



MODELOS ESPAÇO-TEMPORAIS PARA DADOS DE ÁREA NA FAMÍLIA EXPONENCIAL

Juan Carlos Vivar

Tese de Doutorado submetida ao programa de Pós-graduação em Estatística do Instituto de Matemática da Universidade Federal do Rio de Janeiro - UFRJ, como parte dos requisitos necessários à obtenção do título de Doutor em Estatística.

Orientador: **Marco A. R. Ferreira**

Rio de Janeiro

Dezembro, 2007

Livros Grátis

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

Milhares de livros grátis para download.

MODELOS ESPAÇO-TEMPORAIS PARA DADOS DE ÁREA NA FAMÍLIA EXPONENCIAL

Juan Carlos Vivar

Orientador: **Marco A. R. Ferreira**

Tese de Doutorado submetida ao programa de Pós-graduação em Estatística do Instituto de Matemática da Universidade Federal do Rio de Janeiro - UFRJ, como parte dos requisitos necessários à obtenção do título de Doutor em Estatística.

Aprovada por:

Prof. Marco A. R. Ferreira - Presidente
(UFRJ - Univ. of Missouri)

Profa. Alexandra M. Schmidt
(UFRJ)

Profa. Flávia Landim
(UFRJ)

Prof. Dani Gamerman
(UFRJ)

Profa. Marília Sá Carvalho
(FIOCRUZ)

Profa. Mariane Alves
(UERJ)

Rio de Janeiro

Dezembro, 2007

Vivar, Juan Carlos.

Modelos Espaço-Temporais para Dados de Área na Família Exponencial/ Juan Carlos Vivar. - Rio de Janeiro: UFRJ/IM, 2007.

xv, 137f.:il.;31cm.

Orientador: Marco A. R. Ferreira

Tese (doutorado) - UFRJ/IM/Programa de Pós-graduação em Estatística, 2007.

Referências Bibliográficas: f.138-146.

1. Modelos espaço-temporais. 2. Inferência Bayesiana. I. Ferreira, Marco A. R. II. Universidade Federal do Rio de Janeiro, Instituto de Matemática. III. Título.

RESUMO

MODELOS ESPAÇO-TEMPORAIS PARA DADOS DE ÁREA NA FAMÍLIA EXPONENCIAL

Juan Carlos Vivar

Orientador: Marco A. R. Ferreira

Resumo da Tese de Doutorado submetida ao programa de Pós-graduação em Estatística do Instituto de Matemática da Universidade Federal do Rio de Janeiro – UFRJ, como parte dos requisitos necessários à obtenção do título de Doutor em Estatística.

Esta tese teve como objetivo principal propor modelos espaço-temporais para dados de área na família exponencial. Os modelos têm a estrutura dos modelos dinâmicos e a dependência espacial é modelada nas inovações do processo latente através de distribuições de campos aleatórios Markovianos Gaussianos próprios.

Uma vantagem da modelagem espacial com campos aleatórios próprios é permitir dinâmicas espaciais específicas a cada tempo para as inovações. Os campos aleatórios com funções de densidade próprias possibilitam a estimação Bayesiana utilizando métodos Monte Carlo via cadeias de Markov. A metodologia de inferência inclui: um esquema eficiente para obter amostras do processo latente utilizando uma variante do algoritmo *forward filtering backward sampler*, e um método para fazer seleção de modelos através da densidade preditiva.

A aplicação dos modelos espaço-temporais e do procedimento de inferência está desenvolvida para dois conjuntos de dados: o número de homicídios por município no Estado do Espírito Santo nos anos 1979-1998, e o registro de presença ou ausência da *Eurasian Collared-Dove* em regiões dos Estados Unidos durante o período 1986-2003.

Palavras chave: Modelos espaço-temporais; dados de área; modelos lineares generalizados dinâmicos; campos aleatórios Markovianos Gaussianos próprios; seleção de modelos.

Rio de Janeiro

Dezembro, 2007

ABSTRACT

SPATIO-TEMPORAL MODELS FOR AREAL DATA IN THE EXPONENTIAL FAMILY

Juan Carlos Vivar

Orientador: **Marco A. R. Ferreira**

Abstract da Tese de Doutorado submetida ao programa de Pós-graduação em Estatística do Instituto de Matemática da Universidade Federal do Rio de Janeiro – UFRJ, como parte dos requisitos necessários à obtenção do título de Doutor em Estatística.

This thesis had as the main goal to propose spatio-temporal models for areal data in the exponential family. These models have the framework of dynamic models and the spatial dependence is modeled on the latent process innovations through proper Gaussian Markov random fields distributions.

An advantage of spatial modeling with proper random fields is to allow time-specific spatial dynamics for the innovations. Random fields with proper density functions make possible Bayesian estimation using Markov chain Monte Carlo methods. The inference procedure includes: an efficient sampling scheme to obtain samples of the latent process using a version of the *forward filtering backward sampler* algorithm, and a method to perform model selection through predictive density.

Application of spatio-temporal models and inference procedure is developed for two datasets: the number of homicides per county in the State of Espírito Santo in years 1979-1998, and the registries of presence or absence of the *Eurasian Collared-Dove* at regions in the United States during the period 1986-2003.

Keywords: Spatio-temporal models; areal data; dynamic generalized linear models; proper gaussian Markov random fields; model selection.

Rio de Janeiro

Dezembro, 2007

Para Esther e minha família

Agradecimentos

A realização desta tese foi possível devido a vários fatores: a ajuda financeira da CAPES; o valioso apoio do Marco (mesmo de longe!), sempre incentivando; o infinito amor e paciência da Esther, minha inspiração; as experiências vividas com o pessoal do DME, professores e alunos; a colaboração dos professores Oswaldo Cruz e Chris Wikle; o alento, apoio e preocupação incondicional de minha família, que ficou longe em distância mas perto do meu coração; e, é claro, a beleza da cidade do Rio de Janeiro.

Muito obrigado a todos.

Sumário

Lista de Tabelas	xii
Lista de Figuras	xiii
1 Introdução	1
1.1 Objetivos da tese	3
1.2 Organização da tese	4
2 Uma classe de modelos espaço-temporais para dados de área Gaussianos	5
2.1 Introdução	5
2.2 Conceitos preliminares	5
2.3 Modelo espaço-temporal geral	7
2.4 Modelos espaço-temporais particulares	8
2.4.1 Modelo polinomial de primeira ordem	8
2.4.2 Modelo polinomial de segunda ordem	9
2.4.3 Modelo de contaminação	10
2.4.4 Modelo com sazonalidade	10
2.5 Inferência Bayesiana	10
2.5.1 Simulação do vetor de parâmetros	11
2.5.2 Simulação do vetor latente	12
3 Resultados recentes da classe de modelos espaço-temporais Gaussianos	16
3.1 Introdução	16

3.2	Separabilidade	16
3.3	Separabilidade na nossa classe de modelos	17
3.3.1	Modelo polinomial de primeira ordem	17
3.3.2	Modelo polinomial de segunda ordem	18
3.3.3	Modelo de contaminação	19
3.4	Métodos alternativos de estimação do processo latente	20
3.4.1	Filtro de informação	20
3.4.2	Forward information filter backward sampler	21
3.5	Seleção de modelos	22
3.5.1	Fator de Bayes e outros critérios	23
3.5.2	Densidade preditiva	24
4	Modelagem de dados espaço-temporais não Gaussianos utilizando trans- formações	27
4.1	Introdução	27
4.2	Dados de violência	28
4.3	Aproximação Gaussiana	29
4.4	Modelos espaço-temporais propostos	30
4.4.1	Modelo I: Polinomial de primeira ordem	32
4.4.2	Modelo II: Contaminação	32
4.4.3	Modelos III e IV: Polinomiais de segunda ordem	33
4.4.4	Modelo V: Segunda ordem com contaminação na equação da veloci- dade	34
4.4.5	Modelo VI: Segunda ordem com contaminação na equação do nível	34
4.4.6	Modelo VII: Segunda ordem com contaminação nas duas equações de sistema	35
4.4.7	Modelo VIII: Segunda ordem com $\phi_2 = 0$ e velocidade igual para cada município	35
4.5	Ajuste e seleção dos modelos propostos	36

4.6	Resultados para Rio de Janeiro	36
4.6.1	Utilizando todos os tempos ($T = 20$)	37
4.6.2	Seleção de modelos	39
4.7	Resultados para Espírito Santo	48
4.7.1	Utilizando todos os tempos ($T = 20$)	49
4.7.2	Seleção de modelos	51
5	Generalização da classe de modelos para dados de área na família expo-	
	nencial	60
5.1	Introdução	60
5.2	Conceitos preliminares	61
5.2.1	Modelos lineares generalizados	61
5.2.2	Família exponencial	63
5.3	Métodos Bayesianos para modelos lineares generalizados	64
5.4	Modelos lineares generalizados dinâmicos	66
5.4.1	Filtragem e suavização generalizadas	67
5.4.2	Filtro de Kalman estendido	68
5.5	Modelos lineares generalizados dinâmicos para dados de área	69
5.5.1	Inferência Bayesiana	70
5.6	Seleção de modelos via densidade preditiva	73
6	Aplicação da classe de modelos para dados de área: Caso Poisson	76
6.1	Introdução	76
6.2	Violência no Estado do Espírito Santo	76
6.3	Modelo geral	77
6.4	Modelos espaço-temporais propostos	78
6.5	Inferência Bayesiana	79
6.5.1	Especificação das prioris	80
6.5.2	Estimação do processo latente	81
6.6	Resultados da seleção de modelos	82

7	Aplicação da classe de modelos para dados de área: Caso Bernoulli	90
7.1	Introdução	90
7.2	Dispersão da <i>Eurasian Collared-Dove</i> nos EUA	90
7.3	Modelo geral	93
7.4	Modelos espaço-temporais propostos	95
7.5	Dados artificiais	98
7.6	Procedimento de inferência para os dados reais	103
7.6.1	Estimação do processo latente	104
7.6.2	Especificação das priors dos parâmetros	106
7.7	Resultados da seleção de modelos	112
8	Considerações finais	116
A	Rio de Janeiro	119
A.1	Mapa político de 1979	119
A.2	Taxas de mortalidade por homicídio por município	122
A.3	Taxas de mortalidade por homicídio por ano	126
B	Espírito Santo	129
B.1	Mapa político de 1979	129
B.2	Taxa de mortalidade por homicídio por município	132
B.3	Taxas de mortalidade por homicídio por ano	136
	Referências Bibliográficas	138

Lista de Tabelas

4.1	Número de municípios dos Estados do Rio de Janeiro e Espírito Santo.	28
4.2	Especificação de S , T e h para Rio de Janeiro e Espírito Santo.	31
4.3	RJ - Com dependência espacial: Resumos a posteriori e $T = 20$	38
4.4	RJ - Sem dependência espacial: Resumos a posteriori e $T = 20$	40
4.5	RJ - Seleção de modelos: Logaritmo da densidade preditiva.	41
4.6	RJ - Resumos a posteriori: Modelo II com dependência espacial.	41
4.7	ES - Com dependência espacial: Resumos a posteriori e $T = 20$	50
4.8	ES - Sem dependência espacial: Resumos a posteriori e $T = 20$	51
4.9	ES - Seleção de modelos: Logaritmo da densidade preditiva.	52
4.10	ES - Resumos a posteriori: Modelo II com dependência espacial.	52
5.1	Algumas distribuições que pertencem à família exponencial.	63
6.1	Logaritmo da densidade preditiva para os modelos considerados.	82
6.2	ES - Resumos a posteriori: Modelo II.	83
7.1	Resumo da agregação dos dados.	95
7.2	<i>ECD</i> - Exemplo simulado: Resumos a posteriori dos parâmetros.	104
7.3	<i>ECD</i> - Logaritmo da densidade preditiva para os modelos ajustados.	113

Lista de Figuras

4.1	Número agregado de homicídios por ano de estudo: (a) Rio de Janeiro, e (b) Espírito Santo.	28
4.2	Matriz de vizinhança dos municípios de Espírito Santo antes e depois da reordenação.	32
4.3	RJ - Logaritmo da densidade preditiva dos modelos com e sem dependência espacial.	42
4.4	RJ - Modelo II: Logaritmo da densidade preditiva com e sem dependência espacial.	43
4.5	RJ - Modelo II: Histogramas das amostras a posteriori dos parâmetros.	43
4.6	RJ - Modelo II: Processo latente estimado (1979-1989).	44
4.7	RJ - Modelo II: Processo latente estimado (1990-1998).	45
4.8	RJ - Modelo II: Médias a posteriori das inovações (1979-1989).	46
4.9	RJ - Modelo II: Médias a posteriori das inovações (1990-1998).	47
4.10	ES - Logaritmo da densidade preditiva dos modelos com e sem dependência espacial.	53
4.11	ES - Modelo II: Logaritmo da densidade preditiva com e sem dependência espacial.	54
4.12	ES - Modelo II: Histogramas das amostras a posteriori dos parâmetros.	54
4.13	ES - Modelo II: Processo latente estimado (1979-1989).	56
4.14	ES - Modelo II: Processo latente estimado (1990-1998).	57
4.15	ES - Modelo II: Médias a posteriori das inovações (1979-1989).	58
4.16	ES - Modelo II: Médias a posteriori das inovações (1990-1998).	59

6.1	ES - Dados na família exponencial: Logaritmo da densidade preditiva.	83
6.2	ES - Modelo II: Histogramas das amostras a posteriori dos parâmetros.	84
6.3	ES - Modelo II: Taxa de risco por 100 mil habitantes (1979-1989).	86
6.4	ES - Modelo II: Taxa de risco por 100 mil habitantes (1990-1998).	87
6.5	ES - Modelo II: Médias a posteriori das inovações (1979-1989).	88
6.6	ES - Modelo II: Médias a posteriori das inovações (1990-1998).	89
7.1	População total da <i>ECD</i> nos Estados Unidos no período 1986-2003.	91
7.2	Locais onde os dados foram coletados e grade utilizada para agregar os locais.	92
7.3	Presença e ausência da <i>ECD</i> ao longo do tempo.	96
7.4	<i>ECD</i> - Dados simulados: Probabilidades simuladas e estimadas (1986-1990).	99
7.5	<i>ECD</i> - Dados simulados: Probabilidades simuladas e estimadas (1991-1995).	100
7.6	<i>ECD</i> - Dados simulados: Probabilidades simuladas e estimadas (1996-2000).	101
7.7	<i>ECD</i> - Dados simulados: Probabilidades simuladas e estimadas (2001-2003).	102
7.8	<i>ECD</i> - Dados simulados: Histogramas a posteriori dos parâmetros.	104
7.9	<i>ECD</i> - Logaritmo da densidade preditiva dos modelos.	113
7.10	<i>ECD</i> - Modelo I(a): Probabilidade estimada (1986-2000).	114
7.11	<i>ECD</i> - Modelo I(a): Probabilidade estimada (2001-2003).	115
A.1	Mapa político do Estado do Rio de Janeiro (1979).	120
A.2	RJ - Taxas de mortalidade por homicídio - I.	122
A.3	RJ - Taxas de mortalidade por homicídio - II.	123
A.4	RJ - Taxas de mortalidade por homicídio - III.	124
A.5	RJ - Taxas de mortalidade por homicídio - IV.	125
A.6	RJ - Taxas de mortalidade por homicídio - V.	126
A.7	RJ - Mapas das taxas de mortalidade por homicídio. Anos 1979 - 1982.	126
A.8	RJ - Mapas das taxas de mortalidade por homicídio. Anos 1983 - 1990.	127
A.9	RJ - Mapas das taxas de mortalidade por homicídio. Anos 1991 - 1998.	128
B.1	Mapa político do Espírito Santo (1979).	130

B.2	ES - Taxas de mortalidade por homicídio - I.	132
B.3	ES - Taxas de mortalidade por homicídio - II.	133
B.4	ES - Taxas de mortalidade por homicídio - III.	134
B.5	ES - Taxas de mortalidade por homicídio - IV.	135
B.6	ES - Mapas das taxas de mortalidade por homicídio. Anos 1979 - 1990. . . .	136
B.7	ES - Mapas das taxas de mortalidade por homicídio. Anos 1991 - 1998. . . .	137

Capítulo 1

Introdução

A última década teve um incremento no número de pesquisas sobre modelagem espaço-temporal. Isto ocorreu devido à importância do entendimento de fenômenos espaço-temporais como soluções de questões científicas em muitas disciplinas, tais como climatologia, economia e epidemiologia, entre muitas outras. O poder crescente dos computadores tem tornado possível a análise de grandes quantidades de dados espaço-temporais através de modelos cada vez mais realistas. Muitos destes modelos têm sido desenvolvidos para processos contínuos no espaço, conhecidos também como processos geostatísticos (por exemplo, [Handcock e Wallis, 1994](#); [Guttorp et al., 1994](#); [Wikle e Cressie, 1999](#); [Gelfand et al., 2005](#); [Xu et al., 2005](#)). Não obstante, existem conjuntos de dados de área ou por regiões de um domínio geográfico de interesse, como por exemplo totais ou médias de alguma variável de interesse para cada estado de um país, os quais devem ser analisados com outras classes de modelos realistas. Vários modelos para dados de área têm sido desenvolvidos na literatura de mapeamento de doenças (por exemplo, [Bernardinelli et al., 1995](#); [Waller et al., 1997](#); [Knorr-Held e Besag, 1998](#); [Knorr-Held, 2000](#); [Knorr-Held e Richardson, 2003](#); [Schmid e Held, 2004](#)). Uma característica comum destes modelos é a utilização de campos aleatórios Markovianos impróprios como prioris para os efeitos espaciais aleatórios. Mesmo que estes modelos sejam bons para suavização, eles não permitem fazer previsões conjuntas para as áreas no tempo futuro, isto impossibilita o uso de seleção de modelos baseados em previsão.

Nosso principal objetivo na presente tese é propor modelos espaço-temporais para dados

de área na família exponencial estendendo o trabalho de [Vivar \(2004\)](#) no qual introduzimos uma classe de modelos espaço-temporais para dados de área Gaussianos. Estes modelos estão baseados na estrutura dos modelos lineares dinâmicos multivariados ([West e Harrison, 1997](#)). A nossa contribuição está em permitir que as inovações das equações de observação e do processo latente, no caso Gaussiano, e do processo latente, quando os dados são não Gaussianos, tenham dependência espacial seguindo processos de campos aleatórios Markovianos Gaussianos próprios (*CAMGP*). Dizemos que uma variável \mathbf{X} segue um processo de *CAMGP*, $\mathbf{X} \sim \text{CAMGP}(\boldsymbol{\mu}_X, \mathbf{P})$ se

$$p(\mathbf{X}) \propto \exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_X)' \mathbf{P}(\mathbf{x} - \boldsymbol{\mu}_X)\right) \quad (1.1)$$

com vetor de médias $\boldsymbol{\mu}_X$ e matriz de precisão \mathbf{P} . A definição de \mathbf{P} é $\mathbf{P} = \tau(\mathbf{I}_S + \phi \mathbf{M})$ na qual \mathbf{M} é a matriz de vizinhança definida na região de estudo, \mathbf{I}_S é a matriz identidade de ordem S e τ e ϕ são os parâmetros que medem a precisão e a correlação espacial, respectivamente.

Como a nossa classe de modelos herda a flexibilidade da estrutura dos modelos dinâmicos, permite incluir características importantes como tendências polinomiais, sazonalidade, transporte e dispersão. O uso de inovações com densidades de *CAMGP* tem conseqüências tanto na modelagem como na computação. Com respeito à modelagem, permite dinâmicas espaciais específicas a cada tempo para os erros de evolução (e de observação no caso Gaussiano). Computacionalmente, como as inovações com densidades de *CAMGP* têm funções de densidade próprias podemos usar um procedimento de estimação Bayesiana baseada em métodos Monte Carlo via cadeias de Markov (*MCMC*, abreviação do inglês *Markov chain Monte Carlo*), incluindo um esquema eficiente para simular do vetor de estados mediante uma variante do algoritmo *FFBS* (abreviação do inglês *forward filtering backward sampler*) e também um método para fazer seleção de modelos através da densidade preditiva das observações $p(\mathbf{y}_t | \mathbf{y}_{t-1}, \mathbf{y}_{t-2}, \dots, \mathbf{y}_1)$.

Muitos modelos espaço-temporais baseados em modelos lineares dinâmicos têm sido propostos na literatura para aplicações com dados geoestatísticos, como por exemplo meteorologia ([Ghil et al., 1981](#)), análise de ozônio ([Guttorp et al., 1994](#)), previsão da equivalência entre água e neve ([Huang e Cressie, 1996](#)), calibração de radares para dados de precipitação

(Brown et al., 2001), e estatística ambiental (Wikle et al., 1998; Huerta et al., 2004). Estes modelos assumem inovações que são espacialmente independentes ou cuja dependência espacial é modelada da forma usual dos modelos geoestatísticos, como por exemplo, utilizando a classe Matérn (Matérn, 1960). Enquanto estes modelos são úteis para dados espaço-temporais geoestatísticos, não são adequados para analisar dados de área. Em contraste, nossa classe de modelos permite uma análise significativa das dinâmicas espaço-temporais de dados regionais ou de área na família exponencial.

Além do uso de campos aleatórios Markovianos impróprios, vários modelos espaço-temporais para dados de área assumem um nível comum variando no tempo para todo o campo (por exemplo, Waller et al., 1997; Knorr-Held e Besag, 1998; Knorr-Held e Richardson, 2003). Enquanto isto pode ser adequado para algumas aplicações, a suposição do nível comum variando no tempo induz uma função de covariância espaço-temporal que decresce a uma função positiva da distância temporal, quando a distância espacial cresce para o infinito (Brown et al., 2001). Isto é indesejável em muitas aplicações, nas quais diferentes partes do campo têm diferentes tendências temporais. Algumas alternativas propostas por Bernardinelli et al. (1995), Sun et al. (2000) e Assunção et al. (2001) permitem tendências temporais lineares e parabólicas determinísticas com coeficientes variando no espaço, as quais são adequadas para conjuntos de dados com poucos pontos no tempo e inadequadas para predições temporais. Knorr-Held (2000) propôs um modelo espaço-temporal no qual cada região tem um nível que evolui no tempo com um passeio aleatório, com as inovações para as regiões em um determinado tempo seguindo um campo aleatório Markoviano Gaussiano (*CAMG*); como detalhamos no Capítulo 2, este modelo é um caso limite particular dentro da nossa classe de modelos espaço-temporais.

1.1 Objetivos da tese

Como estabelecemos anteriormente, o nosso principal objetivo é (1) propor uma classe de modelos espaço-temporais para dados de área na família exponencial. Outros objetivos gerais são: (2) propor um esquema eficiente de amostragem do processo latente utilizando

uma variante do algoritmo *FFBS*, e (3) propor um procedimento para fazer seleção de modelos baseada na densidade preditiva. Os objetivos específicos são: (i) desenvolver uma metodologia de estimação via métodos *MCMC* para diferentes versões do nosso modelo geral, (ii) utilizar estas diferentes versões do modelo geral em aplicações com conjuntos de dados reais, e (iii) selecionar o melhor modelo em cada aplicação através da densidade preditiva.

1.2 Organização da tese

Esta tese está organizada da seguinte forma: no capítulo seguinte revisamos o trabalho desenvolvido em [Vivar \(2004\)](#) para análise de dados Gaussianos. No Capítulo 3 discutimos alguns resultados sobre separabilidade da estrutura de covariância da nossa classe de modelos; apresentamos também o filtro de informação como uma alternativa ao filtro de Kalman no caso Gaussiano e a distribuição preditiva como uma técnica de seleção de modelos. Resultados para dados na família exponencial via aproximação pela normal são apresentados no Capítulo 4. No Capítulo 5 apresentamos os modelos espaço-temporais para observações na família exponencial. Esta é uma extensão necessária dos modelos lineares dinâmicos quando temos que a distribuição das observações não pode ser assumida como Gaussiana, por exemplo, se temos dados binários ou de contagem. A aplicação dos modelos espaço-temporais e do procedimento de estimação está desenvolvida para dois conjuntos de dados, a saber: o número de homicídios por município no Estado do Espírito Santo entre os anos 1979-1998, no Capítulo 6, e no Capítulo 7, o registro de presença ou ausência da *Eurasian Collared-Dove* em regiões dos Estados Unidos durante o período 1986-2003. Finalmente, no Capítulo 8 estão as conclusões e possíveis extensões desta tese. Os mapas políticos e os nomes dos municípios dos Estados de Rio de Janeiro e Espírito Santo, considerados nas aplicações, estão nos Apêndices A e B, assim como também os gráficos das taxas de mortalidade por homicídio por município e os mapas das mesmas taxas por cada ano de estudo.

Capítulo 2

Uma classe de modelos espaço-temporais para dados de área Gaussianos

2.1 Introdução

Apresentamos neste capítulo um resumo do trabalho realizado durante a dissertação de Mestrado. Nosso objetivo consistiu em desenvolver uma nova classe de modelos espaço-temporais para dados de área Gaussianos. Estes modelos estão baseados nos modelos lineares dinâmicos Bayesianos, cujos erros têm distribuições normais. A nossa contribuição foi introduzir dependência espacial nos erros das equações de observação e de sistema, modelando-os como campos aleatórios Markovianos Gaussianos próprios (*CAMGP*). As seguintes Seções descrevem conceitos preliminares, o modelo espaço-temporal geral e suas principais características, alguns modelos particulares e os procedimentos utilizados para fazer inferência Bayesiana dos parâmetros e do vetor latente.

2.2 Conceitos preliminares

O primeiro conceito trata sobre os modelos lineares dinâmicos (*MLD*), também conhecidos como modelos de espaço de estados, que foram introduzidos por [Harrison e Stevens \(1976\)](#) dentro de uma estrutura Bayesiana. Um amplo estudo desta classe de modelos com numerosas aplicações na análise de séries temporais encontra-se em [West e Harrison \(1997\)](#).

Os *MLD* são modelos com parâmetros variando no tempo e são representados na sua versão multivariada por

$$\mathbf{y}_t = \mathbf{F}'_t \boldsymbol{\beta}_t + \boldsymbol{\epsilon}_t, \quad \boldsymbol{\epsilon}_t \sim N(\mathbf{0}, \mathbf{V}_t), \quad (2.1)$$

$$\boldsymbol{\beta}_t - \boldsymbol{\mu}_\beta = \mathbf{G}_t(\boldsymbol{\beta}_{t-1} - \boldsymbol{\mu}_\beta) + \boldsymbol{\omega}_t, \quad \boldsymbol{\omega}_t \sim N(\mathbf{0}, \mathbf{W}_t), \quad (2.2)$$

chamadas de equação de observação e equação de sistema, respectivamente. Para um estudo mais profundo sobre *MLD* veja [Migon et al. \(2005\)](#). Detalhes de cada componente do modelo são apresentados na seguinte seção.

Outra componente necessária para estabelecer nosso modelo espaço-temporal é a definição de campo aleatório Markoviano Gaussiano (*CAMG*). Esta é uma ampla classe de campos aleatórios Markovianos, conhecida principalmente por pesquisadores da área de estatística espacial ([Cressie, 1993](#); [Besag e Kooperberg, 1995](#); [Wikle et al., 1998](#); [Cressie e Huang, 1999](#); [Banerjee et al., 2004](#)), mas com muitas aplicações além desta, por exemplo, análise de imagens, análise de séries temporais, análise de dados longitudinais e de sobrevivência, estatística espaço-temporal, modelos gráficos e estatística semiparamétrica. Ver [Rue e Held \(2005\)](#) e as referências aí contidas para mais detalhes. Extensões do uso dos *CAMG* podem ser encontrados em [Rue e Tjelmeland \(1999\)](#) e [Dethlefsen \(2003\)](#).

Os *CAMG* também recebem o nome de modelos condicionais auto-regressivos (CAR). O estudo desta classe de modelos foi introduzido nos artigos pioneiros de Besag ([Besag, 1974, 1975](#)) de modelos para interações espaciais, definidos em grades bidimensionais para dados distribuídos no espaço.

Definimos um *CAMG* em uma grade bidimensional se, para cada região k existe um conjunto $N_k = \{l : l \text{ é vizinho de } k\}$ e se a distribuição condicional de cada x_k , o valor do campo na posição k , dados todos os outros valores x 's, somente depende dos valores x 's contidos em N_k . Uma forma de definir N_k é incluir as regiões que têm fronteiras comuns com a região k (vizinhança de primeira ordem). Isto pode ser estendido para incluir todas as áreas de segunda geração que dividem fronteira com os vizinhos originais de primeira ordem (vizinhança de segunda ordem). A estrutura de correlação de um *CAMG* está definida por uma matriz de precisão que é singular por definição, o que produz que a distribuição

conjunta dos x 's seja imprópria. No nosso modelo, esta matriz é definida positiva, isto é, tem inversa e, portanto, temos um *CAMG* próprio. Na próxima seção define-se a forma de esta matriz.

2.3 Modelo espaço-temporal geral

Consideramos um conjunto de regiões indexadas pelos inteiros $1, 2, \dots, S$, formando uma grade dentro de um domínio geográfico de interesse. Assumimos que esta grade está associada a um sistema de vizinhança $\{N_s; s = 1, \dots, S\}$, onde N_s denota o conjunto de regiões que são vizinhas da região s . Para cada tempo t e região s observamos a variável de interesse y_{ts} , $t = 1, \dots, T$, $s = 1, \dots, S$. Denote o campo observado no tempo t como $\mathbf{y}_t = (y_{t1}, \dots, y_{tS})'$. Usando a estrutura dos modelos lineares dinâmicos multivariados (West e Harrison, 1997), Vivar (2004) propôs modelos espaço-temporais com distribuições de campos aleatórios Markovianos Gaussianos próprios (*CAMGP*) nos erros, da seguinte forma

$$\mathbf{y}_t = \mathbf{F}'_t \boldsymbol{\beta}_t + \boldsymbol{\epsilon}_t, \quad \boldsymbol{\epsilon}_t \sim \text{CAMGP}(\mathbf{0}_S, \mathbf{V}_t^{-1}), \quad (2.3)$$

$$\boldsymbol{\beta}_t - \boldsymbol{\mu}_\beta = \mathbf{G}_t(\boldsymbol{\beta}_{t-1} - \boldsymbol{\mu}_\beta) + \boldsymbol{\omega}_t, \quad \boldsymbol{\omega}_t \sim \text{CAMGP}(\mathbf{0}_S, \mathbf{W}_t^{-1}), \quad (2.4)$$

onde $\mathbf{0}_S$ é o vetor nulo S -dimensional e os erros $\boldsymbol{\epsilon}_1, \dots, \boldsymbol{\epsilon}_T, \boldsymbol{\omega}_1, \dots, \boldsymbol{\omega}_T$ são independentes.

As Equações (2.3) e (2.4) são conhecidas como equações de observação e de sistema ou evolução, respectivamente. Nestas equações, $\boldsymbol{\beta}_t$ é o processo espaço-temporal latente, \mathbf{F}_t conecta o processo latente com as observações, \mathbf{G}_t descreve a evolução espaço-temporal do processo, $\boldsymbol{\epsilon}_t$ é o campo do erro de observação, $\boldsymbol{\omega}_t$ é o campo da inovação do estado, e as matrizes \mathbf{V}_t e \mathbf{W}_t descrevem a estrutura de covariância espacial de $\boldsymbol{\epsilon}_t$ e $\boldsymbol{\omega}_t$, respectivamente. Além disso, $\boldsymbol{\mu}_\beta$ é a média do processo latente $\boldsymbol{\beta}_t$ e somente está definida se o processo é estacionário no tempo. Seguindo a notação de Ferreira e De Oliveira (2007), $\mathbf{X} \sim \text{CAMGP}(\boldsymbol{\mu}_X, \mathbf{P})$ significa que a variável \mathbf{X} segue um processo de *CAMGP* com vetor de médias $\boldsymbol{\mu}_X$ e matriz de precisão \mathbf{P} , isto é, a função de densidade de \mathbf{X} é proporcional a $\exp\left(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_X)' \mathbf{P}(\mathbf{x} - \boldsymbol{\mu}_X)\right)$, com $\mathbf{P} = \tau(\mathbf{I}_S + \phi \mathbf{M})$ e \mathbf{M} a matriz de vizinhança com

elementos

$$(\mathbf{M})_{k,l} = \begin{cases} r_k, & k = l \\ -g_{kl}, & k \in N_l \\ 0, & \text{caso contrário.} \end{cases} \quad (2.5)$$

N_l representa o conjunto de vizinhos da região l , $g_{kl} > 0$ é uma medida de similaridade entre regiões k e l , $r_k = \sum_{l \in N_k} g_{kl}$, τ é um parâmetro de escala, \mathbf{I}_S é a matriz identidade $S \times S$ e $\phi \geq 0$ controla o grau de correlação espacial. A matriz \mathbf{P} é dominada pela diagonal, isto é, em cada fila (coluna) de \mathbf{P} , o valor na diagonal é maior que a soma dos valores absolutos fora da diagonal, e por conseqüência, \mathbf{P} é positiva definida (Harville, 1997).

A dependência espacial aumenta com ϕ : quando $\phi = 0$ as regiões são independentes, e quando $\phi \rightarrow \infty$ o *CAMGP* aproxima-se do modelo auto-regressivo intrínseco (Besag et al., 1991; Besag e Kooperberg, 1995).

2.4 Modelos espaço-temporais particulares

As Equações (2.3) e (2.4) definem uma classe de modelos espaço-temporais para dados de área muito flexível, na qual a característica chave para uma aplicação com sucesso destes modelos é a especificação das matrizes \mathbf{F}_t , \mathbf{G}_t , \mathbf{V}_t e \mathbf{W}_t . Tipicamente, estas matrizes dependem de uns poucos parâmetros que não variam no tempo e que denotamos pelo vetor de parâmetros ψ .

Nas seguintes Subsecções assume-se que as variâncias nas Equações (2.3) e (2.4) são constantes no tempo, isto é, $\mathbf{V}_t = \mathbf{V}$ e $\mathbf{W}_t = \mathbf{W}$, $\forall t$.

2.4.1 Modelo polinomial de primeira ordem

O modelo polinomial de primeira ordem (West e Harrison, 1997) é o modelo espaço-temporal mais simples com $\mathbf{F}'_t = \mathbf{I}_S$ e $\mathbf{G}_t = \rho \mathbf{I}_S$. Neste modelo, β_t é o nível do processo no tempo t . Se ρ está no intervalo $(-1, 1)$, o modelo é estacionário, se $\rho = \pm 1$, o modelo é não estacionário.

Um caso particular deste modelo encontra-se em [Knorr-Held \(2000\)](#). A evolução temporal para o processo latente no seu modelo IV é um modelo polinomial de primeira ordem com $\rho = 1$, $\mathbf{W}^{-1} = \tau(\mathbf{I}_S + \phi\mathbf{M})$, $\phi \rightarrow \infty$, $\tau\phi = \lambda < \infty$, $g_{kl} = 1, k \in N_l$, e m_k igual ao número de vizinhos da região k . Para verificar isto, compare a média da distribuição condicional completa de $\beta_{ts}, t = 2, \dots, T - 1$ sob o modelo de primeira ordem, dado por

$$\frac{1}{2}(\beta_{t-1,s} + \beta_{t+1,s}) + \frac{1}{2m_s} \sum_{k \in N_s} (\beta_{tk} - \beta_{t-1,k}) + \frac{1}{2m_s} \sum_{k \in N_s} (\beta_{tk} - \beta_{t+1,k}), \quad (2.6)$$

com a expressão dada na Seção 2.2.4 de [Knorr-Held \(2000\)](#). Esta expressão mostra que modelos polinomiais de primeira ordem podem ser utilizados para suavizar os dados espaço-temporais observados, de forma análoga à redução de ruído no caso da utilização de *CAMGP* na análise de imagens.

2.4.2 Modelo polinomial de segunda ordem

No modelo polinomial de segunda ordem existem dois campos latentes: um deles representa o nível e o outro representa a velocidade do processo de mudança do nível. Além disso,

$$\boldsymbol{\beta}_t = \begin{pmatrix} \beta_{1t} \\ \beta_{2t} \end{pmatrix}, \mathbf{F}'_t = (\mathbf{I}_S, \mathbf{0}_S), \mathbf{G}_t = \begin{pmatrix} \rho_1 \mathbf{I}_S & \rho_1 \mathbf{I}_S \\ \mathbf{0}_S & \rho_2 \mathbf{I}_S \end{pmatrix} \text{ e } \mathbf{W}^{-1} = \begin{pmatrix} \mathbf{W}_1^{-1} & \mathbf{0}_S \\ \mathbf{0}_S & \mathbf{W}_2^{-1} \end{pmatrix}.$$

Para garantir a estacionariedade, o autovalor mais alto de \mathbf{G}_t tem que ser menor que um em valor absoluto ([West e Harrison, 1997](#)). Como os dois autovalores são iguais a ρ_1 e ρ_2 , então $|\rho_1| < 1$ e $|\rho_2| < 1$ garantem a estacionariedade. Neste caso, os parâmetros ρ_1 e ρ_2 nas equações de nível e velocidade, de forma similar ao parâmetro ρ no modelo polinomial de primeira ordem, controlam a taxa de convergência do processo ao nível histórico.

O modelo polinomial de segunda ordem pode ser útil quando, para cada região e para cada tempo, a tendência temporal do processo latente pode ser aproximada por uma equação linear local, a tendência linear muda com o tempo, e as inovações das equações de nível e de velocidade são correlacionadas espacialmente.

2.4.3 Modelo de contaminação

Em muitos processos espaço-temporais, um incremento anormal em uma determinada região em um dado tempo pode se estender às regiões vizinhas nos tempos subseqüentes (por exemplo, processos epidêmicos). Este comportamento pode ser capturado com um modelo de contaminação espaço-temporal, que assume $\mathbf{F}_t = \mathbf{I}_S$ e

$$\mathbf{G}_t = \frac{\rho}{(1 + \kappa h)} \mathbf{H}, \text{ onde } (\mathbf{H})_{kl} = \begin{cases} 1, & k = l, \\ \kappa, & k \in N_l, \\ 0, & \text{caso contrário,} \end{cases} \quad (2.7)$$

e N_l é o conjunto de vizinhos ao longo do tempo da região l , $\kappa > 0$ pode ser considerado como um índice de contaminação, $h = \max_l \sum_{k=1, k \neq l}^S \mathbf{1}(k \in N_l)$ é o número máximo de vizinhos ao longo do tempo de todas as regiões, $\rho \in (0, 1)$ é o parâmetro que mede a persistência temporal. A estrutura da matriz \mathbf{H} é similar à da matriz \mathbf{M} definida na Equação (2.5) e incorpora as interações espaço-temporais entre regiões vizinhas.

2.4.4 Modelo com sazonalidade

Neste modelo é incluído um termo de sazonalidade na equação de observação e a sazonalidade é modelada como em [Harvey \(1989\)](#). O modelo fica da seguinte forma:

$$\begin{aligned} \mathbf{y}_t &= \boldsymbol{\beta}_t + \mathbf{s}_t + \boldsymbol{\epsilon}_t, & \boldsymbol{\epsilon}_t &\sim \text{CAMGP}(\mathbf{0}_S, \mathbf{V}_t^{-1}) \\ \boldsymbol{\beta}_t &= \rho \boldsymbol{\beta}_{t-1} + \boldsymbol{\omega}_{1t}, & \boldsymbol{\omega}_{1t} &\sim \text{CAMGP}(\mathbf{0}_S, \mathbf{W}_{1t}^{-1}) \\ \mathbf{s}_t &= -(\mathbf{s}_{t-1} + \dots + \mathbf{s}_{t-i+1}) + \boldsymbol{\omega}_{2t}, & \boldsymbol{\omega}_{2t} &\sim \text{CAMGP}(\mathbf{0}_S, \mathbf{W}_{2t}^{-1}) \end{aligned} \quad (2.8)$$

onde i é a periodicidade dos dados. Por exemplo, $i = 4$ se os dados são trimestrais ou $i = 12$ se os dados são mensais. A sazonalidade evolui restrita à soma das componentes sazonais de i tempos passados mais um erro que segue um *CAMGP*.

2.5 Inferência Bayesiana

Quando as matrizes \mathbf{F}_t , \mathbf{G}_t , \mathbf{V}_t e \mathbf{W}_t do modelo são completamente conhecidas, o filtro de Kalman pode ser usado para estimar o processo latente $\boldsymbol{\beta}_t$ ([West e Harrison, 1997](#)). Na

prática, estas matrizes dependem de um vetor de parâmetros ψ .

Um dos principais aspectos do nosso modelo espaço-temporal é que os erros das equações de observação e de sistema seguem processos de *CAMGP*, os quais permitem incluir dependência espacial nos erros das equações. Ainda mais importante, como os erros modelados com *CAMGP* têm funções de densidade próprias, nossa classe de modelos pode ser analisada com um esquema eficiente de Monte Carlo via cadeias de Markov (*MCMC*, do inglês, *Markov chain Monte Carlo*). A cadeia de Markov pode ser dividida em dois blocos: simulação do vetor de parâmetros ψ e simulação do vetor latente β_1, \dots, β_T pelo algoritmo *FFBS* (abreviação do inglês de *forward filtering backward sampler*) (Frühwirth-Schnatter, 1994; Carter e Kohn, 1994), para obter amostras dos vetores latentes $\beta_{(1:T)} = (\beta_1, \dots, \beta_T)$ da distribuição a posteriori normal multivariada completa $p(\beta_{(1:T)} | \mathbf{D}_T)$, onde $\mathbf{D}_T = (\mathbf{y}_1, \dots, \mathbf{y}_T)$ é o conjunto completo das observações.

2.5.1 Simulação do vetor de parâmetros

Os métodos *MCMC* são aplicáveis a modelos gerais altamente estruturados e complexos (Green et al., 2003) tais como nossos modelos espaço-temporais para dados de área Gaussianos.

É muito importante construir cadeias de Markov com boas propriedades tais como convergência rápida e baixa autocorrelação entre realizações. Com tal objetivo, a cadeia de Markov deve ser específica para cada modelo estudado e depender da forma como $\mathbf{F}_t, \mathbf{G}_t, \mathbf{V}_t$ e \mathbf{W}_t dependem do vetor de parâmetros ψ .

Para fazer inferência a posteriori do vetor de parâmetros ψ encontra-se a distribuição condicional completa de cada um dos componentes de ψ partindo da distribuição conjunta a posteriori

$$p(\beta_{(1:T)}, \psi | \mathbf{D}_T) \propto \left[\prod_{t=1}^T p(\mathbf{y}_t | \beta_t, \psi) \right] \left[\prod_{t=1}^T p(\beta_t | \beta_{t-1}, \psi) \right] p(\psi). \quad (2.9)$$

Os métodos *MCMC* utilizados para obter amostras destes parâmetros são o amostrador de Gibbs e o algoritmo de Metropolis-Hastings. Para informação mais detalhada destes métodos *MCMC* veja Robert e Casella (1999) e Gamerman e Lopes (2006). Para um estudo

mais detalhado sobre as condicionais completas dos parâmetros do modelo espaço-temporal, veja [Vivar \(2004\)](#).

2.5.2 Simulação do vetor latente

O filtro de Kalman

O termo filtro de Kalman, que primeiramente apareceu em [Kalman \(1960\)](#), refere-se ao procedimento recursivo para fazer inferência em modelos de espaço de estados aplicados a séries temporais, quando as matrizes \mathbf{F}_t , \mathbf{G}_t , \mathbf{V}_t e \mathbf{W}_t são conhecidas. Quando isto ocorre, dados os valores de $\mathbf{D}_{t-1} = (\mathbf{y}_1, \dots, \mathbf{y}_{t-1})$, podemos prever \mathbf{y}_t e estimar os vetores latentes não observáveis $\beta_1, \dots, \beta_{t-1}$ através das distribuições preditiva $(\mathbf{y}_t | \mathbf{D}_{t-1})$ e posteriori $(\beta_{t-1} | \mathbf{D}_{t-1})$, respectivamente.

Para uma introdução à metodologia do filtro de Kalman, veja [Meinhold e Singpurwalla \(1983\)](#). [Harrison e Stevens \(1976\)](#), os primeiros a se interessarem na previsão Bayesiana, utilizam resultados bem conhecidos de estatística multivariada para desenvolver o filtro de Kalman.

A derivação das equações do filtro de Kalman está baseada em indução no tempo t : supõe-se que o modelo é fechado a informações externas, ou seja, $\mathbf{D}_t = \{\mathbf{y}_t, \mathbf{D}_{t-1}\}$ e supõe-se também que a priori inicial em $t = 0$ é normal multivariada

$$(\beta_0 | \mathbf{D}_0) \sim N(\mathbf{m}_0, \mathbf{C}_0),$$

para algum vetor de médias \mathbf{m}_0 e matriz de covariâncias \mathbf{C}_0 conhecidos.

Com estas suposições, as equações de atualização, para cada t , são

- Posteriori em $t - 1$:

Para alguma média \mathbf{m}_{t-1} e matriz de covariâncias \mathbf{C}_{t-1} ,

$$(\beta_{t-1} | \mathbf{D}_{t-1}) \sim N(\mathbf{m}_{t-1}, \mathbf{C}_{t-1}).$$

- Priori em t :

$$(\beta_t | \mathbf{D}_{t-1}) \sim N(\mathbf{a}_t, \mathbf{R}_t),$$

com

$$\mathbf{a}_t = \mathbf{G}_t \mathbf{m}_{t-1} \quad \text{e} \quad \mathbf{R}_t = \mathbf{G}_t \mathbf{C}_{t-1} \mathbf{G}_t' + \mathbf{W}_t.$$

- Previsão um passo a frente:

$$(\mathbf{y}_t | \mathbf{D}_{t-1}) \sim N(\mathbf{f}_t, \mathbf{Q}_t)$$

com

$$\mathbf{f}_t = \mathbf{F}_t' \mathbf{a}_t \quad \text{e} \quad \mathbf{Q}_t = \mathbf{F}_t' \mathbf{R}_t \mathbf{F}_t + \mathbf{V}_t.$$

- Posteriores em t :

$$(\boldsymbol{\beta}_t | \mathbf{D}_t) \sim N(\mathbf{m}_t, \mathbf{C}_t)$$

com

$$\begin{aligned} \mathbf{m}_t &= \mathbf{a}_t + \mathbf{A}_t \mathbf{e}_t, & \mathbf{C}_t &= \mathbf{R}_t - \mathbf{A}_t \mathbf{Q}_t \mathbf{A}_t', \\ \mathbf{A}_t &= \mathbf{R}_t \mathbf{F}_t \mathbf{Q}_t^{-1} & \text{e} \quad \mathbf{e}_t &= \mathbf{y}_t - \mathbf{f}_t. \end{aligned}$$

Guardamos os valores de \mathbf{m}_t e \mathbf{C}_t a cada tempo para sua posterior utilização no algoritmo *forward filtering backward sampler*.

FFBS

A idéia básica do algoritmo *FFBS*, proposta independentemente por [Frühwirth-Schnatter \(1994\)](#) e [Carter e Kohn \(1994\)](#) é descrita a seguir.

Dado o modelo dinâmico $\{\mathbf{F}_t, \mathbf{G}_t, \mathbf{V}_t, \mathbf{W}_t\}$, desejamos amostrar o conjunto completo de vetores de estado $\boldsymbol{\beta}_{(1:T)} = (\boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_T)$ da posteriori normal multivariada completa $p(\boldsymbol{\beta}_{(1:T)} | \mathbf{D}_T)$. Explorando a estrutura Markoviana da equação de sistema do modelo linear dinâmico, podemos escrever

$$\begin{aligned} p(\boldsymbol{\beta}_{(1:T)} | \mathbf{D}_T) &= p(\boldsymbol{\beta}_T | \mathbf{D}_T) \prod_{t=1}^{T-1} p(\boldsymbol{\beta}_t | \boldsymbol{\beta}_{t+1}, \mathbf{D}_t) \\ &= p(\boldsymbol{\beta}_T | \mathbf{D}_T) p(\boldsymbol{\beta}_{T-1} | \boldsymbol{\beta}_T, \mathbf{D}_{T-1}) \dots p(\boldsymbol{\beta}_1 | \boldsymbol{\beta}_2, \mathbf{D}_1) p(\boldsymbol{\beta}_0 | \boldsymbol{\beta}_1, \mathbf{D}_0) \end{aligned} \quad (2.10)$$

Podemos amostrar o vetor $\boldsymbol{\beta}_{(1:T)}$ simulando seqüencialmente os vetores de estado seguindo estes passos:

1. Utilizar o filtro de Kalman para encontrar a média e variância das distribuições $p(\boldsymbol{\beta}_1|\mathbf{D}_1), \dots, p(\boldsymbol{\beta}_T|\mathbf{D}_T)$.
2. Gerar um valor de $p(\boldsymbol{\beta}_T|\mathbf{D}_T)$.
3. Calcular média e variância de $p(\boldsymbol{\beta}_{T-1}|\boldsymbol{\beta}_T, \mathbf{D}_T)$. Gerar $\boldsymbol{\beta}_{T-1}$ desta distribuição.
4. Calcular recursivamente

$$p(\boldsymbol{\beta}_{T-i}|\boldsymbol{\beta}_{T-i+1}, \dots, \boldsymbol{\beta}_T, \mathbf{D}_T) = p(\boldsymbol{\beta}_{T-i}|\boldsymbol{\beta}_{T-i+1}, \mathbf{D}_T).$$

Gerar $\boldsymbol{\beta}_{T-i}$, $i = 1, \dots, T - 1$, desta distribuição.

Dessa forma, cada $\boldsymbol{\beta}_t$ gerado está condicionado ao conjunto completo das observações \mathbf{D}_T . Este método de simulação é mais eficiente que simular estado por estado (Carlin et al., 1992). Em nosso caso a distribuição condicional de $\boldsymbol{\beta}_{t-1}|\boldsymbol{\beta}_t, \mathbf{D}_t$ tem forma de uma normal multivariada com

$$\text{Var}(\boldsymbol{\beta}_{t-1}|\boldsymbol{\beta}_t, \mathbf{D}_t) = (\mathbf{G}'_t \mathbf{W}_t^{-1} \mathbf{G}_t + \mathbf{C}_{t-1}^{-1})^{-1} \quad (2.11)$$

$$E(\boldsymbol{\beta}_{t-1}|\boldsymbol{\beta}_t, \mathbf{D}_t) = \text{Var}(\boldsymbol{\beta}_{t-1}|\boldsymbol{\beta}_t, \mathbf{D}_t)(\mathbf{G}'_t \mathbf{W}_t^{-1} \boldsymbol{\beta}_t + \mathbf{C}_{t-1}^{-1} \mathbf{m}_{t-1}). \quad (2.12)$$

Como precisamos de amostras $\boldsymbol{\beta}_t^*$ de uma normal multivariada $N(\mathbf{b}_t, \mathbf{B}_t)$, aplicamos o seguinte algoritmo

1. Calcular a fatorização de Cholesky, $\mathbf{B}_t = \tilde{\mathbf{L}}\tilde{\mathbf{L}}'$
2. Amostrar $\mathbf{z}_t \sim N(\mathbf{0}_S, \mathbf{I}_S)$
3. Calcular $\mathbf{v}_t = \tilde{\mathbf{L}}\mathbf{z}_t$
4. Calcular $\boldsymbol{\beta}_t^* = \mathbf{b}_t + \mathbf{v}_t$
5. Retornar $\boldsymbol{\beta}_t^*$

Uma análise de dados simulados mostrou boa performance do método de estimação dos parâmetros e do processo latente. Uma ilustração da metodologia desenvolvida para um conjunto de dados reais foi a aplicação aos dados de velocidade do vento no Oceano Pacífico

tropical, analisado também em [Wikle e Cressie \(1999\)](#). Os resultados para os conjuntos de dados simulados e o conjunto de dados reais encontram-se em [Vivar \(2004\)](#).

Neste capítulo fizemos uma revisão da classe de modelos espaço-temporais proposta na dissertação de Mestrado e o procedimento de inferência dos parâmetros e do processo latente. Nos capítulos seguintes apresentaremos os avanços obtidos para esta classe de modelos e a extensão para dados não Gaussianos.

Capítulo 3

Resultados recentes da classe de modelos espaço-temporais Gaussianos

3.1 Introdução

Neste capítulo apresentamos novos tópicos desenvolvidos no estudo da nossa classe de modelos espaço-temporais para dados de área Gaussianos. Iniciamos com a definição de separabilidade na seguinte seção. Na Seção 3.3 discutimos a separabilidade da estrutura de covariância da nossa classe de modelos espaço-temporais. Introduzimos o filtro de informação, como uma técnica alternativa ao filtro de Kalman, que permite acelerar certos cálculos na simulação do vetor latente através de uma variante do algoritmo *FFBS*, na Seção 3.4. Finalmente, uma forma de seleção de modelos utilizando a distribuição preditiva é apresentada em detalhe na Seção 3.5.

3.2 Separabilidade

Suponha que um certo processo aleatório $Z(t; \mathbf{s})$, indexado no tempo e no espaço, tem uma covariância espaço-temporal estacionária $C(i; \mathbf{k})$, onde $i \in \mathbb{R}$ é uma defasagem temporal e $\mathbf{k} \in \mathbb{R}^d$ ($d \geq 1$) é uma defasagem espacial, ou seja

$$\text{Cov}(Z(t; \mathbf{s}), Z(t + i; \mathbf{s} + \mathbf{k};)) = C(i; \mathbf{k} | \boldsymbol{\psi}), \quad (3.1)$$

em que a função C é positiva definida e $\boldsymbol{\psi} \in \Psi \subset \mathbb{R}^q$ é o vetor de parâmetros do modelo.

As covariâncias espaço-temporais separáveis têm a propriedade de poderem ser escritas como o produto de uma covariância puramente espacial e uma covariância puramente temporal, ou seja

$$C(i; \mathbf{k} | \boldsymbol{\psi}) = C^{(1)}(i | \boldsymbol{\psi}_1) C^{(2)}(\mathbf{k} | \boldsymbol{\psi}_2) \quad (3.2)$$

onde $C^{(1)}$ é uma função positiva definida em \mathbb{R}^1 , $C^{(2)}$ é uma função positiva definida em \mathbb{R}^d e $\boldsymbol{\psi}' = (\boldsymbol{\psi}'_1, \boldsymbol{\psi}'_2)$. Se a estrutura de covariância de um modelo pode ser escrita como na Equação (3.2), dizemos que o modelo é separável.

Embora a separabilidade leve a cálculos rápidos, a classe de modelos com essa característica é de aplicação limitada pois separabilidade indica que não há interação entre o espaço e o tempo. Alguns exemplos de classes de funções de covariância não separáveis podem ser encontrados em [Cressie e Huang \(1999\)](#).

3.3 Separabilidade na nossa classe de modelos

Analisaremos a seguir as propriedades de separabilidade dos modelos espaço-temporais propostos na Seção 2.4. Nas seguintes subseções assume-se que as matrizes de covariância são constantes para todo t , isto é, $\mathbf{V}_t = \mathbf{V}$ e $\mathbf{W}_t = \mathbf{W}$, $\forall t$.

3.3.1 Modelo polinomial de primeira ordem

Na Subseção 2.4.1 indica-se que o modelo é estacionário se $\rho \in (-1, 1)$. Quando o modelo é estacionário o valor esperado de \mathbf{y}_t é o vetor de zeros e a sua matriz de covariância $\mathbf{V} + (1 - \rho^2)^{-1} \mathbf{W}$. Além disso, a covariância entre \mathbf{y}_{tk} e $\mathbf{y}_{t+i,l}$ é $(1 - \rho^2)^{-1} \rho^i w_{kl}$, onde w_{kl} é o elemento (k, l) de \mathbf{W} . É evidente que modelos com estas especificações têm funções de covariância que podem fatorar-se como o produto de funções de covariância espacial e temporal, então este é um modelo separável.

3.3.2 Modelo polinomial de segunda ordem

A seguinte proposição determina a estrutura de dependência entre as diversas quantidades de interesse em um modelo polinomial de segunda ordem ou de crescimento.

Teorema 3.3.1 *Considere o modelo polinomial de segunda ordem estacionário com $|\rho_1| < 1$, $|\rho_2| < 1$, $\mathbf{W}_{1t} = \mathbf{W}_1$ e $\mathbf{W}_{2t} = \mathbf{W}_2$, $\forall t$. Então:*

(i) $E(\boldsymbol{\beta}_{1t}) = E(\boldsymbol{\beta}_{2t}) = 0$ e a matriz de covariância de $\boldsymbol{\beta}_{1t}$ e $\boldsymbol{\beta}_{2t}$ é

$$Cov(\boldsymbol{\beta}_{1t}, \boldsymbol{\beta}_{2t}) = \frac{\rho_1 \rho_2}{1 - \rho_1 \rho_2} \mathbf{W}_2. \quad (3.3)$$

(ii) A matriz de covariância $\boldsymbol{\beta}_{1t}$ é

$$Var(\boldsymbol{\beta}_{1t}) = \frac{1}{1 - \rho_1^2} \left[\left(\frac{\rho_1^2}{1 - \rho_2^2} + \frac{2\rho_1^3 \rho_2}{1 - \rho_1 \rho_2} \right) \mathbf{W}_2 + \mathbf{W}_1 \right]. \quad (3.4)$$

(iii) A covariância entre \mathbf{y}_{tk} e $\mathbf{y}_{t+i,l}$ é

$$Cov(\mathbf{y}_{tk}, \mathbf{y}_{t+i,l}) = \rho_1^i \{Var(\boldsymbol{\beta}_{1t})\}_{kl} + \rho_1 \frac{\rho_2^i - \rho_1^i}{\rho_2 - \rho_1} \{Cov(\boldsymbol{\beta}_{1t}, \boldsymbol{\beta}_{2t})\}_{kl} \quad (3.5)$$

Prova:

(i) Dos resultados usuais da análise de séries temporais para processos auto-regressivos de primeira ordem, $E(\boldsymbol{\beta}_{2t}) = 0$ e $Cov(x_{2t}) = (1 - \rho_2^2)^{-1} \mathbf{W}_2$. Aplicando o operador esperança na equação do nível, usando o fato que pela estacionariedade $E(\boldsymbol{\beta}_{1t}) = E(\boldsymbol{\beta}_{1,t-1})$, e resolvendo para $E(\boldsymbol{\beta}_{1t})$, encontramos que $E(\boldsymbol{\beta}_{1t}) = 0$. Além disso,

$$\begin{aligned} E(\boldsymbol{\beta}_{1t} \boldsymbol{\beta}'_{2t}) &= E \{ [\rho_1 (\boldsymbol{\beta}_{1,t-1} + \boldsymbol{\beta}_{2,t-1})] [\rho_2 \boldsymbol{\beta}_{2,t-1} + \mathbf{w}_{2t}]' \} \\ &= \rho_1 \rho_2 E(\boldsymbol{\beta}_{1,t-1} \boldsymbol{\beta}'_{2,t-1}) + \rho_1 \rho_2 \mathbf{W}_2. \end{aligned}$$

Solucionando a expressão acima para $E(\boldsymbol{\beta}_{1t} \boldsymbol{\beta}'_{2t})$, resulta

$$E(\boldsymbol{\beta}_{1t} \boldsymbol{\beta}'_{2t}) = \rho_1 \rho_2 (1 - \rho_1 \rho_2)^{-1} \mathbf{W}_2.$$

(ii) Aplicando o operador variância na equação do nível, obtemos

$$\begin{aligned} Var(\boldsymbol{\beta}_{1t}) &= \rho_1^2 Var(\boldsymbol{\beta}_{1,t-1}) + \rho_1^2 Var(\boldsymbol{\beta}_{2,t-1}) + 2\rho_1^2 E(\boldsymbol{\beta}_{1,t-1}\boldsymbol{\beta}'_{2,t-1}) + \mathbf{W}_1 \\ &= \rho_1^2 Var(\boldsymbol{\beta}_{1t}) + \rho_1^2(1 - \rho_2^2)^{-1}\mathbf{W}_2 + 2\rho_1^3\rho_2(1 - \rho_1\rho_2)^{-1}\mathbf{W}_2 + \mathbf{W}_1. \end{aligned}$$

Solucionando a expressão acima para $Var(\boldsymbol{\beta}_{1t})$, resulta

$$Var(\boldsymbol{\beta}_{1t}) = \frac{1}{1 - \rho_1^2} \left[\left(\frac{\rho_1^2}{1 - \rho_2^2} + \frac{2\rho_1^3\rho_2}{1 - \rho_1\rho_2} \right) \mathbf{W}_2 + \mathbf{W}_1 \right].$$

(iii) Vamos provar o item (iii) por indução. É fácil mostrar que o item (iii) mantém-se para $i = 1$:

$$\begin{aligned} Cov(\mathbf{y}_t, \mathbf{y}_{t+1}) &= E(\boldsymbol{\beta}_{1t}\boldsymbol{\beta}'_{1,t+1}) \\ &= E[\boldsymbol{\beta}_{1t}(\rho_1\boldsymbol{\beta}_{1t} + \rho_1\boldsymbol{\beta}_{2t} + \mathbf{w}_{t+1})'] \\ &= \rho_1 Var(\boldsymbol{\beta}_{1t}) + \rho_1 Cov(\boldsymbol{\beta}_{1t}, \boldsymbol{\beta}_{2t}). \end{aligned}$$

Agora, assumamos que o item (iii) é válido para $i > 1$. Assim:

$$\begin{aligned} Cov(\mathbf{y}_t, \mathbf{y}_{t+i+1}) &= E(\boldsymbol{\beta}_{1t}\boldsymbol{\beta}'_{1,t+i+1}) \\ &= \rho_1 E(\boldsymbol{\beta}_{1t}\boldsymbol{\beta}'_{1,t+i}) + \rho_1\rho_2^i E(\boldsymbol{\beta}_{1t}\boldsymbol{\beta}'_{2t}) \\ &= \rho_1 \left[\rho_1^i Var(\boldsymbol{\beta}_{1t}) + \rho_1 \frac{\rho_2^i - \rho_1^i}{\rho_2 - \rho_1} Cov(\boldsymbol{\beta}_{1t}, \boldsymbol{\beta}_{2t}) \right] \\ &\quad + \rho_1\rho_2^i Cov(\boldsymbol{\beta}_{1t}, \boldsymbol{\beta}_{2t}) \\ &= \rho_1^{i+1} Var(\boldsymbol{\beta}_{1t}) + \rho_1 \frac{\rho_2^{i+1} - \rho_1^{i+1}}{\rho_2 - \rho_1} Cov(\boldsymbol{\beta}_{1t}, \boldsymbol{\beta}_{2t}) \end{aligned}$$

Assim, por indução o item (iii) é válido para todo $i \geq 1$.

Seguindo diretamente do item (iii) podemos afirmar que modelos polinomiais de segunda ordem com $\rho_2 \neq 0$ são não separáveis. \square

3.3.3 Modelo de contaminação

Com já foi notado na Seção 2.4.3, a matriz de evolução do sistema \mathbf{G}_t deste modelo tem uma estrutura similar à da matriz de vizinhança \mathbf{M} , incluindo um parâmetro de contaminação $\kappa > 0$. Sabe-se que para $\rho \in (-1, 1)$ o modelo é estacionário. O'Hagan (1998)

mostrou que um processo Gaussiano estacionário tem estrutura de covariância separável se, e somente se, a esperança condicional do processo em qualquer região no tempo $t + 1$, dado o campo completo no tempo t , depende somente do valor do processo na mesma região no tempo t . Evidentemente, esta condição não é satisfeita por modelos com \mathbf{G}_t definida como na Equação (2.7), pois mediante a matriz de contaminação \mathbf{H} permitimos a interação do processo no tempo t e na região s (β_{ts}) com o processo no tempo $t + 1$ e nas regiões vizinhas de s . Então o modelo de contaminação espaço-temporal é não separável.

3.4 Métodos alternativos de estimação do processo latente

O filtro de Kalman e o algoritmo *FFBS* foram definidos anteriormente na Subseção 2.5.2. Note-se que o *FFBS* requer pelo menos três inversões de matrizes, uma no filtro de Kalman e duas na suavização (Equações (2.11)-(2.12)) para cada tempo, o que é computacionalmente custoso, mais ainda quando o problema tem uma dimensão S grande. Uma alternativa ao uso do *FFBS* é o algoritmo que denominaremos *FIFBS* (do inglês, *forward information filter backward sampler*) comentado nas próximas subseções.

3.4.1 Filtro de informação

Existem diversas variantes do filtro de Kalman, matematicamente equivalentes, que têm vantagens computacionais em casos especiais. Esse é o caso do filtro de informação (Anderson e Moore, 1979) que detalhamos a seguir.

Para aplicar o filtro de informação assume-se, como no filtro de Kalman, que $\mathbf{D}_t = \mathbf{D}_{t-1} \cup \{\mathbf{y}_t\}$ e que a priori $\beta_0 | \mathbf{D}_0 \sim N(\mathbf{m}_0, \mathbf{C}_0)$ com \mathbf{m}_0 e \mathbf{C}_0 conhecidos. Definimos as distribuições

- Priori em t : $\beta_t | \psi, \mathbf{D}_{t-1} \sim N(\mathbf{a}_t, \mathbf{R}_t)$,
- Posteriori em t : $\beta_t | \psi, \mathbf{D}_t \sim N(\mathbf{m}_t, \mathbf{C}_t)$,

com

$$\mathbf{a}_t = \mathbf{G}_t \mathbf{m}_{t-1}, \quad (3.6)$$

$$\mathbf{R}_t^{-1} = \mathbf{W}_t^{-1} - \mathbf{W}_t^{-1} \mathbf{G}_t (\mathbf{G}_t' \mathbf{W}_t^{-1} \mathbf{G}_t + \mathbf{C}_{t-1}^{-1})^{-1} \mathbf{G}_t' \mathbf{W}_t^{-1}, \quad (3.7)$$

$$\mathbf{C}_t^{-1} = \mathbf{R}_t^{-1} + \mathbf{F}_t \mathbf{V}_t^{-1} \mathbf{F}_t', \quad (3.8)$$

$$\mathbf{C}_t^{-1} \mathbf{m}_t = \mathbf{R}_t^{-1} \mathbf{a}_t + \mathbf{F}_t \mathbf{V}_t^{-1} \mathbf{y}_t. \quad (3.9)$$

Note-se que as equações estão definidas em termos das matrizes esparsas \mathbf{V}_t^{-1} e \mathbf{W}_t^{-1} , e a única inversão de matrizes necessária é \mathbf{R}_t^{-1} na Equação (3.8).

3.4.2 Forward information filter backward sampler

De forma análoga ao *FFBS*, o algoritmo *FIFBS* consta de duas etapas: a primeira etapa consiste em aplicar o filtro de informação, descrito na seção anterior. Guardamos a cada tempo os valores da matriz \mathbf{C}_t^{-1} e do vetor $\mathbf{C}_t^{-1} \mathbf{m}_t$, $t = 1, \dots, T$ para utilizarmos na segunda etapa do método.

Ao término do filtro de informação temos \mathbf{C}_T^{-1} e $\mathbf{C}_T^{-1} \mathbf{m}_T$ e iniciamos a segunda etapa simulando um valor do processo no tempo T da sua distribuição a posteriori $\beta_T^* | \boldsymbol{\psi}, \mathbf{D}_T \sim N(\mathbf{m}_T, \mathbf{C}_T)$. A segunda etapa do *FIFBS* continua com a amostragem recursiva retrospectiva de $\beta_t^* | \beta_{t+1}^*, \boldsymbol{\psi}, \mathbf{D}_T \sim N(\mathbf{b}_t, \mathbf{B}_t)$, $t = T - 1, \dots, 1$, com

$$\mathbf{B}_t^{-1} = \mathbf{C}_t^{-1} + \mathbf{G}_{t+1}' \mathbf{W}_{t+1}^{-1} \mathbf{G}_{t+1}, \quad (3.10)$$

$$\mathbf{B}_t^{-1} \mathbf{b}_t = \mathbf{C}_t^{-1} \mathbf{m}_t + \mathbf{G}_{t+1}' \mathbf{W}_{t+1}^{-1} \beta_{t+1}^*. \quad (3.11)$$

Note que com $\mathbf{B}_t^{-1} \mathbf{b}_t$ e \mathbf{B}_t^{-1} podemos usar uma decomposição de Cholesky, duas soluções de sistemas lineares triangulares e uma soma vetorial para simular eficientemente β_t^* . Assim, seja \mathbf{L} a matriz triangular inferior obtida da decomposição de Cholesky de \mathbf{B}_t^{-1} , ou seja $\mathbf{B}_t^{-1} = \mathbf{L}\mathbf{L}'$ e seja \mathbf{z}_t um vetor de variáveis normais padrão independentes, $\mathbf{z}_t \sim N(\mathbf{0}, \mathbf{I})$. Então:

$$\mathbf{B}_t = (\mathbf{L}\mathbf{L}')^{-1}$$

$$\mathbf{B}_t = (\mathbf{L}^{-1})' \mathbf{L}^{-1}.$$

Sabe-se que $\beta_t^* | \beta_{t+1}^*, \psi, \mathbf{D}_T$ tem uma distribuição normal multivariada com vetor de médias \mathbf{b}_t , e matriz de covariância \mathbf{B}_t , então podemos escrever

$$\begin{aligned}\beta_t^* &= (\mathbf{L}^{-1})' \mathbf{z}_t + \mathbf{b}_t \\ \beta_t^* &= (\mathbf{L}')^{-1} \mathbf{z}_t + \mathbf{b}_t \\ \mathbf{L}' \beta_t^* &= \mathbf{z}_t + \mathbf{L}' \mathbf{b}_t.\end{aligned}\tag{3.12}$$

A expressão $\mathbf{L}' \mathbf{b}_t$ é obtida facilmente depois de arranjar o vetor $\mathbf{B}_t^{-1} \mathbf{b}_t$ de (3.11),

$$\begin{aligned}\mathbf{B}_t^{-1} \mathbf{b}_t &= (\mathbf{L} \mathbf{L}') \mathbf{b}_t \\ \mathbf{B}_t^{-1} \mathbf{b}_t &= \mathbf{L} (\mathbf{L}' \mathbf{b}_t) \\ \mathbf{B}_t^{-1} \mathbf{b}_t &= \mathbf{L} \mathbf{v}_t.\end{aligned}\tag{3.13}$$

Encontrando a solução \mathbf{v}_t do sistema linear e substituindo em (3.12)

$$\begin{aligned}\mathbf{L}' \beta_t^* &= \mathbf{z}_t + \mathbf{v}_t \\ \mathbf{L}' \beta_t^* &= \mathbf{u}_t.\end{aligned}\tag{3.14}$$

A solução deste último sistema linear é o valor amostrado de β_t^* . Desta forma, o algoritmo *FIFBS* requer somente uma inversão de matriz dentro do filtro de informação em cada tempo e nenhuma na segunda etapa, resultando assim mais rápido que o *FFBS*.

3.5 Seleção de modelos

A seleção de modelos é uma atividade fundamental na análise de dados. Esta atividade torna-se mais importante com os avanços das ferramentas computacionais que permitem a formulação e o ajuste de modelos cada vez mais complexos. Existe uma grande variedade de procedimentos e sugestões na literatura para comparar modelos Bayesianos com ferramentas de diagnóstico e seleção de modelos baseado em testes ou critérios de seleção. [Gelfand \(1996\)](#), [Raftery \(1996\)](#), [Gelman e Meng \(1996\)](#) e muitos capítulos em [Dey et al. \(2000\)](#) fazem uma revisão dos métodos existentes.

3.5.1 Fator de Bayes e outros critérios

A típica abordagem Bayesiana para seleção de modelos e escolha de variáveis considera um número finito de modelos plausíveis $m = 1, \dots, M$, assumindo que um deles é o “verdadeiro” modelo. Quando os modelos têm as probabilidades a priori iguais, as probabilidades a posteriori são proporcionais às respectivas densidades preditivas. O modelo que tem a maior probabilidade a posteriori é selecionado. No caso particular, quando comparados dois modelos m_1 e m_2 , a razão $p(m_1|Y)/p(m_2|Y)$ é chamada de fator de Bayes. Um fator de Bayes maior do que 1 favorece o modelo m_1 ; um valor menor que 1 favorece o modelo m_2 .

Obviamente, o problema está no cálculo das probabilidades a posteriori para modelos mais complexos. Abordagens atuais, baseadas em simulações *MCMC*, compreendem métodos para estimar diretamente as verossimilhanças marginais (Chib, 1995), amostrar sobre o espaço produto dos parâmetros e das indicadoras dos modelos (Carlin e Chib, 1995; Green, 1995), ou combinar simulações e aproximações assintóticas (Di Ciccio et al., 1997). Dellaportas et al. (2000) discutem procedimentos de escolha de variáveis utilizando fator de Bayes. Uma revisão comparativa de métodos *MCMC* para calcular fatores de Bayes encontra-se em Han e Carlin (2000), incluindo várias referências de trabalhos recentes.

Um outro problema é que os fatores de Bayes não estão bem definidos quando distribuições a priori impróprias são utilizadas. Tais prioris são muito comuns em modelos hierárquicos complexos ou modelos dinâmicos. Conseqüentemente, outros critérios para seleção de modelos, no mesmo espírito dos critérios *AIC* ou *BIC*, têm sido sugeridos na literatura nos últimos anos. Gelfand e Ghosh (1998) sugerem uma abordagem de maximização da utilidade, baseada nas predições a posteriori. Outro critério, proposto por Spiegelhalter et al. (2002) é o *DIC*, que pode ser visto como uma generalização do *AIC*. Ambos os critérios são facilmente calculados utilizando as amostras a posteriori. Não utilizamos o critério *DIC* como nosso critério de seleção de modelos pois ele tende a favorecer os modelos mais complexos.

3.5.2 Densidade preditiva

Quando é preciso comparar as capacidades preditivas de diferentes modelos, uma opção natural são as ferramentas baseadas nas densidades preditivas. [Gelfand et al. \(1992\)](#) propõem vários diagnósticos baseados na distribuição preditiva da validação cruzada (do inglês, *cross-validation*) $p(y_{ts}|\mathbf{y}_{t,-s})$, onde $\mathbf{y}_{t,-s}$ representa o conjunto de dados sem a observação y_{ts} no tempo t ; veja também [Gelfand \(1996\)](#). Outra opção é a densidade preditiva da previsão, para estruturas de séries temporais ([Nandram e Petrucelli, 1997](#)). Abordagens com algumas variantes encontram-se em [Geisser e Eddy \(1979\)](#) e [San Martini e Spezzaferrri \(1984\)](#).

[Gelman et al. \(1995b\)](#) sugerem verificar o modelo com a densidade preditiva a posteriori. Mais recentemente, [Dey et al. \(1998\)](#) sugeriram métodos mais gerais baseados em simulação, requerendo somente a especificação do modelo e técnicas de simulação da posteriori, por exemplo, os métodos *MCMC*. A densidade preditiva a posteriori, $p(\mathbf{y}|\mathbf{y}_{obs})$ é a densidade preditiva de um novo e independente conjunto de valores observados sob o modelo, dados os atuais valores observados ([Rubin, 1984](#); [Aitkin, 1991](#)).

Para estabelecer o nosso critério, supomos que temos que selecionar um entre M possíveis modelos espaço-temporais Gaussianos. O m -ésimo modelo, depois de reescrever as Equações (2.3)-(2.4), fica da seguinte forma:

$$\mathbf{y}_t|\boldsymbol{\beta}_t, \boldsymbol{\psi} \sim N(\mathbf{F}'_t\boldsymbol{\beta}_t, \mathbf{V}_t) \quad (3.15)$$

$$\boldsymbol{\beta}_t|\boldsymbol{\beta}_{t-1}, \boldsymbol{\psi} \sim N(\mathbf{G}_t\boldsymbol{\beta}_{t-1}, \mathbf{W}_t) \quad (3.16)$$

onde $\boldsymbol{\psi}$ representa o vetor de parâmetros.

Seja $p_m(\boldsymbol{\beta}_{1:(t-1)}, \boldsymbol{\psi}|\mathbf{D}_{t-1})$ a distribuição a posteriori conjunta de $\boldsymbol{\psi}$ e de $\boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_{t-1}$ sob o modelo m até o tempo $t-1$, onde $\mathbf{D}_{t-1} = (\mathbf{y}_1, \dots, \mathbf{y}_{t-1})$. Com essas considerações a distribuição preditiva $p_m(\mathbf{y}_t|\mathbf{D}_{t-1})$ é calculada como segue:

$$\begin{aligned} p_m(\mathbf{y}_t|\mathbf{D}_{t-1}) &= \int p_m(\mathbf{y}_t|\boldsymbol{\beta}_t, \boldsymbol{\psi})p_m(\boldsymbol{\beta}_t|\boldsymbol{\beta}_{t-1}, \boldsymbol{\psi})d\boldsymbol{\beta}_t \\ &\quad \times \int p_m(\boldsymbol{\beta}_{1:(t-1)}, \boldsymbol{\psi}|\mathbf{D}_{t-1})d\boldsymbol{\beta}_{1:(t-1)}d\boldsymbol{\psi} \\ p_m(\mathbf{y}_t|\mathbf{D}_{t-1}) &= \int p_m(\mathbf{y}_t|\boldsymbol{\beta}_{t-1}, \boldsymbol{\psi})p_m(\boldsymbol{\beta}_{1:(t-1)}, \boldsymbol{\psi}|\mathbf{D}_{t-1})d\boldsymbol{\beta}_{1:(t-1)}d\boldsymbol{\psi} \end{aligned} \quad (3.17)$$

pois

$$p_m(\mathbf{y}_t | \boldsymbol{\beta}_{t-1}, \boldsymbol{\psi}) = \int p_m(\mathbf{y}_t | \boldsymbol{\beta}_t, \boldsymbol{\psi}) p_m(\boldsymbol{\beta}_t | \boldsymbol{\beta}_{t-1}, \boldsymbol{\psi}) d\boldsymbol{\beta}_t$$

com $\boldsymbol{\beta}_{1:(t-1)} = (\boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_{t-1})$.

Uma das vantagens de trabalhar com distribuições preditivas é a facilidade dos cálculos. Modelos complexos, como os modelos espaço-temporais são ajustados com frequência utilizando métodos de simulação, como o *MCMC*. Uma amostra $(\boldsymbol{\beta}_{t-1}^{(1)}, \boldsymbol{\psi}^{(1)}), \dots, (\boldsymbol{\beta}_{t-1}^{(L)}, \boldsymbol{\psi}^{(L)})$ da distribuição conjunta a posteriori $p_m(\boldsymbol{\beta}_{1:(t-1)}, \boldsymbol{\psi} | \mathbf{D}_{t-1})$ pode ser utilizada para calcular uma estimativa da densidade preditiva de \mathbf{y}_t , dada a informação obtida até o tempo $t - 1$ pelo método Monte Carlo

$$\hat{p}_m(\mathbf{y}_t | \mathbf{D}_{t-1}) = \frac{1}{J} \sum_{j=1}^J p_m(\mathbf{y}_t | \boldsymbol{\beta}_{t-1}^{(j)}, \boldsymbol{\psi}^{(j)}), \quad (3.18)$$

em que J é suficientemente grande e $\boldsymbol{\beta}_{t-1}^{(j)}$ e $\boldsymbol{\psi}^{(j)}$ são amostras obtidas da distribuição a posteriori conjunta após atingir a convergência. Com as Equações (3.15) e (3.16) podemos encontrar os termos do somatório do lado direito da Equação (3.18) para o nosso modelo espaço-temporal geral:

$$\mathbf{y}_t | \boldsymbol{\beta}_{t-1}, \boldsymbol{\psi} \sim N(\mathbf{F}'_t \mathbf{G}_t \boldsymbol{\beta}_{t-1}, \mathbf{V}_t + \mathbf{F}'_t \mathbf{W}_t \mathbf{F}_t), \quad (3.19)$$

isto simplifica o cálculo da densidade preditiva.

Assim, para cada modelo m e para cada tempo t , aplicamos um esquema *MCMC* completo e fazemos a previsão para o tempo seguinte $t + 1$. Utilizando o fato de que a densidade preditiva conjunta de $\mathbf{y}_{t^*+1}, \dots, \mathbf{y}_T$ pode ser escrita como $p_m(\mathbf{y}_{t^*+1}, \dots, \mathbf{y}_T | \mathbf{D}_{t^*}) = \prod_{t=t^*+1}^T p_m(\mathbf{y}_t | \mathbf{D}_{t-1})$, uma estimativa da densidade preditiva conjunta sob o modelo m é

$$\hat{p}_m(\mathbf{y}_{t^*+1}, \dots, \mathbf{y}_T | \mathbf{D}_{t^*}) = \prod_{t=t^*+1}^T \hat{p}_m(\mathbf{y}_t | \mathbf{D}_{t-1}), \quad (3.20)$$

em que t^* é tal que a densidade $p_m(\boldsymbol{\psi} | \mathbf{D}_{t^*})$ é própria para todos os modelos considerados $m = 1, \dots, M$. Este método é uma medida da performance preditiva dos modelos. O modelo escolhido é aquele com a maior densidade preditiva. Desta forma, o modelo selecionado além de ter a probabilidade a posteriori mais alta, também tem a melhor performance preditiva.

Quando algum modelo tem probabilidade a posteriori perto de um, esse modelo é o escolhido. Mas pode ocorrer que vários modelos tenham probabilidades a posteriori similares. Nesse caso, esses vários modelos devem ser reportados e as previsões devem ser calculadas por mistura Bayesiana de modelos (Clyde e George, 2004).

Neste capítulo foram acrescentados alguns tópicos no estudo dos modelos espaço-temporais Gaussianos, tais como a separabilidade, o filtro de informação, o algoritmo *FIFBS* para simular o processo latente e a seleção do melhor modelo, entre um conjunto de possíveis modelos, via a densidade preditiva. Quando temos dados não Gaussianos que pertencem à família exponencial podemos trabalhar com duas abordagens: (i) transformar os dados mediante alguma função real para serem aproximadamente Gaussianos e desta forma os resultados deste capítulo podem ser aplicados sem problema; ou (ii) estender os modelos Gaussianos para uma classe mais geral e, seguindo o mesmo raciocínio apresentado aqui, generalizar o *FIFBS* e a seleção de modelos para dados na família exponencial. Uma aplicação utilizando a primeira abordagem será apresentada no seguinte capítulo.

Capítulo 4

Modelagem de dados espaço-temporais não Gaussianos utilizando transformações

4.1 Introdução

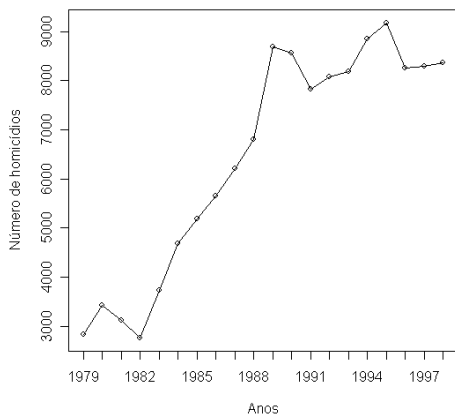
Apresentamos neste capítulo um estudo com dados de violência (homicídios) nos Estados brasileiros de Rio de Janeiro e Espírito Santo. Nossa classe de modelos espaço-temporais discutidos no Capítulo 2 pode ser aplicada a dados Gaussianos ou aproximadamente Gaussianos. Devido à natureza dos dados de violência, as opções para trabalhar com estes dados são duas: (i) transformar o conjunto dos dados utilizando alguma função real para ficar aproximadamente Gaussiano, ou (ii) estender naturalmente a nossa classe de modelos para tratar com observações não Gaussianas que pertençam à família exponencial. Esta última abordagem é apresentada nos próximos capítulos. Nas seguintes Seções desenvolve-se a aplicação aos dados transformados. Para justificar a suposição de estrutura espacial nas inovações, um estudo comparativo com modelos mais simples que não incluem dependência no espaço foi realizado nas aplicações. Este capítulo corresponde em grande parte ao nosso primeiro artigo ([Vivar e Ferreira, 2007](#)).

4.2 Dados de violência

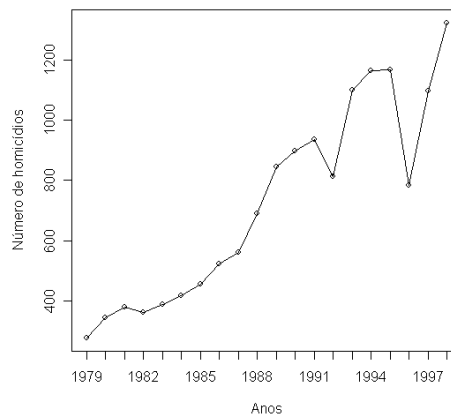
Os dados consistem no número anual de homicídios por município dos Estados de Rio de Janeiro e Espírito Santo, desde o ano 1979 até 1998. Durante as décadas dos anos 80 e 90 foram criados vários novos municípios por divisão ou fusão dos antigos municípios. Por razões de compatibilidade utilizaremos em todas as análises o mapa político do ano 1979 para cada Estado. A relação completa de todos os municípios, a divisão política de 1979, assim como a taxa de mortalidade por homicídio por cem mil habitantes para cada município e para cada ano, encontram-se nos Apêndices A e B.

Tabela 4.1: Número de municípios por Estado.

Estado	Número de municípios em 1979	Número de municípios em 1998
Rio de Janeiro	64	91
Espírito Santo	52	78



(a)



(b)

Figura 4.1: Número agregado de homicídios por ano de estudo: (a) Rio de Janeiro, e (b) Espírito Santo.

A Tabela 4.1 indica o número de municípios de cada Estado nos anos de início e término do estudo. A Figura 4.1 ilustra o número total de homicídios em cada Estado para o período de estudo. Os dois gráficos apresentam uma tendência crescente, embora a magnitude dos dados do Rio de Janeiro seja muito maior que a do Espírito Santo. Estes dados foram cedidos pelo Professor Oswaldo Cruz do PROCC, Fiocruz.

4.3 Aproximação Gaussiana

Para cada município s e ano t , $s = 1, \dots, S$, $t = 1, \dots, T$, tem-se a população estimada n_{ts} e o número observado de homicídios z_{ts} . Como é típico para dados de contagem como estes, assume-se que z_{ts} segue uma distribuição de Poisson. Mais especificamente, assume-se que

$$z_{ts} | \lambda_{ts} \sim Po(n_{ts} \lambda_{ts}), \quad (4.1)$$

onde λ_{ts} é o risco latente no ano t no município s . Assim, supõe-se que o número de homicídios z_{ts} por município são independentes condicionais ao valor de λ_{ts} . O número de homicídios em ambos Estados, lamentavelmente, é alto, é por esta razão que uma aproximação Gaussiana via transformação dos dados deve funcionar adequadamente.

Nosso principal interesse é entender a dinâmica espaço-temporal da taxa de mortalidade por homicídio (TMH) por 100000 habitantes: $TMH_{ts} = 100000 z_{ts}/n_{ts}$, assim, a aproximação é feita da seguinte forma:

$$TMH_{ts} \approx N \left(10^5 \lambda_{ts}, 10^{10} \frac{\lambda_{ts}}{n_{ts}} \right). \quad (4.2)$$

Para normalizar e estabilizar a variância dos dados, utilizamos a transformação raiz quadrada (Anscombe, 1948). Defina y_{ts} como a raiz quadrada da taxa de mortalidade por homicídio por 100000 habitantes, ou seja, $y_{ts} = \sqrt{TMH_{ts}} = \sqrt{10^5 z_{ts}/n_{ts}}$. Para encontrar a distribuição de y_{ts} usaremos o método delta.

Método delta:

Suponha que $E(\mathbf{X}) = \mathbf{a}$, $Var(\mathbf{X}) = \mathbf{A}$ e φ é uma transformação um a um de \mathbf{X} com derivadas bem definidas no ponto \mathbf{a} . Então

$$\mathbf{Y} = \varphi(\mathbf{X}) = \varphi(\mathbf{a}) + (\mathbf{X} - \mathbf{a})' \frac{\partial \varphi(\mathbf{a})}{\partial \mathbf{X}} + o(\mathbf{X} - \mathbf{a}) \quad (4.3)$$

Nesta equação, o termo $|o(\mathbf{u})|/|\mathbf{u}| \rightarrow 0$ quando $\mathbf{u} \rightarrow 0$. Se \mathbf{X} está próximo de \mathbf{a} então o último termo do lado direito pode ser omitido e \mathbf{Y} terá uma relação aproximadamente linear com \mathbf{X} onde

$$E(\mathbf{Y}) \doteq \boldsymbol{\varphi}(\mathbf{a}) \quad (4.4)$$

$$Var(\mathbf{Y}) \doteq \left(\frac{\partial \boldsymbol{\varphi}(\mathbf{a})}{\partial \mathbf{X}} \right)' \mathbf{A} \frac{\partial \boldsymbol{\varphi}(\mathbf{a})}{\partial \mathbf{X}}. \quad (4.5)$$

Se \mathbf{X} tem distribuição Gaussiana então \mathbf{Y} tem distribuição aproximadamente Gaussiana. Este resultado é conhecido como o método delta (Migon e Gamerman, 1999). Conseqüentemente, identificando os valores de $\mathbf{a} = 10^5 \lambda_{ts}$, $\mathbf{A} = 10^{10} \lambda_{ts}/n_{ts}$ e $\boldsymbol{\varphi}(\mathbf{X}) = (\mathbf{X})^{0.5}$ pode-se obter finalmente

$$\begin{aligned} E(y_{ts}) &\doteq (10^5 \lambda_{ts})^{0.5} \\ &\doteq 10^{2.5} (\lambda_{ts})^{0.5} \end{aligned} \quad (4.6)$$

$$\begin{aligned} Var(y_{ts}) &\doteq \left(\frac{1}{2} (10^5 \lambda_{ts})^{-0.5} \right) \left(10^{10} \frac{\lambda_{ts}}{n_{ts}} \right) \left(\frac{1}{2} (10^5 \lambda_{ts})^{-0.5} \right) \\ &\doteq \frac{10^5}{4n_{ts}} \end{aligned} \quad (4.7)$$

Portanto, y_{ts} tem uma distribuição aproximadamente normal

$$y_{ts} \stackrel{a}{\sim} N \left(10^{2.5} \sqrt{\lambda_{ts}}, 10^5 / (4n_{ts}) \right). \quad (4.8)$$

4.4 Modelos espaço-temporais propostos

Usando a Equação (4.8) a nossa classe de modelos espaço-temporais Gaussianos pode ser aplicada diretamente ao conjunto de dados transformados, tendo como matriz de covariância das observações $\mathbf{V}_t = \frac{10^5}{4} \text{diag}(n_{t1}^{-1}, \dots, n_{tS}^{-1})$ e $\mathbf{F}'_t \boldsymbol{\beta}_t = 10^{2.5} (\sqrt{\lambda_{t1}}, \dots, \sqrt{\lambda_{tS}})'$. Assim, o nosso modelo geral para esta aplicação é

$$\mathbf{y}_t = \mathbf{F}'_t \boldsymbol{\beta}_t + \boldsymbol{\epsilon}_t, \quad \boldsymbol{\epsilon}_t \sim N(\mathbf{0}_S, \mathbf{V}_t), \quad (4.9)$$

$$\boldsymbol{\beta}_t = \mathbf{G}_t \boldsymbol{\beta}_{t-1} + \boldsymbol{\omega}_t, \quad \boldsymbol{\omega}_t \sim CAMGP(\mathbf{0}_S, \mathbf{W}_t^{-1}), \quad (4.10)$$

em que S , T e h estão definidos na Tabela 4.2.

Tabela 4.2: Especificação de S , T e h para os Estados.

Característica	RJ	ES
Número de municípios S	64	52
Anos de coleta de dados (1979 - 1998) T	20	20
Número máximo de vizinhos de um município h	9	10

Nas seguintes Subseções descrevemos cada um dos diversos modelos espaço-temporais ajustados baseados no modelo geral (4.9)-(4.10) e derivados daqueles modelos apresentados na Seção 2.4. A matriz de vizinhança \mathbf{M} da matriz de precisão \mathbf{W}_t^{-1} do *CAMGP*, foi elaborada para ter estrutura de vizinhança de primeira ordem, ou seja, considerando vizinhos aqueles municípios que no ano de 1979 tinham fronteiras comuns, isto é, cada elemento de \mathbf{M} é

$$(\mathbf{M})_{k,l} = \begin{cases} r_k, & k = l \\ -1, & k \in N_l \\ 0, & \text{caso contrário,} \end{cases} \quad (4.11)$$

em que r_k é o número de vizinhos de cada município. Esta matriz é esparsa, fato que pode ser aproveitado para acelerar os cálculos. Um algoritmo para reordenar os municípios de forma que a largura de banda (*bandwidth*, em inglês) da matriz seja minimizada encontra-se na biblioteca *GMRFLib*, versão 1.07 de Rue e Follestad (2003). Um exemplo desta redução da largura de banda está na Figura 4.2.

Uma análise exploratória dos dados transformados revela um crescimento na raiz quadrada da *TMH* ao longo do tempo em muitos municípios (ver gráficos da *TMH* por município para os dois Estados nos Apêndices A e B). Esta tendência indica que o processo é não estacionário. Por esta razão, todos os modelos devem ser ajustados com $\rho = 1$. Ao todo, 8 modelos foram especificados com tal característica. O modelo I é o mais simples e é usado para suavizar os dados. O modelo II, de contaminação, é usado, pois os mapas dos Estados em cada ano apresentam grupos de municípios com comportamentos similares. Para modelar o crescimento detectado na análise exploratória, os modelos III, IV, ..., e VIII

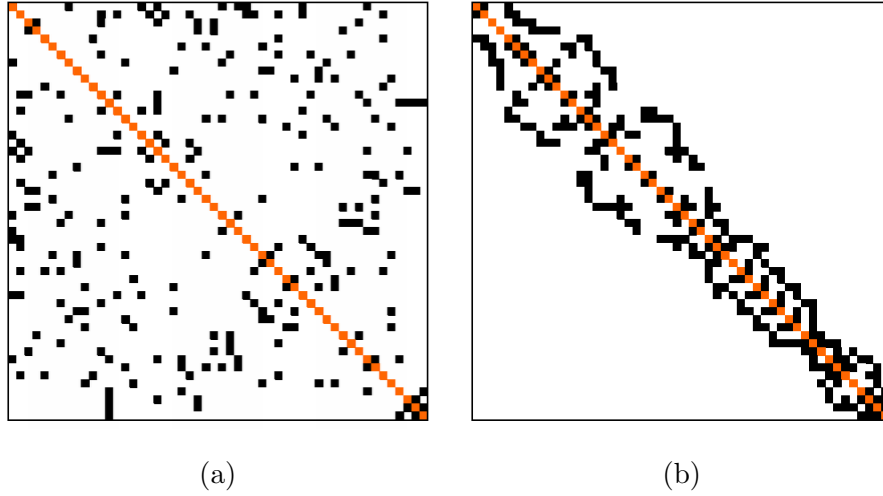


Figura 4.2: Matriz de vizinhança dos municípios de Espírito Santo: (a) Antes da reordenação. (b) Após a reordenação.

são variantes do modelo polinomial de segunda ordem (ver Secção 2.4.2), alguns inclusive modelando contaminação nas equações de velocidade ou de nível.

4.4.1 Modelo I: Polinomial de primeira ordem

Este é o modelo espaço-temporal mais simples, no qual os valores observados em cada tempo \mathbf{y}_t são iguais ao processo latente β_t mais o vetor de erros ϵ_t , e o β_t segue um processo autorregressivo de primeira ordem. O erro de evolução ω_t segue um *CAMGP* com matriz de precisão \mathbf{W}_t^{-1} . As matrizes \mathbf{F}_t e \mathbf{G}_t são iguais a \mathbf{I}_S , matriz identidade de ordem S . O modelo pode ser resumido como:

- $\mathbf{F}'_t = \mathbf{I}_S$ e $\mathbf{G}_t = \mathbf{I}_S$,
- $\mathbf{W}_t^{-1} = \tau(\mathbf{I}_S + \phi\mathbf{M})$.

4.4.2 Modelo II: Contaminação

O modelo de contaminação pode incorporar a informação que temos sobre diversos grupos de municípios vizinhos que apresentam similares taxas de mortalidade por homicídio em cada tempo. Pode ocorrer que um determinado município com uma alta (ou baixa) *TMH*

tenha alguma influência nos seus vizinhos nos tempos seguintes, devido a um incremento (ou redução) na violência. Esta influência é modelada com o parâmetro κ incluído na matriz de evolução \mathbf{G}_t . As matrizes do modelo de contaminação são especificadas a seguir.

- $\mathbf{F}'_t = \mathbf{I}_S$,
- $\mathbf{G}_t = \frac{1}{1+\kappa h} \mathbf{H} \longrightarrow \{\mathbf{H}\}_{kl} = \begin{cases} 1, & k = l, \\ \kappa, & k \in N_l, \text{ Matriz de contaminação} \\ 0, & c.c. \end{cases}$
- $\mathbf{W}_t^{-1} = \tau(\mathbf{I}_S + \phi \mathbf{M})$.

4.4.3 Modelos III e IV: Polinomiais de segunda ordem

O modelo polinomial de segunda ordem, também conhecido como modelo de crescimento linear, tem duas equações de sistema: a equação de nível e a equação de velocidade de crescimento do nível. Como muitos municípios apresentam um \sqrt{THM} crescente este comportamento pode ser capturado por estes modelos. As matrizes correspondentes são

- $\mathbf{F}'_t = (\mathbf{I}_S, \mathbf{0}_S)$,
- $\mathbf{G}_t = \begin{pmatrix} \mathbf{G}_{1t} & \mathbf{G}_{1t} \\ \mathbf{0}_S & \mathbf{G}_{2t} \end{pmatrix}$, $\mathbf{G}_{it} = \mathbf{I}_S, i = 1, 2$,
- $\mathbf{W}_t^{-1} = \begin{pmatrix} \mathbf{W}_{1t}^{-1} & \mathbf{0}_n \\ \mathbf{0}_n & \mathbf{W}_{2t}^{-1} \end{pmatrix}$, $\mathbf{W}_{it}^{-1} = \tau_i(\mathbf{I}_S + \phi_i \mathbf{M}), i = 1, 2$.

O modelo IV considera a estimação de todos os parâmetros (τ_1, ϕ_1, τ_2 , e ϕ_2) das matrizes de precisão da equação de sistema. O modelo III é uma variante que assume, a priori, que não existe dependência espacial na velocidade do processo, ou que esta dependência é pouco significativa, isto é, $\phi_2 = 0$.

4.4.4 Modelo V: Segunda ordem com contaminação na equação da velocidade

Com este modelo tentamos modelar conjuntamente o crescimento observado ao longo do tempo e a contaminação, assumindo que esta ocorre na equação da velocidade e que não existe influência do nível do processo no tempo t sobre os vizinhos de cada região no tempo seguinte. As componentes deste modelo são as seguintes:

- $\mathbf{F}'_t = (\mathbf{I}_S, \mathbf{0}_S)$,
- $\mathbf{G}_t = \begin{pmatrix} \mathbf{G}_{1t} & \mathbf{G}_{1t} \\ \mathbf{0}_S & \mathbf{G}_{2t} \end{pmatrix}$, $\mathbf{G}_{1t} = \mathbf{I}_S$ e $\mathbf{G}_{2t} = \frac{1}{1+\kappa_2 h} \mathbf{H} \longrightarrow \{\mathbf{H}\}_{kl} = \begin{cases} 1, & k = l, \\ \kappa_2, & k \in N_l, \\ 0, & c.c. \end{cases}$
- $\mathbf{W}_t^{-1} = \begin{pmatrix} \mathbf{W}_{1t}^{-1} & \mathbf{0}_n \\ \mathbf{0}_n & \mathbf{W}_{2t}^{-1} \end{pmatrix}$, $\mathbf{W}_{it}^{-1} = \tau_i(\mathbf{I}_S + \phi_i \mathbf{M})$, $i = 1, 2$.

4.4.5 Modelo VI: Segunda ordem com contaminação na equação do nível

Este modelo é similar ao anterior, assumimos que a contaminação acontece na equação do nível e que a equação da velocidade é um processo auto-regressivo. As componentes são as seguintes:

- $\mathbf{F}'_t = (\mathbf{I}_S, \mathbf{0}_S)$,
- $\mathbf{G}_t = \begin{pmatrix} \mathbf{G}_{1t} & \mathbf{G}_{1t} \\ \mathbf{0}_S & \mathbf{G}_{2t} \end{pmatrix}$,
- $\mathbf{G}_{2t} = \mathbf{I}_S$ e $\mathbf{G}_{1t} = \frac{1}{1+\kappa_1 h} \mathbf{H} \longrightarrow \{\mathbf{H}\}_{kl} = \begin{cases} 1, & k = l, \\ \kappa_1, & k \in N_l, \\ 0, & c.c. \end{cases}$
- $\mathbf{W}_t^{-1} = \begin{pmatrix} \mathbf{W}_{1t}^{-1} & \mathbf{0}_n \\ \mathbf{0}_n & \mathbf{W}_{2t}^{-1} \end{pmatrix}$, $\mathbf{W}_{it}^{-1} = \tau_i(\mathbf{I}_S + \phi_i \mathbf{M})$, $i = 1, 2$.

4.4.6 Modelo VII: Segunda ordem com contaminação nas duas equações de sistema

Com este modelo assumimos que existe contaminação no nível e na velocidade do processo latente. Isto é, valores muito altos (ou baixos) do nível e a velocidade do processo em determinados municípios no tempo t influenciam de alguma forma os valores do nível e a velocidade dos seus vizinhos nos tempos seguintes. O modelo fica definido com as seguintes matrizes:

- $F'_t = (\mathbf{I}_S, \mathbf{0}_S)$,
- $G_t = \begin{pmatrix} G_{1t} & G_{1t} \\ \mathbf{0}_S & G_{2t} \end{pmatrix}$, $G_{it} = \frac{1}{1+\kappa_i h} \mathbf{H}_i \longrightarrow \{\mathbf{H}_i\}_{kl} = \begin{cases} 1, & k = l, \\ \kappa_i, & k \in N_l, \quad , i = 1, 2, \\ 0, & c.c. \end{cases}$
- $W_t^{-1} = \begin{pmatrix} W_{1t}^{-1} & \mathbf{0}_n \\ \mathbf{0}_n & W_{2t}^{-1} \end{pmatrix}$, $W_{it}^{-1} = \tau_i(\mathbf{I}_S + \phi_i \mathbf{M})$, $i = 1, 2$.

4.4.7 Modelo VIII: Segunda ordem com $\phi_2 = 0$ e velocidade igual para cada município

Com este modelo assumimos que todos os municípios têm a mesma velocidade de crescimento ou decrescimento a cada tempo e que o vetor das inovações da velocidade não tem dependência espacial. As matrizes correspondentes a este modelo são as seguintes:

- $F'_t = (\mathbf{I}_S, \mathbf{0})$,
- $G_t = \begin{pmatrix} G_{1t} & \mathbf{1} \\ \mathbf{0} & G_{2t} \end{pmatrix}$, $G_{1t} = \mathbf{I}_S$ e $G_{2t} = 1$
- $W_t^{-1} = \begin{pmatrix} W_{1t}^{-1} & \mathbf{0} \\ \mathbf{0} & W_{2t}^{-1} \end{pmatrix}$, $W_{1t}^{-1} = \tau_1(\mathbf{I}_S + \phi_1 \mathbf{M})$ e $W_{2t}^{-1} = \tau_2$.

4.5 Ajuste e seleção dos modelos propostos

O nosso interesse ao ajustar todos os modelos propostos na seção anterior é selecionar aquele com melhor performance preditiva. Utilizando o critério da densidade preditiva conjunta, detalhado na Subseção 3.5.2, junto com a Equação (3.19), obtemos as seguintes formas alternativas para escrever os modelos propostos:

$$\text{Modelos I - II:} \quad \mathbf{y}_t | \boldsymbol{\beta}_{t-1}, \boldsymbol{\psi} \sim N(\mathbf{G}_t \boldsymbol{\beta}_{t-1}, \mathbf{V}_t + \mathbf{W}_t)$$

$$\text{Modelos III - VIII:} \quad \mathbf{y}_t | \boldsymbol{\beta}_{t-1}, \boldsymbol{\psi} \sim N(\mathbf{G}_{1t}(\boldsymbol{\beta}_{1,t-1} + \boldsymbol{\beta}_{2,t-1}), \mathbf{V}_t + \mathbf{W}_{1t})$$

Lembrando que cada modelo tem o vetor de parâmetros $\boldsymbol{\phi}$ específico e a matriz \mathbf{V}_t é a mesma para todos os modelos e é igual $\frac{10^5}{4} \text{diag}(n_{t1}^{-1}, \dots, n_{t5}^{-1})$. Para o modelo I a matriz \mathbf{G}_t é igual à matriz identidade e para o modelo II, \mathbf{G}_t inclui a matriz de contaminação. Analogamente, os modelos III e IV têm $\mathbf{G}_{1t} = \mathbf{I}$. Com o objetivo de determinar a importância de permitir inovações correlacionadas espacialmente, foram ajustados os modelos I - VIII sem dependência espacial ($\phi = 0$), isto é, as inovações são assumidas independentes. Todos os algoritmos de estimação dos parâmetros e a densidade preditiva foram implementados em *Ox*, uma linguagem de programação matricial (Doornik, 2002). As simulações foram executadas em um processador AMD Athlon 3400 de 2.4 GHz. Nas próximas seções descreveremos o procedimento de inferência das quantidades desconhecidas e reportamos os resultados de forma separada para os Estados de Rio de Janeiro e Espírito Santo. Os resultados para Rio de Janeiro correspondem à aplicação desenvolvida em Vivar e Ferreira (2007).

4.6 Resultados para Rio de Janeiro

De acordo com o paradigma Bayesiano, os modelos estão completos com a especificação de distribuições a priori para os parâmetros $\boldsymbol{\beta}_0$, τ_i , ϕ_i e $\kappa_i, i = 1, 2$; a suposição usual de prioris independentes é utilizada aqui. A priori para $\boldsymbol{\beta}_0$, o vetor inicial latente, é uma normal multivariada com vetor de médias zero e matriz de precisão diagonal com elementos próximos de 0, que representa informação vaga. Quando a independência espacial é assumida (isto é, $\phi = 0$) a priori para τ_i é uma distribuição gama $Ga(2, 2)$ que produz uma distribuição

condicional completa gama. No caso de dependência espacial, a priori para τ_i e ϕ_i é a priori de referência conjunta proposta por [Ferreira e De Oliveira \(2007\)](#) para processos *CAMGP*. A priori para κ_i é uma distribuição uniforme no intervalo $(0, 1)$.

Para cada modelo, o esquema *MCMC* discutido na Seção 2.5 foi implementado, incluindo o *FIFBS*. Desta forma, utilizando as prioris mencionadas anteriormente e a posteriori conjunta (2.9) derivam-se as condicionais completas dos parâmetros: a atualização de τ_i foi realizada por passos independentes de Gibbs e para ϕ_i e κ_i as atualizações foram com passos de Metropolis para $\log \phi_i$ e para κ_i .

Geramos cadeias de tamanho 50000 com um período de aquecimento (*burn-in*, em inglês) de 25000 para cada um dos modelos e as amostras são guardadas a passos de comprimento 5, para reduzir a autocorrelação entre elas, obtendo finalmente amostras de tamanho 5000. A convergência dos parâmetros foi monitorada graficamente.

Testes preliminares indicavam que, em alguns modelos, o parâmetro da correlação espacial $\phi \rightarrow 0$ inclusive com prioris informativas. Baseado nestes resultados, o modelo V foi ajustado com $\phi_2 = 0$ e o modelo VII com $\phi_1 = 0$. Apresentamos na seguinte subseção os resultados da estimação dos parâmetros para todos os modelos propostos utilizando o conjunto de dados completo, ($T = 20$) e na Subseção 4.6.2 os resultados da seleção do melhor modelo.

4.6.1 Utilizando todos os tempos ($T = 20$)

A Tabela 4.3 mostra as médias e desvios padrão a posteriori para os modelos propostos com dependência espacial. Para os modelos I, II e VIII o tempo computacional oscilou entre 40-80 minutos e para os modelos III-VII ficou, em média, em 400 minutos para executar as 50000 iterações.

O primeiro fato importante a destacar é quando comparamos os modelos I e II: o índice de contaminação κ do modelo II resulta numa magnitude pequena, mas a inclusão deste parâmetro produz um aumento de mais de 20 vezes no parâmetro ϕ (que mede a correlação espacial) com respeito ao modelo I. Outro resultado interessante é que as médias a posteriori de ϕ_1 e τ_1 dos Modelos III e IV são bem parecidas. Já quando incluímos o ϕ_2 na estimação,

a precisão τ_2 cresce bastante (mais de 20 vezes).

Tabela 4.3: RJ - Modelos com dependência espacial: Médias e desvios padrão a posteriori dos parâmetros e $T = 20$.

Parâmetro	Modelo I	Modelo II	Modelo III	Modelo IV
ϕ (ϕ_1)	0.16 (0.06)	3.71 (1.45)	2.54 (1.39)	2.78 (1.02)
τ (τ_1)	0.39 (0.05)	0.37 (0.11)	0.53 (0.17)	0.40 (0.10)
ϕ_2	-	-	-	0.02 (0.01)
τ_2	-	-	40.7 (5.62)	898 (360)
κ (κ_1)	-	0.003 (0.001)	-	-
κ_2	-	-	-	-
Parâmetro	Modelo V	Modelo VI	Modelo VII	Modelo VIII
ϕ (ϕ_1)	12.8 (9.11)	1250 (535)	-	1.12 (0.40)
τ (τ_1)	0.35 (0.13)	0.002 (0.001)	5.89 (1.18)	0.84 (0.19)
ϕ_2	-	0.018 (0.008)	23.5 (18.1)	-
τ_2	11.0 (2.44)	0.229 (0.057)	0.34 (0.17)	315.8 (12.4)
κ (κ_1)	-	0.720 (0.032)	0.004 (0.001)	-
κ_2	0.81 (0.11)	-	0.750 (0.123)	-

Os modelos de segunda ordem que incluem contaminação em uma ou nas duas equações de sistema (V, VI e VII) exibem resultados diferenciados. No modelo V, as precisões τ_1 e τ_2 indicam moderada variabilidade entre as inovações do nível e pouca variabilidade nas inovações da velocidade. Isto é confirmado pelo alto índice de contaminação κ_2 que mede a influência da velocidade dos municípios nos seus vizinhos. O modelo VI tem uma dependência espacial no nível muito alta e muito baixa na velocidade. As precisões são baixas, revelando principalmente uma alta variabilidade nas inovações do nível. O parâmetro κ_2 é alto, similar ao κ_1 do modelo V. O modelo VII é o único no qual a dependência espacial entre as inovações é forte. A contaminação resulta significativa na velocidade e pequena no nível e as precisões mostram comportamentos diferentes: pouca variabilidade no nível e moderada na velocidade.

O parâmetro de precisão τ_2 do modelo VIII tem um alto valor, indicando que a velocidade a cada tempo varia muito pouco com respeito ao seu valor médio. Os parâmetros do *CAMGP* da equação do nível são comparáveis aos obtidos nos modelos II, III e IV.

Analogamente, na Tabela 4.4 estão os resumos a posteriori dos modelos sem considerar correlação espacial. O tempo computacional para cada um resultou menor que no caso dos modelos com dependência espacial, pois existem menos parâmetros a serem estimados. Os resultados para τ dos modelos I e II são praticamente idênticos, aparentemente o reduzido valor de κ não influencia na estimação do parâmetro de precisão. Os resultados dos modelos III e IV são iguais, pois sem os parâmetros que medem a dependência espacial, ambos são equivalentes. Nestes modelos a mais alta variabilidade ocorre no nível. De modo similar, o modelo VI só se diferencia do III e do IV porque estimamos o parâmetro de contaminação κ_1 e este valor não modifica a estimação de τ_1 e τ_2 . A inclusão da contaminação na equação de velocidade através de κ_2 nos modelos V e VII revela estimativas semelhantes para os parâmetros de precisão, outra vez com um pequeno valor de κ_1 no caso do modelo VII. Este valor resultou consideravelmente alto. No modelo VIII, o valor estimado de τ_2 não foi diferente ao observado no mesmo modelo considerando dependência espacial. A diferença notável ocorre na estimativa da precisão τ_1 que foi quatro vezes maior, isto é, a variabilidade das inovações do nível ficou quatro vezes menor.

4.6.2 Seleção de modelos

Para determinar qual o melhor modelo entre os oito propostos utilizamos o critério da densidade preditiva (Subseção 3.5.2) como medida da performance preditiva de cada modelo. O modelo selecionado é aquele com a mais alta densidade preditiva. Note-se da Equação (3.20) que o valor de t^* deve ser determinado. Realizamos testes considerando $t^* = 5$, mas o processo de estimação teve uma melhor performance quando consideramos os primeiros sete tempos ($t^* = 7$), pois todos os modelos conseguiram aprender melhor sobre os parâmetros e a convergência foi atingida para todos os parâmetros de todos os modelos. Os logaritmos das densidades preditivas para todos os modelos estão na Tabela 4.5.

Tabela 4.4: RJ - Modelos sem dependência espacial: Médias e desvios padrão a posteriori dos parâmetros e $T = 20$.

Parâmetro	Modelo I	Modelo II	Modelo III	Modelo IV
τ (τ_1)	1.64 (0.12)	1.62 (0.12)	2.45 (0.31)	2.45 (0.31)
τ_2	-	-	13.4 (1.71)	13.4 (1.71)
κ (κ_1)	-	0.0006 (0.0005)	-	-
κ_2	-	-	-	-
Parâmetro	Modelo V	Modelo VI	Modelo VII	Modelo VIII
τ (τ_1)	4.82 (0.76)	2.47 (0.32)	4.76 (0.80)	3.12 (0.32)
τ_2	4.89 (0.74)	13.4 (1.70)	4.73 (0.70)	315.0 (12.5)
κ (κ_1)	-	0.0006 (0.0005)	0.0008 (0.0005)	-
κ_2	0.92 (0.07)	-	0.92 (0.07)	-

A segunda e a terceira colunas desta Tabela mostram

$$\log p_m(\mathbf{y}_8, \dots, \mathbf{y}_{20} | \mathbf{D}_7) = \sum_{t=8}^{20} \log p_m(\mathbf{y}_t | \mathbf{D}_{t-1}), \quad (4.12)$$

e o maior valor obtido foi com o Modelo II, o modelo de contaminação. Outro resultado destacado da Tabela 4.5 é verificar que os modelos com dependência espacial conseguiram densidades preditivas mais altas que os seus respectivos modelos que não consideram a dependência espacial. Concluimos que os parâmetros ϕ influenciam positivamente a densidade preditiva dos modelos e, portanto, é importante considerar a estrutura espacial das inovações. A Figura 4.3 representa de forma gráfica o logaritmo das densidades preditivas um passo a frente para cada tempo, para ambos conjuntos de modelos (com e sem considerar dependência espacial). Nota-se claramente que o modelo II é o melhor em quase todos os tempos para os dois conjuntos. Outra conclusão obtida do gráfico da densidade preditiva dos modelos sem dependência espacial é que, concordando com os resultados da estimação dos parâmetros considerando todos os tempos observados, os modelos III(IV) e VI têm a densidade preditiva muito parecida, assim como os modelos V e VII. Adicionalmente, na Figura 4.4 compara-se a performance somente do modelo II com e sem correlação espacial,

Tabela 4.5: RJ - Seleção de modelos: Logaritmo da densidade preditiva.

Modelo	Com dependência	Sem dependência
	espacial	espacial
I	-1521.2	-1548.3
II	-1370.0	-1545.9
III	-1590.2	-2089.5
IV	-1415.4	-2089.5
V	-1454.6	-1832.7
VI	-1665.4	-2085.8
VII	-1427.4	-1831.6
VIII	-1539.3	-1573.2

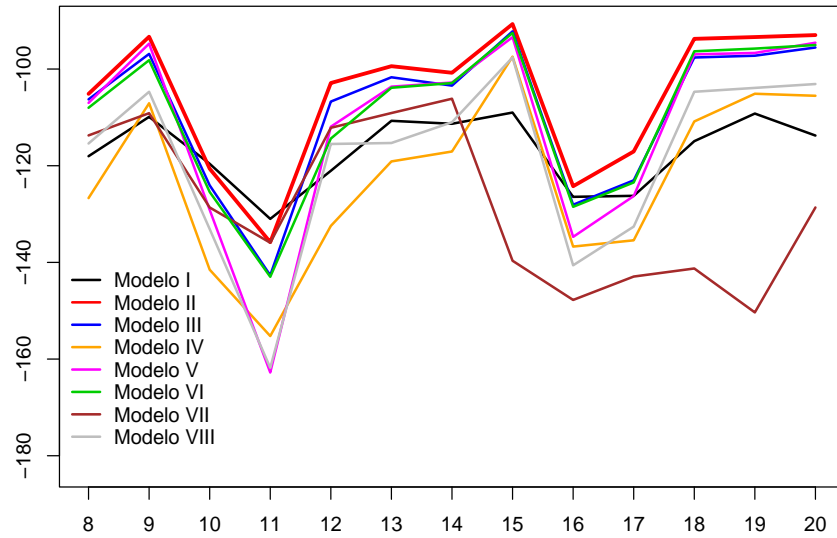
Tabela 4.6: RJ - Medidas resumo a posteriori do modelo II com dependência espacial.

Parâmetro	Média	Desvio	2.5%	50%	97.5%
ϕ	3.71	1.45	1.76	3.42	7.21
τ	0.37	0.11	0.19	0.36	0.59
κ	0.003	0.001	0.001	0.003	0.005

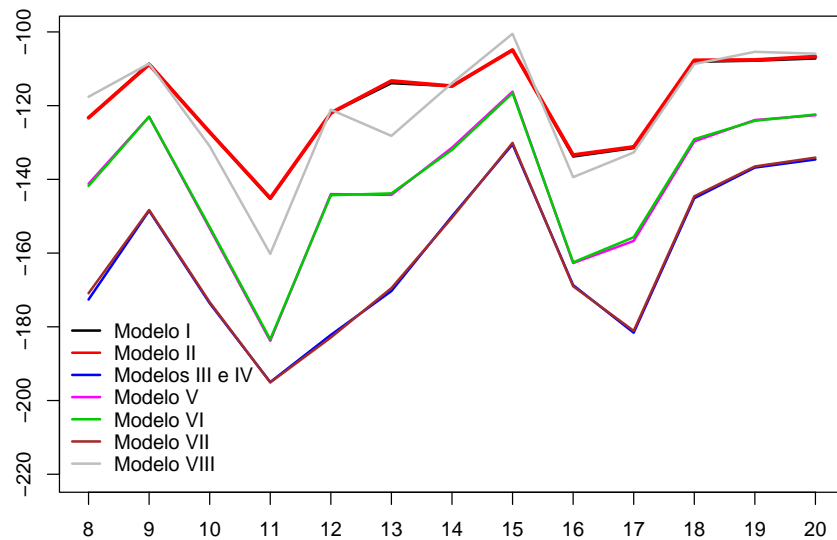
resultando evidente que o modelo com inovações espacialmente dependentes vence em todos os tempos.

O critério selecionou como o melhor entre os modelos propostos o modelo de contaminação com dependência espacial entre as inovações. A Tabela 4.6 mostra os resultados da inferência a posteriori incluindo os quantis para elaborar intervalos de credibilidade de 95%. A convergência foi determinada pela observação do traço de cada cadeia. A Figura 4.5 mostra as marginais a posteriori dos parâmetros deste modelo.

Os resultados da estimação do vetor latente obtido pelo *FIFBS* para todo o período de estudo encontram-se ilustrados nas Figuras 4.6 e 4.7. Estes mapas representam a média a posteriori do nível do risco em uma escala do branco (menor risco) ao vermelho obscuro



(a)



(b)

Figura 4.3: RJ - Logaritmo da densidade preditiva dos modelos. (a) Com dependência espacial e (b) Sem dependência espacial.

(maior risco). Nos primeiros anos, o nível do risco aumenta na área metropolitana da cidade do Rio de Janeiro (região centro-sul), capital do Estado. Ao longo dos anos 90, essa violência disseminou-se nos outros municípios, permanecendo o nível do risco em um alto patamar. Em muito poucos municípios acontece uma redução no nível do risco nos anos de estudo.

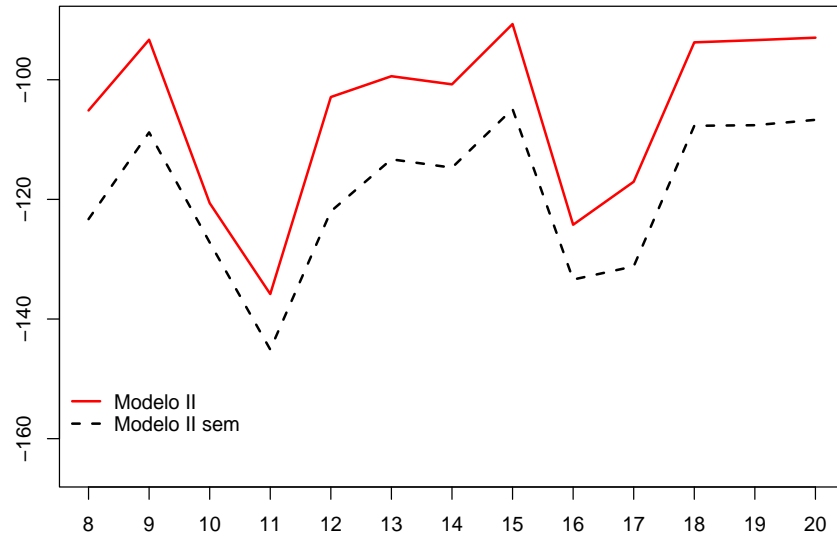


Figura 4.4: RJ - Modelo II: Logaritmo da densidade preditiva com e sem dependência espacial.

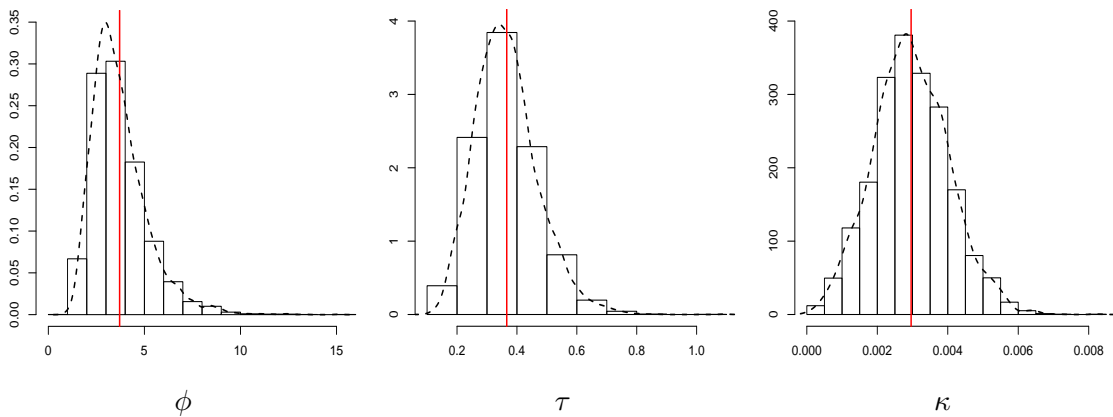


Figura 4.5: RJ - Modelo II: Histogramas das amostras a posteriori dos parâmetros. Em cada painel, a linha vermelha representa a média a posteriori e a linha tracejada representa a densidade marginal.

Os mapas das Figuras 4.8 e 4.9 representam a médias a posteriori das inovações em uma escala de cores que representam uma alta redução no risco (azul obscuro) até um alto incremento no risco (vermelho obscuro). A cor branca indica uma certa estabilidade de um ano para outro, com inovações muito pequenas tanto positivas como negativas. Estes mapas

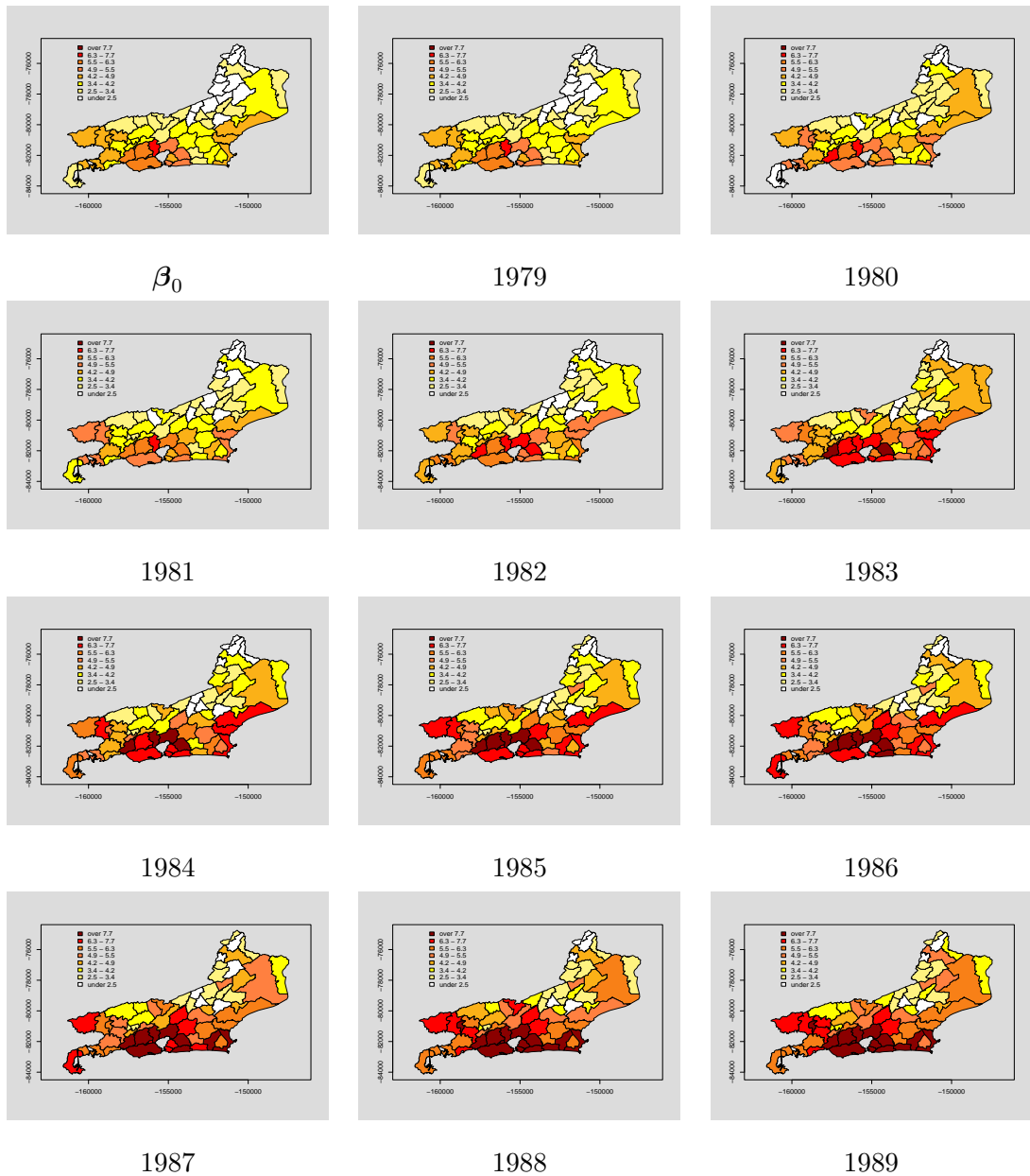


Figura 4.6: RJ - Modelo II: Mapas representando o processo latente estimado. Estado inicial β_0 e anos 1979-1989.

das inovações são muito informativos, pois eles representam estimativas dos efeitos espacialmente estruturados específicos para cada ano do estudo. Por exemplo, depois de tomar conta do efeito da contaminação, no ano 1983 praticamente teve aumento em todos os municípios.

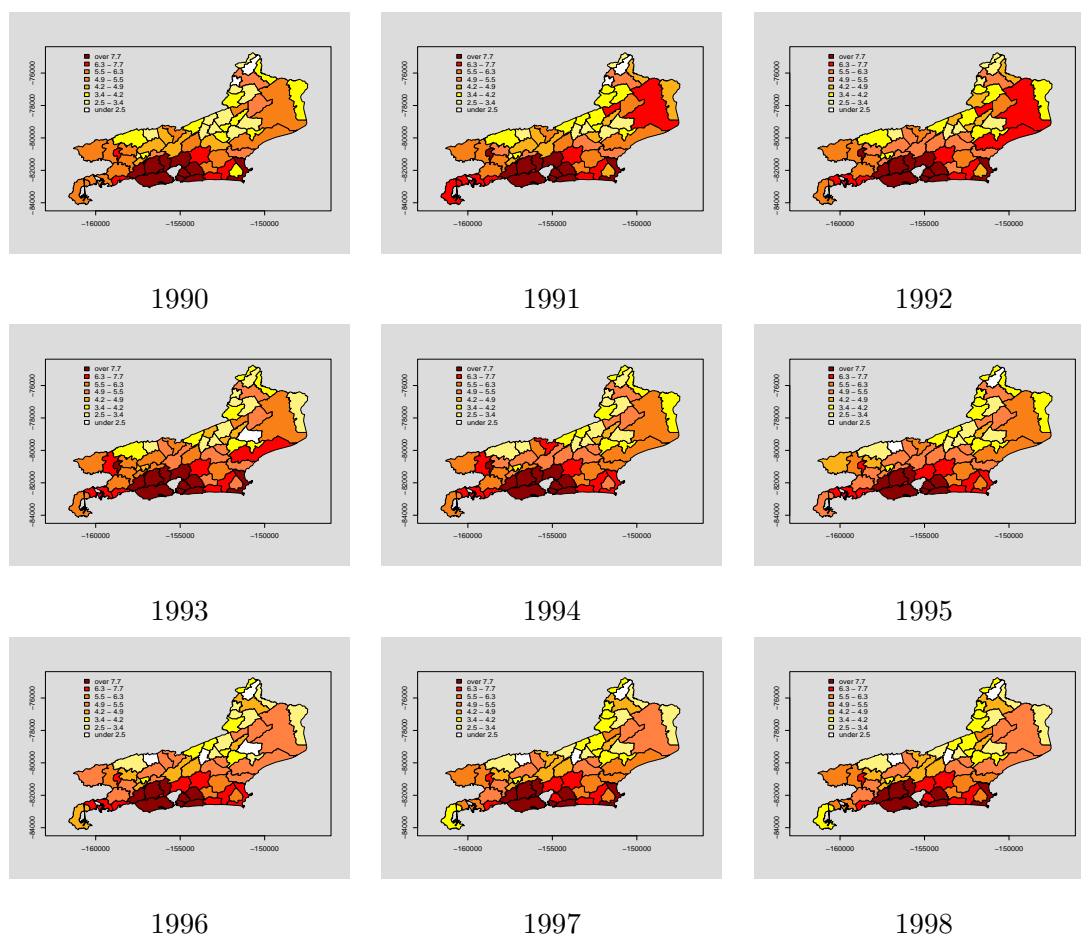


Figura 4.7: RJ - Modelo II: Mapas representando o processo latente estimado. Anos 1990-1998.

Outro exemplo, em 1989 houve um aumento específico do nível do risco nas regiões oeste e nordeste e uma redução na região central do Estado do Rio de Janeiro. A partir deste ano, observa-se que as inovações para o município do Rio de Janeiro ficam estáveis, com um par de anos de ligeira redução. Este padrão é diferente para outros municípios, que em alguns anos formam grupos com a mesma tendência. Estes agrupamentos espaciais podem ser o resultado de ações combinadas de municípios vizinhos contra a violência, no caso de redução do risco, ou ações intermunicipais de grupos fora-da-lei quando ocorre incremento do risco. Estes mapas podem ser utilizados pelas autoridades municipais ou estaduais para estabelecer prioridades ou avaliar o impacto de políticas de segurança.

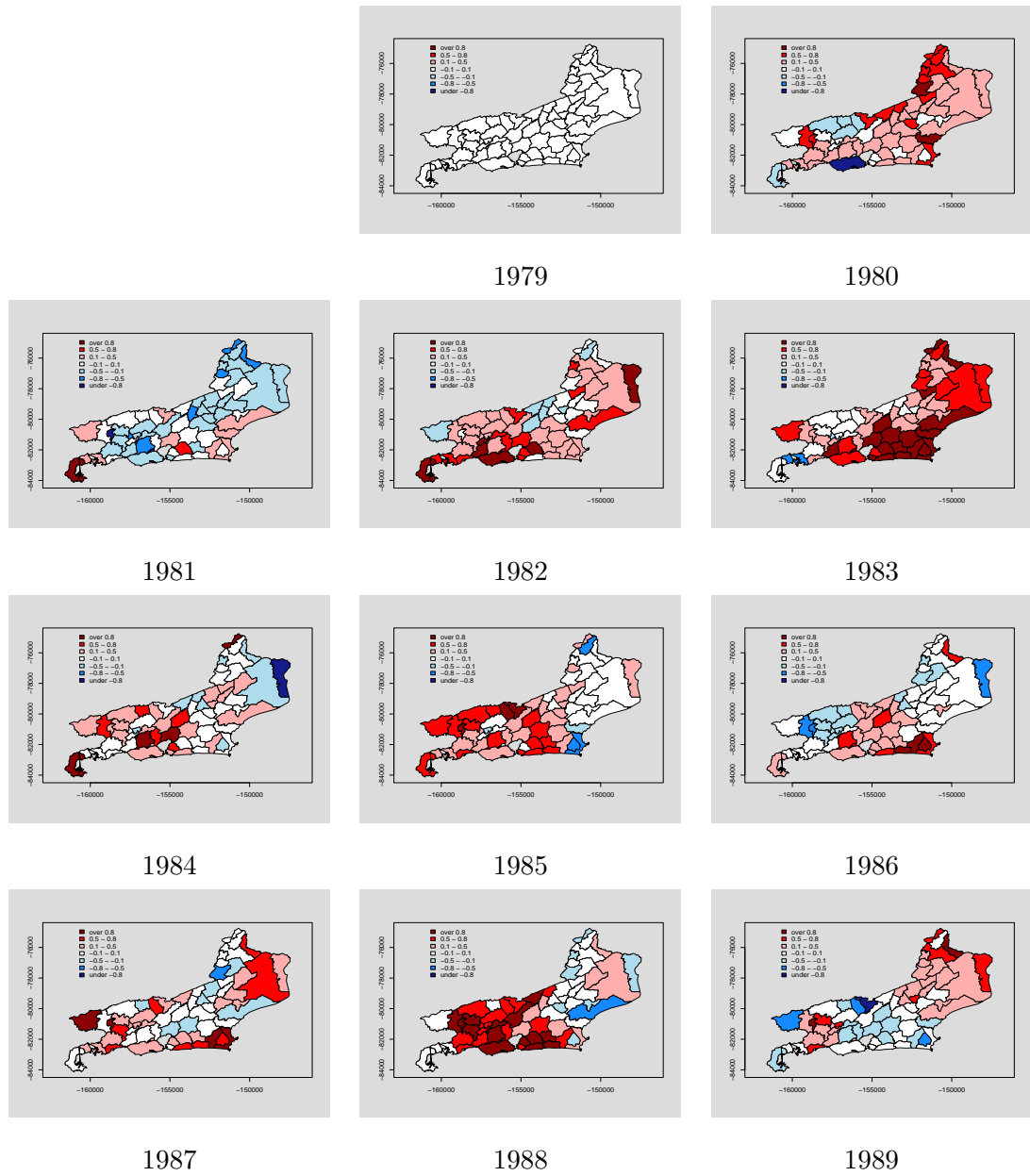


Figura 4.8: RJ - Modelo II: Mapas representando as estimativas das médias a posteriori das inovações. Anos 1979-1989.

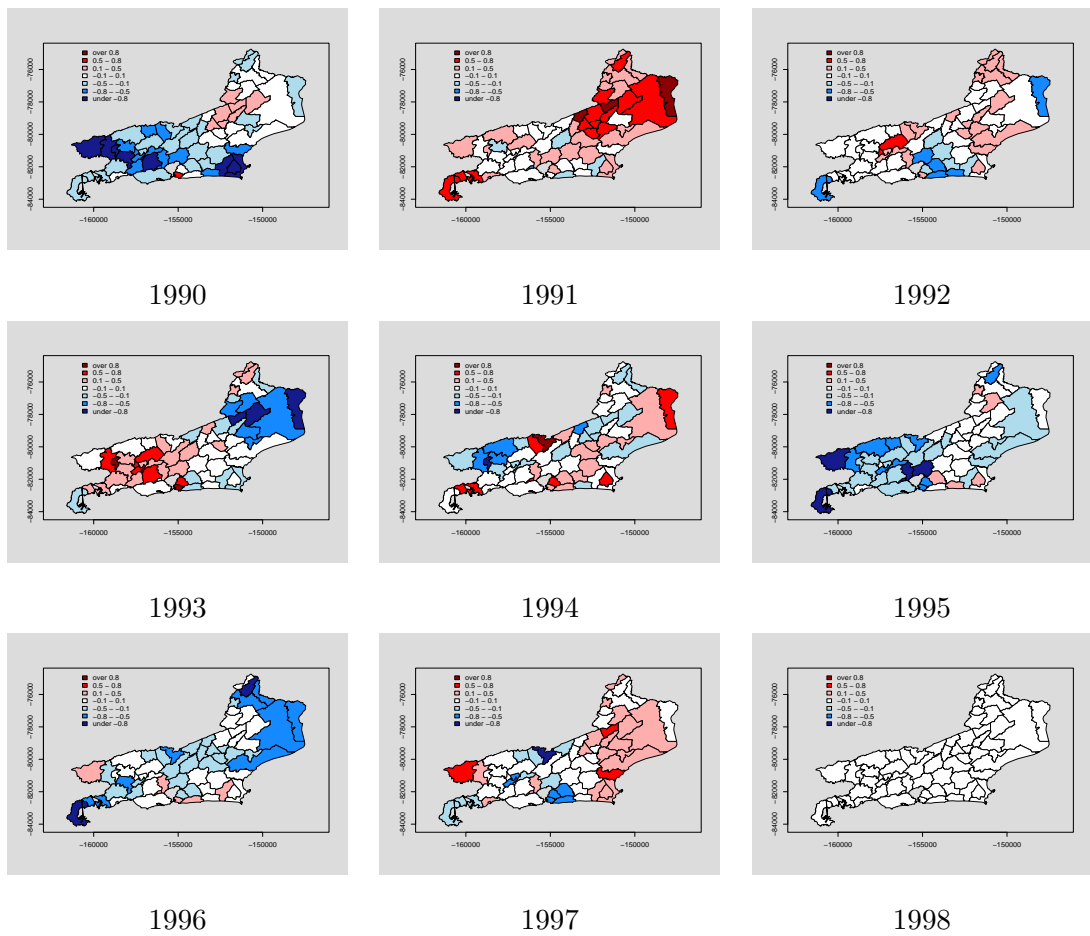


Figura 4.9: RJ - Modelo II: Mapas representando as estimativas das médias a posteriori das inovações. Anos 1990-1998.

4.7 Resultados para Espírito Santo

Para os dados de homicídio do Espírito Santo foram feitas algumas mudanças no processo de estimação. Quando os modelos mais complexos (V, VI e VII) foram ajustados, os parâmetros do *CAMGP*, τ_i e ϕ_i , $i = 1, 2$ não convergiam, então foi necessário colocar um pouco mais de informação nas distribuições a priori destes parâmetros. Assim, temos que a priori de τ_i é gama $Ga(3, 3)$ e

$$\phi_i \propto \begin{cases} 1, & 0 < \phi_i < 1 \\ \phi_i^{-b}, & \phi_i \geq 1 \end{cases}$$

com o expoente $b = \{3, 4, 5\}$ dependendo do modelo. Continua válida a suposição de prioris independentes, isto é $p(\boldsymbol{\beta}_0, \tau_i, \phi_i, \kappa_i) = p(\boldsymbol{\beta}_0)p(\tau_i)p(\phi_i)p(\kappa_i)$, $i = 1, 2$. As distribuições a priori para o vetor latente inicial $\boldsymbol{\beta}_0$ e κ_i são mantidas: normal multivariada com vetor de médias zero e matriz de precisão diagonal com elementos próximos de 0 e uniforme no intervalo $(0, 1)$, respectivamente.

De novo, para cada modelo, seja estudando os dados completos ou a seleção de modelos, a inferência é realizada amostrando das distribuições condicionais completas de cada elemento do vetor de parâmetros $\boldsymbol{\psi}$, derivadas da posteriori conjunta (2.9), e o vetor latente $\boldsymbol{\beta}_{(1:T)}$ é amostrado utilizando o algoritmo *FIFBS*. O esquema *MCMC* inclui o amostrador de Gibbs para τ_i , $i = 1, 2$ cuja condicional completa tem forma conhecida (gama); e passos de Metropolis-Hastings para $\log \phi_i$ e κ_i , cujas condicionais completas não têm forma fechada.

Geramos duas cadeias independentes partindo de pontos iniciais dispersos, de tamanho 90000 e com um *burn-in* de 10000 para cada um dos modelos. As amostras são guardadas a passos de comprimento 20, para reduzir a autocorrelação, obtendo amostras finais de tamanho 4000. Um dos métodos recomendados para monitorar a convergência está baseado em detectar quando as cadeias de Markov “esquecem” os seus valores iniciais, comparando algumas seqüências extraídas de diferentes pontos iniciais e verificando que elas são indistinguíveis. Este método é conhecido também como “critério visual de convergência”.

Uma abordagem mais quantitativa está baseada na análise de variância: diagnostica-se convergência aproximada quando a variância entre as diferentes seqüências não é maior do

que a variância entre cada seqüência individual (Gelman, 1996). A estatística de Gelman e Rubin, o \hat{R} é a razão das variâncias estimadas (Gelman e Rubin, 1992). Quando a simulação converge, o \hat{R} decai para 1, isto significa que as cadeias de Markov paralelas são essencialmente indistinguíveis. Se o \hat{R} é alto, então existem razões para acreditar que continuando com mais simulações podemos melhorar a nossa inferência sobre a distribuição objetivo. Na prática, as simulações são executadas até que os valores de \hat{R} são todos menores que 1.1 ou 1.2. Para verificar a convergência das cadeias utilizamos a biblioteca *CODA* (Plummer et al., 2006) do software *R* (R Development Core Team, 2005). O *CODA* é um conjunto de análises e diagnósticos de convergência baseados nas amostras a posteriori obtidas via *MCMC*. Os critérios mais populares são o de Raftery e Lewis (Raftery e Lewis, 1996) e a estatística \hat{R} de Gelman e Rubin (Gelman e Rubin, 1992).

4.7.1 Utilizando todos os tempos ($T = 20$)

A Tabela 4.7 mostra as médias e desvios padrão a posteriori para os modelos propostos considerando dependência espacial. Para os modelos I, II e VIII o tempo computacional variou entre 70-110 minutos e para os modelos III-VII ficou em 440 minutos, aproximadamente. Esta tabela também mostra a estatística de Gelman e Rubin o \hat{R} para cada parâmetro dos modelos ajustados. De acordo com este critério, todas as cadeias atingiram a convergência.

Para todos os modelos, a precisão das inovações do nível τ_1 tem um valor baixo e isto implica que a variabilidade é alta, comparada a magnitude das inovações. Da mesma forma, a correlação espacial entre as inovações do nível ϕ_1 é forte, obtendo o maior valor para o modelo polinomial de primeira ordem. No modelo II, ao incluirmos o índice de contaminação, a precisão τ_1 diminuiu, se comparada com o modelo I, e esta contaminação não é muito alta. Nos modelos polinomiais de segunda ordem, a velocidade do processo apresenta baixa variabilidade quando ϕ_2 não é estimado (modelo III) e o contrário acontece quando estimamos ϕ_2 (modelo IV). Inclusive, este modelo evidencia uma forte dependência espacial.

Os modelos polinomiais de segunda ordem que incluem contaminação têm as estimativas do parâmetro de contaminação parecidas com as estimativas do modelo VII: altos valores de κ_2 nos modelos V e VII e baixos valores de κ_1 nos modelos VI e VII. Destaca no modelo VI

Tabela 4.7: ES - Modelos com dependência espacial: Médias e desvios padrão a posteriori dos parâmetros e $T = 20$.

Parâmetro	I	\hat{R}	II	\hat{R}	III	\hat{R}	IV	\hat{R}
ϕ (ϕ_1)	5.15 (2.04)	1.0	4.20 (1.51)	1.0	3.22 (2.20)	1.0	1.48 (0.77)	1.0
τ (τ_1)	0.26 (0.07)	1.0	0.24 (0.06)	1.0	0.38 (0.14)	1.0	0.52 (0.17)	1.0
ϕ_2	-	-	-	-	-	-	34.1 (28.2)	1.0
τ_2	-	-	-	-	11.2 (1.58)	1.0	2.49 (0.93)	1.0
κ (κ_1)	-	-	0.01 (0.002)	1.0	-	-	-	-
κ_2	-	-	-	-	-	-	-	-
Parâmetro	V	\hat{R}	VI	\hat{R}	VII	\hat{R}	VIII	\hat{R}
ϕ (ϕ_1)	3.73 (3.18)	1.1	1.58 (0.81)	1.0	2.51 (1.73)	1.0	2.70 (1.44)	1.0
τ (τ_1)	0.52 (0.25)	1.0	0.51 (0.17)	1.0	0.35 (0.21)	1.0	0.41 (0.15)	1.0
ϕ_2	3.22 (4.39)	1.2	10.9 (7.22)	1.1	0.94 (0.42)	1.0	-	-
τ_2	1.78 (0.74)	1.0	3.43 (1.06)	1.0	1.56 (0.59)	1.0	256 (11.2)	1.0
κ (κ_1)	-	-	0.01 (0.002)	1.0	0.02 (0.01)	1.0	-	-
κ_2	0.77 (0.13)	1.0	-	-	0.78 (0.13)	1.0	-	-

a forte dependência espacial expressada por ϕ_2 . No modelo VIII, o parâmetro de precisão do vetor de inovações da velocidade τ_2 tem um alto valor estimado, como no caso do Rio de Janeiro, implicando que a velocidade a cada tempo varia muito pouco com respeito ao valor médio.

Na Tabela 4.8 estão os resumos a posteriori dos modelos sem considerar correlação espacial nas inovações e os respectivos valores da estatística de Gelman e Rubin (\hat{R}). Verifica-se também que as cadeias convergiram. Os modelos I e II apresentam as estimativas a posteriori de τ muito parecidas, e o índice de contaminação κ do modelo II é pequeno, similar ao caso quando modelamos com estrutura espacial. Os modelos III e IV são equivalentes quando não estimamos os parâmetros de dependência espacial. As estimativas do parâmetro de contaminação κ_2 dos modelos V e VII são semelhantes, assim como κ_1 dos modelos VI e VII. Os parâmetros de precisão dos modelos V e VII implicam uma variabilidade moderada,

de acordo com a magnitude das inovações. O modelo VI apresenta uma variabilidade maior nas inovações do nível e uma pequena variabilidade nas inovações da velocidade.

Tabela 4.8: ES - Modelos sem dependência espacial: Médias e desvios padrão a posteriori dos parâmetros e $T = 20$.

Parâm.	I	\hat{R}	II	\hat{R}	III	\hat{R}	IV	\hat{R}
τ (τ_1)	1.50 (0.13)	1.0	1.40 (0.13)	1.0	2.12 (0.31)	1.0	2.12 (0.31)	1.0
τ_2	-	-	-	-	11.0 (1.55)	1.0	11.0 (1.55)	1.0
κ (κ_1)	-	-	0.003 (0.001)	1.0	-	-	-	-
κ_2	-	-	-	-	-	-	-	-
Parâm.	V	\hat{R}	VI	\hat{R}	VII	\hat{R}	VIII	\hat{R}
τ (τ_1)	4.35 (0.92)	1.0	2.20 (0.33)	1.0	3.93 (0.75)	1.0	2.62 (0.32)	1.0
τ_2	4.70 (0.91)	1.0	13.8 (2.11)	1.0	3.89 (0.69)	1.0	254 (11.1)	1.0
κ (κ_1)	-	-	0.003 (0.001)	1.0	0.004 (0.001)	1.0	-	-
κ_2	0.84 (0.11)	1.0	-	-	0.85 (0.11)	1.0	-	-

4.7.2 Seleção de modelos

Analogamente ao caso do Rio de Janeiro, utilizamos o critério da densidade preditiva como medida para selecionar o melhor modelo entre todos os ajustados. Determinamos que com $t^* = 10$ pontos no tempo todos os modelos produzem distribuições a posteriori próprias. Os resultados estão na Tabela 4.9.

O modelo de contaminação (II) resulta com a melhor performance preditiva, como também ocorreu no caso dos dados do Rio de Janeiro. Na Figura 4.10 mostra-se o valor do logaritmo da densidade preditiva um passo a frente. Este é um recurso visual que permite observar o comportamento preditivo dos modelos em cada ponto do tempo. A Figura 4.11 mostra a comparação entre os modelos de contaminação com e sem dependência espacial, verificando-se que o claro vencedor em todos os tempos é o modelo que considera a estrutura espacial nas inovações. Portanto, pode-se afirmar que a modelagem das inovações com

Tabela 4.9: ES - Seleção de modelos: Logaritmo da densidade preditiva.

Modelo	Com dependência	Sem dependência
	espacial	espacial
I	-1002.8	-1044.3
II	-985.3	-1037.9
III	-1433.3	-1382.0
IV	-1193.8	-1382.0
V	-1100.5	-1193.3
VI	-1193.6	-1347.8
VII	-1067.3	-1185.1
VIII	-1032.2	-1085.0

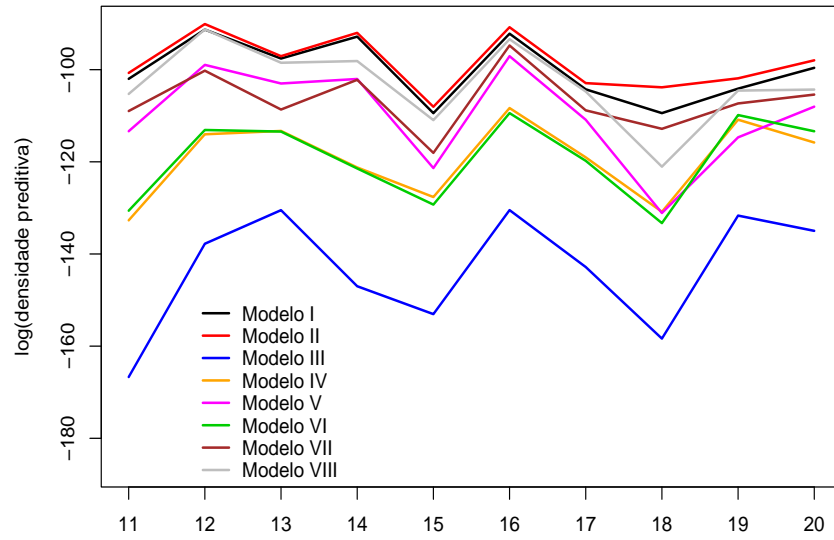
Tabela 4.10: ES - Medidas resumo a posteriori do modelo II com dependência espacial.

Parâmetro	Média	Desvio	2.5%	50%	97.5%	\hat{R}
ϕ	4.20	1.51	2.09	3.93	7.94	1.0
τ	0.24	0.06	0.13	0.23	0.38	1.0
κ	0.005	0.001	0.002	0.005	0.009	1.0

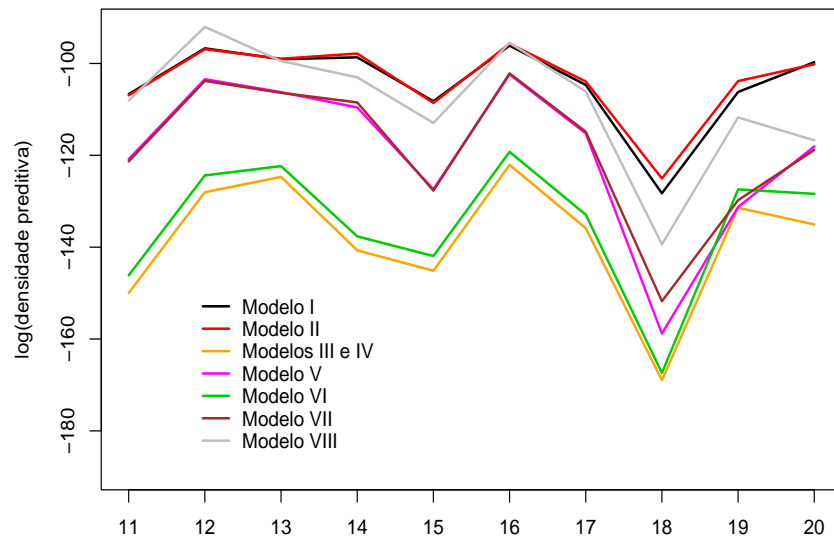
dependência espacial é relevante para este conjunto de dados. É interessante notar o fato que todos os modelos tiveram uma queda na performance preditiva no tempo $t = 18$, isto é, quando utilizamos os primeiros $t^* = 17$ anos para estimar os parâmetros e o processo latente.

A Tabela 4.10 mostra algumas medidas resumo dos parâmetros do modelo de contaminação com dependência espacial: médias e desvios padrão a posteriori e quantis das distribuições a posteriori para construir intervalos de credibilidade de 95%. Os histogramas das amostras a posteriori dos estão na Figura 4.12.

Os mapas das Figuras 4.13 e 4.14 mostram a média a posteriori do nível do risco para todos os anos, estimado utilizando o algoritmo *FIFBS*. A escala de cores começa em branco



(a)



(b)

Figura 4.10: ES - Logaritmo da densidade preditiva dos modelos. (a) Com dependência espacial e (b) Sem dependência espacial.

(menor nível do risco) e termina em vermelho obscuro (maior nível do risco). Nestes mapas podemos observar que os municípios do Espírito Santo têm comportamentos diferenciados nos primeiros anos de estudo. A região sul apresenta a maior quantidade de municípios com os menores níveis do risco ao longo do tempo, até o ano 1995. Nos últimos três anos

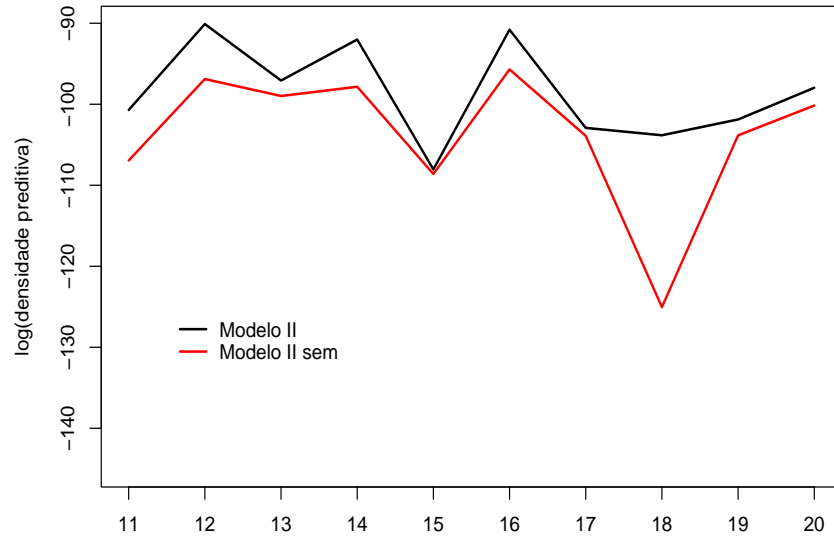


Figura 4.11: ES - Modelo II: Logaritmo da densidade preditiva com e sem dependência espacial.

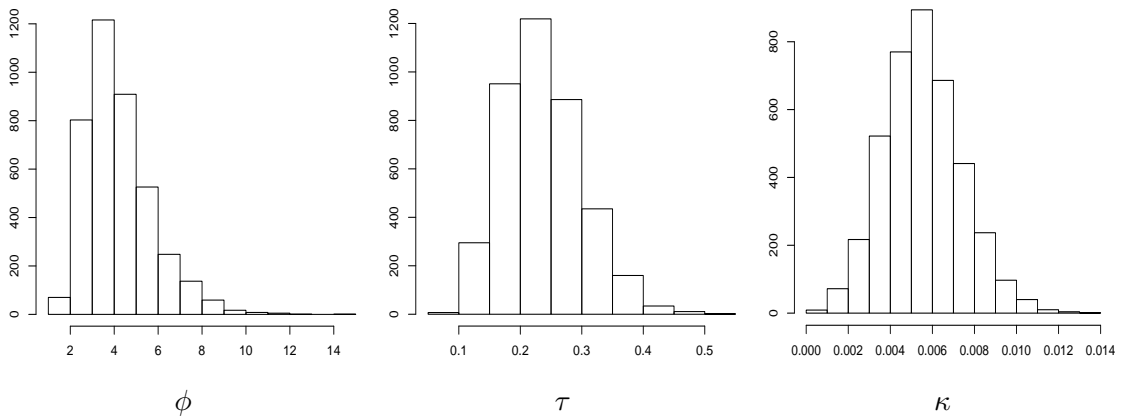


Figura 4.12: ES - Modelo II: Histogramas das amostras a posteriori dos parâmetros.

(1996-1998), a maior parte dos municípios têm baixos níveis de risco, mas a região central do Estado mantém os níveis altos apresentados ao longo dos anos. A cidade de Vitória e os municípios vizinhos mantém ano após ano o nível do risco crescente.

As Figuras 4.15 e 4.16 ilustram as inovações da equação de sistema do modelo selecionado, o modelo de contaminação com dependência espacial. Estes mapas mostram os efeitos espaciais específicos para cada ano de estudo, em uma escala de cores do vermelho obscuro (valores positivos) até o azul obscuro (valores negativos). A cor branca representa

o que podemos denominar de estabilidade, pois os valores positivos ou negativos, são muito pequenos. Podemos encontrar resultados interessantes, por exemplo, depois de subtrair o efeito da contaminação, no ano 1982 quase todos os municípios tiveram um impacto negativo ou estável, mas no ano seguinte tiveram um efeito positivo. No ano 1986, os municípios das regiões norte e sul e alguns da parte central tiveram um forte efeito positivo, deixando um grupo de municípios vizinhos na região central incluindo a cidade de Vitória, com efeitos negativos ou estáveis. A partir do ano 1994 os municípios tinham mais efeitos negativos que positivos, indicando uma estabilidade ou uma redução nos níveis do risco. Mas no ano 1998 os níveis em quase todo o Estado tiveram um incremento. Estas sucessivas variações podem ter explicações políticas, sociais, policiais, populacionais ou de outro tipo. Não faz parte dos objetivos desta tese apresentar as explicações políticas, sociais, entre outras de tais variações, mas a inclusão de covariáveis apropriadas pode ajudar a encontrar estas explicações.

Em este capítulo tratamos com a aproximação Gaussiana como uma forma de trabalhar com dados não normais aplicando a nossa classe de modelos espaço-temporais de área Gaussianos. Duas aplicações com o número de homicídios nos Estados do Rio de Janeiro e de Espírito Santo foram desenvolvidas, incluindo estimação de parâmetros e do processo latente e seleção do melhor modelo em um conjunto de modelos propostos. A metodologia proposta produziu bons resultados com estes dados aproximados. No capítulo seguinte estudaremos a extensão desta classe de modelos para dados na família exponencial.

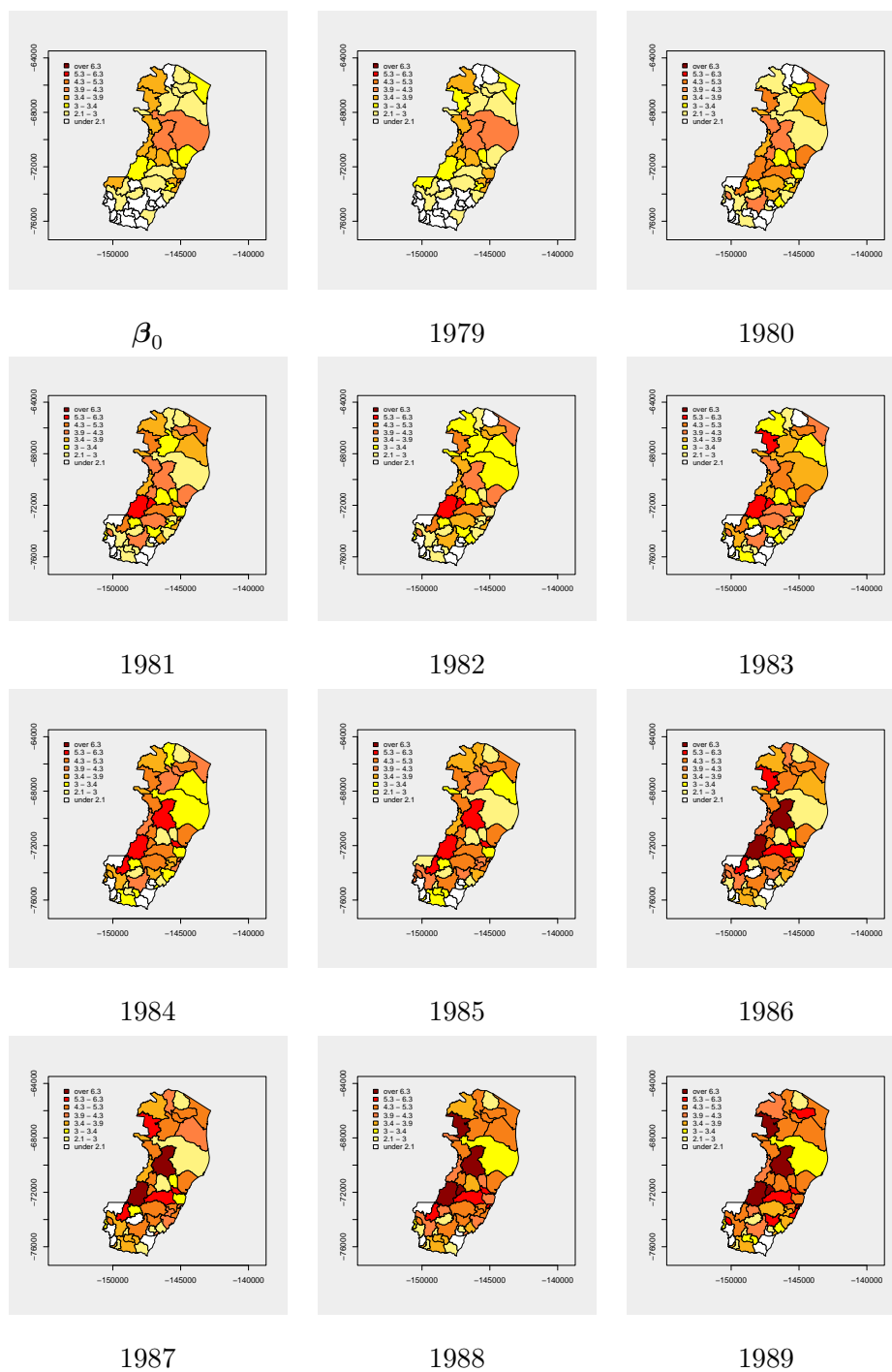


Figura 4.13: ES - Modelo II: Mapas representando o processo latente estimado. Estado inicial β_0 e anos 1979-1989.

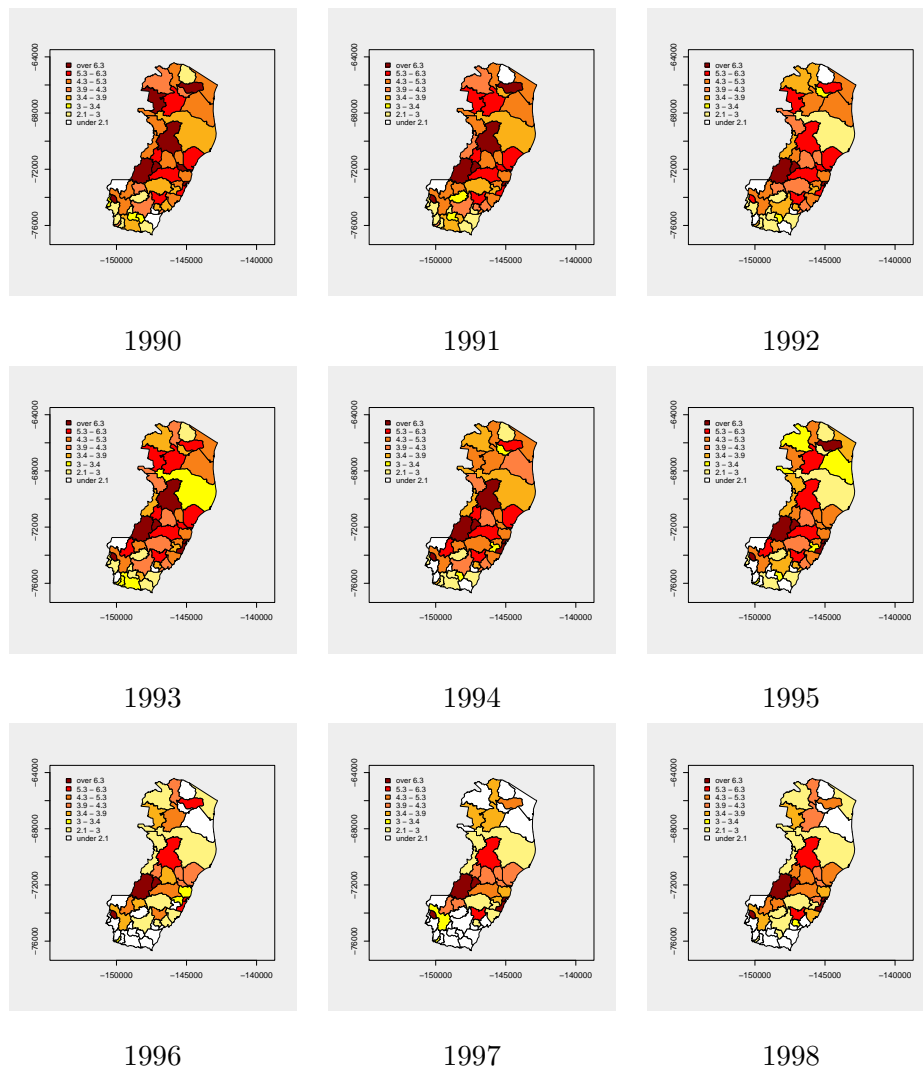


Figura 4.14: ES - Modelo II: Mapas representando o processo latente estimado. Anos 1990-1998.

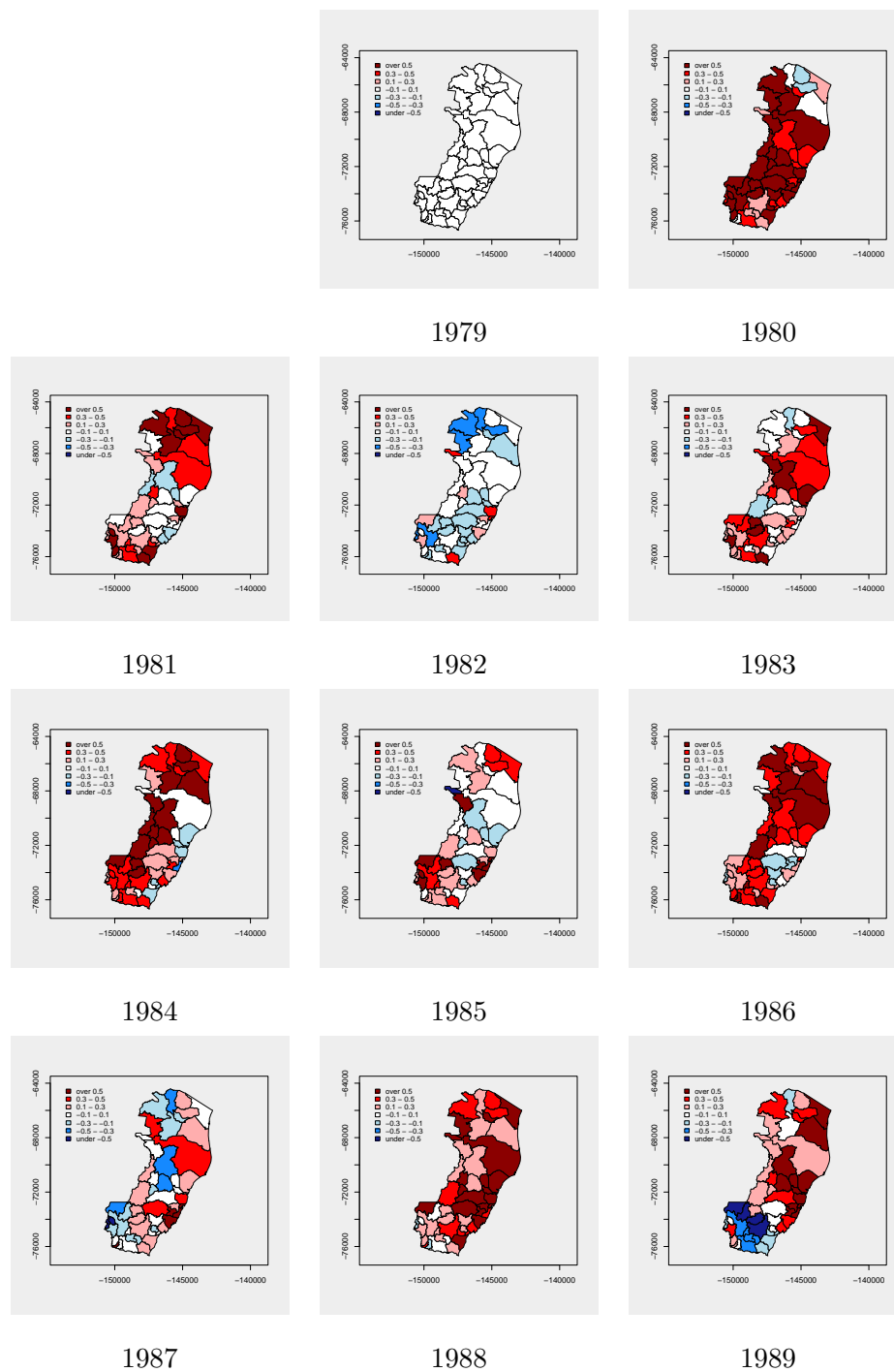


Figura 4.15: ES - Modelo II: Mapas representando as estimativas das médias a posteriori das inovações. Anos 1979-1989.

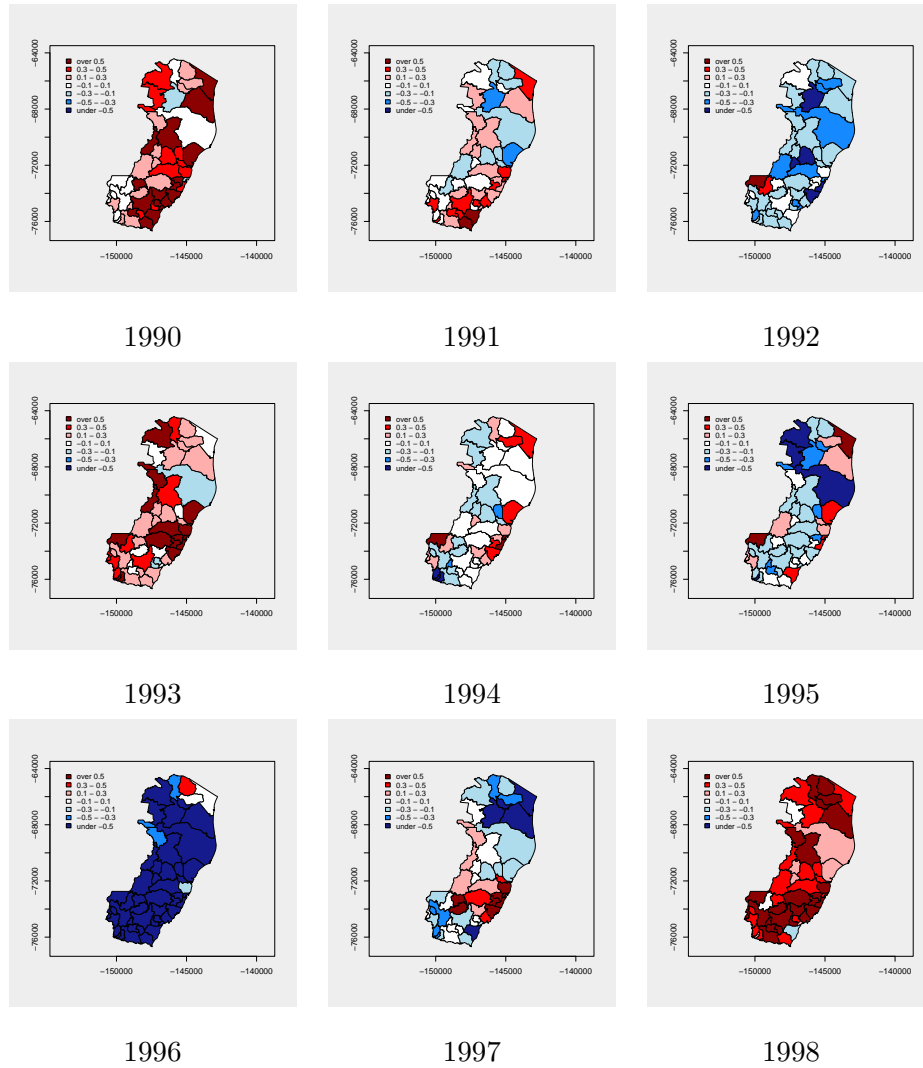


Figura 4.16: RJ - Modelo II: Mapas representando as estimativas das médias a posteriori das inovações. Anos 1990-1998.

Capítulo 5

Generalização da classe de modelos para dados de área na família exponencial

5.1 Introdução

No capítulo anterior estabelecemos que a modelagem de dados espaço-temporais não Gaussianos é possível utilizando nossa classe de modelos, quando as observações são transformadas para que resultem aproximadamente Gaussianas. Em muitas ocasiões estas aproximações não são adequadas, pois os resultados do processo de estimação acabam sendo difíceis de interpretar. Por exemplo, no capítulo anterior, o processo β_t foi chamado de “nível do risco”. Neste capítulo estendemos nossos modelos espaço-temporais para dados de área na família exponencial. Esta classe generalizada de modelos permite uma abordagem direta para estas observações sem precisar de transformações nem assumir normalidade, tendo como resultado uma análise mais natural dos dados e uma melhor interpretação dos resultados. Iniciamos com uma compilação sobre modelos lineares generalizados e detalhes da família exponencial na seguinte seção. A Seção 5.3 refere-se aos métodos Bayesianos para modelos lineares generalizados. A extensão dinâmica dos modelos lineares generalizados está na Seção 5.4. Finalmente, na Seção 5.5 é introduzida nossa classe de modelos lineares generalizados dinâmicos para dados de área, baseada nos modelos tratados na Seção 5.4.

5.2 Conceitos preliminares

5.2.1 Modelos lineares generalizados

Os modelos estatísticos clássicos para regressão, séries temporais e análise de dados longitudinais são geralmente úteis em situações onde os dados são aproximadamente Gaussianos e podem ser explicados por alguma estrutura linear. Estes modelos são fáceis de interpretar e os métodos são teoricamente bem entendidos e estudados. Porém, as suposições subjacentes podem ser muito restritivas e a aplicação dos métodos pode produzir erros em situações nas quais os dados são evidentemente não normais, tais como dados binários ou de contagem. A modelagem estatística fornece ferramentas mais flexíveis para a análise de dados desta natureza, como os modelos lineares generalizados.

Os modelos lineares generalizados foram definidos formalmente por [Nelder e Wedderburn \(1972\)](#) como uma família unificadora de modelos para análise de regressão com respostas não Gaussianas, isto é, dados categóricos, de contagem ou não negativos. Exemplos típicos de dados desta natureza são: registros diários de ocorrência de chuva em uma região, o número de incidências de dengue por semana ou medidas diárias de SO_2 (dióxido sulfúrico). Estes modelos têm um grande impacto tanto na teoria como na prática da Estatística, e dominaram, junto com outros modelos relativamente próximos, a literatura aplicada por duas décadas. Uma das principais fontes de informação sobre os modelos lineares generalizados é [McCullagh e Nelder \(1989\)](#), em que, em particular, o caso univariado é considerado de forma extensa.

Os modelos lineares generalizados são uma extensão dos modelos lineares clássicos. Assim, começamos descrevendo esta última classe de modelos. Assume-se que o vetor de observações \mathbf{y} com n componentes é uma realização da variável aleatória \mathbf{Y} que tem componentes independentemente distribuídos com média $\boldsymbol{\mu}$ e variância constante σ^2 e

$$E(\mathbf{Y}) = \boldsymbol{\mu} = \mathbf{F}'\boldsymbol{\beta}$$

onde \mathbf{F}' e $\boldsymbol{\beta}$ representam respectivamente, a matriz de covariáveis, denotadas por x_1, x_2, \dots, x_q e o vetor de parâmetros desconhecidos, respectivamente.

Para facilitar a transição para o modelo linear generalizado, especifica-se o modelo linear clássico em três partes:

1. *Componente aleatória*: os componentes de \mathbf{Y} tem distribuições normais independentes com $E(\mathbf{Y}) = \boldsymbol{\mu}$ e variância constante σ^2 .
2. *Componente sistemática*: as covariáveis x_1, x_2, \dots, x_q produzem um preditor linear $\boldsymbol{\theta}$ dado por

$$\boldsymbol{\theta} = \mathbf{F}'\boldsymbol{\beta}$$

onde \mathbf{F}' é uma matriz $n \times q$ que é função conhecida de x_1, x_2, \dots, x_q e $\boldsymbol{\beta}$ é um vetor de q parâmetros.

3. *Ligação ou link* entre as componentes aleatória e sistemática:

$$\boldsymbol{\mu} = \boldsymbol{\theta}.$$

Esta terceira componente pode ser escrita como uma função para cada elemento do preditor

$$\theta_i = g(\mu_i),$$

então $g(\cdot)$ será chamada de *função de ligação*. Nesta formulação, os modelos lineares clássicos têm distribuição normal na componente aleatória e a função identidade como função de ligação na terceira componente. Os modelos lineares generalizados permitem duas extensões desta especificação dos modelos clássicos: a primeira, a distribuição na componente aleatória pode pertencer à família exponencial (a distribuição normal forma parte desta família) e a segunda, a função de ligação da terceira componente pode ser qualquer função monótona diferenciável.

Duas funções adicionais são definidas: a inversa da função de ligação $g^{-1}(\cdot) = f(\cdot)$ é chamada de função resposta, assim

$$\mu_i = f(\theta_i)$$

e a função $\delta(\cdot)$, que transforma diretamente o preditor linear θ_i no parâmetro natural ou canônico η_i da família exponencial

$$\eta_i = \delta(\theta_i).$$

5.2.2 Família exponencial

Assume-se que cada componente da variável aleatória \mathbf{Y} tem uma distribuição na família exponencial, se sua densidade discreta ou contínua com respeito a uma medida σ -finita tem a forma

$$p(y|\eta, v) = \exp\{(y\eta - b(\eta))/a(v) + c(y, v)\} \quad (5.1)$$

para certas funções $a(\cdot), b(\cdot)$ e $c(\cdot)$. Se o parâmetro de dispersão $v > 0$ é conhecido, esta é uma família exponencial com parâmetro natural ou canônico η . A função $c(\cdot)$ é maior ou igual a zero e é mensurável. A função $b(\cdot)$ é convexa e duas vezes diferenciável e está relacionada com a média e a variância de y mediante

$$E(y|\eta) = \mu = \frac{\partial b(\eta)}{\partial \eta} = b'(\eta) \quad (5.2)$$

$$Var(y|\eta) = \Sigma = \frac{\partial^2 b(\eta)a(v)}{\partial \eta^2} = b''(\eta)a(v) \quad (5.3)$$

A variância Σ depende somente do parâmetro canônico (e, portanto, da média μ) e é chamada de função de variância.

A Tabela 5.1 mostra alguns membros importantes da família exponencial e suas respectivas componentes:

Tabela 5.1: Algumas distribuições que pertencem à família exponencial.

Distribuição	η	$b(\eta)$	$b'(\eta)$
Normal $N(\mu, \sigma^2)$	μ	$\eta^2/2$	$\mu = \eta$
Bernoulli $Ber(\pi)$	$\log(\pi/(1 - \pi))$	$\log(1 + \exp(\eta))$	$\pi = \exp(\eta)/(1 + \exp(\eta))$
Poisson $Po(\lambda)$	$\log(\lambda)$	$\exp(\eta)$	$\lambda = \exp(\eta)$
Gama $Ga(\alpha, \beta)$	$-1/\alpha$	$-\log(-\eta)$	$\alpha = -1/\eta$
Inv. Gaussiana $IG(\mu, \sigma^2)$	$-1/2\mu^2$	$-(-2\eta)^{1/2}$	$\mu = (-2\eta)^{-1/2}$

As funções de ligação canônicas produzem modelos com propriedades matemáticas e estatísticas convenientes. Porém, esta não deve ser a única razão para trabalhar com elas, pois em determinadas aplicações, as funções de ligação não naturais podem ser mais apropriadas.

Como ilustração, a distribuição de Poisson com média λ será representada na forma da família exponencial.

$$\begin{aligned} (y|\lambda) &\sim Po(\lambda) \\ p(y|\lambda) &= \frac{\lambda^y \exp(-\lambda)}{y!}, \quad y = 0, 1, 2, \dots \\ &= \exp\{y \log(\lambda) - \lambda - \log(y!)\}. \end{aligned}$$

Esta densidade pertence à família exponencial com parâmetro canônico $\eta = \log(\lambda)$, $b(\eta) = \lambda = \exp(\eta)$, $a(\phi) = 1$ e $c(y) = -\log(y!)$. Assim, os primeiros momentos da distribuição são:

$$\begin{aligned} \mu &= b'(\eta) = \exp(\eta) = \lambda, \\ \Sigma &= b''(\eta) = \exp(\eta) = \lambda. \end{aligned}$$

5.3 Métodos Bayesianos para modelos lineares generalizados

Métodos Bayesianos para analisar modelos lineares generalizados são de origem recente e têm crescido muito rapidamente desde a última década, devido principalmente ao enorme avanço nos métodos para calcular as distribuições a posteriori requeridas, tais como os métodos de simulação usando *MCMC*. Outra razão é a flexibilidade da modelagem hierárquica Bayesiana que estende os modelos lineares generalizados a problemas mais complexos como modelos lineares generalizados semiparamétricos, efeitos aleatórios generalizados e modelos de misturas, abordagens dinâmicas ou de espaço de estados, de séries temporais não normais, dados longitudinais e processos espaço-temporais e de sobrevivência.

Vários resultados para os modelos lineares generalizados Bayesianos e suas extensões encontram-se no livro de referência editado por [Dey et al. \(2000\)](#). Um estudo mais completo do ponto de vista Bayesiano é o capítulo de [Gelfand e Ghosh \(2000\)](#).

Nos modelos Bayesianos assume-se que $\boldsymbol{\beta}$ é um vetor aleatório com densidade a priori $p(\boldsymbol{\beta})$, e as estimativas de Bayes estão baseadas na densidade a posteriori $p(\boldsymbol{\beta}|\mathbf{Y} = (y_1, \dots, y_n))$ de $\boldsymbol{\beta}$ dado o conjunto de dados. O teorema de Bayes relaciona as densidades a priori e a

posteriori através da relação

$$p(\boldsymbol{\beta}|\mathbf{Y}) = \frac{p(\mathbf{Y}|\boldsymbol{\beta})p(\boldsymbol{\beta})}{\int p(\mathbf{Y}|\boldsymbol{\beta})p(\boldsymbol{\beta})d\boldsymbol{\beta}} \quad (5.4)$$

onde $p(\mathbf{Y}|\boldsymbol{\beta})$ é a função de verossimilhança. As densidades marginais a posteriori para os componentes de $\boldsymbol{\beta}$, as médias a posteriori e as matrizes de covariância, etc., podem ser obtidas da distribuição a posteriori integrando com respeito às quantidades desconhecidas (ou somando no caso discreto). Por exemplo,

$$E(\boldsymbol{\beta}|\mathbf{Y}) = \int \boldsymbol{\beta} p(\boldsymbol{\beta}|\mathbf{Y}) d\boldsymbol{\beta} \quad (5.5)$$

é a média a posteriori, que é a estimativa ótima de $\boldsymbol{\beta}$ supondo perda quadrática, e

$$Var(\boldsymbol{\beta}|\mathbf{Y}) = \int (\boldsymbol{\beta} - E(\boldsymbol{\beta}|\mathbf{Y}))(\boldsymbol{\beta} - E(\boldsymbol{\beta}|\mathbf{Y}))' p(\boldsymbol{\beta}|\mathbf{Y}) d\boldsymbol{\beta} \quad (5.6)$$

é a matriz de covariâncias a posteriori associada, que é uma medida da precisão da estimativa da média a posteriori. Desta forma, aparentemente, o paradigma Bayesiano parece ser facilmente implementado. Porém, soluções analíticas exatas de essas integrais podem ser obtidas somente para modelos especiais, por exemplo, para o modelo linear normal. Para a maior parte dos modelos importantes, por exemplo, modelos logit binomiais com pelo menos uma covariável, não existe nenhuma priori conjugada que permita um tratamento analítico conveniente. Por conseguinte, a implementação direta da abordagem de estimação Bayesiana através das Equações (5.5) e (5.6) requer integração numérica ou por Monte Carlo. Como estas integrais têm a dimensão do vetor $\boldsymbol{\beta}$ (possivelmente dimensões grandes) esta não é uma tarefa simples. Para modelos muito mais complexos, técnicas de simulação *MCMC* estão disponíveis e são de uso comum, por exemplo:

- *Estimação da moda a posteriori*: A estimativa da moda a posteriori $\tilde{\boldsymbol{\beta}}_q$ maximiza a densidade a posteriori q ou equivalentemente o logaritmo da verossimilhança a posteriori

$$l_q(\boldsymbol{\beta}|\mathbf{Y}) = l(\boldsymbol{\beta}) + \log p(\boldsymbol{\beta}) \quad (5.7)$$

em que $l(\boldsymbol{\beta})$ é o logaritmo da verossimilhança do modelo linear generalizado em consideração. Tem sido proposta por vários autores, como por exemplo, [Leonard \(1972\)](#), [Land \(1978\)](#), e [Zellner e Rossi \(1984\)](#).

- *Inferência completamente Bayesiana via MCMC*: Está baseada em amostras extraídas da distribuição a posteriori e aproxima médias e variâncias a posteriori pelo esquema *MCMC*. Para detalhes nos métodos *MCMC* como o amostrador de Gibbs e o algoritmo de Metropolis-Hastings veja [Gamerman e Lopes \(2006\)](#).

5.4 Modelos lineares generalizados dinâmicos

Quando os dados consistem em observações dependentes como as séries temporais, a classe dos modelos lineares generalizados não pode ser utilizada diretamente. Por exemplo, o número de homicídios no Estado de Espírito Santo em um período de tempo é evidentemente não normal. Outro exemplo, a presença de uma determinada ave em vários locais dos Estados Unidos é registrada como (1 = presença, 0 = ausência) ao longo dos anos. Este é um problema com observações Bernoulli.

Uma opção muito utilizada na prática para analisar séries temporais é a classe de modelos lineares dinâmicos (Seção 2.2). Geralmente eles assumem observações Gaussianas, mas existe uma generalização para a família exponencial formalizada por [West et al. \(1985\)](#), chamada modelos lineares generalizados dinâmicos. Esta classe também pode ser vista como uma generalização dos modelos lineares generalizados, com os parâmetros mudando no tempo.

A extensão dinâmica dos modelos lineares generalizados é obtida substituindo a equação de observação do modelo linear dinâmico pela equação do parâmetro natural supondo uma distribuição na família exponencial para cada componente s do vetor \mathbf{Y}_t , $s = 1, \dots, S$. As inovações da equação de sistema seguem uma distribuição Gaussiana.

$$\begin{aligned}
 p(y_{ts}|\eta_{ts}) &\propto \exp\{y_{ts}\eta_{ts} - b(\eta_{ts}) + c(y_{ts})\}, & (5.8) \\
 \eta_{ts} &= \delta(\theta_{ts}), \\
 \boldsymbol{\theta}_t &= \mathbf{F}'_t \boldsymbol{\beta}_t, \\
 \boldsymbol{\beta}_t &= \mathbf{G}_t \boldsymbol{\beta}_{t-1} + \mathbf{w}_t, \quad \mathbf{w}_t \sim N(\mathbf{0}, \mathbf{W}_t),
 \end{aligned}$$

em que $\boldsymbol{\theta}_t = (\theta_{t1}, \theta_{t2}, \dots, \theta_{tS})'$. Esta é uma extensão evidente do modelo linear dinâmico,

pois a equação de observação pode ser não Gaussiana e a regressão linear dinâmica expressada através de $\boldsymbol{\theta}_t$ pode afetar a equação de observação de uma forma não linear.

5.4.1 Filtragem e suavização generalizadas

Embora a formulação dos modelos lineares dinâmicos generalizados seja direta, o problema da filtragem e a suavização é mais complicada. A análise Bayesiana baseada em integrações numéricas, geralmente requer integrações multidimensionais repetidas e pode chegar a ser muito complicada.

Sejam $\boldsymbol{\psi}$ o vetor de parâmetros das matrizes \mathbf{F}_t , \mathbf{G}_t e \mathbf{W}_t e $\boldsymbol{\eta}_t = (\eta_{t1}, \dots, \eta_{tS})$, então a distribuição a posteriori conjunta é proporcional a

$$p(\boldsymbol{\beta}_{(1:T)}, \boldsymbol{\psi} | \mathbf{D}_T) \propto \left[\prod_{t=1}^T p(\mathbf{y}_t | \boldsymbol{\theta}_t, \boldsymbol{\eta}_t) \right] \left[\prod_{t=1}^T p(\boldsymbol{\beta}_t | \boldsymbol{\beta}_{t-1}, \boldsymbol{\psi}) \right] p(\boldsymbol{\psi}) \quad (5.9)$$

A estimação do vetor latente está baseada nas densidades a posteriori como (5.9) ou $p(\boldsymbol{\beta}_t | \mathbf{y}_T)$ para a suavização, ou $p(\boldsymbol{\beta}_t | \mathbf{y}_{t-1})$, $p(\boldsymbol{\beta}_t | \mathbf{y}_t)$ para previsão e filtragem. Nestas densidades a dependência em $\boldsymbol{\psi}$ foi omitida. Podemos distinguir três abordagens de estimação:

- Análise conjugada priori-posteriori, que procura resolver analiticamente as integrais necessárias no teorema de Bayes, talvez fazendo aproximações adicionais, como em [West et al. \(1985\)](#). Ao contrário de uma equação de transição explícita, eles assumem distribuições conjugadas priori-posteriori para o parâmetro natural e impõem uma equação de previsão para $\boldsymbol{\beta}_t$, envolvendo o conceito de fator de desconto para contornar a estimação das matrizes de covariância do erro desconhecidas.
- Estimação da moda a posteriori, evitando os procedimentos de integração ou amostragem.
- Inferência completamente Bayesiana, ou pelo menos, análise das médias a posteriori baseada em integrações numéricas ou em métodos Monte Carlo ou *MCMC*.

Os dois últimos tópicos foram revisados no final da Seção 5.3. [Ferreira e Gamerman \(2000\)](#) apresentam uma revisão de diferentes metodologias estatísticas propostas para tratar com os modelos lineares generalizados dinâmicos desde um ponto de vista Bayesiano.

5.4.2 Filtro de Kalman estendido

Desde que o filtro de Kalman linear não é mais aplicável para problemas não Gaussianos, vários filtros aproximados que realizam estimações satisfatórias em muitas situações têm sido propostos: o filtro de Kalman estendido, o filtro de segunda ordem, o filtro da raiz quadrada, etc. (Anderson e Moore, 1979).

O filtro de Kalman estendido é um método para avaliar o vetor de estados nos modelos dinâmicos descritos na Equação (5.8). Para aproximar a média (moda) a posteriori do processo latente $\beta_{(1:T)}$ a função resposta $f(\theta_t)$ é linearizada e a equação de observação é aproximada por uma equação de observação Gaussiana. O modelo converte-se então em um modelo Gaussiano aproximado e é possível aplicar o bem conhecido filtro de Kalman. Esta metodologia é conhecida como o filtro de Kalman estendido. A seguir detalharemos os passos necessários para sua aplicação.

A priori no tempo $t = 0$ é $(\beta_0 | \mathbf{D}_0) \sim N(\mathbf{m}_0, \mathbf{C}_0)$. Para $t = 1, \dots, T$ aplicamos os seguintes passos:

- Conhecemos $(\beta_{t-1} | \psi, \mathbf{D}_{t-1}) \sim N(\mathbf{m}_{t-1}, \mathbf{C}_{t-1})$ do tempo anterior.
- Priori em t :

$$(\beta_t | \psi, \mathbf{D}_{t-1}) \sim N(\mathbf{a}_t, \mathbf{R}_t),$$

onde

$$\mathbf{a}_t = \mathbf{G}_t \mathbf{m}_{t-1} \quad \text{e} \quad \mathbf{R}_t = \mathbf{G}_t \mathbf{C}_{t-1} \mathbf{G}_t' + \mathbf{W}_t^{-1}.$$

- Linearizamos a função resposta f usando expansão de Taylor em torno do ponto: $\hat{\theta}_t = \mathbf{F}_t' \mathbf{a}_t$. Assim podemos aproximar cada elemento do vetor de médias mediante

$$\mu_{ts} = f(\theta_{ts}) \approx f(\hat{\theta}_{ts}) + f'(\hat{\theta}_{ts})(\theta_{ts} - \hat{\theta}_{ts}) \quad (5.10)$$

e a variância $\Sigma_t = \text{diag}(\Sigma_{ts})$ por $\hat{\Sigma}_{ts} = \Sigma_{ts}(\hat{\theta}_{ts})$, $s = 1, \dots, S$.

- Calculamos o vetor de observações artificiais $\hat{\mathbf{y}}_t = (\hat{y}_{t1}, \hat{y}_{t2}, \dots, \hat{y}_{tS})'$, com

$$\hat{y}_{ts} = \left[f'(\hat{\theta}_{ts}) \right]^{-1} \left\{ y_{ts} - f(\hat{\theta}_{ts}) \right\} + \hat{\theta}_{ts} \quad (5.11)$$

e a matriz de covariância aproximada

$$\hat{\mathbf{V}}_t = \text{diag} \left(\frac{\hat{\Sigma}_{ts}}{[f'(\hat{\theta}_{ts})]^2} \right). \quad (5.12)$$

- Previsão um passo a frente:

$$(\hat{\mathbf{y}}_t | \boldsymbol{\psi}, \mathbf{D}_{t-1}) \sim N(\mathbf{f}_t, \mathbf{Q}_t)$$

onde

$$\mathbf{f}_t = \mathbf{F}'_t \mathbf{a}_t \quad \text{e} \quad \mathbf{Q}_t = \mathbf{F}'_t \mathbf{R}_t \mathbf{F}_t + \hat{\mathbf{V}}_t.$$

- Posteriori em t :

$$(\boldsymbol{\beta}_t | \boldsymbol{\psi}, \mathbf{D}_t) \sim N(\hat{\mathbf{m}}_t, \hat{\mathbf{C}}_t)$$

com

$$\hat{\mathbf{m}}_t = \mathbf{a}_t + \mathbf{A}_t(\hat{\mathbf{y}}_t - \mathbf{f}_t) \quad \text{e} \quad \hat{\mathbf{C}}_t = \mathbf{R}_t - \mathbf{A}_t \mathbf{Q}_t \mathbf{A}'_t,$$

onde

$$\mathbf{A}_t = \mathbf{R}_t \mathbf{F}_t \mathbf{Q}_t^{-1}.$$

Em [Fahrmeir \(1992\)](#) e [Fahrmeir e Tutz \(2001\)](#) encontram-se expressões equivalentes do filtro de Kalman estendido para modelos lineares generalizados dinâmicos multivariados, para estimar o processo latente através da moda a posteriori. Uma vez obtidos os primeiros momentos aproximados, $\hat{\mathbf{m}}_t$ e $\hat{\mathbf{C}}_t$, para cada tempo, é possível utilizar a amostragem recursiva (*backward sampler*).

5.5 Modelos lineares generalizados dinâmicos para dados de área

Considere uma área de interesse, dividida em S regiões. Estas regiões podem formar uma grade regular ou irregular e estão relacionadas por um sistema de vizinhança $\{N_s, s = 1, \dots, S\}$. Seja y_{ts} , $t = 1, \dots, T$, $s = 1, \dots, S$ uma observação com distribuição na família

exponencial de parâmetro canônico η_{ts} , no tempo t e na região s . Nosso modelo espaço-temporal para dados desta natureza é

$$\begin{aligned}
p(y_{ts}|\eta_{ts}) &\propto \exp\{y_{ts}\eta_{ts} - b(\eta_{ts}) + c(y_{ts})\}, \\
\eta_{ts} &= \delta(\theta_{ts}), \\
\boldsymbol{\theta}_t &= \mathbf{F}'_t \boldsymbol{\beta}_t, \quad \boldsymbol{\theta}_t = (\theta_{t1}, \dots, \theta_{tS})' \\
\boldsymbol{\beta}_t &= \mathbf{G}_t \boldsymbol{\beta}_{t-1} + \boldsymbol{\omega}_t, \quad \boldsymbol{\omega}_t \sim \text{CAMGP}(\mathbf{0}, \mathbf{W}_t^{-1}),
\end{aligned} \tag{5.13}$$

O nosso modelo está baseado na Equação (5.8), na qual a distribuição normal da equação de evolução é substituída pela distribuição do campo aleatório Markoviano Gaussiano próprio, com matriz de precisão \mathbf{W}_t^{-1} . O principal interesse é fazer inferência para o processo latente $\boldsymbol{\beta}_{(1:T)}$. As características e propriedades dos modelos para dados Gaussianos propostos na Seção 4.4 podem ser estendidos para a classe de modelos generalizados dinâmicos deste capítulo.

5.5.1 Inferência Bayesiana

Como a distribuição a posteriori em (5.9) não tem forma fechada conhecida, nossa proposta é utilizar métodos *MCMC* para amostrá-la em dois blocos: amostragem do vetor de parâmetros $\boldsymbol{\psi}$ e amostragem do processo latente $\boldsymbol{\beta}_{(1:T)} = (\boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_T)'$.

Amostragem de $\boldsymbol{\psi}$

O vetor de parâmetros $\boldsymbol{\psi}$ inclui as quantidades desconhecidas das matrizes de desenho \mathbf{F}_t , de evolução \mathbf{G}_t e de precisão do *CAMGP*, \mathbf{W}_t^{-1} e, eventualmente, outras quantidades que sejam incorporadas ao modelo. Para todos os elementos de $\boldsymbol{\psi}$, utilizamos os métodos *MCMC* usuais: o amostrador de Gibbs para aqueles que tenham a sua distribuição condicional completa conhecida e passos de Metropolis-Hastings com densidades propostas adequadas para aqueles cujas condicionais completas não têm forma fechada. Esta amostragem é específica para cada modelo.

Amostragem do processo latente

Recentemente, [Ravines \(2006\)](#) propôs um esquema de amostragem para modelos lineares generalizados denominado *CUBS*, em que uma priori conjugada para o parâmetro canônico é introduzida para aproximar os primeiros momentos, $\hat{\mathbf{m}}_t$ e $\hat{\mathbf{C}}_t$, e o processo latente é estimado recursivamente através do *backward sampler*. Isto resulta muito eficiente para o caso de observações univariadas em cada tempo. A aplicação do *CUBS* na nossa classe de modelos dependeria da especificação de uma priori conjugada multivariada para o parâmetro canônico. Isto implicaria dificuldades adicionais na resolução do problema, razão pela qual decidimos utilizar uma metodologia alternativa.

Seguindo o raciocínio da Seção 3.4, um método alternativo ao filtro de Kalman estendido e ao *FFBS* pode ser elaborado também para simular o processo latente $\beta_{(1:T)}$ do modelo linear generalizado dinâmico proposto. O passo de filtragem, que utiliza o filtro de Kalman quando o modelo dinâmico é normal e o filtro de Kalman estendido quando o modelo dinâmico é linear generalizado, é substituído pelo que denominaremos de filtro de informação estendido, para aproveitar as matrizes de precisão esparsas do nosso modelo. O passo de amostragem, *backward sampler* continua sendo múltiplo para aumentar a velocidade de convergência.

Filtro de informação estendido

- Priori em t : $\beta_t | \psi, \mathbf{D}_{t-1} \sim N(\mathbf{a}_t, \mathbf{R}_t)$, com

$$\mathbf{a}_t = \mathbf{G}_t \mathbf{m}_{t-1}, \quad (5.14)$$

$$\mathbf{R}_t^{-1} = \mathbf{W}_t^{-1} - \mathbf{W}_t^{-1} \mathbf{G}_t (\mathbf{G}_t' \mathbf{W}_t^{-1} \mathbf{G}_t + \mathbf{C}_{t-1}^{-1})^{-1} \mathbf{G}_t' \mathbf{W}_t^{-1} \quad (5.15)$$

- Função resposta f linearizada, usando expansão de Taylor em torno do ponto $\hat{\boldsymbol{\theta}}_t$, $\hat{\boldsymbol{\theta}}_t = \mathbf{F}_t' \mathbf{a}_t$, $\hat{\boldsymbol{\theta}}_t = (\hat{\theta}_{t1}, \dots, \hat{\theta}_{tS})'$. O vetor de médias é aproximado por

$$\mu_{ts} = f(\theta_{ts}) \approx f(\hat{\theta}_{ts}) + f'(\hat{\theta}_{ts})(\theta_{ts} - \hat{\theta}_{ts})$$

e a matriz de covariância por $\boldsymbol{\Sigma}_t = \text{diag}(\boldsymbol{\Sigma}_{ts})$ por $\hat{\boldsymbol{\Sigma}}_{ts} = \boldsymbol{\Sigma}_{ts}(\hat{\theta}_{ts})$, $s = 1, \dots, S$.

- Vetor de observações artificiais como na Equação (5.11):

$$\hat{y}_{ts} = \left[f'(\hat{\theta}_{ts}) \right]^{-1} \left\{ y_{ts} - f(\hat{\theta}_{ts}) \right\} + \hat{\theta}_{ts}, \quad \hat{\mathbf{y}}_t = (\hat{y}_{t1}, \dots, \hat{y}_{tS}) \quad (5.16)$$

e a matriz de precisão aproximada

$$\hat{\mathbf{V}}_t^{-1} = \text{diag} \left(\frac{\left[f'(\hat{\theta}_{ts}) \right]^2}{\hat{\Sigma}_{ts}} \right) \quad (5.17)$$

- Posteriori em t : $\beta_t | \psi, \mathbf{D}_t \sim N(\hat{\mathbf{m}}_t, \hat{\mathbf{C}}_t)$, com

$$\hat{\mathbf{C}}_t^{-1} = \mathbf{R}_t^{-1} + \mathbf{F}_t \hat{\mathbf{V}}_t^{-1} \mathbf{F}_t', \quad (5.18)$$

$$\hat{\mathbf{C}}_t^{-1} \hat{\mathbf{m}}_t = \mathbf{R}_t^{-1} \mathbf{a}_t + \mathbf{F}_t \hat{\mathbf{V}}_t^{-1} \hat{\mathbf{y}}_t. \quad (5.19)$$

Amostragem do $\beta_{(1:T)}$

O filtro de Kalman estendido e o filtro de informação estendido são métodos aproximados para calcular os vetores de médias e as matrizes de covariância das distribuições a posteriori $p(\beta_t | \psi, \mathbf{D}_t)$. Analogamente ao caso do *FIFBS* (Subseção 3.4.2) o β_T^* é simulado da distribuição $N(\hat{\mathbf{m}}_T, \hat{\mathbf{C}}_T)$, resultado obtido do filtro de informação estendido e para $t = T - 1, \dots, 0$ a amostragem recursiva retrospectiva de $\beta_t^* | \beta_{t+1}^*, \psi, \mathbf{D}_T$ é realizada da densidade condicional $N(\hat{\mathbf{b}}_t, \hat{\mathbf{B}}_t)$, onde

$$\hat{\mathbf{B}}_t^{-1} = \hat{\mathbf{C}}_t^{-1} + \mathbf{G}'_{t+1} \mathbf{W}_{t+1}^{-1} \mathbf{G}_{t+1}, \quad (5.20)$$

$$\hat{\mathbf{B}}_t^{-1} \hat{\mathbf{b}}_t = \hat{\mathbf{C}}_t^{-1} \hat{\mathbf{m}}_t + \mathbf{G}'_{t+1} \mathbf{W}_{t+1}^{-1} \beta_{t+1}^*. \quad (5.21)$$

A densidade conjunta dos β_t^* pode ser escrita como o produto das densidades retrospectivas $p(\beta_t^* | \beta_{t+1}^*, \psi, \mathbf{D}_T)$ (Frühwirth-Schnatter, 1994; Carter e Kohn, 1994). O vetor $\beta_{(1:T)}^*$ obtido desta distribuição é aceito com uma certa probabilidade utilizando o algoritmo de Metropolis-Hastings, pois a posteriori $\beta_{(1:T)} | \psi, \mathbf{D}_T$ não tem forma fechada. No algoritmo de Metropolis-Hastings é preciso construir uma densidade proposta para calcular a probabilidade de aceitação, geralmente uma que produza taxas de aceitação, por exemplo, entre 0.3 e 0.7. Nós assumimos que a distribuição $N(\hat{\mathbf{b}}_t, \hat{\mathbf{B}}_t)$ fornece uma boa aproximação da

verdadeira distribuição a posteriori de β_t , por isso aceitamos o vetor simulado com probabilidade 1. Desta forma evitamos avaliar o vetor do processo latente na probabilidade de aceitação em cada passo do esquema *MCMC*. Esta suposição resulta adequada nas aplicações apresentadas nos Capítulos 6 e 7.

Este método aproximado, similar ao *FIFBS* (*FFBS*) que substitui no passo de filtragem o filtro de informação (filtro de Kalman) pelo filtro de informação estendido (filtro de Kalman estendido) é denominado *EF(I)FBS* (abreviação de *extended forward (information) filter backward sampler*).

O esquema geral de amostragem para a classe de modelos espaço-temporais generalizados dinâmicos é:

1. Dado o modelo espaço-temporal generalizado dinâmico, identificar as seguintes funções:
 - (a) O parâmetro natural η_{ts} , a média μ_{ts} e a variância Σ_{ts} ;
 - (b) O preditor linear θ_{ts} ;
 - (c) A função de ligação $g(\cdot)$ e a função resposta $f(\cdot)$;
 - (d) A função $\delta(\cdot)$ que transforma o preditor linear diretamente no parâmetro natural;
2. Aplicar o primeiro passo do *EFIFBS*, o filtro de informação estendido, calculando \hat{y}_{ts} e $\hat{\mathbf{V}}_t^{-1}$;
3. Aplicar o segundo passo do *EFIFBS*, amostrando o $\beta_{(1:T)}$;
4. Simular da posteriori $p(\psi | \beta_{(1:T)}, \mathbf{D}_T)$ de acordo com a especificação de cada modelo.

5.6 Seleção de modelos via densidade preditiva

Dado um conjunto de modelos para um determinado problema, busca-se selecionar o melhor entre eles baseados em algum critério. Na Subseção 3.5.2 introduzimos o critério de densidade preditiva para seleção de modelos. Supondo que existem M modelos todos com iguais probabilidades a priori de ser o “verdadeiro”, as probabilidades a posteriori são proporcionais

a suas respectivas densidades preditivas. Estas densidades dependem da distribuição a priori do vetor de parâmetros $\boldsymbol{\psi}$. Para contornar esta dificuldade, utilizamos a abordagem da amostra piloto (Frühwirth-Schnatter, 1995) das primeiras t^* observações no tempo; isto produz prioris calibradas para os parâmetros de cada modelo. Então, podemos utilizar integração Monte Carlo para calcular a distribuição preditiva sob cada modelo para as restantes $T - t^*$ observações.

Seja a densidade observacional $p_m(\mathbf{y}_t | \boldsymbol{\eta}_t(\boldsymbol{\beta}_t))$ e a densidade da evolução $p_m(\boldsymbol{\beta}_t | \boldsymbol{\beta}_{t-1}, \boldsymbol{\psi})$. Note que as definições de $\boldsymbol{\beta}_t$ e $\boldsymbol{\psi}$ podem ser (e geralmente o são) diferentes sob cada modelo, mas esta distinção é omitida para manter simples a notação. Seja $p_m(\boldsymbol{\beta}_{1:t-1}, \boldsymbol{\psi} | \mathbf{D}_{t-1})$ que denota a distribuição a posteriori conjunta de $\boldsymbol{\psi}$ e $\boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_{t-1}$ sob o modelo m até o tempo $t - 1$. Então, a distribuição preditiva de \mathbf{y}_t sob o modelo m dada a informação até o tempo $t - 1$ é

$$\begin{aligned} p_m(\mathbf{y}_t | \mathbf{D}_{t-1}) &= \int p_m(\mathbf{y}_t | \boldsymbol{\eta}_t(\boldsymbol{\beta}_t)) p_m(\boldsymbol{\beta}_t | \boldsymbol{\beta}_{t-1}, \boldsymbol{\psi}) d\boldsymbol{\beta}_t \\ &\quad \times \int p_m(\boldsymbol{\beta}_{1:t-1}, \boldsymbol{\psi} | \mathbf{D}_{t-1}) d\boldsymbol{\beta}_{1:t-1} d\boldsymbol{\psi} \\ &= \int p_m(\mathbf{y}_t | \boldsymbol{\beta}_{t-1}, \boldsymbol{\psi}) p_m(\boldsymbol{\beta}_{1:t-1}, \boldsymbol{\psi} | \mathbf{D}_{t-1}) d\boldsymbol{\beta}_{1:t-1} d\boldsymbol{\psi} \end{aligned} \quad (5.22)$$

pois

$$p_m(\mathbf{y}_t | \boldsymbol{\beta}_{t-1}, \boldsymbol{\psi}) = \int p_m(\mathbf{y}_t | \boldsymbol{\eta}_t(\boldsymbol{\beta}_t)) p_m(\boldsymbol{\beta}_t | \boldsymbol{\beta}_{t-1}, \boldsymbol{\psi}) d\boldsymbol{\beta}_t.$$

O esquema de amostragem esquematizado na seção anterior pode ser utilizado para simular uma amostra $(\boldsymbol{\beta}_{t-1}^{(1)}, \boldsymbol{\psi}^{(1)}), \dots, (\boldsymbol{\beta}_{t-1}^{(L)}, \boldsymbol{\psi}^{(L)})$, com L suficientemente grande, da distribuição a posteriori conjunta $p_m(\boldsymbol{\beta}_{1:t-1}, \boldsymbol{\psi} | \mathbf{D}_{t-1})$. Então uma estimativa da densidade preditiva de \mathbf{y}_t dada a informação até o tempo $t - 1$ é

$$\hat{p}_m(\mathbf{y}_t | \mathbf{D}_{t-1}) = \frac{1}{L} \sum_{l=1}^L p_m(\mathbf{y}_t | \boldsymbol{\beta}_{t-1}^{(l)}, \boldsymbol{\psi}^{(l)}) \quad (5.23)$$

Desta forma, ajustamos um esquema *MCMC* para cada um dos m modelos e para cada tempo t e fazemos previsão para o tempo $t + 1$. Usando o fato que a densidade preditiva conjunta das observações $\mathbf{y}_{t^*+1}, \dots, \mathbf{y}_T$ pode ser escrita como $p_m(\mathbf{y}_{t^*+1}, \dots, \mathbf{y}_T | \mathbf{D}_{t^*}) =$

$\prod_{t=t^*+1}^T p_m(\mathbf{y}_t | \mathbf{D}_{t-1})$, uma estimativa da densidade preditiva conjunta sob o modelo m é

$$\hat{p}_m(\mathbf{y}_{t^*+1}, \dots, \mathbf{y}_T | \mathbf{D}_{t^*}) = \prod_{t=t^*+1}^T \hat{p}_m(\mathbf{y}_t | \mathbf{D}_{t-1}),$$

onde t^* é tal que a densidade a posteriori de $\boldsymbol{\psi}$ $p_m(\boldsymbol{\psi} | \mathbf{D}_{t^*})$ é própria para todo $m = 1, \dots, M$.

Como resultado, selecionamos o modelo com a densidade preditiva mais alta. Assim, o modelo selecionado não só tem a probabilidade a posteriori mais alta, mas também com a melhor performance preditiva.

Neste capítulo foi introduzida a classe de modelos espaço-temporais generalizados dinâmicos para dados de área e uma metodologia para estimar o processo latente, estendendo o algoritmo *FIFBS* através do filtro de informação estendido, junto com uma técnica de seleção de modelos baseada em densidades preditivas.

Capítulo 6

Aplicação da classe de modelos para dados de área: Caso Poisson

6.1 Introdução

Neste capítulo apresentamos uma aplicação do nosso modelo linear generalizado dinâmico para dados espaço-temporais de contagem. Na Seção 6.2 revisamos os dados de homicídios no Estado de Espírito Santo, que estudamos via transformação no Capítulo 4. Descrevemos o modelo geral para dados Poisson na Seção 6.3, e o conjunto de modelos propostos para estes dados de violência na Seção 6.4. O procedimento para fazer inferência Bayesiana é ilustrado na Seção 6.5 dividindo o procedimento para o vetor de parâmetros e para o processo latente. Os resultados da seleção de modelos estão na Seção 6.6.

6.2 Violência no Estado do Espírito Santo

Os dados consistem no número anual de homicídios por município do Estado de Espírito Santo, desde o ano de 1979 até o ano de 1998. Ao longo deste período vários novos municípios foram criados por divisão ou fusão dos antigos. Por razões de compatibilidade utilizaremos em todas as análises o mapa político do ano de 1979 com um total de 52 municípios.

Para cada município s e ano t , $s = 1, \dots, S = 52$, $t = 1, \dots, T = 20$, tem-se a população estimada n_{ts} e o número observado de homicídios z_{ts} . Como é típico para dados de contagem

como estes, assumimos que o número de homicídios y_{ts} em cada município de Espírito Santo tem distribuição de Poisson, com média $\lambda_{ts}n_{ts}$, isto é

$$y_{ts}|\lambda_{ts} \sim Po(\lambda_{ts}n_{ts}),$$

A distribuição de Poisson pertence à família exponencial, pois a sua densidade pode ser escrita da mesma forma que em (5.1):

$$\begin{aligned} (y_{ts}|\lambda_{ts}) &\sim Po(\lambda_{ts}n_{ts}) \\ p(y_{ts}|\lambda_{ts}) &= \frac{(\lambda_{ts}n_{ts})_{ts}^{y_{ts}} \exp(-\lambda_{ts}n_{ts})}{y_{ts}!}, \quad y_{ts} = 0, 1, 2, \dots \\ &= \exp\{y_{ts} \log(\lambda_{ts}n_{ts}) - \lambda_{ts}n_{ts} - \log(y_{ts}!)\}. \end{aligned} \quad (6.1)$$

6.3 Modelo geral

Da Equação (6.1) podem-se derivar as componentes da distribuição: o parâmetro natural $\eta_{ts} = \log(\lambda_{ts}n_{ts})$ e as funções $b(\eta_{ts}) = \lambda_{ts}n_{ts} = \exp(\eta_{ts})$ e $c(y_{ts}) = -\log(y_{ts}!)$. Com estes resultados, é simples identificar as seguintes identidades:

- A média da distribuição $\mu_{ts} = b'(\eta_{ts}) = \exp(\eta_{ts})$
- A variância da distribuição $\Sigma_{ts} = b''(\eta_{ts}) = \exp(\eta_{ts})$.
- O preditor linear: $\theta_{ts} = \log(\lambda_{ts})$.
- A função de ligação: $g(\mu_{ts}) = \log(\mu_{ts}/n_{ts}) = \theta_{ts}$.
- A função resposta: $f(\theta_{ts}) = n_{ts} \exp(\theta_{ts}) = \mu_{ts}$ e a primeira derivada $f'(\theta_{ts}) = n_{ts} \exp(\theta_{ts}) = \mu_{ts}$.
- A função $\delta(\theta_{ts})$:

$$\begin{aligned} \eta_{ts} &= \log(\lambda_{ts}n_{ts}) \\ &= \log(\lambda_{ts}) + \log(n_{ts}) \\ &= \theta_{ts} + \log(n_{ts}) \\ \eta_{ts} &= \delta(\theta_{ts}) \end{aligned}$$

O modelo espaço-temporal generalizado dinâmico para dados de Poisson está definido da seguinte forma

$$\begin{aligned}
p(y_{ts}|n_{ts}) &\propto \exp\{y_{ts}\eta_{ts} - \exp(\eta_{ts}) + \log(y_{ts}!)\}, & (6.2) \\
\eta_{ts} &= \theta_{ts} + \log(n_{ts}), \\
\boldsymbol{\theta}_t &= \mathbf{F}'_t \boldsymbol{\beta}_t, & \boldsymbol{\theta}_t = (\theta_{t1}, \dots, \theta_{tS})' \\
\boldsymbol{\beta}_t &= \mathbf{G}_t \boldsymbol{\beta}_{t-1} + \boldsymbol{\omega}_t, & \boldsymbol{\omega}_t \sim \text{CAMGP}(\mathbf{0}, \mathbf{W}_t^{-1}),
\end{aligned}$$

em que $\boldsymbol{\beta}_t = (\beta'_{t1}, \dots, \beta'_{tS})'$ é o processo latente de interesse; a matriz \mathbf{F}_t é a matriz de desenho que conecta o processo latente com o preditor linear; \mathbf{G}_t é a matriz da evolução espaço-temporal; $\boldsymbol{\omega}_t$ é o campo de inovações da equação de sistema e \mathbf{W}_t descreve a estrutura de covariância do vetor das inovações ou erros de sistema.

6.4 Modelos espaço-temporais propostos

Esta seção apresenta vários casos especiais do modelo espaço-temporal geral (6.2). Uma análise exploratória dos dados de homicídios por município determina que este é claramente um processo não estacionário e inclusive nota-se uma tendência crescente para muitos deles (ver gráficos da taxa de mortalidade por homicídios no Apêndice B), então esta característica tem que ser incluída nos modelos propostos. Ajustaremos os modelos que foram detalhados na Seção 4.4 pelas razões explicadas na referida seção.

- I - Polinomial de primeira ordem
- II - Contaminação
- III - Polinomial de segunda ordem com $\phi_2 = 0$
- IV - Polinomial de segunda ordem
- V - Polinomial de segunda ordem com contaminação na equação da velocidade
- VI - Polinomial de segunda ordem com contaminação na equação do nível

VII - Polinomial de segunda ordem com contaminação nas duas equações de sistema

VIII - Polinomial de segunda ordem com $\phi_2 = 0$ e velocidade igual para cada município.

6.5 Inferência Bayesiana

Estudaremos como fazer inferência para esta classe de modelos. Vamos supor que estamos ajustando o modelo VII, o polinomial de segunda ordem com contaminação nas duas equações de sistema.

- $\mathbf{F}'_t = (\mathbf{I}_S, \mathbf{0}_S)$,
- $\mathbf{G}_t = \begin{pmatrix} \mathbf{G}_{1t} & \mathbf{G}_{1t} \\ \mathbf{0}_S & \mathbf{G}_{2t} \end{pmatrix}$, $\mathbf{G}_{it} = \frac{1}{1+\kappa_i h} \mathbf{H}_i \longrightarrow \{\mathbf{H}_i\}_{kl} = \begin{cases} 1, & k = l, \\ \kappa_i, & k \in N_l, \quad , i = 1, 2, \\ 0, & c.c. \end{cases}$
- $\mathbf{W}_t^{-1} = \begin{pmatrix} \mathbf{W}_{1t}^{-1} & \mathbf{0}_n \\ \mathbf{0}_n & \mathbf{W}_{2t}^{-1} \end{pmatrix}$, $\mathbf{W}_{it}^{-1} = \tau_i(\mathbf{I}_S + \phi_i \mathbf{M})$, $i = 1, 2$.

A inferência pode ser realizada de forma análoga para qualquer outro modelo. Então, seja $\boldsymbol{\psi}$ o vetor de parâmetros das matrizes \mathbf{F}_t , \mathbf{G}_t e \mathbf{W}_t^{-1} . Como \mathbf{F}'_t é constante, temos:

$$\boldsymbol{\psi} = (\{\kappa_1, \kappa_2\}, \{\tau_1, \phi_1, \tau_2, \phi_2\}, \boldsymbol{\beta}_0).$$

Assim, supondo prioris independentes, a distribuição a posteriori conjunta é proporcional a

$$\begin{aligned} p(\boldsymbol{\beta}_{(1:T)}, \boldsymbol{\psi} | \mathbf{D}_T) &\propto \left[\prod_{t=1}^T p(\mathbf{y}_t | \boldsymbol{\theta}_t, \boldsymbol{\eta}_t) \right] \left[\prod_{t=1}^T p(\boldsymbol{\beta}_t | \boldsymbol{\beta}_{t-1}, \boldsymbol{\psi}) \right] p(\boldsymbol{\psi}) \\ p(\boldsymbol{\beta}_{(1:T)}, \boldsymbol{\psi} | \mathbf{D}_T) &\propto \left[\prod_{t=1}^T p(\mathbf{y}_t | \boldsymbol{\theta}_t, \boldsymbol{\eta}_t) \right] \left[\prod_{t=1}^T p(\boldsymbol{\beta}_{1,t} | \boldsymbol{\beta}_{1,t-1}, \tau_1, \phi_1, \kappa_1) \right] p(\tau_1) p(\phi_1) p(\kappa_1) \\ &\quad \times \left[\prod_{t=1}^T p(\boldsymbol{\beta}_{2,t} | \boldsymbol{\beta}_{2,t-1}, \tau_2, \phi_2, \kappa_2) \right] p(\tau_2) p(\phi_2) p(\kappa_2) p(\boldsymbol{\beta}_{1,0}) p(\boldsymbol{\beta}_{2,0}) \quad (6.3) \end{aligned}$$

com $\boldsymbol{\beta}_{(1:T)} = (\boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_T)'$ e $\boldsymbol{\beta}_t = (\boldsymbol{\beta}'_{1,t}, \boldsymbol{\beta}'_{2,t})'$.

6.5.1 Especificação das prioris

O modelo está completo dentro do paradigma Bayesiano quando as prioris para as quantidades desconhecidas são especificadas. A seguir, detalhamos as distribuições a priori para os parâmetros do modelo VII:

τ_1, τ_2 : A priori para cada um destes parâmetros de precisão é gama $Ga(a, a)$, com média igual a 1 e variância igual a $1/a$. Depois de alguns testes, $a = 4$ foi escolhido, pois com valores de a menores os $\tau_i, i = 1, 2$ não pareciam convergir. Com valores maiores de a as prioris foram consideradas muito informativas. A condicional completa resultante é gama e utilizamos o amostrador de Gibbs para simular desta distribuição.

ϕ_1, ϕ_2 : As prioris para os parâmetros que controlam a correlação espacial são da forma

$$p_0(\phi_i) \propto \begin{cases} 1, & 0 < \phi_i < 1 \\ \phi_i^{-b}, & \phi_i \geq 1 \end{cases}$$

com os expoentes $b = 3$ e $b = 4$ para ϕ_1 e ϕ_2 , respectivamente. Esta priori foi considerada depois de vários testes com prioris diferentes, incluindo aquela proposta por [Ferreira e De Oliveira \(2007\)](#), mas as estimativas dos $\phi_i, i = 1, 2$ tinham um comportamento crescente. Esta distribuição a priori penaliza os valores altos de $\phi_i, i = 1, 2$, e quanto maior o valor de b , mais rápido o decaimento exponencial, isto é, penaliza os valores altos de $\phi_i, i = 1, 2$ com mais severidade. A condicional completa não tem forma conhecida e amostramos com passos de Metropolis, com densidade proposta para $\log(\phi_i), i = 1, 2$.

κ_1, κ_2 : As prioris para os índices de contaminação são $U(0,1)$, mas as condicionais completas também não têm forma fechada. As densidades propostas para estes parâmetros são normais truncadas entre 0 e 1, $NTr_{[0,1]}(\kappa^*, \Delta_\kappa)$, com média igual ao valor atual κ^* e uma variância Δ_κ constante que produza taxas de aceitação razoáveis (entre 0.3 e 0.7). Os valores das variâncias para este modelo foram $\Delta_{\kappa_1} = 1 \times 10^{-5}$ e $\Delta_{\kappa_2} = 5 \times 10^{-4}$.

β_0 : A distribuição a priori para o vetor inicial β_0 é normal multivariada com vetor de médias zero e matriz de covariância diagonal com valores pequenos, da ordem de

10^{-3} . Quando amostramos o processo latente com o *EFIFBS*, incluímos um passo adicional para gerar valores de β_0 .

6.5.2 Estimação do processo latente

Para gerar amostras da distribuição a posteriori conjunta utilizamos o *EFIFBS* detalhado na Subseção 5.5.1. Precisamos calcular as seguintes quantidades: as observações aproximadas

$$\begin{aligned}\hat{y}_{ts} &= [f'(\hat{\theta}_{ts})]^{-1} \{y_{ts} - f(\hat{\theta}_{ts})\} + \hat{\theta}_{ts}, & \hat{\mathbf{y}}_t &= (\hat{y}_{t1}, \dots, \hat{y}_{tS}) \\ &= [n_{ts} \exp(\hat{\theta}_{ts})]^{-1} \{y_{ts} - n_{ts} \exp(\hat{\theta}_{ts})\} + \hat{\theta}_{ts} \\ \hat{y}_{ts} &= y_{ts} \exp(-\hat{\theta}_{ts})/n_{ts} + \hat{\theta}_{ts} - 1\end{aligned}\tag{6.4}$$

e a matriz de precisão aproximada

$$\begin{aligned}\hat{\mathbf{V}}_t^{-1} &= \text{diag} \left(\frac{[f'(\hat{\theta}_{ts})]^2}{\hat{\Sigma}_{ts}} \right) \\ &= \text{diag} \left(\frac{[n_{ts} \exp(\hat{\theta}_{ts})]^2}{\exp(\hat{\theta}_{ts} + \log(n_{ts}))} \right) \\ \hat{\mathbf{V}}_t^{-1} &= \text{diag} \left(\frac{n_{ts}}{\exp(-\hat{\theta}_{ts})} \right)\end{aligned}$$

com $\hat{\boldsymbol{\theta}}_t = (\hat{\theta}_{t1}, \hat{\theta}_{t2}, \dots, \hat{\theta}_{tS})' = \mathbf{F}'_t \mathbf{a}_t$. Desta forma podemos gerar valores de β_t condicionado no valor β_{t+1} da densidade aproximada $N(\hat{\mathbf{b}}_t, \hat{\mathbf{B}}_t)$ (Equações (5.20) e (5.21)).

Diversos conjuntos de dados artificiais com estrutura espacial foram simulados para testar a metodologia de estimação. Obtivemos bons resultados que não são apresentados aqui. Para o conjunto de dados reais, dentro do esquema *MCMC*, geramos duas cadeias independentes partindo de pontos iniciais dispersos, de tamanho 90000 e com um *burn-in* de 10000 para todos os modelos. As amostras finais de tamanho 4000 são obtidas guardando conservadoramente as amostras a cada 20 iterações do método *MCMC*. Estes códigos estão implementados em *Ox* e verificamos a convergência das cadeias utilizando os diagnósticos de Gelman e Rubin (1992) e de Raftery e Lewis (1996) calculados usando a biblioteca *CODA*

do R como na Seção 4.7. Todos os diagnósticos indicaram convergência. Aqui mostramos somente o diagnóstico de Gelman e Rubin.

6.6 Resultados da seleção de modelos

Aplicando os resultados obtidos na Seção 5.6 para todos os modelos e considerando os dez primeiros tempos como amostra piloto ($t^* = 10$) podemos selecionar o modelo com melhor performance preditiva e maior probabilidade a posteriori.

A Tabela 6.1 mostra o logaritmo da densidade preditiva para o conjunto de modelos ajustados. Para cada modelo m , temos

$$\log p_m(\mathbf{y}_{11}, \dots, \mathbf{y}_{20} | \mathbf{D}_{10}) = \sum_{t=11}^{20} \log p_m(\mathbf{y}_t | \mathbf{D}_{t-1}), \quad (6.5)$$

Tabela 6.1: Logaritmo da densidade preditiva para os modelos considerados.

Modelo	Log(densidade preditiva)
I	-1881.40
II	-1873.36
III	-4254.69
IV	-2364.12
V	-5815.18
VI	-3365.69
VII	-2227.64
VIII	-2008.27

A Tabela 6.1 mostra que o melhor modelo é o modelo de contaminação (modelo II), resultado que coincide com aquele obtido na Seção 4.7.2 quando ajustamos os modelos espaço-temporais Gaussianos aos dados de violência transformados. Para mostrar a diferença de performance preditiva entre os modelos a cada tempo t , a Figura 6.1 representa as densidades preditivas um passo a frente para os modelos ajustados. O modelo II é o melhor em quase todos os tempos.

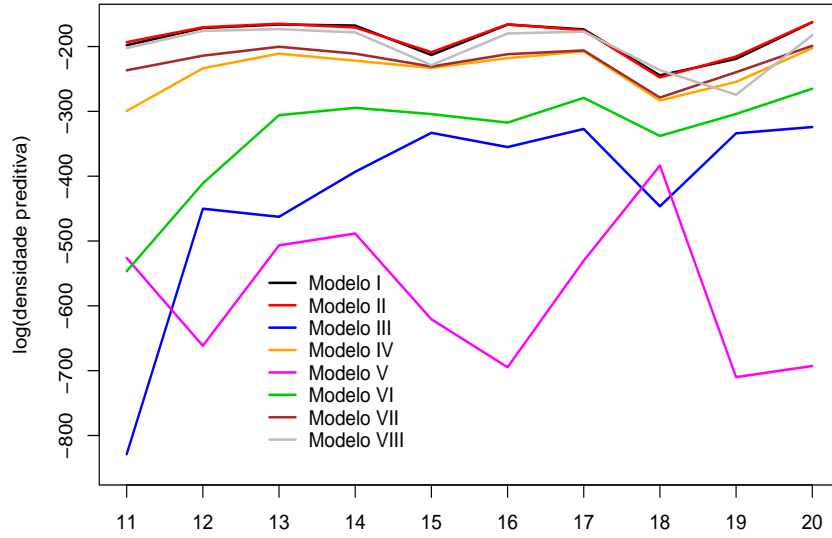


Figura 6.1: ES - Dados na família exponencial: Logaritmo da densidade preditiva.

Tabela 6.2: ES - Família exponencial: Medidas resumo a posteriori dos parâmetros do modelo II.

Parâmetro	Média	Desvio	2.5%	50%	97.5%	\hat{R}
ϕ	6.47	2.38	3.08	6.11	10.2	1.0
τ	1.52	0.44	0.08	1.46	2.51	1.0
κ	0.0003	0.0002	0.00001	0.0003	0.0008	1.0

Na Tabela 6.2 estão os sumários da posteriori dos parâmetros do modelo de contaminação: média, moda, desvio padrão e quantis para construir intervalos de credibilidade de 95% para cada parâmetro. Estes parâmetros representam uma alta dependência espacial entre as inovações e uma moderada precisão. O índice de contaminação κ é muito pequeno, mas a sua inclusão no modelo melhorou a densidade preditiva do modelo I. Também o valor da estatística de Gelman e Rubin para cada parâmetro indica que as cadeias convergiram. Os histogramas das amostras a posteriori desses parâmetros estão na Figura 6.2.

As Figuras 6.3 e 6.4 representam a taxa do risco por 100 mil habitantes, que é uma forma mais natural de interpretar o processo latente $\beta_{(1:T)}$, derivada do modelo através do preditor linear: para cada tempo t e região s , o risco λ_{ts} é igual a $\exp(\theta_{ts})$ e $\theta_t = \exp(\mathbf{F}_t' \beta_t)$,

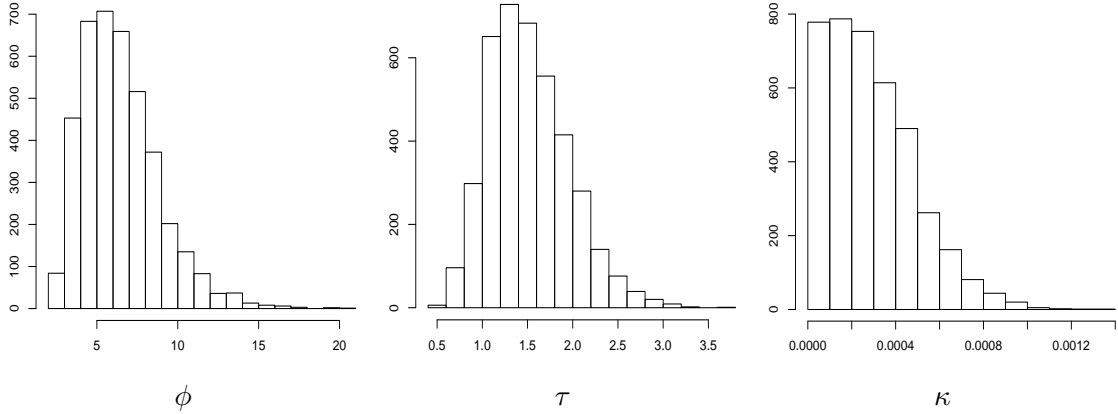


Figura 6.2: ES - Família exponencial: Histogramas das amostras a posteriori dos parâmetros do modelo II.

com $\theta_t = (\theta_{t1}, \dots, \theta_{tS})'$. No caso dos modelos espaço-temporais para dados Gaussianos a interpretação do $\beta_{(1:T)}$ não era direta e os mapas representavam simplesmente a média do processo latente. A escala destes mapas vai desde a cor branca (taxa de risco menor que 5 por 100 mil habitantes) até o vermelho obscuro (taxa de risco de mais de 60 por 100 mil habitantes). A taxa de risco inicial para todos os municípios está entre 10 e 20 por 100 mil. Nos anos seguintes a taxa aumenta principalmente nos municípios do interior do Estado. No ano 1987 a taxa de município de Vitória está entre 20 e 30 por 100 mil e no ano seguinte cresce para 30-40 por 100 mil. Chegando ao final do período de estudo, a taxa para Vitória chegou ao nível máximo da escala, embora em muitos outros municípios a taxa de risco de morte por homicídio diminuiu e em alguns manteve-se constante.

Os mapas das Figuras 6.5 e 6.6 representam a média a posteriori dos campos das inovações para cada ano de estudo. Estes mapas podem ser muito úteis pois eles ilustram os impactos espacialmente estruturados para cada ano específico depois de retirar o efeito da contaminação. As cores indicam forte, moderado e pequeno impactos positivos (tons vermelhos) e negativos (tons azuis). A cor branca indica um impacto quase nulo, seja positivo ou negativo. Podemos destacar o mapa do ano 1980 no qual os efeitos negativos aparecem nas zonas norte e sul do Estado, enquanto na parte central muitos municípios mantiveram-se constantes ou apresentaram impactos positivos. Nos anos 1983, 1984 e 1986

a maior parte dos municípios das zonas norte e central teve um efeito positivo, e no ano 1988 estes efeitos positivos ocuparam as zonas sul e central, incluindo a capital, Vitória. Na década dos 90, com exceção dos anos 1990, 1993 e 1998, a maior parte dos municípios teve impactos nulos ou negativos no risco. No último ano de estudo, os municípios da zona sul e alguns da parte central do Estado incluindo Vitória e seus vizinhos, têm efeitos positivos. Nos municípios restantes o impacto dos efeitos foi quase nulo.

Neste capítulo apresentamos uma aplicação do nosso modelo espaço-temporal generalizado dinâmico para dados de Poisson, o número de homicídios nos municípios do Espírito Santo. O modelo geral para dados de contagem foi definido e vários outros modelos derivados foram ajustados a esses dados de violência para testar o procedimento de inferência Bayesiana que incluiu o algoritmo denominado *EFIFBS*, uma variante do *FIFBS*, para estimar o processo latente. O método de seleção de modelos baseado na densidade preditiva foi aplicado ao conjunto de modelos propostos. Para o modelo escolhido, o modelo de contaminação, mostramos os resultados da inferência dos parâmetros e do processo latente, na forma de mapas que ilustram a taxa do risco de cada município por 100 mil habitantes.

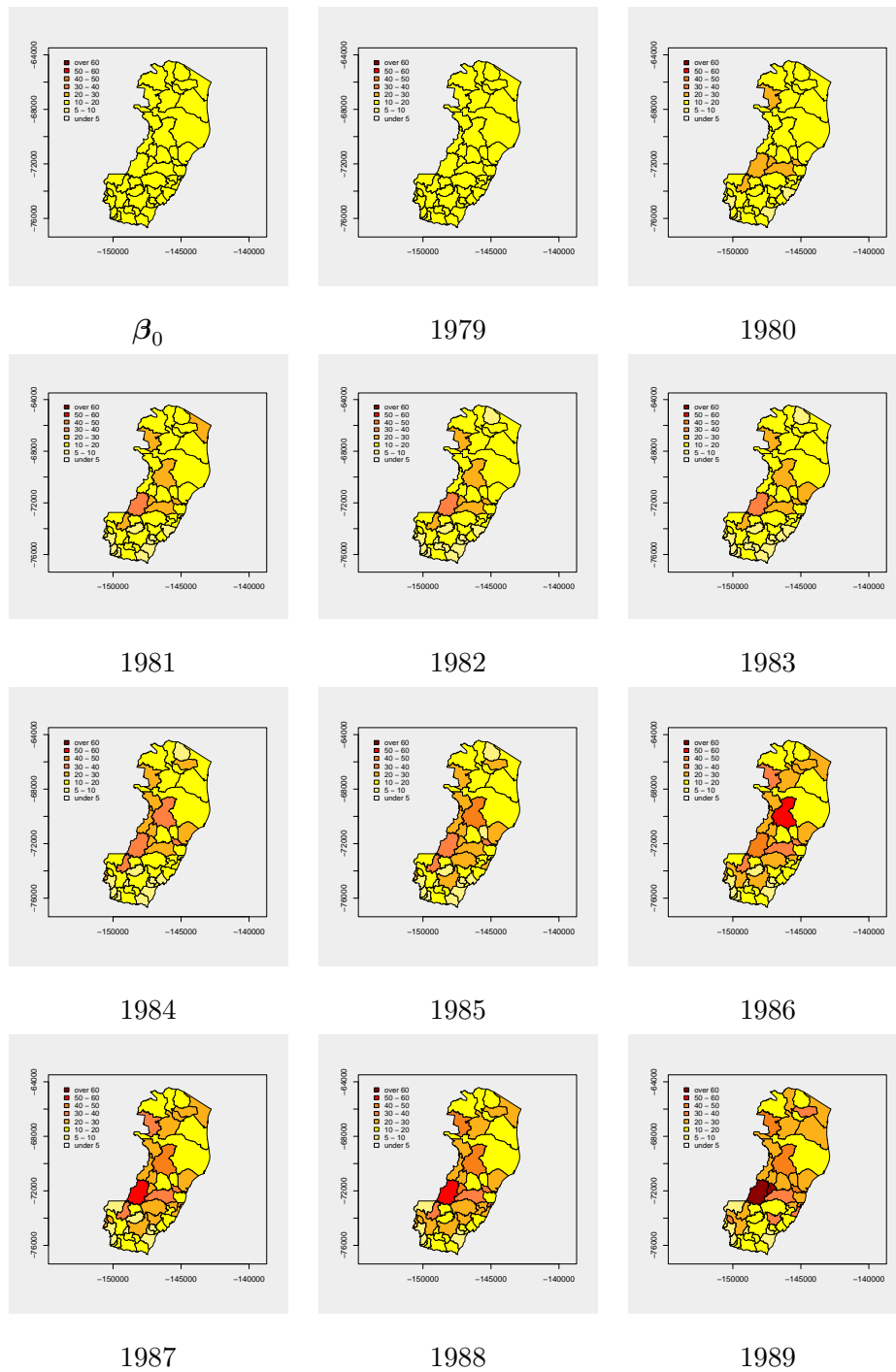


Figura 6.3: ES - Modelo II: Mapas representando a taxa de risco por 100 mil habitantes. Estado inicial β_0 e anos 1979-1989.

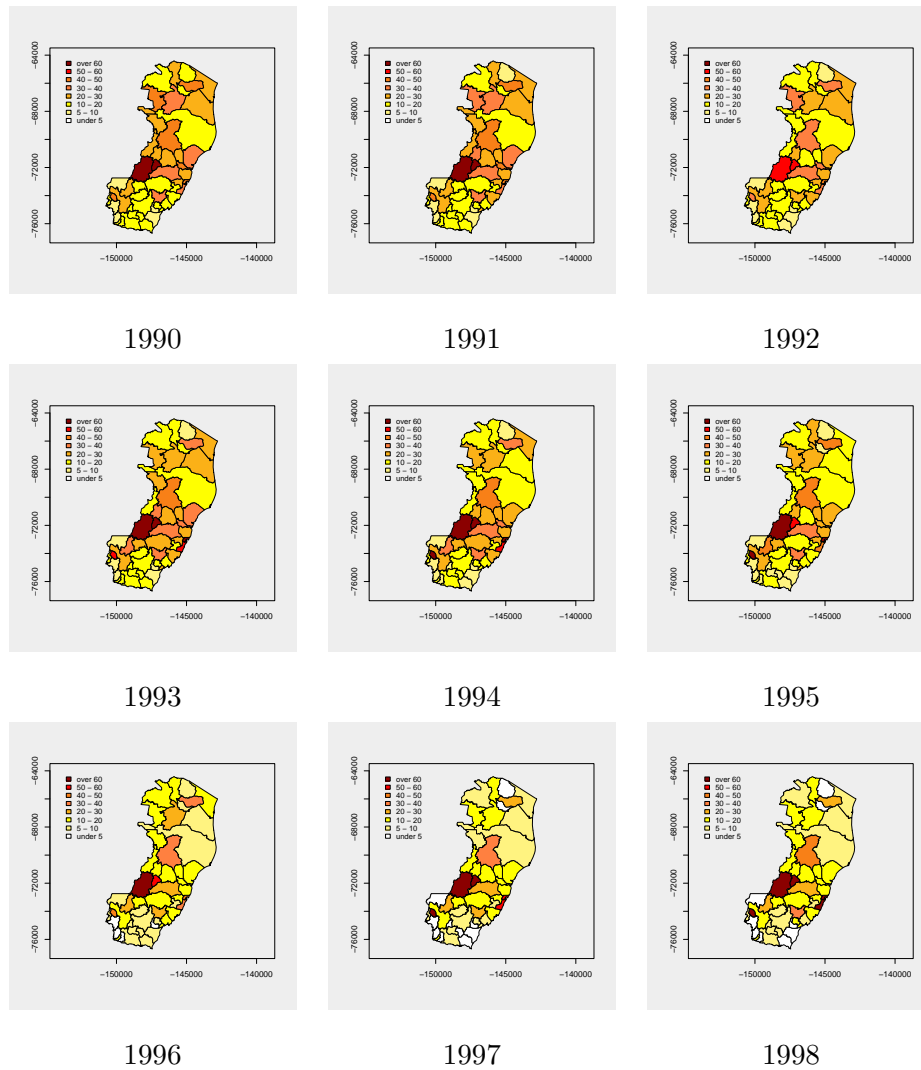


Figura 6.4: ES - Modelo II: Mapas representando a taxa de risco por 100 mil habitantes. Anos 1990-1998.

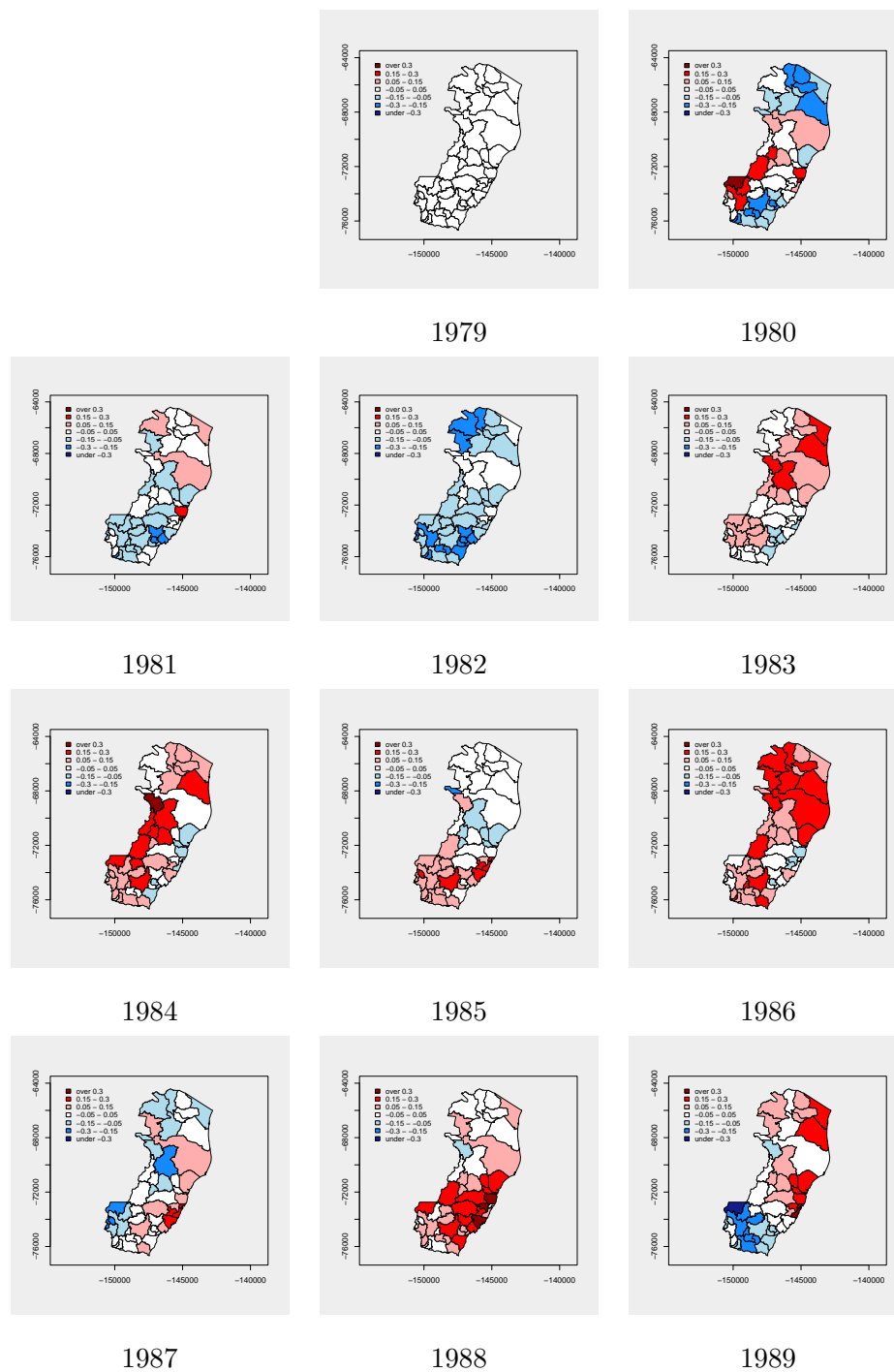


Figura 6.5: ES - Modelo II: Mapas representando as estimativas das médias a posteriori das inovações. Anos 1979-1989.

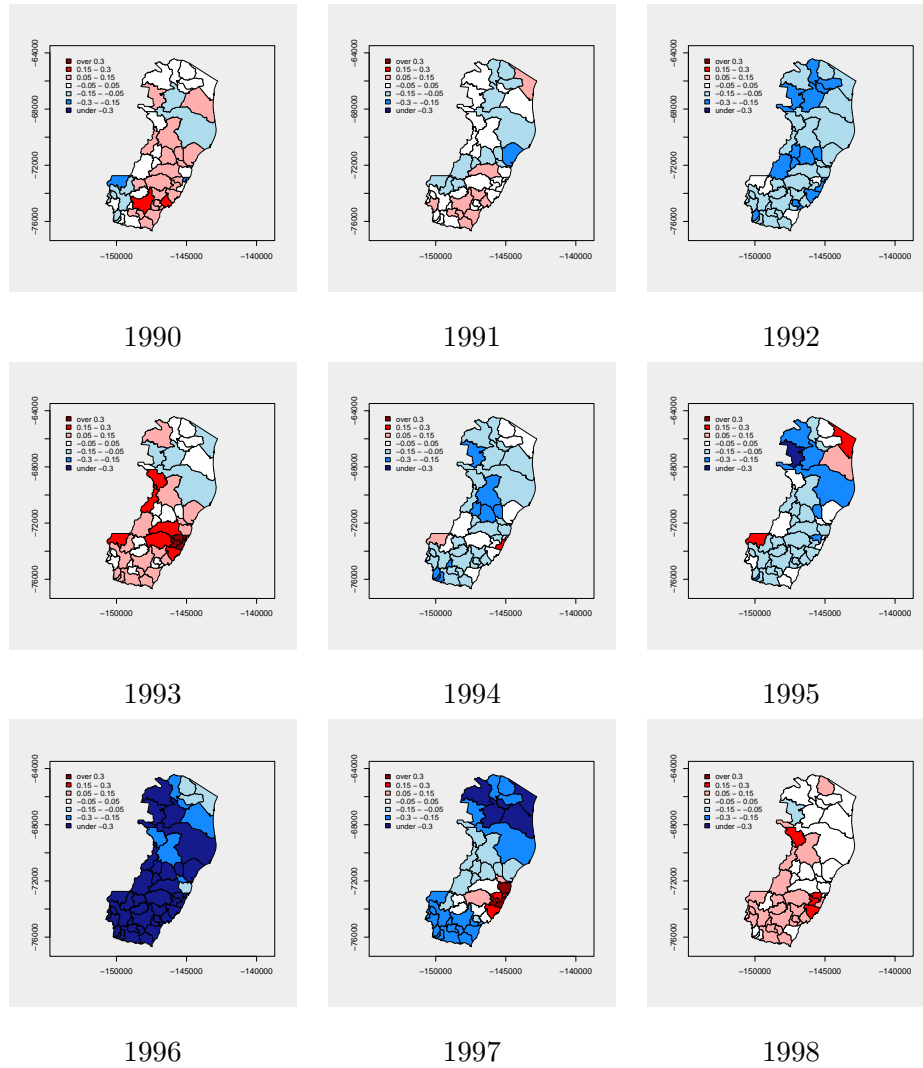


Figura 6.6: RJ - Modelo II: Mapas representando as estimativas das médias a posteriori das inovações. Anos 1990-1998.

Capítulo 7

Aplicação da classe de modelos para dados de área: Caso Bernoulli

7.1 Introdução

Apresentamos uma segunda aplicação do nosso modelo linear generalizado dinâmico. Os dados consistem na dispersão de certa ave, a *Eurasian Collared-Dove* nos Estados Unidos. Como é explicado na próxima seção, os dados originais consistem no número de aves registrado em certos locais de observação. Estes dados foram convertidos em dados 0 – 1, para serem modelados com a distribuição de Bernoulli. O modelo geral para dados de natureza 0 – 1 encontra-se na Seção 7.3 e o conjunto particular de modelos propostos para este problema está especificado na Seção 7.4. Um exemplo com dados simulados com especificações similares aos dados reais é desenvolvido na Seção 7.5. O procedimento de inferência Bayesiana para os parâmetros e o processo latente, utilizando os dados reais, está explicado na Seção 7.6. Finalmente, na Seção 7.7, são apresentados os resultados da seleção do melhor modelo.

7.2 Dispersão da *Eurasian Collared-Dove* nos EUA

A Eurasian Collared-Dove (abreviatura em inglês, *ECD*, nome científico: *Streptopelia decaocto*) é uma espécie de ave, originária da Ásia, que migrou para Europa na primeira metade

do século passado e depois de décadas passou das Bahamas para o Estado da Flórida, nos Estados Unidos, observando-se os primeiros exemplares na metade da década do 80. Está considerada como uma espécie invasora e como uma potencial ameaça aos ecossistemas nativos dos EUA, devido a sua adaptação e dispersão, em especial pela região sudeste dos Estados Unidos, e também pelo seu rápido crescimento populacional. (Figura 7.1)

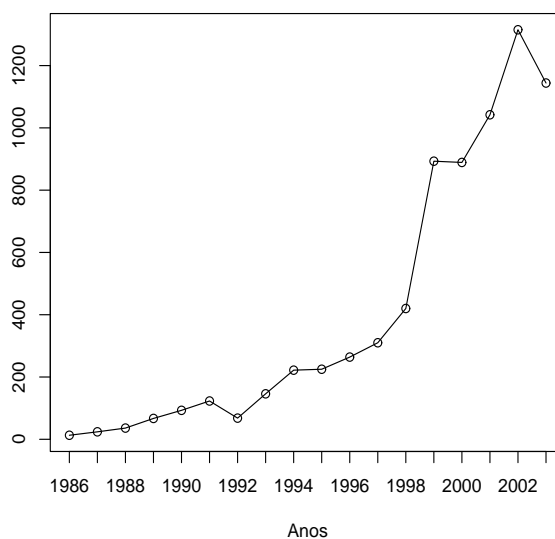


Figura 7.1: População total da *ECD* nos Estados Unidos no período 1986-2003.

Os dados sobre a *ECD* foram obtidos de uma pesquisa norte-americana denominada *North American Breeding Bird Survey* e foram coletados por observadores voluntários durante cada período ao longo de determinadas rotas. As unidades amostrais da pesquisa são rotas de aproximadamente 40 km, que estão na borda das estradas, ao longo das quais o observador faz 50 paradas e registra o número de aves que observa e escuta por um período de três minutos. Existem mais de 4000 rotas no estudo, mas não todas são percorridas a cada ano. Além disso, existe uma grande incerteza para estas observações, dadas as diferenças em experiência e habilidade dos observadores voluntários. No caso da *ECD*, esta incerteza está incrementada pelo fato de que estas aves são muito parecidas com outra espécie. Embora existam diferenças fundamentais nas respectivas aparências e cantos, os observadores as confundem com frequência. Provavelmente este foi o caso nos primeiros anos da invasão,

quando os observadores tinham menos experiência em distinguir entre as duas espécies.

Consideramos 18 anos da pesquisa, desde 1986 até 2003. A Figura 7.2(a) indica os 253 locais onde os dados foram coletados durante o período de estudo. Nosso objetivo com estes dados é aplicar o nosso modelo espaço-temporal generalizado dinâmico para dados de área, assim, o primeiro passo é dividir o território em regiões. Esta divisão é realizada construindo uma grade regular. Várias configurações diferentes foram testadas e finalmente uma grade de 13×17 foi selecionada (Figura 7.2(b)). Cada uma das regiões da grade agrega diferentes quantidades de locais.

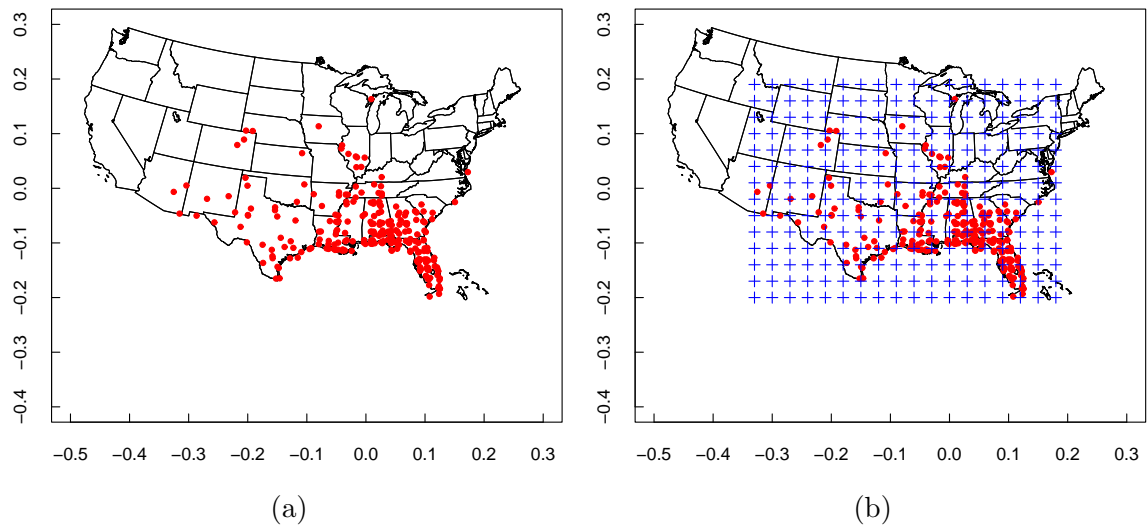


Figura 7.2: (a) Locais onde os dados foram coletados durante o período de estudo (1986-2003) e (b) Grade regular utilizada para agregar espacialmente os locais (13 linhas e 17 colunas).

Os dados dos locais agregados para cada área s representam o número de *ECD* registrados para cada tempo t . Estes podem ser modelados como no Capítulo 6, mas a nossa abordagem é considerar dados do tipo Bernoulli, isto é, registrar para cada área se a *ECD* está presente (codificado como 1) ou ausente (codificado como 0). Esta codificação é feita da seguinte forma: para cada região s e tempo t , se foi registrado um número positivo da *ECD* considera-se “presença: $y_{ts} = 1$ ”, e se nenhuma *ECD* foi observada ou a região s não foi selecionada naquele ano t , o nosso dado será “ausência: $y_{ts} = 0$ ”. Esta modelagem pode ser de interesse devido à natureza invasora da espécie, para conhecer seu comportamento, em que regiões

está e para quais regiões vai se deslocar.

Então, para cada região s e ano t , $s = 1, \dots, S = 221$, $t = 1, \dots, T = 18$, temos o valor $y_{ts} = 0$ (ausência da *ECD*) ou $y_{ts} = 1$ (presença da *ECD*). Assume-se que dados de natureza 0 – 1 têm distribuição de Bernoulli, com uma certa probabilidade π_{ts} , isto é

$$y_{ts}|\pi_{ts} \sim Ber(\pi_{ts}).$$

A distribuição de Bernoulli também pertence à família exponencial, pois escrevemos sua função de densidade de probabilidade na forma da família exponencial (Equação (5.1)):

$$\begin{aligned} (y_{ts}|\pi_{ts}) &\sim Ber(\pi_{ts}) \\ p(y_{ts}|\pi_{ts}) &= \pi_{ts}^{y_{ts}} (1 - \pi_{ts})^{(1-y_{ts})} \\ p(y_{ts}|\pi_{ts}) &= \exp \{y_{ts} \log(\pi_{ts}) + (1 - y_{ts}) \log(1 - \pi_{ts})\} \\ &= \exp \{y_{ts} \log(\pi_{ts}) + \log(1 - \pi_{ts}) - y_{ts} \log(1 - \pi_{ts})\} \\ &= \exp \left\{ y_{ts} \log \left(\frac{\pi_{ts}}{1 - \pi_{ts}} \right) + \log(1 - \pi_{ts}) \right\}. \end{aligned} \quad (7.1)$$

7.3 Modelo geral

Da Equação (7.1) derivamos as seguintes componentes da densidade da família exponencial: o parâmetro natural

$$\eta_{ts} = \log \left(\frac{\pi_{ts}}{1 - \pi_{ts}} \right) \rightarrow \pi_{ts} = \frac{\exp(\eta_{ts})}{1 + \exp(\eta_{ts})},$$

as funções $c(y_{ts}) = 0$ e

$$\begin{aligned} b(\eta_{ts}) &= -\log(1 - \pi_{ts}) = -\log \left(1 - \frac{\exp(\eta_{ts})}{1 + \exp(\eta_{ts})} \right) = -\log \left(\frac{1}{1 + \exp(\eta_{ts})} \right) \\ b(\eta_{ts}) &= \log(1 + \exp(\eta_{ts})), \end{aligned}$$

e as seguintes identidades:

- A média da distribuição $\mu_{ts} = b'(\eta_{ts}) = \exp(\eta_{ts}) / (1 + \exp(\eta_{ts})) = \pi_{ts}$.

- A variância da distribuição

$$\begin{aligned}\Sigma_{ts} &= b''(\eta_{ts}) = \frac{\exp(\eta_{ts})(1 + \exp(\eta_{ts})) - (\exp(\eta_{ts}))^2}{(1 + \exp(\eta_{ts}))^2} \\ &= \frac{\exp(\eta_{ts})}{(1 + \exp(\eta_{ts}))^2} \\ \Sigma_{ts} &= \pi_{ts}(1 - \pi_{ts}).\end{aligned}$$

- O preditor linear: $\theta_{ts} = \text{logit}(\pi_{ts}) = \eta_{ts}$.
- A função de ligação: $g(\mu_{ts}) = g(\pi_{ts}) = \text{logit}(\pi_{ts}) = \theta_{ts}$.
- A função resposta: $f(\theta_{ts}) = \frac{\exp(\theta_{ts})}{(1 + \exp(\theta_{ts}))}$ e a primeira derivada

$$f'(\theta_{ts}) = \frac{\exp(\theta_{ts})}{(1 + \exp(\theta_{ts}))^2}.$$

- A função $\delta(\theta_{ts})$: $\delta(\theta_{ts}) = \theta_{ts} = \eta_{ts}$.

Nosso modelo espaço-temporal generalizado dinâmico geral para dados binários fica definido da seguinte forma

$$\begin{aligned}p(y_{ts}|\eta_{ts}) &= \exp\{y_{ts}\eta_{ts} - \log(1 + \exp(\eta_{ts}))\}, & (7.2) \\ \eta_{ts} &= \theta_{ts}, \\ \boldsymbol{\theta}_t &= \mathbf{F}'_t \boldsymbol{\beta}_t, & \boldsymbol{\theta}_t = (\theta_{t1}, \dots, \theta_{tS})' \\ \boldsymbol{\beta}_t &= \mathbf{G}_t \boldsymbol{\beta}_{t-1} + \boldsymbol{\omega}_t, & \boldsymbol{\omega}_t \sim \text{CAMGP}(\mathbf{0}, \mathbf{W}_t^{-1}),\end{aligned}$$

no qual $\boldsymbol{\beta}_t = (\boldsymbol{\beta}'_{t1}, \dots, \boldsymbol{\beta}'_{tS})'$ é o processo latente sobre o qual deseja-se fazer inferência; a matriz \mathbf{F}_t conecta o processo latente com o preditor linear $\boldsymbol{\theta}_t$; \mathbf{G}_t é a matriz da evolução espaço-temporal; $\boldsymbol{\omega}_t$ é o campo de inovações vetorizado da equação de sistema e \mathbf{W}_t^{-1} descreve a estrutura espacial do vetor das inovações ou erros de sistema. Uma abordagem hierárquica Bayesiana para o conjunto de dados original, o número de aves registradas em cada local, encontra-se em [Wikle e Hooten \(2006\)](#) e [Hooten e Wikle \(2007\)](#).

7.4 Modelos espaço-temporais propostos

A Figura 7.3 mostra as regiões nas quais foi registrada a presença da *Eurasian Collared-Dove* (*ECD*) e na Tabela 7.1 está a percentagem relativa de regiões nas quais foi registrada a presença da *ECD* sobre o total de regiões. Com análises preliminares da aplicação de modelos polinomiais de segunda ordem observou-se que a estimação de vários parâmetros ficou instável. Isto deve ter ocorrido por não dispormos de muita informação ao longo do tempo (poucas “presenças”).

Tabela 7.1: Resumo da agregação dos dados: N é o total de regiões com presença de aves e % a percentagem relativa às 221 regiões.

Ano	1986	1987	1988	1989	1990	1991	1992	1993	1994
N	1	1	2	2	1	2	5	7	7
%	0.5	0.5	0.9	0.9	0.5	0.9	2.3	3.2	3.2
Ano	1995	1996	1997	1998	1999	2000	2001	2002	2003
N	10	14	17	18	27	30	37	48	64
%	4.5	6.3	7.7	8.1	12.2	13.6	16.7	21.7	29.0

Para incluir informação da região no instante de tempo anterior na modelagem, modificamos a equação de sistema do modelo espaço temporal generalizado dinâmico (7.2) como segue:

$$\boldsymbol{\beta}_t = \boldsymbol{\beta}_{t-1} + \gamma \mathbf{y}_{t-1} + \mathbf{H}^* \mathbf{y}_{t-1} + \boldsymbol{\omega}_t, \quad \boldsymbol{\omega}_t \sim \text{CAMGP}(\mathbf{0}, \mathbf{W}_t^{-1}), \quad (7.3)$$

na qual γ é um escalar e os elementos da matriz \mathbf{H}^* são

$$(\mathbf{H}^*)_{kl}^* = \begin{cases} 0, & k = l, \\ \nu/(1 + \nu h), & k \in N_l, \\ 0, & \text{caso contrário,} \end{cases} \quad (7.4)$$

em que ν é um parâmetro no intervalo $[0,1]$ que leva informação aos membros da vizinhança N_l da região l e h é o número máximo de vizinhos.

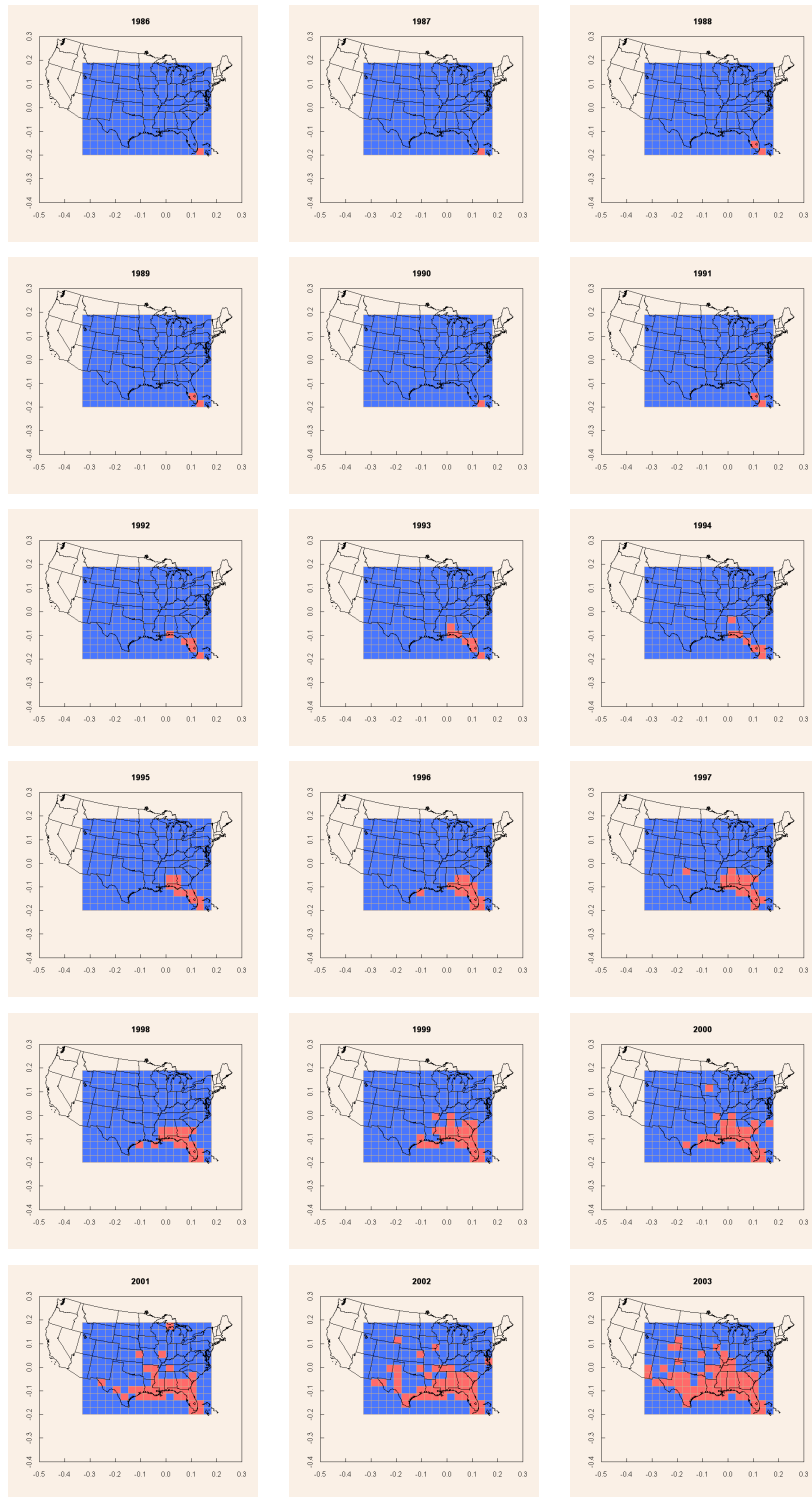


Figura 7.3: Presença (vermelho) e ausência (azul) do *Eurasian Collared-Dove* em cada região da grade ao longo do tempo.

Assumimos uma vizinhança de primeira ordem na grade regular construída, então, o número máximo de vizinhos por região é $h = 4$. Quando $\mathbf{F}_t = \mathbf{I}_S$, cada elemento do vetor latente β_t é igual ao parâmetro canônico η_{ts} , e portanto, diretamente relacionado à probabilidade de sucesso π_{ts} ; isto é, $\beta_{ts} = \eta_{ts} = \text{logit}(\pi_{ts})$. Assim, fica claro que quando a presença da *ECD* é registrada em uma determinada região s em um tempo dado t ($y_{ts} = 1$), pela nova equação de sistema (7.3) o valor do processo latente na mesma região no tempo seguinte é $\beta_{t+1,s} = \beta_{ts} + \gamma + \omega_{ts}$ e para os vizinhos de s , $\beta_{t+1,k} = \beta_{tk} + \gamma y_{tk} + \nu/(1 + 4\nu) + \omega_{tk}$, com $k \in N_s$. Desta forma, a presença em uma região vai levar informação para os vizinhos no tempo seguinte. Se não foi registrada presença ($y_{ts} = 0$), o processo latente no tempo $t + 1$ fica $\beta_{t+1,s} = \beta_{ts} + \omega_{ts}$.

Um modelo similar é elaborado se γ é eliminado da equação de sistema e \mathbf{H}^* é modificado da seguinte forma:

$$(\mathbf{H}^*)_{kl} = \begin{cases} 1/(1 + \nu h), & k = l, \\ \nu/(1 + \nu h), & k \in N_l, \\ 0, & \text{caso contrário.} \end{cases} \quad (7.5)$$

Desta forma, existe uma restrição em como o \mathbf{y}_{t-1} afeta o β_t .

Os modelos propostos para estes dados são:

- I - Modelo de primeira ordem sem estrutura espacial ($\phi = 0$)
- II - Modelo de primeira ordem com estrutura espacial
- III - Modelo de primeira ordem incluindo $\mathbf{H}^* \mathbf{y}_{t-1}$, com \mathbf{H}^* modificado
- IV - Modelo de primeira ordem incluindo $\gamma \mathbf{y}_{t-1} + \mathbf{H}^* \mathbf{y}_{t-1}$.

Resultados preliminares da aplicação destes modelos ao conjunto de dados reais indicaram alguns problemas na estimação dos parâmetros. Decidimos criar um conjunto de dados simulados para testar a nossa metodologia de estimação proposta. Apresentamos este estudo com dados simulados na seção seguinte.

7.5 Dados artificiais

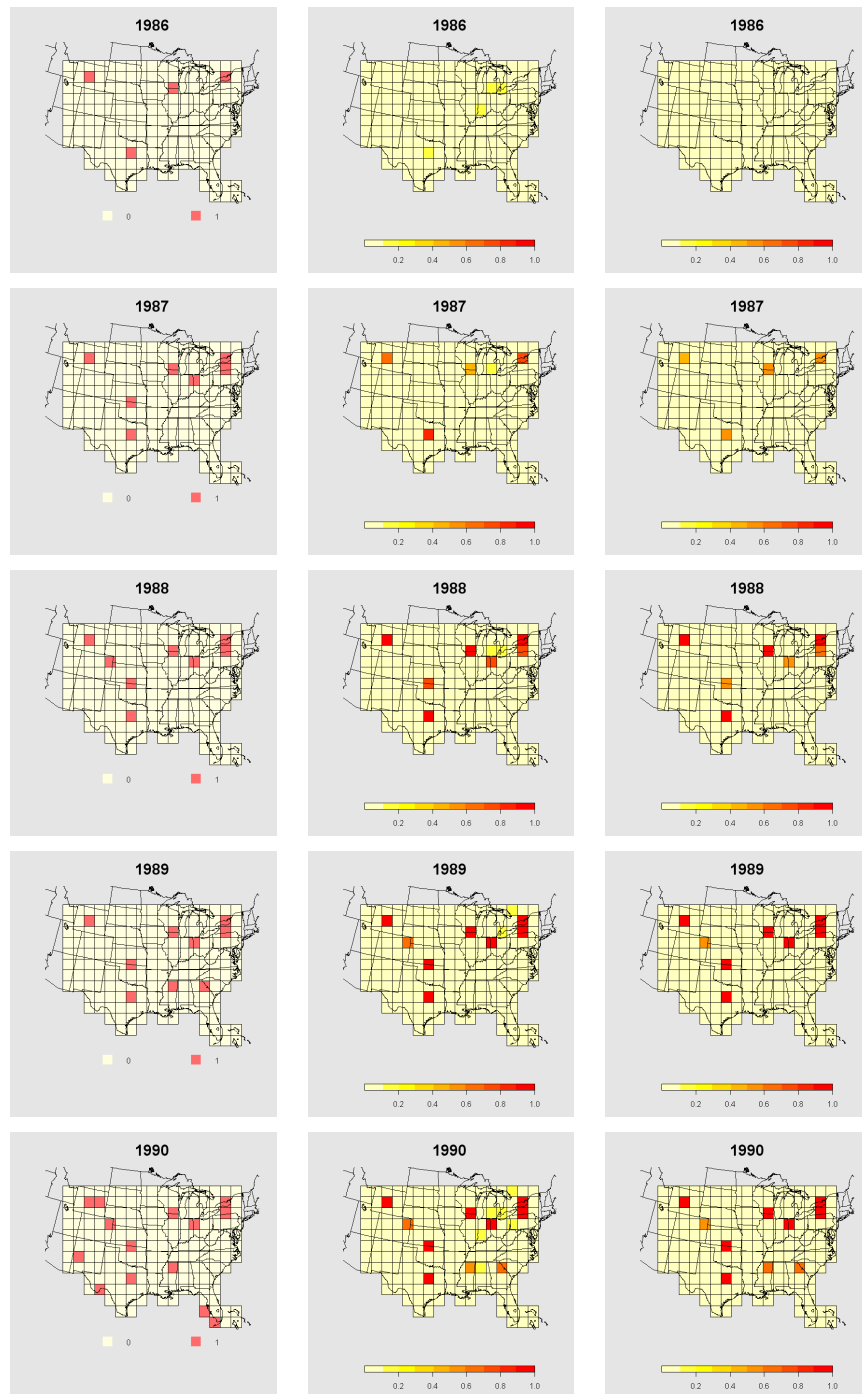
Os dados foram simulados com a mesma estrutura dos dados reais, com a mesma grade de vizinhança e o mesmo período de tempo ($T = 18$). O modelo escolhido para simular os dados foi o modelo IV, de primeira ordem com $\gamma \mathbf{y}_{t-1} + \mathbf{H}^* \mathbf{y}_{t-1}$ na equação de sistema e com $\mathbf{F}_t = \mathbf{G}_t = \mathbf{I}_S$, isto é

$$\begin{aligned} p(y_{ts}|\eta_{ts}) &= \exp\{y_{ts}\eta_{ts} - \log(1 + \exp(\eta_{ts}))\}, & (7.6) \\ \eta_{ts} &= \theta_{ts}, \\ \boldsymbol{\theta}_t &= \boldsymbol{\beta}_t, & \boldsymbol{\theta}_t = (\theta_{t1}, \dots, \theta_{tS})' \\ \boldsymbol{\beta}_t &= \boldsymbol{\beta}_{t-1} + \gamma \mathbf{y}_{t-1} + \mathbf{H}^* \mathbf{y}_{t-1} + \boldsymbol{\omega}_t, & \boldsymbol{\omega}_t \sim \text{CAMGP}(\mathbf{0}, \mathbf{W}_t^{-1}), \end{aligned}$$

e $\mathbf{W}_t^{-1} = \tau(\mathbf{I}_S + \phi \mathbf{M})$, \mathbf{I}_S é a matriz identidade. Além disso, os elementos da matriz \mathbf{H}^* estão definidos como em (7.4)

$$(\mathbf{H}^*)_{kl}^* = \begin{cases} 0, & k = l, \\ \nu/(1 + \nu h), & k \in N_l, \\ 0, & \text{caso contrário.} \end{cases}$$

Os valores do vetor de parâmetros $\boldsymbol{\psi} = (\tau, \phi, \gamma, \nu)$ foram (0.9, 5.0, 4.0, 0.5), respectivamente. Os valores de $\tau = 0.9$ e $\phi = 5.0$, produzem erros $\boldsymbol{\omega}_t$ aproximadamente entre (-1,1). Já $\gamma = 4.0$ e $\nu = 0.5$ vão produzir um efeito crescente no processo latente $\boldsymbol{\beta}_t$. O vetor de médias inicial é negativo $\mathbf{m}_0 = -4.0 \times \mathbf{1}'_S$ para garantir um início de processo com poucos sucessos ($y_{ts} = 1$), similar ao observado com os dados da dispersão da *ECD*. A coluna (a) das Figuras 7.4, 7.5, 7.6, e 7.7, representa o conjunto de dados simulados de “presença” e “ausência” da *ECD* para todos os anos (1986-2003). A coluna (b) das mesmas Figuras mostra as probabilidades de sucesso simuladas π_{ts} para cada tempo t e região s , lembrando que as probabilidades estão diretamente relacionadas ao processo latente $\pi_{ts} = \exp(\beta_{ts})/(1 + \exp(\beta_{ts}))$.



(a)

(b)

(c)

Figura 7.4: *ECD* - Exemplo: (a) Dados simulados, (b) probabilidade simulada e (c) probabilidade estimada para anos 1986-1990.

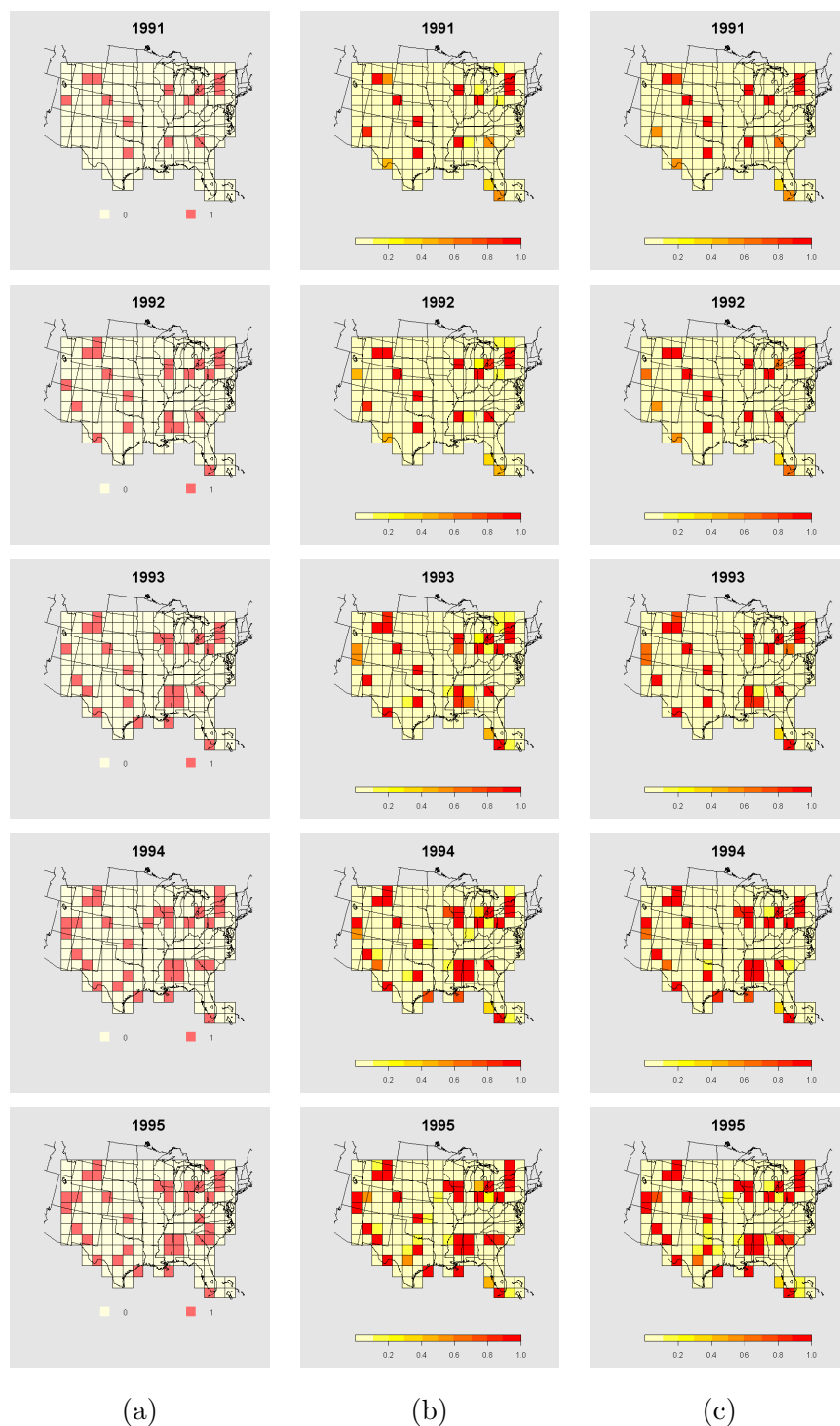
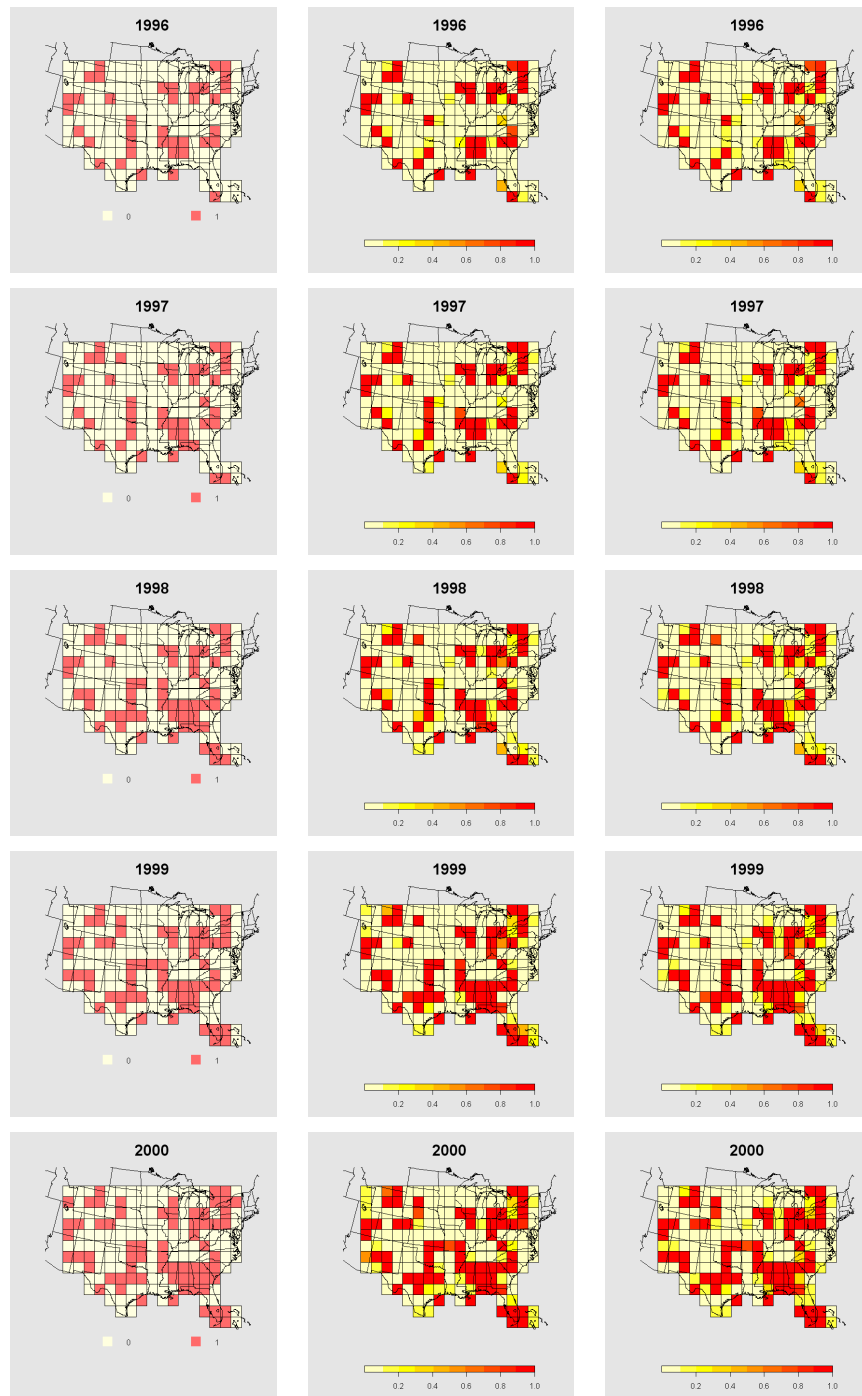


Figura 7.5: *ECD* - Exemplo: (a) Dados simulados, (b) probabilidade simulada e (c) probabilidade estimada para anos 1991-1995.



(a)

(b)

(c)

Figura 7.6: *ECD* - Exemplo: (a) Dados simulados, (b) probabilidade simulada e (c) probabilidade estimada para anos 1996-2000.

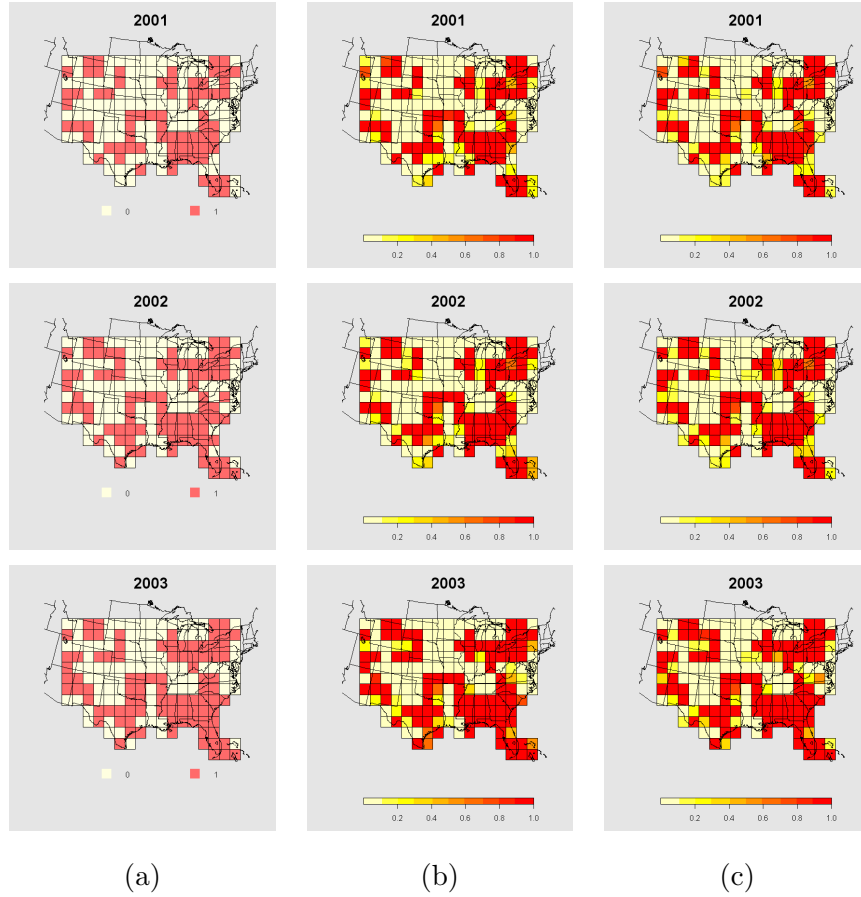


Figura 7.7: *ECD* - Exemplo: (a) Dados simulados, (b) probabilidade simulada e (c) probabilidade estimada para anos 2001-2003.

Inferência Bayesiana

Para completar o modelo de acordo com o paradigma Bayesiano, especificamos as distribuições a priori para os parâmetros do modelo IV, $\psi = (\tau, \phi, \gamma, \nu)$. As prioris para τ e ϕ são semelhantes as utilizadas para o problema do número de homicídios no Espírito Santo, isto é, a priori, $\tau \sim Ga(3, 3)$ e

$$p_0(\phi) \propto \begin{cases} 1, & 0 < \phi < 1 \\ \phi^{-3}, & \phi \geq 1. \end{cases}$$

A priori para γ é $p_0(\gamma) = N(0, 10)$ e para ν , $p_0(\nu) = U(0, 1)$.

O processo latente estimado $\beta_{(1:T)}$ através do *EFIFBS*, transformado em probabilidades estimadas, está representado na coluna (c) das Figuras 7.4, 7.5, 7.6 e 7.7. Com a nossa metodologia conseguimos uma estimação muito razoável do processo latente, embora nos primeiros anos as estimativas tenham pouca variabilidade e subestimem os valores mais altos e superestimem os mais baixos.

Dentro do esquema *MCMC*, foram construídas duas cadeias para obter as amostras a posteriori do vetor de parâmetros ψ e verificamos a convergência com os diagnósticos incluídos na biblioteca *CODA* do *R*. As amostras finais tinham tamanho 4000 e os resumos a posteriori, junto com o critério de Gelman e Rubin para convergência estão na Tabela 7.2. O critério indica que todas as cadeias convergiram. Embora o parâmetro de precisão τ seja superestimado e o parâmetro de dependência espacial ϕ seja subestimado pontualmente, os intervalos de credibilidade de 95% contêm os verdadeiros valores destes parâmetros. E a estimação dos parâmetros γ e ν é bastante aceitável. Os histogramas das amostras a posteriori de cada um destes parâmetros estão ilustrados na Figura 7.8.

7.6 Procedimento de inferência para os dados reais

Depois de testar a metodologia de estimação de parâmetros e o processo latente para o conjunto de dados simulados, vamos examinar os dados reais. Mantendo a suposição de prioris independentes para os parâmetros do modelo IV, a distribuição a posteriori conjunta é proporcional a

$$\begin{aligned}
 p(\beta_{(1:T)}, \psi | \mathbf{D}_T) &\propto \left[\prod_{t=1}^T p(\mathbf{y}_t | \boldsymbol{\theta}_t, \boldsymbol{\eta}_t) \right] \left[\prod_{t=1}^T p(\beta_t | \beta_{t-1}, \psi) \right] p(\psi) \\
 p(\beta_{(1:T)}, \psi | \mathbf{D}_T) &\propto \left[\prod_{t=1}^T p(\mathbf{y}_t | \boldsymbol{\theta}_t, \boldsymbol{\eta}_t) \right] \left[\prod_{t=2}^T p(\beta_t | \beta_{t-1}, \tau, \phi, \gamma, \nu) \right] \\
 &\quad \times p_0(\tau) p_0(\phi) p_0(\gamma) p_0(\nu) p(\beta_1).
 \end{aligned} \tag{7.7}$$

Tabela 7.2: *ECD* - Exemplo simulado: Médias, desvios e quantis a posteriori dos parâmetros. Critério de convergência de Gelman e Rubin, \hat{R} , indica que as cadeias convergiram.

Parâmetro	Média	Desvio	2.5%	50%	97.5%	\hat{R}
$\tau = 0.9$	1.61	0.64	0.67	1.50	3.12	1.0
$\phi = 5.0$	3.60	2.72	0.41	2.92	9.89	1.1
$\gamma = 4.0$	3.81	0.31	3.18	3.83	4.41	1.1
$\nu = 0.5$	0.49	0.16	0.25	0.46	0.88	1.0

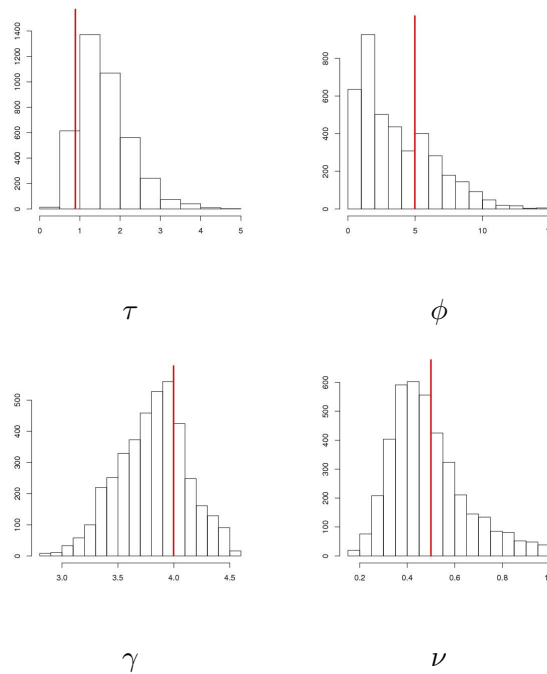


Figura 7.8: *ECD* - Dados simulados: Histogramas das amostras a posteriori dos parâmetros. As linhas vermelhas representam os valores utilizados: $\tau = 0.9$, $\phi = 5.0$, $\gamma = 4.0$ e $\nu = 0.5$.

7.6.1 Estimação do processo latente

Os resultados anteriores do ajuste do modelo IV com dados simulados indicavam que a escassa informação que temos nos primeiros tempos, correspondentes aos primeiros anos da presença da *ECD* nos Estados Unidos, refletia em uma estimação inadequada do processo latente nesses anos. A priori utilizada para β_1 foi a priori de uso comum nos modelos dinâmicos

para o vetor inicial: normal multivariada com vetor de médias igual a uma constante vezes um vetor coluna de uns e uma matriz de precisão (ou covariância) diagonal. Decidimos usar uma priori para o vetor latente inicial que contivesse mais informação extraída dos dados. Uma solução proposta foi utilizar o procedimento de Bayes empírico (Gelman et al., 1995a) para a distribuição a priori de β_1 . O procedimento está esquematizado como segue:

- Proporção de presenças no tempo 1: $\bar{y}_1 = \frac{\sum_{s=1}^S y_{1s}}{S}$
- $\left(\sum_{s=1}^S y_{1s}\right) | \pi_{1s}$ tem distribuição binomial $bin(S, \pi_{1s})$
- Informação observada: $J(\pi_{1s}) = \frac{S}{\bar{y}_1(1-\bar{y}_1)}$
- Aproximação analítica para S grande: $\pi_{1s} | \mathbf{y}_1 \sim N\left(\bar{y}_1, \frac{\bar{y}_1(1-\bar{y}_1)}{S}\right)$
- Como $\beta_{1s} = \theta_{1s} = \log\left(\frac{\pi_{1s}}{1-\pi_{1s}}\right) = g(\pi_{1s})$, utilizamos o método delta para encontrar a distribuição a priori de β_{1s} :

$$E(\beta_{1s} | \mathbf{y}_1) \approx g(\bar{y}_1) = \log\left(\frac{\bar{y}_1}{1-\bar{y}_1}\right) \quad (7.8)$$

$$\begin{aligned} V(\beta_{1s} | \mathbf{y}_1) &\approx \left(\frac{\partial g(\bar{y}_1)}{\partial \pi_{1s}}\right)^2 V(\pi_{1s} | \mathbf{y}_1) \\ &\approx \left(\frac{-1}{\bar{y}_1(1-\bar{y}_1)}\right)^2 \frac{\bar{y}_1(1-\bar{y}_1)}{S} \\ V(\beta_{1s} | \mathbf{y}_1) &\approx \frac{1}{\bar{y}_1(1-\bar{y}_1)S} \end{aligned} \quad (7.9)$$

- Assim, a priori de β_1 é

$$\beta_1 \sim N\left(\log\left(\frac{\bar{y}_1}{1-\bar{y}_1}\right) \times \mathbf{1}_S, \frac{1}{\bar{y}_1(1-\bar{y}_1)S} \times \mathbf{I}_S\right), \quad (7.10)$$

assumindo que todas as regiões têm o mesmo “risco de presença” e que a presença ou ausência em cada região são independentes.

Definida a distribuição a priori para β_1 , utilizamos o algoritmo *EFIFBS* descrito na Subseção 5.5.1 para obter as amostras do processo latente. Para simular da distribuição a posteriori conjunta utilizamos as seguintes quantidades: as observações aproximadas:

$$\begin{aligned}
\hat{y}_{ts} &= [f'(\hat{\theta}_{ts})]^{-1} \{y_{ts} - f(\hat{\theta}_{ts})\} + \hat{\theta}_{ts} \\
&= \frac{(1 + \exp(\hat{\theta}_{ts}))^2}{\exp(\hat{\theta}_{ts})} \left\{ y_{ts} - \frac{\exp(\hat{\theta}_{ts})}{(1 + \exp(\hat{\theta}_{ts}))} \right\} + \hat{\theta}_{ts} \\
&= y_{ts}(\exp(-\hat{\theta}_{ts}))(1 + \exp(\hat{\theta}_{ts}))^2 - (1 + \exp(\hat{\theta}_{ts})) + \hat{\theta}_{ts} \\
\hat{y}_{ts} &= y_{ts}(\exp(-\hat{\theta}_{ts}))(1 + \exp(\hat{\theta}_{ts}))^2 - \exp(\hat{\theta}_{ts}) - 1 + \hat{\theta}_{ts}, \tag{7.11}
\end{aligned}$$

e a matriz de precisão aproximada

$$\begin{aligned}
\hat{\mathbf{V}}_t^{-1} &= \text{diag} \left(\frac{[f'(\hat{\theta}_{ts})]^2}{\hat{\Sigma}_{ts}} \right) \\
&= \text{diag} \left(\frac{[\exp(\eta_{ts})/(1 + \exp(\eta_{ts}))^2]^2}{\exp(\eta_{ts})/(1 + \exp(\eta_{ts}))^2} \right) \\
&= \text{diag} \left(\frac{\exp(\hat{\theta}_{ts})}{(1 + \exp(\hat{\theta}_{ts}))^2} \right) \\
&= \text{diag} \left(1 / \left[\exp(-\hat{\theta}_{ts})(1 + \exp(\hat{\theta}_{ts}))^2 \right] \right) \\
\hat{\mathbf{V}}_t^{-1} &= \text{diag} \left(1 / \left[\exp(-\hat{\theta}_{ts}) + \exp(\hat{\theta}_{ts}) + 2 \right] \right). \tag{7.12}
\end{aligned}$$

O $\hat{\theta}_t$ representa $\hat{\theta}_t = (\hat{\theta}_{t1}, \hat{\theta}_{t2}, \dots, \hat{\theta}_{tS})' = \mathbf{F}'_t \mathbf{a}_t$. Desta forma podemos gerar valores de β_t condicionado no valor β_{t+1} da densidade aproximada $N(\hat{\mathbf{b}}_t, \hat{\mathbf{B}}_t)$. Para o modelo IV, isto é:

$$\mathbf{B}_t^{-1} = \mathbf{C}_t^{-1} + \mathbf{W}_t^{-1} \tag{7.13}$$

$$\mathbf{B}_t^{-1} \mathbf{b}_t = \mathbf{C}_t^{-1} \mathbf{m}_t + \mathbf{W}_t^{-1} (\beta_{t+1} - \gamma \mathbf{y}_t - \mathbf{H}^* \mathbf{y}_t) \tag{7.14}$$

7.6.2 Especificação das prioris dos parâmetros

Dando continuidade ao procedimento de estimação do modelo IV, a seguir apresentamos as distribuições a priori e as condicionais completas de cada parâmetro do modelo. Inicialmente,

as priors para os parâmetros são semelhantes as utilizadas para o exemplo simulado da seção anterior.

ν : A priori para ν é $U(0, 1)$, similar ao caso do parâmetro κ do modelo de contaminação discutido na Seção 4.6

$$p(\nu | \dots) \propto \left[\prod_{t=1}^T p(\boldsymbol{\beta}_t | \boldsymbol{\beta}_{t-1}, \dots) \right] p_0(\nu)$$

$$p(\nu | \dots) \propto \left[\prod_{t=1}^T \exp \{ -0.5 \boldsymbol{\epsilon}_t' \mathbf{W}_t^{-1} \boldsymbol{\epsilon}_t \} \right] \mathbb{I}_{[0,1]}(\nu)$$

com $\boldsymbol{\epsilon}_t = (\boldsymbol{\beta}_t - \boldsymbol{\beta}_{t-1} - \gamma \mathbf{y}_{t-1} - \mathbf{H}^* \mathbf{y}_{t-1})$. Lembrando que

$$(\mathbf{H}^*)_{kl}^* = \begin{cases} 0, & k = l, \\ \nu / (1 + \nu h), & k \in N_l, \\ 0, & \text{caso contrário,} \end{cases}$$

Como a condicional completa não tem forma conhecida utilizamos o algoritmo de Metropolis-Hastings com uma proposta Normal truncada em $[0, 1]$

$$\nu^{(prop)} \sim N_{[0,1]}(\nu^{old}, \Delta_\nu)$$

com Δ_ν de forma tal que a probabilidade de aceitação seja razoável. Assim, o $\nu^{(prop)}$ é aceito com probabilidade

$$A = \min \left\{ 1, \frac{p(\nu^{(prop)} | \dots) \left[\Phi \left(\frac{1 - \nu^{(old)}}{\sqrt{\Delta_\nu}} \right) - \Phi \left(\frac{0 - \nu^{(old)}}{\sqrt{\Delta_\nu}} \right) \right]}{p(\nu^{(old)} | \dots) \left[\Phi \left(\frac{1 - \nu^{(prop)}}{\sqrt{\Delta_\nu}} \right) - \Phi \left(\frac{0 - \nu^{(prop)}}{\sqrt{\Delta_\nu}} \right) \right]} \right\}.$$

γ : A priori para γ é $p_0(\gamma) = N(u_0, v_0)$ e a condicional completa é também normal

$$\begin{aligned}
p(\gamma|\dots) &\propto \left[\prod_{t=1}^T p(\boldsymbol{\beta}_t|\boldsymbol{\beta}_{t-1}, \dots) \right] p_0(\gamma) \\
&\propto \left[\prod_{t=1}^T \exp \left\{ -0.5 \boldsymbol{\epsilon}_t' \mathbf{W}_t^{-1} \boldsymbol{\epsilon}_t \right\} \right] \left[\exp \left\{ -\frac{1}{2v_0} (\gamma - u_0)^2 \right\} \right] \\
&\propto \exp \left\{ -\frac{1}{2} \left[\sum_{t=1}^T (\gamma^2 \mathbf{y}'_{t-1} \mathbf{W}_t^{-1} \mathbf{y}_{t-1} \right. \right. \\
&\quad \left. \left. - 2\gamma (\boldsymbol{\beta}_t - \boldsymbol{\beta}_{t-1} - \mathbf{H}^* \mathbf{y}_{t-1})' \mathbf{W}_t^{-1} \mathbf{y}_{t-1} \right) \right] \times \frac{(\gamma - u_0)^2}{v_0} \Big\} \\
&\propto \exp \left\{ -\frac{1}{2} \left[\gamma^2 \left(\left(\sum_{t=1}^T \mathbf{y}'_{t-1} \mathbf{W}_t^{-1} \mathbf{y}_{t-1} \right) + v_0^{-1} \right) \right. \right. \\
&\quad \left. \left. - 2\gamma \left(\sum_{t=1}^T (\boldsymbol{\beta}_t - \boldsymbol{\beta}_{t-1} - \mathbf{H}^* \mathbf{y}_{t-1})' \mathbf{W}_t^{-1} \mathbf{y}_{t-1} \right) + \frac{u_0}{v_0} \right] \right\} \\
p(\gamma|\dots) &\sim N(u_1, v_1)
\end{aligned}$$

com $\boldsymbol{\epsilon}_t = (\boldsymbol{\beta}_t - \boldsymbol{\beta}_{t-1} - \gamma \mathbf{y}_{t-1} - \mathbf{H}^* \mathbf{y}_{t-1})$ e variância e média

$$\begin{aligned}
v_1 &= \left(\left(\sum_{t=1}^T \mathbf{y}'_{t-1} \mathbf{W}_t^{-1} \mathbf{y}_{t-1} \right) + v_0^{-1} \right)^{-1} \\
u_1 &= v_1 \left[\left(\sum_{t=1}^T (\boldsymbol{\beta}_t - \boldsymbol{\beta}_{t-1} - \mathbf{H}^* \mathbf{y}_{t-1})' \mathbf{W}_t^{-1} \mathbf{y}_{t-1} \right) + \frac{u_0}{v_0} \right].
\end{aligned}$$

ϕ : A priori para ϕ é a mesma utilizada para o problema com os dados simulados

$$p_0(\phi) \propto \begin{cases} 1, & 0 < \phi < 1 \\ \phi^{-b}, & \phi \geq 1 \end{cases}$$

com o expoente $b = 2.5$. Assim, a distribuição condicional completa de ϕ é

$$\begin{aligned}
p(\phi|\dots) &\propto \left[\prod_{t=1}^T p(\boldsymbol{\beta}_t|\boldsymbol{\beta}_{t-1}, \dots) \right] p_0(\phi) \\
&\propto \left[\prod_{t=1}^T |\mathbf{W}_t|^{-0.5} \exp \{ -0.5 \boldsymbol{\epsilon}_t' \mathbf{W}_t^{-1} \boldsymbol{\epsilon}_t \} \right] p_0(\phi) \\
&\propto \left[\prod_{t=1}^T |\tau(\mathbf{I} + \phi \mathbf{M})|^{0.5} \exp \{ -0.5 \boldsymbol{\epsilon}_t' (\tau(\mathbf{I} + \phi \mathbf{M})) \boldsymbol{\epsilon}_t \} \right] p_0(\phi) \\
&\propto \left[\prod_{t=1}^T |(\mathbf{I} + \phi \mathbf{Z})|^{0.5} \exp \{ -0.5 \tau \phi \boldsymbol{\epsilon}_t' \mathbf{M} \boldsymbol{\epsilon}_t \} \right] p_0(\phi) \\
&\propto \left[\prod_{t=1}^T \left(\prod_{i=1}^S (\phi \lambda_i + 1) \right)^{0.5} \exp \left\{ -\frac{\tau \phi}{2} \boldsymbol{\epsilon}_t' \mathbf{M} \boldsymbol{\epsilon}_t \right\} \right] p_0(\phi) \\
p(\phi|\dots) &\propto \left(\prod_{i=1}^S (\phi \lambda_i + 1) \right)^{T/2} \exp \left\{ -\frac{\tau \phi}{2} \sum_{t=1}^T \boldsymbol{\epsilon}_t' \mathbf{M} \boldsymbol{\epsilon}_t \right\} p_0(\phi)
\end{aligned}$$

com $\boldsymbol{\epsilon}_t = (\boldsymbol{\beta}_t - \boldsymbol{\beta}_{t-1} - \gamma \mathbf{y}_{t-1} - \mathbf{H}^* \mathbf{y}_{t-1})$. Os λ_i são os autovalores da matriz \mathbf{M} e $\mathbf{Z} = \text{diag}(\lambda_1, \dots, \lambda_S)$.

Como $p(\phi|\dots)$ não tem forma fechada, usamos o algoritmo de Metropolis-Hastings com uma proposta para ϕ como segue: $\phi^{(prop)} = \exp(\tilde{\phi}^{(prop)})$, onde $\exp(\tilde{\phi}^{(prop)}) \sim U[\log(\phi^{(old)}) - \Delta_\phi, \log(\phi^{(old)}) + \Delta_\phi]$, com $\Delta_\phi > 0$ como parâmetro de sintonia. Assim, o $\phi^{(prop)}$ é aceito com probabilidade

$$A = \min \left\{ 1, \frac{p(\phi^{(prop)}|\dots)\phi^{(prop)}}{p(\phi^{(old)}|\dots)\phi^{(old)}} \right\}.$$

τ : A primeira distribuição a priori especificada para o parâmetro de precisão foi gama $Ga(a_0, b_0)$ para diversas combinações de (a_0, b_0) . Desta forma a condicional completa derivada de (7.7) é gama também.

$$\begin{aligned}
p(\tau|\dots) &\propto \left[\prod_{t=1}^T p(\boldsymbol{\beta}_t|\boldsymbol{\beta}_{t-1},\dots) \right] p_0(\tau) \\
&\propto \left[\prod_{t=1}^T |\mathbf{W}_t|^{-0.5} \exp \{ -0.5 \boldsymbol{\epsilon}_t' \mathbf{W}_t^{-1} \boldsymbol{\epsilon}_t \} \right] \tau^{a_0-1} \exp(-b_0\tau) \\
&\propto \left[\prod_{t=1}^T |\tau(\mathbf{I} + \phi\mathbf{M})|^{0.5} \exp \{ -0.5 \boldsymbol{\epsilon}_t' (\tau(\mathbf{I} + \phi\mathbf{M})) \boldsymbol{\epsilon}_t \} \right] \tau^{a_0-1} \exp(-b_0\tau) \\
&\propto \tau^{S.T/2} \exp \left\{ -\frac{\tau}{2} \sum_{t=1}^T \boldsymbol{\epsilon}_t' (\mathbf{I} + \phi\mathbf{M}) \boldsymbol{\epsilon}_t \right\} \tau^{a_0-1} \exp(-b_0\tau) \\
p(\tau|\dots) &\propto \tau^{a_0+S.T/2-1} \exp \left\{ -\tau \left[b_0 + \frac{1}{2} \sum_{t=1}^T \boldsymbol{\epsilon}_t' (\mathbf{I} + \phi\mathbf{M}) \boldsymbol{\epsilon}_t \right] \right\} \\
(\tau|\dots) &\sim Ga(a_1, b_1),
\end{aligned}$$

com

$$\begin{aligned}
a_1 &= a_0 + S.T/2 \\
b_1 &= b_0 + \frac{1}{2} \sum_{t=1}^T \boldsymbol{\epsilon}_t' (\mathbf{I} + \phi\mathbf{M}) \boldsymbol{\epsilon}_t
\end{aligned}$$

e $\boldsymbol{\epsilon}_t = (\boldsymbol{\beta}_t - \boldsymbol{\beta}_{t-1} - \gamma\mathbf{y}_{t-1} - \mathbf{H}^*\mathbf{y}_{t-1})$.

Muitas dessas combinações (a_0, b_0) produziam prioris vagas e semi-informativas, mas a estimativa de $\tau \rightarrow 0$ em todos os casos, pois o b_1 aumentava em cada simulação, provocando erros no algoritmo *MCMC*, devido a que a variância das inovações crescia para o infinito. Combinações de (a_0, b_0) que produziam prioris muito informativas foram descartadas.

Aparentemente, a verossimilhança exige condições fortes na priori. Com outra especificação da priori tentamos solucionar este problema, truncando a priori de τ , de forma tal que valores pequenos de τ não sejam permitidos, isto é:

$$p_0(\tau) \propto \begin{cases} \tau^{-1}, & \text{se } \tau > \tau_0, \\ 0, & \text{caso contrário,} \end{cases}$$

e o limite inferior τ_0 foi calculado como segue. Para simplificar, vamos considerar o seguinte modelo (sem dependência espacial):

$$y_{ts}|\pi_{ts} \sim Ber(\pi_{ts})$$

$$\eta_{ts} = \log(\pi_{ts}/(1 - \pi_{ts})) = \alpha + \omega_{ts}, \omega_{ts} \sim N(0, \sigma^2)$$

Assumimos que o parâmetro π_{ts} tem alta probabilidade de estar no intervalo (0.01, 0.99), porque com estes dados em particular é difícil diferenciar $\pi_{ts} = 0.005$ e $\pi_{ts} = 0.01$. Então, seja $P(\pi_{ts} \in (0.01, 0.99)) = 0.95$. Isto implica que a probabilidade $P(\eta_{ts} \in (-4.6, 4.6))$ é aproximadamente igual a 0.95. Digamos que $\alpha \sim N(0, 1)$. A estrutura espacial vai permitir valores absolutos de η_{ts} bem maiores que 4.6, mas esta suposição sobre α ajuda na simplificação. Então,

$$\sqrt{(\sigma^2 + 1)} < 4.6/2 \longrightarrow \sigma^2 < 4.29$$

Esta condição foi aproximada para $\sigma^2 < 5$. Portanto, o parâmetro de precisão $\tau = \sigma^{-2} > 1/5 = 0.2$.

Com uma condição menos conservadora, por exemplo, se $P(\pi_{ts} \in (0.01, 0.99)) = 0.99$ e sem arredondar o limite superior para σ^2 , o limite para τ seria $\tau > 0.45$. Sendo mais estritos com as exigências para π_{ts} , o limite inferior para τ seria ainda maior.

Com esta nova priori, a condicional completa do τ resulta uma gama truncada inferiormente por τ_0 . Testamos diferentes valores para $\tau_0 = \{0.20, 0.45, 0.70 \text{ e } 1.0\}$ e erros aconteceram nas iterações iniciais, pois os valores simulados de τ ficaram muito próximos de τ_0 , provocando instabilidade numérica. Optamos por não estimar e fixar o valor de τ em $\{0.20, 0.45, 0.70 \text{ e } 1.0\}$ e incluir modelos com diferentes valores de τ para fazer a seleção.

Utilizando os métodos *MCMC*, geramos cadeias de tamanho 5500 e com um *burn-in* de 1000 para todos os modelos. Assim, as amostras finais têm tamanho 4500. Estes códigos estão implementados em *Ox*. Verificamos a convergência das cadeias utilizando os diagnósticos citados na Subseção 6.5.2.

7.7 Resultados da seleção de modelos

Aplicando o método de seleção de modelos baseado na densidade preditiva da Seção 5.6 para os modelos polinomiais de primeira ajustados:

I(a) - Sem estrutura espacial e $\tau = 0.20$

I(b) - Sem estrutura espacial e $\tau = 0.45$

I(c) - Sem estrutura espacial e $\tau = 0.70$

I(d) - Sem estrutura espacial e $\tau = 1.0$

II - Com estrutura espacial e $\tau = 0.20$

III - Incluindo $\mathbf{H}^*\mathbf{y}_{t-1}$, com \mathbf{H}^* modificado e $\tau = 0.20$

IV - Incluindo $\gamma\mathbf{y}_{t-1} + \mathbf{H}^*\mathbf{y}_{t-1}$ e $\tau = 0.20$

A amostra piloto considerada foi $t^* = 13$, pois com menores valores a convergência do parâmetro da correlação espacial ϕ ficou prejudicada. A Tabela 7.3 mostra o logaritmo da densidade preditiva para o conjunto de modelos ajustados. Para cada modelo m , temos

$$\log p_m(\mathbf{y}_{14}, \dots, \mathbf{y}_{18} | \mathbf{D}_{13}) = \sum_{t=14}^{18} \log p_m(\mathbf{y}_t | \mathbf{D}_{t-1}), \quad (7.15)$$

O critério de seleção indicou o modelo I(a), o polinomial de primeira ordem sem estrutura espacial com $\tau = 0.20$ como o melhor, embora o modelo II que considera a dependência espacial e também com $\tau = 0.20$ tenha um comportamento similar ao longo dos tempos. Isto pode ser observado na Figura 7.9. As Figuras 7.10 e 7.11 mostram as probabilidades estimadas através do processo latente β_t , a sua vez estimado utilizando o algoritmo *EFIFBS* para todos os anos. Observa-se uma saturação depois de alguns anos, pois as regiões que aparentemente eram ideais para a proliferação da *ECD* do ponto de vista ambiental, devido ao clima, ao relevo, à vegetação, etc., já tinham sido colonizadas por estes pássaros. Uma possível explicação é que o processo latente nas regiões onde registrou-se a presença nos últimos anos foi estimado apenas considerando os dados, sem a dependência espacial.

Como nos primeiros anos não temos informação suficiente, fica difícil corrigir e estimar corretamente o processo latente.

Tabela 7.3: *ECD* - Logaritmo da densidade preditiva para os modelos ajustados.

Modelo	Log(densidade preditiva)
I(a)	-116.26
I(b)	-126.97
I(c)	-132.40
I(d)	-141.59
II	-116.45
III	-119.29
IV	-124.05

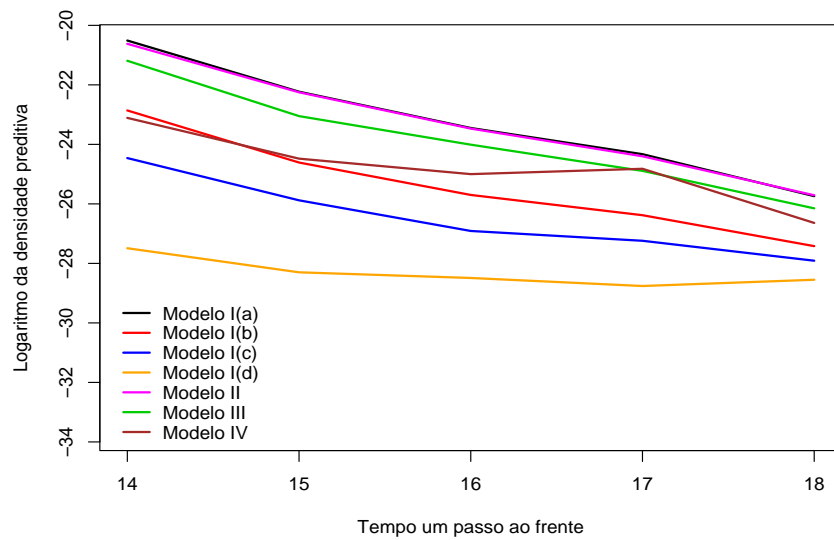


Figura 7.9: *ECD* - Logaritmo da densidade preditiva dos modelos.

Neste capítulo apresentamos uma aplicação do modelo espaço-temporal generalizado dinâmico para um conjunto de dados binários, relativos à presença ou ausência da *Eurasian Collared-Dove* nos Estados Unidos. Originalmente estes eram dados de contagem, pelo que houve necessidade de agregar os locais de observação em uma grade regular. O modelo

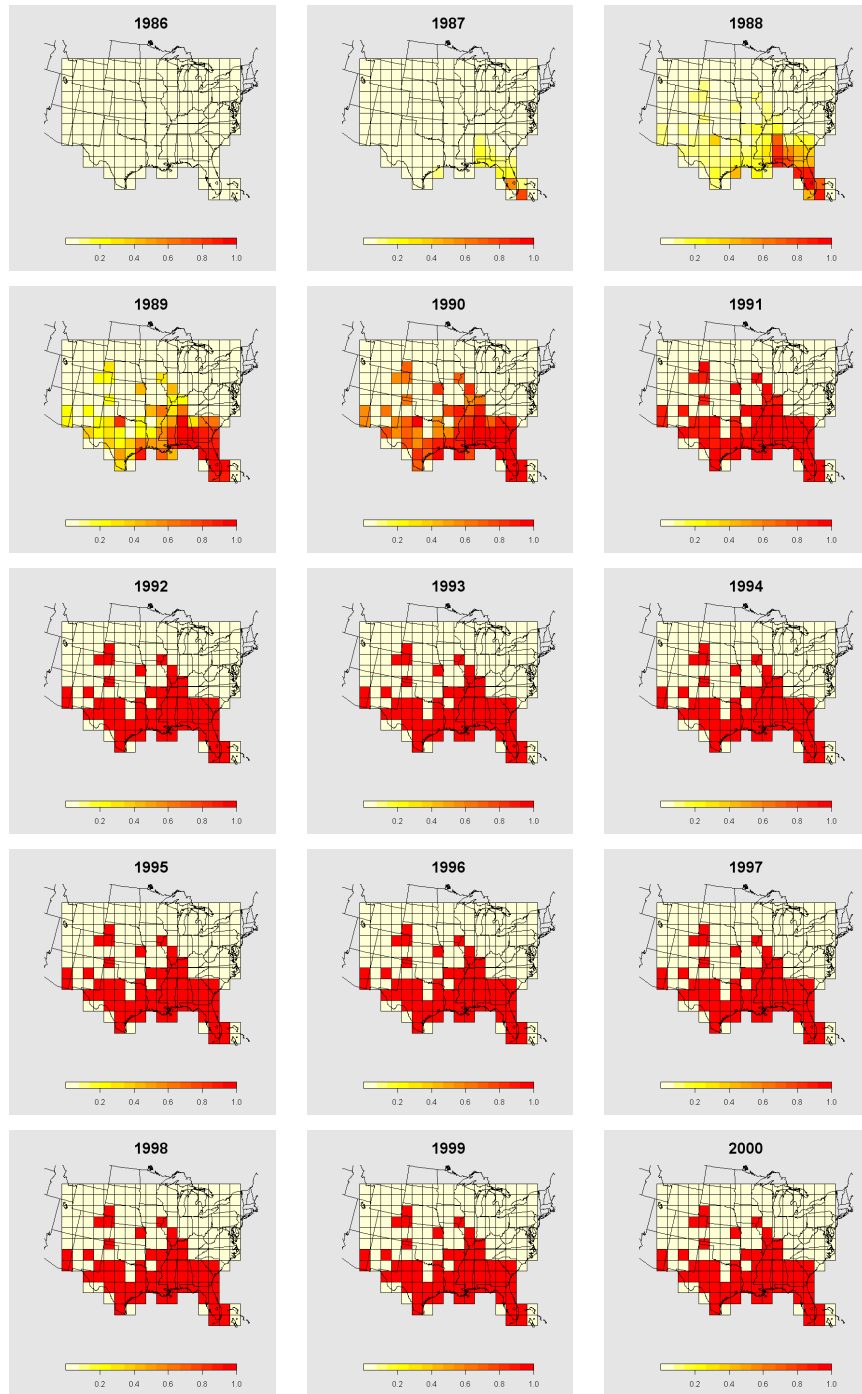


Figura 7.10: ECD - Modelo I(a): Probabilidade estimada para os anos 1986-2000.

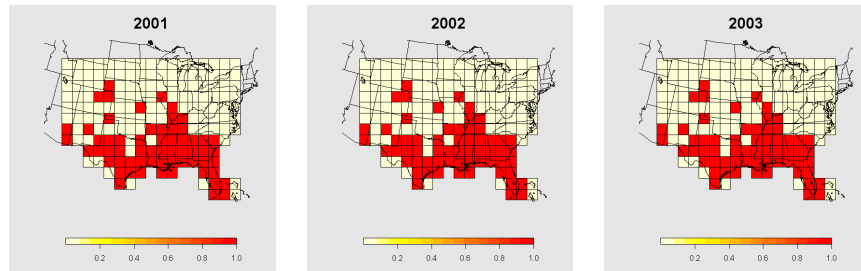


Figura 7.11: *ECD* - Modelo I(a): Probabilidade estimada para os anos 2001-2003.

geral para dados desta natureza foi definido e diversos modelos foram derivados desse modelo geral. Estes dados agregados não resultaram muito informativos, e elaboramos um exemplo com dados simulados para testar a metodologia para fazer inferência Bayesiana para os parâmetros e o algoritmo *EFIFBS* para estimar o processo latente. Finalmente, realizamos seleção de modelos baseada na densidade preditiva. Os resultados para os dados simulados foram promissores, pelo que concluímos que com dados mais informativos a metodologia proposta é adequada.

Capítulo 8

Considerações finais

Esta tese teve como objetivo propor uma classe de modelos espaço-temporais para dados de área na família exponencial, como por exemplo, dados de contagem ou dados binários, entre outros, estendendo naturalmente a classe de modelos proposta em [Vivar \(2004\)](#), a qual era útil para observações Gaussianas. Para trabalhar com observações que pertencem à família exponencial existem diversas abordagens, nesta tese descrevemos duas delas. Na primeira, utilizamos uma transformação nos dados para deixá-los aproximadamente normais. Quando esta aproximação é boa, é possível aplicar a classe de modelos espaço-temporais para dados Gaussianos introduzida na dissertação do Mestrado. Uma aplicação transformando os dados do número de homicídios por municípios nos Estados do Rio de Janeiro e do Espírito Santo obteve bons resultados. A transformação utilizada neste caso foi a raiz quadrada da taxa de mortalidade por homicídio (TMH) por cem mil habitantes para cada município. A aproximação funcionou adequadamente devido a que o número de homicídios é alto nesses Estados. Os resultados obtidos para os dois estados encontram-se no [Capítulo 4](#).

A segunda abordagem consiste em desenvolver um esquema de amostragem eficiente para o processo latente da classe de modelos espaço-temporais generalizados dinâmicos. Este foi outro dos objetivos atingidos com sucesso. O esquema de amostragem é denominado *EFIFBS* (*extended forward information filter backward sampler*). No primeiro passo do *EFIFBS* (passo de filtragem) utilizamos o filtro de informação estendido, que é uma generalização do filtro de informação, que por sua vez é uma alternativa para o filtro de

Kalman. A principal característica do filtro de informação é que as equações recursivas que o definem estão em função das matrizes de precisão dos *CAMGP*. Esta é uma vantagem computacional, pois precisamos de menos inversões de matrizes a cada iteração do *MCMC*, diminuindo o tempo computacional. O segundo passo (amostragem recursiva), é a bem conhecida amostragem em bloco do algoritmo *FFBS* (Carter e Kohn, 1994; Frühwirth-Schnatter, 1994). Introduzimos os modelos espaço-temporais generalizados dinâmicos para dados de área e a metodologia de inferência Bayesiana para os parâmetros no Capítulo 5.

Duas aplicações desta classe de modelos, a primeira, utilizando os dados do número de homicídios no Espírito Santo, (caso Poisson); e, a segunda, com dados de presença-ausência da ave *Eurasian Collared-Dove* em regiões dos Estados Unidos (caso Bernoulli) foram detalhadas nos Capítulos 6 e 7. Uma das principais vantagens de modelar diretamente os dados espaço-temporais na família exponencial é que os parâmetros estimados ganham uma interpretação natural, diferentemente do que ocorre com os dados transformados. Por exemplo, no caso do número de homicídios, o processo latente pode ser descrito como a taxa de risco por cem mil habitantes.

Uma metodologia para selecionar modelos através da densidade preditiva foi desenvolvida para esta classe de modelos espaço-temporais. A densidade preditiva leva em conta naturalmente a incerteza associada a estimação dos parâmetros do modelo. O critério para selecionar o melhor modelo entre todos os propostos está baseado na densidade preditiva conjunta, e está descrito na Subseção 3.5.2 no caso de dados Gaussianos, e na Seção 5.6 para dados na família exponencial. O melhor modelo é aquele com a maior densidade preditiva conjunta.

Para todas as aplicações, especificamos diferentes modelos espaço-temporais para selecionar o melhor com o nosso critério. Além disso, para o caso do número de homicídios utilizando aproximação Gaussiana, um exercício com modelos com e sem estrutura espacial nas inovações foi desenvolvido, para determinar se é razoável ou não supor a dependência espacial. Os resultados foram muito claros: todos os modelos considerando dependência espacial resultaram melhores que os seus pares sem estrutura espacial, e entre aqueles, o melhor foi o modelo chamado de contaminação. Também foi o modelo de contaminação

o selecionado entre os modelos generalizados dinâmicos para o número de homicídios no Espírito Santo. No caso dos dados de presença-ausência da *Eurasian Collared-Dove*, não contamos com muita informação e dos modelos especificados foi selecionado o mais simples. Um exemplo com dados artificiais gerados utilizando um modelo mais complexo determinou que a metodologia de estimação é apropriada quando os dados têm algo mais de informação.

Os próximos trabalhos incluem a elaboração de modelos mais complexos, na medida do possível, para o problema da *Eurasian Collared-Dove* e outras aplicações com dados na família exponencial. A modelagem incluindo covariáveis adequadas é uma tarefa que pode ser abordada no futuro. A metodologia de estimação proposta nesta tese, os resultados destas e outras aplicações e o método de seleção de modelos serão reportados em futuros artigos.

Apêndice A

Rio de Janeiro

A.1 Mapa político de 1979

O mapa político do Estado do Rio de Janeiro do ano 1979 está ilustrado na Figura A.1. Este foi a base para definir a sua estrutura espacial utilizada no Capítulo 4. A seguir está uma lista com todos os nomes dos municípios atuais. Os nomes dados às regiões que são resultado da “união” de municípios estão em *itálico*.

- | | |
|--|--|
| 1. Angra dos Reis | 10. Campos dos Goytacazes + <i>Italva</i> +
Cardoso Moreira |
| 2. Araruama | 11. Cantagalo |
| 3. Barra do Piraí | 12. Carmo |
| 4. <i>Barra Mansa</i> + Quatis | 13. <i>Casimiro de Abreu</i> + Rio das Ostras |
| 5. Bom Jardim | 14. Conceição de Macabu |
| 6. Bom Jesus do Itabapoana | 15. <i>Cordeiro</i> + Macuco |
| 7. Cabo Frio + <i>Arraial do Cabo</i> +
Armação de Búzios | 16. Duas Barras |
| 8. Cachoeiras de Macacu | 17. Duque de Caxias |
| 9. <i>Cambuci</i> + São José de Ubá | 18. Engenheiro Paulo de Frontin |

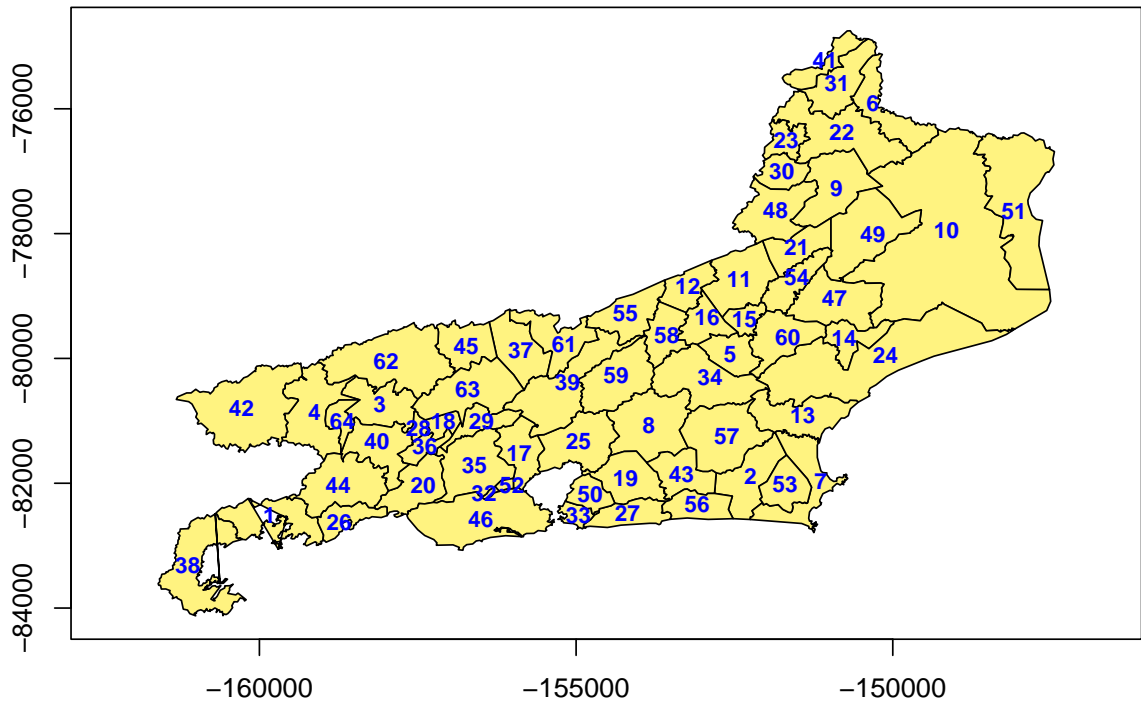


Figura A.1: Mapa político do Estado do Rio de Janeiro (1979).

- | | |
|---|--|
| 19. Itaboraí + Tanguá | 29. Miguel Pereira |
| 20. Itaguaí + Seropédica | 30. Miracema |
| 21. Itaocara | 31. <i>Natividade</i> + Varre-Sai |
| 22. Itaperuna | 32. Nilópolis |
| 23. Laje do Muriaé | 33. Niterói |
| 24. <i>Macaé</i> + Quissamã + Carapebus | 34. Nova Friburgo |
| 25. <i>Magé</i> + Guapimirim | 35. <i>Nova Iguaçu</i> + Belford Roxo + Japeri + Queimados |
| 26. Mangaratiba | 36. Paracambi |
| 27. Maricá | 37. Paraiba do Sul |
| 28. Mendes | |

- | | |
|---|--|
| <p>38. Parati</p> <p>39. <i>Petrópolis</i> + São José do Vale do Rio Preto</p> <p>40. <i>Piraí</i> + Pinheiral</p> <p>41. Porciúncula</p> <p>42. Resende + <i>Itatiaia</i> + Porto Real</p> <p>43. Rio Bonito</p> <p>44. Rio Claro</p> <p>45. Rio das Flores</p> <p>46. Rio de Janeiro</p> <p>47. Santa Maria Madalena</p> <p>48. <i>Santo Antônio de Pádua</i> + Aperibé</p> <p>49. São Fidélis</p> <p>50. São Gonçalo</p> <p>51. <i>São João da Barra</i> + São Francisco de Itabapoana</p> | <p>52. São João de Meriti</p> <p>53. <i>São Pedro da Aldeia</i> + Iguaba Grande</p> <p>54. São Sebastião do Alto</p> <p>55. Sapucaia</p> <p>56. Saquarema</p> <p>57. Silva Jardim</p> <p>58. Sumidouro</p> <p>59. Teresópolis</p> <p>60. Trajano de Moraes</p> <p>61. <i>Três Rios</i> + Areal + Comendador Levy Gasparian</p> <p>62. Valença</p> <p>63. Vassouras + <i>Paty do Alferes</i></p> <p>64. Volta Redonda</p> |
|---|--|

A.2 Taxas de mortalidade por homicídio por município

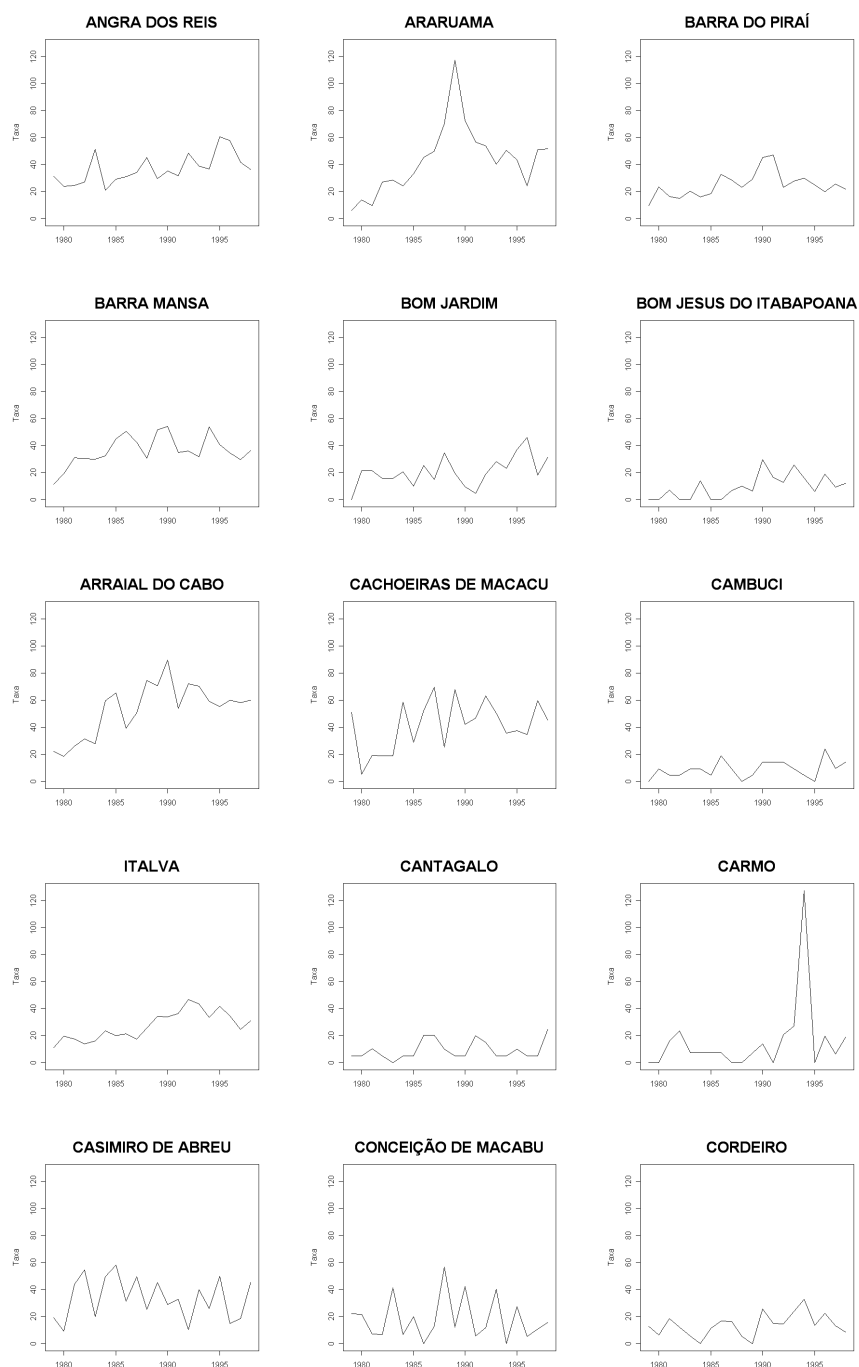


Figura A.2: TMH por 100000 habitantes por município do Rio de Janeiro - I.

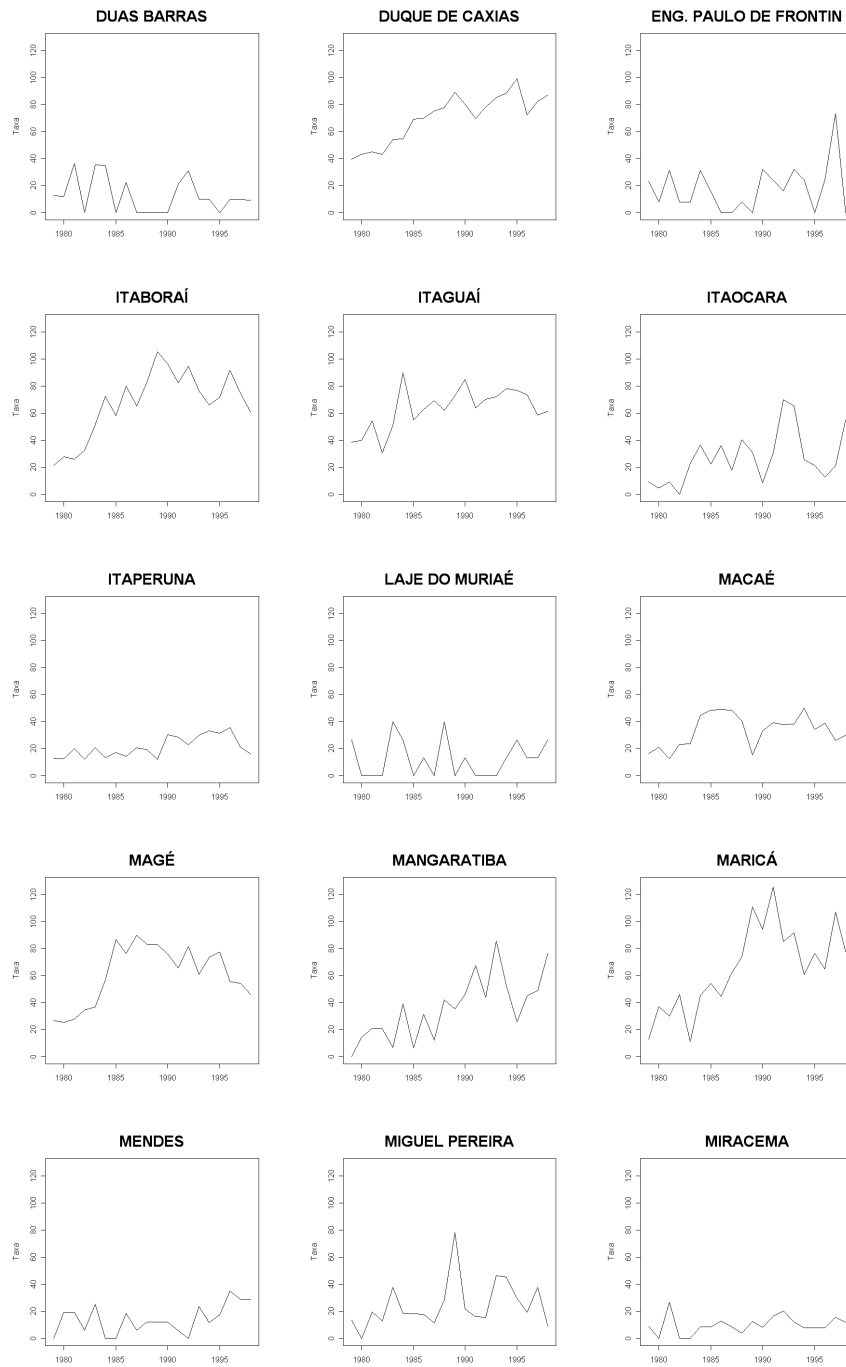


Figura A.3: TMH por 100000 habitantes por município do Rio de Janeiro - II.

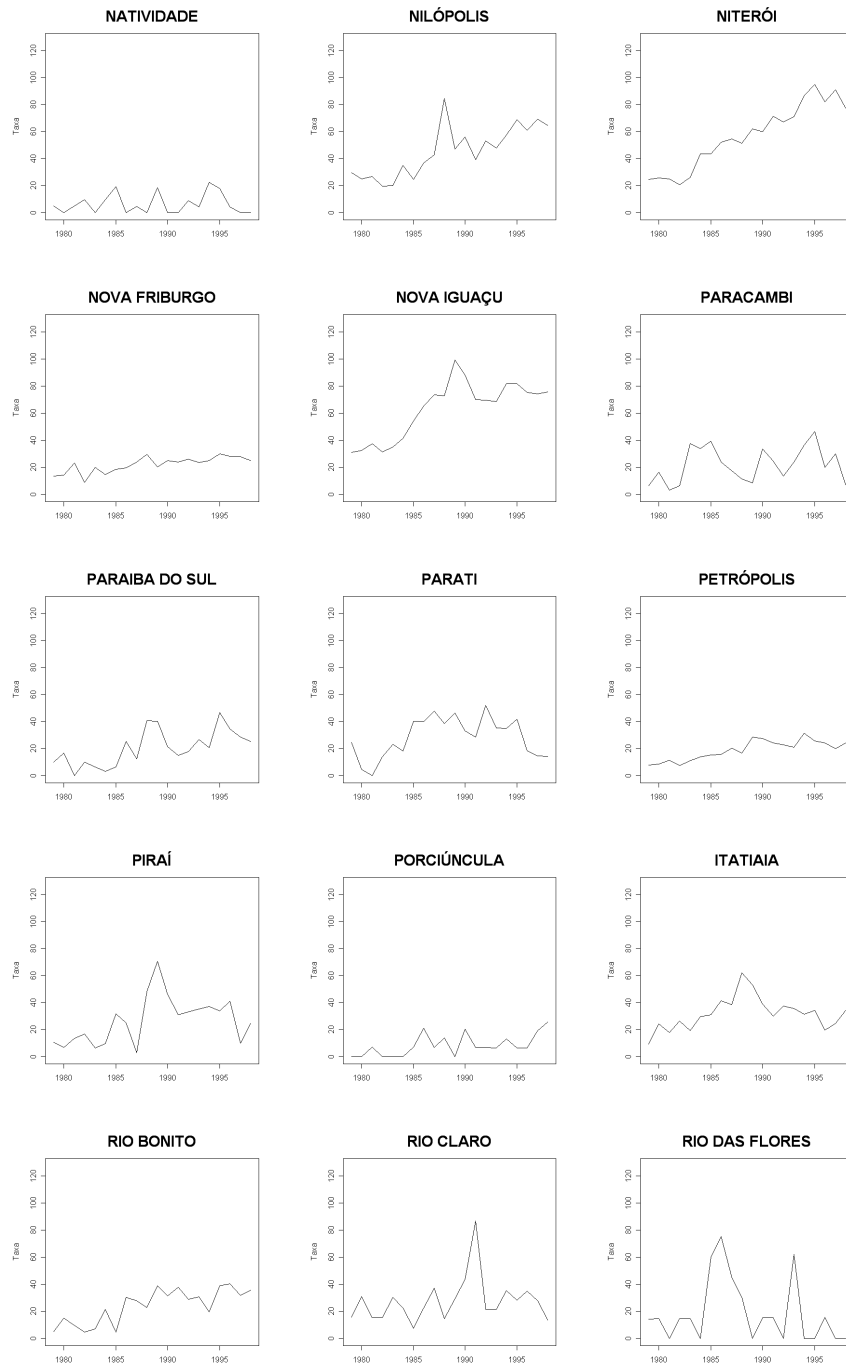


Figura A.4: TMH por 100000 habitantes por município do Rio de Janeiro - III.

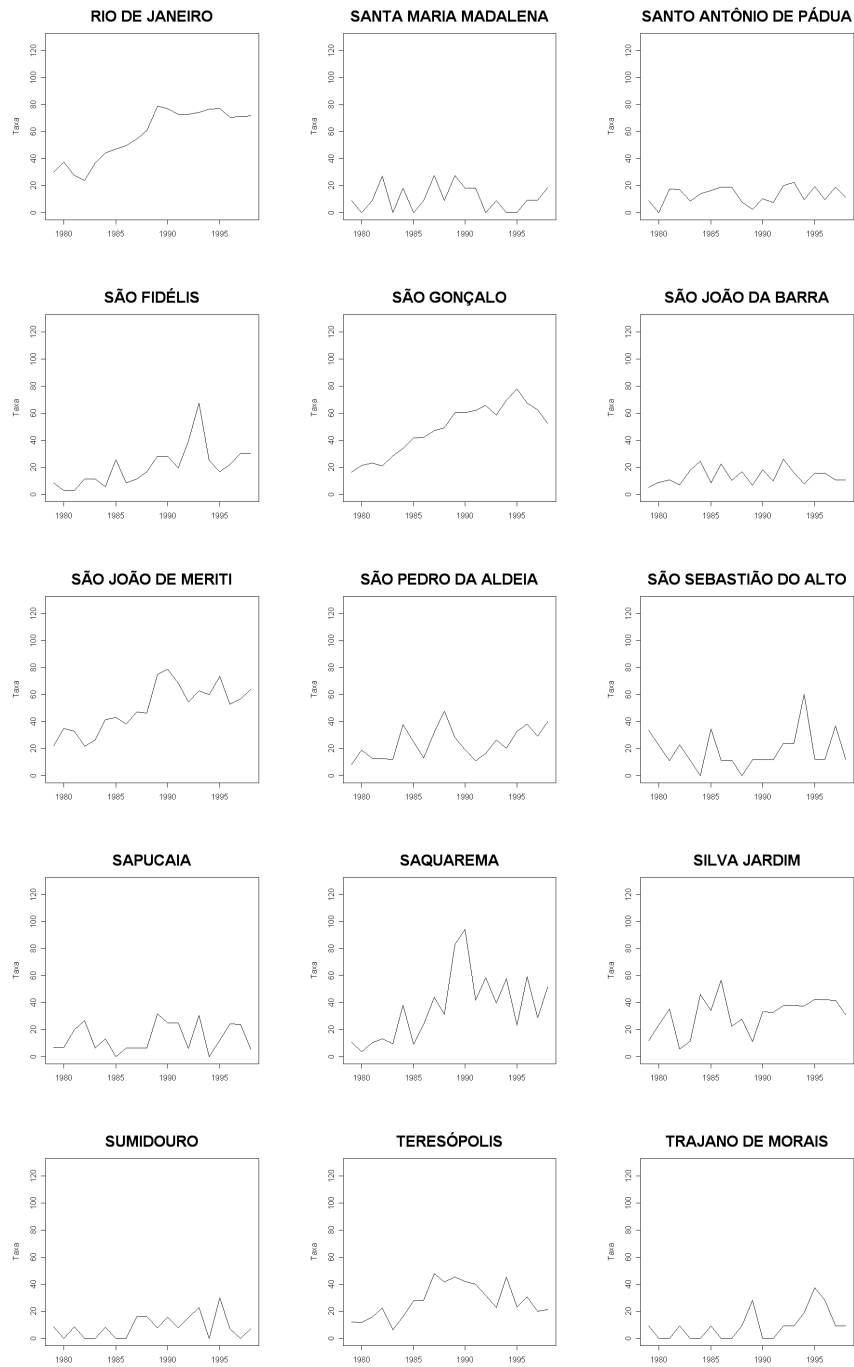


Figura A.5: TMH por 100000 habitantes por município do Rio de Janeiro - IV.

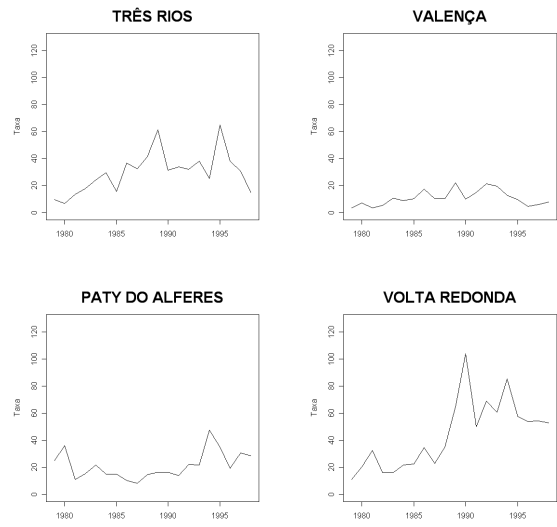


Figura A.6: TMH por 100000 habitantes por município do Rio de Janeiro - V.

A.3 Taxas de mortalidade por homicídios por ano

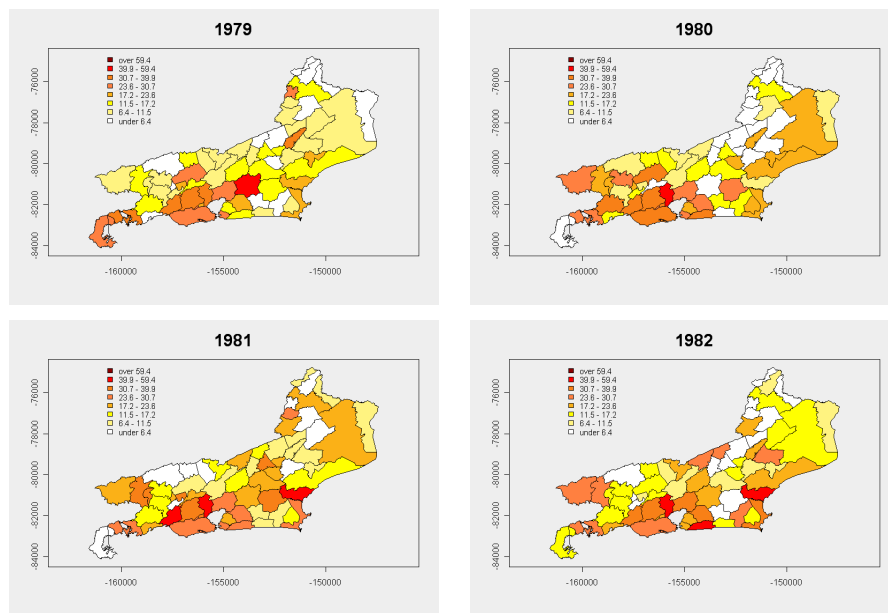


Figura A.7: Mapas da TMH por 100000 habitantes por município do Rio de Janeiro - Anos 1979 - 1982.

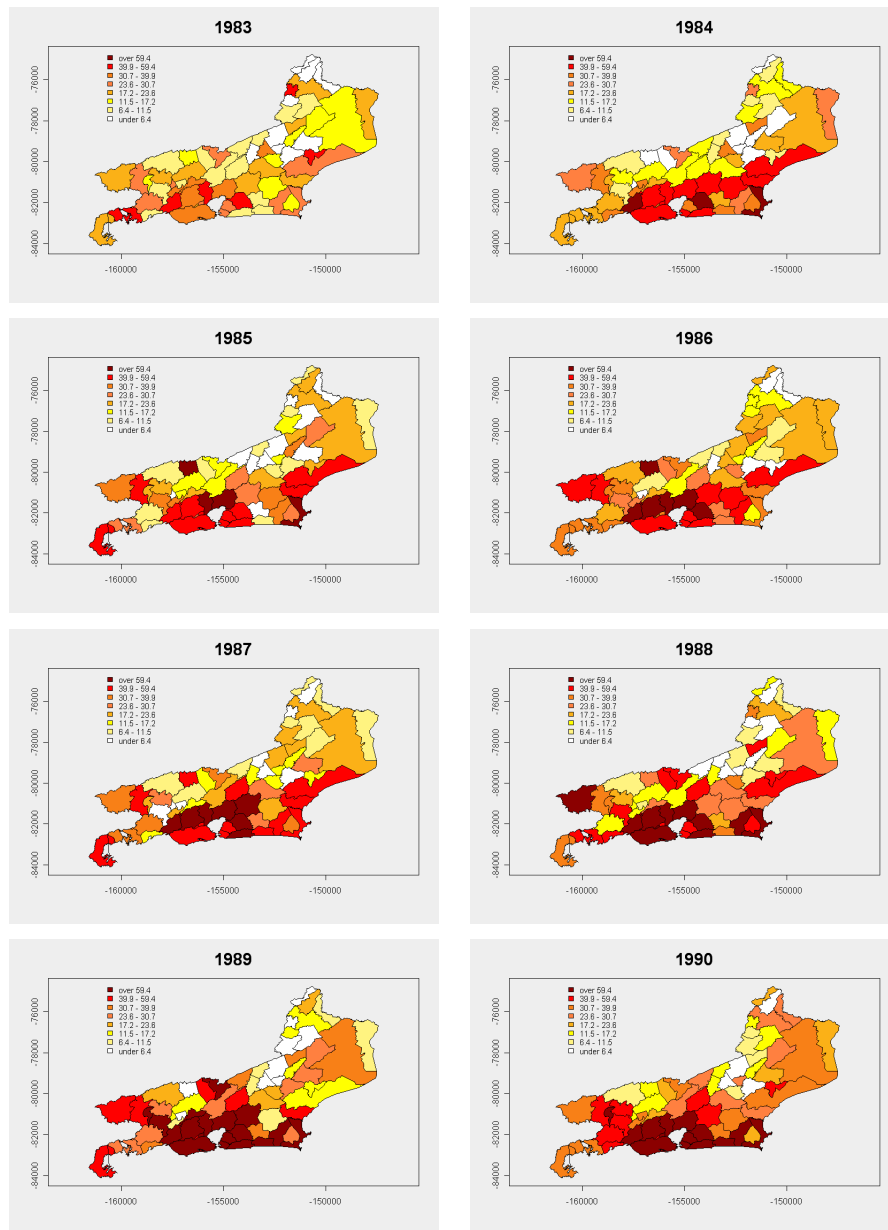


Figura A.8: Mapas da TMH por 100000 habitantes por município do Rio de Janeiro - Anos 1983 - 1990.

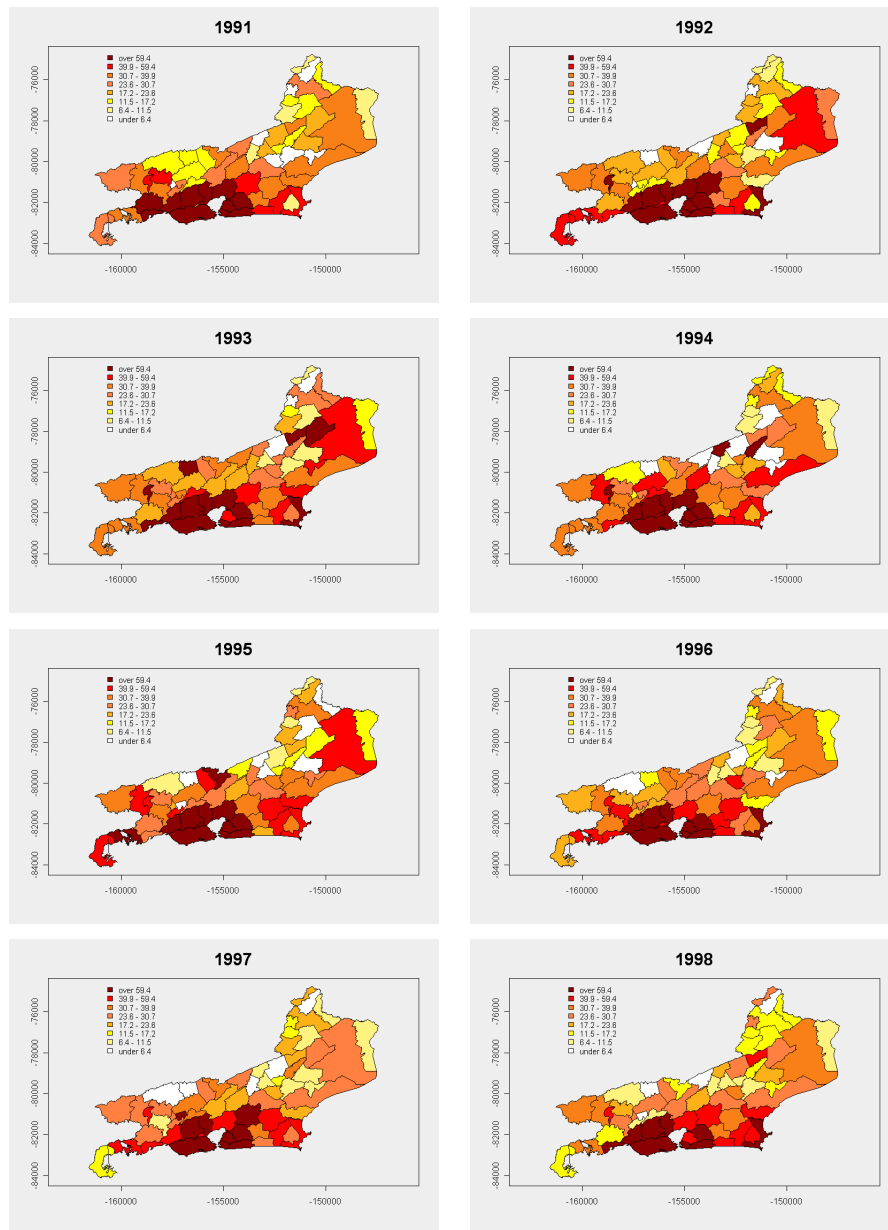


Figura A.9: Mapas da TMH por 100000 habitantes por município do Rio de Janeiro - Anos 1991 - 1998.

Apêndice B

Espírito Santo

B.1 Mapa político de 1979

O mapa político do Estado do Espírito Santo do ano 1979 está ilustrado na Figura B.1. Este foi a base para definir a matriz de vizinhança no Capítulo 4. A seguir está uma lista com todos os nomes dos municípios atuais. Os nomes dados às regiões que são resultado da “união” de municípios estão em *itálico*.

- | | |
|---|--|
| 1. <i>Afonso Cláudio</i> + Laranja da Terra + Brejetuba | 9. <i>Barra de São Francisco</i> + Água Doce do Norte |
| 2. <i>Alegre</i> + Ibitirama | 10. Boa Esperança |
| 3. Alfredo Chaves | 11. Bom Jesus do Norte |
| 4. Anchieta | 12. <i>Cachoeiro de Itapemirim</i> + Vargem Alta |
| 5. Apiacá | 13. Cariacica |
| 6. Aracruz | 14. Castelo |
| 7. Atilio Vivacqua | 15. <i>Marilândia</i> + São Domingos do Norte + Governador Lindenberg + Colatina |
| 8. Baixo Guandu | |

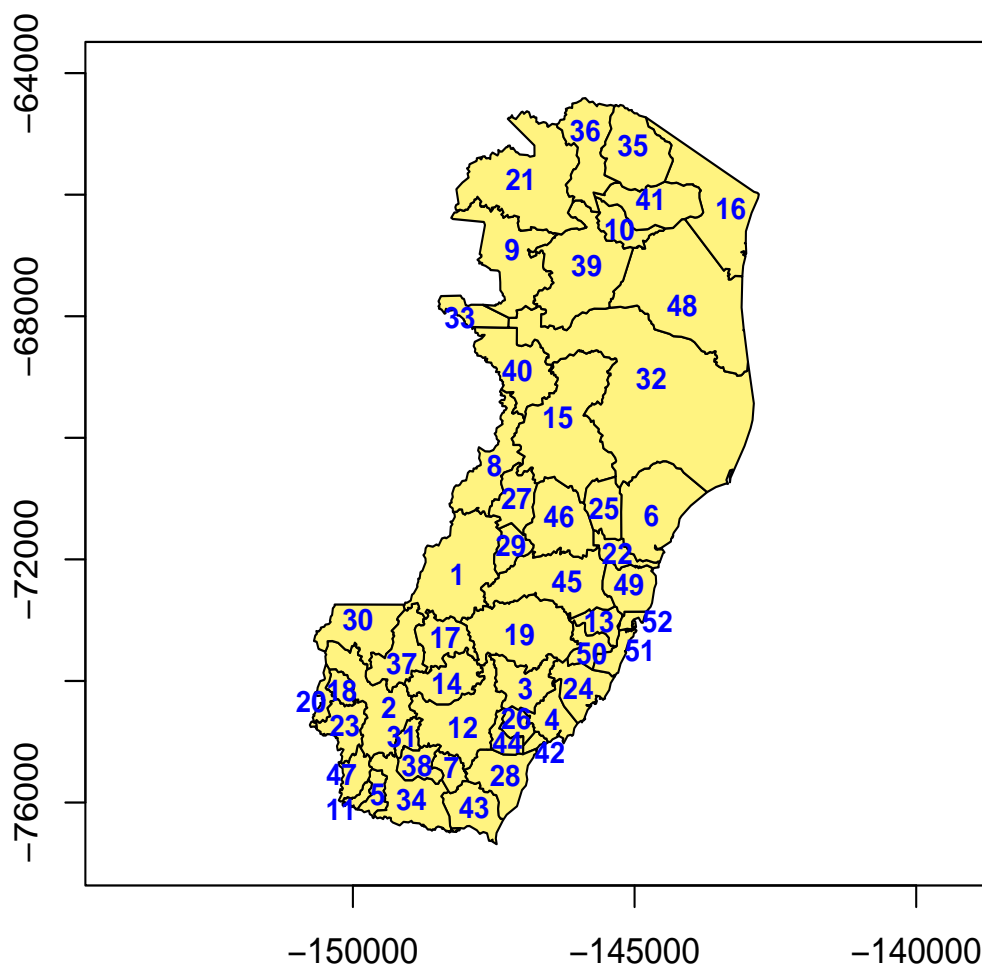


Figura B.1: Mapa político do Espírito Santo (1979).

16. <i>Conceição da Barra</i> + Pedro Canário	21. Ecoporanga
17. <i>Venda Nova do Imigrante</i> + Conceição do Castelo	22. Fundão
18. Divino de São Lourenço	23. Guaçuí
19. <i>Domingos Martins</i> + Marechal Floriano	24. Guarapari
20. Dores do Rio Preto	25. <i>Ibiraçu</i> + João Neiva
	26. Iconha

- | | |
|--|--|
| 27. Itaguaçu | 40. <i>Pancas</i> + Alto Rio Novo |
| 28. <i>Itapemirim</i> + Marataízes | 41. Pinheiros |
| 29. Itarana | 42. Piúma |
| 30. <i>Ibatiba</i> + Irupi + Iúna | 43. Presidente Kennedy |
| 31. Jerônimo Monteiro | 44. Rio Novo do Sul |
| 32. <i>Águia Branca</i> + São Gabriel da Palha
+ Vila Valério + Sooretama + Rio Ba-
nanal + Linhares | 45. <i>Santa Maria de Jetibá</i> + Santa
Leopoldina |
| 33. Mantenópolis | 46. <i>Santa Teresa</i> + São Roque de Canaã |
| 34. Mimoso do Sul | 47. São José do Calçado |
| 35. Montanha | 48. <i>Jaguaré</i> + São Mateus |
| 36. <i>Mucurici</i> + Ponto Belo | 49. Serra |
| 37. Muniz Freire | 50. Viana |
| 38. Muqui | 51. Vila Velha |
| 39. <i>Nova Venécia</i> + Vila Pavão | 52. Vitória |

B.2 Taxa de mortalidade por homicídio por município

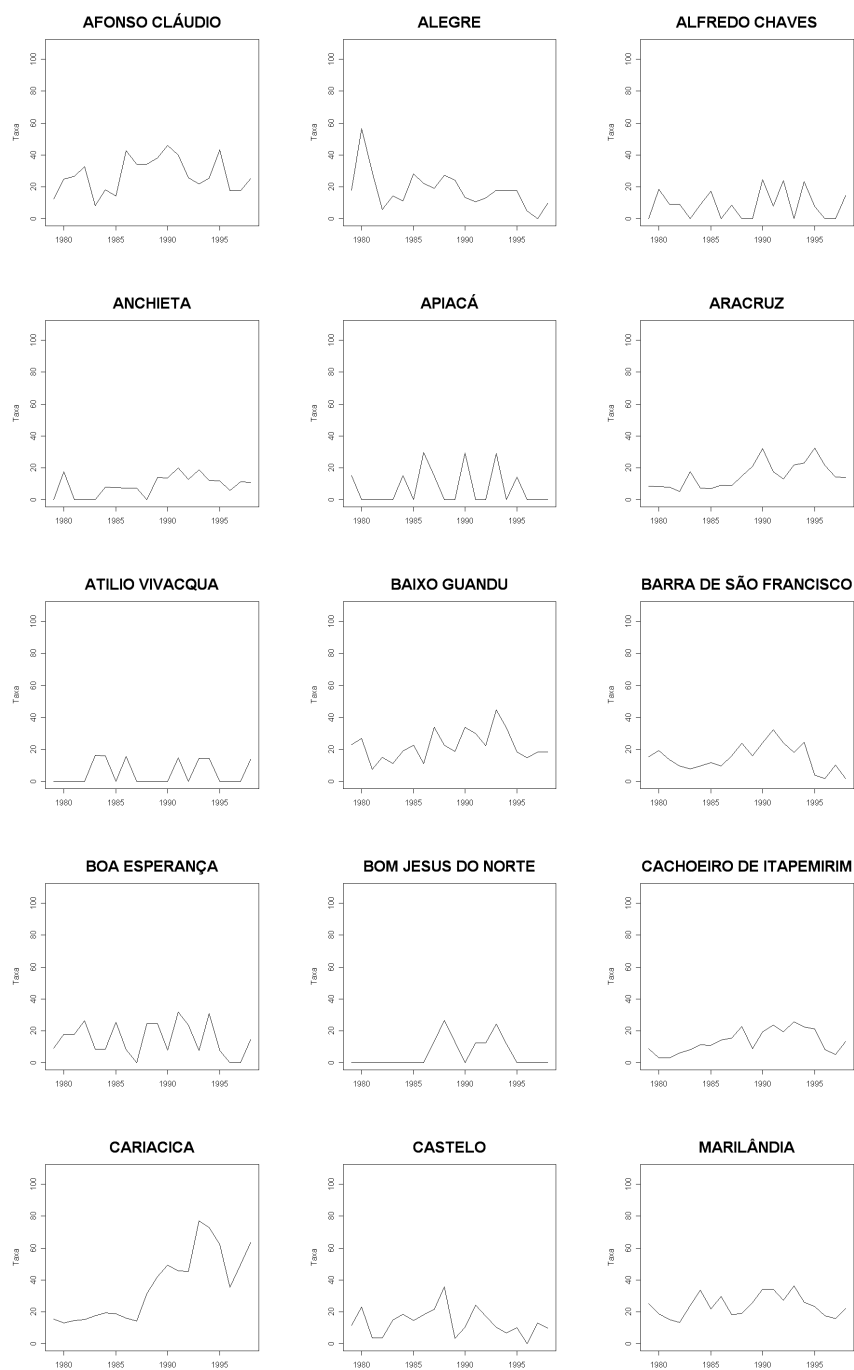


Figura B.2: TMH por 100000 habitantes por município do Espírito Santo - I.

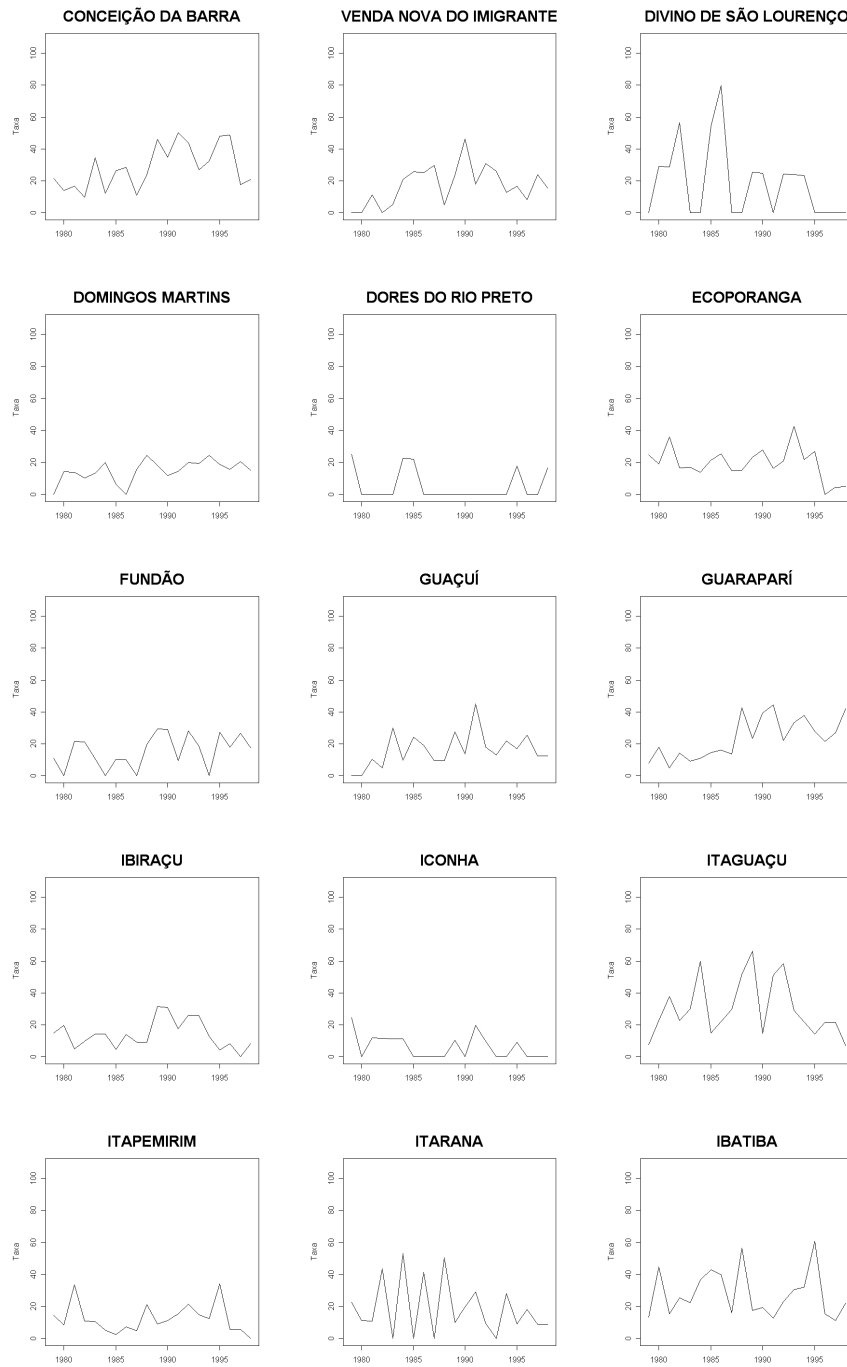


Figura B.3: TMH por 100000 habitantes por município do Espírito Santo - II.

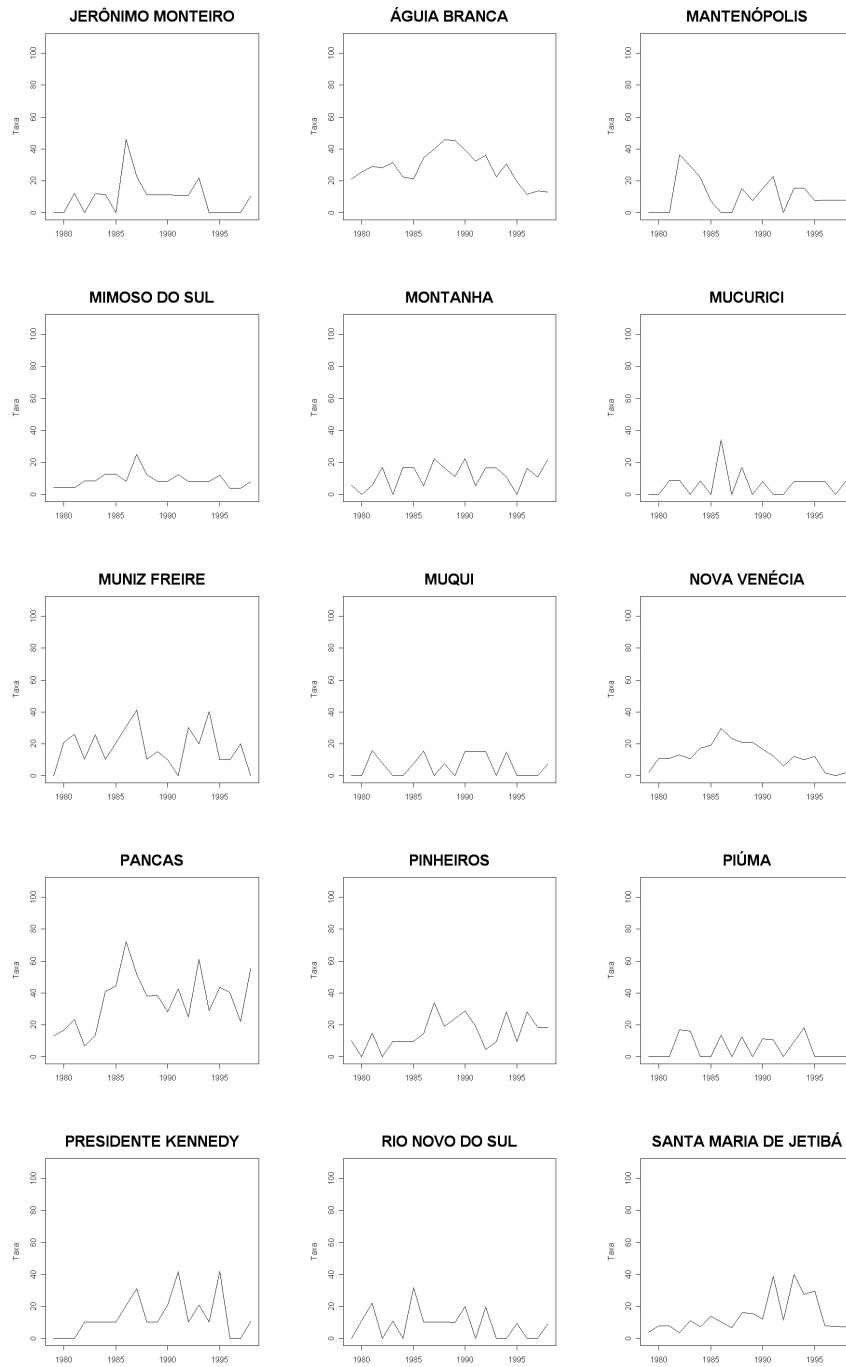


Figura B.4: TMH por 10000 habitantes por município do Espírito Santo - III.

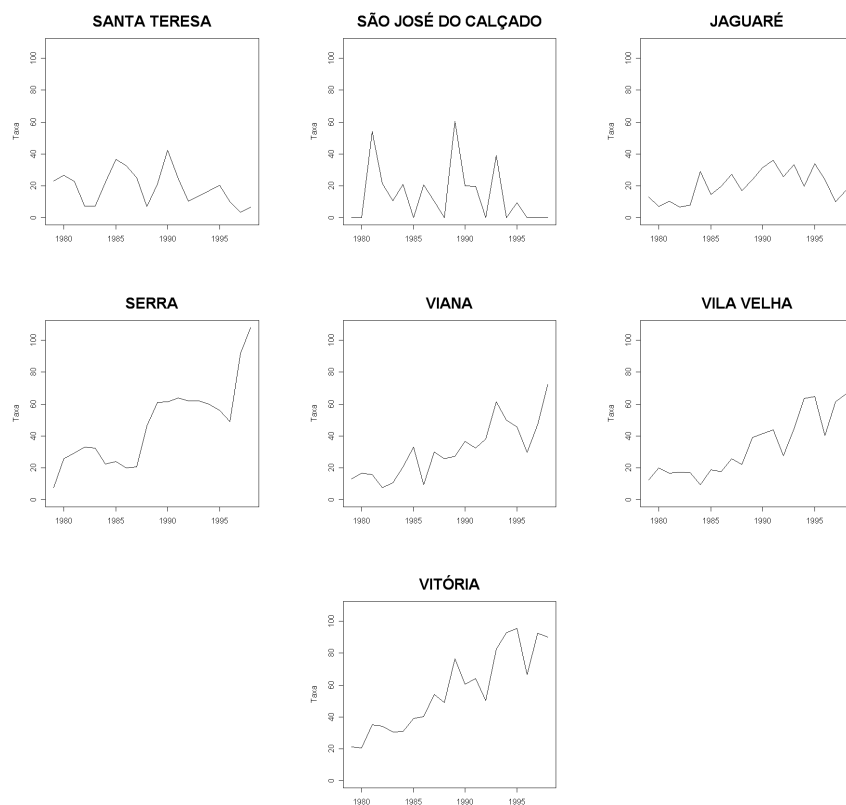


Figura B.5: TMH por 100000 habitantes por município do Espírito Santo - IV.

B.3 Taxas de mortalidade por homicídio por ano

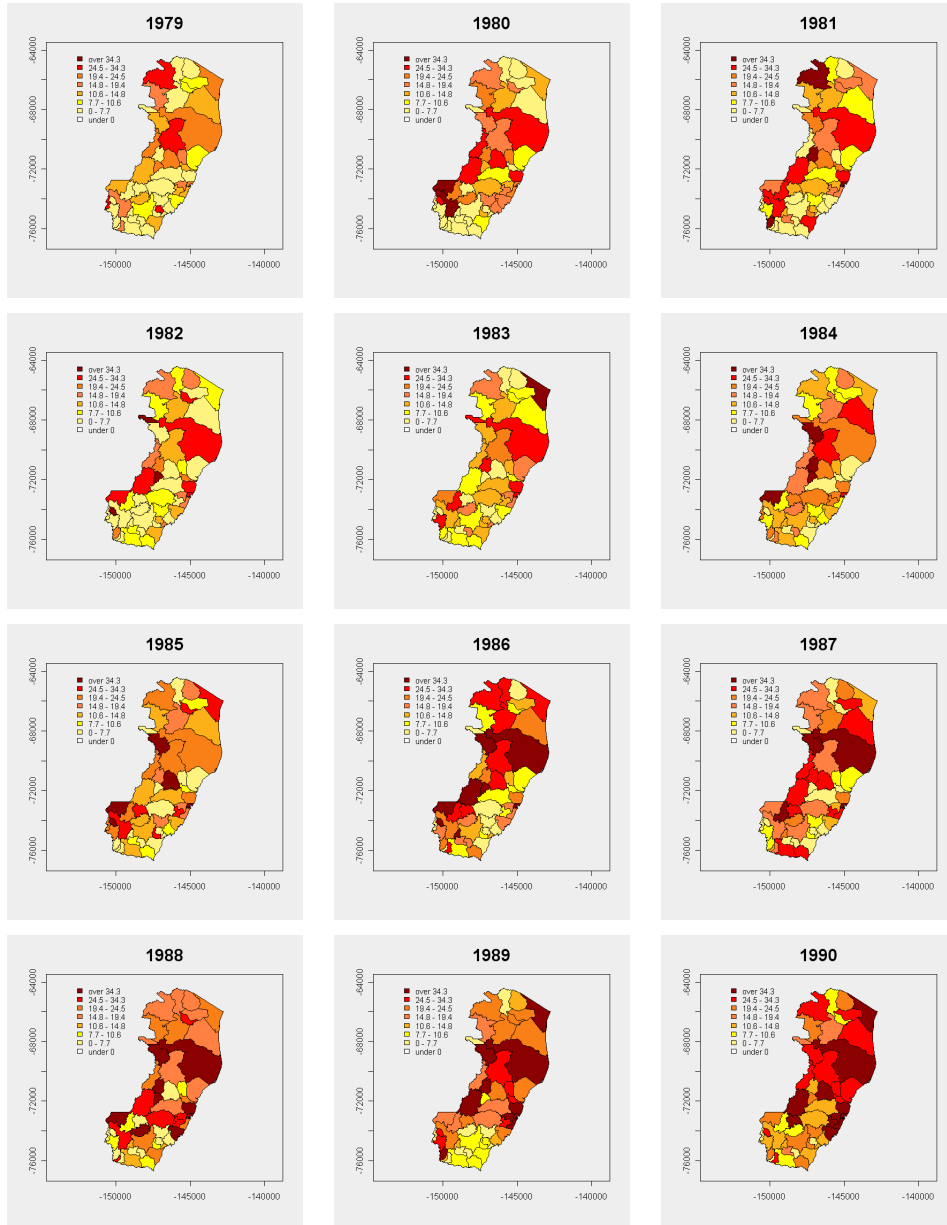


Figura B.6: Mapas da TMH por 100000 habitantes por município do Espírito Santo - Anos 1979 - 1990.

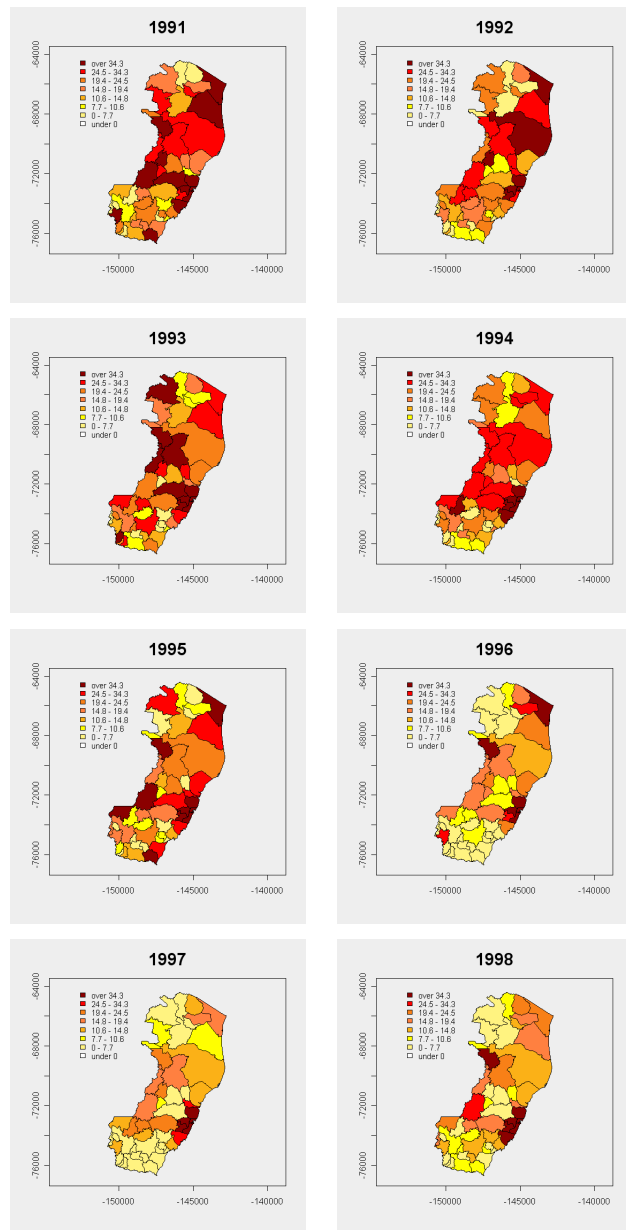


Figura B.7: Mapas da TMH por 100000 habitantes por município do Espírito Santo - Anos 1991 - 1998.

Referências Bibliográficas

- Aitkin, M. (1991). Posterior Bayes factors. *Journal of the royal Statistical Society, Series B*, 53(1):111–142.
- Anderson, B. D. & Moore, J. B. (1979). *Optimal Filtering*. Prentice Hall, New Jersey.
- Anscombe, F. (1948). The transformation of Poisson, binomial and negative binomial data. *Biometrika*, 35:246–254.
- Assunção, R. M., Reis, I. A., & Oliveira, C. D. L. (2001). Diffusion and prediction of leishmaniasis in a large metropolitan area in Brazil with a Bayesian space-time model. *Statistics in Medicine*, 20:2319–2335.
- Banerjee, S., Carlin, B. P., & Gelfand, A. E. (2004). *Hierarchical Modeling and Analysis for Spatial Data*, volume 101 de *Monographs on Statistics and Applied Probability*. Chapman and Hall, London.
- Bernardinelli, L., Clayton, D., Pascutto, C., Montomoli, C., Ghislandi, M., & Songini, M. (1995). Bayesian analysis of space-time variation in disease risk. *Statistics in Medicine*, 14:2433–2443.
- Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems (with discussion). *Journal of the Royal Statistical Society, Series B*, 36:192–236.
- Besag, J. (1975). Statistical analysis of non-lattice data. *The Statistician*, 24(3):171–195.
- Besag, J. E. & Kooperberg, C. (1995). On conditional and intrinsic autoregressions. *Biometrika*, 82:733–746.

- Besag, J. E., York, J., & Mollié, A. (1991). Bayesian image restoration with two applications in spatial statistics (with discussion). *Annals of the Institute of Statistical Mathematics*, 43:1–59.
- Brown, P. E., Diggle, P. J., Lord, M. E., & Young, P. C. (2001). Space-time calibration of radar-rainfall data. *Applied Statistics*, 50:221–242.
- Carlin, B. & Chib, S. (1995). Bayesian model choice via Markov chain Monte Carlo methods. *Journal of the Royal Statistical Society, Series B*, 57:473–484.
- Carlin, B. P., Polson, N. G., & Stoffer, D. S. (1992). A Monte Carlo approach to nonnormal and nonlinear state space modeling. *Journal of the American Statistical Association*, 87:493–500.
- Carter, C. K. & Kohn, R. (1994). On Gibbs sampling for state space models. *Biometrika*, 81:541–553.
- Chib, S. (1995). Marginal likelihood from the Gibbs output. *Journal of the American Statistical Association*, 90:1313–1321.
- Clyde, M. & George, E. I. (2004). Model uncertainty. *Statistical Science*, 19:81–94.
- Cressie, N. A. C. (1993). *Statistics for Spatial Data*. Wiley, New York, 2nd. edition.
- Cressie, N. A. C. & Huang, H. C. (1999). Classes of nonseparable, spatio-temporal stationary covariance functions. *Journal of the American Statistical Association*, 94(448):1330–1340.
- Dellaportas, R., Forster, J., & Ntzoufras, J. (2000). Bayesian Variable Selection using the Gibbs Sampler. Em *Generalized Linear Models: a Bayesian Perspective*, Dey, D. P., Ghosh, S., & Mallick, B., editores, New York. Marcel Dekker.
- Dethlefsen, C. (2003). Markov Random Field Eextension using State Space Models. Em *Bayesian Statistics 7*, Bernardo, J., Berger, J., Dawid, A., & Smith, A., editores. London: Oxford University Press.

- Dey, D. K., Gelfand, A. E., Swartz, T., & Vlachos, P. (1998). Simulation-based model checking for hierarchical models. *Test*, 7:325–346.
- Dey, D. P., Ghosh, S., & Mallick, B., editores (2000). *Generalized Linear Models: a Bayesian Perspective*. Marcel Dekker, New York.
- Di Ciccio, T., Kass, R., Raftery, A., & Wasserman, L. (1997). Computing Bayes factors by combining simulations and asymptotic approximations. *Journal of the American Statistical Association*, 92:903–915.
- Doornik, J. A. (2002). *Object-Oriented Matrix Programming Using Ox*. Timberlake Consultants Press and Oxford, London, 3rd. edition.
- Fahrmeir, L. (1992). Posterior mode estimation by extended Kalman filter for multivariate dynamic generalized linear models. *Journal of the American Statistical Association*, 87:501–509.
- Fahrmeir, L. & Tutz, G. (2001). *Multivariate Statistical Modelling based on Generalized Linear Models*. Springer, New York, 2nd. edition.
- Ferreira, M. A. R. & De Oliveira, V. (2007). Bayesian reference analysis for Gaussian Markov Random Fields. *Journal of Multivariate Analysis*, 98:789–812.
- Ferreira, M. A. R. & Gamerman, D. (2000). Dynamic Generalized Linear Models. Em *Generalized Linear Models: a Bayesian Perspective*, Dey, D. P., Ghosh, S., & Mallick, B., editores, páginas 57–72, New York. Marcel Dekker.
- Frühwirth-Schnatter, S. (1994). Data augmentation and dynamic linear models. *Journal of Time Series*, 15(2):183–202.
- Frühwirth-Schnatter, S. (1995). Bayesian model discrimination and Bayes factors for linear Gaussian state-space models. *Journal of the Royal Statistics Society, Series B*, 57:237–246.

- Gamerman, D. & Lopes, H. (2006). *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*. Chapman and Hall, London, 2nd edition.
- Geisser, S. & Eddy, W. F. (1979). A predictive approach to model selection. *Journal of the American Statistical Association*, 74:153–160.
- Gelfand, A. & Ghosh, S. (1998). Spatio-temporal modeling of residential sales markets. *Journal of Business and Economic Statistics*, 16:312–321.
- Gelfand, A. E. (1996). Model determination using sample-based methods. Em *Markov Chain Monte Carlo in practice*, Gilks, W., Richardson, S., & Spiegelhalter, D. J., editores, páginas 145–161, London. Chapman and Hall.
- Gelfand, A. E., Banerjee, S., & Gamerman, D. (2005). Spatial process modelling for univariate and multivariate dynamic spatial data. *Environmetrics*, 16:465–479.
- Gelfand, A. E., Dey, D. K., & Chang, H. (1992). Model Determination using Predictive Distributions with Implementation via Sampling-Based Methods (with discussion). Em *Bayesian Statistics 4*, Bernardo, J., Berger, J., Dawid, A., & Smith, A., editores. London: Oxford University Press.
- Gelfand, A. E. & Ghosh, M. (2000). Generalized Linear Models: A Bayesian View. Em *Generalized Linear Models: a Bayesian Perspective*, Dey, D. P., Ghosh, S., & Mallick, B., editores, páginas 3–22, New York. Marcel Dekker.
- Gelman, A. (1996). Influence and Monitoring Convergence. Em *Markov Chain Monte Carlo in practice*, Gilks, W., Richardson, S., & Spiegelhalter, D. J., editores, páginas 131–143, London. Chapman and Hall.
- Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (1995a). *Bayesian Data Analysis*. Chapman and Hall, London.
- Gelman, A. & Meng, X.-L. (1996). Model Checking and Model Improvement. Em *Markov Chain Monte Carlo in practice*, Gilks, W., Richardson, S., & Spiegelhalter, D. J., editores, páginas 189–202, London. Chapman and Hall.

- Gelman, A., Meng, X.-L., & Stern, H. (1995b). Posterior predictive assessment of model fitness via realized discrepancies (with discussion). *Statistica Sinica*, 6:733–807.
- Gelman, A. & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, 7(4):457–511.
- Ghil, M., Cohn, S., Tavantzis, J., Bube, K., & Isaacson, E. (1981). Applications of Estimation Theory to Numerical Weather Prediction. Em *Dynamic Meteorology: Data Assimilation Methods*, Bengtsson, L., Ghil, M., & Källén, E., editores, páginas 139–224. New York: Springer-Verlag.
- Green, P. J. (1995). Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika*, 82:711–732.
- Green, P. J., Hjort, N. L., & Richardson, S. (2003). *Highly Structured Stochastic Systems*. Oxford University Press.
- Guttorp, P., Meiring, W., & Sampson, P. D. (1994). A space-time analysis of ground-level ozone data. *Environmetrics*, 5:241–254.
- Han, C. & Carlin, B. (2000). Mcmc methods for computing Bayes factors: a comparative review. Technical report, University of Minnesota.
- Handcock, M. S. & Wallis, J. R. (1994). An approach to statistical spatial-temporal modeling of meteorological fields (with discussion). *Journal of the American Statistical Association*, 89(426):368–390.
- Harrison, P. J. & Stevens, C. F. (1976). Bayesian forecasting (with discussion). *Journal of the Royal Statistical Society, Series B*, 38:205–247.
- Harvey, A. C. (1989). *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge University Press, Cambridge.
- Harville, D. A. (1997). *Matrix Algebra from a Statistician's Perspective*. Springer, New York.

- Hooten, M. & Wikle, C. (2007). A hierarchical Bayesian non-linear spatio-temporal model for the spread of invasive species with application to the Eurasian Collared-Dove. *Environmental and Ecological Statistics (to appear)*.
- Huang, H. C. & Cressie, N. A. C. (1996). Spatio-temporal prediction of snow water equivalent using the Kalman filter. *Computational Statistics and Data Analysis*, 22:159–175.
- Huerta, G., Sansó, B., & Stroud, J. R. (2004). A spatio-temporal model for Mexico City ozone levels. *Journal of the Royal Statistical Society, Series C*, 53(2):231–248.
- Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *Journal of Basic Engineering*, 82(Series D):34–45.
- Knorr-Held, L. (2000). Bayesian modelling of inseparable space-time variation in disease risk. *Statistics in Medicine*, 19:2555–2567.
- Knorr-Held, L. & Besag, J. (1998). Modelling risk from a disease in time and space. *Statistics in Medicine*, 17:2045–2060.
- Knorr-Held, L. & Richardson, S. (2003). A hierarchical model for space-time surveillance data on meningococcal disease incidence. *Applied Statistics*, 52:169–183.
- Land, N. M. (1978). Empirical Bayes methods for two-way contingency tables. *Biometrika*, 65:581–590.
- Leonard, T. (1972). Bayesian methods for binomial data. *Biometrika*, 59:869–874.
- Matérn, B. (1960). *Spatial variation*. Springer-Verlag, Berlin, 2nd edition.
- McCullagh, P. & Nelder, J. A. (1989). *Generalized Linear Models*, volume 37 de *Monographs on Statistics and Applied Probability*. Chapman and Hall, London, 2nd. edition.
- Meinhold, R. J. & Singpurwalla, N. (1983). Understanding the Kalman filter. *American Statistician*, 37:123–127.

- Migon, H., Gamerman, D., Lopes, H., & Ferreira, M. A. R. (2005). Dynamic Models. Em *Handbook of Statistics: Bayesian Thinking, Modeling and Computation*, Dey, D. K. & Rao, C., editores, volume 25, páginas 553–588. North Holland.
- Migon, H. S. & Gamerman, D. (1999). *Statistical Inference: an integrated approach*. Arnold, London.
- Nandram, B. & Petrucci, J. (1997). A Bayesian analysis of autoregressive time series panel data. *Journal of Business and Economic Statistics*, 15(3):328–334.
- Nelder, J. A. & Wedderburn, R. (1972). Generalized linear models. *Journal of the Royal Statistical Society, Series A*, 135(3):370–384.
- O’Hagan, A. (1998). A Markov property for covariance structures. Technical report, University of Nottingham.
- Plummer, M., Best, N., Cowles, K., & Vines, K. (2006). *CODA: Output analysis and diagnostics for MCMC*. R package version 0.10-5.
- R Development Core Team (2005). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Viena, Austria. ISBN 3-900051-07-0.
- Raftery, A. E. (1996). Hypothesis Testing and Model Selection. Em *Markov Chain Monte Carlo in practice*, Gilks, W., Richardson, S., & Spiegelhalter, D. J., editores, páginas 163–187, London. Chapman and Hall.
- Raftery, A. E. & Lewis, S. M. (1996). Implementing MCMC. Em *Markov Chain Monte Carlo in practice*, Gilks, W., Richardson, S., & Spiegelhalter, D. J., editores, páginas 115–130, London. Chapman and Hall.
- Ravines, R. R. (2006). *Um esquema eficiente de amostragem em modelos dinâmicos generalizados com aplicações em funções de transferência*. PhD thesis, Departamento de Métodos Estatísticos, IM – UFRJ, Rio de Janeiro, Brasil.
- Robert, C. P. & Casella, G. (1999). *Monte Carlo Statistical Methods*. Springer-Verlag.

- Rubin, D. B. (1984). Bayesianly justifiable and relevant frequency calculations for the applied statistician. *Annals of Statistics*, 12:1151–1172.
- Rue, H. & Follstad, T. (2003). GMRFLib: a C-library for fast and exact simulation of Gaussian Markov random fields. version 1.07. URL: <http://www.math.ntnu.no/~hrue/GMRFLib>.
- Rue, H. & Held, L. (2005). *Gaussian Markov Random Fields: Theory and Applications*, volume 104 de *Monographs on Statistics and Applied Probability*. Chapman and Hall/CRC, London.
- Rue, H. & Tjelmeland, H. (1999). Fitting Gaussian Markov random fields to Gaussian fields. Statistics Report 16, Trondheim, Norway, Department of Mathematical Sciences, Norwegian University of Science and Technology.
- San Martini, A. & Spezzaferri, F. (1984). A predictive model selection criterion. *Journal of the Royal Statistical Society, Series B*, 46:296–303.
- Schmid, V. & Held, L. (2004). Bayesian extrapolation of space-time trends in cancer registry data. *Biometrics*, 60:1034–1042.
- Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & van der Linde, A. (2002). Bayesian measures of model complexity and fit (with discussion and rejoinder). *Journal of the royal Statistical Society, Series B*, 64:583–640.
- Sun, D., Tsutakawa, R. K., Kim, H., & He, Z. (2000). Spatio-temporal interaction with disease mapping. *Statistics in Medicine*, 19:2015–2035.
- Vivar, J. C. (2004). Uma nova classe de modelos espaço-temporais para dados de área. Dissertação de Mestrado, Departamento de Métodos Estatísticos, IM – UFRJ, Rio de Janeiro, Brasil.
- Vivar, J. C. & Ferreira, M. A. R. (2007). Spatio-temporal models for Gaussian areal data. (*Submitted to Journal of Computational and Graphical Statistics*).

- Waller, L. A., Carlin, B. P., Xia, H., & Gelfand, A. E. (1997). Hierarchical spatio-temporal mapping of disease rates. *Journal of the American Statistical Association*, 92:607–617.
- West, M. & Harrison, P. J. (1997). *Bayesian Forecasting and Dynamic Models*. Springer, New York, 2nd. edition.
- West, M., Harrison, P. J., & Migon, H. S. (1985). Dynamic generalized linear models and bayesian forecasting (with discussion). *Journal of the American Statistical Association*, 80:73–96.
- Wikle, C. & Hooten, M. (2006). Hierarchical Bayesian Spatio-Temporal Models for Population Spread. Em *Hierarchical Modelling for the Environmental Sciences: Statistical methods and applications*, Clark, J. A. & Gelfand, A. E., editores, páginas 145–169, Oxford. Oxford University Press.
- Wikle, C. K., Berliner, L. M., & Cressie, N. A. C. (1998). Hierarchical Bayesian space-time models. *Environmental and Ecological Statistics*, 5(2):117–154.
- Wikle, C. K. & Cressie, N. A. C. (1999). A dimension-reduced approach to space-time Kalman filtering. *Biometrika*, 86(4):815–829.
- Xu, K., Wikle, C., & Fox, N. (2005). A kernel-based spatio-temporal dynamical model for nowcasting radar precipitation. *Journal of the American Statistical Association*, 100:1133–1144.
- Zellner, A. & Rossi, P. E. (1984). Bayesian analysis of dichotomous quantal response models. *Journal of Econometrics*, 25:365–393.

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](#)