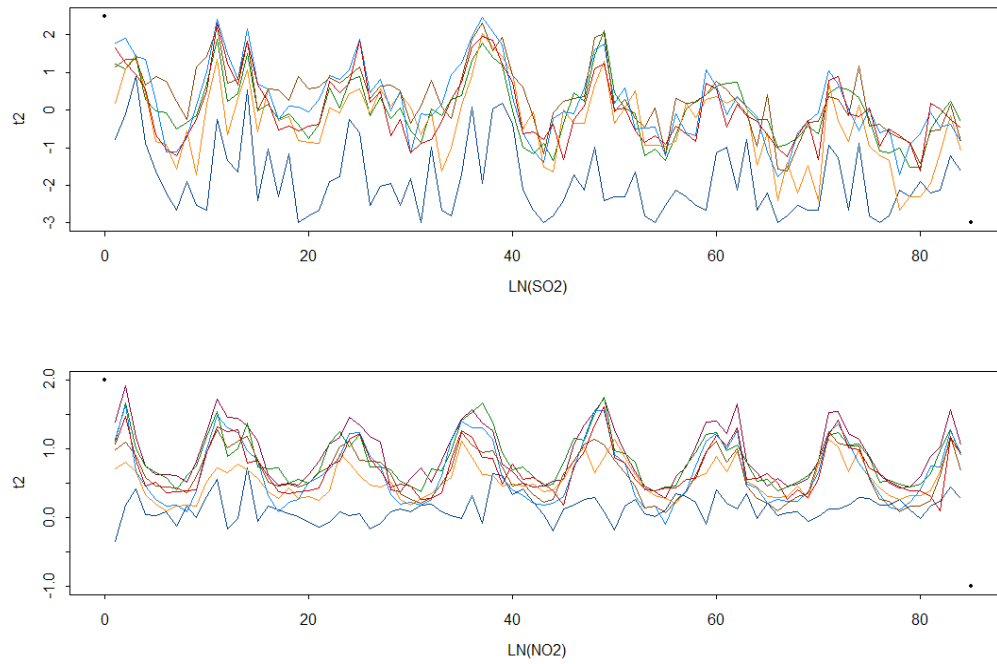


Figura 3.10: Variáveis resposta  $Y_1(LN(SO_2))$  e  $Y_2(LN(NO_2))$  ao longo do tempo, por estação de monitoramento.

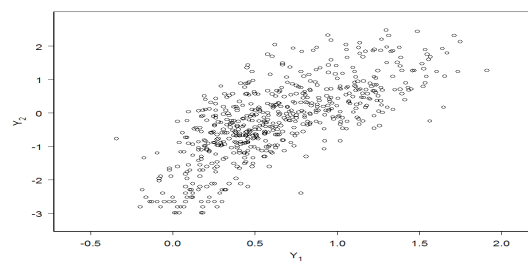


Figura 3.11: Plot de  $Y_1$  contra  $Y_2$



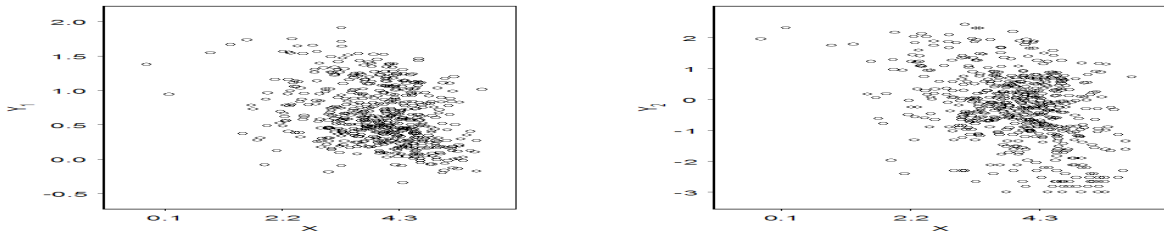


Figura 3.12: Plot de  $X$  contra  $Y_1$ ; plot de  $X$  contra  $Y_2$

Na Figura 3.11 temos os gráficos de  $X$  contra  $Y_1$  e  $X$  contra  $Y_2$ , mostrando correlações não muito altas entre as variáveis resposta e a variável explicativa. A estimativa dessas correlações é de  $-0.37$  (para  $Y_1$ ) e  $-0.31$  (para  $Y_2$ ). Análises preliminares mostram que a correlação entre os resíduos obtidos pela regressão linear de  $Y_1$  contra  $X$ , e  $Y_2$  contra  $X$ , é de  $0.69$ . Esse resultado faz sentido apenas para o caso em que as medições são independentes no tempo e no espaço.

O modelo proposto para  $Y$  é o modelo (3.2) com as mesmas especificações utilizadas no exemplo com dados simulados. Sendo assim, a estrutura espacial do modelo é incorporada na correlação entre os coeficientes de regressão, que nessa aplicação incluem o intercepto e o coeficiente angular relativo a  $X$ , como explicitado em (3.5). Como uma análise preliminar, foram feitas 14 regressões lineares (para cada variável resposta, e para cada estação monitoradora), supondo independência dos dados no tempo. A Figura 3.13 mostra os coeficientes estimados por essas regressões ao longo do espaço para  $Y_1$  e  $Y_2$ , mostrando uma provável estrutura espacial do intercepto e coeficiente angular da equação de regressão de  $Y_1$ . O intercepto da equação de  $Y_2$  parece se comportar de forma mais independente no espaço, enquanto o coeficiente angular permanece praticamente fixo.

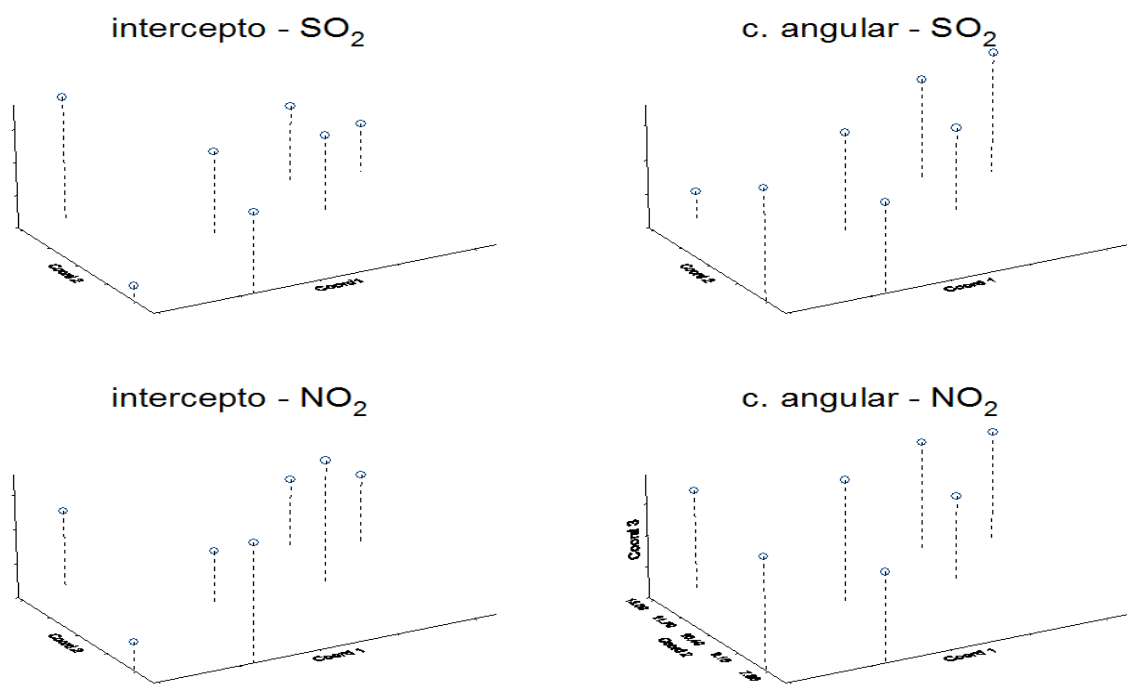


Figura 3.13: Coeficientes de regressão linear ao longo do espaço

Como no exemplo com dados simulados, as equações de regressão no primeiro e segundo níveis da hierarquia são especificadas pelas equações (3.5) e (3.6), e  $G$  se move no tempo de acordo com um passeio aleatório definido em (3.8). Sendo assim,  $F_{1t}$  e  $\Theta_{1t}$ ,  $t = 1, \dots, 84$ , são dados por

$$F_{1t} = \begin{bmatrix} 1 & X_{1,t} & 0 & 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & 1 & X_{2,t} & 0 & 0 & \dots & 0 \\ & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & X_{6,t} \end{bmatrix}, \quad \Theta_{1t} = \begin{bmatrix} \alpha_{(1,1),t} & \alpha_{(1,2),t} \\ \beta_{(1,1),t} & \beta_{(1,2),t} \\ \alpha_{(2,1),t} & \alpha_{(2,2),t} \\ \beta_{(2,1),t} & \beta_{(2,2),t} \\ \vdots & \vdots \\ \alpha_{(6,1),t} & \alpha_{(6,2),t} \\ \beta_{(6,1),t} & \beta_{(6,2),t} \end{bmatrix}$$

$\Theta_{2t}$  é dado em (3.7), e  $F_{2t}$  tem dimensão  $16 \times 2$ , e é dados por  $F_{2t} = [1, 1, \dots, 1] \otimes I_2$ . As matrizes  $\Sigma$ ,  $V$  e  $W$  são desconhecidas, e  $C$  é especificada pela equação (3.9), dependendo de parâmetros  $\lambda = (b, k)$ . Finalmente, o modelo é completado com a especificação das prioris, sendo  $b \sim U[0, 2]$  e  $k \sim G(0.01, 0.01)$ , e as restantes dadas em (3.3) com parâmetros especificados em (3.10).

Amostras das distribuições a posteriori dos parâmetros foram obtidas utilizando o algoritmo descrito na seção 3.3. As figuras 3.14, 3.15 e 3.16 mostram respectivamente os histogramas obtidos para os elementos das matrizes  $\Sigma$ ,  $V$  e  $W$ .

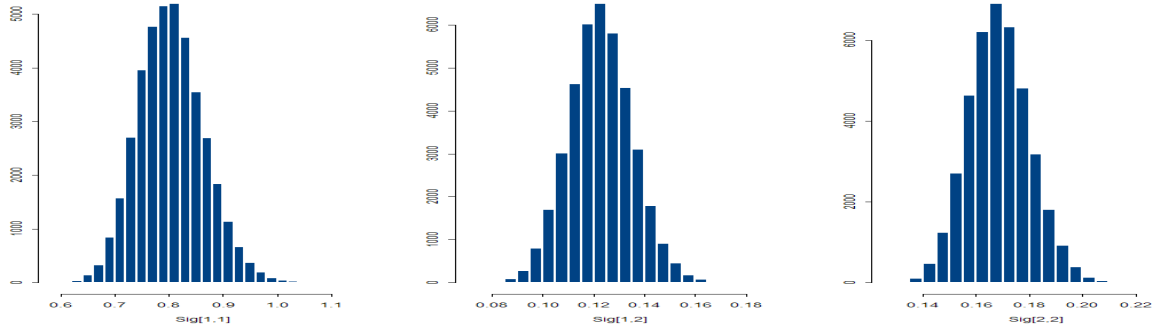


Figura 3.14: Histograma da amostra da distribuição a posteriori obtida para os elementos da matriz  $\Sigma$

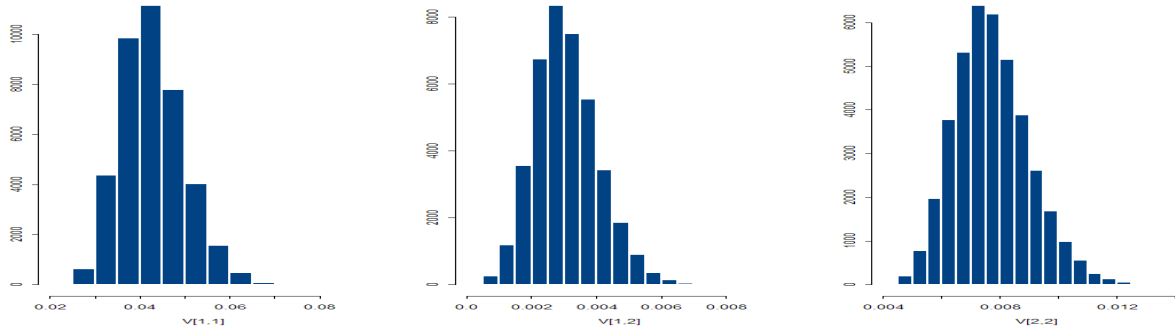


Figura 3.15: Histograma da amostra da distribuição a posteriori obtida para os elementos da matriz  $V$

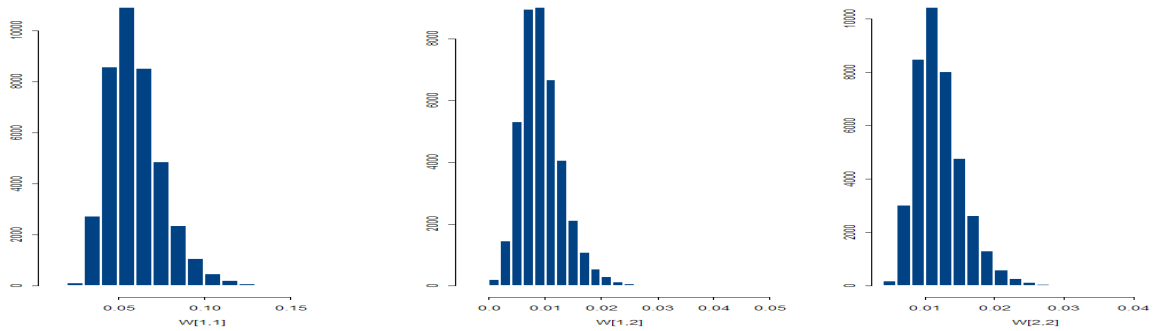


Figura 3.16: Histograma da amostra da distribuição a posteriori obtida para os elementos da matriz  $W$

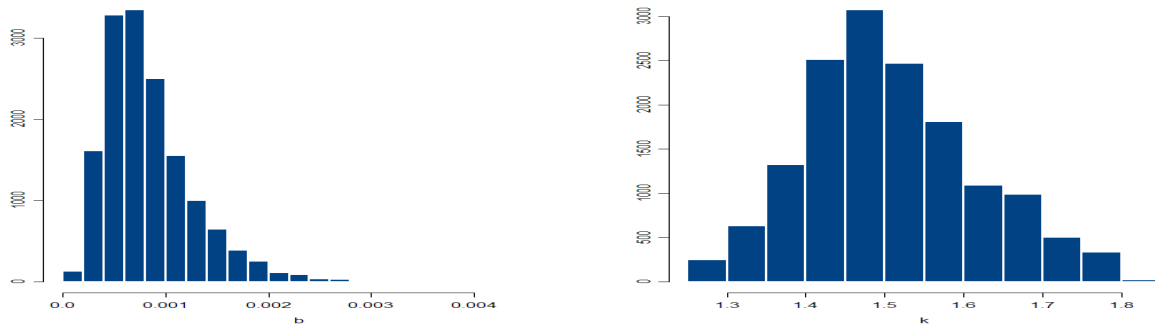


Figura 3.17: Histograma da amostra da distribuição a posteriori obtida para os parâmetros  $b$  e  $k$

Supondo que nenhuma correlação espacial e temporal esteja presente nos dados, seria esperado obtermos uma estimativa para  $\Sigma_{1,2}$  próxima ao valor da correlação estimada entre os resíduos obtidos pela regressão linear simples de  $Y_1$  e  $Y_2$  contra  $X$ , que foi de 0.69. O histograma da amostra da posteriori de  $\Sigma_{1,2}$ , no entanto, está concentrada em torno do valor 0.22, estando bem longe de incluir o valor 0.69, como pode ser visto na Figura 3.17. Dessa forma podemos concluir que uma boa parte da correlação entre as variáveis resposta, tirando o efeito da variável explicativa, foi explicada pela estrutura espaço temporal do modelo. A Figura 3.15 mostra o histograma dos parâmetros  $b$  e  $k$  da função de covariância que define a matriz  $C$ . Visto que a distância entre as estações de monitoramento varia de 70 a 870 km, as amostras geradas das posterioris de  $b$  e  $k$  definem uma matriz de correlações  $C$  com valores muito altos, variando em média de 0.89 a 0.99.

A Figura 3.18 mostra a média da distribuição a posteriori dos elementos da matriz  $\Theta_2$  ao longo do tempo. Como era de se esperar, a sazonalidade presente nas variáveis resposta está sendo absorvida por esses elementos, que se mostram claramente sazonais.

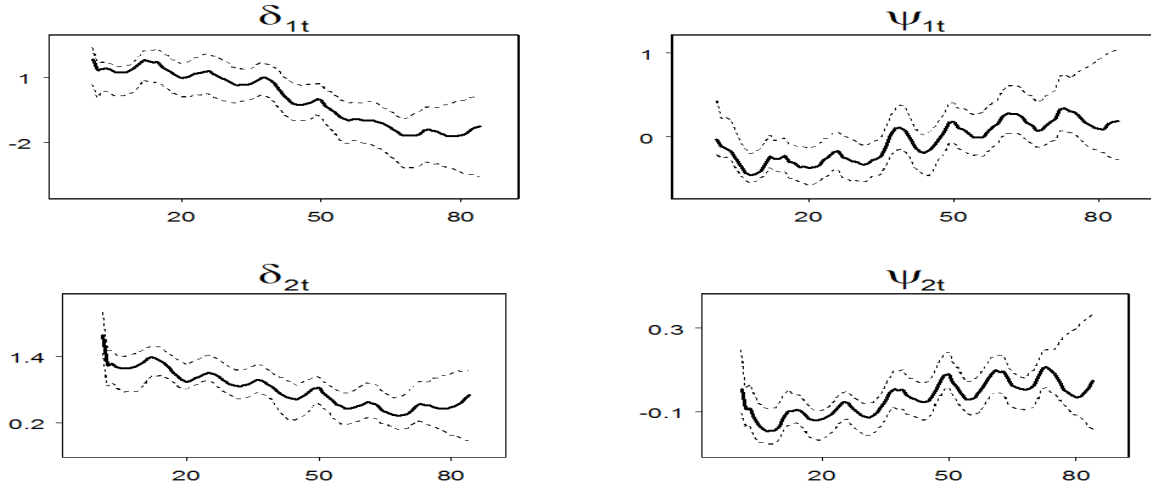


Figura 3.18: Quantis de 2.5%, 50% e 97.5% da amostra da distribuição a posteriori dos elementos de  $\Theta_{2t}$ ,  $t = 1, \dots, 84$ .

Não foram obtidos resultados satisfatórios na interpolação e previsão dos processos  $Y_1$  e  $Y_2$ , e por esse motivo esses resultados não serão reportados nessa análise. Fica clara a necessidade de modelar a sazonalidade dos dados de forma paramétrica, de forma a produzir previsões realistas no tempo. Além disso, seria provavelmente mais realistas (como sugerido pela Figura 3.11), relaxar a restrição de que a correlação espacial nos elementos de  $\Theta_1$  é a mesma ( $C$ ). Também é importante ressaltar a dificuldade do modelo de perceber uma estrutura espacial baseado em apenas 7 estações de monitoramento.

### 3.8 Conclusão

Nesse capítulo foi proposto um modelo dinâmico hierárquico para respostas multivariadas, sendo nosso objetivo principal ter um modelo capaz de perceber a estrutura espaço-temporal presente nos dados, permitindo fazer interpolação espacial e previsão temporal. O modelo aqui proposto é uma restrição do modelo proposto em Landim e Gamerman (2000), onde adotamos uma especificação paramétrica para a estrutura de covariância

espacial para os coeficientes de regressão no primeiro nível da hierarquia. Com essa restrição, temos um modelo de regressão multivariado cujos parâmetros além de variarem de forma dinâmica no tempo, variam de forma suave no espaço. Esse modelo tem a vantagem de permitir a interpolação das variáveis resposta e coeficientes de regressão para qualquer conjunto de pontos não observados, levando em consideração a variação espacial.

Trabalhamos sob o ponto de vista Bayesiana, propondo um algoritmo de MCCM para a geração de amostras das posteriores dos parâmetros desconhecidos do modelo. Foram feitas aplicações a dois conjuntos de dados. Primeiramente trabalhamos com dados simulados, gerando duas variáveis resposta e uma variável explicativa em 100 locais igualmente espaçados e 70 períodos de tempo. Consideramos apenas 16 locais no espaço e 60 períodos de tempo para fazer inferência sobre os parâmetros do modelo, e interpolar e prever para os locais e tempos restantes. Obtivemos resultados em geral satisfatórios para a inferência dos parâmetros. Os parâmetros  $\lambda$ , que definem a matriz de covariância espacial dos coeficientes de regressão  $\Theta_1$ , foram bem estimados, apesar de serem parâmetros geralmente de difícil estimação. Os elementos fora da diagonal principal das matrizes de covariância  $V$  e  $W$ , entretanto, foram superestimados. Considerando o período de tempo fixo em  $t = 61$ , as superfícies interpoladas das variáveis  $Y_1$  e  $Y_2$  apresentaram um padrão espacial bem próximo ao padrão "verdadeiro", apesar do mesmo não ter ocorrido com as superfícies dos parâmetros  $\alpha$  e  $\beta$ . Isso mostra a dificuldade que o modelo tem de diferenciar a variabilidade espacial proveniente desses parâmetros.

O segundo conjunto de dados analisado é um conjunto de dados reais de poluição na Alemanha. As variáveis resposta,  $LN(SO_2)$  e  $LN(NO_2)$ , e a variável explicativa, logaritmo do índice pluviométrico (mm), foram medidos em 7 estações de monitoramento em 83 períodos de tempo. As variáveis resposta apresentam uma forte sazonalidade que não foi tratada de forma paramétrica pelo modelo, de forma que esta foi absorvida pelo componente dinâmico temporal. O modelo não se mostrou adequado para fazer previsão

temporal e interpolação no espaço para esse conjunto de dados, provavelmente devido à sazonalidade dos dados e pouca quantidade de informação na variação espacial (apenas 7 estações monitoradoras). Análises preliminares também mostraram que supor que a estrutura de correlação espacial seja a mesma em todos os parâmetros de regressão não seja talvez adequada. Uma possível solução para esse problema é trabalhar com uma extensão desse modelo, levando em conta diferentes estruturas de correlação espacial para os coeficientes de regressão, e incluindo um componente sazonal que explique parte da variação temporal presente nos dados de forma paramétrica.



# Capítulo 4

## Análise de Processos Temporais de Cox e a Influência da Agregação dos Dados Observados

### 4.1 Introdução

Nesse capítulo estudamos modelos que tratam a variação temporal da incidência de eventos. Um exemplo típico é o estudo da incidência de doenças. Para problemas desse tipo, pode-se analisar os próprios tempos de observação de eventos, ou contagens do número de eventos em intervalos de tempo disjuntos. Nessa pesquisa exploramos diferentes níveis de agregação da variável resposta e conseqüentemente diferentes quantidades de informação trazidas pelo conjunto de dados, analisando os efeitos obtidos no ajuste do modelo proposto.

Para modelar respostas não normais, Breslow e Clayton (1993) propõem os Modelos Lineares Generalizados Mistos (MLGMs), extendendo os Modelos Lineares Generaliza-

dos de McCullagh and Nelder (1989). Uma abordagem Bayesiana para os MLGM's pode ser vista em Gamerman (1997b). Os MLGM's são úteis para acomodar sobredispersão, que é muitas vezes observada em respostas binomiais (Williams, 1982) ou poisson (Breslow, 1984), e também para fazer estimativas em problemas de múltiplos parâmetros, como a construção de mapas da incidência de doenças em pequenas áreas (Clayton e Kaldor, 1987; Manton et al, 1989). Christensen e Waagepetersen (2001) consideram os MLGM's para modelar a intensidade do número de sementes presentes no solo em uma determinada região. Condicionamente a um campo aleatório gaussiano  $S$ , o número de sementes é por hipótese independente no espaço e tem distribuição de poisson. O modelo log-poisson resultante pode ser visto como uma aproximação para a distribuição associada a um processo pontual de Cox log gaussiano (Moller, Syversveen and Weegepetersen, 1998; Brix and Moller, 2001).

Os processos de Cox são largamente utilizados para modelar padrões pontuais, visto eles muitas vezes refletem a heterogeneidade presente em dados ambientais. Brix e Diggle (2001) propõem um processo de Cox num estudo simulado, desenvolvendo métodos baseados em momentos para a estimação dos parâmetros desconhecidos do modelo. Esse artigo é motivado pelo problema de monitorar possíveis mudanças no padrão de incidência espaço-temporal de infecções gastro-intestinais, dado os registros feitos diariamente de casos ocorridos por local no espaço. Esse é o objeto de interesse do projeto AEGISS (Ascertainment Enhancement of Gastrointestinal Infection Surveillance and Statistics), que tem como objetivo o uso de métodos estatísticos para identificar anomalias na distribuição espaço-temporal das doenças gastro-intestinais na Grã-Bretanha.

Nesse capítulo propomos um processo de Cox contínuo no tempo para modelar infecções gastro-intestinais sob uma abordagem Bayesiana. Utilizamos para isso um conjunto de dados do projeto AEGISS, onde o número de casos de infecções é registrado com base diária. Como os dados não são observados de forma contínua no tempo (não há informação sobre o horário em que cada caso foi reportado), é necessário fazer apro-

ximações para o ajuste do modelo. Um dos nossos objetivos é entender como essa falta de informação afeta a modelagem e a predição da incidência de doenças para tempos futuros.

Trabalhamos também com um conjunto de dados simulado, analisamos a performance da estimação do processo de Cox considerando variáveis resposta em diferentes níveis de agregação: casos observados em tempo contínuo, e contagens observadas a cada dia e a cada semana. Predições para tempos futuros também foram feitas com base em cada um dos casos, a critério de comparação.

Na seção 4.2 é definido o processo de Cox, e apresentamos as abordagens 1 e 2, que consideram respectivamente respostas contínuas e discretas (agregadas em intervalos de tempo), com resultados de inferência. A seção 4.3 descreve os métodos de MCCM (Monte Carlo em Cadeias de Markov) utilizados para a estimação dos parâmetros do modelo sob as duas diferentes abordagens. Na seção 4.4 apresentamos um exercício com processos simulados, comparando a performance de modelagem e previsão sob diferentes níveis de agregação da variável resposta. Na seção 4.5 temos uma breve descrição do projeto AEGISS e dos dados sobre doenças gastro-intestinais, e a aplicação do modelo a esse conjunto de dados. Finalmente as conclusões e extensões são apresentadas na seção 4.6.

## 4.2 Processo de Cox

O Processo de Poisson homogêneo desempenha um papel importante na teoria de processos pontuais. Entretanto, por ser muito simplista, não é geralmente uma representação realista de processos reais. Uma extensão natural pode ser obtida considerando a função de intensidade do processo de Poisson como uma realização de um campo aleatório. Essa generalização é conhecida como processo de Cox, e é definida por Diggle (2003) como se segue:

**Definição 1.** Seja  $S \in R^d$ . Suponha que  $\{Z = Z(\xi) : \xi \in S\}$  é um campo aleatório não negativo tal que com probabilidade um,  $\xi \rightarrow Z(\xi)$  é uma função localmente integrável. Se a distribuição condicional de  $X$  dado  $Z$  é um processo de poisson em  $S$  com função de intensidade  $Z$ , então  $X$  segue um processo de Cox guiado por  $Z$ .

Aqui assumimos um processo temporal, com  $S \in R^+$ . Em nosso caso particular, definimos a função de intensidade  $\Lambda$  como o produto da intensidade populacional  $\rho$ , que será considerada conhecida, e uma função de risco  $\Pi$ . Propomos um modelo log linear para a função de risco  $\Pi$ , incorporando um processo gaussiano e covariáveis que tratam a variação temporal não explicada por  $\rho$ . Esse processo é chamado processo log-gaussiano, e discutido em Moller et al. (1998). Temos que

$$\Lambda(t) = \rho(t)\Pi(t), \quad t \in R^+ \quad (4.1)$$

onde  $\rho(t)$  é conhecido,  $\Pi(t) = \exp[\gamma(t) + \beta X(t)]$ ,  $X(t)$  é um vetor de tamanho  $p$  de variáveis explicativas no tempo  $t$ ,  $\beta$  são coeficientes desconhecidos associados a  $X$ , e  $\gamma$  é um processo temporal que varia suavemente no tempo. Fixamos  $E[\exp(\gamma(t) + \beta X(t))] = 1$ , de forma que  $\exp(\gamma(t))$  e  $\beta X(t)$  têm o papel de modelar desvios da média do processo de  $\rho(t)$ .

A dependência de  $\gamma$  no tempo é especificada por uma estrutura autoregressiva, dada por:

$$[\gamma(t_j) + 0.5\sigma^2] = \varphi_{t_i, t_j}[\gamma(t_i) + 0.5\sigma^2] + e(t_i, t_j), \quad e(t_i, t_j) \sim N(0, \tau_{t_i, t_j}^2). \quad (4.2)$$

onde  $\varphi_{t_i, t_j}$  é a correlação entre  $\gamma(t_i)$  e  $\gamma(t_j)$ , e  $\tau_{t_i, t_j}^2 = \sigma^2(1 - \varphi_{t_i, t_j}^2)$ ,  $t_i < t_j \in R^+$ . Note que  $E[\gamma(t)] = 0.5\sigma^2$  é uma restrição imposta para assegurar que  $E[\exp(\gamma(t) + \beta X(t))] = 1$ , para todo  $t$ .

Aqui  $\varphi$  é definido como uma função exponencial com parâmetro de decaimento  $\phi$ , que é proporcional à distância temporal entre dois eventos:

$$\varphi_{t_i, t_j} = \exp\left[-\frac{t_i - t_j}{\phi}\right], \quad t_i < t_j \in R^+ \quad \phi > 0. \quad (4.3)$$

Sendo assim,  $\tau_{t_i, t_j}^2 = \sigma^2[1 - \exp(-2(t_j - t_i)/\phi)]$ .

A seguir, nas subseções 4.2.1 e 4.2.2, apresentamos duas abordagens para o processo de Cox, baseadas em variáveis resposta contínuas e agregadas em intervalos de tempo. Em muitos casos, dados são reportados de hora em hora, diariamente, por semana, etc. Um dos nossos objetivos nesse projeto é comparar a modelagem baseada no conjunto completo de tempos de observação, com a modelagem baseada em conjuntos de dados agregados. A perda de precisão devido à agregação depende fortemente dos valores dos parâmetros do modelo combinados com o grau de agregação dos dados.

### 4.2.1 Abordagem 1: respostas no tempo contínuo

Seja  $z$  o conjunto de tempos  $z = (z(1), \dots, z(T))$  de ocorrência de casos de uma certa doença. Assume-se que  $z$  é uma realização de um processo de Cox  $Z$ , cuja intensidade é dada pelo mesmo produto dado em (4.1):  $\Lambda(z(t)) = \rho(z(t))\Pi(z(t))$ ,  $t = 1, \dots, T$ ,  $z(t) \in R^+$ , onde  $\rho(z(t))$  é conhecido, e  $\Pi(z(t)) = \exp[\gamma(z(t)) + \beta X(z(t))]$ .

É fácil mostrar que dado  $\Lambda$  e  $z(t-1)$ , a função de distribuição acumulada do tempo de ocorrência do próximo evento  $z(t)$  é a exponencial da integral da função de intensidade

$\Lambda$  entre os períodos de tempo  $z(t-1)$  e  $z(t)$ :

$$F(z(t)|z(t-1), \Lambda) = \exp\left(-\int_{z(t-1)}^{z(t)} \Lambda(s)ds\right). \quad (4.4)$$

A dependência de  $\gamma$  no tempo é definida em (4.2). Dados  $\phi$  e  $\sigma^2$ ,  $z(t)$  somente dependerá das observações passadas através de  $z(t-1)$ , como mostra a equação abaixo:

$$\begin{aligned} [\gamma(z(t)) + 0.5\sigma^2] &= \varphi_{z(t-1), z(t)}[\gamma(z(t-1)) + 0.5\sigma^2] + e(z(t-1), z(t)), \quad (4.5) \\ e(z(t-1), z(t)) &\sim N(0, \tau_{z(t-1), z(t)}^2), \quad t = 2, \dots, T \end{aligned}$$

onde  $\varphi_{z(t-1), z(t)}^2 = \exp[-(z(t) - z(t-1))/\phi]$ ,  $\tau_{z(t-1), z(t)}^2 = \sigma^2(1 - \varphi_{z(t-1), z(t)}^2)$  e  $\gamma(z(1)) \sim N(-0.5\sigma^2, \sigma^2)$ .

### Inferência: Distribuições a Priori e Posteriori

A inferência feita sob o ponto de vista Bayesiano começa com a especificação de distribuições a priori para as quantidades desconhecidas. No modelo descrito em (4.4) e (4.5), essas quantidades são os parâmetros  $\beta$ , o processo  $\gamma$ , e os hiperparâmetros  $\phi$  e  $\sigma^2$ . A priori para  $\beta_i$ ,  $i = 1, \dots, p$  é especificada por  $\beta_i \sim N(\mu_b, \sigma_b^2)$ . Para  $\phi$  especificamos uma priori uniforme no intervalo  $(a_\phi, b_\phi)$ . Especificamos uma distribuição a priori para a precisão  $\sigma^{-2}$  ao invés de especificar uma para  $\sigma^2$ . Essa é uma prática comum pois geralmente leva à propriedade de conjugação quando assume-se uma priori gamma para o inverso da precisão. Apesar deste não ser o caso em nosso modelo, pois  $\sigma^2$  também afeta a média do processo  $\gamma$ , seguiremos a especificação usual de que  $\sigma^{-2} \sim G(a_\sigma, b_\sigma)$ .

Dados  $\beta$ ,  $\phi$ ,  $\sigma^2$  e  $\gamma$ , a função de distribuição de  $z$  pode ser escrita como:

$$f(z|\beta, \gamma, \phi, \sigma^2) = \prod_{t=2}^T f(z(t)|z(t-1), \beta, \gamma, \phi, \sigma^2) f(z(1)|\beta, \gamma, \phi, \sigma^2) \quad (4.6)$$

onde  $F(z(t)|z(t-1), \beta, \gamma, \phi, \sigma^2)$  é dado em (4.4).

Agora seja  $\gamma^*$  o conjunto da valores de  $\gamma$  associados aos tempos de observação de eventos, ou seja,  $\gamma^* = [\gamma(z(1)), \gamma(z(2)), \dots, \gamma(z(T))]$ , com  $\gamma^*(t) = \gamma(z(t))$ ,  $t = 1, \dots, T$ , e seja  $\gamma^{-*}$  todos os  $\gamma$  restantes, definidos por  $\gamma^{-*} = \{\gamma - \gamma^*\}$ , tal que  $\gamma = (\gamma^*, \gamma^{-*})$ .  $f(z|\beta, \gamma, \phi, \sigma^2)$  somente depende de  $\gamma$  através de  $\gamma^*$ , como visto pela equação (4.6). A posteriori de  $(\beta, \gamma, \phi, \sigma^2)$  pode ser escrita explorando essa propriedade:

$$f(\beta, \gamma, \phi, \sigma^2|z) \propto f(\beta, \gamma^*, \phi, \sigma^2|z) f(\gamma^{-*}|\gamma^*, \phi, \sigma^2) \quad (4.7)$$

$f(\beta, \gamma^*, \phi, \sigma^2|z)$  é encontrado combinando as prioris descritas previamente e a verossimilhança em (4.6):

$$f(\beta, \gamma^*, \phi, \sigma^2|z) \propto f(z|\beta, \gamma^*, \phi, \sigma^2) f(\gamma^*|\phi, \sigma^2) f(\phi) f(\sigma^2) f(\beta) \quad (4.8)$$

Dados  $\phi$  e  $\sigma^2$ , a distribuição de  $\gamma^*$  é completamente especificada, e pode ser escrita por

$$f(\gamma^*|\phi, \sigma^2) \propto f(\gamma^*(T)|\gamma^*(T-1), \phi, \sigma^2) f(\gamma^*(T-1)|\gamma^*(T-2), \phi, \sigma^2) \dots f(\gamma^*(1)|\sigma^2) \quad (4.9)$$

com  $(\gamma^*(t)|\gamma^*(t-1), \phi, \sigma^2)$  dado em (4.5).

$f(\gamma^*|\gamma^*, \phi, \sigma^2)$  pode ser encontrado extendendo a idéia usada para se obter  $f(\gamma(k)|\gamma^*, \phi, \sigma^2)$ ,  $k \in (z(t-1), z(t))$ ,  $t = 2, \dots, T$ . É fácil mostrar que a distribuição de  $\gamma(k)$  dados  $\gamma^*$ ,  $\phi$  e  $\sigma^2$  somente depende de  $\gamma^*$  através de  $\gamma^*(t)$  e  $\gamma^*(t-1)$ . Essa distribuição pode ser escrita como

$$f(\gamma(k)|\gamma^*, \phi, \sigma^2) = f(\gamma(k)|\gamma^*(t), \gamma^*(t-1), \phi, \sigma^2) \propto f(\gamma^*(t)|\gamma(k), \phi, \sigma^2) f(\gamma(k)|\gamma^*(t-1), \phi, \sigma^2)$$

Usando propriedades da distribuição normal, temos que

$$[\gamma(k)|\gamma^*(t), \gamma^*(t-1), \phi, \sigma^2] \sim N \left( \frac{\gamma^*(t-1)\tau_{z(t-1),k}^{-2} + \gamma^*(t)\tau_{k,z(t)}^{-2}}{\tau_{z(t-1),k}^{-2} + \tau_{k,z(t)}^{-2}}, \frac{1}{\tau_{z(t-1),k}^{-2} + \tau_{k,z(t)}^{-2}} \right) \quad (4.10)$$

Não há forma fechada para a posteriori em (4.8), portanto é necessário usar métodos numéricos para amostrar dessa distribuição. Usamos métodos de MCCM que são descritos na seção 4.3. Uma vez obtidas amostras das distribuições a posteriori dos parâmetros, pode-se fazer inferência a respeito dos mesmos.

## Predição

Em geral, outro grande interesse nas aplicações é o de prever respostas futuras dadas as respostas observadas no presente. Sob o ponto de vista Bayesiano, os valores a serem previstos são incorporados de forma natural no modelo, sendo tratados como qualquer outro parâmetro desconhecido. Toda a informação necessária para a predição é a distribuição a posteriori dadas as observações passadas.



Suponha que estamos interessados em prever o tempo de ocorrência do próximo caso a ser reportado, isto é, queremos obter a distribuição preditiva de  $z(T+1)$ .  $z(T+1)$  só depende dos valores  $z$  através de  $z(T)$ , e a densidade acumulada de  $z(T+1)|z(T)$  pode ser escrita como:

$$F(z(T+1)|z(T)) = \int \exp \left[ - \int_{z(T)}^{z(T+1)} \Lambda(s) d(s) \right] d\beta d\phi d\sigma^2, \quad (4.11)$$

onde  $\Lambda(T+1)$  é dado por  $\Lambda(T+1) = \rho(T+1)\Pi(T+1)$ , sendo  $\Pi(T+1) = \exp[\gamma^*(T+1) + \beta X(z(T+1))]$ , com  $\gamma^*(T+1) = \gamma(z(T+1))$ . A distribuição de  $\gamma^*(T+1)$  dados os valores passados  $\gamma^*(1), \dots, \gamma^*(T)$  só depende de  $\gamma^*(T)$  e é dada por

$$\begin{aligned} [\gamma^*(T+1) + 0.5\sigma^2] &= \varphi_{z(T), z(T+1)} [\gamma^*(T) + 0.5\sigma^2] + e(z(T), z(T+1)) \\ e(z(T), z(T+1)) &\sim N(0, \tau_{z(T), z(T+1)}^2), \end{aligned} \quad (4.12)$$

onde  $\varphi_{z(T), z(T+1)}$  é definido em (4.3) e  $\tau_{z(T), z(T+1)}^2 = \sigma^2(1 - \varphi_{z(T), z(T+1)}^2)$ .

A distribuição preditiva em (4.11) não tem uma forma fechada, e sendo assim é necessário utilizar métodos numéricos para obter uma amostra dessa distribuição.

## 4.2.2 Abordagem 2: dados de contagem

Suponha que temos uma realização do processo de Cox  $Z$  definido na seção 4.2.1, mas não temos acesso à informação do momento exato quando cada caso foi reportado. Ao invés disso os dados são observados agregados em intervalos de tempo, e nosso processo resposta é o número de casos reportados durante esses intervalos. Sem perda de generalidade,

vamos chamar cada um desses intervalos de "dia". Seja  $y(i)$  o número de casos da doença em questão observados no dia  $i$ ,  $i = 1, \dots, T$ . A série temporal  $y = y(1), \dots, y(T)$  é uma realização do processo de poisson  $Z$  com função de intensidade dada por  $\int_{t-1}^t \Lambda(s) ds = \int_{t-1}^t \rho(s) \Pi(s) ds$ ,  $t = 1, \dots, T$

Pode ser mostrado que

$$y(t) \sim \text{Pois} \left( \int_{t-1}^t \Lambda(s) ds \right), t = 1, \dots, T \quad (4.13)$$

onde  $\text{Pois}(\lambda)$  denota a distribuição de poisson com intensidade  $\lambda$ ,  $\rho(t)$  é conhecido, e  $\Pi(t) = \exp[\gamma(t) + \beta X(t)]$ .

Como as observações são feitas diariamente, o modelo não faz distinção entre  $\Lambda(t_i)$  e  $\Lambda(t_j)$  se  $t_i$  e  $t_j$  são períodos de tempo pertencentes a um mesmo dia. Essa perda de precisão se dá pela falta de informação do conjunto de dados agregados, que não mais informa o tempo de ocorrência de cada evento. Dessa forma, não há nenhuma perda adicional em aproximar  $\int_{t-1}^t \Lambda(s) ds$  por  $\Lambda_{apx}(t - 1)$ . A equação (4.13) pode ser aproximada por:

$$y(t) \sim \text{Pois}(\Lambda_{apx}(t)), t = 1, \dots, T \quad (4.14)$$

Note que a interpretação para  $\Lambda_{apx}(t)$  não é a intensidade  $\Lambda$  no início do dia  $t$ , mas sim a intensidade média durante o dia  $t$ .

O processo temporal  $\gamma$  também é aproximado por uma constante para cada dia:  $\int_{t-1}^t \gamma(s) ds \approx \gamma_{apx}(t - 1)$ . A dependência de  $\gamma$  no tempo é especificada pela mesma estrutura autoregressiva dada em (4.2), mas agora definida no tempo discreto:

$$[\gamma_{apx}(t) + 0.5\sigma^2] = \varphi[\gamma_{apx}(t-1) + 0.5\sigma^2] + e(t), \quad e(t) \sim N(0, \tau^2), \quad t = 2, \dots, T \quad (4.15)$$

onde  $\varphi = \exp(-1/\phi)$  é a correlação entre  $\gamma_{apx}(t)$  e  $\gamma_{apx}(t-1)$ , que é constante para todo  $t = 2, \dots, T$ , e  $\tau^2 = \sigma^2(1 - \varphi^2)$ .

### Inferência: Distribuições a Priori e Posteriori

As quantidades desconhecidas no modelo descrito por (4.14) e (4.15) são  $\beta$ ,  $\gamma_{apx}$ ,  $\phi$  e  $\sigma^2$ . Mais uma vez especificamos prioris para  $\beta_i$ ,  $i = 1, \dots, p$ ,  $\phi$  e  $\sigma^{-2}$  como  $\beta_i \sim N(\mu_b, \sigma_b^2)$ ,  $\phi \sim U(a_\phi, b_\phi)$  e  $\sigma^{-2} \sim G(a_\sigma, b_\sigma)$ .

Vamos definir  $\gamma_{apx} = (\gamma_{apx}(1), \gamma_{apx}(2), \dots, \gamma_{apx}(T))$ . A distribuição de  $\gamma_{apx} | \sigma^2, \phi$  é uma versão regularmente espaçada de (4.9), e é dada por

$$f(\gamma_{apx} | \phi, \sigma^2) \propto f(\gamma_{apx}(T) | \gamma_{apx}(T-1), \phi, \sigma^2) f(\gamma_{apx}(T-1) | \gamma_{apx}(T-2), \phi, \sigma^2) \dots f(\gamma_{apx}(1) | \sigma^2) \quad (4.16)$$

onde  $(\gamma_{apx}(t) | \gamma_{apx}(t-1), \phi, \sigma^2)$  é especificado em (4.15).

Dados  $\beta$ ,  $\phi$ ,  $\sigma^2$  e  $\gamma_{apx}$ ,  $y$  tem distribuição de poisson, definindo a seguinte função de verossimilhança:

$$l(\beta, \gamma_{apx}, \phi, \sigma^2; y) \propto \prod_{i=1}^T (\rho(i) \exp[\gamma_{apx}(i) + \beta X])^{y_i} \exp[-\rho(i) \exp[\gamma_{apx}(i) + \beta X]]$$

A distribuição a posteriori para o conjunto de parâmetros  $(\beta, \gamma_{apx}, \phi, \sigma^2)$  pode ser encontrada pelo produto dessa verossimilhança e das prioris especificadas previamente:

$$f(\gamma_{apx}, \phi, \sigma^2 | y) \propto l(\beta, \gamma_{apx}, \phi, \sigma^2; y) f(\gamma_{apx} | \phi, \sigma^2) f(\sigma^2) f(\phi) f(\beta) \quad (4.17)$$

Essa distribuição a posteriori, assim como a definida em (4.8), não possui forma fechada, e métodos de MCMC (descritos na seção 4.3) são usados para obter amostras da mesma.

### Predição

Suponha que estamos interessados em prever o número de casos que serão reportados no tempo  $T + 1$ . Supondo  $X(T + 1)$  conhecido, e dados os parâmetros  $\gamma(T + 1)$  e  $\beta$ ,  $y(T + 1)$  tem distribuição de poisson, como em (4.14), com parâmetro  $\Lambda_{apx}(T + 1) = \rho(T + 1)\Pi(T + 1)$ , onde  $\Pi(T + 1) = \exp[\gamma(T + 1) + \beta X(T + 1)]$ . A distribuição preditiva  $f(y(T + 1) | y(T))$  não tem forma fechada.

A distribuição de  $\gamma(T + 1)$  dadas as observações passadas  $\gamma(1), \dots, \gamma(T)$  só depende de  $\gamma(T)$  e pode ser escrita como

$$[\gamma(T + 1) + 0.5\sigma^2] = \phi[\gamma(T) + 0.5\sigma^2] + e(T + 1), \quad e(T + 1) \sim N(0, \tau^2).$$

Mais uma vez é necessário utilizar métodos numéricos para obter amostras de  $y(T + 1)$  e fazer inferência sobre esse parâmetro.

## 4.3 Estimação por métodos de MCCM

Devido à complexidade das densidades a posteriori em (4.8) e (4.17), métodos numéricos devem ser usados para estimar os parâmetros desconhecidos do modelo. Aqui usamos métodos de MCCM para resolver esse problema, tendo como objetivo propor um algoritmo simples, mas capaz de fornecer resultados em um período razoável de tempo.

### 4.3.1 Estimando parâmetros sob a abordagem 1

A primeira abordagem, descrita na seção 4.2.1, possui três parâmetros desconhecidos ( $\sigma$ ,  $\phi$  e  $\beta$ ), além de um processo dinâmico desconhecido  $\gamma$ . A densidade a posteriori  $f(\beta, \gamma, \phi, \sigma^2 | z)$  pode ser escrita pelo produto de  $f(\beta, \gamma^*, \phi, \sigma^2 | z)$  e  $f(\gamma^{-*} | \gamma^*, \phi, \sigma^2)$ , como visto em (4.7). Combinando esse resultado com (4.8), essa posteriori pode ser escrita como:

$$f(\beta, \gamma, \phi, \sigma^2 | z) \propto f(z | \beta, \gamma^*, \phi, \sigma^2) f(\gamma^{-*} | \gamma^*, \phi, \sigma^2) f(\gamma^* | \phi, \sigma^2) f(\phi) f(\sigma^2) f(\beta)$$

Assim sendo, podemos obter uma amostra de  $f(\beta, \gamma, \phi, \sigma^2 | z)$  gerando de cada parcela do produto acima. Propomos passos de Metropolis-Hastings para gerar de  $\sigma^{-2}$ ,  $\phi$  e  $\beta$ .  $\sigma^{-2}$  e  $\phi$  são testados conjuntamente para evitar problemas de correlação, e os parâmetros  $\beta$  são testados conjuntamente.  $\gamma^* = (\gamma^*(1), \dots, \gamma^*(T))$  também são amostrados por Metropolis-Hastings, mas devido à grande dimensão de T (T é tipicamente grande), esses parâmetros não são testados conjuntamente. Depois que amostramos valores para  $\gamma^*$ ,  $\phi$  e  $\sigma^2$ ,  $f(\gamma^{-*} | \gamma^*, \phi, \sigma^2)$  pode ser obtido usando propriedades da distribuição normal, e  $f(z | \beta, \gamma^*, \phi, \sigma^2)$  pode ser amostrado usando as equações (4.4) e (4.5).

Os passos do algoritmo de MCCM são descritos abaixo:

- **Dê valores iniciais para os parâmetros**

Seja  $i$  o índice que indica a posição atual da cadeia. O primeiro passo é fazer  $i = 0$ , e dar valores iniciais para os parâmetros  $\beta^{(0)}$ ,  $\sigma^{-2(0)}$ ,  $\phi^{(0)}$  e  $\gamma^{*(0)}$ .

- **Amostre  $\beta$**

Faça  $i = i + 1$ .

$\beta$  é amostrado por Metropolis-Hastings. A proposta é um passeio aleatório, centrado no valor anterior de  $\beta$  na cadeia, ou seja, o valor proposto para  $\beta^{(i)}$  é amostrado de uma  $N(\beta^{(i-1)}, h_\beta)$ , onde  $h_\beta$  é a variância que controla a razão de aceitação.

- **Amostre  $\sigma^{-2}$  e  $\phi$**

Se a média do processo  $\gamma$  fosse conhecida, e não dependesse de  $\sigma^2$ , seria fácil mostrar que uma priori gamma para  $\sigma^{-2}$  seria conjugada, ou seja, a posteriori de  $\sigma^{-2}$  também teria distribuição gamma. Dessa forma,  $\sigma^{-2}$  poderia ser gerada pelo amostrador de Gibbs. A proposta escolhida para  $\sigma^{-2}$  é baseada na sua condicional completa dado que  $E[\gamma^*(t)]$ ,  $t = 1, \dots, T$  é conhecido.

Vamos supor que  $E[\gamma^*(t)] = -0.5\sigma^{*2}$ ,  $t = 1, \dots, T$ , onde  $\sigma^{*2}$  é conhecida. Note que esse é o mesmo modelo que tínhamos anteriormente, apenas substituindo  $\sigma^2$  por  $\sigma^{*2}$  na equação que define a esperança de  $\gamma$ . Sob esse modelo, a condicional completa de  $\eta = \sigma^{-2}$  pode ser escrita como

$$f(\eta|\gamma, \phi, \sigma^{*2}) = f(\eta|\gamma^*, \phi, \sigma^{*2}) \propto \frac{\eta^{1/2}}{\sqrt{2\pi}} \exp\left(-\frac{\eta}{2}(\gamma^*(1) + 0.5\sigma^{*2})^2\right) \cdot \prod \frac{\eta^{1/2}}{\sqrt{2\pi}\sqrt{1-\phi^2}} \exp\left(-\frac{\eta}{2(1-\phi^2)}(\gamma^*(t) - \gamma^*(t-1) - (\phi-1)0.5\sigma^{*2})^2\right) \eta^{a_\sigma-1} \exp(-b_\sigma\eta) \propto \eta^{T/2+a_\sigma-1} \exp\left(-\eta\left[\frac{(\gamma^*(1) + 0.5\sigma^{*2})^2}{2} + \frac{(\gamma^*(t) - \gamma^*(t-1) - (\phi-1)0.5\sigma^{*2})^2}{2(1-\phi^2)} + b_\sigma\right]\right).$$

Sendo assim,  $(\eta|\gamma^*, \phi, \sigma^{*2})$  tem distribuição  $G(a^*, b^*)$  onde  $a^* = T/2 + a_\sigma$  e  $b^* =$

$$\frac{(\gamma^*(1)+0.5\sigma^{*2})^2}{2} + \frac{(\gamma^*(t)-\gamma^*(t-1)-(\phi-1)0.5\sigma^{*2})^2}{2(1-\phi^2)} + b_\sigma.$$

A proposta para  $\eta$  nos passos de Metropolis-Hastings é  $G(a^*, b^*)$ , com  $\sigma^{*2}$  dado pelo valor anterior de  $\sigma^2$  na cadeia. Em outras palavras, o valor proposto para  $\eta^{(i)}$  é gerado de  $f(\eta|\gamma^{*(i-1)}, \phi^{(i-1)}, \sigma^{2(i-1)})$ . A proposta para  $\phi$  é um passeio aleatório centrado no seu valor anterior :  $\phi^{(i)} \sim N(\phi^{(i-1)}, h_\phi)$ , com  $h_\phi$  sendo especificado de forma a controlar a razão de aceitação. O movimento proposto para os parâmetros  $(\eta, \phi)$  a cada passo da cadeia é testado conjuntamente.

- **Amostre**  $\gamma^*(t), t = 1, \dots, T$

A proposta para cada  $\gamma^*(t), t = 1, \dots, T$  é um passeio aleatório centrado em seu valor anterior na cadeia. Cada elemento de  $\gamma^*$  é testado separadamente, pois se torna extremamente difícil aceitar um passo de um bloco de parâmetros de alta dimensionalidade, que é o caso aqui, já que  $T$  é um número tipicamente alto. O custo de propor gerações separadas para cada  $\gamma^*(t)$  é o de possivelmente acarretar em alta correlação entre os valores amostrados de cada elemento de  $\gamma^*$  na cadeia, causando uma razão de convergência baixa. Uma possibilidade de se lidar com esse problema é dividir  $\gamma^*$  em blocos menores e testar cada bloco conjuntamente. Como não houve problemas de convergência em nossos modelos simulados e na aplicação com dados reais, não exploramos essa possibilidade.

O algoritmo proposto para gerar de  $\gamma^*$  é dado a seguir:

1. faça  $t = 0$
2. faça  $t = t + 1$  se  $t < T$ , caso contrário fim do algoritmo.
3. gere  $\gamma^{*(i)}(t)$  de uma  $N(\gamma^{*(i-1)}(t), h_\gamma)$ , onde  $h_\gamma$  é a variância que ajusta a razão de aceitação.
4. a proposta gerada para  $\gamma^{*(i-1)}(t)$  é aceita com probabilidade

$$\min \left( 1, \frac{h'(\gamma^{*(i)}(t)|z(t), z(t-1), z(t+1), \gamma^{*(i)}(t-1), \gamma^{*(i-1)}(t+1), \phi^{(i)}, \sigma^{2(i)})}{h'(\gamma^{*(i-1)}(t)|z(t), z(t-1), z(t+1), \gamma^{*(i)}(t-1), \gamma^{*(i-1)}(t+1), \phi^{(i)}, \sigma^{2(i)})} \right),$$

onde

$$\begin{aligned} h'(\gamma^*(t)|z(t), z(t-1), z(t+1), \gamma^*(t-1), \gamma^*(t+1), \phi, \sigma^2) = \\ \frac{-[\Lambda(z(t)) + \Lambda(z(t-1))]}{2} \exp \left[ \frac{-[\Lambda(z(t)) + \Lambda(z(t-1))]}{2} (z(t) - z(t-1)) \right] \\ \frac{-[\Lambda(z(t+1)) + \Lambda(z(t))]}{2} \exp \left[ \frac{-[\Lambda(z(t+1)) + \Lambda(z(t))]}{2} (z(t+1) - z(t)) \right] \end{aligned}$$

Detalhes de como essas equações foram obtidas podem ser vistos no apêndice.

5. Volte para o passo 2.

- **Execute o algoritmo até obter convergência**

O algoritmo deve ser executado até a convergência da cadeia para distribuições estáveis. Depois disso obtemos amostras das distribuições a posteriori de cada parâmetro, e inferência pode ser feita a respeito deles.

Para fazer previsão para valores futuros, basta adicionar um passo ao algoritmo descrito acima. Supondo  $X$  conhecido, e tendo gerado  $\beta^{(i)}, \gamma^{*(i)}(1), \dots, \gamma^{*(i)}(T), \sigma^{2(i)}$  e  $\phi^{(i)}$ , geramos  $\gamma^{(i)}(z(T) + \delta), \gamma^{(i)}(z(T) + 2\delta), \dots, \gamma^{(i)}(z(T) + k_T\delta)$ , para  $k_T \in Z^+$  suficientemente grande, de acordo com a equação (4.2). Daí, pode-se obter  $\Pi^{(i)}(z(T) + k\delta) = \exp[\gamma^{(i)}(z(T) + k\delta) + \beta^{(i)}X(z(T) + k\delta)]$ , e  $\Lambda^{(i)}(z(T) + k\delta) = \rho(z(T) + k\delta)\Pi^{(i)}(z(T) + k\delta)$ , para  $k = 1, \dots, k_T$ .

Geramos  $x^{(k)}, k = 1, \dots, k_T$ , de  $f(x|[z(T) + (k-1)\delta], [\Lambda^{(i)}(z(T) + (k-1)\delta)])$ , que pode ser aproximada por uma densidade exponencial, usando a seguinte aproximação para a equação (4.4):

$$F(x^{(k)}|[z(T) + (k-1)\delta], [\Lambda^{(i)}(z(T) + (k-1)\delta)]) = \exp \left[ -\delta\Lambda^{(i)}(z(T) + (k-1)\delta) \right]. \quad (4.18)$$



As gerações são feitas sucessivamente, até obter  $x^{(k^*)} < z(T) + k\delta$ . O valor gerado para  $z(T + 1)$  é dado por  $z(T + 1) = x^{(k^*)}$ .

Previsões h passos a frente podem ser feitas de forma similar, gerando amostras sucessivas de  $z(T + 1), \dots, z(T + h)$ .

### 4.3.2 Estimando parâmetros sob a abordagem 2

O algoritmo utilizado para obter amostras dos parâmetros do modelo para a segunda abordagem é muito similar à usada na abordagem 1. As quantidades desconhecidas são as mesmas que tínhamos anteriormente:  $\sigma$ ,  $\phi$ ,  $\beta$  e  $\gamma$ . O algoritmo de MCCM é dado abaixo:

- **Dê valores iniciais para os parâmetros**

Dê valores iniciais para os parâmetros  $\beta^{(0)}$ ,  $\sigma^{-2(0)}$ ,  $\phi^{(0)}$  e  $\gamma^{*(0)}$ . Faça  $i = 0$ .

- **Amostre  $\beta$ ,  $\sigma^{-2}$  e  $\phi$**

Faça  $i = i + 1$ . Os parâmetros  $\beta^{(i)}$ ,  $\sigma^{-2(i)}$  e  $\phi^{(i)}$  são amostrados como no algoritmo descrito na seção 4.3.1.

- **Amostre  $\gamma(t)$ ,  $t = 1, \dots, T$**

A amostragem de  $\gamma^{(i)}(t)$ ,  $t = 1, \dots, T$  é feita de forma similar à da seção 4.3.1, não havendo aqui, entretanto, a necessidade de usar aproximações. Utilizamos o seguinte algoritmo:

1. faça  $t = 0$

2. faça  $t = t + 1$  se  $t < T$ , caso contrário fim do algoritmo.

3. gere  $\gamma^{(i)}(t)$  de uma  $N(\gamma^{(i-1)}(t), h_\gamma)$ , onde  $h_\gamma$  é a variância que ajusta a razão de aceitação.

4. aceitamos a proposta para  $\gamma^{(i-1)}(t)$  com probabilidade dada abaixo :

$$\min \left( 1, \frac{h^*(\gamma^{(i-1)}(t+1)|\gamma^{(i)}(t), \phi^{(i)}, \sigma^{2(i)})h(\gamma^{(i)}(t)|\gamma^{(i)}(t-1), \phi^{(i)}, \sigma^{2(i)})}{h^*(\gamma^{(i-1)}(t+1)|\gamma^{(i-1)}(t), \phi^{(i)}, \sigma^{2(i)})h(\gamma^{(i-1)}(t)|\gamma^{(i)}(t-1), \phi^{(i)}, \sigma^{2(i)})} \right)$$

onde  $h^*(\gamma^{(i)}(t)|\cdot)$  é a densidade da distribuição  $f(\gamma^{(i)}(t)|\cdot)$ , definida em (4.14), avaliada em  $\gamma^{(i)}(t)$ .

5. Volte para o passo 2.

Fazer previsão para o número de casos em tempos futuros é trivial nesse caso. Para gerar uma amostra de  $y(T+1)$ , basta adicionar um passo no algoritmo. Supondo  $X$  conhecido, e tendo gerado  $\beta^{(i)}, \gamma^{(i)}(1), \dots, \gamma^{(i)}(T), \sigma^{2(i)}$  e  $\phi^{(i)}$ , geramos  $\gamma^{(i)}(T+1)$  segundo a equação 4.15, e obtemos  $\Pi^{(i)}(T+1) = \exp[\gamma^{(i)}(T+1) + \beta^{(i)}X(T+1)]$  e  $\Lambda^{(i)}(T+1) = \rho(T+1)\Pi^{(i)}(T+1)$ . Finalmente podemos gerar  $y^{(i)}(T+1)$  de uma distribuição de poisson, definida na equação (4.14), com  $\Lambda_{apx} = \Lambda^{(i)}(T+1)$ .

Previsões h passos a frente podem ser feitas de forma similar, gerando amostras sucessivas de  $y^{(i)}(T+1), \dots, y^{(i)}(T+h)$ .

## 4.4 Análise de dados simulados

Nessa seção estudamos o efeito de agregar o processo resposta, através de um experimento simulado, e validamos o algoritmo proposto comparando o valor verdadeiro dos parâmetros com os estimados.

Considere o modelo descrito em (4.3). Para simular desse modelo é necessário discretizar o tempo onde o processo é definido, em intervalos pequenos, tal que  $\Lambda(t)$  possa ser considerado constante dentro de cada um deles. O conceito de "intervalo pequeno" não é claro nesse contexto, e vai depender do processo que está sendo descrito. Quanto

mais alta a intensidade do processo, menor deve ser o tamanho de cada intervalo de forma a assegurar uma aproximação razoável.

Simulamos uma realização do processo durante  $T_1 = 800$  dias, supondo  $\Lambda(t) = \rho(t)\exp(\gamma(t))$  constante para cada intervalo de hora  $t = 1, \dots, 19200$ , que é uma aproximação razoável para o processo aqui gerado. Variáveis explicativas não foram incluídas nesse modelo. A intensidade populacional  $\rho$  é definida pela combinação linear de senos e cossenos, com periodicidade de sete dias, e considerada conhecida nessa análise. Os outros parâmetros são especificados por  $\sigma^2 = 2$ ,  $\phi = 200$ , e o processo  $\gamma(t)$ ,  $t = 1, \dots, 19200$  é simulado recursivamente segundo as distribuições  $\gamma^{(0)}$  e  $(\gamma^*(t)|\gamma^*(t-1), \sigma, \phi)$  dadas em (4.5). A Figura 4.1 mostra  $\exp(\gamma(t))$  e  $\rho$  ao longo do tempo, assim como intensidade resultante do produto dessas séries,  $\lambda$ , mostrando alta correlação entre  $\exp(\gamma)$  e  $\lambda$ .

O processo simulado originou  $T=6855$  casos. Nosso objetivo nessa análise é comparar a inferência feita sob as Abordagens 1 e 2 descritas na seção 4.2. Para isso vamos definir  $z(i)$  como sendo o tempo de observação do caso  $i$ ,  $i=1, \dots, T$ , tal que  $z = (z(1), z(2), \dots, z(T))$ ;  $y_1(j)$  é definido como o número de casos no dia  $j$ ,  $j = 1, \dots, T_1$ , tal que  $y_1 = (y_1(1), y_1(2), \dots, y_1(T_1))$ ; e finalmente definimos  $y_2$  como o número de casos na semana  $k$ ,  $k = 1, \dots, T_2$ , tal que  $y_2 = (y_2(1), y_2(2), \dots, y_2(T_2))$  e  $T_2 = 114$ .

A modelagem é feita para três casos, primeiramente considerando  $z$  conhecido (abordagem 1), depois considerando  $z$  desconhecido e  $y_1$  conhecido (abordagem 2.1), e por fim considerando apenas  $y_2$  conhecido (abordagem 2.2). A seguir temos os resultados obtidos em cada caso.

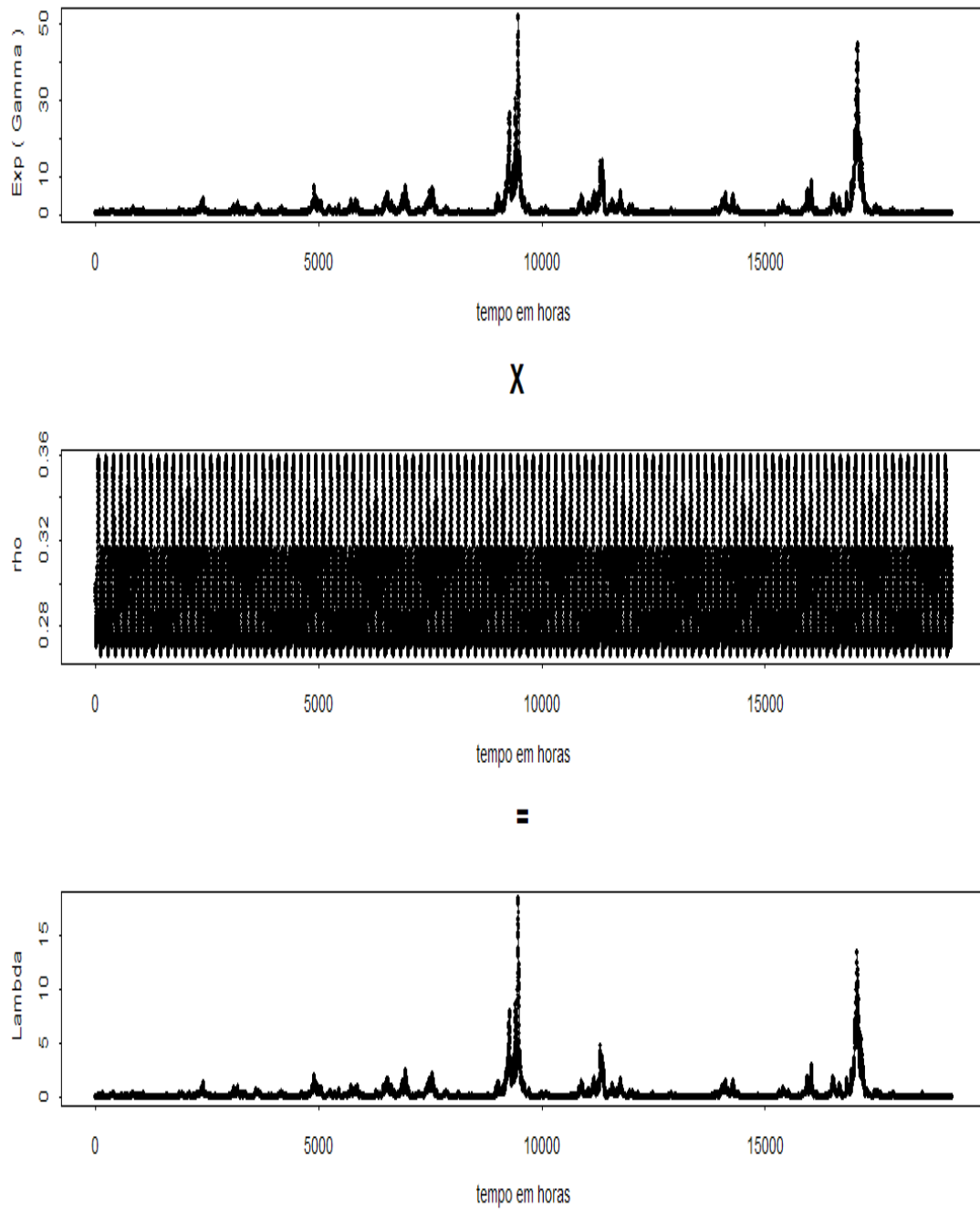


Figura 4.1: valores simulados dos processos  $\text{exp}(\gamma)$ ,  $\rho$  e  $\lambda$  ao longo do tempo.

### 4.4.1 Modelando $z$ - abordagem 1

Suponha que  $z$ , os tempos de ocorrência dos casos de interesse, são observados, de modo que é possível trabalhar com respostas contínuas. O modelo proposto para  $z$  é especificado pelas equações (4.4) e (4.5), e os parâmetros são estimados usando os métodos de MCCM descritos na seção 4.3, considerando as prioris definidas na seção 4.2.1, com parâmetros escolhidos de forma que elas não sejam muito informativas:

$$\sigma^{-2} \sim G(0.1, 0.1) \quad \phi \sim U(0, 1000)$$

As estimativas para os parâmetros  $\sigma^2$  e  $\rho$ , obtidas pela média das suas respectivas amostras da posteriori, são dadas por  $\hat{\sigma}^2 = 2.22$  e  $\hat{\rho} = 199.07$ . A figura 4.2 apresenta os histogramas das amostras da posteriori obtidos para esses parâmetros, mostrando que há uma grande massa de probabilidade ao redor dos valores verdadeiros,  $\sigma^2 = 2$  e  $\rho = 200$ .

A Figura 4.3 mostra a trajetória do processo  $\gamma$  (verdadeiro) no tempo, com os intervalos de credibilidade de 95% da amostra a posteriori desses parâmetros. O intervalo contém a maioria dos valores verdadeiros, com exceção de alguns pontos, que estão em sua maioria abaixo de -4. A justificativa para isso está no fato de que há menos informação nos dados que correspondem à trajetória da intensidade do processo quando esta apresenta valores baixos, visto que intensidades baixas tipicamente geram poucos casos.

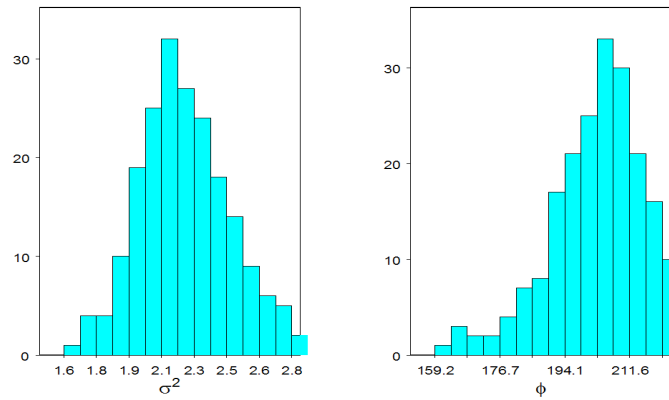
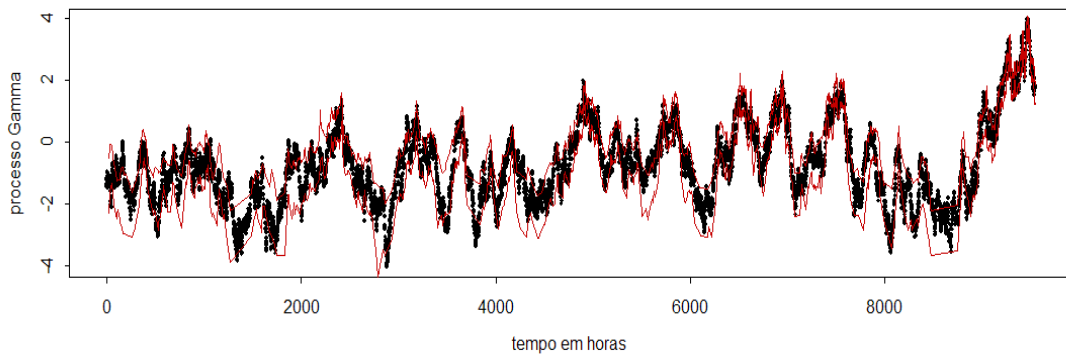


Figura 4.2: Histograma das amostras da posteriori de  $\sigma^2$  e  $\phi$

(A)



(B)

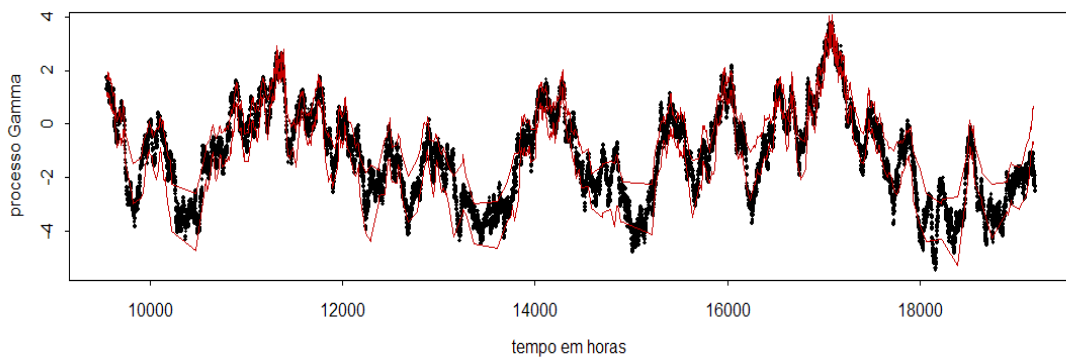


Figura 4.3: Trajetória real de  $\gamma$  com intervalos de credibilidade da amostra da posteriori (em vermelho): (A) nos tempos 1 a 9600; (B) nos tempos 9601 a 19200.

Por fim apresentamos resultados de predição para o número de ocorrências nas 2184 últimas horas (13 últimas semanas), supondo na análise que esses valores são desconhecidos. A Figura 4.4 mostra os valores verdadeiros e o Intervalo de credibilidade de 95% obtido sob dois diferentes cenários. Primeiramente sem atualizar a informação recebida, ou seja, os valores observados permanecem constantes (cenário A). No segundo caso o algoritmo é atualizado a cada novo caso registrado (cenário B). Podemos ver pela figura que sob o cenário A a maioria do número de casos obtidos nas primeiras 500 horas não está contida nos intervalos de credibilidade, mostrando a inadequação do modelo para prever mudanças no processo muitos passos a frente. A previsão sob o cenário B se mostra bem mais eficiente, dando origem a intervalos de credibilidade que contém a grande maioria dos valores reais.

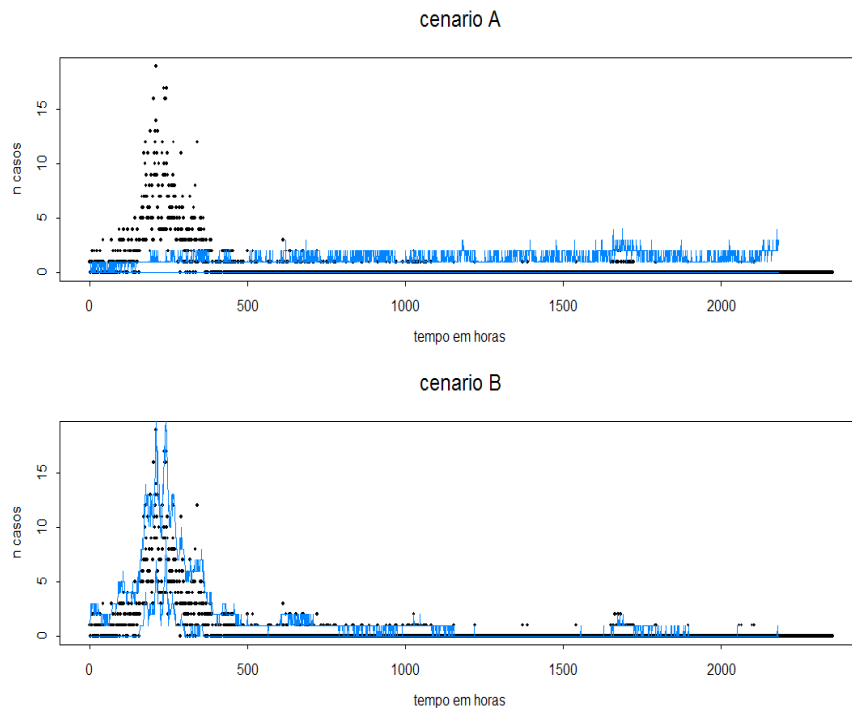


Figura 4.4: Previsão para o número de casos nas 1008 últimas horas geradas sob as abordagens A e B

#### 4.4.2 Modelando $y_1$ - abordagem 2.1

Vamos supor agora que os tempos de observação  $z$  são desconhecidos, mas observa-se o número de casos ocorridos a cada dia  $y_1(j)$ ,  $j=1, \dots, 800$ . O modelo proposto para  $y_1$  é especificado pelas equações (4.14) e (4.15). Note que o parâmetro  $\phi$  não mais corresponde ao  $\phi$  especificado na geração dos dados e na abordagem 1. Portanto vamos chamar esses parâmetro de  $\phi_1$ , com índice 1 indicando que os dados foram agregados em um período de 1 dia.  $\phi_1$  é dado por  $\phi/24$ , visto que os dados agora são agregados a cada 24 horas. O parâmetro  $\sigma^2$  permanece o mesmo.

As prioris são especificadas como na abordagem anterior, com excessão da priori para  $\phi_1$ , que agora é uma uniforme no intervalo  $(0, 1000/24)$ . As estimativas para os parâmetros  $\sigma^2$  e  $\phi_1$ , dadas pela média das suas amostras a posteriori, são  $\hat{\sigma}^2 = 2.45$  e  $\hat{\phi}_1 = 14.83$ , que corresponde a  $\hat{\phi} = 355.92$ . Essas estimativas estão mais longe dos valores reais que as estimativas obtidas pela abordagem 1, o que é uma consequência da perda de informação pela agregação dos dados. Comparando os histogramas das amostras da posteriori obtidos anteriormente com os obtidos nessa análise (para  $\sigma^2$  e  $\phi = 24\phi_1$ ), apresentados na figura 4.5, podemos ver ainda mais claramente a consequência da agregação. Os histogramas agora estão mais espalhados, e as estimativas, consequentemente, menos precisas. Sendo assim, o processo de Cox é melhor estimado sob a abordagem 1.

A figura 4.6 mostra a trajetória do processo  $\gamma$  agregada por dia ( $\gamma_1$ ), e os intervalos de credibilidade de 95% obtidos nessa análise. Como no caso anterior, a grande maioria dos valores reais está contida nos intervalos.



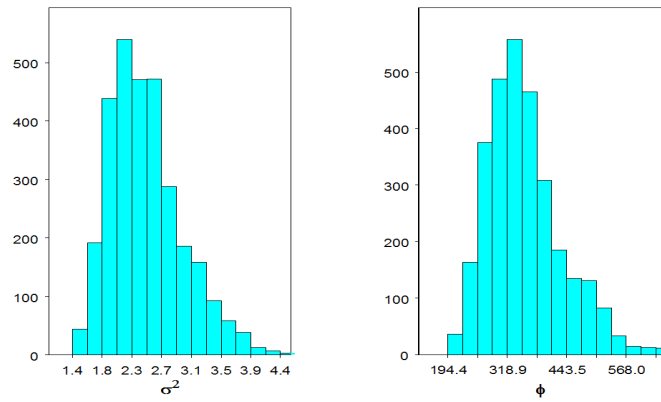
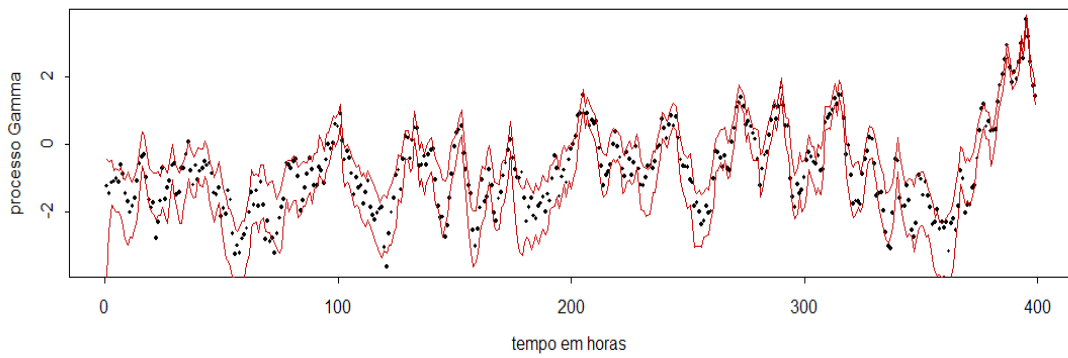


Figura 4.5: Histograma das amostras da posteriori de  $\sigma^2$  e  $\phi$ .

(A)



(B)

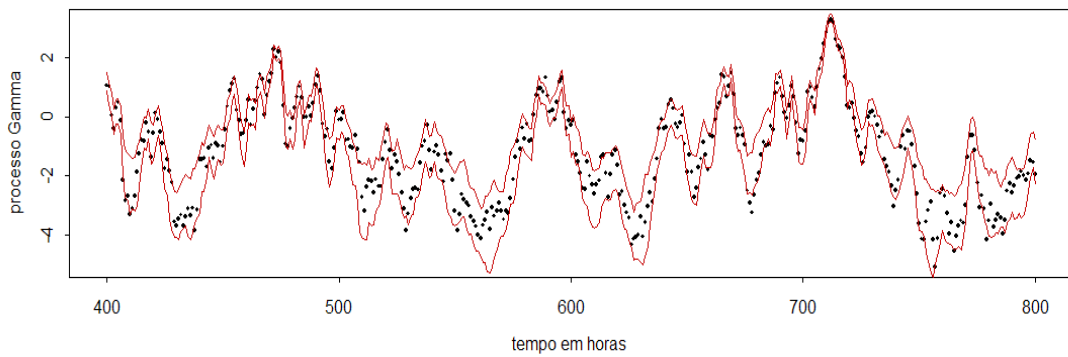


Figura 4.6: Trajetória real de  $\gamma_1$  com intervalos de credibilidade da amostra da posteriori (em vermelho):(A) nos tempos 1 a 400; (B) nos tempos 401 a 800.

A predição para os últimos 91 dias (que correspondem às 2184 últimas horas e 13 últimas semanas) foram feitas e comparadas com as previsões feitas sob a abordagem 1 agregadas por dia. As Figuras 4.7 e 4.8 mostram essa comparação para os dois cenários considerados anteriormente: (A) quando a predição é feita baseada em dados observados no passado, e que não são atualizados no tempo, e (B) quando ela é feita baseada em dados atualizados (a cada evento na abordagem 1, e a cada dia na abordagem 2.1). Os gráficos novamente mostram que previsões sem atualização dos dados observados são muito pobres. O intervalo de confiança obtido com atualizações em tempo contínuo (abordagem 1) se mostra mais apertado em torno dos valores reais quando comparado ao intervalo obtido com atualizações diárias, retratando a maior precisão obtida com a maior quantidade de informação vinda dos dados desagregados.

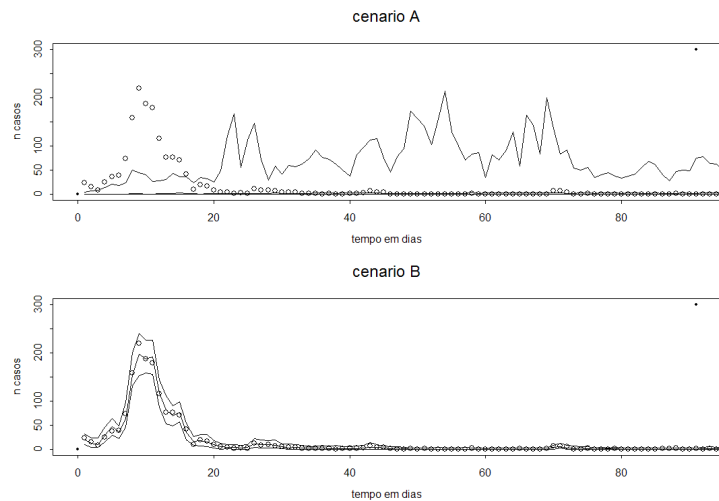


Figura 4.7: Previsão para o número de casos nos 42 últimos dias - abordagem 1.

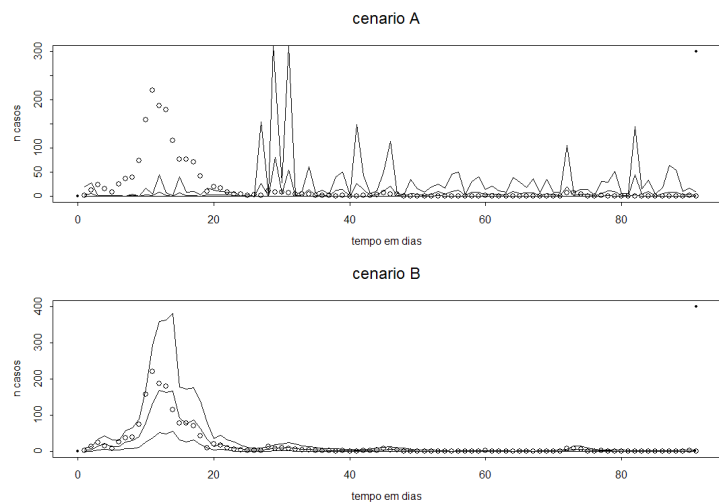


Figura 4.8: Previsão para o número de casos nos 42 últimos dias - abordagem 2.1.

### 4.4.3 Modelando $y_2$ - abordagem 2.2

Nesse caso, as variáveis resposta observadas são o número de casos ocorridos a cada semana  $y_2(j)$ ,  $j=1, \dots, 114$ . O modelo proposto para  $y_2$  é o mesmo especificado na abordagem 2.1, dado pelas equações (4.14) e (4.15). Agora o parâmetro  $\phi$  será chamado de  $\phi_7$ , com índice 7 indicando que os dados foram agregados em um período de 7 dias.  $\phi_7$  é dado por  $\phi/168$ , já que nesse caso os dados são agregados a cada 168 horas.

As priors são especificadas como nas abordagens 1 e 2.1, exceto a priori para  $\phi_7$ , que é especificada como sendo uma uniforme no intervalo  $(0, 1000/168)$ . Amostras das distribuições a posteriori dos parâmetros são obtidas pelo algoritmo descrito na seção 4.3. As médias a posteriori obtidas para os parâmetros  $\sigma^2$  e  $\phi_7$  foram  $\hat{\sigma}^2 = 2.34$  e  $\hat{\phi}_7 = 2.42$ , correspondendo a  $\hat{\phi} = 406.56$ . A estimativa de  $\phi$  ficou mais distante do valor real que a obtida pela abordagem 2.1, mas o mesmo não ocorreu com o parâmetro  $\sigma^2$ . Comparando os histogramas das amostras da posteriori desses parâmetros, dados na figura 4.9, com os histogramas obtidos pela abordagem 2.1, podemos ver que o histograma de  $\sigma^2$  é bastante similar ao da figura 4.5, apresentando uma cauda direita ligeiramente maior. O histograma de  $\phi$  está ainda mais espalhado, e centrado em valores mais distantes do valor real. Esse fato retrata a menor quantidade de informação contida em  $y_2$  em relação a  $y_1$ . Vale lembrar que a agregação de  $z$  para  $y_1$  é de ordem 24, enquanto a agregação de  $y_1$  para  $y_2$  é de ordem 7. Sendo assim, é esperado uma maior diferença na qualidade dos resultados obtidos segundo a abordagem 1 em relação as demais abordagens.

Na figura 4.10 temos a trajetória do processo  $\gamma$  agregada por semana ( $\gamma_7$ ), com os intervalos de credibilidade de 95% obtidos pela amostra a posteriori desses parâmetros. Também nesse caso a maioria dos valores reais está contida nos intervalos.

A intensidade do processo  $\Lambda$  estimado é comparada na Figura 4.11 (no log) com a intensidade  $\Lambda$  estimadas pelas abordagens 1 e 2.1, após serem feitas as devidas agregações. Os pontos no gráfico representam os valores reais. As linhas verde, vermelha e azul são os

resultados obtidos respectivamente pelas abordagens 1, 2.1 e 2.2. O resultado reportado na figura mostra existir muito pouca diferença entre as abordagens quanto a estimação de  $\Lambda$ , sendo as três "linhas" quase indistintas na maior parte da trajetória do processo. Resultados similares foram obtidos comparando o processo estimado para  $\Lambda$  agregado por dia, sob as abordagens 1 e 1.2. Podemos concluir que se a finalidade é a de estimar a média do processo em uma determinada agregação, não é vantagem estimá-lo a níveis mais desagregados.

Finalmente fazemos a comparação da predição do número de casos a cada semana utilizando as abordagens 1, 2.1 e 2.2, e para os cenários A e B. A figura 4.12 mostra essa comparação, e por ela podemos concluir o mesmo que já havia sido observado na comparação entre as abordagens 1 e 2.1 para as previsões a nível diário: as previsões sob o cenário A são extremamente pobres, e as previsões sob o cenário B são melhores sob a abordagem 1. Não há muita diferença, entretanto, nas previsões feitas sob as abordagens 2.1 e 2.2.

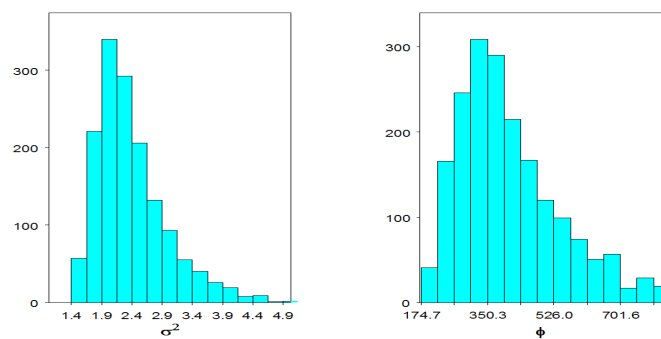


Figura 4.9: Histograma das amostras da posteriori de  $\sigma_1^2$  e  $\phi_1$ .

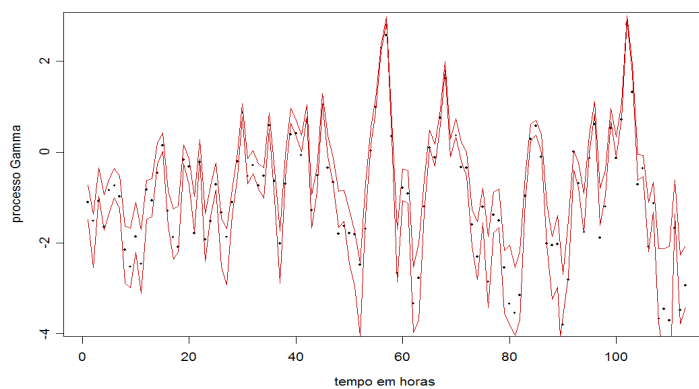


Figura 4.10: Trajetória real de  $\gamma_7$  com intervalos de credibilidade da amostra da posteriori (em vermelho).

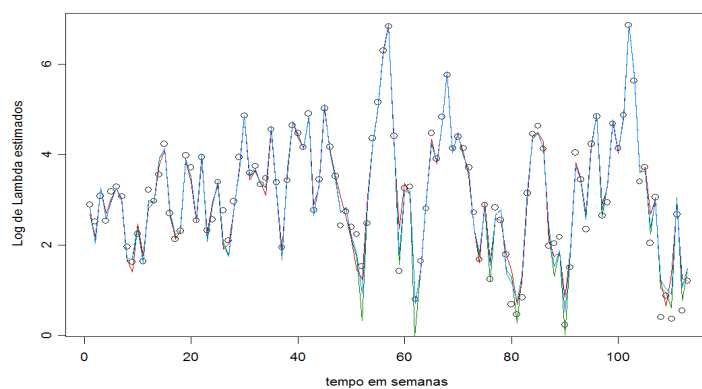


Figura 4.11: comparação das estimativas do processo  $\log(\Lambda)$  agregado por semana, pelas três abordagens.

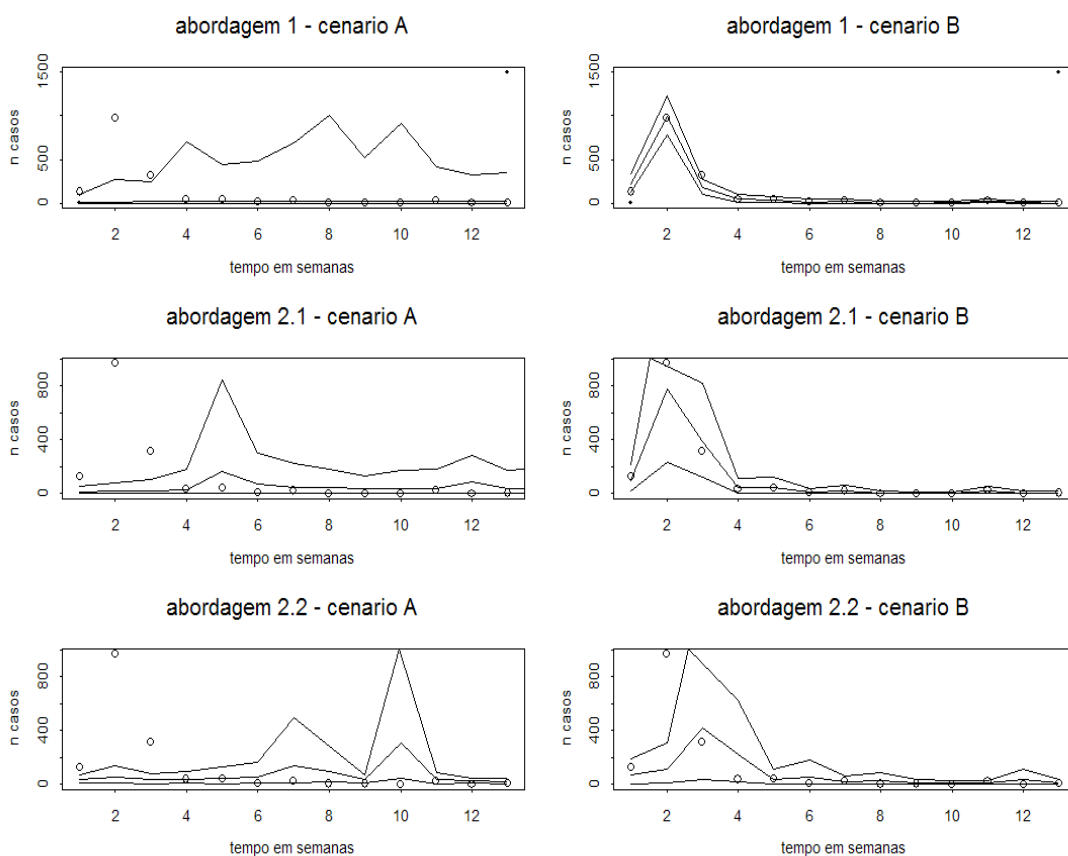


Figura 4.12: Previsão para o número de casos nas 13 últimas semanas usando as abordagens 1, 2.1 e 2.2 e cenários A e B.

## 4.5 Análise de casos de doenças gastro-intestinais

### 4.5.1 Motivação e análises preliminares

O projeto AEGISS (Ascertainment Enhancement of Gastrointestinal Infection Surveillance and Statistics) vem sendo desenvolvido através de uma colaboração entre cinco autoridades de saúde na Grã-Bretanha: o "Wessex Primary Care Research Network" associado a comitês médicos locais; o Departamento de Saúde Ambiental; as Universidade de Lancaster e Southampton; e o serviço de saúde pública. Esse projeto é patrocinado pela Agência de Padrões de Alimentares e pelo Departamento de Saúde.

O objetivo do AEGISS é reduzir a ocorrência de infecções gastro-intestinais na comunidade, através de melhorias na qualidade de informação sobre a ocorrência de casos, assim como proporcionar o fornecimento mais rápido dessas informações para análise. Os médicos do serviço de saúde público são informados sobre casos suspeitos de intoxicação alimentar, de forma que eles podem investigar a fonte e evitar que outras pessoas sejam infectadas. Os agentes de saúde são instruídos para reportar o número de incidência diariamente. A pesquisa de rotina requer um compromisso entre a sensibilidade e a especificidade de detecção, e a qualidade dos dados derivados da rotina diária de atividades clínicas. Infelizmente os métodos atuais dão origem a informação variável e podem levar muito tempo para atingir as autoridades de saúde pública. Dessa forma a investigação é atrasada deixando a fonte de infecção se espalhar. Outra limitação dos dados é o tempo decorrido entre a intoxicação do indivíduo e o momento em que o caso foi efetivamente reportado.

Uma fonte de dados nesse projeto é o NHS Direto, que é um serviço de coleta de dados através de ligações 24 horas. Nesse trabalho analisamos parte dos dados do NHS Direto que cobre o condado de Hampshire e inclui um total de  $T = 6754$  casos reportados de janeiro de 1992 a dezembro de 1993. Análises de casos no tempo e espaço podem ajudar



a identificar mudanças no padrão de ocorrência de infecções gastro-intestinais e a causa potencial de problemas emergentes, melhorando atividades de intervenção e prevenção, e limitando o alastramento de doenças gastro-intestinais a um período mais curto do que é atualmente possível. Nessa pesquisa nos limitamos a modelar a variação temporal dos casos reportados, sendo possível considerar o caso espaço-temporal usando uma extensão do modelo proposto.

A Figura 4.13 mostra o número de casos na região de interesse ao longo do tempo. O maior número de casos foi reportado no dia 29 de dezembro de 1992, com 29 casos. A Figura 4.14 mostra o histograma da contagem diária, mostrando que na maioria das vezes foram reportados de 5 a 10 casos.

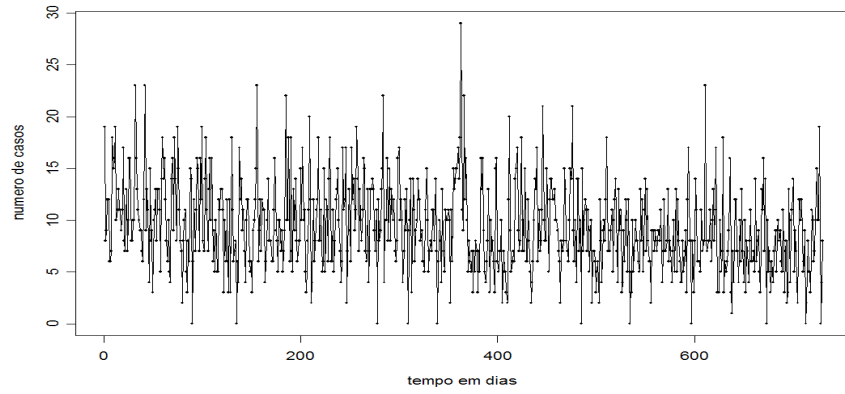


Figura 4.13: Casos reportados ao longo do tempo.

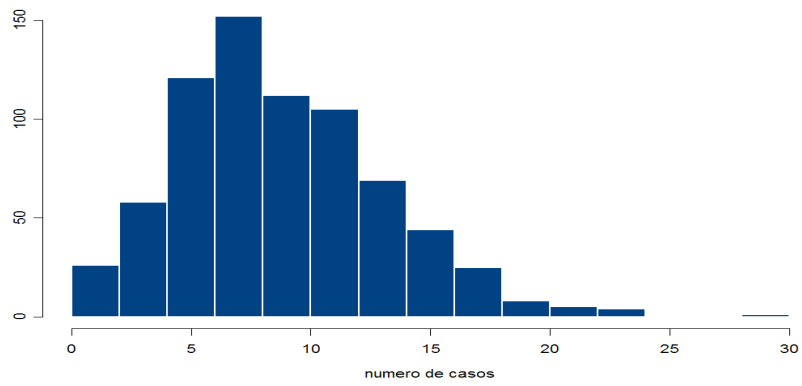


Figura 4.14: Histograma dos casos reportados por dia.

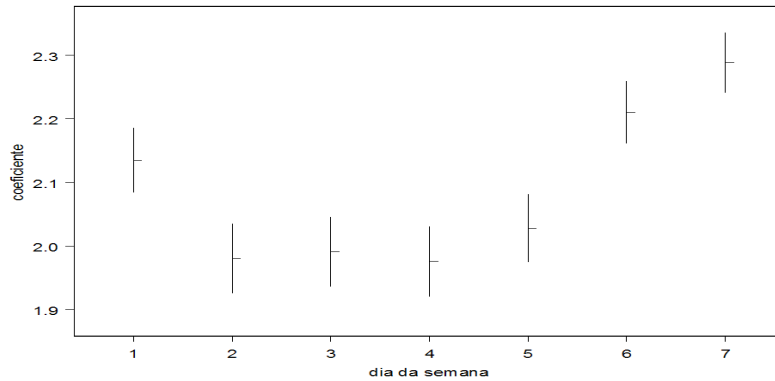


Figura 4.15: Intervalos de 95% de credibilidade dos coeficientes (obtido no ajuste de um MLG) para os indicadores no dia da semana de segunda a domingo.

Análises exploratórias mostram que a incidência de casos aumenta nos fins de semana. Há um aumento no número de casos de sexta-feira para sábado, e o mesmo ocorre de sábado para domingo. A incidência decresce de domingo para segunda-feira, e novamente de segunda para terça-feira, se mantendo estável até sexta-feira. A Figura 4.15 mostra os coeficientes obtidos no ajuste de um MLG de poisson aos dados de contagem ( $y$ ) usando como variáveis explicativas indicadores do dia da semana. Esses indicadores recebem valor 1 se a contagem foi feita no dia da semana em questão, e zero caso contrário.

## 4.5.2 Modelagem e resultados

Seja  $y(t)$  o número de casos observados no dia  $t$ ,  $t = 1, \dots, 730$ . Assumimos que  $y = (y(1), y(2), \dots, y(730))$  segue um processo de Cox, e o modelo para  $y$  é definido pelas equações (4.14) e (4.15). A intensidade populacional  $\rho$  não é conhecida nesse caso, e supomos razoável considerá-la constante no tempo ao considerarmos os indicadores de dia da semana como variáveis explicativas. Assim, definimos a intensidade do processo como sendo  $\Lambda(s) = \exp[\gamma(s) + \beta X(s)]$ ,  $s \in (1, 730)$ , sendo a versão discretizada  $\Lambda_{apx}$  dada por  $\Lambda_{apx}(t) = \exp[\gamma(t) + \beta X(t)]$ ,  $t = 1, \dots, 730$ .  $X(t)$  é definido pela coleção de indicadores do

dia da semana  $X(t) = (SEG(t), TER(t), QUA(t), QUI(t), SEX(t), SAB(t), DOM(t))$ , sendo  $\beta$ , portanto, um vetor de tamanho 7.

Num segundo exercício, trabalhamos com uma versão desagregada de  $y$ , com a geração de tempos de observação  $z = (z(1), z(2), \dots, z(T))$  a partir da suposição de que dados são reportados uniformemente ao longo de cada dia. Essa suposição foi usada para uma simples ilustração do processo contínuo, sendo necessário um estudo mais aprofundado para se obter uma suposição realista.

Com base nos dois conjuntos de variáveis resposta,  $z$  e  $y$ , o modelo proposto foi estimado sob as abordagens 1 e 2, respectivamente, considerando as mesmas distribuições a priori definidas no exercício com dados simulados com frequência diária, ou seja,

$$\sigma^{-2} \sim G(0.1, 0.1) \quad \phi \sim U(0, 1000/24),$$

sendo que nesse caso  $\phi$  é o parâmetro que controla o decaimento da correlação temporal em ambas as abordagens, visto que as variáveis resposta estão na mesma unidade de tempo.

A Figura 4.16 mostra os histogramas das amostras das distribuições a posteriori dos parâmetros  $\sigma$  e  $\phi$ . Podemos ver que, ao contrário do que ocorreu para os dados simulados, não houve grande diferença aqui entre os resultados obtidos pelas diferentes abordagens. Isso retrata o fato de que não há perda de informação quando agregações são feitas em intervalos cuja taxa de incidência de eventos é constante.

A Figura 4.17 apresenta os intervalos de 95% de credibilidade obtidos sob as duas abordagens para os parâmetros  $\beta$ , mostrando que o modelo (nos dois casos) capturou o efeito dos dias da semana observado pela análise exploratória de dados feita na seção 4.5.1. Podemos ver que há um aumento significativo do parâmetro  $\beta$  à medida que nos afastamos do meio da semana.

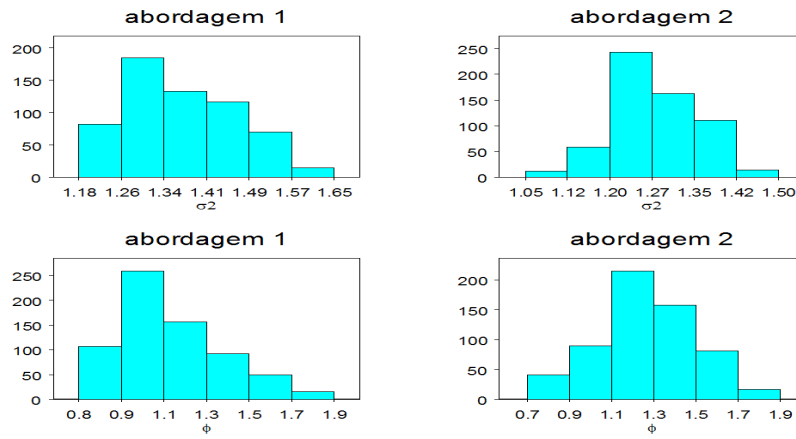


Figura 4.16: Histogramas das amostras da posteriori de  $\sigma^2$  e  $\phi$  sob as abordagens 1 e 2.

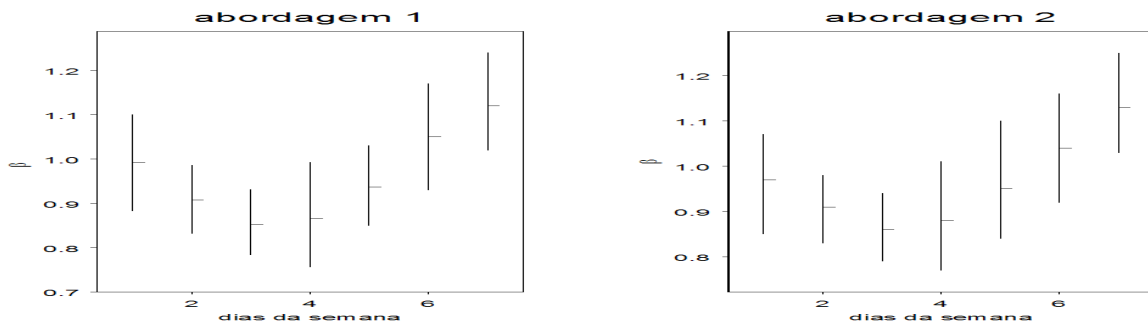


Figura 4.17: Intervalos de 95% de credibilidade dos parâmetros  $\beta$  correspondendo aos indicadores de segunda a domingo - abordagem 1.

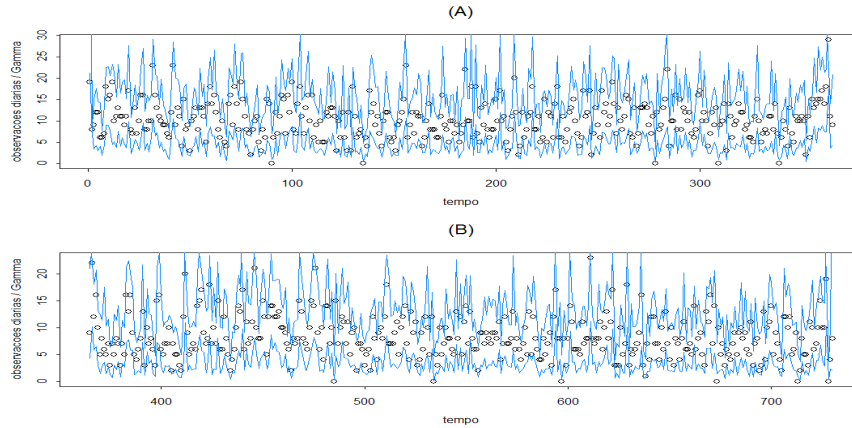


Figura 4.18: Intervalos de 95% de credibilidade da trajetória do processo  $\Lambda$  e número de casos observados por dia: (A) do dia 1 a 365; (B) do dia 366 a 730.

Na Figura 4.18 temos o intervalo de 95% de credibilidade da trajetória do processo  $\Lambda$  (agregado por dia), estimado sob a abordagem 2, sendo comparado com o número de casos observados por dia. Podemos ver que na maioria das vezes a contagem observada está dentro do intervalo, mostrando uma ajuste adequado. O intervalo obtido sob a abordagem 1 é bastante similar ao apresentado na Figura 4.18, e por isso não foi incluído aqui.

Por fim, a previsão para o número de casos para os 10 últimos dias (no modelo estimado excluindo esses valores) é mostrada na Figura 4.19, baseada em dados atualizados (a cada evento na abordagem 1 e a cada dia na abordagem 2). Como a previsão baseada em dados não atualizados se mostrou muito ineficaz na análise de dados simulados, ela não foi considerada nessa aplicação. Podemos ver pelo gráfico que também aqui a previsão feita sob a abordagem 1 é superior à feita sob a abordagem 2, visto que as atualizações são feitas com mais frequência.

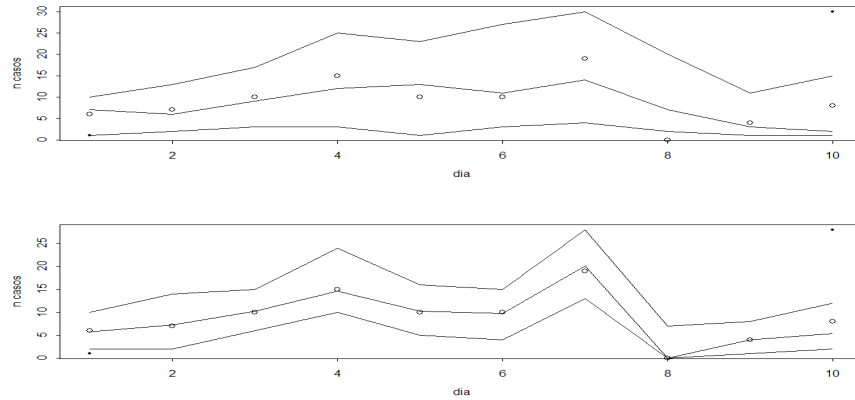


Figura 4.19: Previsão para o número de casos nos 10 últimos dias usando as abordagens 1 e 2

## 4.6 Conclusão

O processo de Cox é bastante utilizado para tratar a variação temporal (e/ou espacial) na incidência de eventos, como por exemplo a incidência de doenças. Para fazer a estimação do processo idealmente teríamos os próprios tempos de observação de eventos como variáveis resposta, de forma a não haver perda de informação por parte das observações. Em dados reais, entretanto, observa-se em geral o número de casos agregados em intervalos de tempo, como o número de ocorrências por dia, semana, mês, ano, e etc. Nesse capítulo analisamos a estimação e previsão do processo de Cox no tempo, sob o ponto de vista Bayesiano, considerando duas abordagens distintas: (1) variáveis resposta contínuas (tempos de observação); e (2) variáveis resposta discretas (observações agregadas em intervalos de tempo). Um dos nossos objetivos com essa análise é de comparar os resultados obtidos sob as duas abordagens, estudando os efeitos da agregação na estimação do modelo e previsão de valores futuros.

Para essa análise foi feito um estudo simulado, onde um processo de Cox contínuo no tempo foi gerado durante um período de 800 dias, dando origem a 6855 casos. A mo-

delagem é feita para três casos, primeiramente considerando os tempos de observação  $z$  conhecidos (abordagem 1), e depois considerando apenas as contagens de eventos por dia ( $y_1$ ), e por semana ( $y_2$ ) conhecidas (abordagens 2.1 e 2.2 respectivamente). Obtivemos melhores estimativas (mais próximas dos valores reais) para os parâmetros  $\sigma^2$  e  $\phi$ , que definem o processo temporal  $\gamma$ , sob a abordagem 1. Os histogramas obtidos das amostras a posteriori desses parâmetros também estão mais concentrados ao redor dos seus respectivos valores reais sob essa abordagem. Podemos concluir que os parâmetros do processo de Cox são melhor estimados quando há mais informação nos dados observados (abordagem 1). As estimativas e histogramas de  $\sigma^2$  e  $\phi$  obtidos sob as abordagens 2.1 e 2.2 estão menos distantes entre si, visto que a agregação feita de  $y_1$  para  $y_2$  é bem menor que a feita de  $z$  para  $y_1$ . Entretanto, comparando estimativas feitas para o processo  $\Lambda$  sob as três abordagens, concluímos que se o interesse é de estimar pontualmente a média do processo para uma determinada agregação temporal, não é vantagem estimá-lo a níveis mais desagregados.

Também com base nas três abordagens consideradas para os dados simulados, foram feitas previsões futuras sob dois cenários distintos: (A) quando a predição é feita baseada em dados observados no passado, e que não são atualizados no tempo, e (B) quando ela é feita baseada em dados atualizados (a cada evento na abordagem 1, a cada dia na abordagem 2.1, e a cada semana na abordagem 2.2). As previsões feitas sem atualização dos dados observados é muito ruim independente da abordagem utilizada, mostrando mais uma vez que a modelagem autoregressiva de primeira ordem para descrever a variação do processo não é adequada para fazer previsões para um futuro pouco próximo. As previsões obtidas com a atualizações dos dados observados foram melhores sob a abordagem 1, devido ao maior número de informação utilizada nas previsões para esse caso.

Finalmente o modelo foi aplicado a dados de infecções gastro-intestinais na Inglaterra, sendo o número de casos da doença registrados diariamente, durante um período de 730 dias. Como uma ilustração, trabalhamos também com uma versão desagregada desse



conjunto de dados, com a geração de tempos de observação a partir da suposição de que dados são reportados uniformemente ao longo de cada dia. Variáveis indicadoras do dia da semana foram utilizadas como variáveis explicativas, visto que análises exploratórias mostraram haver maior incidência de eventos no final de semana.

Os resultados obtidos na aplicação mostram pouca diferença na estimação dos parâmetros do processo, provavelmente devido à desagregação dos dados não ter sido feita de forma realista. Seria necessário um estudo mais aprofundado sobre o processo de coleta dos dados para fazer uma suposição mais adequada sobre a distribuição dos casos ao longo do dia.

## 4.7 Apêndice - aproximação usada na geração de $\gamma$ no algoritmo de MCCM sob a abordagem 1

Seja  $\gamma[t-1, t+1]$  o intervalo fechado dos valores de  $\gamma$  entre os pontos  $z(t-1)$  e  $z(t+1)$ , ou seja,  $\gamma[t-1, t+1] = [\gamma^*(t-1), \gamma^*(t+1)]$ . Vamos definir  $\gamma^{(-t)}[t-1, t+1]$  pelo intervalo igual a  $\gamma[t-1, t+1]$ , excluindo o ponto  $\gamma^*(t)$ , tal que  $\{\gamma^{(-t)}(t-1, t+1), \gamma^*(t)\} = [\gamma^*(t-1), \gamma^*(t+1)]$ . Vamos definir também  $\gamma^*(-t)$  como sendo o conjunto  $[\gamma(0), \gamma(T)]$  menos o ponto  $\gamma^*(t)$ .

$\gamma^*(t)$  só depende de  $\gamma^*(-t)$  e  $z$  através de  $\gamma^{(-t)}[t-1, t+1]$ ,  $z(t-1)$ ,  $z(t)$  e  $z(t+1)$ . Dessa forma, a aceitação do passo de  $\gamma^{*(i-1)}(t)$  para  $\gamma^{*(i)}(t)$  só depende da densidade  $f(\gamma^*(t)|z(t-1), z(t), z(t+1), \gamma^*(-t)[t-1, t+1], \phi^{(i)}, \sigma^{2(i)})$ . Usando o teorema de Bayes e notando que dado  $\gamma^{(-t)}[t-1, t+1]$ ,  $\gamma^*(t)$  não depende de  $z(t-1)$  e  $z(t+1)$ , temos o seguinte resultado:

$$f(\gamma^*(t)|z(t-1), z(t), z(t+1), \gamma^*(-t)[t-1, t+1], \phi, \sigma^2) \propto f(z(t)|z(t-1), z(t+1), \gamma[t-1, t+1], \phi, \sigma^2) f(\gamma^*(t)|\gamma^*(-t)[t-1, t+1], \phi, \sigma^2) \quad (4.19)$$

Como  $\gamma$  é uma função contínua, temos que

$$P(\gamma^{*(t)}) = \lim_{\delta \rightarrow 0} \gamma^*(t + \delta) = \lim_{\delta \rightarrow 0} \gamma^*(t - \delta) = 1.$$

Sendo assim, podemos escrever (4.18) como:

$$f(\gamma^*(t)|z(t-1), z(t), z(t+1), \gamma^*(-t)[t-1, t+1], \phi, \sigma^2) \propto$$

$$f(z(t)|z(t-1), z(t+1), \gamma[t-1, t+1], \phi, \sigma^2),$$

onde

$$f(z(t)|z(t-1), z(t+1), \gamma[t-1, t+1], \phi, \sigma^2) = \left( \int_{z(t-1)}^{z(t)} \Lambda(s) ds \right) \left( \int_{z(t)}^{z(t+1)} \Lambda(s) ds \right) \cdot \exp \left[ -[z(t) - z(t-1)] \left( \int_{z(t-1)}^{z(t)} \Lambda(s) ds \right) \right] \exp \left[ -[z(t+1) - z(t)] \left( \int_{z(t)}^{z(t+1)} \Lambda(s) ds \right) \right]$$

Avaliar essa densidade não é trivial, e por isso é necessário usar uma aproximação para tal. Uma solução seria particionar  $\gamma[t-1, t+1]$  em intervalos menores, e supor que  $\gamma$  é constante dentro de cada um deles. Aqui propomos uma alternativa simplificada porém eficaz, que é considerar uma transição linear entre  $\Lambda(z(t-1))$  e  $\Lambda(z(t))$ . Propomos a seguinte aproximação:

$$f(z(t)|z(t-1), \gamma, \phi, \sigma^2) \approx \frac{-[\Lambda(z(t)) + \Lambda(z(t-1))]}{2} \exp \left[ \frac{-[\Lambda(z(t)) + \Lambda(z(t-1))]}{2} (z(t) - z(t-1)) \right].$$

A distribuição  $f(\gamma^*(t)|z(t-1), z(t), z(t+1), \gamma^*(-t)[t-1, t+1], \phi, \sigma^2)$  é proporcional a  $f(z(t)|z(t-1), z(t+1), \gamma[t-1, t+1], \phi, \sigma^2)$ , e portanto sua densidade avaliada em  $\gamma^*(t)$  pode ser aproximada pela função  $h'(\gamma^*(t)|z(t), z(t-1), z(t+1), \gamma^*(t-1), \gamma^*(t+1), \phi, \sigma^2)$ , dada por:

$$h'(\gamma^*(t)|z(t), z(t-1), z(t+1), \gamma^*(t-1), \gamma^*(t+1), \phi, \sigma^2) = \frac{-[\Lambda(z(t)) + \Lambda(z(t-1))]}{2} \exp \left[ \frac{-[\Lambda(z(t)) + \Lambda(z(t-1))]}{2} (z(t) - z(t-1)) \right] \cdot \frac{-[\Lambda(z(t+1)) + \Lambda(z(t))]}{2} \exp \left[ \frac{-[\Lambda(z(t+1)) + \Lambda(z(t))]}{2} (z(t+1) - z(t)) \right]$$

Essa equação é usada para testar o movimento de um passo para o outro na cadeia. Aceitamos o valor gerado para  $\gamma^{*(i)}(t)$  com probabilidade:

$$\min \left( 1, \frac{h'(\gamma^{*(i)}(t)|z(t), z(t-1), z(t+1), \gamma^{*(i)}(t-1), \gamma^{*(i-1)}(t+1), \phi^{(i)}, \sigma^{2(i)})}{h'(\gamma^{*(i-1)}(t)|z(t), z(t-1), z(t+1), \gamma^{*(i)}(t-1), \gamma^{*(i-1)}(t+1), \phi^{(i)}, \sigma^{2(i)})} \right)$$

# Capítulo 5

## Extensões e conclusões

Esse capítulo está dividido em duas seções. Na seção 5.1 apresentamos possíveis extensões para as modelagens propostas nos capítulos 2, 3 e 4, extensões estas que podem ser usadas como base para o desenvolvimento de pesquisas futuras na área de modelos espaço-temporais. Na seção 5.2 apresentamos as conclusões gerais, resumindo os principais resultados obtido nesse trabalho.

### 5.1 Extensões

Uma série de hipóteses simplificadoras foram consideradas nas análises aqui realizadas. Uma delas é considerar que os processos espaço-temporais contínuos (objetos de estudo nos capítulos 2 e 3), ou uma determinada transformação dos mesmos, são gaussianos. A hipótese de campos aleatórios gaussianos prevalece na geoestatística clássica, e também nas abordagens Bayesianas de predição espacial, como pode ser visto em Brown, Le, e Zidek (1994), Gaudard, Karson, Linder e Sinha (1995), Handcock e Stein (1993), Handcock e Wallis (1994) e Kitanidis (1986). Entretanto, muitos conjuntos de dados ambientais possuem características fortemente não gaussianas, como serem não negativos, e

terem distribuição assimétrica, geralmente com caudas pesadas.

Muitas vezes aplicar transformações como "raiz quadrada" ou "log" não é um método eficaz para aproximar os dados observados a um conjunto de dados com distribuição normal. Uma forma natural de tratar a não-normalidade dos dados é assumir que o campo aleatório de interesse é resultado de uma transformação não linear desconhecida de um campo aleatório gaussiano, como proposto por De Oliveira, Kedem e Short (1997). A família de transformações mais utilizada é a de Box-Cox (Box and Cox, 1964), frequentemente usada para "normalizar" dados positivos:

$$g_{\lambda}(x) = \begin{cases} \frac{x^{\lambda}-1}{\lambda}, & \text{se } \lambda \neq 0 \\ \log(x), & \text{se } \lambda = 0 \end{cases}$$

Outros exemplos de famílias de transformação podem ser vistos em Aranda-Ordaz (1981), Manly (1976) e Sweeting (1985).

A predição é então feita para um campo aleatório gaussiano transformado, com a transformação utilizada pertencendo a uma determinada família paramétrica de transformações. A proposta é de estimar os parâmetros da transformação conjuntamente com os parâmetros do modelo. Em alguns casos, entretando, não é possível levar dados não-normais à normalidade através de transformação. Nesse caso a não normalidade pode ser tratada através da definição de distribuições dentro da família exponencial, assim como foi feito no capítulo 4 dessa tese. Um caso típico para a uso dessa abordagem é o de dados de contagem, que geralmente são especificados como realizações da distribuição de Poisson.

Hipóteses simplificadoras também foram utilizadas na especificação de estruturas de correlação espacial nos capítulos 2 e 3. São elas hipóteses de homoscedasticidade ( $\sigma_i = \sigma, \forall i$ ), estacionariedade e isotropia. Trabalhamos aqui com a família exponencial de potência (De Oliveira et al., 1997), pressupondo uma estrutura de correlação estacionária e isotrópica. Muitas outras possibilidades podem ser exploradas para abordar

a variação espacial dos dados. Outras famílias de funções de correlação que satisfazem essas propriedades são a família esférica (Wackernagel, 1995) e a família Matérn (Matérn, 1986), entre outras.

Provavelmente o método mais estudado para processos espaciais não estacionários e anisotrópicos é a abordagem de deformação de Sampson e Guttorp (1992). Eles introduzem uma abordagem não paramétrica para a estimação de estruturas de covariância espaciais para um processo  $Z(s,t)$  observado repetidamente nos tempos  $t$  ( $t=1,\dots,T$ ) para um número finito de estações de monitoramento  $s$  ( $s=1,\dots,N$ ) no plano. Primeiramente define-se a função de correlação do processo espacial referente a um espaço latente, denotado por  $D$ , onde se aplicam estacionariedade e isotropia. O espaço das estações de monitoramento originais é chamado de  $G$ .

Seguindo essa idéia, Schmidt e O'Hagan (2003) propõem uma abordagem Bayesiana na qual o mapeamento entre os espaços  $G$  e  $D$  é representado por uma função  $d(\cdot)$ . Eles definem um processo gaussiano como priori para  $d(\cdot)$ , e usam o método MCCM para obter amostras das distribuições a posteriori dos parâmetros do modelo.

O modelo proposto no capítulo 3 é uma extensão do modelo proposto no capítulo 2 onde os coeficientes de regressão variam tanto no espaço quanto no tempo. Entretanto, a estrutura de correlação espacial proposta no capítulo 3 para tratar a variação dos parâmetros de regressão nessa dimensão é por hipótese constante, ou seja, todos os parâmetros do modelo têm, supostamente, a mesma estrutura de correlação espacial. Essa suposição em geral não é realista, visto que a natureza da variação espacial do efeito de cada covariável e do intercepto não é a mesma. Da mesma forma, muitas vezes pode não ser razoável supor que a estrutura de correlação seja a mesma para diferentes variáveis resposta. Uma generalização para os modelos propostos no capítulo 3 é a de considerar diferentes estruturas de correlação espacial para cada coeficiente que varia nessa dimensão. Uma limitação dessa extensão é a dificuldade de estimar parâmetros de uma matriz de correlações espacialmente estruturada quando temos pouca quantidade de

informação sobre a variação espacial. No caso da aplicação feita no capítulo 3, dados são observados em apenas 7 estações de monitoramento, sendo portanto difícil, nesse caso, estimar uma estrutura espacial distinta para cada coeficiente do modelo. Uma alternativa para isso é a de explorar similaridades entre as estruturas, supondo, por exemplo, o mesmo decaimento mas variâncias distintas para cada caso. Outra extensão que pode ser feita para a modelagem proposta no capítulo 3 é através da especificação de estruturas de dinâmicas temporais mais elaboradas para explicar a variação dos coeficientes no tempo.

Uma extensão para a modelagem proposta no capítulo 4 pode ser feita para acomodar observações feitas não apenas no tempo, mas também no espaço, levando em consideração a correlação existente nessas duas dimensões. Para tal, a função de intensidade  $\Lambda$  do processo, definida em 4.1, passa a ser variável também no espaço, podendo ser escrita como:

$$\Lambda(t, s) = \rho(t, s)\Pi(t, s), \quad t \in R^+, \quad s \in R^2. \quad (5.1)$$

onde  $\rho(t, s)$  é conhecido, e  $\Pi(t, s) = \exp[\gamma(t, s) + \beta X(t, s)]$ .

O processo  $\gamma(t, s)$  é uma generalização do processo  $\gamma(t)$ , cuja estrutura de correlação é definida na equação 4.2. Uma possível forma de modelar  $\gamma(t, s)$  é através da hipótese de separabilidade nas dimensões espaço/tempo, levando a  $\gamma(t, s) = \gamma_1(t)\gamma_2(s)$ . O processo  $\gamma_1(t)$  pode ser definido por uma estrutura similar a definida para  $\gamma(t)$  na equação (4.2). Uma possível especificação para  $\gamma_2(s)$  é assumir que esse parâmetro é um campo aleatório gaussiano, com função de correlação paramétrica. Dessa forma é fácil fazer interpolação espacial e previsão no tempo para o processo  $\gamma(t, s)$ , e previsão e interpolação para  $\Pi(t, s)$  podem também ser feitos de forma análoga a previsão feita para os modelos propostos no capítulo 4.

Brix e Diggle (2001) propõem um processo espacial de Ornstein-Uhlenbeck, que são



uma classe flexível de processos gaussianos contínuos espaço-temporais, para descrever a variação de  $\gamma(t, s)$  no tempo e espaço. Por essa abordagem a função de covariância incondicional de  $\gamma$  pode ser escrita como

$$\text{cov}\{\gamma(t_1, s_1), \gamma(t_2, s_2)\} = \sigma^2 \exp\{-B(t_2 - t_1)\} r(s_1, s_2; \phi), \quad t_1 < t_2, \quad s_1, s_2 \in J \quad (5.2)$$

onde  $B$  é uma matriz não singular, e  $r$  denota a função de correlação espacial indexada por um ou mais parâmetros  $\phi$ . Quando é feita a suposição de que a função de covariância é separável no espaço/tempo,  $B = \beta$  é um escalar.

Inspirado nas modelagens feitas nos capítulos 2 e 3, também podemos supor que nesse modelo os coeficientes ligados aos processos explicativos variam no tempo/espaço. A proposta é definir  $\Pi(t, s) = \exp[\gamma(t, s) + \beta(t, s)X(t, s)]$ .  $\beta(t, s)$  pode ser modelado como os coeficientes de regressão da modelagem proposta no capítulo 3, por exemplo, tendo estrutura de evolução dinâmica no tempo e correlação espacial definida por uma estrutura paramétrica.

As extensões aqui descritas são apenas uma amostra da infinita gama de possibilidades para estender os modelos propostos e analisados nesse trabalho.

## 5.2 Conclusões

Nesse trabalho foram feitas três diferentes análises de modelos para tratar processos humanos e ambientais, observados ao longo do tempo e ou espaço. O objetivo foi de propor abordagens que levassem em consideração a estrutura de dependência dos dados nessas duas dimensões.

Os modelos propostos são estruturados da forma hierárquica e apresentam parâmetros variando no tempo/espaço. A variação temporal é definida através da especificação de

estruturas dinâmicas no tempo, enquanto a variação espacial é definida através de funções de correlação paramétricas sob a suposição de que os parâmetros são um campo aleatório gaussiano.

Na primeira proposta foi feito um estudo da incorporação de incertezas em um modelo de regressão que permitia que tanto o intercepto quanto o coeficiente angular variassem no espaço, estendendo os modelos propostos em Paez e Gamerman (2003). Nessa pesquisa analisamos dados simulados, onde além da incorporação da variação espacial nos parâmetros de regressão, a incorporação de incertezas no modelo foi analisada através da comparação da interpolação espacial nos casos onde as covariáveis eram (1) conhecidas para todo o espaço de interesse, e (2) conhecidas em apenas um conjunto de pontos. As superfícies interpoladas sob esses dois cenários apresentaram padrões similares, porém com maior precisão por parte da análise feita baseada no conjunto de dados mais informativo (variável explicativa conhecida em todo o espaço). Além disso análises foram feitas com base em diferentes padrões amostrais de observação no espaço, onde comparamos resultados obtidos por amostras ordenadas e aleatórias de tamanho 25 e 81. A vantagem de se trabalhar com amostras de tamanho 81 foi evidente para a maior parte dos casos analisados, mas a vantagem de se trabalhar com dados ordenados foi apenas percebida na interpolação do processo  $X$ , que é, ao contrário dos coeficientes de regressão, replicado no tempo.

Nessa análise também foram comparadas diferentes especificações das distribuições a priori para os parâmetros dos campos aleatórios gaussianos, visto que estes são em geral parâmetros de difícil estimação. Nesse estudo pudemos concluir que há uma melhora substancial na estimação das variâncias do processo quando prioris centradas em valores estimados por análises preliminares são utilizados, quando comparada à estimação considerando prioris não informativas. Não existe vantagem, entretanto, em considerar prioris altamente informativas centradas nessas estimativas preliminares.

A aplicação feita para a primeira modelagem proposta foi a análise de um conjunto de

dados de poluição na Região Metropolitana do Rio de Janeiro. Nessa análise tínhamos como variável resposta a raiz quadrada de níveis de concentração de  $PM_{10}$  observado em diferentes locais no espaço e períodos de tempo, e como variáveis explicativas indicadores do dia da semana e a temperatura máxima diária. O modelo proposto estende o modelo de Paez e Gamerman (2003), permitindo que não só o intercepto varie no espaço, mas também o coeficiente da temperatura. Como resultado obtivemos superfícies interpoladas com um padrão parecido para o intercepto e coeficiente de temperatura, concluindo, como em Paez e Gamerman (2003), que os níveis de  $PM_{10}$  são geralmente mais baixos no sudeste da área de estudo.

A segunda proposta pode ser vista como uma extensão dos modelos da primeira proposta, onde os parâmetros de regressão variam não somente no espaço mas também no tempo. Além disso ela abrange o caso de respostas multivariadas. O modelo proposto aqui é uma restrição do modelo proposto por Landim e Gamerman (2000), onde impomos uma estrutura paramétrica para as matrizes de covariância que expressam a variação espacial presente nos coeficientes.

Um algoritmo de MCCM foi proposto para estimar os parâmetros do modelo e fazer previsões para tempos futuros e interpolação espacial. Para validação desse algoritmo trabalhamos com um conjunto de dados simulados com duas variáveis resposta e uma variável explicativa em 100 locais igualmente espaçados e 70 períodos de tempos. A inferência sobre os parâmetros foi feita com base em 25 locais no espaço e os 60 primeiros períodos de tempo, com objetivo de interpolar o processo nos locais restantes e 10 últimos períodos de tempo, e comparar com os valores reais. Os resultados obtidos na estimação dos parâmetros do modelo foram em geral satisfatórios. Para um período de tempo fixo, as superfícies interpoladas dos coeficientes de regressão não apresentaram um padrão similar ao verdadeiro. Apesar disso o padrão das variáveis resposta interpoladas ficou bem próximo ao padrão real, mostrando que o modelo tem dificuldade em diferenciar a variabilidade espacial proveniente de diferentes parâmetros.

Para a segunda proposta também foi feita uma aplicação a um conjunto de dados de poluição, dessa vez considerando o logaritmo dos níveis de  $NO_2$  e  $SO_2$  como variáveis resposta, e o índice pluviométrico como variável explicativa. Esses dados foram observados na Alemanha em apenas 7 estações de monitoramento e 83 períodos de tempo. O modelo não se mostrou adequado para fazer previsão e interpolação para esse conjunto de dados, provavelmente devido à sazonalidade presente nas variáveis resposta que não foi incorporada na média do modelo, sendo portanto absorvida no componente dinâmico temporal.

A última proposta difere das outras duas por não tratar de respostas normais. Nesse caso é proposto um processo de Cox no tempo para duas diferentes abordagens, a primeira supondo variáveis resposta contínuas formadas por tempos de observação de um determinado evento, e a segunda supondo variáveis resposta discretas formadas por contagens em intervalos de tempo igualmente espaçados. Um dos objetivos dessa proposta é de comparar os resultados obtidos através dessas duas abordagens, analisando os efeitos da perda de informação pela agregação da variável resposta.

Num estudo simulado geramos um processo de Cox contínuo no tempo durante um período de 800 dias, dando origem a 6855 casos. A modelagem é feita considerando três agregações da variável resposta. Primeiramente a variável resposta é dada pelos tempos de observação de eventos; depois consideramos a contagem de casos diários como variável resposta; e por fim a resposta é dada pela contagem de casos a cada semana. A conclusão obtida é de que os parâmetros do processo de Cox são melhor estimados quando a resposta é dada pelos tempos de observação, ou seja, quando informação sobre o processo não foi perdida pela agregação dos dados. Entretanto as estimativas feitas para a média do processo agregado por semana foram praticamente as mesmas sob as três abordagens, mostrando que se o objetivo da aplicação é a de estimar a média do processo para uma dada agregação não há vantagem de trabalhar com observações mais desagregadas. Também com base nessas três abordagens foram feitas previsões para

tempos futuros sob dois cenários distintos baseados em dados passados que (A) não são atualizados no tempo, e (B) são atualizados no tempo. Os resultados baseados no cenário A foram muito pobres mostrando que o modelo não é adequado para fazer previsões para períodos de tempo em um futuro distante. Os resultados obtidos com base no cenário B foram superiores usando a variável resposta mais desagregada, cujas atualizações no tempo são feitas mais frequentemente.

Uma aplicação foi feita a um conjunto de dados de infecções gastro-intestinais observados ao longo de dois anos na Inglaterra. Esses dados são coletados em forma de contagens diárias, não havendo informação sobre o tempo de ocorrência de cada caso. Para um exercício de comparação consideramos também a análise de variáveis resposta dadas por tempos de observação que foram gerados a partir da suposição de os casos são reportados uniformemente a cada dia. Dessa vez os resultados obtidos mostram pouca diferença para a estimação do processo, refletindo o fato de que não há perda de informação na agregação de eventos em períodos de tempo com taxa de incidência constante.

De um modo geral, os modelos propostos foram estimados de maneira satisfatória nos exemplos com dados simulados, e foram razoavelmente bem ajustados aos conjuntos de dados reais, se mostrando bastante úteis para explicar variações espaço-temporais. Entretanto, previsões feita para períodos de tempo em um futuro distante, foram pobres, mostrando a necessidade de atualizar a informação proveniente dos dados para prever tempos futuros. Além disso pudemos comprovar através de análises feitas nos capítulos 2 e 4 a superioridade da estimação e previsão de modelos quando há maior informação no conjunto de dados.

# Referências Bibliográficas

- [1] Aguilar, O. e West, M. (1998). Analysis of hospital quality monitors using hierarchical time series models. *Bayesian Statistics in Science e Technology: Case studies 4*. C. Gatsonis et al, Springer-Verlag.
- [2] Area-Ordaz, F.J. (1981). On Two Families of Transformations to Additivity for Binary Response Data, *Biometrika*, 68, 357-363.
- [3] Assunção, J.J., Gamerman, D. e Assunção, R.M. (1999). Regional differences in factor productivities of Brazilian agriculture: A space varying parameter approach. Technical report, Statistical Laboratory, Universidade Federal do Rio de Janeiro.
- [4] Berger, J.O., De Oliveira, V. e Sansó, B. (2001). Objective Bayesian Analysis of Spatially Correlated Data. *Journal of the American Statistical Association*, **96**, 1361-1374.
- [5] Box, G.E.P. e Cox, D.R. (1964). An analysis of transformations (with discussion). *Journal of the Royal Statistics Society, Series B*, **26**, 211-52.
- [6] Breslow N.E. e Clayton (1993). Approximate inference in generalized linear mixed models, *Journal of the American Statistical Association*, **88**, 9-25.
- [7] Brix, A. e Diggle, P.J. (2001). Spatiotemporal prediction for log-Gaussian Cox processes. *Journal of the Royal Statistical Society, Series B*, **51**, 81-92.

- [8] Brix, A. e Moller, J. (2001). Space-time multitype log Gaussian Cox processes with a view to modelling weed data. *Scandinavian Journal of Statistics*, **28**, 471-488.
- [9] Brown, P.E., Diggle, P.J., Lord, M. E. e Young, P.C. (2001). Space-time calibration of radar-rainfall data. *Applied Statistics*, **50**, 221-242
- [10] Brown, P.J., Le, N.D., e Zidek, J.V.(1994). Multivariate Spatial Interpolation e Exposure to Air Pollutants, *Canadian Journal of Statistics*, **4**, 489-509.
- [11] Cargnoni, C., Muller, P., West, M. (1997). Bayesian forecasting of multinomial time series through conditionally Gaussian dynamic models. *Journal of the American Statistical Association*, **87**, 493-500.
- [12] Christensen, O.F. e Waagepetersen, R. (2001). Bayesian prediction os spatial count data using generalised linear mixed models. *Biometrics*, **58**, 280-286.
- [13] Clayton, D. e Kaldor, J. (1987). Empirical Bayes estimates of age-steardized relative risks for use in disease mapping. *Biometrics*, **43**, 671-681.
- [14] Cressie, N.A.C. (1993). *Statistics for Spatial Data* (rev. ed.), New York: Wiley.
- [15] Dawid, A.P. (1981). Some matrix-variate distribution theory: notational considerations e a Baysian application. *Biometrika*, **68**, 265-274.
- [16] De Oliveira, V., Kedem, B. e Short, D.A. (1997). Bayesian prediction of transformed Gaussian reom fields. *Journal of the American Statistical Association*, **92**, 1422-1433.
- [17] Diggle, P.J. (2003). *Statistical Analysis for Spatial Point Patterns*, Segunda Edição. London: Arnold.
- [18] Eynon, B.P. e Switzer, P. (1983). The variability of rainfall acidity. *Canadian Journal of Statistics*, **11**, 11-24.

- [19] Frühwirth-Schnatter, S. (1995). Bayesian model discrimination e Bayesian factors for linear Gaussian state space models. *Journal of the Royal Statistical Society Series B*, **57**, 237-246.
- [20] Gamerman, D. (1997a). *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*. London: Chapman & Hall.
- [21] Gamerman, D. (1997b). Sampling from the posterior distribution in generalized linear mixed models. *Statistics and Computing*, **7**, 57-68.
- [22] Gamerman, D. e Migon, H.S. (1993). Dynamic hierarchical models. *Journal of the Royal Statistical Society Series B*, **55**, 3, 629-642.
- [23] Gamerman, D., Moreira, A.R.B., e Rue, H. (2003). Space-varying regression models: specifications e simulation. *Computational Statistics Data Analysis*, **42**, 513-533.
- [24] Gaudard, M., Karson, M., Linder, E. e Sinha, D. (1995). Bayesian Spatial Prediction, Technical Report, Whittermore School, University of New Hampshire.
- [25] Gelfand, A.E., Banerjee, S., Gamerman, D. (2003). Spatial Process Modelling for Univariate e Multivariate Dynamic Spatial Data. To appear in *Environmetrics*.
- [26] Gelfand, A.E., Kim, H.-J., Sirmans, C.F. e Banerjee, S. (2003). Spatial modeling with spatially varying coefficient processes. *Journal of the American Statistical Association*, **98**, 387-396.
- [27] Handcock, M.S. e Stein, M.L. (1993). A Bayesian analysis of Kriging, *Technometrics*, **35**, 403-410.
- [28] Handcock, M.S. e Wallis, J.R. (1994). An approach to statistical spatio-temporal modeling of meteorological fields (with discussion). *Journal of the American Statistical Association*, **89**, 368-378.



- [29] Harrison, P.J. e Stevens, C.F. (1976). Bayesian forecasting. *Journal of the Royal Statistical Society*, B, **38**, 205-247 (with discussion).
- [30] Hastings, W.K. (1970). Monte Carlo sampling methods using Markov chains e their applications. *Biometrika*, **57**, 97-109.
- [31] Holle, D.M., De Oliveira, V., Cox, L.H., Smith, R.L. (2000). Estimation of regional trends in sulfur dioxide over the eastern United States. *Environmetrics*, **11**, 373-393.
- [32] Huang, H.C. e Cressie, N.A.C. (1996). Spatiotemporal prediction of snow water equivalent using the Kalman filter. *Computational Statistics e Data Analysis*, **22**, 157-175.
- [33] Huerta, G., Sansó, B. e Stroud, J.R. (2004). A spatio temporal model for Mexico city ozone levels". *Applied Statistics*, **53**, 231-248.
- [34] Kitanidis, P.K. (1986). Parameter uncertainty in estimation of spatial functions: Bayesian analysis. *Water Resources Research*, **22**, 449-507.
- [35] Kyriakidis, P.C. e Journel, A.G. (1999). Geostatistical space-time models. *Mathematical Geology*, **31**, 651-684.
- [36] Landim, F.M.P.F. (1998). Modelos dinâmicos hierárquicos matriz-variados, Tese de doutorado, COPPE, Universidade Federal do Rio de Janeiro.
- [37] Landim, F.M.P.F. e Gamerman, D. (2000). Dynamic hierarchical models - an extension to matrix variate observations. *Computational Statistics e Data Analysis*, **35**, 11-42.
- [38] Le, N.D. e Zidek, J.V. (1992). Interpolation with uncertain spatial covariances: A Bayesian alternative to Kriging. *Journal of Multivariate Analysis*, **43**, 351-374.
- [39] Manly, B.F.J. (1976). Exponential data transformation to Normality, *Biometrika*, **68**, 35-43.

- [40] Manton, K.G., Woodbury, M.A., Stallard, E., et al. (1989). Empirical Bayes procedures for stabilizing maps of U.S. cancer mortality rates. *Journal of the American Statistical Association*, **84**, 637-650.
- [41] Mardia, D., Goodall, C., Redfern, E. e Alonso, F. (1998). The kriged Kalman filter. *Test*, **7**, 217-285.
- [42] Mátern, B. (1986). *Spatial Variation*, Segunda edição. Berlin: Springer-Verlag.
- [43] McCullagh, P., Nelder, J.A. (1989). *Generalized Linear Models*, Segunda Edição. New York: Chapman & Hall.
- [44] Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H. e Teller, E. (1953). Equation of state calculations by fast computing machine. *Journal of chemical physics*, **21**, 1087-91.
- [45] Moller, J., Syversveen, A.R. e Waagepetersen, R. (1998). Log Gaussian Cox process. *Scandinavian Journal of Statistics*, **25**, 451-482.
- [46] Oehlert, G.W. (1993). Regional trends in sulfate wet deposition. *Journal of the American Statistical Association*, **88**, 390-399.
- [47] Paez, M.S. (2000). Estudo dos efeitos espaço-temporais na concentração de poluentes na Região Metropolitana do Rio de Janeiro. Dissertação de mestrado, Universidade Federal do Rio de Janeiro.
- [48] Paez, M.S. e Gamerman, D. (2003). Study of the space-time effects in the concentration of airborne pollutants in the Metropolitan Region of Rio de Janeiro. *Environmetrics*, **14**, 387-408.
- [49] Quintana, J. (1987). Multivariate Bayesian forecasting models, unpublished Ph.D. Thesis, University of Warwick.

- [50] Sampson, P.D. e Guttorp, P.(1992). Nonparametric estimation of nonstationary spatial covariance structure. *Journal of the American Statistical Association*, **87**, 108-119.
- [51] Sansó, B. e Guenni, L. (1999). Venezuelan rainfall data analysed using a Bayesian space-time model. *Applied Statistics*, **48**, 345-362.
- [52] Schmidt, A.M. e O'Hagan, A.O. (2003). Bayesian Inference for Nonstationary Spatial Covariance Structures via Spatial Deformations. *Journal of the Royal Statistical Society Series B*, **65**, 743-758.
- [53] Shaddick, G. e Wakefield, J. (2001). Modeling multivariate pollutant data at multiple sites. *Applied Statistics*, **51**, 351-372.
- [54] Stroud, J.R., Muller, P. e Sansó, B. (2001). Dynamic models for spatiotemporal data. *Journal of the Royal Statistical Society Series B*, **63**, 673-689.
- [55] Sweeting, T.J. (1985). Consistent Prior Distributions for Transformed Models, *Bayesian Statistics 2*, eds. J. M. Bernardo, M. H. DeGroot, D. V. Lindley, e A. F. M. Smith, Valencia University Press, 755-762.
- [56] Tonellato, S.F. (2001). A multivariate time series model for the analysis e prediction of carbon monoxide atmospheric concentrations. *Applied Statistics*, **50**, 187-200.
- [57] Wackernagel, H. (1995). *Multivariate Geostatistics*. Berlin: Springer-Verlag.
- [58] West, M. e Harrison, R.J. (1997). *Bayesian Forecasting e Dinamic Models*. New York: Springer Verlag.
- [59] Wikle, C.K., Berliner, L.M. e Cressie, N. (1998). Hierarchical Bayesian space-time models. *Environmental e Ecological Statistics*, **5**, 117-154.
- [60] Wikle, C.K. e Cressie, N. (1999). A dimension-reduction approach to space-time Kalman filtering. *Biometrika*, **86**, 815-829.

- [61] Williams, D.A. (1982). Extra-binomial variation in logistic linear models. *Applied Statistics*, **31**, 144-148.

# Livros Grátis

( <http://www.livrosgratis.com.br> )

Milhares de Livros para Download:

[Baixar livros de Administração](#)

[Baixar livros de Agronomia](#)

[Baixar livros de Arquitetura](#)

[Baixar livros de Artes](#)

[Baixar livros de Astronomia](#)

[Baixar livros de Biologia Geral](#)

[Baixar livros de Ciência da Computação](#)

[Baixar livros de Ciência da Informação](#)

[Baixar livros de Ciência Política](#)

[Baixar livros de Ciências da Saúde](#)

[Baixar livros de Comunicação](#)

[Baixar livros do Conselho Nacional de Educação - CNE](#)

[Baixar livros de Defesa civil](#)

[Baixar livros de Direito](#)

[Baixar livros de Direitos humanos](#)

[Baixar livros de Economia](#)

[Baixar livros de Economia Doméstica](#)

[Baixar livros de Educação](#)

[Baixar livros de Educação - Trânsito](#)

[Baixar livros de Educação Física](#)

[Baixar livros de Engenharia Aeroespacial](#)

[Baixar livros de Farmácia](#)

[Baixar livros de Filosofia](#)

[Baixar livros de Física](#)

[Baixar livros de Geociências](#)

[Baixar livros de Geografia](#)

[Baixar livros de História](#)

[Baixar livros de Línguas](#)

[Baixar livros de Literatura](#)  
[Baixar livros de Literatura de Cordel](#)  
[Baixar livros de Literatura Infantil](#)  
[Baixar livros de Matemática](#)  
[Baixar livros de Medicina](#)  
[Baixar livros de Medicina Veterinária](#)  
[Baixar livros de Meio Ambiente](#)  
[Baixar livros de Meteorologia](#)  
[Baixar Monografias e TCC](#)  
[Baixar livros Multidisciplinar](#)  
[Baixar livros de Música](#)  
[Baixar livros de Psicologia](#)  
[Baixar livros de Química](#)  
[Baixar livros de Saúde Coletiva](#)  
[Baixar livros de Serviço Social](#)  
[Baixar livros de Sociologia](#)  
[Baixar livros de Teologia](#)  
[Baixar livros de Trabalho](#)  
[Baixar livros de Turismo](#)