

MODELOS ESPAÇO-TEMPORAIS MULTI-ESCALA

Adelmo Inácio Bertolde

Tese de Doutorado submetida ao programa de Pós-Graduação em Estatística do Instituto de Matemática da Universidade Federal do Rio de Janeiro - UFRJ, como requisito necessário à obtenção do título de Doutor em Estatística.

Orientador: **Marco Antonio Rosa Ferreira**

Rio de Janeiro

Dezembro, 2007

Livros Grátis

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

Milhares de livros grátis para download.

MODELOS ESPAÇO-TEMPORAIS MULTI-ESCALA

Aluno: **Adelmo Inácio Bertolde**

Orientador: **Marco Antonio Rosa Ferreira**

Tese de Doutorado submetida à banca examinadora do programa de Pós-graduação em Estatística do Instituto de Matemática da Universidade Federal do Rio de Janeiro, como requisito necessário à obtenção do título de Doutor em Estatística.

Aprovada por:

Prof. Marco Antonio Rosa Ferreira

Prof. Dani Gamerman

Prof. Helio dos Santos Migon

Prof. Glaura da Conceição Franco

Prof. Ricardo Sandes Ehlers

Rio de Janeiro

Dezembro, 2007

Resumo

Este texto apresenta uma proposta de modelagem espaço-temporal multi-escala para dados de área. Tal modelo pode ser útil em diversas aplicações tais como em análise ambiental, economia, agricultura e etc.. Fazemos uso da metodologia multi-escala desenvolvida por Kolaczyk e Huang (2001) e decomposos o processo espacial a cada tempo. Modelamos os coeficientes multi-escala ao longo do tempo através de modelos dinâmicos. São realizadas duas aplicações do modelo a dados sócio-econômicos do estado do Espírito Santo, Brasil. Os resultados indicam que o modelo capta bem os movimentos espaço-temporais das séries estudadas.

Palavras chaves: Modelos multi-escala; Inferência Bayesiana; Modelagem espaço-temporal; Dados de área; Resoluções múltiplas.

Abstract

This text presents a multiscale modeling approach for space-time for areal data. Such model may be useful in diverse applications such as in environments, economic, agriculture and others. An uses of the methodology becomes multiscale developed for Kolaczyk and Huang (2001) and we decompose the space process at each time. We modeled the multiscale coefficients along the time through dinamic models. The model was applied two socio-economic the data of the state of the Espirito Santo, Brazil. The results indicate that the model captures well the space-time movements of the studied series.

Keywords: Multiscale models; Bayesian Inference; Space-time modeling; areal data; Multiple resolutions.

Agradecimentos

Inicialmente agradeço à Deus pela oportunidade de crescimento pessoal e profissional, tendo como cenário a linda cidade do Rio de Janeiro, obra prima da natureza.

Várias pessoas colaboraram na realização desse trabalho. Agradeço especialmente ao meu orientador Marco Antonio, que me estimulou, me corrigiu e me proporcionou toda a estrutura para que eu pudesse realizar as tarefas de forma organizada, não apenas como o grande profissional que é, mas também como um grande amigo. À minha esposa Tereza Cristina pelo apoio constante, companheirismo e também pelo seu incentivo nos meus momentos de desânimo. A todos os meus amigos e colegas de curso, particularmente Valmária, Marcelo, Laninha, Juan e Esther. Foram muitas horas de estudo, de sacrifício mas também de muita diversão, aprendizado de vida e trocas de experiências. Agradeço aos professores da pós-graduação do DME pela dedicação e interesse no aprendizado dos alunos. Ao Departamento de Estatística da UFES, à Faperj, à CAPES e ao DME pelo suporte financeiro e oportunidade de aperfeiçoamento acadêmico.

Finalmente, gostaria de dedicar esse trabalho aos meus pais Gabriel e Gicelda, que mesmo sendo pouco alfabetizados sempre me incentivaram nos estudos. Os difíceis, porém, saudosos anos em que lavrei a terra com minha família estão firmemente marcados na minha memória, e me dão a certeza de que é possível a realização dos nossos sonhos, por mais inalcançáveis que possam parecer.

Encerro com uma mensagem que resume um pouco da experiência vivida nos últimos anos:

”Há muito tempo que eu saí de casa. Há muito tempo que eu caí na estrada. Há muito tempo que eu estou na vida. Foi assim que eu quis, e assim eu sou feliz.

Principalmente por poder voltar a todos os lugares onde já cheguei. Pois lá deixei um prato de comida, um abraço amigo, um canto prá dormir e sonhar.

E aprendi que se depende sempre de tanta muita diferente gente. Toda pessoa sempre é as marcas das lições diárias de outras tantas pessoas.

E é tão bonito quando a gente entende que a gente é tanta gente onde quer que a gente vá.

E é tão bonito quando a gente sente que nunca está sozinho por mais que a gente pense estar.

E é tão bonito quando a gente pisa firme nessas linhas que estão nas palmas de nossas mãos.

E é tão bonito quando a gente vai a vida nos caminhos onde bate mais forte o coração.”

Caminhos do Coração (Gonzaguinha)

Sumário

Lista de Tabelas	ix
Lista de Figuras	x
1 Introdução	1
2 Modelagem Multi-escala	4
2.1 Introdução	4
2.2 Estrutura Espacial	5
2.3 Verossimilhança Fatorizada	7
2.4 Distribuições a Priori	9
2.5 Estimação	10
2.5.1 Estimação do Elemento Excluído	11
2.6 Aplicações	12
2.6.1 Aplicação a Dados Simulados	12
2.6.2 Aplicação ao PIB do Espírito Santo	14
3 Modelos Lineares Dinâmicos	20
3.1 Introdução	20
3.2 Modelo Linear Dinâmico	21
3.3 O Filtro de Kalman	22
3.4 Princípios de Inferência Bayesiana	24

3.5	Simulação de Monte Carlo Via Cadeias de Markov	26
3.5.1	Diagnósticos de Convergência	27
3.6	Estimação com Variâncias Desconhecidas	28
3.6.1	Distribuições a Priori	28
3.6.2	Distribuições a Posteriori	29
3.6.3	Simulação de $\boldsymbol{\mu}_{1:T}$ Via FFBS	29
3.6.4	Simulação de V e W	31
4	Modelagem Espaço-Temporal Multi-Escala	33
4.1	Introdução e Notação	33
4.2	Modelo Geral	36
4.3	Um Exemplo Didático	37
5	Modelagem Espaço-Temporal Multi-Escala Com Variâncias Obser-	
	vacionais Constantes	39
5.1	Densidade Conjunta do Modelo	40
5.2	Distribuições a Priori	44
5.3	Simulação e Estimação de $\mu_{0,k,t}$	45
5.3.1	Distribuições a Posteriori	45
5.4	Simulação e Estimação de $\boldsymbol{\theta}_{j,k,t}$	48
5.4.1	Distribuições a Posteriori	48
5.5	Estimação de $\boldsymbol{\mu}_{J-1,t}$	52
5.6	Simulação de Um Processo Multi-Escala	53
5.7	Casos Particulares	54
5.7.1	Variâncias Conhecidas	55
5.7.2	Variância de Sistema Desconhecida	58
5.8	Aplicação aos Dados de PIB <i>per capita</i>	59

6	Modelagem Espaço-Temporal Multi-Escala com Variâncias Observacionais Distintas	72
6.1	Construção do Modelo	72
6.2	Densidade Conjunta do Modelo	74
6.3	Distribuições a Priori	75
6.4	Simulação e Estimação de $\mu_{0,k,t}$	75
6.4.1	Distribuições a Posteriori	75
6.5	Simulação e Estimação de $\theta_{j,k,t}$	76
6.5.1	Distribuições a Posteriori	76
6.5.2	Simulação e Estimação de σ^2	77
6.6	Aplicação aos Dados de Produção Agrícola	77
7	Conclusões	88
7.1	Meta	89
	Referências Bibliográficas	90
A	Algumas Propriedades da Distribuição Normal Multivariada	93
B	Descrição Multi-Escala Para o Caso Normal	95
C	Banco de Dados	98
C.1	Dados Sócio-Econômicos	98
C.2	Estrutura de Partição	99

Lista de Tabelas

5.1	Estatísticas a posteriori para $\psi_{1,k}$	59
5.2	Estatísticas a posteriori para ψ_k	62
5.3	Estatísticas a posteriori para $\psi_{0,k}$	63
5.4	Estatísticas a posteriori para $\psi_{1,k}$	63
6.1	Estatísticas a posteriori para ψ_k	79
6.2	Estatísticas a posteriori para $\psi_{0,k}$	79
6.3	Estatísticas a posteriori para $\psi_{1,k}$	80

Lista de Figuras

1.1	Divisão político administrativa do estado do Espírito Santo por municípios (a), micro-regiões (b) e macro-regiões (c) (ano base 1979).	2
2.1	Processo multi-escala representado do nível $j = 0$ mais agregado (a) até o nível $j = 2$ mais desagregado (c), considerando 4 descendentes para cada ancestral. . .	5
2.2	Função $(2x+10)$. Nível médio (linha), simulações (círculos) e estimativas (linha pontilhada)	13
2.3	Função $\text{Seno}(x)$. Nível médio (linha), simulações (círculos) e estimativas (linha pontilhada)	14
2.4	Função x^2 . Nível médio (linha), simulações (círculos) e estimativas (linha pontilhada)	14
2.5	Log do PIB <i>per capita</i> do E.S. - 1970	15
2.6	Log do PIB <i>per capita</i> do E.S. - 1980	16
2.7	Log do PIB <i>per capita</i> do E.S. - 1985	16
2.8	Log do PIB <i>per capita</i> do E.S. - 1996	17
2.9	Log do PIB <i>per capita</i> do E.S. - 1999	17
2.10	Log do PIB <i>per capita</i> do E.S. - 2000	18
2.11	Log do PIB <i>per capita</i> do E.S. - 2001	18
2.12	Log do PIB <i>per capita</i> do E.S. - 2002	19
2.13	Log do PIB <i>per capita</i> do E.S. - 2003	19

5.1	Resultados comparativos de $\mathbf{y}_{2,..,t}$ (círculos), verdadeiro valor de $\boldsymbol{\mu}_{2,..,t}$ (vermelho), $\hat{\boldsymbol{\mu}}_{2,..,t}$ (preto) e estimativa suavizada $\mathbf{a}_{j,k,T}(-r)$ (azul). Tempos $t = 1, 4, 8, 12$, respectivamente.	55
5.2	Resultados comparativos de $\mathbf{y}_{2,..,t}$ (círculos), verdadeiro valor de $\boldsymbol{\mu}_{2,..,t}$ (vermelho), $\hat{\boldsymbol{\mu}}_{2,..,t}$ (preto) e estimativa suavizada $\mathbf{a}_{j,k,T}(-r)$ (azul). Tempos $t = 15, 18, 22, 25$, respectivamente.	56
5.3	Valores de $y_{2,k,t}$ observados (a), $\boldsymbol{\mu}_{2,k,t}$ verdadeiros (b) e suavizados (c) por $\mathbf{a}_{j,k,T}(-r)$. Tempo $t = 1$	57
5.4	Valores de $y_{2,k,t}$ observados (a), $\boldsymbol{\mu}_{2,k,t}$ verdadeiros (b) e suavizados (c) por $\mathbf{a}_{j,k,T}(-r)$. Tempo $t = 8$	57
5.5	Valores de $y_{2,k,t}$ observados (a), $\boldsymbol{\mu}_{2,k,t}$ verdadeiros (b) e suavizados (c) por $\mathbf{a}_{j,k,T}(-r)$. Tempo $t = 15$	57
5.6	Valores de $y_{2,k,t}$ observados (a), $\boldsymbol{\mu}_{2,k,t}$ verdadeiros (b) e suavizados (c) por $\mathbf{a}_{j,k,T}(-r)$. Tempo $t = 22$	58
5.7	Valores de $y_{2,k,t}$ observados (a), $\boldsymbol{\mu}_{2,k,t}$ verdadeiros (b) e suavizados (c) por $\mathbf{a}_{j,k,T}(-r)$. Tempo $t = 25$	58
5.8	Resultados comparativos de $\mathbf{y}_{2,k,.}$ (círculos), verdadeiro valor de $\boldsymbol{\mu}_{2,k,.}$ (vermelho) e média a posteriori para $\boldsymbol{\mu}_{2,k,.}$ (preto) para 6 áreas fictícias, ao longo do tempo. Linhas tracejadas inferior e posterior indicam os percentis 2.5 e 97.5 de $\hat{\boldsymbol{\mu}}_{2,k,.}$, respectivamente.	60
5.9	Análise Exploratória por períodos do PIB <i>per capita</i> municipal do E.S. Fonte: IBGE/IPEA.	61
5.10	Análise Exploratória por períodos do Log(PIB <i>per capita</i>) municipal do E.S. Fonte: IBGE/IPEA.	61
5.11	Trajectoria (a) e histograma (b) para valores de σ^2 gerados a posteriori. Estatística de diagnóstico de Gelman e Rubin igual a 1.02 e de Geweke iguais a -0.12 e 0.41, respectivamente.	62

5.12	Respectivamente, trajetória e histograma para $\mu_{2,k,1}$ a posteriori nos municípios de Guarapari (a e a'), Conceição da Barra (b e b'), Colatina (c e c') e Castelo (d e d'). Estatísticas de diagnóstico de Gelman iguais a 1, 1.01, 1.01 e 1.01, respectivamente.	64
5.13	Resultados comparativos de $y_{0,k,.}$ (círculos) e média a posteriori de $\mu_{0,k,.}$ (preto) por macro-região. Linhas tracejadas inferior e posterior indicam os percentis 2.5 e 97.5 de $\hat{\mu}_{0,k,.}$, respectivamente.	65
5.14	Resultados comparativos de $y_{2,k,.}$ (círculos) e média a posteriori de $\mu_{2,k,.}$ (preto) para alguns municípios da região metropolitana de Vitória. Linhas tracejadas inferior e posterior indicam os percentis 2.5 e 97.5 de $\hat{\mu}_{2,k,.}$, respectivamente.	66
5.15	Resultados comparativos de $y_{1,k,.}$ (círculos) e média a posteriori de $\mu_{1,k,.}$ (preto) para as micro-regiões 4 e 5. Linhas tracejadas inferior e posterior indicam os percentis 2.5 e 97.5 de $\hat{\mu}_{1,k,.}$, respectivamente.	67
5.16	Log(PIB per capita) do E.S. - 1970	67
5.17	Log(PIB per capita) do E.S. - 1980	68
5.18	Log(PIB per capita) do E.S. - 1985	68
5.19	Log(PIB per capita) do E.S. - 1996	69
5.20	Log(PIB per capita) do E.S. - 1999	69
5.21	Log(PIB per capita) do E.S. - 2000	70
5.22	Log(PIB per capita) do E.S. - 2001	70
5.23	Log(PIB per capita) do E.S. - 2002	71
5.24	Log(PIB per capita) do E.S. - 2003	71
6.1	Análise Exploratória do logaritmo do valor municipal da produção agrícola para o estado do E.S. Fonte: IBGE/IPEA.	78
6.2	Trajetoária (a) e histograma (b) para valores de σ^2 gerados a posteriori. Estatística de diagnóstico de Gelman e Rubin igual a 1.02 e de Geweke iguais a 1.98 e 0.42, respectivamente.	79

6.3	Resultados comparativos de $\mathbf{y}_{0,k..}$ (círculos) e média a posteriori de $\boldsymbol{\mu}_{0,k..}$ (preto) por macro-região. Linhas tracejadas inferior e posterior indicam os percentis 2.5 e 97.5 de $\hat{\boldsymbol{\mu}}_{0,k..}$, respectivamente.	81
6.4	Resultados comparativos de $\mathbf{y}_{2,k..}$ (círculos) e média a posteriori de $\boldsymbol{\mu}_{2,k..}$ (preto) por município. Linhas tracejadas inferior e posterior indicam os percentis 2.5 e 97.5 de $\hat{\boldsymbol{\mu}}_{2,k..}$, respectivamente.	82
6.5	Respectivamente, trajetória e histograma para $\boldsymbol{\mu}_{2,k,2}$ a posteriori nos municípios de Cariacica (a e a'), Guarapari (b e b'), Santa Leopoldina (c e c') e Nova Venécia (d e d'). Estatísticas de diagnóstico de Gelman iguais a 1.05, 1, 1.01 e 1.09, respectivamente.	83
6.6	Valor de produção agrícola do E.S. (em milhares de reais) - 1991	84
6.7	Valor de produção agrícola do E.S. (em milhares de reais) - 1993	84
6.8	Valor de produção agrícola do E.S. (em milhares de reais) - 1995	85
6.9	Valor de produção agrícola do E.S. (em milhares de reais) - 1997	85
6.10	Valor de produção agrícola do E.S. (em milhares de reais) - 1999	86
6.11	Valor de produção agrícola do E.S. (em milhares de reais) - 2001	86
6.12	Valor de produção agrícola do E.S. (em milhares de reais) - 2003	87
6.13	Valor de produção agrícola do E.S. (em milhares de reais) - 2005	87

Capítulo 1

Introdução

As variações aleatórias presentes nos fenômenos naturais são objeto de estudo do homem há muitos séculos e objetivando contribuir para a explicação dos mesmos foram criados os modelos estatísticos. O interesse no comportamento desses fenômenos ao longo do espaço e do tempo motivou o surgimento de uma metodologia voltada para os estudos dos processos espaço-temporais.

Os modelos estatísticos desenvolvidos para dados espaço-temporais apresentaram um grande avanço nos últimos 15 anos, com trabalhos que vão desde aplicações em ciências ambientais, análise de dados econômicos e mapeamento de doenças até a taxas de violência, entre outros. Os fatores que mais têm impulsionado tal avanço são a quantidade cada vez maior de dados disponíveis e o progresso significativo dos métodos de simulação bayesianos, este advindo do progresso computacional.

Este trabalho tem por objetivo apresentar uma proposta de modelagem espaço-temporal multi-escala para dados de área. Utilizamos a metodologia multi-escala desenvolvida por Kolaczyk e Huang (2001) e decomposmos o processo espacial a cada tempo. Modelamos os coeficientes multi-escala ao longo do tempo através de modelos dinâmicos. Com base na estrutura hierárquica das observações, o modelo

multi-escala induz dependência espacial ao definir os grupos de vizinhos, tendo como consequência uma simplificação no trato das matrizes de covariância envolvidas.

Realizamos três aplicações da modelagem proposta: a primeira a um processo espaço-temporal multi-escala simulado e as outras duas fazendo uso de dados sócio-econômicos relativos ao estado do Espírito Santo, Brasil. Como motivação inicial, temos, como objeto de estudo, o mapa do estado do Espírito Santo dividido em 3 níveis de partições respectivas: municípios, micro-regiões e macro-regiões.

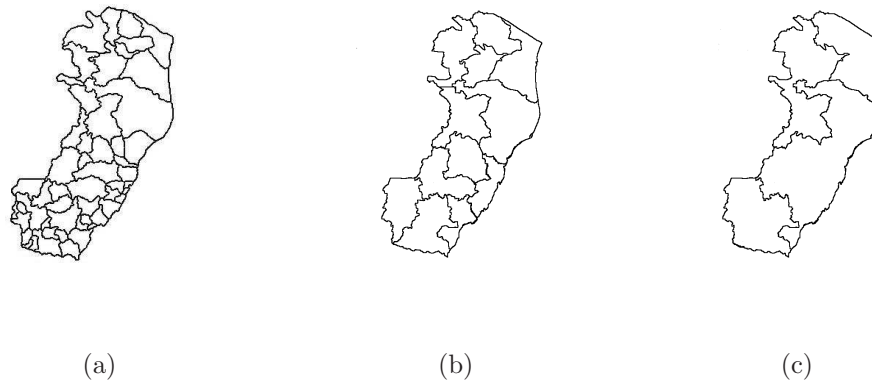


Figura 1.1: Divisão político administrativa do estado do Espírito Santo por municípios (a), micro-regiões (b) e macro-regiões (c) (ano base 1979).

O capítulo 2 descreve a metodologia multi-escala proposta por Kolaczyk e Huang (2001) e relata alguns trabalhos correlacionados desenvolvidos. São apresentados os conceitos básicos sobre partição hierárquica e a definição de verossimilhança em um processo multi-escala.

No capítulo 3 apresentamos um resumo de um modelo dinâmico linear, onde descrevemos as equações do filtro de Kalman, juntamente com os conceitos de Inferência Bayesiana, de técnicas de simulação estocástica e de convergência utilizados. Os capítulos 4 e 5 tratam da proposta do Modelo Multi-Escala Espaço-Temporal

objeto desse trabalho, onde supomos variâncias proporcionais e iguais para cada elemento do nível mais desagregado. Realizamos uma aplicação do modelo a um processo multi-escala simulado e a dados de PIB *per capita*.

Fazemos uma extensão do modelo no capítulo 6, onde supomos variâncias envolvidas distintas. Ao final apresentamos uma aplicação a dados de valor de produção agrícola do estado do Espírito Santo. Finalmente, no capítulo 7 relatamos as conclusões e metas desse trabalho. Ao final estão as referências bibliográficas e anexos utilizados no trabalho de pesquisa de tese.

Capítulo 2

Modelagem Multi-escala

2.1 Introdução

Este capítulo traz um breve resumo da metodologia multi-escala desenvolvida por Kolaczyk e Huang (2001), que faz uso da inferência para analisar processos observados em mais de um nível de escala. Ressalta-se a importância de tal método, visto que a maioria dos estudos envolvendo mensurações em múltiplas escalas se presta somente à análise descritiva (Kolaczyk e Huang, 2001). Em particular, o trabalho de Kolaczyk e Huang (2001) é importante para o entendimento do conteúdo deste texto pois utilizamos sua decomposição multi-escala na construção de nossos modelos espaço-temporais multi-escala.

A hipótese estrutural é a de existência de uma partição definida hierarquicamente, correspondendo a sucessivas agregações no domínio dos dados. A inter-relação entre os níveis de resolução permite eficiente suavização espacial do fenômeno em estudo (Kolaczyk e Huang, 2001). Neste capítulo, ilustramos esta metodologia com uso de dados simulados e com dados de PIB *per capita* por município do estado do Espírito Santo, Brasil. Mais detalhes a respeito da estrutura multi-escala do estado do Espírito Santo são encontrados no anexo C.

2.2 Estrutura Espacial

A modelagem multi-escala assume inicialmente um particionamento hierárquico de uma região S , onde S tipicamente é tratada como um sub-espaco bidimensional. Considere um conjunto de medidas em um domínio S onde se está interessado em conhecer a média $[\mu(s) : s \in S]$ do processo. Matematicamente, expressa-se essa idéia através da hipótese da existência de uma partição (B_1, \dots, B_n) de S , onde $B_k \subset S$ tal que $(B_k \cap B_{k'}) = \emptyset$, para $k \neq k'$, sendo $\cup_k B_k = S$ e $k = 1, \dots, n$. Para cada elemento B_k desta partição, assume-se a existência de uma medida y_k , relacionada à média do processo $\mu(s)$ através do valor esperado $E[y_k] = \mu_k$, onde $\mu_k \equiv \mu(B_k) \equiv \int_{B_k} \mu(s) ds$. Assume-se também que os y_k 's são condicionalmente independentes dado os μ_k 's respectivos.

Formalmente, tal estrutura hierárquica multi-escala caracteriza-se em termos de um conjunto B composto por J partições de S , para algum inteiro $J > 1$. Para reforçar o entendimento, vejamos a figura 2.1, que apresenta um processo multi-escala fictício para $J = 3$ níveis de agregações e 4 descendentes para cada ancestral:

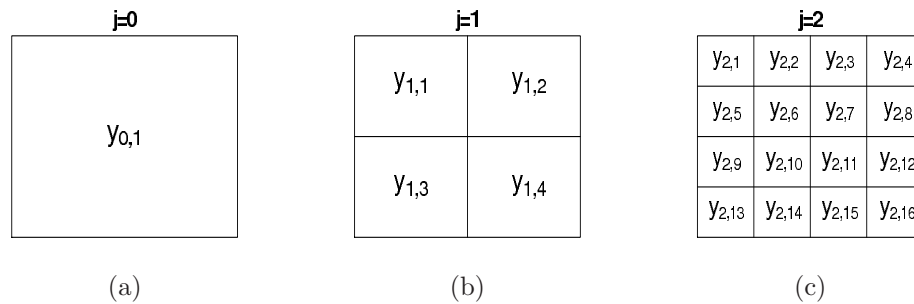


Figura 2.1: Processo multi-escala representado do nível $j = 0$ mais agregado (a) até o nível $j = 2$ mais desagregado (c), considerando 4 descendentes para cada ancestral.

Denota-se os elementos da partição no nível j de resolução por $B_{j,1}, \dots, B_{j,n_j}$.

Para cada um dos níveis de escala j , $j = 0, 1, \dots, J-1$, a coleção $\{B_{j,k}\}_{k=1}^{n_j}$ é definida satisfazendo a duas condições:

i) formar uma partição própria de S e,

ii) que cada elemento $B_{j,k}$ seja expresso como uma união única dos elementos da partição na escala $(j+1)$. Isto significa que $B_{j,k} = \cup_{(j+1,k') \in D_{j,k}} B_{j+1,k'}$, onde $D_{j,k}$ refere-se à coleção de índices espaciais dos descendentes de $B_{j,k}$, para $j = 0, \dots, J-2$. Por exemplo, a partir da figura 2.1, podemos verificar que o conjunto dos índices espaciais de $B_{0,1}$ é dado por

$$D_{0,1} = \{(1, 1), (1, 2), (1, 3), (1, 4)\}$$

e o conjunto dos índices espaciais de $B_{1,3}$ é dado por

$$D_{1,3} = \{(2, 9), (2, 10), (2, 13), (2, 14)\}.$$

Ressalta-se que ao se determinar quais são os descendentes de cada respectivo ancestral, fica estabelecida a relação de vizinhança entre as áreas em cada nível de partição.

Suponha definida a partição hierárquica $\{B_{j,k}\}_{k=1}^{n_j}$ e considere um elemento $B_{j,k}$ desta partição. Kolaczyk e Huang (2001) consideram agregação por somas, ou seja, a variável $y_{j,k}$ é definida como

$$y_{j,k} = \sum_{(j+1,k') \in D_{j,k}} y_{j+1,k'}. \quad (2.1)$$

Consequentemente, o valor esperado $\mu_{j,k}$ da variável $y_{j,k}$ satisfaz

$$\mu_{j,k} = \sum_{(j+1,k') \in D_{j,k}} \mu_{j+1,k'}. \quad (2.2)$$

A partir da expressão (2.1) conclui-se que as medidas $\{y_{j,k}\}_{k=1}^{n_j}$ do nível j estão numa escala mais agregada do que $\{y_{j+1,k'}\}_{k'=1}^{n_{j+1}}$. Há, portanto, uma perda de informação

ao se utilizar as medidas num nível j mais agregado. Segundo Kolaczyk e Huang (2001), tal perda poderá variar espacialmente, dependendo do grau de homogeneidade da média do processo $\mu(s)$ no domínio S . Logo, torna-se relevante considerar o efeito de mudança dos níveis de escala dos dados no processo de modelagem. Com este objetivo, Kolaczyk e Huang (2001) propuseram a construção da chamada *verossimilhança fatorizada*, apresentada na seção seguinte.

2.3 Verossimilhança Fatorizada

Note que a distribuição de probabilidade dos descendentes de $B_{j,k}$, dado μ e $y_{j,k}$, é degenerada. Uma forma de resolver esta degeneração é considerar os descendentes, menos um deles; o descendente é excluído de forma arbitrária e não altera a análise. Assim sendo, consideremos a seguinte definição:

- $D_{j,k}^*$ é a coleção de índices espaciais dos descendentes de $B_{j,k}$ excluindo um deles, sendo $j = 0, \dots, J - 2$.

Logo, $\mathbf{y}_{D_{j,k}^*}$ é o vetor de medidas dos descendentes do elemento $B_{j,k}$ excluindo um deles, com definição análoga para $\boldsymbol{\mu}_{D_{j,k}^*}$. Daí que, explorando a independência condicional dos $y_{j,k}$, tem-se que:

$$\prod_{k=1}^{n_{J-1}} f(y_{J-1,k}|\mu) = \prod_{k=1}^{n_{J-2}} f(y_{J-2,k}|\mu) \times \prod_{k=1}^{n_{J-2}} f(\mathbf{y}_{D_{j-2,k}^*}|y_{J-2,k}, \mu). \quad (2.3)$$

O lado esquerdo de (2.3) é a verossimilhança das observações no nível de escala $J - 1$, enquanto que o primeiro termo do lado direito é a verossimilhança no nível de escala $J - 2$. Logo, o segundo termo do lado direito contém a perda de informação devida a agregação realizada da escala $J - 1$ para $J - 2$. Kolaczyk e Huang (2001) definem, portanto, a expressão para a verossimilhança fatorizada como: assumindo que o vetor $(y_{J-1,1}, \dots, y_{J-1,n_{J-1}})'$ e a estruturação de partições hierárquicas estejam

definidos, a verossimilhança fatorizada pode ser descrita por:

$$\prod_{k=1}^{n_{J-1}} f(y_{J-1,k}|\mu) = \prod_{k=1}^{n_0} f(y_{0,k}|\mu) \times \prod_{j=0}^{J-2} \prod_{k=1}^{n_j} f(\mathbf{y}_{D_{j,k}^*} | y_{j,k}, \mu). \quad (2.4)$$

Recorde que, dado um vetor $\mathbf{y}_j = (y_{j,1}, \dots, y_{j,n_j})'$ com elementos independentes e normalmente distribuídos, pode-se expressá-lo por $\mathbf{y}_j | \boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j \sim N(\boldsymbol{\mu}_j, \boldsymbol{\Sigma}_j)$, onde $\boldsymbol{\mu}_j = (\mu_{j,1}, \dots, \mu_{j,n_j})'$ e $\boldsymbol{\Sigma}_j = \text{diag}(\boldsymbol{\sigma}_j^2)$, sendo $\boldsymbol{\sigma}_j^2 = (\sigma_{j,1}^2, \dots, \sigma_{j,n_j}^2)'$. No caso do vetor $\mathbf{y}_{J-1} = (y_{J-1,1}, \dots, y_{J-1,n_{J-1}})'$ ser modelado por um processo $N(\boldsymbol{\mu}_{J-1}, \boldsymbol{\Sigma}_{J-1})$, a fatorização (2.4) toma a seguinte forma:

$$f(\mathbf{y}_{J-1} | \boldsymbol{\mu}_{J-1}, \boldsymbol{\Sigma}_{J-1}) = f(\mathbf{y}_0 | \boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0) \prod_{j=0}^{J-2} \prod_{k=1}^{n_j} f(\mathbf{y}_{D_{j,k}^*} | y_{j,k}, \boldsymbol{\theta}_{j,k}, \boldsymbol{\Omega}_{j,k}),$$

onde:

- i) $\mathbf{y}_0 | \boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0$ tem distribuição $N(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$.
- ii) $\mathbf{y}_{D_{j,k}^*} | y_{j,k}, \boldsymbol{\theta}_{j,k}, \boldsymbol{\Omega}_{j,k} \sim N(\boldsymbol{\nu}_{j,k} y_{j,k} + \boldsymbol{\theta}_{j,k}, \boldsymbol{\Omega}_{j,k})$, sendo

$$\boldsymbol{\nu}_{j,k} = \frac{\boldsymbol{\sigma}_{D_{j,k}^*}^2}{\sigma_{j,k}^2},$$

$$\boldsymbol{\theta}_{j,k} = \boldsymbol{\mu}_{D_{j,k}^*} - \boldsymbol{\nu}_{j,k} \mu_{j,k} \text{ e}$$

$$\boldsymbol{\Omega}_{j,k} = \boldsymbol{\Sigma}_{D_{j,k}^*} - \left[\frac{\boldsymbol{\sigma}_{D_{j,k}^*}^2 (\boldsymbol{\sigma}_{D_{j,k}^*}^2)'}{\sigma_{j,k}^2} \right],$$

considerando que:

- $\boldsymbol{\sigma}_{D_{j,k}^*}^2$ é o vetor de variâncias dos descendentes de $B_{j,k}$ excluindo um deles;
- $\boldsymbol{\Sigma}_{D_{j,k}^*}$ é uma matriz diagonal formada pelos elementos de $\boldsymbol{\sigma}_{D_{j,k}^*}^2$ e
- $n_{D_{j,k}^*}$ é o número de descendentes de $B_{j,k}$ menos um.

2.4 Distribuições a Priori

A fatorização (2.4) reparametriza o modelo normal inicialmente parametrizado pelo vetor de médias $\boldsymbol{\mu}_{J-1}$ em termos do vetor $(\boldsymbol{\mu}_0, \boldsymbol{\theta}_0, \dots, \boldsymbol{\theta}_{J-2})'$. Esta é uma decomposição multi-escala da média do processo análoga à decomposição em ondaletas, onde $\boldsymbol{\theta}_{j,k}$ tem uma estrutura similar. Mais especificamente, $\boldsymbol{\theta}_{j,k}$ explica como a medida para cada nó é dividida entre os respectivos descendentes. Tal como na análise por ondaletas, Kolaczyk e Huang (2001) atribuem uma priori independente para cada elemento de $(\boldsymbol{\mu}_0, \boldsymbol{\theta}_0, \dots, \boldsymbol{\theta}_{J-2})'$.

Uma escolha natural para a média do nível mais agregado é assumir uma priori conjugada $\boldsymbol{\mu}_0 \sim N(m\boldsymbol{\mu}_0, \boldsymbol{\Phi}_0)$, sendo para $\boldsymbol{\Phi}_0$ uma distribuição pouco informativa. Para os $\boldsymbol{\theta}_{j,k}$'s existem algumas possíveis especificações que implicam em diferentes níveis de homogeneidade para o processo latente $\mu(s)$. Numa primeira especificação, Kolaczyk e Huang (2001) assumem uma priori conjugada

$$\boldsymbol{\theta}_{j,k} \mid \boldsymbol{\Phi}_{j,k} \sim N(\mathbf{0}, \boldsymbol{\Phi}_{j,k}), \quad (2.5)$$

onde $\boldsymbol{\Phi}_{j,k}$, $j = 0, \dots, J-1$, $k = 1, \dots, n_j$ são matrizes de covariâncias conhecidas. O uso dessa priori permite um rápido cálculo das estatísticas a posteriori para as médias e variâncias. Kolaczyk e Huang (2001) também consideram uma priori baseada numa distribuição mista, isto é,

$$\boldsymbol{\theta}_{j,k} \mid \eta_{j,k}, \boldsymbol{\Phi}_{j,k}^0, \boldsymbol{\Phi}_{j,k}^1 \sim (1 - \eta_{j,k})N(\mathbf{0}, \boldsymbol{\Phi}_{j,k}^0) + \eta_{j,k}N(\mathbf{0}, \boldsymbol{\Phi}_{j,k}^1), \quad (2.6)$$

onde $\eta_{j,k} \sim \text{Bernoulli}(p_{j,k})$, e $\boldsymbol{\Phi}_{j,k}^0$ e $\boldsymbol{\Phi}_{j,k}^1$ são matrizes de covariâncias conhecidas tal que $\boldsymbol{\Phi}_{j,k}^0$ implica em uma distribuição bastante concentrada em torno da média $\mathbf{0}$, enquanto que $\boldsymbol{\Phi}_{j,k}^1$ impõe uma distribuição bastante dispersa. Esta especificação considera, imparcialmente, processos latentes $\mu(s)$ não homogêneos.

2.5 Estimação

A estimação do processo latente para o nível mais desagregado $\boldsymbol{\mu}_{J-1}$ é executada assumindo uma função de perda quadrática. Em consequência, o estimador Bayesiano é a média a posteriori de $\boldsymbol{\mu}_{J-1}$.

Um importante resultado oriundo da construção realizada por Kolaczyk e Huang (2001) é que a distribuição a posteriori da média do processo para o nível mais desagregado $\boldsymbol{\mu}_{J-1}$ é fatorizada em termos de uma nova reparametrização, expressa por:

$$f(\boldsymbol{\mu}_{J-1}|y_{J-1}) = f(\boldsymbol{\mu}_0|\mathbf{y}_0) \prod_{j=0}^{J-2} \prod_{k=1}^{n_j} f(\boldsymbol{\theta}_{j,k} | \mathbf{y}_{D_{j,k}^*}, y_{j,k}). \quad (2.7)$$

Em consequência de (2.7), a média a posteriori para $\boldsymbol{\mu}_{J-1}$ pode ser recursivamente calculada usando a expressão

$$\hat{\boldsymbol{\mu}}_{D_{j,k}^*} = \hat{\boldsymbol{\theta}}_{j,k} + \boldsymbol{\nu}_{j,k} \hat{\mu}_{j,k},$$

onde $\hat{\boldsymbol{\theta}}_{j,k} = E[\boldsymbol{\theta}_{j,k} | \mathbf{y}_{D_{j,k}^*}, y_{j,k}]$. O procedimento recursivo se inicia com o cálculo da média a posteriori do processo latente para o nível mais agregado:

$$\hat{\boldsymbol{\mu}}_0 = (\boldsymbol{\Sigma}_0^{-1} + \boldsymbol{\Phi}_0^{-1})^{-1}(\boldsymbol{\Sigma}_0^{-1}y_0 + \boldsymbol{\Phi}_0^{-1}\mathbf{m}\boldsymbol{\mu}_0).$$

Note que quando a priori para $\boldsymbol{\mu}_0$ é pouco informativa, a média a posteriori será próxima de \mathbf{y}_0 .

Quando a priori conjugada normal (2.5) é usada, a distribuição a posteriori para $\boldsymbol{\theta}_{j,k}$ tem densidade normal com média dada por

$$\hat{\boldsymbol{\theta}}_{j,k} = \boldsymbol{\Phi}_{j,k}(\boldsymbol{\Phi}_{j,k} + \boldsymbol{\Omega}_{j,k})^{-1}(\mathbf{y}_{D_{j,k}^*} - \boldsymbol{\nu}_{j,k}y_{j,k}). \quad (2.8)$$

Quando a priori de misturas de normais (2.6) é usada, a distribuição a posteriori de $\boldsymbol{\theta}_{j,k}$ é também uma mistura de normais. Nesse caso, a média a posteriori será

expressa por

$$\hat{\boldsymbol{\theta}}_{j,k} = q_{j,k} \hat{\boldsymbol{\theta}}_{j,k}^1 + (1 - q_{j,k}) \hat{\boldsymbol{\theta}}_{j,k}^0,$$

onde $\hat{\boldsymbol{\theta}}_{j,k}^0$ e $\hat{\boldsymbol{\theta}}_{j,k}^1$ são calculados através de (2.8) usando $\boldsymbol{\Phi}_{j,k}^0$ e $\boldsymbol{\Phi}_{j,k}^1$, respectivamente, em substituição a $\boldsymbol{\Phi}_{j,k}$. Além disso,

$$\begin{aligned} q_{j,k} &= P(\eta_{j,k} = 1 \mid \mathbf{y}_{D_{j,k}^*}, y_{j,k}) \\ &= \frac{O_{j,k}}{1 + O_{j,k}}, \end{aligned}$$

onde $O_{j,k}$ é a razão de chances a posteriori para cada nó (j, k) , variando entre 0 e 1 e controlando a componente de mistura, isto é,

$$\begin{aligned} O_{j,k} &= \frac{p_{j,k}}{1 - p_{j,k}} \times \\ &\frac{|\boldsymbol{\Phi}_{j,k}^0 + \boldsymbol{\Omega}_{j,k}|^{0.5} \exp\{0.5(\mathbf{y}_{D_{j,k}^*} - \boldsymbol{\nu}_{j,k} y_{j,k})' (\boldsymbol{\Phi}_{j,k}^0 + \boldsymbol{\Omega}_{j,k})^{-1} (\mathbf{y}_{D_{j,k}^*} - \boldsymbol{\nu}_{j,k} y_{j,k})\}}{|\boldsymbol{\Phi}_{j,k}^1 + \boldsymbol{\Omega}_{j,k}|^{0.5} \exp\{0.5(\mathbf{y}_{D_{j,k}^*} - \boldsymbol{\nu}_{j,k} y_{j,k})' (\boldsymbol{\Phi}_{j,k}^1 + \boldsymbol{\Omega}_{j,k})^{-1} (\mathbf{y}_{D_{j,k}^*} - \boldsymbol{\nu}_{j,k} y_{j,k})\}}. \end{aligned}$$

2.5.1 Estimação do Elemento Excluído

Observe que se σ_j^2 for conhecido, $\boldsymbol{\Omega}_{j,k}$ também o será. Nesse caso, somente restaria estimar o vetor $(\boldsymbol{\mu}_0, \boldsymbol{\theta}_0, \dots, \boldsymbol{\theta}_{J-2})'$, que é uma reparametrização do vetor de médias $\boldsymbol{\mu}_{J-1}$.

Após a obtenção do estimador de $(\boldsymbol{\mu}_0, \boldsymbol{\theta}_0, \dots, \boldsymbol{\theta}_{J-2})'$, o passo seguinte é encontrar, recursivamente, os estimadores de $(\boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_{J-1})'$. Para tal, necessita-se, além dos estimadores dos $\boldsymbol{\theta}$'s, dos estimadores das médias do ancestral e dos descendentes. Recorde que a eliminação de um dos elementos no proceso de inferência se faz necessária, pois do contrário, a verossimilhança fatorizada envolveria uma função degenerada, visto que $y_{j,k} = \sum_{(j+1,k') \in D_{j,k}} y_{j+1,k'}$.

Observe que os $\boldsymbol{\theta}$'s e os $\boldsymbol{\mu}_{D_{j,k}^*}$'s têm dimensões iguais a $n_{D_{j,k}^*}$. Assim sendo, o cálculo da estimativa de $\boldsymbol{\mu}_{D_{j,k}}$ pode ser realizado da seguinte forma:

- Inicialmente toma-se o valor de $\nu_{j,k}$ e o estimador de $\mu_{j,k}$.
- Em seguida, estima-se $\theta_{j,k}$, este também possuindo dimensão igual a $n_{D_{j,k}^*}$.
- Posteriormente, com uso da expressão $\theta_{j,k} = \mu_{D_{j,k}^*} - \nu_{j,k}\mu_{j,k}$, calcula-se o estimador de $\mu_{D_{j,k}^*}$.
- Finalmente, em decorrência da expressão (2.2), observe que o descendente excluído pode ser estimado pela diferença entre o estimador de $\mu_{j,k}$ e a soma dos elementos do estimador de $\mu_{D_{j,k}^*}$. Isto é, igual a

$$\hat{\mu}_{j,k} - \sum_{(j+1,k') \in D_{j,k}^*} \hat{\mu}_{j+1,k'}.$$

Desse modo, o estimador de $\mu_{D_{j,k}^*}$ poderá ser calculado sucessivamente ao longo do tempo t .

2.6 Aplicações

Nesta seção apresentamos duas aplicações iniciais da modelagem multi-escala proposta por Kolaczyk e Huang (2001): a primeira, utilizando os dados simulados de 3 curvas (seno, parábola e uma reta) e a segunda fazendo uso de dados do PIB *per capita* do estado do Espírito Santo. Os dados referentes ao PIB e população por município estão disponibilizados na base de dados mantida pelo IPEA (Instituto de Pesquisa Econômica e Aplicada) e podem ser obtidos no endereço eletrônico <http://www.ipeadata.gov.br>.

2.6.1 Aplicação a Dados Simulados

Geramos dados de uma reta, de um seno e de uma parábola. Em cada caso as gerações foram realizadas em 3 níveis de resolução. Cada ancestral do nível $j = 0$ e $j = 1$ possui, respectivamente, 3 descendentes, totalizando 45 áreas no

nível de maior desagregação. Os ruídos são normalmente distribuídos. Para cada nível j consideramos a mesma estrutura de covariância dentro de cada grupo de descendentes e o vetor $\nu_{j,k}$ constante para todos os níveis. Os dados simulados foram, respectivamente:

- Valores simulados da função $(2x + 10)$, com ruído $N(0,3)$.
- Valores simulados da função $\text{seno}(x)$, com ruído $N(0,0.5)$.
- Valores simulados da função x^2 , com ruído $N(0,9)$.

Para cada uma das curvas, valores observados e estimados para as 45 áreas do nível mais desagregado são mostrados nas figuras 2.2 a 2.4. Destas, pode-se concluir que a estimação multi-escala mostrou-se satisfatória, sendo pouco sensível a dados extremos.

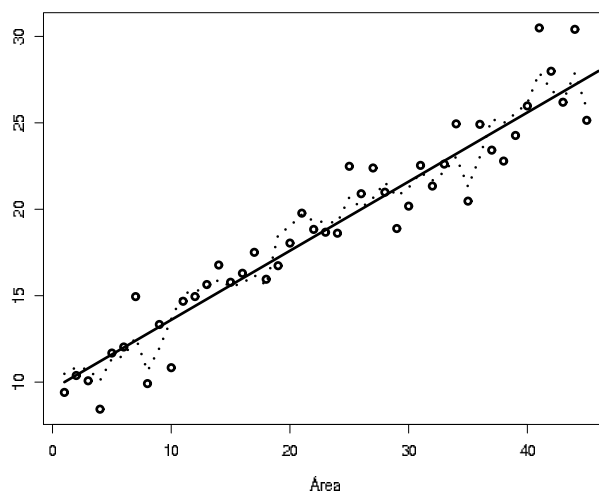


Figura 2.2: Função $(2x+10)$. Nível médio (linha), simulações (círculos) e estimativas (linha pontilhada)

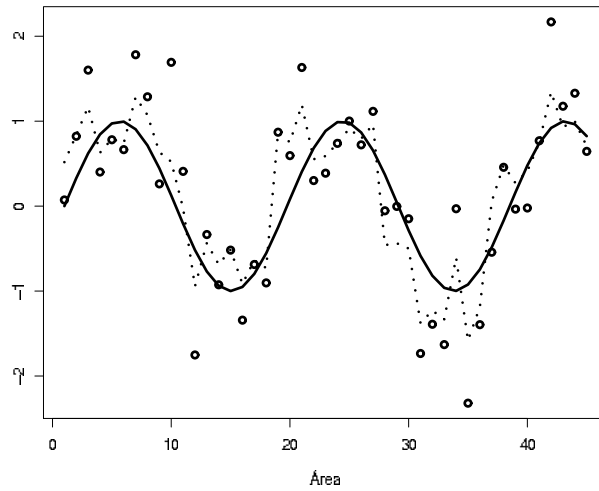


Figura 2.3: Função $\text{Sen}(x)$. Nível médio (linha), simulações (círculos) e estimativas (linha pontilhada)

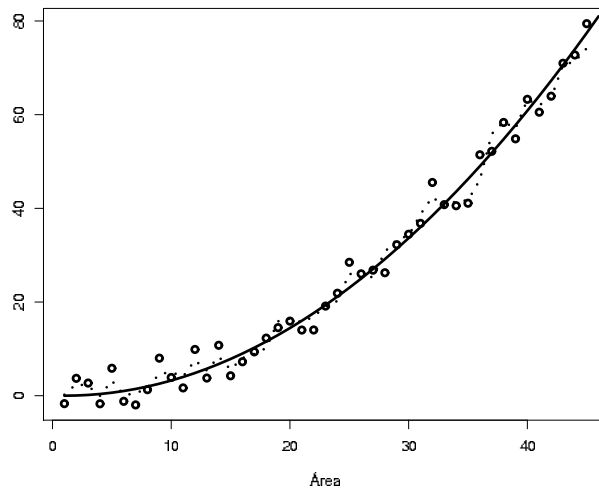


Figura 2.4: Função x^2 . Nível médio (linha), simulações (círculos) e estimativas (linha pontilhada)

2.6.2 Aplicação ao PIB do Espírito Santo

A hierarquia espacial é utilizada aqui em 3 níveis, a saber: no nível mais agregado estão as macro-regiões, em um total de 4; o nível seguinte é o das micro-regiões, composto por 12 elementos, e, por último, o nível mais desagregado que é formado por 52 municípios. Utilizamos a divisão político administrativa municipal do Espírito

Santo relativa ao ano de 1979 (para mais detalhes ver anexo C).

Os dados do PIB estão em milhares de reais tendo 2000 como ano base, deflacionados por um índice implícito do PIB nacional. O período utilizado compreende os anos de 1970, 1980, 1985, 1996, 1999, 2000, 2001, 2002 e 2003. Para minimizar efeitos de assimetria bem como de valores discrepantes, aplicamos o logaritmo aos dados.

Visualizando-se os mapas observados e estimados para o log do PIB *per capita* (figuras 2.5 a 2.13), observa-se que há uma dinâmica na distribuição espaço-temporal da variável em questão. O que nos propomos nesse trabalho é modelar tal dinâmica através de uma modelagem multi-escala que contemple a interdependência espaço-temporal. Para dar suporte ao que se pretende, no próximo capítulo fazemos uma revisão da metodologia utilizada na análise temporal e da parte de simulação do modelo.

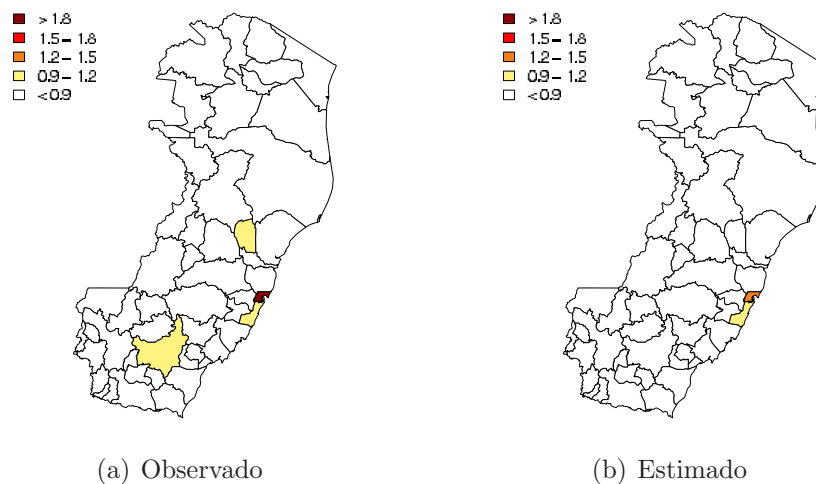


Figura 2.5: Log do PIB *per capita* do E.S. - 1970

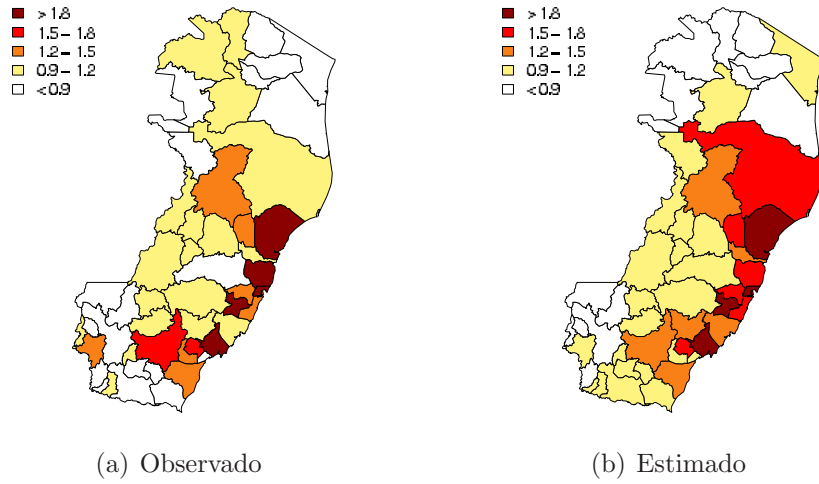


Figura 2.6: Log do PIB *per capita* do E.S. - 1980

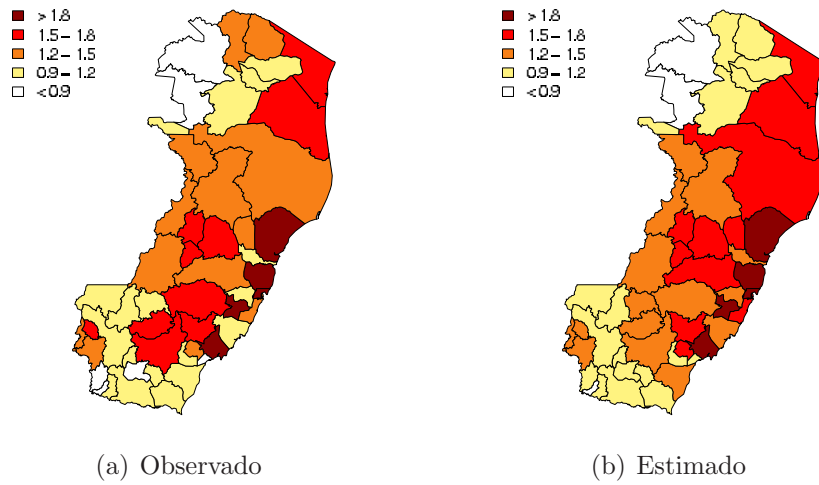


Figura 2.7: Log do PIB *per capita* do E.S. - 1985

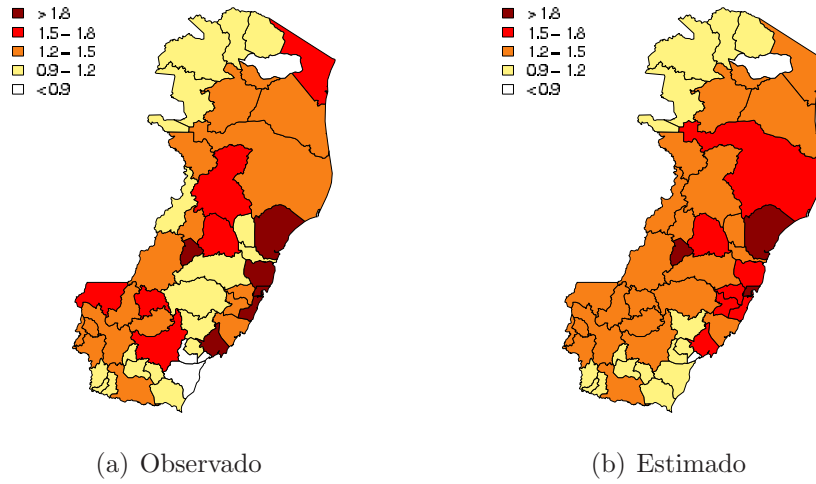


Figura 2.8: Log do PIB *per capita* do E.S. - 1996

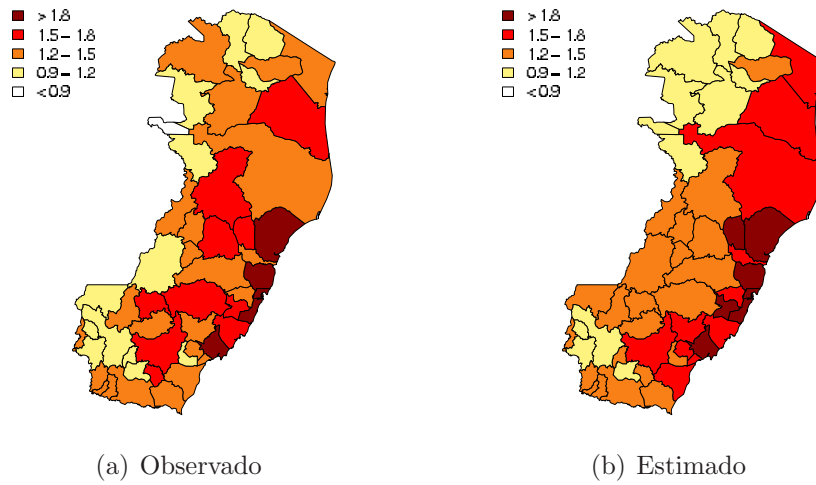


Figura 2.9: Log do PIB *per capita* do E.S. - 1999

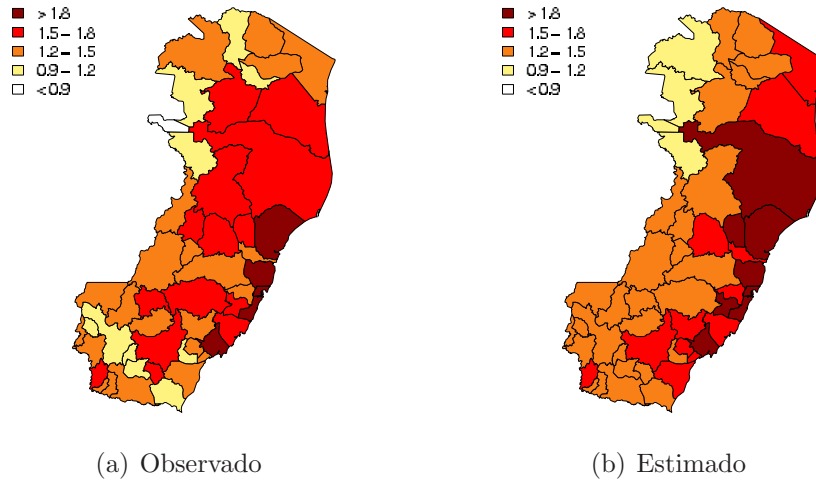


Figura 2.10: Log do PIB *per capita* do E.S. - 2000

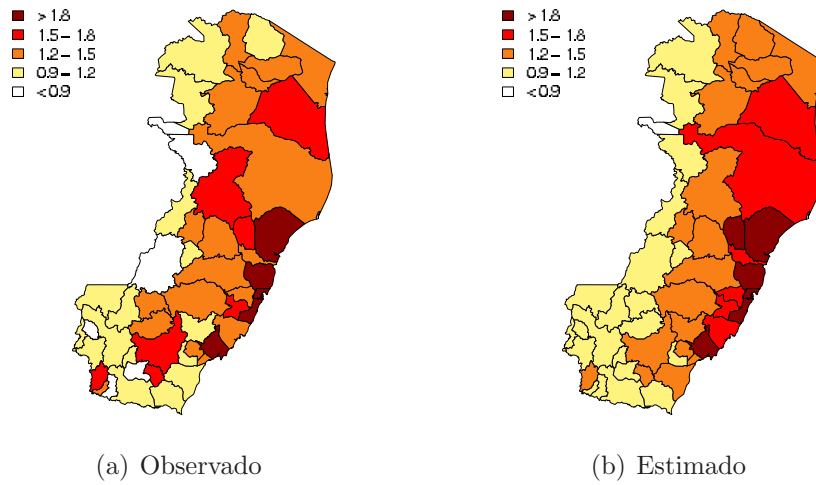


Figura 2.11: Log do PIB *per capita* do E.S. - 2001

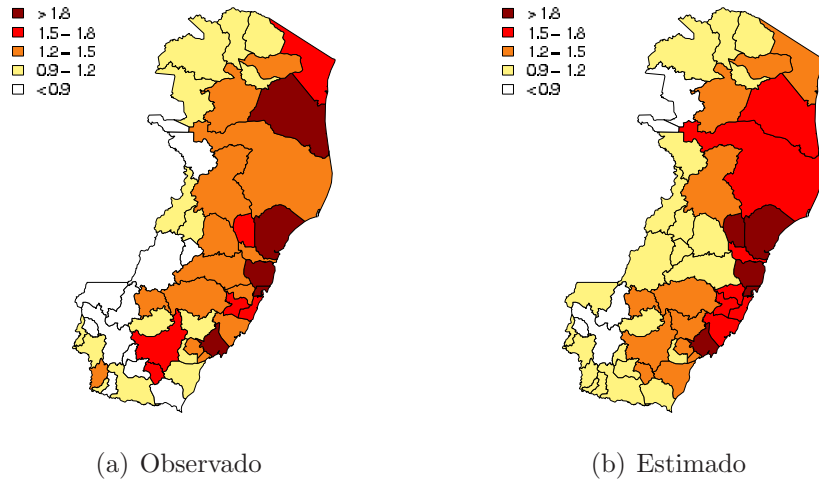


Figura 2.12: Log do PIB *per capita* do E.S. - 2002

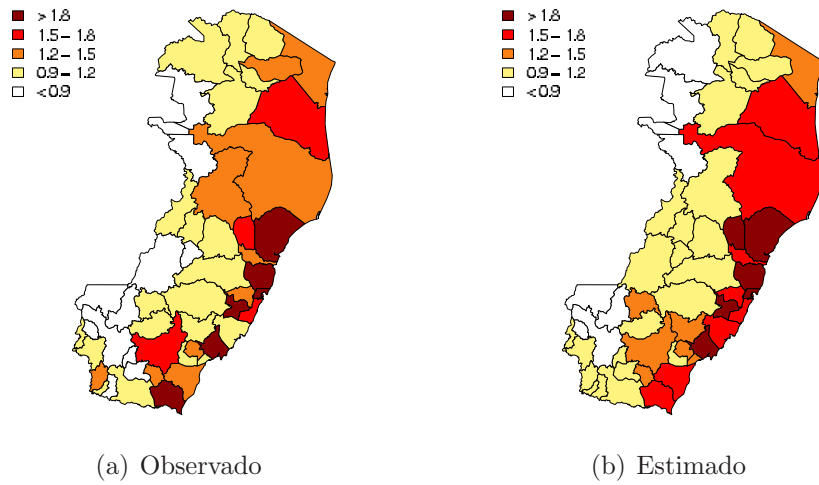


Figura 2.13: Log do PIB *per capita* do E.S. - 2003

Capítulo 3

Modelos Lineares Dinâmicos

3.1 Introdução

É bastante intuitivo o uso de uma estrutura de modelagem sequencial no estudo de séries temporais, na qual pretende-se analisar o comportamento futuro da série baseado em um conjunto de informações existentes. Um modelo de séries temporais é uma estrutura que descreve a evolução temporal de uma série, permitindo a realização de previsões. Tal modelo é probabilístico devido à aleatoriedade presente na variável de interesse. Destaca-se que um dos principais objetivos de uma análise de séries temporais é o entendimento do mecanismo de como estas são geradas, bem como, a predição para tempos futuros. O conhecimento sobre o mecanismo de geração das séries possibilita uma melhor descrição das mesmas, enquanto que a previsão contribui para a tomada de decisões.

Uma possibilidade de modelagem para uma série temporal, seguindo o enfoque Bayesiano, são os chamamos Modelos Lineares Dinâmicos (MLD's), que foram introduzidos por Harrison e Stevens (1976). Os MLD's também são conhecidos como modelos de espaço de estados e objetivam analisar uma variável latente com base em uma variável observável. Mais informações sobre análise de séries temporais com

enfoque bayesiano podem ser encontradas em West e Harrison (1997) e Migon *et al* (2005).

3.2 Modelo Linear Dinâmico

Um dos modelos lineares dinâmicos mais simples é o de primeira ordem. O princípio básico que norteia tal modelo é a flutuação aleatória dos valores da série temporal em torno de um ponto médio, sendo este também sujeito a variações ao longo do tempo. Basicamente, atribui-se às observações um modelo normal $y_t|\mu_t \sim N(\mu_t, V_t)$, onde V_t é conhecido. A modelagem de μ_t segue um passeio aleatório tal como $\mu_t = \mu_{t-1} + w_t$, com $w_t \sim N(0, W_t)$. Esse modelo, apesar de ser relativamente simples, incorpora o conceito de evolução temporal através de w_t , de forma que a média possa variar ao longo do tempo. Um dos fatores relevantes nesse modelo é a possibilidade de previsões a s -passos baseados na expressão $E(y_{t+s}|\mu_t) = E(\mu_{t+s} + v_{t+s}|\mu_t) = \mu_t$, sendo μ_t dependente de observações passadas, para todo $s > 0$.

Considere então um vetor de séries temporais denotado por $\{\mathbf{y}_t, t \in T\}$. A variável de interesse é \mathbf{y}_t , podendo ser discreta ou contínua, e T é o conjunto de índices, em geral representando o tempo. Nesse contexto, suponha que $t = 0$ seja o tempo inicial do processo e que o respectivo conjunto de informações iniciais seja definido como D_0 , que por sua vez representa todo o histórico de informações iniciais relevantes a respeito do processo temporal em questão. Analogamente, para um instante de tempo t , D_t representa o conjunto de informações disponíveis no tempo t . Quando \mathbf{y}_t é a única informação adicional no tempo t , tem-se então $D_t = (D_0, \mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_t)$.

Um modelo linear dinâmico, para cada tempo t , pode ser caracterizado formal-

mente pelas seguintes equações:

Equação de Observação:

$$\mathbf{y}_t = \mathbf{F}_t' \boldsymbol{\mu}_t + \mathbf{v}_t, \quad \mathbf{v}_t \sim N(\mathbf{0}, \mathbf{V}_t). \quad (3.1)$$

Equação de Sistema:

$$\boldsymbol{\mu}_t = \mathbf{G}_t \boldsymbol{\mu}_{t-1} + \mathbf{w}_t, \quad \mathbf{w}_t \sim N(\mathbf{0}, \mathbf{W}_t), \quad (3.2)$$

onde a informação inicial é dada pela distribuição $(\boldsymbol{\mu}_0 | D_0) \sim N(\mathbf{m}_0, \mathbf{C}_0)$, com \mathbf{v}_t e \mathbf{w}_t mutuamente e internamente independentes, e também independentes de $(\boldsymbol{\mu}_0 | D_0)$. \mathbf{F}_t é a matriz que relaciona as observações com o processo latente e \mathbf{G}_t é a matriz que descreve a evolução do processo latente ao longo do tempo.

A idéia central desse modelo é que, à medida que se avança no tempo, há um aumento na incerteza de $\boldsymbol{\mu}_t$, sugerindo, portanto, que se introduza um termo de evolução temporal para tal parâmetro. Pode-se observar que a equação de sistema é responsável pela evolução, onde $\boldsymbol{\mu}_t$ é o vetor de estados no tempo t , cuja distribuição a priori pode ser expressa por $p(\boldsymbol{\mu}_t | D_{t-1})$. A atualização dá-se através da incorporação da observação de \mathbf{y}_t com uso do Teorema de Bayes, descrita por $p(\boldsymbol{\mu}_t | D_t)$. A previsão, por sua vez, é realizada a partir da distribuição preditiva de $(\mathbf{y}_t | D_{t-1})$. Para o caso acima, onde os erros são normalmente distribuídos, essas distribuições possuem formas analíticas conhecidas, resultando nos chamados Modelos Lineares Dinâmicos.

3.3 O Filtro de Kalman

O Filtro de Kalman refere-se a um conjunto de procedimentos recursivos para amostragem em modelos de espaço de estados. Sua utilização em MLD's é adequada quando \mathbf{F}_t , \mathbf{G}_t , \mathbf{V}_t e \mathbf{W}_t forem conhecidos. Foi originalmente proposto por Kalman

(1960) e, apesar de bastante difundido em revistas científicas da área de engenharia, é de grande interesse de parte dos pesquisadores da área de estatística face à similaridade com os modelos de regressão linear. Porém, diferentemente destes últimos, no Filtro de Kalman não se assume um valor constante para os coeficientes ao longo do tempo, permitindo que os mesmos possam variar por meio de uma equação de evolução. Ressalta-se que para as aplicações nesse texto, as matrizes \mathbf{F}_t e \mathbf{G}_t são iguais à matriz identidade, com suas respectivas dimensões.

Embasado em resultados de estatística multivariada (ver anexo A) e Inferência Bayesiana, o Filtro de Kalman transformou-se em uma importante ferramenta para amostragem em modelos de espaço de estados, com aplicações em modelos espaço-temporais. Tal ferramenta fornece a distribuição condicional da variável de interesse dada toda a informação disponível até o tempo t de maneira relativamente simples. Basicamente, assume-se que uma medida observável \mathbf{y}_t , no tempo t , depende de uma quantidade não observável $\boldsymbol{\mu}_t$, conhecida como o *estado da natureza*. Mais detalhes podem ser encontrados em Meinhold e Singpurwalla (1983).

Seja então \mathbf{y}_t uma série temporal. O Filtro de Kalman no modelo linear dinâmico, com passos temporais de uma unidade, é construído conforme o seguinte esquema de inferência:

- Posteriori em $t - 1$:

$$\boldsymbol{\mu}_{t-1} | D_{t-1} \sim N(\mathbf{m}_{t-1}, \mathbf{C}_{t-1}).$$

- Priori em t :

$$\boldsymbol{\mu}_t | D_{t-1} \sim N(\mathbf{a}_t, \mathbf{R}_t),$$

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

- Previsão à 1 passo:

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

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

- Posteriori em t :

$$\boldsymbol{\mu}_t | D_t \sim N(\mathbf{m}_t, \mathbf{C}_t),$$

onde $\mathbf{m}_t = \mathbf{a}_t + \mathbf{A}_t \mathbf{e}_t$, $\mathbf{C}_t = \mathbf{R}_t - \mathbf{A}_t \mathbf{Q}_t \mathbf{A}'_t$, $\mathbf{A}_t = \mathbf{R}_t \mathbf{F}_t \mathbf{Q}_t^{-1}$ e $\mathbf{e}_t = \mathbf{y}_t - \mathbf{f}_t$.

- Suavização:

$$\boldsymbol{\mu}_{T-r} | D_T \sim N(\mathbf{a}_T(-r), \mathbf{R}_T(-r)),$$

onde:

$$\mathbf{a}_T(-r) = \mathbf{m}_{T-r} + \mathbf{B}_{T-r} [\mathbf{a}_T(-r+1) - \mathbf{a}_{T-r+1}],$$

$$\mathbf{R}_T(-r) = \mathbf{C}_{T-r} - \mathbf{B}_{T-r} [\mathbf{R}_{T-r+1} - \mathbf{R}_T(-r+1)] \mathbf{B}'_{T-r},$$

com $\mathbf{B}_{T-r} = \mathbf{C}_{T-r} \mathbf{G}'_{t+1} \mathbf{R}_{T-r+1}^{-1}$ e valores iniciais $\mathbf{a}_T(0) = \mathbf{m}_T$ e $\mathbf{R}_T(0) = \mathbf{C}_T$.

As duas próximas seções apresentam sucintamente alguns conceitos importantes de inferência e de simulação usados ao longo do texto.

3.4 Princípios de Inferência Bayesiana

A Inferência Estatística é o suporte para a formulação de modelos probabilísticos que possibilitam compreender e resumir as informações disponíveis, criando cenários que representem satisfatoriamente bem a realidade. A construção daqueles, porém, é uma tarefa árdua, pois, ao mesmo tempo em que se deseja incluir o máximo de informação nos modelos, não é uma boa ação torná-los demasiadamente complexos.

Em geral, um boa escolha seria criar um modelo que comporte as principais características do fenômeno, verificando posteriormente sua validade.

Considere então uma quantidade desconhecida θ , onde $\theta \in \Theta$, e X uma variável aleatória observável e dependente de θ . No caso da Inferência Clássica, a quantidade desconhecida θ é fixa, sendo sua estimativa baseada nos valores amostrais. Por outro lado, seguindo os preceitos da Inferência Bayesiana, θ não é mais tratado como um valor fixo mas, como uma variável aleatória, sendo sua distribuição de probabilidade $p(\theta)$ chamada de *distribuição a priori*. Dentro deste enfoque, a especificação da distribuição a priori é uma parte importante no processo de estimação.

A Inferência Bayesiana contém dois elementos essenciais: a distribuição amostral $p(X|\theta)$ e a priori $p(\theta)$, esta comumente dependendo de outros parâmetros (hiperparâmetros) e passível de subjetividade em sua especificação. Uma das maneiras de se encontrar uma distribuição a priori é através do método de famílias conjugadas. Para o caso das famílias exponenciais, essa tarefa torna-se relativamente simplificada.

Após ter observado a amostra, X passa a ser conhecida e fixa, e a atualização de θ se dará através da chamada *distribuição a posteriori*, definida por:

$$p(\theta|x) = \frac{p(x|\theta)p(\theta)}{p(x)},$$

onde

$$p(x) = \int p(x|\theta)p(\theta)d\theta.$$

Para o caso desse trabalho, como a distribuição dos dados supostamente segue um modelo normal, foram utilizadas as distribuições a priori conjugadas normal e gama inversa (GI). Para mais detalhes ver Migon e Gamerman (1999).

3.5 Simulação de Monte Carlo Via Cadeias de Markov

As simulações de Monte Carlo via cadeias de Markov (MCMC) são realizadas a partir das densidades condicionais completas a posteriori das variáveis envolvidas (ver mais detalhes em Gamerman e Lopes (2006)). Podem ser citadas como técnicas bastante úteis de MCMC, o amostrador de Gibbs (Geman e Geman, 1984) e o algoritmo de Metropolis-Hastings (Metropolis *et al.*, 1953). Tais ferramentas exigem que se forneçam os valores iniciais, gerando-se as simulações sequencialmente a partir das densidades completas a posteriori. Cabe citar que para utilizar o amostrador de Gibbs, necessita-se conhecer a forma analítica das densidades completas a posteriori. Caso isso não sendo possível, pode-se usar o algoritmo de Metropolis-Hastings para contornar tal obstáculo. No caso do amostrador de Gibbs, as distribuições condicionais completas compõem o núcleo de transição.

Nesse texto, todas as distribuições condicionais completas são conhecidas. Sendo assim, segue uma descrição resumida do amostrador de Gibbs utilizado:

Suponha um conjunto de variáveis aleatórias W_1, \dots, W_u e suas respectivas distribuições condicionais completas e conhecidas, dadas por $p(w_i | w_j, j \neq i)$, $i, j = 1, \dots, u$. O algoritmo de Gibbs utiliza um sistema de amostragem descrito como:

- Arbitre os valores iniciais $w_{1(0)}, \dots, w_{u(0)}$ da cadeia.
- Dados os valores iniciais, amostre sucessivamente valores para $w_{1(l)}, \dots, w_{u(l)}$

a partir das distribuições condicionais respectivas, dadas por:

$$\begin{aligned}
 w_{1(l)} &\sim p(w_1 \mid w_{2(l-1)}, \dots, w_{u(l-1)}) \\
 w_{2(l)} &\sim p(w_2 \mid w_{1(l)}, w_{3(l-1)}, \dots, w_{u(l-1)}) \\
 w_{3(l)} &\sim p(w_3 \mid w_{1(l)}, w_{2(l)}, w_{4(l-1)}, \dots, w_{u(l-1)}) \\
 &\vdots \\
 w_{u(l)} &\sim p(w_u \mid w_{1(l)}, w_{2(l)}, \dots, w_{u-1(l)})
 \end{aligned}$$

- À medida que $l \rightarrow \infty$ e, dadas algumas condições de regularidade, pode-se mostrar que $w_{1(l)}, \dots, w_{u(l)}$ convergem em distribuição para W_1, \dots, W_u . Logo, várias características de W_1, \dots, W_u podem ser estimadas.

Na subseção que segue descrevemos, em linhas gerais, as técnicas utilizadas nesse texto para análise das cadeias quanto à convergência.

3.5.1 Diagnósticos de Convergência

Para analisar as cadeias quanto à convergência, utilizamos o programa CODA (Plummer, Best, Cowles e Vines, 2007), disponível em <http://www-fis.iarc.fr/coda>. O mesmo disponibiliza ferramentas de análise descritiva e testes de convergência (ver Gelman e Rubin (1992) e Geweke (1992)) das séries geradas via MCMC. Em termos gerais, pode-se assim descrever tais critérios:

- **Gelman e Rubin** - suponha m cadeias inicializadas em pontos distintos. Considere as seguintes formas de estimar a variância da distribuição de equilíbrio: uma calculando a variância entre as cadeias, V_e , e a outra a partir da variância calculada dentro das cadeias, V_d . No caso de convergências das m cadeias, um estimador não viciado para a variância das m cadeias combinadas é dado por $V_c = \frac{n-1}{n}V_d + \frac{1}{n}V_e$, sendo n o número de iterações para cada cadeia. O teste utiliza a estatística

$$Gr = \sqrt{\frac{V_c}{V_d}},$$

indicando convergência quando Gr estiver próximo de 1.

- **Geweke** - Seja uma cadeia gerada com um número N suficientemente grande de iterações. A idéia é testar a igualdade das médias \bar{x}_1 e \bar{x}_2 , calculadas, respectivamente, a partir da fração $0.1N$ inicial e $0.5N$ final da amostra. Considerando os respectivos estimadores das variâncias assintóticas de \bar{x}_1 e \bar{x}_2 , dados por $V(\bar{x}_1)$ e $V(\bar{x}_2)$, tem-se que, quando $N \rightarrow \infty$

$$Gk = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{V(\bar{x}_1)/0.1N + V(\bar{x}_2)/0.5N}} \rightarrow N(0, 1).$$

Assim, valores extremos de Gk indicam falta de convergência.

3.6 Estimação com Variâncias Desconhecidas

Em um MLD a estimação dos parâmetros através do Filtro de Kalman fica impossibilitada quando as matrizes \mathbf{F}_t , \mathbf{G}_t , \mathbf{V}_t e \mathbf{W}_t forem desconhecidas, sendo necessário, nesse caso, o uso de aproximações numéricas na obtenção dos estimadores desejados. Destaca-se para esse fim os métodos de Monte Carlo via cadeias de Markov (MCMC).

3.6.1 Distribuições a Priori

Considere inicialmente as variâncias constantes no tempo, isto é, $\mathbf{V}_t = \mathbf{V}$ e $\mathbf{W}_t = \mathbf{W}$. As suas respectivas distribuições a priori podem ser dadas por uma distribuição Wishart Inversa, ou seja, $\mathbf{V} \sim WI(\mathbf{\Lambda}, \lambda)$ e $\mathbf{W} \sim WI(\mathbf{\Delta}, \delta)$. No caso de \mathbf{V} e \mathbf{W} serem univariadas, tem-se respectivamente, as seguintes distribuições a priori Gamas Inversas:

$$V \sim GI\left(\frac{\alpha_0}{2}, \frac{\beta_0}{2}\right)$$

$$W \sim GI\left(\frac{\alpha_1}{2}, \frac{\beta_1}{2}\right).$$

A escolha dos hiperparâmetros deve refletir o conhecimento do pesquisador a respeito dos mesmos.

3.6.2 Distribuições a Posteriori

Considere $\boldsymbol{\mu}_{1:T} = (\boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_T)$ o vetor de estados da variável latente. A distribuição conjunta a posteriori de $\boldsymbol{\mu}_t$, \mathbf{V} e \mathbf{W} pode ser expressa como:

$$f(\boldsymbol{\mu}_{1:T}, \mathbf{V}, \mathbf{W} | D_T) \propto \prod_{t=1}^T f(\mathbf{y}_t | \boldsymbol{\mu}_t, \mathbf{V}) f(\boldsymbol{\mu}_t | \boldsymbol{\mu}_{t-1}, \mathbf{W}) f(\mathbf{V}) f(\mathbf{W}) \quad (3.3)$$

As respectivas distribuições condicionais completas serão obtidas a partir de (3.3).

3.6.3 Simulação de $\boldsymbol{\mu}_{1:T}$ Via FFBS

Para estimar o vetor $\boldsymbol{\mu}_{1:T}$, se faz necessário gerar amostras a partir de sua distribuição condicional completa $f(\boldsymbol{\mu}_{1:T} | \mathbf{V}, \mathbf{W}, D_T)$. Pode-se utilizar o amostrador de Gibbs ou o algoritmo de Metropolis-Hastings, quando for o caso e simular, estado por estado, a partir de $f(\boldsymbol{\mu}_t | \boldsymbol{\mu}_{-t}, \mathbf{V}, \mathbf{W}, D_T)$, sendo $\boldsymbol{\mu}_{-t} = (\boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_{t-1}, \boldsymbol{\mu}_{t+1}, \dots, \boldsymbol{\mu}_T)$. Tal caminho, porém, não é o recomendável, visto que há uma forte correlação a posteriori entre os valores gerados de $(\boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_T)$. Por ser mais eficiente computacionalmente, iremos utilizar o algoritmo FFBS - *Forward Filtering Backward Sampling* - (ver Carter e Kohn (1994) e Fruwirth-Schnatter (1994)). A seguir faz-se uma breve descrição do referido algoritmo.

Algoritmo FFBS

Seja um MLD de primeira ordem tal como definido em (3.1) e (3.2), sendo \mathbf{V} e \mathbf{W} desconhecidos. Deseja-se amostrar o vetor completo de estado $\boldsymbol{\mu}_{1:T} = (\boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_T)$ a partir da distribuição condicional completa $f(\boldsymbol{\mu}_{1:T} | \mathbf{V}, \mathbf{W}, D_T)$. Apoiando-se nas

propriedades markovianas, pode-se escrever essa distribuição como:

$$\begin{aligned} f(\boldsymbol{\mu}_{1:T}|\mathbf{V}, \mathbf{W}, D_T) &= \prod_{t=1}^T f(\boldsymbol{\mu}_{t-1}|\boldsymbol{\mu}_t, \mathbf{V}, \mathbf{W}, D_t) f(\boldsymbol{\mu}_T|\mathbf{V}, \mathbf{W}, D_T) \\ &= f(\boldsymbol{\mu}_1|\boldsymbol{\mu}_2, \mathbf{V}, \mathbf{W}, D_1) f(\boldsymbol{\mu}_2|\boldsymbol{\mu}_3, \mathbf{V}, \mathbf{W}, D_2) \\ &\quad \times \cdots \times f(\boldsymbol{\mu}_{T-1}|\boldsymbol{\mu}_T, \mathbf{V}, \mathbf{W}, D_{T-1}) f(\boldsymbol{\mu}_T|\mathbf{V}, \mathbf{W}, D_T) \end{aligned}$$

A amostragem sequencial de $\boldsymbol{\mu}_{1:T}$ pode ser realizada tal como:

1. Inicialmente, utilizar o Filtro de Kalman descrito na seção 3.3 para obter a média e a variância das distribuições $f(\boldsymbol{\mu}_1|\mathbf{V}, \mathbf{W}, D_1), \dots, f(\boldsymbol{\mu}_T|\mathbf{V}, \mathbf{W}, D_T)$. Observar que, pelo fato de $f(\boldsymbol{\mu}_t|\mathbf{V}, \mathbf{W}, D_t)$ ser normalmente distribuído, basta conhecer a média e a variância para que sejam completamente determinadas as respectivas distribuições.
2. Em seguida, gerar um valor para $\boldsymbol{\mu}_T$ a partir de $(\boldsymbol{\mu}_T|\mathbf{V}, \mathbf{W}, D_T) \sim N(\mathbf{m}_T, \mathbf{C}_T)$, onde \mathbf{m}_T e \mathbf{C}_T estão definidos conforme a seção 3.3.
3. Após obter a média e a variância de $f(\boldsymbol{\mu}_{T-1}|\boldsymbol{\mu}_T, \mathbf{V}, \mathbf{W}, D_T)$, gerar $\boldsymbol{\mu}_{T-1}$ a partir dessa distribuição.
4. Obter, recursivamente

$$f(\boldsymbol{\mu}_{T-r}|\boldsymbol{\mu}_{T-r+1}, \dots, \boldsymbol{\mu}_T, D_T) = f(\boldsymbol{\mu}_{T-r}|\boldsymbol{\mu}_{T-r+1}, D_T) \sim N(\mathbf{h}_T(r), \mathbf{H}_T(r)),$$

sendo $r = 1, \dots, T$,

$$\mathbf{h}_T(r) = \mathbf{m}_{T-r} + \mathbf{B}_{T-r}(\boldsymbol{\mu}_{T-r+1} - \mathbf{a}_{T-r+1}),$$

$$\mathbf{H}_T(r) = \mathbf{C}_{T-r} - \mathbf{B}_{T-r}\mathbf{R}_{T-r+1}\mathbf{B}'_{T-r},$$

com $\mathbf{B}_{T-r} = \mathbf{C}_{T-r}\mathbf{G}'_t\mathbf{R}_{T-r+1}^{-1}$. Em seguida, gerar $\boldsymbol{\mu}_{T-r}$ dessa distribuição. Vale ressaltar nesse ponto que $f(\boldsymbol{\mu}_{T-r}|\boldsymbol{\mu}_{T-r+1}, D_T) = f(\boldsymbol{\mu}_{T-r}|\boldsymbol{\mu}_{T-r+1}, D_{T-r})$.

Com o algoritmo anterior, cada valor gerado de $\boldsymbol{\mu}_t$ estará condicionado a $(\mathbf{V}, \mathbf{W}, D_T)$ e os estados serão gerados todos de uma vez, não tendo a desvantagem da ordenação passo a passo no tempo.

3.6.4 Simulação de V e W

A simulação dos valores de \mathbf{V} e \mathbf{W} será realizada a partir das respectivas distribuições condicionais completas, descritas como:

Condicional completa de \mathbf{V} (caso de \mathbf{V} ser uma matriz $l \times l$)

$$\begin{aligned}
 f(\mathbf{V}|\boldsymbol{\mu}_{1:T}, D_T) &\propto \prod_{t=1}^T f(\mathbf{y}_t|\boldsymbol{\mu}_t, \mathbf{V}, D_{t-1})f(\mathbf{V}|\boldsymbol{\Lambda}, \lambda) \\
 &\propto \left[\prod_{t=1}^T f(\mathbf{y}_t|\boldsymbol{\mu}_t, \mathbf{V}, D_{t-1}) \right] \left[|\mathbf{V}|^{-\frac{\lambda+l+1}{2}} \exp\left(-\frac{1}{2}\text{tr}(\boldsymbol{\Lambda}\mathbf{V}^{-1})\right) \right] \\
 f(\mathbf{V}|\boldsymbol{\mu}_{1:T}, D_T) &\propto |\mathbf{V}|^{-\frac{1}{2}(\lambda+l+1)}|\mathbf{V}|^{-\frac{T}{2}} \\
 &\times \exp\left[-\frac{1}{2}\left(\text{tr}(\boldsymbol{\Lambda}\mathbf{V}^{-1}) + \sum_{t=1}^T \text{tr}((\mathbf{y}_t - \boldsymbol{\mu}_t)'\mathbf{V}^{-1}(\mathbf{y}_t - \boldsymbol{\mu}_t))\right)\right] \\
 f(\mathbf{V}|\boldsymbol{\mu}_{1:T}, D_T) &\propto |\mathbf{V}|^{-\frac{1}{2}(T+\lambda+l+1)} \\
 &\times \exp\left[-\frac{1}{2}\left[\text{tr}(\boldsymbol{\Lambda}\mathbf{V}^{-1}) + \text{tr}\left(\sum_{t=1}^T (\mathbf{y}_t - \boldsymbol{\mu}_t)(\mathbf{y}_t - \boldsymbol{\mu}_t)'\mathbf{V}^{-1}\right)\right]\right] \\
 f(\mathbf{V}|\boldsymbol{\mu}_{1:T}, D_T) &\propto |\mathbf{V}|^{-\frac{1}{2}(T+\lambda+l+1)} \\
 &\times \exp\left[-\frac{1}{2}\text{tr}\left[\left(\boldsymbol{\Lambda} + \sum_{t=1}^T (\mathbf{y}_t - \boldsymbol{\mu}_t)(\mathbf{y}_t - \boldsymbol{\mu}_t)'\right)\mathbf{V}^{-1}\right]\right] \\
 (\mathbf{V}|\boldsymbol{\mu}_{1:T}, D_T) &\sim WI(\boldsymbol{\Lambda}^*, \lambda^*),
 \end{aligned}$$

onde $\lambda^* = T + \lambda$ e $\boldsymbol{\Lambda}^* = \boldsymbol{\Lambda} + \sum_{t=1}^T (\mathbf{y}_t - \boldsymbol{\mu}_t)(\mathbf{y}_t - \boldsymbol{\mu}_t)'$. Vale lembrar que para V univariado tem-se:

$$(V|\mu_{1:T}, D_T) \sim GI\left(\frac{\alpha_0^*}{2}, \frac{\beta_0^*}{2}\right),$$

onde $\alpha_0^* = T + \alpha_0$ e $\beta_0^* = \beta_0 + \sum_{t=1}^T (y_t - \mu_t)^2$.

Condicional completa de \mathbf{W} (caso de \mathbf{W} ser uma matriz $l \times l$)

$$\begin{aligned}
f(\mathbf{W}|\boldsymbol{\mu}_{1:T}, D_T) &\propto \prod_{t=1}^T f(\boldsymbol{\mu}_t|\boldsymbol{\mu}_{t-1}, \mathbf{W})f(\mathbf{W}|\boldsymbol{\Delta}, \delta) \\
&\propto \left[\prod_{t=1}^T f(\boldsymbol{\mu}_t|\boldsymbol{\mu}_{t-1}, \mathbf{W}) \right] \left[|\mathbf{W}|^{-\frac{\delta+l+1}{2}} \exp\left(-\frac{1}{2}\text{tr}(\boldsymbol{\Delta}\mathbf{W}^{-1})\right) \right] \\
f(\mathbf{W}|\boldsymbol{\mu}_{1:T}, D_T) &\propto |\mathbf{W}|^{-\frac{1}{2}(\delta+l+1)} |\mathbf{W}|^{-\frac{T}{2}} \\
&\times \exp\left[-\frac{1}{2}\left(\text{tr}(\boldsymbol{\Delta}\mathbf{W}^{-1}) + \sum_{t=1}^T \text{tr}((\boldsymbol{\mu}_t - \boldsymbol{\mu}_{t-1})'\mathbf{W}^{-1}(\boldsymbol{\mu}_t - \boldsymbol{\mu}_{t-1}))\right)\right] \\
f(\mathbf{W}|\boldsymbol{\mu}_{1:T}, D_T) &\propto |\mathbf{W}|^{-\frac{1}{2}(T+\delta+l+1)} \\
&\times \exp\left[-\frac{1}{2}\left[\text{tr}(\boldsymbol{\Delta}\mathbf{W}^{-1}) + \text{tr}\left(\sum_{t=1}^T (\boldsymbol{\mu}_t - \boldsymbol{\mu}_{t-1})(\boldsymbol{\mu}_t - \boldsymbol{\mu}_{t-1})'\mathbf{W}^{-1}\right)\right]\right] \\
f(\mathbf{W}|\boldsymbol{\mu}_{1:T}, D_T) &\propto |\mathbf{W}|^{-\frac{1}{2}(T+\delta+l+1)} \\
&\times \exp\left[-\frac{1}{2}\text{tr}\left[\left(\boldsymbol{\Delta} + \sum_{t=1}^T (\boldsymbol{\mu}_t - \boldsymbol{\mu}_{t-1})(\boldsymbol{\mu}_t - \boldsymbol{\mu}_{t-1})'\right)\mathbf{W}^{-1}\right]\right] \\
(\mathbf{W}|\boldsymbol{\mu}_{1:T}, D_T) &\sim WI(\boldsymbol{\Delta}^*, \delta^*),
\end{aligned}$$

onde $\delta^* = T + \delta$ e $\boldsymbol{\Delta}^* = \boldsymbol{\Delta} + \sum_{t=1}^T (\boldsymbol{\mu}_t - \boldsymbol{\mu}_{t-1})(\boldsymbol{\mu}_t - \boldsymbol{\mu}_{t-1})'$. Para o caso de W univariado tem-se:

$$(W|\mu_{1:T}, D_T) \sim GI\left(\frac{\alpha_1^*}{2}, \frac{\beta_1^*}{2}\right),$$

onde $\alpha_1^* = T + \alpha_1$ e $\beta_1^* = \beta_1 + \sum_{t=1}^T (\mu_t - \mu_{t-1})^2$.

Capítulo 4

Modelagem Espaço-Temporal Multi-Escala

4.1 Introdução e Notação

Neste capítulo, descrevemos a proposta de modelo desse trabalho. Especificamente, pretendemos apresentar um modelo de espaço de estados com estrutura multi-escala, que vise ao estudo de fenômenos observados em um conjunto de áreas e ao longo do tempo.

Uma referência inicial de modelagem espaço-temporal multi-escala foi a realizada por Berliner, Wikle e Milliff (1999). Este usa uma decomposição em ondaletas para o processo latente, onde os coeficientes evoluem com o tempo. É feita uma aplicação a dados relativos ao vento. O presente trabalho pode ser visto como uma extensão do trabalho supra citado, porém, permitindo aplicações em grades espaciais irregulares e não apresentando a limitação de apenas dois descendentes para cada ancestral da partição. Uma outra referência, esta mais recente, é o trabalho de Johannesson, Cressie e Huang (2007), que também se difere da proposta desse texto pois, usa uma forma empírica para estimar as covariâncias envolvidas, além disso, inclui a

componente dinâmica temporal apenas no nível mais agregado do espaço. Outro trabalho recente que pode ser citado é Ferreira, West, Lee, and Higdon (2006), que propõe uma classe de modelos multi-escala na coordenada temporal. Uma boa revisão sobre o assunto pode ser encontrada em Ferreira and Lee (2007).

Inicialmente, seja um processo espaço-temporal multi-escala no nível mais desagregado representado pelo vetor aleatório $\mathbf{y}_{J-1,t} = (y_{J-1,1,t}, \dots, y_{J-1,n_{J-1},t})'$, com $t = 1, \dots, T$, tal que

$$\mathbf{y}_{J-1,t} | \boldsymbol{\mu}_{J-1,t}, \mathbf{V}_{J-1,t} \sim N(\boldsymbol{\mu}_{J-1,t}, \mathbf{V}_{J-1,t}).$$

Observe que $J-1$ é o nível mais desagregado do processo. O vetor de médias $\boldsymbol{\mu}_{J-1,t}$ e a matriz de covariância $\mathbf{V}_{J-1,t}$ são desconhecidos e expressos, respectivamente, por:

$$\boldsymbol{\mu}_{J-1,t} = (\mu_{J-1,1,t}, \dots, \mu_{J-1,n_{J-1},t})'$$

$$\mathbf{V}_{J-1,t} = \text{diag}(\sigma_{J-1,1,t}^2, \dots, \sigma_{J-1,n_{J-1},t}^2).$$

Recorde que:

- $D_{j,k}^*$ é a coleção de índices espaciais relativos aos descendentes do ancestral $B_{j,k}$, com exclusão de um dos elementos, sendo $j = 0, \dots, J-2$.
- $n_{D_{j,k}^*}$ é o número de descendentes de $B_{j,k}$ menos um.

Algumas considerações importantes com relação ao processo espaço-temporal multi-escala em questão são:

1. J é a quantidade de níveis de escala do processo. Para o caso do estado do Espírito Santo, J é igual a 3 e refere-se aos níveis de município, micro-região e macro-região (para mais detalhes ver anexo C).

2. O processo no nível mais agregado, descrito pelo vetor $\mathbf{y}_{0,t} = (y_{0,1,t}, \dots, y_{0,n_0,t})'$, é expresso por $\mathbf{y}_{0,t} | \boldsymbol{\mu}_{0,t}, \mathbf{V}_{0,t} \sim N(\boldsymbol{\mu}_{0,t}, \mathbf{V}_{0,t})$, onde $\boldsymbol{\mu}_{0,t} = (\mu_{0,1,t}, \dots, \mu_{0,n_0,t})'$ e $\mathbf{V}_{0,t} = \text{diag}(\sigma_{0,1,t}^2, \dots, \sigma_{0,n_0,t}^2)$.
3. O vetor aleatório dos descendentes do elemento $B_{J-2,k}$, excluindo-se um dos descendentes, é expresso por $\mathbf{y}_{D_{J-2,k}^*,t}$.
4. $\mathbf{y}_{D_{j,k}^*,t} | y_{j,k,t}, \boldsymbol{\theta}_{j,k,t}, \boldsymbol{\Omega}_{j,k,t} \sim N(\boldsymbol{\nu}_{j,k,t} y_{j,k,t} + \boldsymbol{\theta}_{j,k,t}, \boldsymbol{\Omega}_{j,k,t})$, onde

$$\boldsymbol{\nu}_{j,k,t} = \frac{\boldsymbol{\sigma}_{D_{j,k}^*,t}^2}{\sigma_{j,k,t}^2}, \quad (4.1)$$

$$\boldsymbol{\theta}_{j,k,t} = \boldsymbol{\mu}_{D_{j,k}^*,t} - \boldsymbol{\nu}_{j,k,t} \mu_{j,k,t} \quad (4.2)$$

$$\boldsymbol{\Omega}_{j,k,t} = \mathbf{V}_{D_{j,k}^*,t} - \left[\frac{\boldsymbol{\sigma}_{D_{j,k}^*,t}^2 \left(\boldsymbol{\sigma}_{D_{j,k}^*,t}^2 \right)'}{\sigma_{j,k,t}^2} \right], \quad (4.3)$$

sendo $\mathbf{V}_{D_{j,k}^*,t}$ uma matriz diagonal formada pelo elementos do vetor $\boldsymbol{\sigma}_{D_{j,k}^*,t}^2$, com

$$\sigma_{j,k,t}^2 = \sum_{(j+1,k') \in D_{j,k}^*} \sigma_{j+1,k',t}^2. \quad (4.4)$$

Observe que $\boldsymbol{\nu}_{j,k,t}$ é um vetor que relaciona a variância do ancestral com as dos seus descendentes, e suas coordenadas são tanto mais próximas de $\frac{1}{n_{D_{j,k}^*} + 1}$ quanto mais homogêneas forem as variâncias dos respectivos descendentes. Por sua vez, $\boldsymbol{\theta}_{j,k,t}$ é o vetor de diferenças relativas entre a média do ancestral e a dos seus descendentes. Os procedimentos para se obter as expressões (4.1), (4.2) e (4.3) podem ser encontrados no Anexo B.

4.2 Modelo Geral

Nesta seção introduzimos um conjunto de equações que, de um modo geral, definem o Modelo Espaço-Temporal Multi-Escala. Observe que a expressão (4.2) possui grande importância, pois, a partir do conhecimento da mesma e do vetor $\boldsymbol{\mu}_{0,t}$, pode-se conhecer, de forma recursiva, a média $\boldsymbol{\mu}_{J-1,t}$ do processo. Tome então

$$\boldsymbol{\theta}_{j,k,t}^e = (\mathbf{y}_{D_{j,k,t}^*} - \boldsymbol{\nu}_{j,k,t} y_{j,k,t})$$

como um estimador empírico para $\boldsymbol{\theta}_{j,k,t}$. Desse modo, escrevemos um Modelo Espaço-temporal Multi-Escala através das seguintes equações:

Equação de Observação:

$$\mathbf{y}_{J-1,t} = \boldsymbol{\mu}_{J-1,t} + \mathbf{v}_{J-1,t}, \quad \mathbf{v}_{J-1,t} | \mathbf{V}_{J-1,t} \sim N(\mathbf{0}, \mathbf{V}_{J-1,t}). \quad (4.5)$$

$$\text{Note que } \mathbf{V}_{J-1,t} = \text{diag}(\sigma_{J-1,1,t}^2, \dots, \sigma_{J-1,n_{J-1,t}}^2).$$

Como se pretende utilizar a estrutura multi-escala dos dados, se faz necessário decompor a expressão (4.5) em equações observacionais que representem os respectivos níveis de escala. Logo, as equações observacionais e de sistema do nível mais agregado até o mais desagregado são expressas respectivamente por:

Decomposição Multi-escala da Equação de Observação:

$$y_{0,k,t} = \mu_{0,k,t} + v_{0,k,t}, \quad v_{0,k,t} | \sigma_{0,k,t}^2 \sim N(0, \sigma_{0,k,t}^2). \quad (4.6)$$

$$\boldsymbol{\theta}_{j,k,t}^e = \boldsymbol{\theta}_{j,k,t} + \mathbf{V}_{j,k,t}, \quad \mathbf{V}_{j,k,t} | \boldsymbol{\Omega}_{j,k,t} \sim N(\mathbf{0}, \boldsymbol{\Omega}_{j,k,t}). \quad (4.7)$$

Equações de Sistema:

$$\mu_{0,k,t} = \mu_{0,k,t-1} + w_{0,k,t}, \quad w_{0,k,t} | W_{0,k,t} \sim N(0, W_{0,k,t}). \quad (4.8)$$

$$\boldsymbol{\theta}_{j,k,t} = \boldsymbol{\theta}_{j,k,t-1} + \mathbf{w}_{j,k,t}, \quad \mathbf{w}_{j,k,t} | \mathbf{W}_{j,k,t} \sim N(\mathbf{0}, \mathbf{W}_{j,k,t}). \quad (4.9)$$

4.3 Um Exemplo Didático

Suponha as variâncias observacionais $\sigma_{J-1,k,t}^2$ constantes ao longo do tempo, e iguais para todos os n_{J-1} elementos no nível $J-1$. Temos então que

$$\mathbf{V}_{J-1,t} = \sigma^2 \mathbf{I}_{n_{J-1} \times n_{J-1}}.$$

Ou seja, $\sigma_{J-1,1}^2 = \sigma_{J-1,2}^2 = \dots = \sigma_{J-1,n_{J-1}}^2 = \sigma^2$. Isso resulta que, para o nível $J-2$,

$$\boldsymbol{\nu}_{J-2,k} = \left(\frac{\mathbf{1}_{n_{D_{J-2,k}^*}}}{n_{D_{J-2,k}^*} + 1} \right),$$

sendo $\mathbf{1}_{n_{D_{J-2,k}^*}}$ um vetor coluna de 1's e,

$$\boldsymbol{\Omega}_{J-2,k} = \sigma^2 \left[\mathbf{I}_{(n_{D_{J-2,k}^*}) \times (n_{D_{J-2,k}^*})} - \left(\frac{\mathbf{1}_{n_{D_{J-2,k}^*}} \mathbf{1}'_{n_{D_{J-2,k}^*}}}{n_{D_{J-2,k}^*} + 1} \right) \right].$$

Note que para qualquer nível $j < J-2$, as expressões (4.1), (4.2) e (4.3) permanecem válidas.

Para facilitar o entendimento, vejamos como ficariam os valores de $\boldsymbol{\nu}_{J-2,k}$, $\boldsymbol{\theta}_{J-2,k,t}$ e $\boldsymbol{\Omega}_{J-2,k}$ em uma situação particular. Tome como referência a estrutura de partição hierárquica do estado do Espírito Santo, relativa à divisão política do ano de 1979. A mesma possui 52 municípios no nível mais desagregado, 12 micro-regiões no nível intermediário e 4 macro-regiões no nível mais agregado (ver figura 1.1). Recorde que se deve desconsiderar, ainda que apenas inicialmente, um dos descendentes de cada ancestral no processo de estimação. Vejamos os dois exemplos seguintes:

- Exemplo 1: seja a micro-região $k = 2$ do nível $j = 1$. Esta micro-região possui 3 municípios descendentes. Logo, tem-se que:

$$\boldsymbol{\nu}_{1,2} = \left(\frac{1}{3}, \frac{1}{3} \right)'$$

e

$$\boldsymbol{\Omega}_{1,2} = \sigma^2 \left[\mathbf{I}_{2 \times 2} - \frac{1}{3} \begin{pmatrix} 1 \\ 1 \end{pmatrix} (1, 1) \right].$$

Daí que, se supormos $\hat{\mu}_{1,2} = 10$ e $\hat{\boldsymbol{\mu}}_{D_{1,2}^*} = (2, 4)'$, então teremos

$$\hat{\boldsymbol{\theta}}_{1,2} = \left(\frac{-4}{3}, \frac{2}{3} \right)'.$$

- Exemplo 2: considere a macro-região $k = 1$ do nível $j = 0$. Essa macro-região possui 5 micro-regiões descendentes, sendo que cada uma delas possui, respectivamente, 7, 3, 5, 3 e 4 municípios. Portanto, tem-se que:

$$\boldsymbol{\nu}_{0,1} = \left(\frac{7}{22}, \frac{3}{22}, \frac{5}{22}, \frac{3}{22} \right)'$$

e

$$\boldsymbol{\Omega}_{0,1} = \sigma^2 \left[\mathbf{I}_{4 \times 4} - \frac{1}{22} \begin{pmatrix} 7 \\ 3 \\ 5 \\ 3 \end{pmatrix} (7, 3, 5, 3) \right].$$

Se supormos $\hat{\mu}_{0,1} = 7$ e $\hat{\boldsymbol{\mu}}_{D_{0,1}^*} = (3, 1, 1, 1)'$, então teremos

$$\hat{\boldsymbol{\theta}}_{0,1} = \left(\frac{17}{22}, \frac{1}{22}, \frac{-13}{22}, \frac{1}{22} \right)'.$$

No próximo capítulo, considerando que as variâncias observacionais e de sistema guardam entre si uma relação de proporcionalidade, descrevemos mais detalhadamente, inclusive com aplicações, o modelo exposto no presente capítulo.

Capítulo 5

Modelagem Espaço-Temporal Multi-Escala Com Variâncias Observacionais Constantes

Neste capítulo, propomos um modelo espaço-temporal multi-escala considerando as variâncias observacionais constantes para cada elemento do nível mais desagregado. Ou seja, iremos propor um modelo para o caso em que $\mathbf{V}_{J-1} = \sigma^2 \mathbf{I}_{n_{J-1} \times n_{J-1}}$.

Considere o modelo geral apresentado no capítulo anterior. Além de constantes no tempo, suponha que as variâncias de sistema sejam proporcionais às respectivas variâncias observacionais. Ou seja, que sejam expressas por:

$$W_{0,k} = \psi_k \sigma_{0,k}^2$$

e

$$\mathbf{W}_{j,k} = \psi_{j,k} \boldsymbol{\Omega}_{j,k}.$$

Observe que os hiperparâmetros ψ_k e $\psi_{j,k}$ são uma fração das respectivas variâncias

observacionais. Logo, podemos reescrever o modelo a partir das expressões (4.5), (4.6), (4.7), (4.8) e (4.9) tal como:

Equação de Observação:

$$\mathbf{y}_{J-1,t} = \boldsymbol{\mu}_{J-1,t} + \mathbf{v}_{J-1}, \quad \mathbf{v}_{J-1} \mid \mathbf{V}_{J-1} \sim N(\mathbf{0}, \mathbf{V}_{J-1}), \quad (5.1)$$

$$\text{onde } \mathbf{V}_{J-1} = \sigma^2 \mathbf{I}_{n_{J-1} \times n_{J-1}}.$$

Decomposição Multi-escala da Equação de Observação:

$$y_{0,k,t} = \mu_{0,k,t} + v_{0,k}, \quad v_{0,k} \mid \sigma_{0,k}^2 \sim N(0, \sigma_{0,k}^2). \quad (5.2)$$

$$\boldsymbol{\theta}_{j,k,t}^e = \boldsymbol{\theta}_{j,k,t} + \mathbf{V}_{j,k}, \quad \mathbf{V}_{j,k} \mid \boldsymbol{\Omega}_{j,k} \sim N(\mathbf{0}, \boldsymbol{\Omega}_{j,k}). \quad (5.3)$$

Equações de Sistema:

$$\mu_{0,k,t} = \mu_{0,k,t-1} + w_{0,k}, \quad w_{0,k} \mid W_{0,k} \sim N(0, \psi_k \sigma_{0,k}^2). \quad (5.4)$$

$$\boldsymbol{\theta}_{j,k,t} = \boldsymbol{\theta}_{j,k,t-1} + \mathbf{w}_{j,k}, \quad \mathbf{w}_{j,k} \mid \mathbf{W}_{j,k} \sim N(\mathbf{0}, \psi_{j,k} \boldsymbol{\Omega}_{j,k}). \quad (5.5)$$

5.1 Densidade Conjunta do Modelo

Suponha que as medidas $\mathbf{y}_{J-1,t} = (y_{J-1,1,t}, \dots, y_{J-1,n_{J-1},t})'$ e o esquema de estruturação de partições hierárquicas estejam definidos. Considerando o modelo descrito pelas expressões (5.1) a (5.5), podemos escrever, para o nível mais desagregado $J-1$, a densidade conjunta $f(\mathbf{y}_{J-1}, \boldsymbol{\mu}_0, \boldsymbol{\theta}, \sigma^2, \psi_k, \psi_{j,k})$ do Modelo Espaço-Temporal Multi-Escala como igual a:

$$\begin{aligned} & \prod_{t=1}^T \left\{ \prod_{k=1}^{n_0} f(y_{0,k,t} \mid \mu_{0,k,t}, \sigma^2) \prod_{j=0}^{J-2} \prod_{k=1}^{n_j} f(\mathbf{y}_{D_{j,k,t}^*} \mid y_{j,k,t}, \boldsymbol{\theta}_{j,k,t}, \sigma^2) \right. \\ & \left. \prod_{k=1}^{n_0} f(\mu_{0,k,t} \mid \mu_{0,k,t-1}, \psi_k) \prod_{j=0}^{J-2} \prod_{k=1}^{n_j} f(\boldsymbol{\theta}_{j,k,t} \mid \boldsymbol{\theta}_{j,k,t-1}, \psi_{j,k}) \right\} \\ & \prod_{k=1}^{n_0} f(\mu_{0,k,0}) f(\psi_k) \prod_{j=0}^{J-2} \prod_{k=1}^{n_j} f(\boldsymbol{\theta}_{j,k,0}) f(\psi_{j,k}) f(\sigma^2). \end{aligned}$$

Condicionado em $(\sigma^2, \psi_k, \psi_{j,k})$, a densidade conjunta $f(\mathbf{y}_{J-1}, \boldsymbol{\mu}_0, \boldsymbol{\theta} | \sigma^2, \psi_k, \psi_{j,k})$ torna-se:

$$f(\mathbf{y}_{J-1}, \boldsymbol{\mu}_0, \boldsymbol{\theta} | \sigma^2, \psi_k, \psi_{j,k}) = \prod_{t=1}^T \left\{ \prod_{k=1}^{n_0} f(y_{0,k,t} | \mu_{0,k,t}) \prod_{j=0}^{J-2} \prod_{k=1}^{n_j} f(\mathbf{y}_{D_{j,k,t}^*} | y_{j,k,t}, \boldsymbol{\theta}_{j,k,t}) \right. \\ \left. \prod_{k=1}^{n_0} f(\mu_{0,k,t} | \mu_{0,k,t-1}) \prod_{j=0}^{J-2} \prod_{k=1}^{n_j} f(\boldsymbol{\theta}_{j,k,t} | \boldsymbol{\theta}_{j,k,t-1}) \right\} \\ \prod_{k=1}^{n_0} f(\mu_{0,k,0}) \prod_{j=0}^{J-2} \prod_{k=1}^{n_j} f(\boldsymbol{\theta}_{j,k,0}),$$

que, rearranjando os produtórios, chega-se a $f(\mathbf{y}_{J-1}, \boldsymbol{\mu}_0, \boldsymbol{\theta} | \sigma^2, \psi_k, \psi_{j,k})$ igual a:

$$\prod_{k=1}^{n_0} \left\{ f(\mu_{0,k,0}) \prod_{t=1}^T f(y_{0,k,t} | \mu_{0,k,t}) f(\mu_{0,k,t} | \mu_{0,k,t-1}) \right\} \prod_{j=0}^{J-2} \prod_{k=1}^{n_j} \\ \left\{ f(\boldsymbol{\theta}_{j,k,0}) \prod_{t=1}^T f(\mathbf{y}_{D_{j,k,t}^*} | y_{j,k,t}, \boldsymbol{\theta}_{j,k,t}) f(\boldsymbol{\theta}_{j,k,t} | \boldsymbol{\theta}_{j,k,t-1}) \right\}, \quad (5.6)$$

onde:

$$\begin{aligned} \boldsymbol{\mu}_0 &= (\boldsymbol{\mu}'_{0,,0}, \dots, \boldsymbol{\mu}'_{0,,T})' \\ \boldsymbol{\mu}_{0,,t} &= (\mu_{0,1,t}, \dots, \mu_{0,n_0,t})' \\ \boldsymbol{\mu}_{0,k,.} &= (\mu_{0,k,0}, \dots, \mu_{0,k,T})' \\ \boldsymbol{\theta} &= (\boldsymbol{\theta}'_1, \dots, \boldsymbol{\theta}'_{J-2})' \\ \boldsymbol{\theta}_j &= (\boldsymbol{\theta}'_{j,,0}, \dots, \boldsymbol{\theta}'_{j,,T})' \\ \boldsymbol{\theta}_{j,,t} &= (\boldsymbol{\theta}'_{j,1,t}, \dots, \boldsymbol{\theta}'_{j,n_j,t})'. \end{aligned}$$

Podemos expressar os termos contidos nos respectivos produtórios da equação (5.6) como:

$$f(y_{0,k,t} | \mu_{0,k,t}) = \frac{1}{(2\pi\sigma_{0,k,t}^2)^{1/2}} e^{-\frac{1}{2\sigma_{0,k,t}^2}(y_{0,k,t} - \mu_{0,k,t})^2} \\ f(\mathbf{y}_{D_{j,k,t}^*} | y_{j,k,t}, \boldsymbol{\theta}_{j,k,t}) = \frac{e^{-\mathbf{a}'(\boldsymbol{\Omega}_{j,k,t})^{-1}\mathbf{a}/2}}{(2\pi)^{n_{D_{j,k,t}^*}/2} |\boldsymbol{\Omega}_{j,k,t}|^{1/2}},$$

com $\mathbf{a} = \left(\mathbf{y}_{D_{j,k,t}^*} - (\nu_{j,k,t} y_{j,k,t} + \boldsymbol{\theta}_{j,k,t}) \right)$.

$$f(\mu_{0,k,t} | \mu_{0,k,t-1}) = \frac{1}{(2\pi W_{0,k,t})^{1/2}} e^{-\frac{1}{2W_{0,k,t}^2}(\mu_{0,k,t} - \mu_{0,k,t-1})^2}$$

$$f(\boldsymbol{\theta}_{j,k,t} | \boldsymbol{\theta}_{j,k,t-1}) = \frac{|W_{j,k,t}|^{-1/2}}{(2\pi)^{1/2}} e^{-\frac{1}{2}(\boldsymbol{\theta}_{j,k,t} - \boldsymbol{\theta}_{j,k,t-1})'(W_{j,k,t})^{-1}(\boldsymbol{\theta}_{j,k,t} - \boldsymbol{\theta}_{j,k,t-1})}$$

$$f(\mu_{0,k,0} | m_0, c_0) = \frac{1}{(2\pi c_0)^{1/2}} e^{-\frac{1}{2c_0^2}(\mu_{0,k,0} - m_0)^2}$$

$$f(\boldsymbol{\theta}_{j,k,0} | \mathbf{n}_0, \mathbf{d}_0) = \frac{|\mathbf{d}_0|^{-1/2}}{(2\pi)^{1/2}} e^{-\frac{1}{2}(\boldsymbol{\theta}_{j,k,0} - \mathbf{n}_0)'(\mathbf{d}_0)^{-1}(\boldsymbol{\theta}_{j,k,0} - \mathbf{n}_0)}$$

Além da notação já apresentada, sejam também as seguintes expressões:

$$\boldsymbol{\theta}_{j,k,\cdot} = (\boldsymbol{\theta}'_{j,k,0}, \dots, \boldsymbol{\theta}'_{j,k,T})'$$

$$\boldsymbol{\theta}_{\cdot,k,t} = (\boldsymbol{\theta}'_{0,k,t}, \dots, \boldsymbol{\theta}'_{J-2,k,t})'$$

$$\boldsymbol{\theta}_{\dots} = (\boldsymbol{\theta}_{0,1,\cdot}, \dots, \boldsymbol{\theta}_{0,n_0,\cdot}, \boldsymbol{\theta}_{1,1,\cdot}, \dots, \boldsymbol{\theta}_{1,n_1,\cdot}, \dots, \boldsymbol{\theta}_{J-2,1,\cdot}, \dots, \boldsymbol{\theta}_{J-2,n_{J-2},\cdot})'$$

O resultado que segue é extremamente importante pois fornece o embasamento teórico necessário à construção do Filtro de Kalman para o Modelo Espaço-Temporal Multi-Escala apresentado.

Teorema 1 - Considere $f(\mathbf{y}_{J-1}, \boldsymbol{\mu}_0, \boldsymbol{\theta} | \sigma^2, \psi_k, \psi_{j,k})$ a função de densidade conjunta do Modelo Espaço-Temporal Multi-Escala descrita em (5.6). Pode-se afirmar que, à posteriori, os vetores $\boldsymbol{\mu}_{0,1,\cdot}, \dots, \boldsymbol{\mu}_{0,n_0,\cdot}, \boldsymbol{\theta}_{0,1,\cdot}, \dots, \boldsymbol{\theta}_{0,n_0,\cdot}, \dots, \boldsymbol{\theta}_{J-2,1,\cdot}, \dots, \boldsymbol{\theta}_{J-2,n_{J-2},\cdot}$ são condicionalmente independentes e normalmente distribuídos, dado $(\sigma^2, \psi_k, \psi_{j,k})$.

Demonstração: Para provar a independência, observe que a distribuição à posteriori de $\boldsymbol{\mu}_{0,1,\cdot}, \dots, \boldsymbol{\mu}_{0,n_0,\cdot}, \boldsymbol{\theta}_{0,1,\cdot}, \dots, \boldsymbol{\theta}_{0,n_0,\cdot}, \dots, \boldsymbol{\theta}_{J-2,1,\cdot}, \dots, \boldsymbol{\theta}_{J-2,n_{J-2},\cdot}$ é proporcional à expressão (5.6). A distribuição a posteriori $f(\boldsymbol{\mu}_0, \boldsymbol{\theta} | \sigma^2, \psi_k, \psi_{j,k}, \mathbf{y}_{J-1})$ pode ser

escrita como:

$$\begin{aligned}
& \propto \prod_{t=1}^T \left\{ \prod_{k=1}^{n_0} f(y_{0,k,t} | \mu_{0,k,t}) \prod_{j=0}^{J-2} \prod_{k=1}^{n_j} f(\mathbf{y}_{D_{j,k,t}^*} | y_{j,k,t}, \boldsymbol{\theta}_{j,k,t}) \right. \\
& \quad \left. \prod_{k=1}^{n_0} f(\mu_{0,k,t} | \mu_{0,k,t-1}) \prod_{j=0}^{J-2} \prod_{k=1}^{n_j} f(\boldsymbol{\theta}_{j,k,t} | \boldsymbol{\theta}_{j,k,t-1}) \right\} \\
& \quad \prod_{k=1}^{n_0} f(\mu_{0,k,0}) \prod_{j=0}^{J-2} \prod_{k=1}^{n_j} f(\boldsymbol{\theta}_{j,k,0}) \\
& = \prod_{k=1}^{n_0} \left\{ f(\mu_{0,k,0}) \prod_{t=1}^T f(y_{0,k,t} | \mu_{0,k,t}) f(\mu_{0,k,t} | \mu_{0,k,t-1}) \right\} \prod_{j=0}^{J-2} \prod_{k=1}^{n_j} \\
& \quad \left\{ f(\boldsymbol{\theta}_{j,k,0}) \prod_{t=1}^T f(\mathbf{y}_{D_{j,k,t}^*} | y_{j,k,t}, \boldsymbol{\theta}_{j,k,t}) f(\boldsymbol{\theta}_{j,k,t} | \boldsymbol{\theta}_{j,k,t-1}) \right\} \\
& \propto \prod_{k=1}^{n_0} f(\mu_{0,k,\cdot} | \mathbf{y}_{0,k,\cdot}) \prod_{j=0}^{J-2} \prod_{k=1}^{n_j} f(\boldsymbol{\theta}_{j,k,\cdot} | \boldsymbol{\theta}_{j,k,\cdot}^e)
\end{aligned}$$

Verifica-se que $f(\boldsymbol{\mu}_0, \boldsymbol{\theta} | \sigma^2, \psi_k, \psi_{j,k}, \mathbf{y}_{J-1})$ fatora-se no produtório das distribuições a posteriori de $\boldsymbol{\mu}_{0,1,\cdot}, \dots, \boldsymbol{\mu}_{0,n_0,\cdot}, \boldsymbol{\theta}_{0,1,\cdot}, \dots, \boldsymbol{\theta}_{0,n_0,\cdot}, \dots, \boldsymbol{\theta}_{J-2,1,\cdot}, \dots, \boldsymbol{\theta}_{J-2,n_{J-2},\cdot}$, provando, portanto, a independência a posteriori entre tais vetores.

Por outro lado, utilizando a equação de evolução juntamente com a distribuição a priori normal para $\mu_{0,k,\cdot}$ e $\boldsymbol{\theta}_{\dots,\cdot}$, tem-se como resultado uma priori normal multivariada para os vetores $\boldsymbol{\mu}_{0,1,\cdot}, \dots, \boldsymbol{\mu}_{0,n_0,\cdot}, \boldsymbol{\theta}_{0,1,\cdot}, \dots, \boldsymbol{\theta}_{0,n_0,\cdot}, \dots, \boldsymbol{\theta}_{J-2,1,\cdot}, \dots, \boldsymbol{\theta}_{J-2,n_{J-2},\cdot}$. Logo, como tanto a distribuição dos dados $\mathbf{y}_{j,t}$ quanto a priori para os vetores citados são normais multivariadas, conclui-se que a distribuição a posteriori para cada um dos vetores também é normal multivariada, o que resulta na prova do teorema.

Esse teorema permite que, dado $(\sigma^2, \psi_k, \psi_{j,k})$, as respectivas distribuições a posteriori de $\boldsymbol{\mu}_{0,1,\cdot}, \dots, \boldsymbol{\mu}_{0,n_0,\cdot}, \boldsymbol{\theta}_{0,1,\cdot}, \dots, \boldsymbol{\theta}_{0,n_0,\cdot}, \dots, \boldsymbol{\theta}_{J-2,1,\cdot}, \dots, \boldsymbol{\theta}_{J-2,n_{J-2},\cdot}$ sejam descritas individualmente, a partir da distribuição conjunta expressa em (5.6). As respectivas médias e variâncias são fornecidas pelas equações do Filtro de Kalman. Estas

equações, por sua vez, serão obtidas equivalentemente ao modelo linear dinâmico já conhecido. Nas próximas seções é apresentada a inferência em detalhes.

5.2 Distribuições a Priori

Observe que de (4.4), σ^2 se faz presente nas expressões da variância para todos os níveis de agregação. Uma distribuição a priori, dados δ e λ , pode ser:

$$\sigma^2 \sim GI\left(\frac{\delta}{2}, \frac{\lambda}{2}\right).$$

Como citado anteriormente, os fatores de proporcionalidade ψ_k e $\psi_{j,k}$ são frações das respectivas variâncias observacionais. Na literatura de MLD's, sabe-se que essa medida é bem difícil de ser estimada. Na prática, espera-se que ψ_k e $\psi_{j,k}$ sejam razoavelmente menores do que 1. Logo, priori's pouco informativas tendem a gerar resultados ruins. Assim sendo, dados τ e κ , as distribuições a priori para ψ_k e $\psi_{j,k}$ serão dadas, de forma idêntica, por:

$$GI\left(\frac{\tau}{2}, \frac{\kappa}{2}\right).$$

Finalmente, tomando como conhecidos os hiperparâmetros $m_{0,k,0}$, $C_{0,k,0}$, $m_{j,k,0}$ e $C_{j,k,0}$, as distribuições a priori para $\mu_{0,j,0}$ e $\theta_{j,k,0}$ são dadas, respectivamente, por:

$$\mu_{0,k,0}|D_0 \sim N(m_{0,k,0}, C_{0,k,0})$$

e

$$\theta_{j,k,0}|D_0 \sim N(\mathbf{m}_{j,k,0}, \mathbf{C}_{j,k,0}).$$

5.3 Simulação e Estimação de $\mu_{0,k,t}$

Inicia-se a estimação para o nível inicial mais agregado da partição, isto é, $y_{0,k,t}$. Considere ψ_k e σ^2 conhecidos para todo k . Tem-se então as seguintes equações:

Equação de Observação:

$$y_{0,k,t} = \mu_{0,k,t} + v_{0,k}, \quad v_{0,k} | \sigma_{0,k}^2 \sim N(0, \sigma_{0,k}^2)$$

Equação de Sistema:

$$\mu_{0,k,t} = \mu_{0,k,t-1} + w_{0,k}, \quad w_{0,k} | \psi_k \sim N(0, \psi_k \sigma_{0,k}^2).$$

5.3.1 Distribuições a Posteriori

A distribuição conjunta a posteriori de $\mu_{0,k,1:T}$ e ψ_k pode ser expressa por:

$$f(\mu_{0,k,1:T}, \psi_k | \sigma^2, D_T) \propto \prod_{t=1}^T f(\mathbf{y}_{0,k,t} | \mu_{0,k,t}, \sigma^2) f(\mu_{0,k,t} | \mu_{0,k,t-1}, \psi_k, \sigma^2) f(\psi_k)$$

Nos itens que se seguem são apresentadas as respectivas distribuições condicionais completas de $\mu_{0,k,1:T}$ e ψ_k , das quais serão realizadas as simulações para fins de estimação.

Simulação de $\mu_{0,k,1:T}$ Via FFBS

Considere $\mu_{0,k,1:T} = (\mu_{0,k,1}, \dots, \mu_{0,k,T})$. A partir de sua distribuição condicional completa $f(\mu_{0,k,1:T} | \sigma^2, \psi_k, D_T)$ geram-se amostras e procede-se a estimação do vetor $\mu_{0,k,1:T}$. A amostragem sequencial de $\mu_{0,k,1:T}$ pode ser realizada via FFBS tal como:

1. Inicialmente, utilizar o Filtro de Kalman para obter a média e a variância das distribuições normais $f(\mu_{0,k,1} | \sigma^2, \psi_k, D_1), \dots, f(\mu_{0,k,T} | \sigma^2, \psi_k, D_T)$. Dados σ^2 e ψ_k , essa operação pode ser realizada a partir da construção do Filtro de Kalman, como mostra o texto a seguir.

Construção do Filtro de Kalman para $\mu_{0,k,t}$

- Posteriori em $t - 1$:

$$\mu_{0,k,t-1} | D_{0,k,t-1} \sim N(m_{0,k,t-1}, C_{0,k,t-1})$$

- Priori em t :

$$\mu_{0,k,t} | D_{0,k,t-1} \sim N(a_{0,k,t}, R_{0,k,t})$$

onde:

$$\begin{aligned} a_{0,k,t} &= m_{0,k,t-1} \\ R_{0,k,t} &= C_{0,k,t-1} + \psi_k \sigma_{0,k}^2 \end{aligned}$$

- Previsão à 1 passo:

$$y_{0,k,t} | D_{0,k,t-1} \sim N(f_{0,k,t}, Q_{0,k,t})$$

onde:

$$\begin{aligned} f_{0,k,t} &= m_{0,k,t-1} \\ Q_{0,k,t} &= R_{0,k,t} + \sigma_{0,k}^2 \end{aligned}$$

- Posteriori em t :

$$\mu_{0,k,t} | D_{0,k,t} \sim N(m_{0,k,t}, C_{0,k,t}) \tag{5.7}$$

onde:

$$\begin{aligned} m_{0,k,t} &= m_{0,k,t-1} + A_{0,k,t} e_{0,k,t} \\ C_{0,k,t} &= R_{0,k,t} - A_{0,k,t}^2 Q_{0,k,t} \\ A_{0,k,t} &= R_{0,k,t} / Q_{0,k,t} \\ e_{0,k,t} &= y_{0,k,t} - f_{0,k,t} \end{aligned}$$

- Suavização:

$$\mu_{0,k,T-r} | D_{0,k,T} \sim N(a_{0,k,T}(-r), R_{0,k,T}(-r)),$$

onde:

$$a_{0,k,T}(-r) = m_{0,k,T-r} + B_{0,k,T-r}[a_{0,k,T}(-r+1) - a_{0,k,T-r+1}]$$

$$R_{0,k,T}(-r) = C_{0,k,T-r} - B_{0,k,T-r}^2[R_{0,k,T-r+1} - R_{0,k,T}(-r+1)],$$

com $B_{0,k,T-r} = C_{0,k,T-r}R_{0,k,T-r+1}^{-1}$ e valores iniciais $a_{0,k,T}(0) = m_{0,k,T}$ e $R_{0,k,T}(0) = C_{0,k,T}$.

2. Gere um valor para $\mu_{0,k,T}$ a partir de $\mu_{0,k,T} | \sigma^2, \psi_k, D_T \sim N(m_{0,k,T}, C_{0,k,T})$.
3. Calcule a média e a variância de $f(\mu_{0,k,T-1} | \mu_{0,k,T}, \sigma^2, \psi_k, D_T)$ e gere $\mu_{0,k,T-1}$ a partir dessa distribuição.
4. Obtenha recursivamente

$$\begin{aligned} f(\mu_{0,k,T-r} | \mu_{0,k,T-r+1}, \dots, \mu_{0,k,T}, D_T) &= f(\mu_{0,k,T-r} | \mu_{0,k,T-r+1}, D_T) \\ &\sim N(h_{0,k,T}(r), H_{0,k,T}(r)), \end{aligned}$$

para $r = 1, \dots, T-1$, sendo:

$$h_{0,k,T}(r) = m_{0,k,T-r} + B_{0,k,T-r}(\mu_{0,k,T-r+1} - a_{0,k,T-r+1})$$

e

$$H_{0,k,T}(r) = C_{0,k,T-r} - B_{0,k,T-r}^2 R_{0,k,T-r+1},$$

sendo $B_{0,k,T-r} = C_{0,k,T-r}R_{0,k,T-r+1}^{-1}$. Em seguida, gerar $\mu_{0,k,T-r}$ dessa distribuição.

A partir do algoritmo apresentado são gerados os valores de $\mu_{0,k,1:T}$ condicionados em (σ^2, ψ_k, D_T) .

Condicional Completa de ψ_k

$$f(\psi_k | \mu_{0,k,1:T}, \sigma^2, D_T) \propto \prod_{t=1}^T f(\mu_{0,k,t} | \mu_{0,k,t-1}, \sigma^2, \psi_k) f(\psi_k | \tau, \kappa)$$

$$f(\psi_k | \mu_{0,k,1:T}, \sigma^2, D_T) \propto \psi_k^{-\left(\frac{T+\tau}{2}+1\right)} \exp \left[-\frac{1}{2\psi_k} \left(\kappa + \frac{1}{\sigma_{0,k}^2} \sum_{t=1}^T (\mu_{0,k,t} - \mu_{0,k,t-1})^2 \right) \right]$$

$$\psi_k | \mu_{0,k,1:T}, \sigma^2, D_T \sim GI \left(\frac{\tau_k^*}{2}, \frac{\kappa_k^*}{2} \right),$$

onde $\tau_k^* = T + \tau$ e $\kappa_k^* = \kappa + \frac{1}{\sigma_{0,k}^2} \sum_{t=1}^T (\mu_{0,k,t} - \mu_{0,k,t-1})^2$.

5.4 Simulação e Estimação de $\theta_{j,k,t}$

A descrição nesta seção envolve todos os demais níveis de escala do processo. Sejam então as seguintes equações do modelo:

Equação de Observação:

$$\theta_{j,k,t}^e = \theta_{j,k,t} + \mathbf{v}_{j,k}, \quad \mathbf{v}_{j,k} | \boldsymbol{\Omega}_{j,k} \sim N(\mathbf{0}, \boldsymbol{\Omega}_{j,k})$$

Note que:

$$\boldsymbol{\Omega}_{J-2,k} = \sigma^2 \left[\mathbf{I}_{(n_{D_{J-2,k}^*}) \times (n_{D_{J-2,k}^*})} - \left(\frac{\mathbf{1}_{n_{D_{J-2,k}^*}} \mathbf{1}'_{n_{D_{J-2,k}^*}}}{n_{D_{J-2,k}^*} + 1} \right) \right].$$

Equação de Sistema:

$$\theta_{j,k,t} = \theta_{j,k,t-1} + \mathbf{w}_{j,k}, \quad \mathbf{w}_{j,k} | \sigma^2, \psi_{j,k} \sim N(\mathbf{0}, \psi_{j,k} \boldsymbol{\Omega}_{j,k}).$$

5.4.1 Distribuições a Posteriori

A distribuição conjunta a posteriori de $\theta_{j,k,1:T}$ e $\psi_{j,k}$ pode ser expressa por:

$$f(\theta_{j,k,1:T}, \psi_{j,k} | \sigma^2, D_T) \propto \prod_{t=1}^T f(\mathbf{y}_{j,k,t} | \theta_{j,k,t}, \sigma^2) f(\theta_{j,k,t} | \theta_{j,k,t-1}, \sigma^2, \psi_{j,k}) f(\psi_{j,k})$$

A seguir são apresentados os respectivos procedimentos de simulação de $\boldsymbol{\theta}_{j,k,1:T}$ e $\psi_{j,k}$.

Simulação de $\boldsymbol{\theta}_{j,k,1:T}$ Via FFBS

Considere $\boldsymbol{\theta}_{j,k,1:T} = (\boldsymbol{\theta}_{j,k,1}, \dots, \boldsymbol{\theta}_{j,k,T})$. Analogamente para o caso do nível $j = 0$, dados σ^2 e $\psi_{j,k}$, a simulação e estimação de $\boldsymbol{\theta}_{j,k,t}$ será realizada a partir da construção do Filtro de Kalman e do uso do FFBS, descrito a seguir.

1. O passo inicial será utilizar o Filtro de Kalman para obter a média e a variância das distribuições normais $f(\boldsymbol{\theta}_{j,k,1}|\sigma^2, \psi_{j,k}, D_1), \dots, f(\boldsymbol{\theta}_{j,k,T}|\sigma^2, \psi_{j,k}, D_T)$.

Esquema para $\boldsymbol{\theta}_{j,k,t}$.

Considere σ^2 e $\psi_{j,k}$ conhecidos para todo k . Logo, temos as seguintes expressões:

- Posteriori em $t - 1$:

$$\boldsymbol{\theta}_{j,k,t-1}|D_{j,k,t-1} \sim N(\mathbf{m}_{j,k,t-1}, \mathbf{C}_{j,k,t-1})$$

- Priori em t :

$$\boldsymbol{\theta}_{j,k,t}|D_{j,k,t-1} \sim N(\mathbf{a}_{j,k,t}, \mathbf{R}_{j,k,t})$$

onde:

$$\begin{aligned} \mathbf{a}_{j,k,t} &= \mathbf{m}_{j,k,t-1} \\ \mathbf{R}_{j,k,t} &= \mathbf{C}_{j,k,t-1} + \psi \boldsymbol{\Omega}_{j,k} \end{aligned}$$

- Previsão à 1 passo:

$$\mathbf{y}_{j,k,t}|D_{j,k,t-1} \sim N(\mathbf{f}_{j,k,t}, \mathbf{Q}_{j,k,t})$$

onde:

$$\begin{aligned}\mathbf{f}_{j,k,t} &= \mathbf{m}_{j,k,t-1} \\ \mathbf{Q}_{j,k,t} &= \mathbf{R}_{j,k,t} + \mathbf{\Omega}_{j,k}\end{aligned}$$

- Posteriori em t :

$$\boldsymbol{\theta}_{j,k,t} | D_{j,k,t} \sim N(\mathbf{m}_{j,k,t}, \mathbf{C}_{j,k,t}) \quad (5.8)$$

onde:

$$\begin{aligned}\mathbf{e}_{j,k,t} &= \boldsymbol{\theta}_{j,k,t}^e - \mathbf{f}_{j,k,t} \\ \mathbf{m}_{j,k,t} &= \mathbf{m}_{j,k,t-1} + \mathbf{A}_{j,k,t} \mathbf{e}_{j,k,t} \\ \mathbf{C}_{j,k,t} &= \mathbf{R}_{j,k,t} - \mathbf{A}_{j,k,t} \mathbf{Q}_{j,k,t} \mathbf{A}'_{j,k,t} \\ \mathbf{A}_{j,k,t} &= \mathbf{R}_{j,k,t} \mathbf{Q}_{j,k,t}^{-1}\end{aligned}$$

- Suavização:

$$\boldsymbol{\theta}_{j,k,T-r} | D_{j,k,T} \sim N(\mathbf{a}_{j,k,T}(-r), \mathbf{R}_{j,k,T}(-r)),$$

onde:

$$\begin{aligned}\mathbf{a}_{j,k,T}(-r) &= \mathbf{m}_{j,k,T-r} + \mathbf{B}_{j,k,T-r} [\mathbf{a}_{j,k,T}(-r+1) - \mathbf{a}_{j,k,T-r+1}] \\ \mathbf{R}_{j,k,T}(-r) &= \mathbf{C}_{j,k,T-r} - \mathbf{B}_{j,k,T-r} [\mathbf{R}_{j,k,T-r+1} - \mathbf{R}_{j,k,T}(-r+1)] \mathbf{B}'_{j,k,T-r},\end{aligned}$$

com $\mathbf{B}_{j,k,T-r} = \mathbf{C}_{j,k,T-r} \mathbf{R}_{j,k,T-r+1}^{-1}$ e valores iniciais $\mathbf{a}_{j,k,T}(0) = \mathbf{m}_{j,k,T}$ e $\mathbf{R}_{j,k,T}(0) = \mathbf{C}_{j,k,T}$.

2. Simule um valor para $\boldsymbol{\theta}_{j,k,T}$ a partir de $(\boldsymbol{\theta}_{j,k,T} | \sigma^2, \psi_{j,k}, D_T) \sim N(\mathbf{m}_{j,k,T}, \mathbf{C}_{j,k,T})$.
3. Calcule a média e a variância de $f(\boldsymbol{\theta}_{j,k,T-1} | \boldsymbol{\theta}_{j,k,T}, \sigma^2, \psi_{j,k}, D_T)$ e gere $\boldsymbol{\theta}_{j,k,T-1}$ a partir dessa distribuição.

4. Obtenha recursivamente

$$\begin{aligned} f(\boldsymbol{\theta}_{j,k,T-r} | \boldsymbol{\theta}_{j,k,T-r+1}, \dots, \boldsymbol{\theta}_{j,k,T}, D_T) &= f(\boldsymbol{\theta}_{j,k,T-r} | \boldsymbol{\theta}_{j,k,T-r+1}, D_T) \\ &\sim N(\mathbf{h}_{j,k,T}(r), \mathbf{H}_{j,k,T}(r)), \end{aligned}$$

sendo $r = 1, \dots, T$,

$$\mathbf{h}_{j,k,T}(r) = \mathbf{m}_{j,k,T-r} + \mathbf{B}_{j,k,T-r}(\boldsymbol{\theta}_{j,k,T-r+1} - \mathbf{a}_{j,k,T-r+1})$$

e

$$\mathbf{H}_{j,k,T}(r) = \mathbf{C}_{j,k,T-r} - \mathbf{B}_{j,k,T-r} \mathbf{R}_{j,k,T-r+1} \mathbf{B}'_{j,k,T-r},$$

com $\mathbf{B}_{j,k,T-r} = \mathbf{C}_{j,k,T-r} \mathbf{R}_{j,k,T-r+1}^{-1}$. Em seguida, gerar $\boldsymbol{\theta}_{j,k,T-r}$ dessa distribuição.

A partir do algoritmo são gerados os valores de $\boldsymbol{\theta}_{j,k,1:T}$ condicionados em $(\sigma^2, \psi_{j,k}, D_T)$.

Condicional Completa de $\psi_{j,k}$

$$\begin{aligned} f(\psi_{j,k} | \boldsymbol{\theta}_{j,k,1:T}, \sigma^2, D_T) &\propto \prod_{t=1}^T f(\boldsymbol{\theta}_{j,k,t} | \boldsymbol{\theta}_{j,k,t-1}, \psi_{j,k}, \sigma^2, D_{t-1}) f(\psi_{j,k} | \tau, \kappa) \\ f(\psi_{j,k} | \boldsymbol{\theta}_{j,k,1:T}, \sigma^2, D_T) &\propto \psi_{j,k}^{-\frac{1}{2} (n_{D_{j,k}^*} T + \tau) - 1} \exp \left[-\frac{1}{2\psi_{j,k}} (\kappa_{j,k}) \right] \\ &\times \exp \left[-\frac{1}{2\psi_{j,k}} \left(\sum_{t=1}^T (\boldsymbol{\theta}_{j,k,t} - \boldsymbol{\theta}_{j,k,t-1})' \boldsymbol{\Omega}_{j,k}^{-1} (\boldsymbol{\theta}_{j,k,t} - \boldsymbol{\theta}_{j,k,t-1}) \right) \right] \\ \psi_{j,k} | \boldsymbol{\theta}_{j,k,1:T}, \sigma^2, D_T &\sim GI \left(\frac{\tau_{j,k}^*}{2}, \frac{\kappa_{j,k}^*}{2} \right), \end{aligned}$$

onde

$$\tau_{j,k}^* = n_{D_{j,k}^*} T + \tau \text{ e } \kappa_{j,k}^* = \kappa + \sum_{t=1}^T (\boldsymbol{\theta}_{j,k,t} - \boldsymbol{\theta}_{j,k,t-1}) \boldsymbol{\Omega}_{j,k}^{-1} (\boldsymbol{\theta}_{j,k,t} - \boldsymbol{\theta}_{j,k,t-1})'.$$

Condicional Completa de σ^2

$$\sigma^2 | \boldsymbol{\mu}_{J-1,t}, D_T \sim GI \left(\frac{\delta^*}{2}, \frac{\lambda^*}{2} \right),$$

onde $\delta^* = n_{J-1}T + \delta$ e $\lambda^* = \lambda + \sum_{t=1}^T \sum_{k=1}^{n_{J-1}} (y_{J-1,k,t} - \mu_{J-1,k,t})^2$.

5.5 Estimação de $\boldsymbol{\mu}_{J-1,t}$

Um dos objetivos deste capítulo é o de estudar o comportamento do fator latente do nível mais desagregado do processo ao longo do tempo, expresso por $\boldsymbol{\mu}_{J-1,t} = (\mu_{J-1,1,t}, \dots, \mu_{J-1,n_{J-1},t})'$. Logo, se faz necessário encontrar o estimador para $(\boldsymbol{\mu}_{0,,t}, \dots, \boldsymbol{\mu}_{J-2,,t})'$. O caminho inicial será encontrar um estimador para $\boldsymbol{\mu}_{0,,t}$ e, fazendo uso recursivo da expressão (4.2), obter um estimador para $\boldsymbol{\mu}_{J-1,t}$. Observe que, a partir de (4.2) tem-se que

$$\boldsymbol{\mu}_{D_{j,k}^*,t} = \boldsymbol{\theta}_{j,k,t} + \boldsymbol{\nu}_{j,k,t} \mu_{j,k,t}.$$

O procedimento de estimação é realizado da seguinte forma:

- Estimador para $\boldsymbol{\mu}_{0,,t}$.

A partir da média da densidade a posteriori expressa em (5.7), tem-se que um estimador para $\mu_{0,k,t}$ é dado por $\hat{\mu}_{0,k,t} = m_{0,k,t-1} + A_{0,k,t} e_{0,k,t}$. Daí que $\hat{\boldsymbol{\mu}}_{0,,t} = (\hat{\mu}_{0,1,t}, \dots, \hat{\mu}_{0,n_0,t})'$.

- Estimador para $\boldsymbol{\mu}_{D_{j,k}^*,t}$.

Tomando-se a média da distribuição apresentada em (5.8), tem-se que $\hat{\boldsymbol{\theta}}_{j,k,t} = \mathbf{m}_{j,k,t}$. Logo, um estimador para $\boldsymbol{\mu}_{D_{j,k}^*,t}$ é dado pela sua média a posteriori, calculada por

$$\hat{\boldsymbol{\mu}}_{D_{j,k}^*,t} = \hat{\boldsymbol{\theta}}_{j,k,t} + \boldsymbol{\nu}_{j,k,t} \hat{\mu}_{j,k,t}. \quad (5.9)$$

Observe, porém, que conforme descrito no capítulo 2, pelo fato de os cálculos serem gerados com um elemento a menos na dimensão do vetor, se faz necessário calcular o estimador para esse elemento faltante, o que pode ser realizado da seguinte maneira:

- Após estimar $\mu_{j,k,t}$ e $\theta_{j,k,t}$, este possuindo dimensão igual a $n_{D_{j,k}^*}$, calcula-se o estimador de $\boldsymbol{\mu}_{D_{j,k}^*,t} = \boldsymbol{\theta}_{j,k,t} + \boldsymbol{\nu}_{j,k,t}\mu_{j,k,t}$, que por sua vez, também possui dimensão igual a $n_{D_{j,k}^*}$.
- Em seguida, estima-se o elemento inicialmente excluído pela expressão

$$\hat{\boldsymbol{\mu}}_{j,k,t} - \sum_{j+1,k' \in D_{j,k}^*} \hat{\boldsymbol{\mu}}_{j+1,k',t}. \quad (5.10)$$

Desse modo, calculamos o estimador para a média do processo latente para todos os descendentes de $B_{j,k}$.

Note que, recursivamente, o uso das expressões (5.9) e (5.10) irá fornecer um estimador para $\boldsymbol{\mu}_{J-1,t}$.

5.6 Simulação de Um Processo Multi-Escala

Simulamos dados de um processo multi-escala tomando como referência a divisão político administrativa do estado do Espírito Santo no ano de 1979. O processo tem três níveis hierárquicos, com a seguinte estrutura de partição:

- O primeiro nível, $j = 0$, é formado por 4 elementos (macro-regiões).
- O segundo nível, $j = 1$, contém 12 elementos (micro-regiões).
- O nível mais desagregado, $j = 2$, é formado por 52 elementos (municípios).

Os dados foram simulados como segue:

1. Inicialmente, foram fornecidos os valores de $\mu_{0,k,0}$, para $k = 1, \dots, 4$. Isto é, o vetor de médias no nível $j = 0$ para o tempo $t = 0$.

- Em seguida foram simulados, com erro normal dado por (4.8), os valores de $\mu_{0,..,t}$, para $t = 1, \dots, T$.
2. O passo seguinte foi fornecer valores para $\theta_{0,..,0}$ e para as respectivas matrizes de covariância, conforme descritos em (4.9).
 - Posteriormente simulou-se, com erro normal, os valores de $\theta_{0,..,t}$ para todos instantes de tempos.
 - Tendo gerados os valores de $\mu_{0,..,t}$ e $\theta_{0,..,t}$, calculou-se os valores de $\mu_{1,..,t}$ conforme (4.2).
 3. Em seguida, forneceu-se o valor de $\theta_{1,..,0}$ e suas respectivas matrizes de covariância.
 - Após, gerou-se os valores de $\theta_{1,..,t}$, com erro normalmente distribuído.
 - Seguiu-se com o cálculo de $\mu_{2,..,t}$ conforme (4.2).
 4. Dados os valores de $\mu'_{2,..,t}$ gerados, em seguida foram simulados, com erro normalmente distribuído, os valores da variável $\mathbf{y}_{2,..,t}$ conforme (4.5).
 5. Gerados os valores de $\mathbf{y}_{2,..,t}$, finalmente calcula-se os valores de $\mathbf{y}_{1,..,t}$ e $\mathbf{y}_{0,..,t}$ com base na expressão 2.1.

5.7 Casos Particulares

O objetivo dessa seção é apresentar resultados iniciais do modelo espaço-temporal multi-escala de forma a analisar o grau de eficiência do mesmo no ajuste aos dados simulados.

5.7.1 Variâncias Conhecidas

Uma situação particular ocorre quando as variâncias observacionais e de sistema são conhecidas. Nesse caso, com uso análogo do Teorema 1, pode-se aplicar diretamente o Filtro de Kalman para fins de estimação dos parâmetros envolvidos. As matrizes de covariância de sistema utilizadas na geração dos dados são da forma $\mathbf{W}_{j,k} = 0.10\mathbf{\Omega}_{j,k}$ e $\mathbf{V}_2 = \text{diag}(100, \dots, 100)$. Para inicializar o Filtro de Kalman foram imputadas as priori's $\mu_{0,k,0}|D_0 \sim N(0, 10^9)$ e $\theta_{j,k,0}|D_0 \sim N((0, \dots, 0)', \text{diag}(10^9))$ nas suas respectivas dimensões.

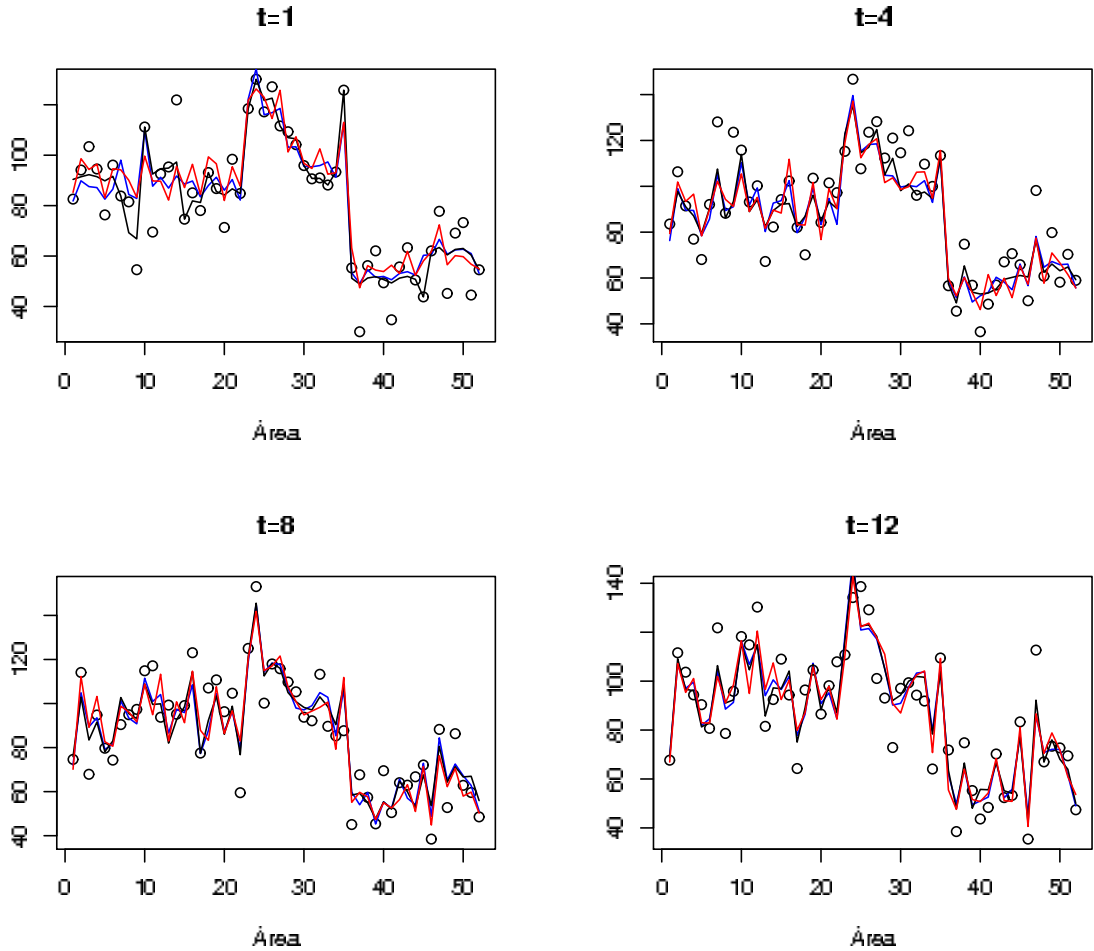


Figura 5.1: Resultados comparativos de $y_{2,..,t}$ (círculos), verdadeiro valor de $\mu_{2,..,t}$ (vermelho), $\hat{\mu}_{2,..,t}$ (preto) e estimativa suavizada $\mathbf{a}_{j,k,T}(-r)$ (azul). Tempos $t = 1, 4, 8, 12$, respectivamente.

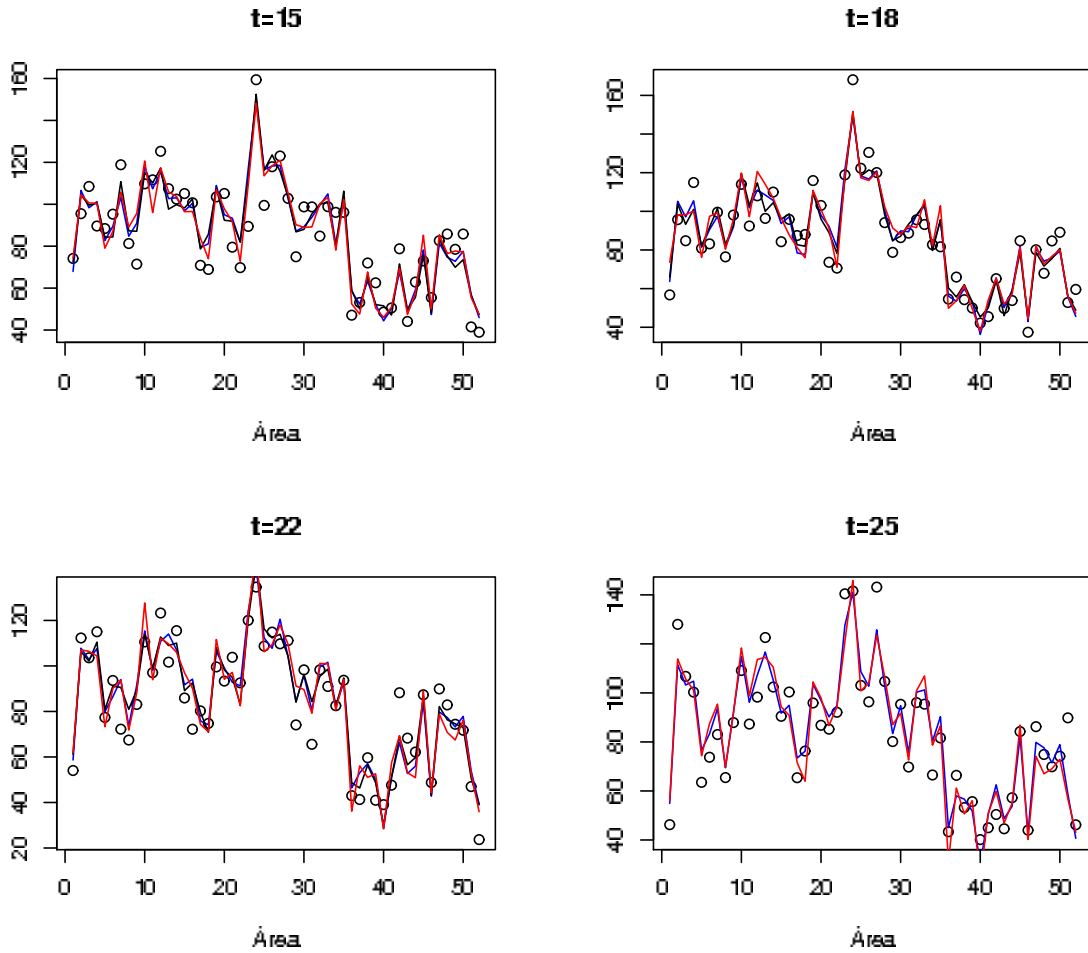


Figura 5.2: Resultados comparativos de $y_{2,..,t}$ (círculos), verdadeiro valor de $\mu_{2,..,t}$ (vermelho), $\hat{\mu}_{2,..,t}$ (preto) e estimativa suavizada $\mathbf{a}_{j,k,T}(-r)$ (azul). Tempos $t = 15, 18, 22, 25$, respectivamente.

As figuras 5.1 e 5.2 mostram os resultados obtidos com os dados simulados para as 52 áreas hipotéticas e para tempos intercalados, onde comparou-se $y_{2,..,t}$, $\mu_{2,..,t}$ verdadeiros, estimativas para $\mu_{2,..,t}$ e a suavização, expressa por $\mathbf{a}_{j,k,T}(-r)$. Observe que as estimativas ficam próximas dos respectivos verdadeiros valores médios do fator latente e pouco sensíveis a valores discrepantes, mostrando um ajuste satisfatório. Para o caso das figuras 5.3 a 5.7, apresentamos as comparações das mesmas medidas utilizando porém o mapa do estado do Espírito Santo, comparações estas que reforçam o bom ajuste do modelo.

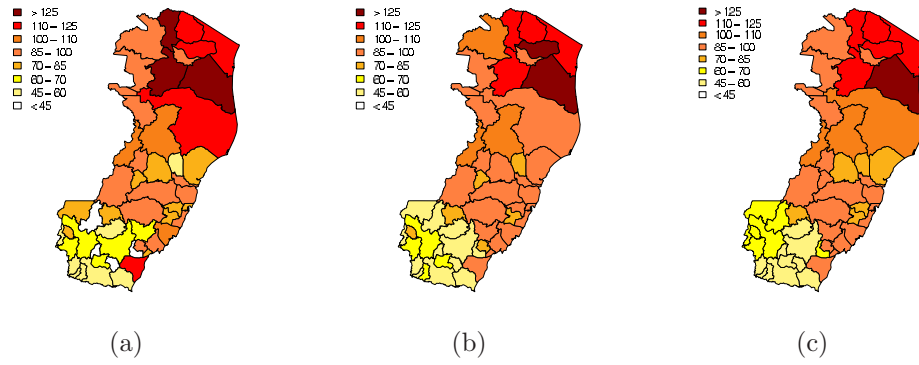


Figura 5.3: Valores de $y_{2,k,t}$ observados (a), $\mu_{2,k,t}$ verdadeiros (b) e suavizados (c) por $\mathbf{a}_{j,k,T}(-r)$. Tempo $t = 1$.

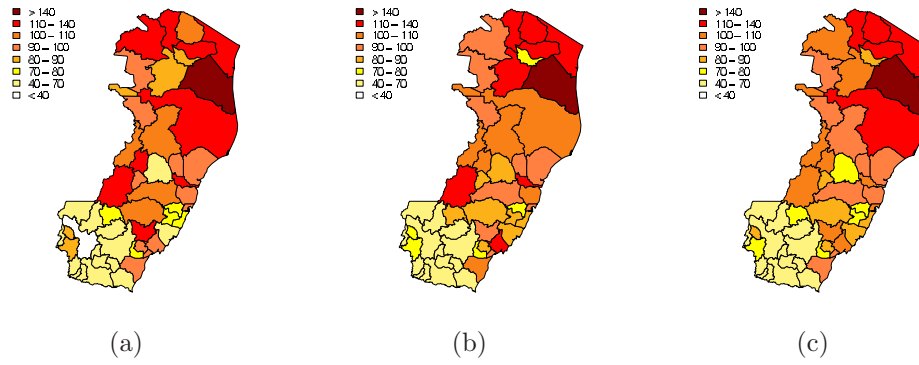


Figura 5.4: Valores de $y_{2,k,t}$ observados (a), $\mu_{2,k,t}$ verdadeiros (b) e suavizados (c) por $\mathbf{a}_{j,k,T}(-r)$. Tempo $t = 8$.

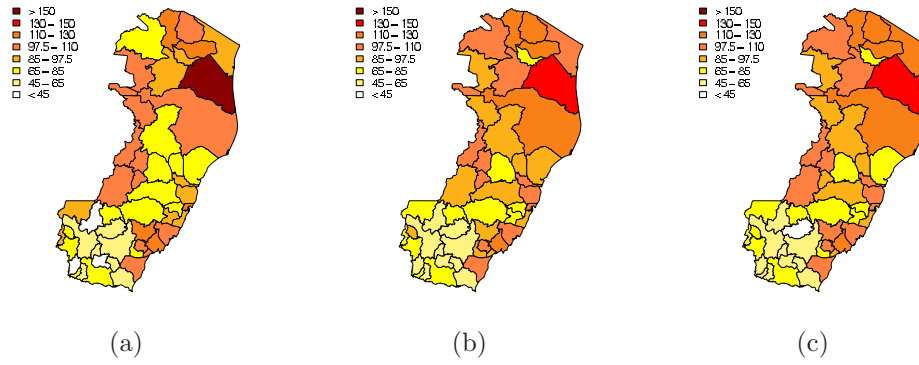


Figura 5.5: Valores de $y_{2,k,t}$ observados (a), $\mu_{2,k,t}$ verdadeiros (b) e suavizados (c) por $\mathbf{a}_{j,k,T}(-r)$. Tempo $t = 15$.

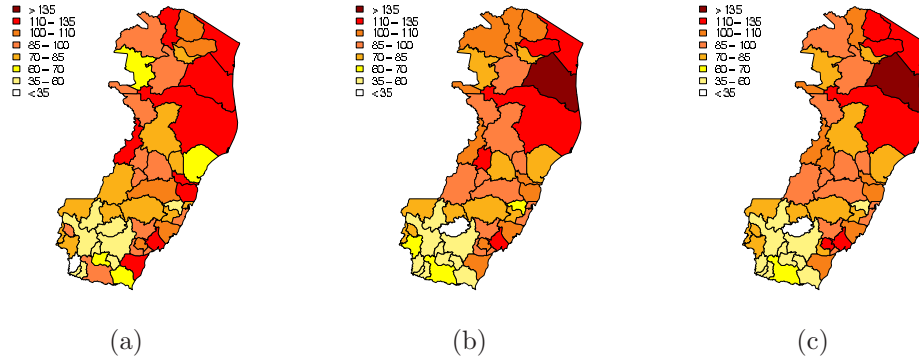


Figura 5.6: Valores de $y_{2,k,t}$ observados (a), $\mu_{2,k,t}$ verdadeiros (b) e suavizados (c) por $\mathbf{a}_{j,k,T}(-r)$. Tempo $t = 22$.

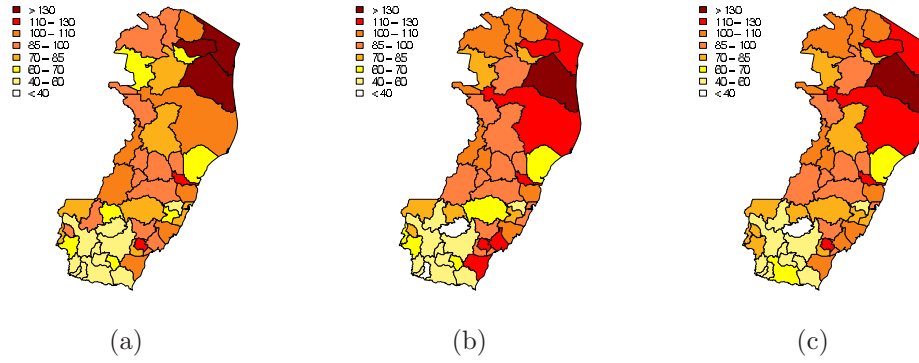


Figura 5.7: Valores de $y_{2,k,t}$ observados (a), $\mu_{2,k,t}$ verdadeiros (b) e suavizados (c) por $\mathbf{a}_{j,k,T}(-r)$. Tempo $t = 25$.

5.7.2 Variância de Sistema Desconhecida

Nessa subseção, supomos desconhecidos os fatores de proporcionalidade ψ_k e $\psi_{j,k}$ e imputamos uma distribuição à priori $GI(3, 0.3)$ para os mesmos. Geramos 5.000 amostras, sendo descartado o primeiro 1/3. São mostrados resultados a posteriori para $\psi_{1,k}$ (ver tabela 5.1) e para a média a posteriori do fator latente para 6 áreas escolhidas ao acaso (ver figura 5.8). Os resultados aqui mostram que as estimativas para o fator $\psi_{1,k}$ e para a média do fator latente estão próximas dos verdadeiros valores. Ressalta-se que, por se tratar de comparações de valores verdadeiros com valores simulados, não realizamos testes de convergência das cadeias geradas.

Tabela 5.1: Estatísticas a posteriori para $\psi_{1,k}$

Macro-região	Média	Percentis (2.5, 97.5)	$\psi_{1,k}$ verdadeiro
k=1	0.06	(0.03,0.10)	0.1
k=2	0.10	(0.04,0.24)	0.1
k=3	0.10	(0.05,0.21)	0.1
k=4	0.09	(0.03,0.23)	0.1
k=5	0.08	(0.03,0.16)	0.1
k=6	0.13	(0.04,0.35)	0.1
k=7	0.09	(0.04,0.25)	0.1
k=8	0.08	(0.03,0.16)	0.1
k=9	0.08	(0.03,0.19)	0.1
k=10	0.10	(0.04,0.24)	0.1
k=11	0.07	(0.03,0.12)	0.1
k=12	0.07	(0.03,0.13)	0.1

5.8 Aplicação aos Dados de PIB *per capita*

Inicia-se aqui uma primeira aplicação a dados reais do modelo proposto, no caso, dados do PIB *per capita* do estado do Espírito Santo para os anos de 1970, 1980, 1985, 1996, 1999, 2000, 2001, 2002 e 2003. Para diminuir o efeito de assimetria, foi aplicada a transformação logarítmica dos dados. Os valores estão em reais com base na data de 1º de agosto de 2000.

Nas figuras 5.9 e 5.10 fazemos uma análise exploratória dos dados. Conclui-se que, de um modo geral, o PIB *per capita* municipal apresenta um crescimento significativo a partir de 1970, com uma queda suave ao final do período de estudo. Ressalta-se que os municípios que apresentaram valores extremos são Vitória, Serra, Aracruz e Anchieta.

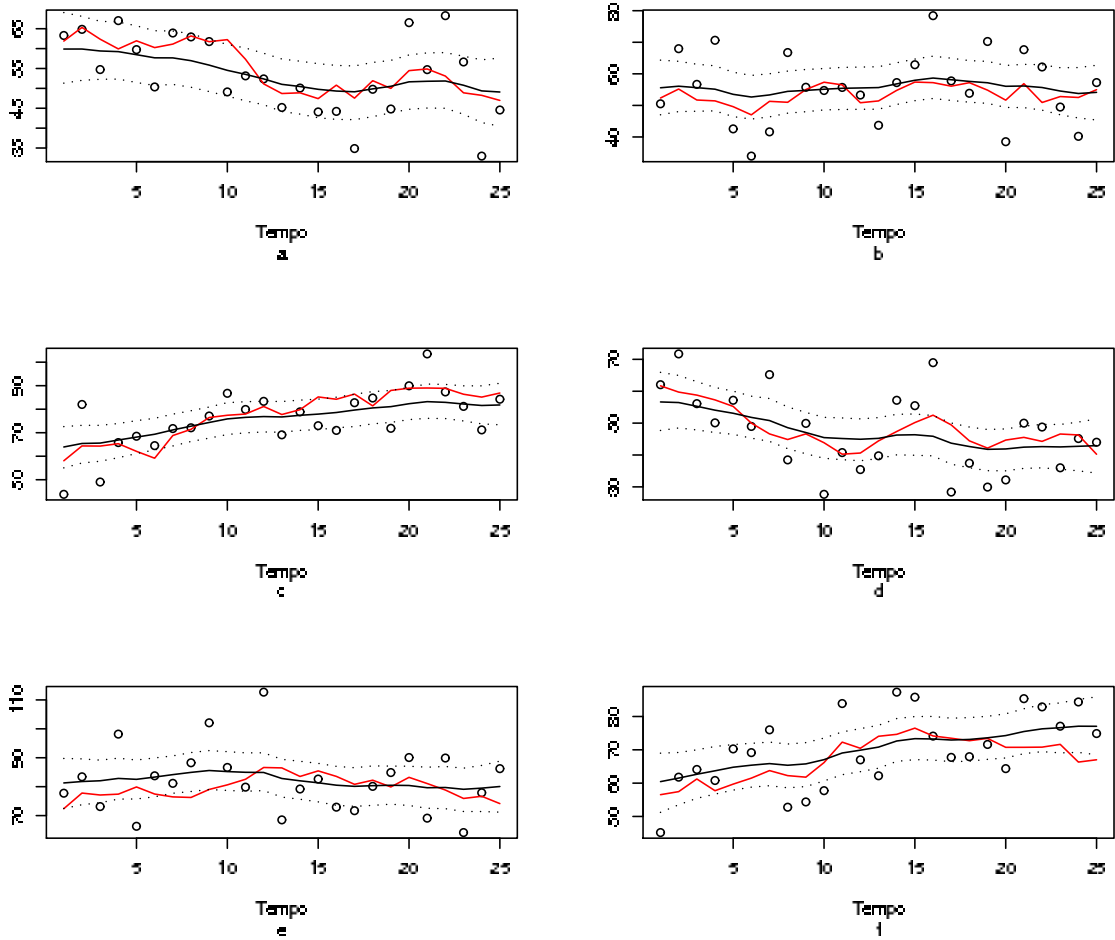


Figura 5.8: Resultados comparativos de $y_{2,k}$, (círculos), verdadeiro valor de $\mu_{2,k}$, (vermelho) e média a posteriori para $\mu_{2,k}$, (preto) para 6 áreas fictícias, ao longo do tempo. Linhas tracejadas inferior e superior indicam os percentis 2.5 e 97.5 de $\hat{\mu}_{2,k}$, respectivamente.

No processo de simulação foram geradas 8.000 amostras, sendo descartadas as primeiras 2.000. A priori utilizada para ψ_k e $\psi_{j,k}$ foi uma $GI(4, 0.15)$ e para σ^2 uma $GI(0, 0)$, sendo proporcional a $(\sigma^2)^{-1}$. As distribuições a priori para inicializar o FK são idênticas às da seção anterior. As tabelas 5.2, 5.3 e 5.4 mostram algumas medidas a posteriori para os fatores de proporcionalidade ψ_k , $\psi_{0,k}$ e $\psi_{1,k}$. Verifica-se que as médias a posteriori são em geral menores do que 1, sendo coerentes com aquilo

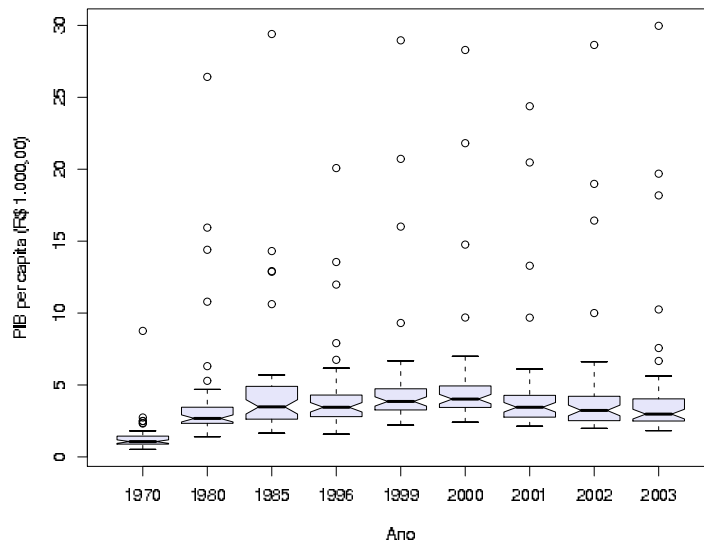


Figura 5.9: Análise Exploratória por períodos do PIB *per capita* municipal do E.S.
 Fonte: IBGE/IPEA.

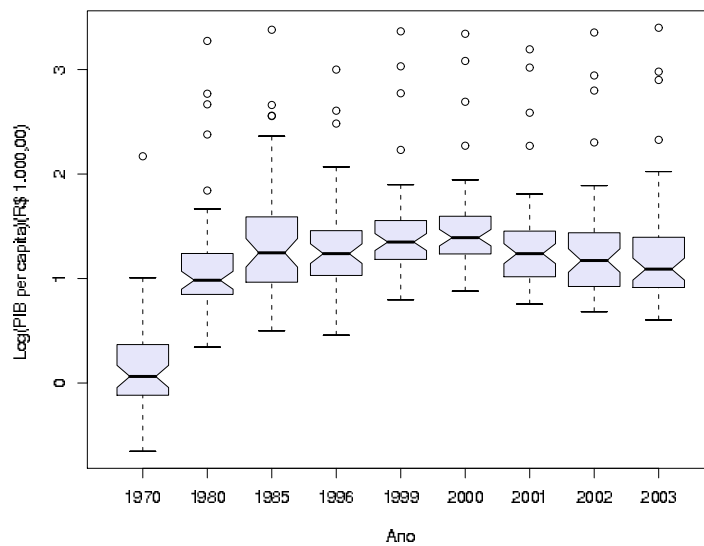


Figura 5.10: Análise Exploratória por períodos do Log(PIB *per capita*) municipal do E.S. Fonte: IBGE/IPEA.

que se espera para esses fatores. Uma exceção ocorre para as médias a posteriori de ψ_k (ver tabela 5.2). Para este, concluímos haver pouca informação nos dados que permita estimá-lo de forma satisfatória.

A figura 5.11 apresenta a trajetória e o histograma para os valores a posteriori para σ^2 . Observe que o traço indica convergência das cadeias. O histograma mostra uma distribuição unimodal, indicando um bom ajuste para esse parâmetro.

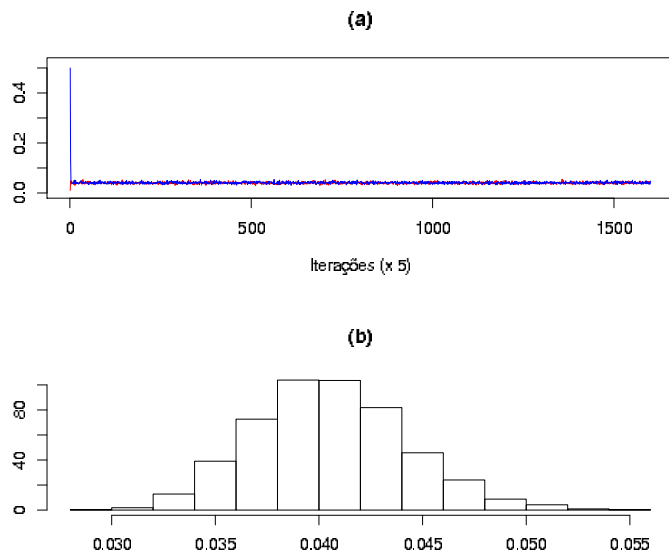


Figura 5.11: Trajetória (a) e histograma (b) para valores de σ^2 gerados a posteriori. Estatística de diagnóstico de Gelman e Rubin igual a 1.02 e de Geweke iguais a -0.12 e 0.41, respectivamente.

Tabela 5.2: Estatísticas a posteriori para ψ_k

Macro-região	Média	Percentis (2.5,97.5)
k=1	44.9	(20.1,93.2)
k=2	9.4	(4.7,19.2)
k=3	11.2	(4.6,24.7)
k=4	23.1	(10.3,48.1)

A figura 5.12 mostra a trajetória e o histograma da média a posteriori do fator latente para $t = 1$, para 1 município escolhido ao acaso em cada macro-região. Concluimos que há convergência das cadeias geradas e as distribuições a posteriori são concentradas e unimodais, indicando bom ajuste.

Tabela 5.3: Estatísticas a posteriori para $\psi_{0,k}$

Macro-região	Média	Percentis (2.5,97.5)
k=1	0.71	(0.36,1.23)
k=2	0.24	(0.06,0.65)
k=3	0.24	(0.08,0.53)
k=4	0.21	(0.02,0.86)

Tabela 5.4: Estatísticas a posteriori para $\psi_{1,k}$

Micro-região	Média	Percentis (2.5,97.5)
k=1	0.09	(0.05,0.16)
k=2	0.32	(0.09,0.81)
k=3	0.10	(0.51,1.66)
k=4	0.09	(0.07,0.68)
k=5	0.08	(0.03,0.97)
k=6	0.13	(0.06,0.81)
k=7	0.22	(0.04,0.65)
k=8	0.17	(0.06,0.42)
k=9	0.41	(0.14,0.92)
k=10	0.26	(0.06,0.69)
k=11	0.60	(0.38,0.92)
k=12	0.55	(0.30,0.94)

As figuras 5.13, 5.14 e 5.15 apresentam os resultados comparativos obtidos para a variável observável e para a média a posteriori do fator latente para, respectivamente, as 4 macro-regiões, para 4 municípios aleatoriamente escolhidos e para as micro-regiões 4 e 5. Suas análises permitem concluir pelo bom ajuste temporal do modelo.

As figuras 5.16 a 5.24 mostram, para espaço e tempo, os valores observados e respectivas estimativas a posteriori para a média dos fatores latentes. A partir das mesmas, concluímos que o modelo se presta bem ao ajuste das séries estudadas. Visualizamos também um movimento de concentração da renda *per capita* em direção ao entorno da região metropolitana de Vitória.

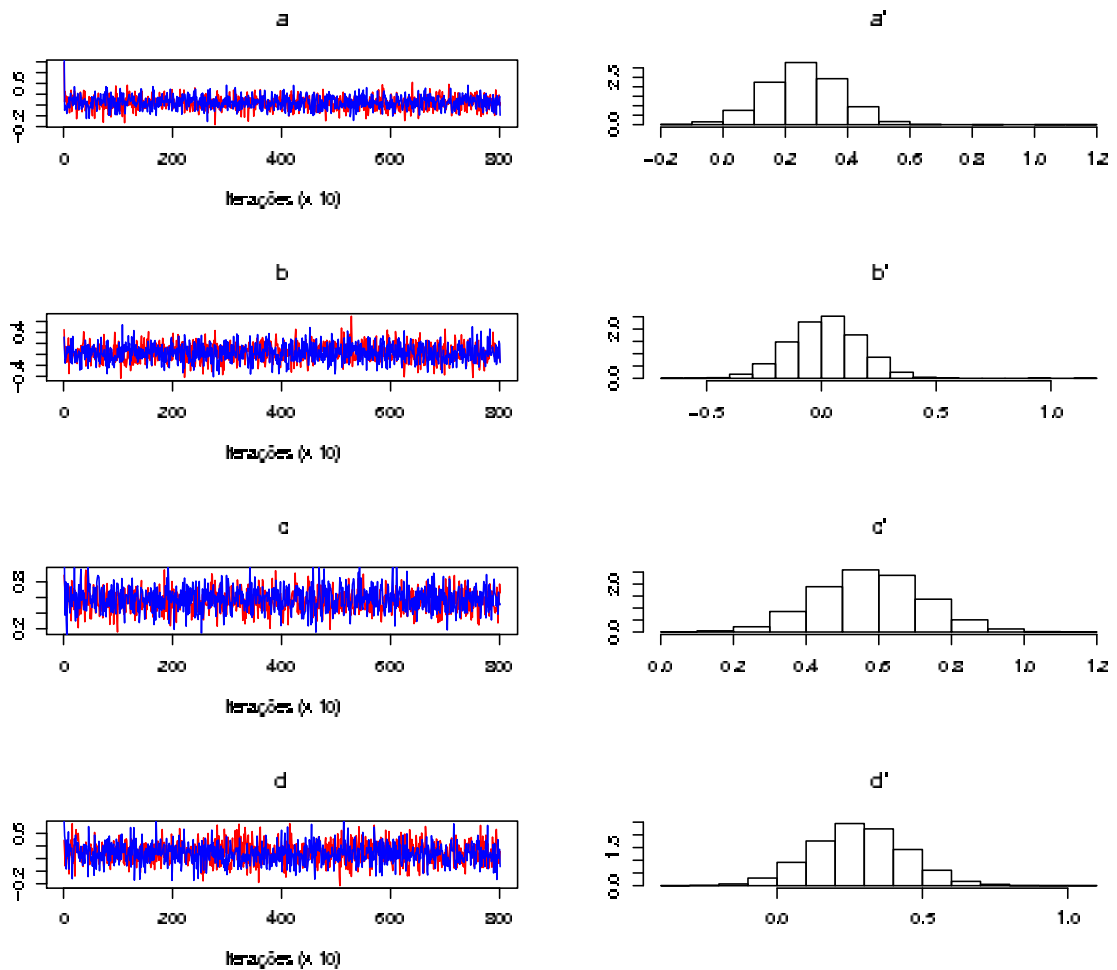


Figura 5.12: Respectivamente, trajetória e histograma para $\mu_{2,k,1}$ a posteriori nos municípios de Guarapari (a e a'), Conceição da Barra (b e b'), Colatina (c e c') e Castelo (d e d'). Estatísticas de diagnóstico de Gelman iguais a 1, 1.01, 1.01 e 1.01, respectivamente.

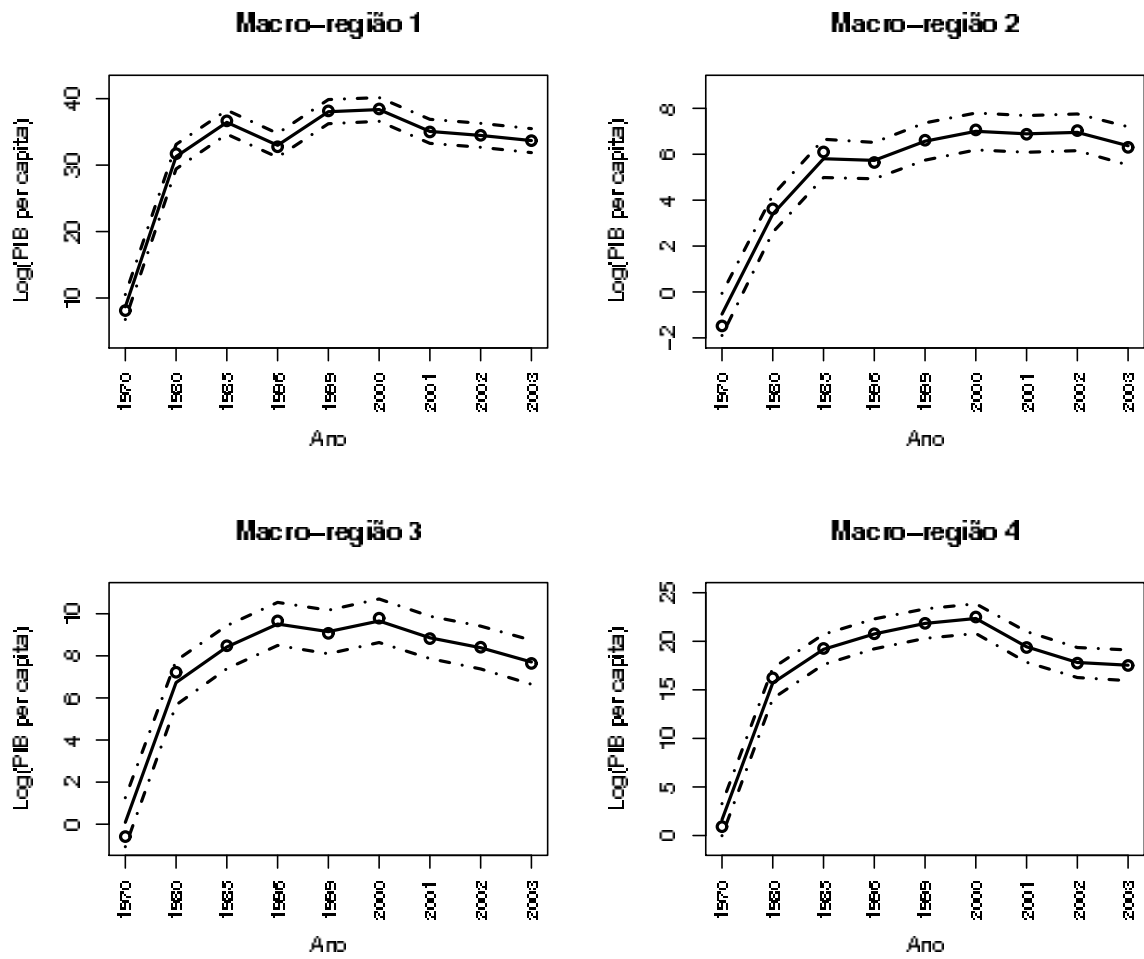


Figura 5.13: Resultados comparativos de $y_{0,k,\cdot}$ (círculos) e média a posteriori de $\mu_{0,k,\cdot}$ (preto) por macro-região. Linhas tracejadas inferior e posterior indicam os percentis 2.5 e 97.5 de $\hat{\mu}_{0,k,\cdot}$, respectivamente.

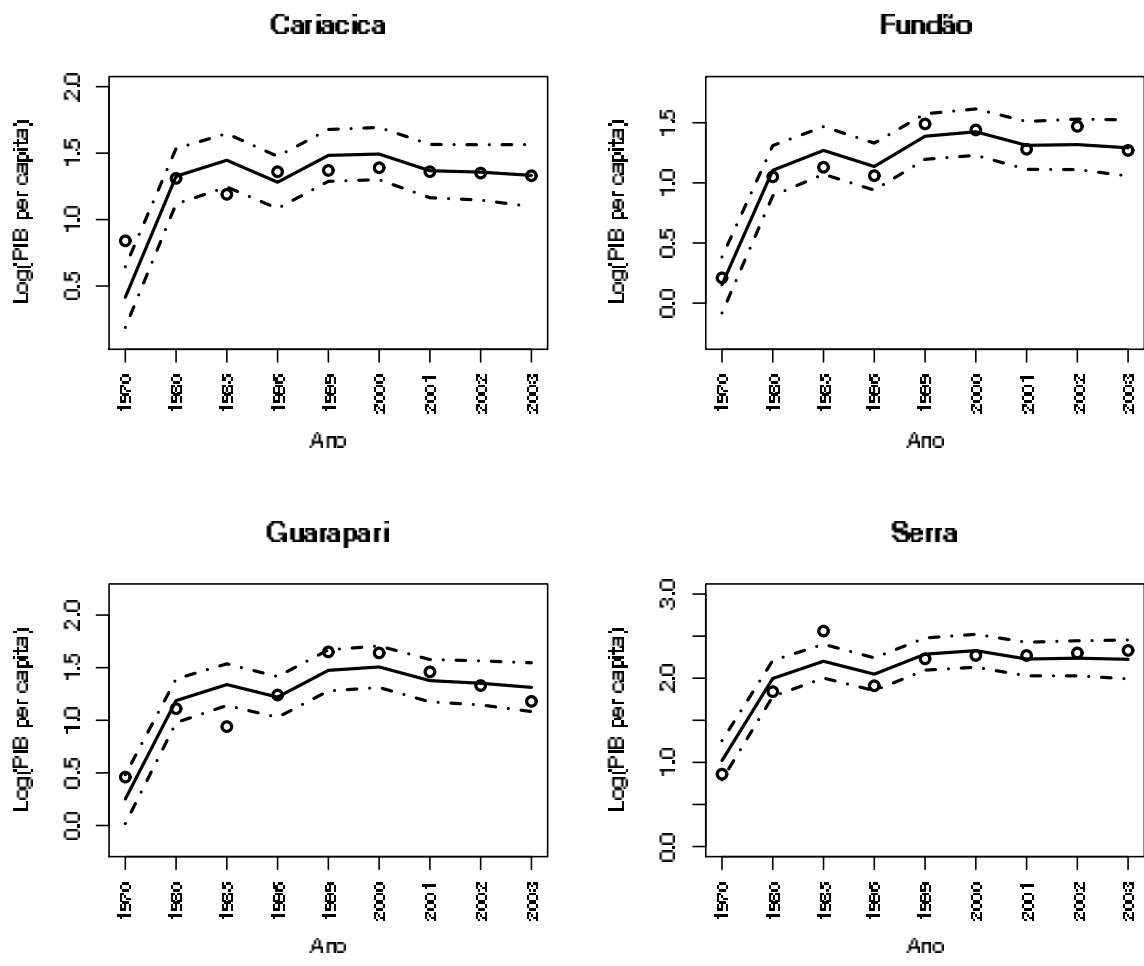


Figura 5.14: Resultados comparativos de $y_{2,k,t}$ (círculos) e média a posteriori de $\mu_{2,k,t}$ (preto) para alguns municípios da região metropolitana de Vitória. Linhas tracejadas inferior e posterior indicam os percentis 2.5 e 97.5 de $\hat{\mu}_{2,k,t}$, respectivamente.

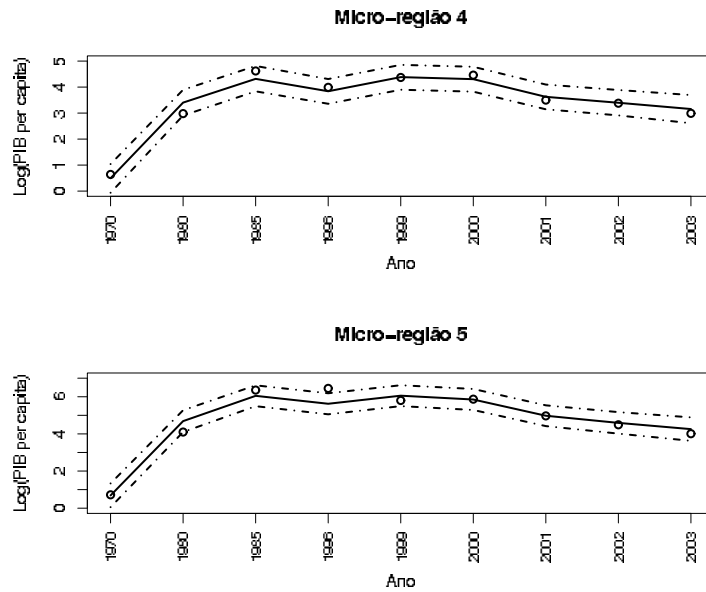


Figura 5.15: Resultados comparativos de $y_{1,k,.}$ (círculos) e média a posteriori de $\mu_{1,k,.}$ (preto) para as micro-regiões 4 e 5. Linhas tracejadas inferior e posterior indicam os percentis 2.5 e 97.5 de $\hat{\mu}_{1,k,.}$, respectivamente.

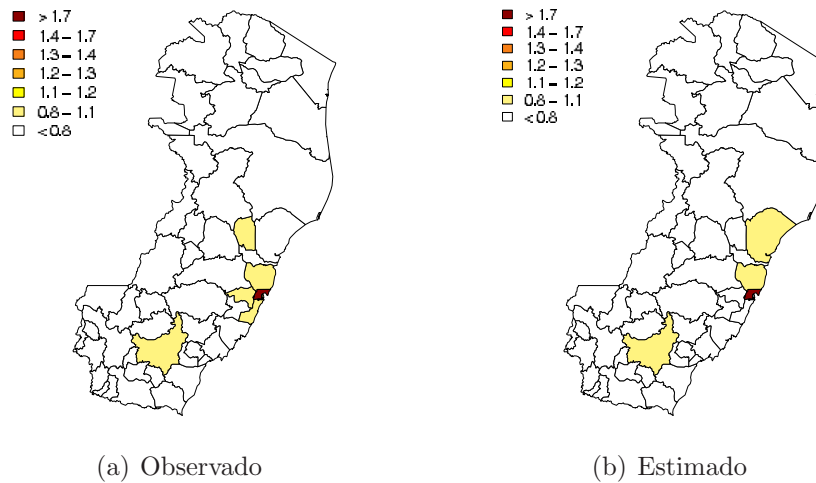


Figura 5.16: Log(PIB per capita) do E.S. - 1970

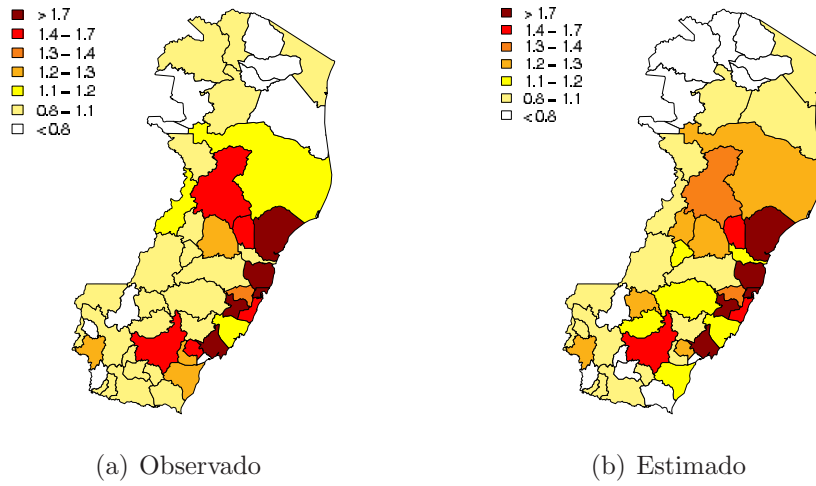


Figura 5.17: Log(PIB per capita) do E.S. - 1980

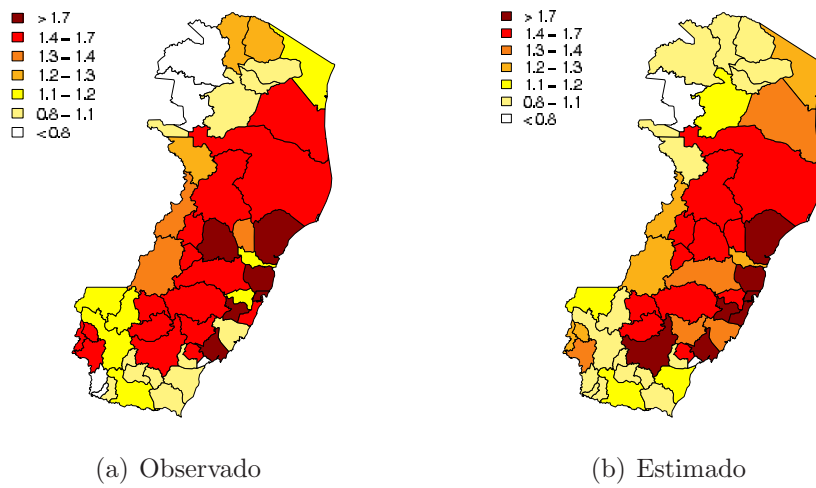


Figura 5.18: Log(PIB per capita) do E.S. - 1985

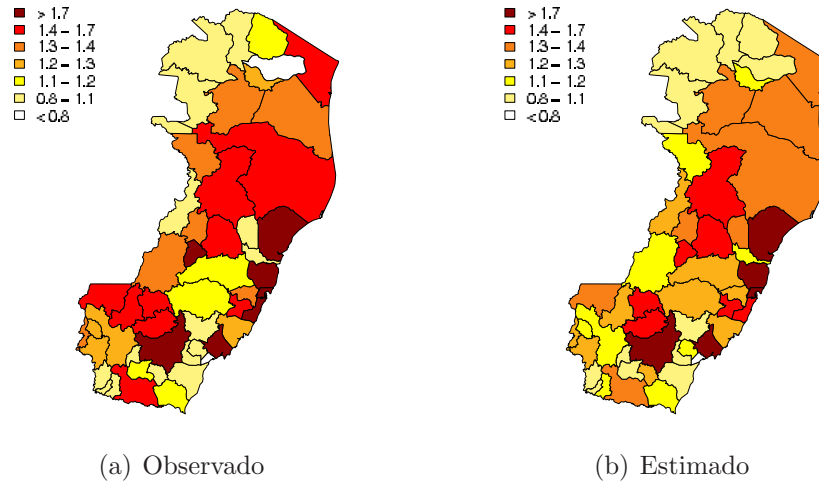


Figura 5.19: Log(PIB per capita) do E.S. - 1996

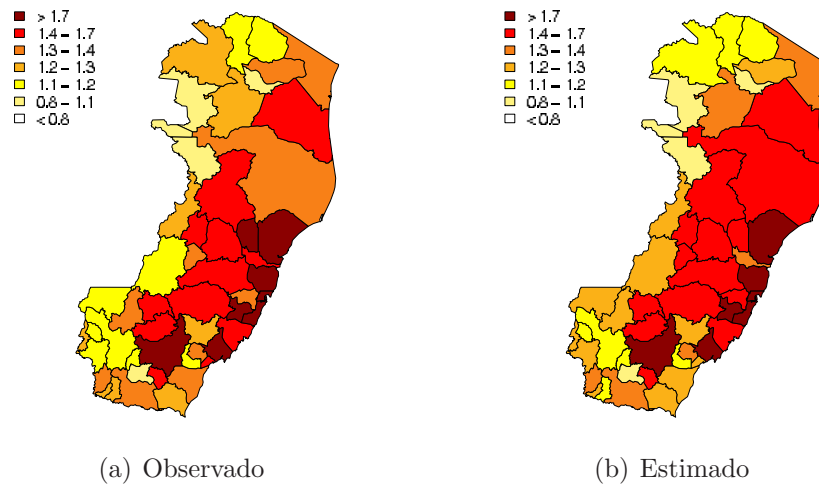


Figura 5.20: Log(PIB per capita) do E.S. - 1999

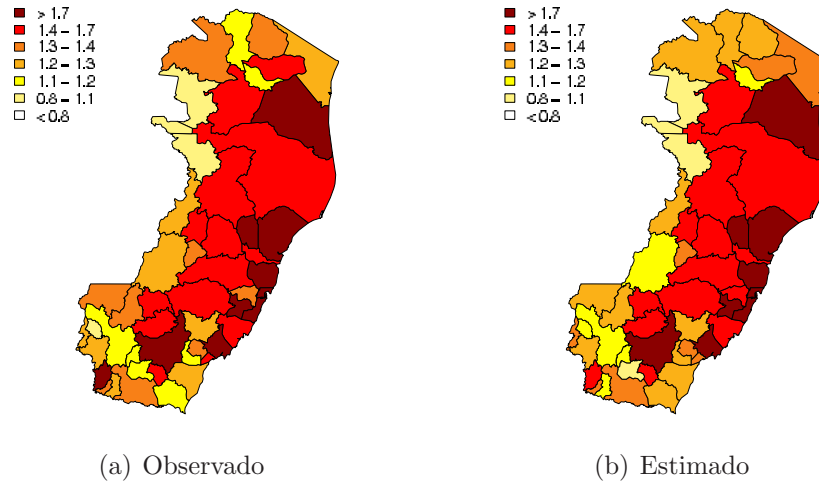


Figura 5.21: Log(PIB per capita) do E.S. - 2000

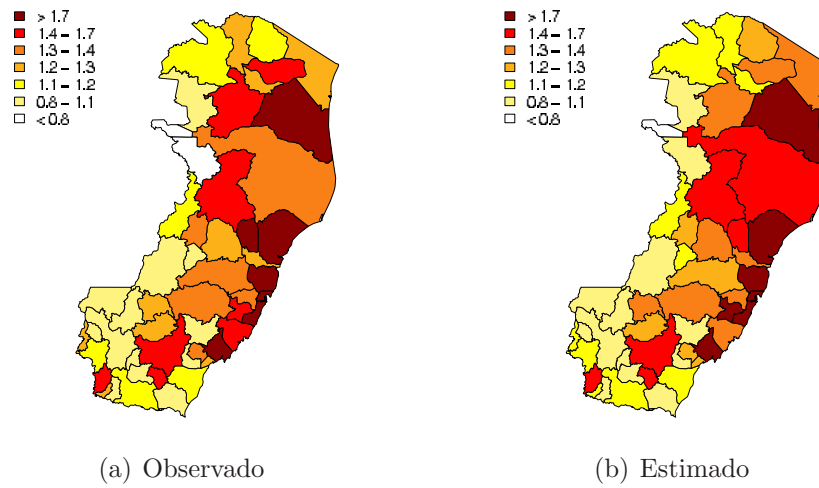


Figura 5.22: Log(PIB per capita) do E.S. - 2001

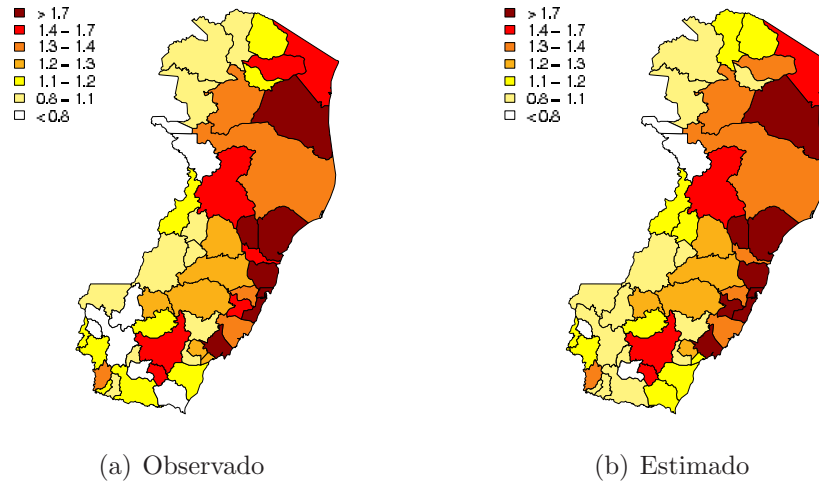


Figura 5.23: Log(PIB per capita) do E.S. - 2002

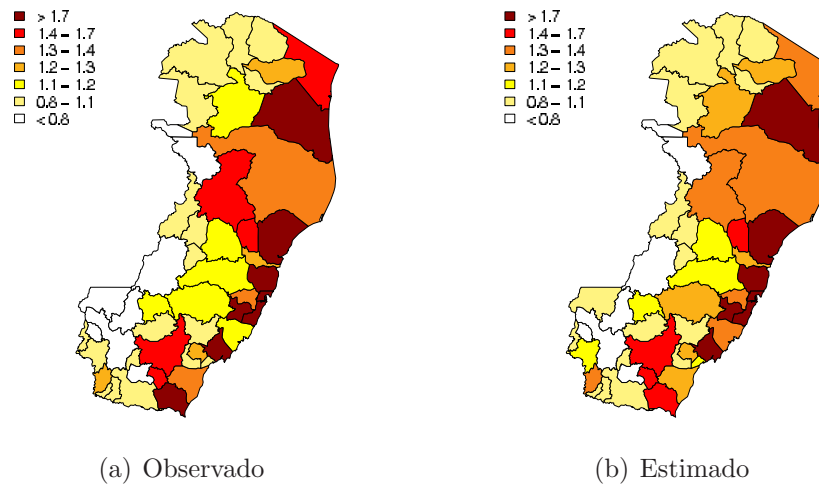


Figura 5.24: Log(PIB per capita) do E.S. - 2003

Capítulo 6

Modelagem Espaço-Temporal Multi-Escala com Variâncias Observacionais Distintas

Neste capítulo propomos um modelo espaço-temporal multi-escala considerando variâncias observacionais distintas para as áreas do nível mais desagregado. Essa distinção se dá pelo uso de pesos nas variâncias que podem assumir valores diferentes entre as áreas.

6.1 Construção do Modelo

Considere novamente o vetor aleatório $\mathbf{y}_{J-1,t} = (y_{J-1,1,t}, \dots, y_{J-1,n_{J-1},t})'$ como um processo espaço-temporal multi-escala tal que $y_{J-1,k,t}$ seja formado por $N_{J-1,k}$ componentes (pesos) constantes no tempo para todo k , com $k = 1, \dots, n_{J-1}$. Sejam então os pesos $N_{J-1,1}, \dots, N_{J-1,n_{J-1}}$ associados aos respectivos elementos do nível $J-1$. Tais pesos podem ser a população, a área plantada, o número de consumidores e etc.. No caso aqui, iremos considerá-los conhecidos.

Defina a variável aleatória $y_{J-1,k,l}^*$ como sendo a observação do l -ésimo compo-

nente para o k -ésimo elemento do nível $J - 1$, de tal forma que $y_{J-1,k,l}^* | \mu_{J-1,k,l}^*, \sigma^2 \sim N(\mu_{J-1,k,l}^*, \sigma^2)$, onde $l = 1, \dots, N_{J-1,k}$, com σ^2 desconhecido. Daí que $y_{J-1,k} = \sum_{l=1}^{N_{J-1,k}} y_{J-1,k,l}^* | \mu_{J-1,k}, \sigma^2 \sim N(\mu_{J-1,k}, N_{J-1,k} \sigma^2)$, onde $\mu_{J-1,k} = N_{J-1,k} \mu_{J-1,k,l}^*$. Para o tempo t , tem-se que

$$y_{J-1,k,t} | \mu_{J-1,k,t}, \sigma^2 \sim N(\mu_{J-1,k,t}, N_{J-1,k} \sigma^2).$$

Por consequência, o vetor de observações $(y_{J-1,1,t}, \dots, y_{J-1,n_{J-1},t})'$ segue a distribuição $N(\boldsymbol{\mu}_{J-1,t}, \mathbf{V}_{J-1})$, onde $\mathbf{V}_{J-1} = \sigma^2 \text{diag}(N_{J-1,1}, \dots, N_{J-1,n_{J-1}})$. Considerando

$$\sigma_{j,k}^2 = \sigma^2 \sum_{(j+1,k') \in D_{j,k}^*} N_{j+1,k'}$$

e

$$N_{j,k} = \sum_{(j+1,k') \in D_{j,k}} N_{j+1,k'}$$

podemos escrever:

i) $\mathbf{y}_{0,t} | \boldsymbol{\mu}_{0,t}, \mathbf{V}_0 \sim N(\boldsymbol{\mu}_{0,t}, \mathbf{V}_0)$, onde $\mathbf{V}_0 = \sigma^2 \text{diag}(N_{0,1}, \dots, N_{0,n_0})$.

ii) $\mathbf{y}_{D_{j,k}^*} | y_{j,k,t}, \boldsymbol{\theta}_{j,k,t}, \boldsymbol{\Omega}_{j,k} \sim N(\boldsymbol{\nu}_{j,k} y_{j,k,t} + \boldsymbol{\theta}_{j,k,t}, \boldsymbol{\Omega}_{j,k})$, onde

$$\boldsymbol{\nu}_{j,k} = \frac{\mathbf{N}'_{D_{j,k}^*}}{\sum_{(j+1,k') \in D_{j,k}} N_{j+1,k'}} \quad (6.1)$$

e

$$\boldsymbol{\Omega}_{j,k} = \sigma^2 \left[\text{diag}(\mathbf{N}'_{D_{j,k}^*}) - \frac{\mathbf{N}'_{D_{j,k}^*} \mathbf{N}_{D_{j,k}^*}}{\sum_{(j+1,k') \in D_{j,k}} N_{j+1,k'}} \right], \quad (6.2)$$

com $\mathbf{N}'_{D_{j,k}^*}$ sendo o vetor de pesos dos descendentes de $B_{j,k}$ excluindo um deles. Considerando ainda as variâncias de sistema da forma $W_{0,k} = \psi_k N_{0,k} \sigma^2$ e $\mathbf{W}_{j,k} = \psi_{j,k} \boldsymbol{\Omega}_{j,k}$, descrevemos o Modelo Espaço-Temporal Multi-Escala com Variâncias Observacionais Distintas pelas seguintes equações:

Equação de Observação:

$$\mathbf{y}_{J-1,t} = \boldsymbol{\mu}_{J-1,t} + \mathbf{v}_{J-1}, \quad \mathbf{v}_{J-1} \mid \mathbf{V}_{J-1} \sim N(\mathbf{0}, \mathbf{V}_{J-1}),$$

Decomposição Multi-escala da Equação de Observação:

$$y_{0,k,t} = \mu_{0,k,t} + v_{0,k}, \quad v_{0,k} \mid \sigma^2 \sim N(0, N_{0,k} \sigma^2).$$

$$\boldsymbol{\theta}_{j,k,t}^e = \boldsymbol{\theta}_{j,k,t} + \mathbf{V}_{j,k}, \quad \mathbf{V}_{j,k} \mid \boldsymbol{\Omega}_{j,k} \sim N(\mathbf{0}, \boldsymbol{\Omega}_{j,k}).$$

Equações de Sistema:

$$\mu_{0,k,t} = \mu_{0,k,t-1} + w_{0,k}, \quad w_{0,k} \mid W_{0,k} \sim N(0, \psi_k N_{0,k} \sigma^2).$$

$$\boldsymbol{\theta}_{j,k,t} = \boldsymbol{\theta}_{j,k,t-1} + \mathbf{w}_{j,k}, \quad \mathbf{w}_{j,k} \mid \mathbf{W}_{j,k} \sim N(\mathbf{0}, \psi_{j,k} \boldsymbol{\Omega}_{j,k}).$$

6.2 Densidade Conjunta do Modelo

Fazendo uso análogo do Teorema 1, verifica-se que a distribuição a posteriori conjunta $f(\boldsymbol{\mu}_0, \boldsymbol{\theta}, \psi_k, \psi_{j,k} \mid \sigma^2, \mathbf{y}_{J-1})$ fatora-se no produtório das distribuições a posteriori de $(\mu_{0,k}, \psi_k)$ e de $(\boldsymbol{\theta}_{j,k}, \psi_{j,k})$ para todo k e j , sendo, portanto, condicionalmente independentes dado σ^2 . Com efeito, os procedimentos de inferência dessas variáveis serão análogos aos do capítulo anterior.

6.3 Distribuições a Priori

Novamente se faz necessário o uso de distribuições a priori para σ^2 , ψ_k e $\psi_{j,k}$. As respectivas distribuições a priori são dadas por:

$$\sigma^2 | \delta, \lambda \sim GI \left(\frac{\delta}{2}, \frac{\lambda}{2} \right).$$

Dados τ e κ , as distribuições a priori para ψ_k e $\psi_{j,k}$ serão identicamente expressas por:

$$GI \left(\frac{\tau}{2}, \frac{\kappa}{2} \right).$$

Finalmente, as distribuições a priori para $\mu_{0,j,0}$ e $\theta_{j,k,0}$ são expressas tal como no capítulo anterior.

6.4 Simulação e Estimação de $\mu_{0,k,t}$

A modelagem de $\mu_{0,k,t}$ será realizada com uso das seguintes equações:

Equação de Observação:

$$y_{0,k,t} = \mu_{0,k,t} + v_{0,k}, \quad v_{0,k} | \sigma^2 \sim N(0, N_{0,k} \sigma^2)$$

Equação de Sistema:

$$\mu_{0,k,t} = \mu_{0,k,t-1} + w_{0,k}, \quad w_{0,k} | \psi_k, \sigma^2 \sim N(0, \psi_k N_{0,k} \sigma^2).$$

A inferência para $\mu_{0,k,t}$, ψ_k e $\psi_{j,k}$ será realizada através da técnica de MCMC.

6.4.1 Distribuições a Posteriori

A distribuição conjunta a posteriori de $\mu_{0,k,1:T}$ e ψ_k pode ser expressa por:

$$f(\mu_{0,k,1:T}, \psi_k | \sigma^2, D_T) \propto \prod_{t=1}^T f(\mathbf{y}_{0,k,t} | \mu_{0,k,t}, \sigma^2) f(\mu_{0,k,t} | \mu_{0,k,t-1}, \psi_k, \sigma^2) f(\psi_k).$$

Analogamente ao capítulo anterior, a simulação de $\mu_{0,k,1:T}$ é realizada via FFBS, enquanto que a simulação de ψ_k é realizada a partir de sua condicional completa, dada por:

Condicional Completa de ψ_k

$$\psi_k | \mu_{0,k,1:T}, \sigma^2, D_T \sim GI\left(\frac{\tau_k^*}{2}, \frac{\kappa_k^*}{2}\right),$$

onde $\tau_k^* = T + \tau_k$ e $\kappa_k^* = \kappa_k + \frac{1}{\sigma^2 N_{0,k}} \sum_{t=1}^T (\mu_{0,k,t} - \mu_{0,k,t-1})^2$.

6.5 Simulação e Estimação de $\theta_{j,k,t}$

Suponha novamente conhecidos todos os hiperparâmetros envolvidos. Tem-se então as seguintes equações do modelo:

Equação de Observação:

$$\theta_{j,k,t}^e = \theta_{j,k,t} + \mathbf{v}_{j,k}, \quad \mathbf{v}_{j,k} | \Omega_{j,k} \sim N(\mathbf{0}, \Omega_{j,k})$$

Equação de Sistema:

$$\theta_{j,k,t} = \theta_{j,k,t-1} + \mathbf{w}_{j,k}, \quad \mathbf{w}_{j,k} | \psi_{j,k} \sim N(\mathbf{0}, \psi_{j,k} \Omega_{j,k}).$$

6.5.1 Distribuições a Posteriori

A distribuição conjunta a posteriori de $\theta_{j,k,1:T}$ e $\psi_{j,k}$ dado σ^2 pode ser expressa como:

$$f(\theta_{j,k,1:T}, \psi_{j,k} | \sigma^2, D_T) \propto \prod_{t=1}^T f(\mathbf{y}_{j,k,t} | \theta_{j,k,t}, \sigma^2) f(\theta_{j,k,t} | \theta_{j,k,t-1}, \psi_{j,k}, \sigma^2) f(\psi_{j,k})$$

Assim como para o nível $j = 0$, os procedimentos de simulação de $\theta_{j,k,1:T}$ foram realizados a partir do uso do FFBS. Para o caso de $\psi_{j,k}$, realizou-se as simulações a partir da condicional completa dada por:

Condicional Completa de $\psi_{j,k}$

$$\psi_{j,k} | \boldsymbol{\theta}_{j,k,1:T}, \sigma^2, D_T \sim GI(\tau_{j,k}^*, \kappa_{j,k}^*),$$

onde $\tau_{j,k}^* = n_{D_{j,k}^*} T + \tau_{j,k}$ e $\kappa_{j,k}^* = \kappa_{j,k} + \sum_{t=1}^T (\boldsymbol{\theta}_{j,k,t} - \boldsymbol{\theta}_{j,k,t-1}) \boldsymbol{\Omega}_{j,k}^{-1} (\boldsymbol{\theta}_{j,k,t} - \boldsymbol{\theta}_{j,k,t-1})'$.

6.5.2 Simulação e Estimação de σ^2

Finalmente, para que toda a inferência do processo de modelagem seja concluída, atualizamos o valor de σ^2 a partir da distribuição dada por:

$$\sigma^2 | \boldsymbol{\mu}_{J-1,t}, D_T \sim GI\left(\frac{\delta^*}{2}, \frac{\lambda^*}{2}\right),$$

onde $\delta^* = n_{J-1} T + \delta$ e $\lambda^* = \lambda + \sum_{t=1}^T \left[\sum_{k=1}^{n_{J-1}} \frac{1}{N_{J-1,k}} (y_{J-1,k,t} - \mu_{J-1,k,t})^2 \right]$.

6.6 Aplicação aos Dados de Produção Agrícola

Realizamos uma aplicação do modelo com os valores de produção agrícola municipal relativos ao estado do Espírito Santo. São dados anuais da produção da agricultura permanente compreendido no período de 1990 a 2005, com valores em milhares de reais deflacionados pelo IGP-M (Índice Geral de Preços de Mercado), ano base 2005 (disponível em <http://www.fgvdados.fgv.br>). Tomamos a área (em hectare) média municipal plantada no período, como peso. Vale ressaltar que pelo fato de o município de Vitória apresentar valor de produção apenas para o início da série aqui utilizada (1990 e 1991), o mesmo foi excluído da análise. Os resultados que seguem foram obtidos após a geração de duas cadeias com 14.000 amostras cada, descartando o primeiro 1/4. As priori's utilizadas são idênticas às do capítulo anterior.

A figura 6.1 mostra a evolução do logaritmo dos valores municipais de produção agrícola para o estado do E.S.. Verificamos uma pequena oscilação no nível entre 1993 e 1994, tendo um suave crescimento em tendência a partir de 2001.

A figura 6.2 apresenta a trajetória e o histograma para os valores a posteriori para σ^2 , que indicam convergência das cadeias e uma distribuição a posteriori bem definida.

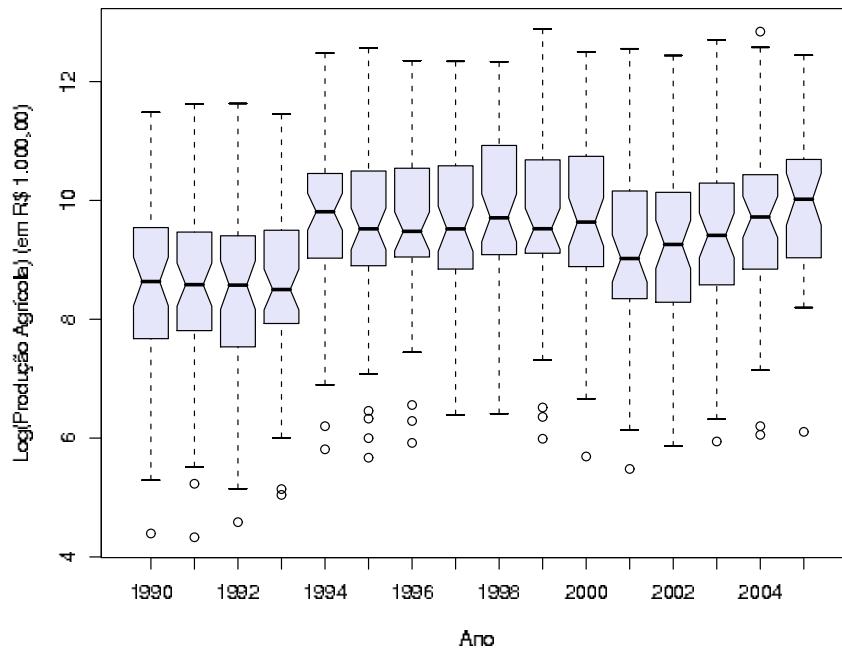


Figura 6.1: Análise Exploratória do logaritmo do valor municipal da produção agrícola para o estado do E.S. Fonte: IBGE/IPEA.

De forma semelhante aos resultados obtidos na aplicação ao PIB *per capita*, as estatísticas a posteriori para os fatores de proporcionalidade ψ_k , $\psi_{0,k}$ e $\psi_{1,k}$ são menores do que 1 (ver tabelas 6.1 a 6.3), exceto novamente para o nível de macro-regiões.

Nas figuras 6.3 e 6.4 vemos os resultados comparativos obtidos para os dados de produção e para a média a posteriori do fator latente para as 4 macro-regiões

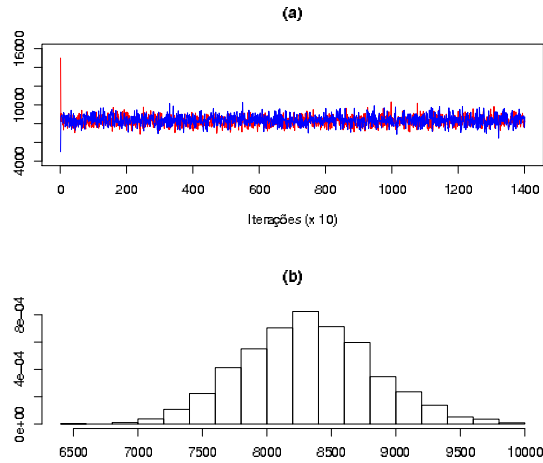


Figura 6.2: Trajetória (a) e histograma (b) para valores de σ^2 gerados a posteriori. Estatística de diagnóstico de Gelman e Rubin igual a 1.02 e de Geweke iguais a 1.98 e 0.42, respectivamente.

Tabela 6.1: Estatísticas a posteriori para ψ_k

Macro-região	Média	Percentis (2.5,97.5)
k=1	5.7	(2.41,11.67)
k=2	6.0	(2.82,11.87)
k=3	0.3	(0.07,0.73)
k=4	7.0	(3.38,13.57)

Tabela 6.2: Estatísticas a posteriori para $\psi_{0,k}$

Macro-região	Média	Percentis (2.5,97.5)
k=1	0.26	(0.07,0.59)
k=2	1.32	(0.40,3.13)
k=3	0.17	(0.04,0.43)
k=4	0.22	(0.04,0.62)

e para 4 municípios escolhidos ao acaso, respectivamente. Concluímos pelo bom ajuste temporal do modelo.

Tabela 6.3: Estatísticas a posteriori para $\psi_{1,k}$

Micro-região	Média	Percentis (2.5,97.5)
k=1	0.050	(0.01,0.11)
k=2	0.053	(0.01,0.13)
k=3	0.051	(0.01,0.11)
k=4	0.056	(0.02,0.14)
k=5	0.053	(0.01,0.12)
k=6	0.057	(0.02,0.15)
k=7	0.180	(0.04,0.47)
k=8	0.050	(0.01,0.13)
k=8	0.050	(0.01,0.13)
k=9	0.050	(0.02,0.12)
k=10	0.057	(0.01,0.16)
k=11	0.051	(0.02,0.11)
k=12	0.052	(0.01,0.11)

A figura 6.5 mostra a trajetória e histograma da média a posteriori do fator latente para o segundo instante de tempo, com municípios escolhidos ao acaso em cada macro-região. Os resultados indicam a convergência das cadeias.

Os mapas das figuras 6.6 a 6.13 mostram, para espaço e tempo, os valores observados e as respectivas estimativas a posteriori para a média dos fatores latentes para os dados de valor de produção agrícola municipal. Os resultados indicam um ajuste satisfatório do modelo aos dados em questão. Note que o processo latente estimado considera não somente as observações para o tempo t em questão, mas também outros instantes de tempo e municípios em estudo.

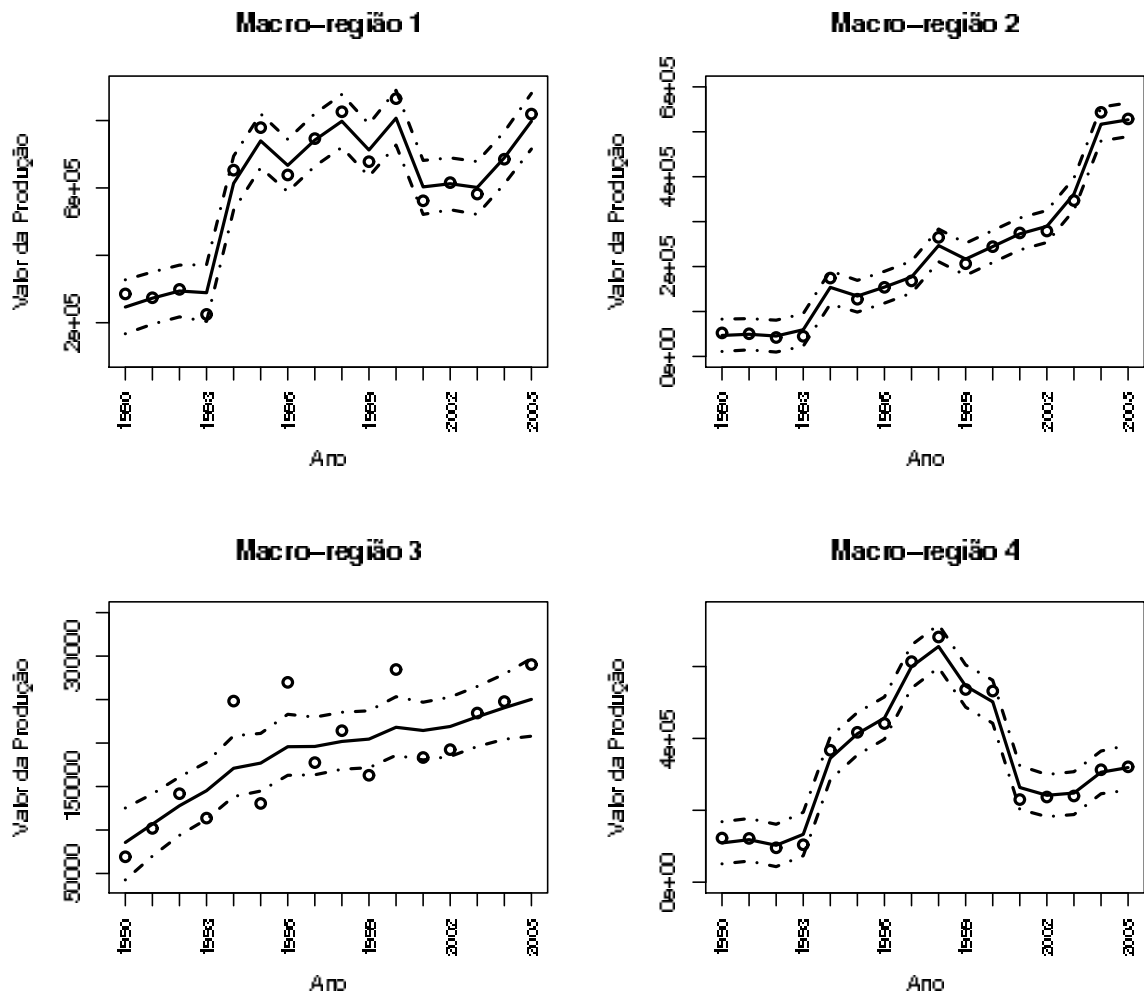


Figura 6.3: Resultados comparativos de $y_{0,k,}$ (círculos) e média a posteriori de $\mu_{0,k,}$ (preto) por macro-região. Linhas tracejadas inferior e posterior indicam os percentis 2.5 e 97.5 de $\hat{\mu}_{0,k,}$, respectivamente.

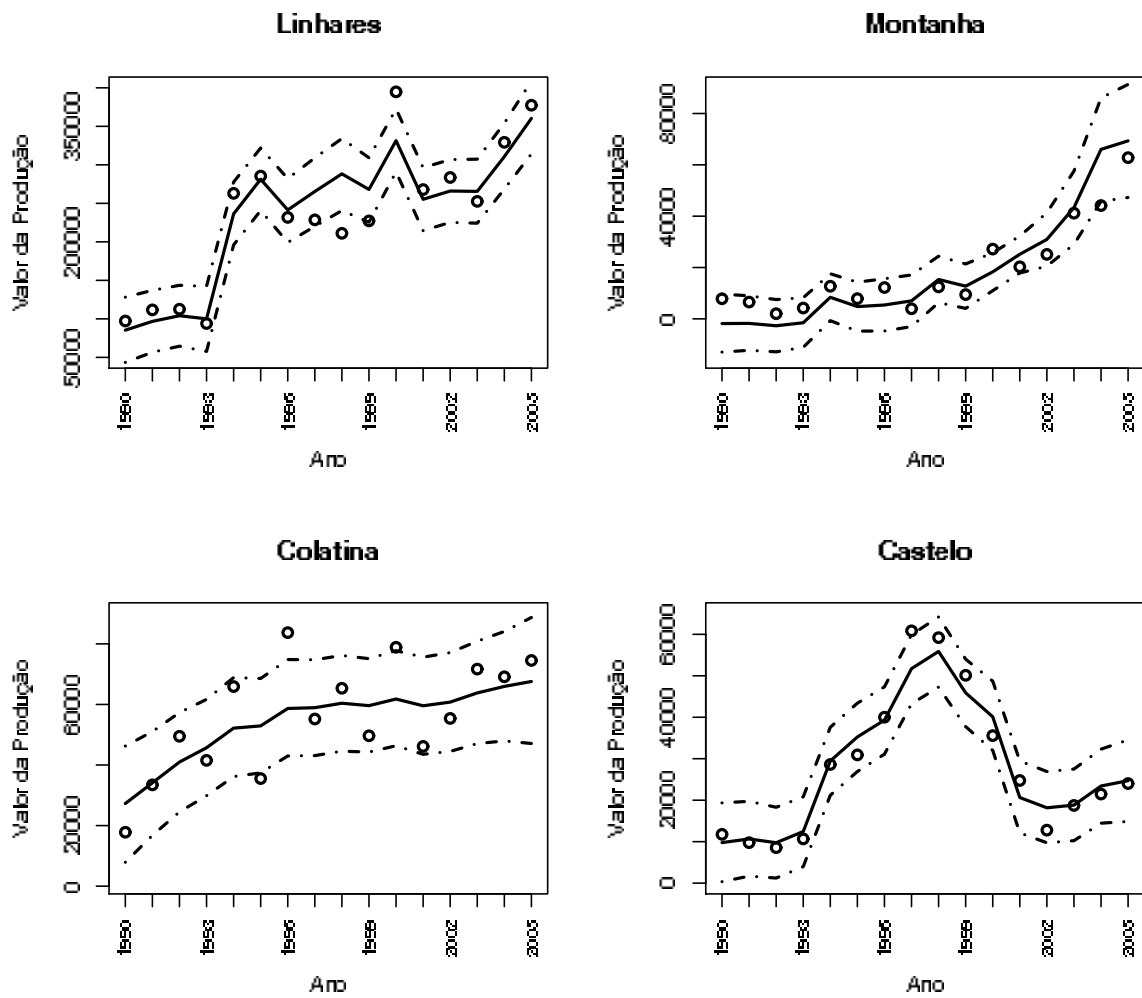


Figura 6.4: Resultados comparativos de $y_{2,k,}$ (círculos) e média a posteriori de $\mu_{2,k,}$ (preto) por município. Linhas tracejadas inferior e posterior indicam os percentis 2.5 e 97.5 de $\hat{\mu}_{2,k,}$, respectivamente.

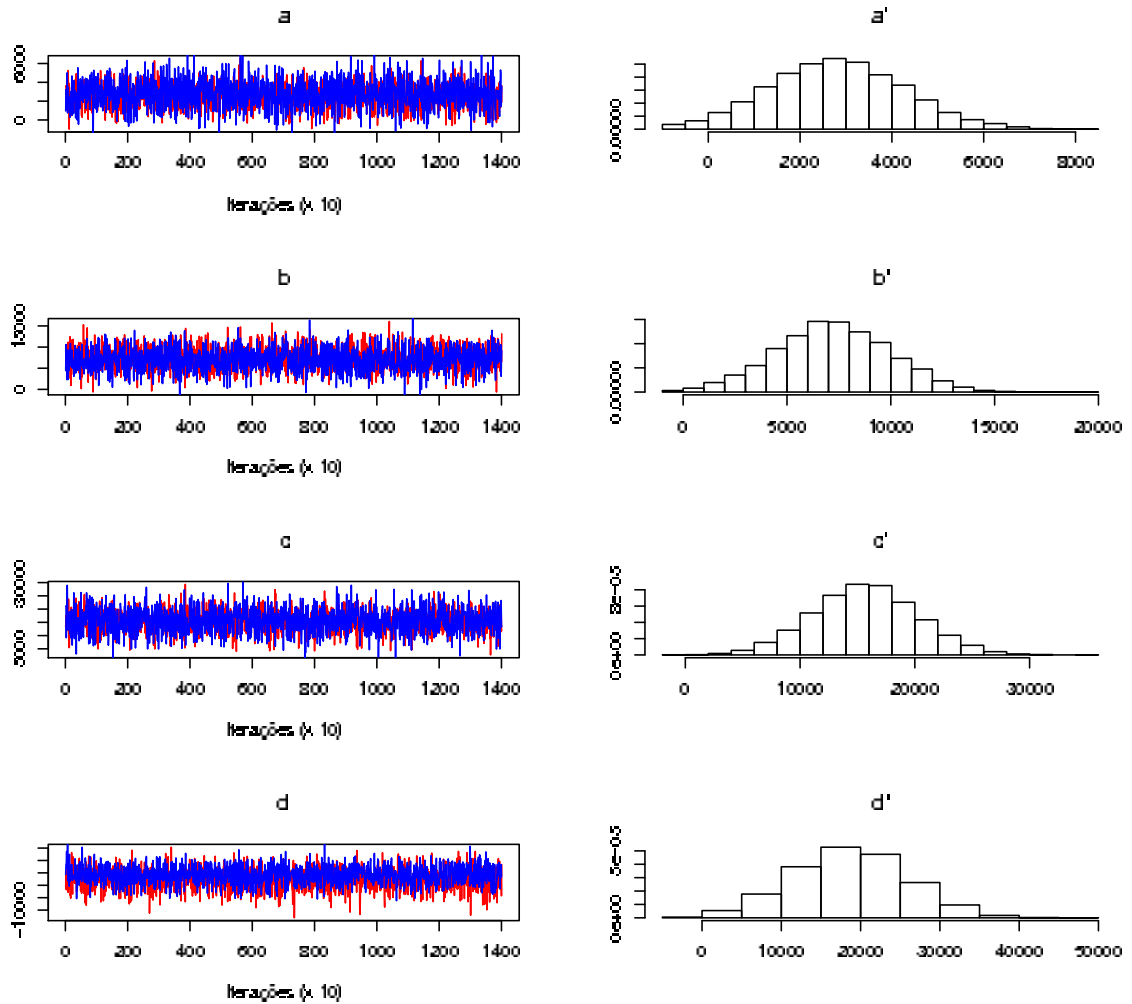


Figura 6.5: Respectivamente, trajetória e histograma para $\mu_{2,k,2}$ a posteriori nos municípios de Cariacica (a e a'), Guarapari (b e b'), Santa Leopoldina (c e c') e Nova Venécia (d e d'). Estatísticas de diagnóstico de Gelman iguais a 1.05, 1, 1.01 e 1.09, respectivamente.

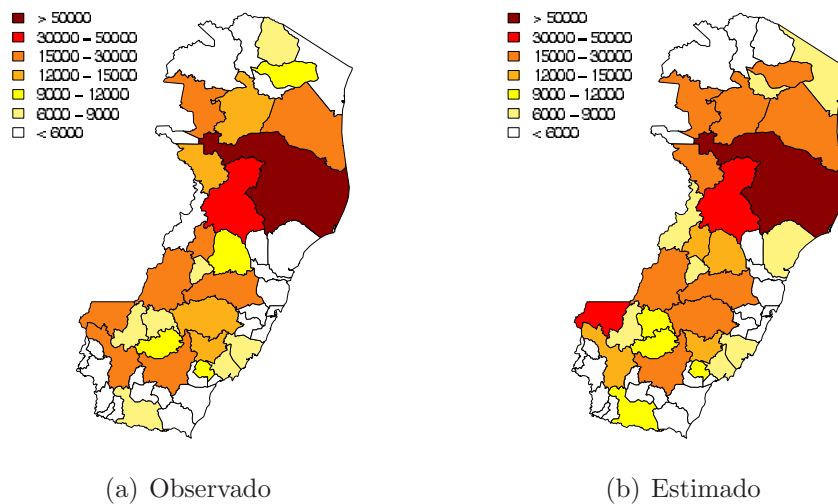


Figura 6.6: Valor de produção agrícola do E.S. (em milhares de reais) - 1991

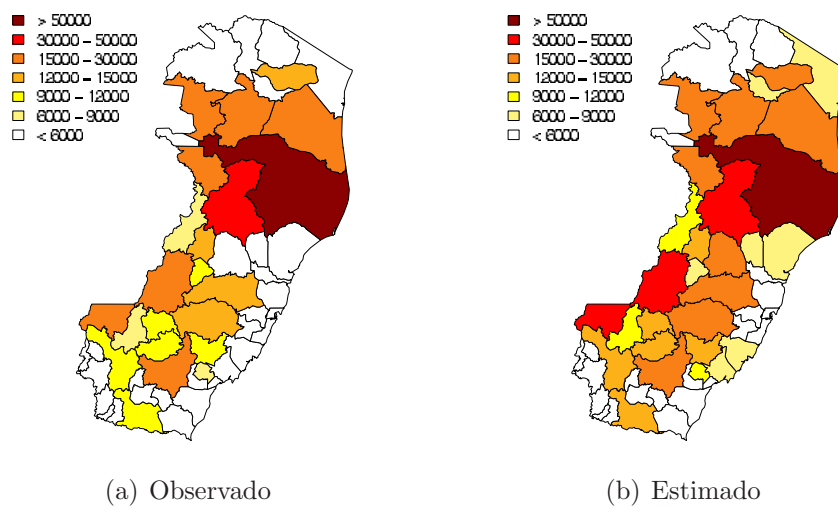


Figura 6.7: Valor de produção agrícola do E.S. (em milhares de reais) - 1993

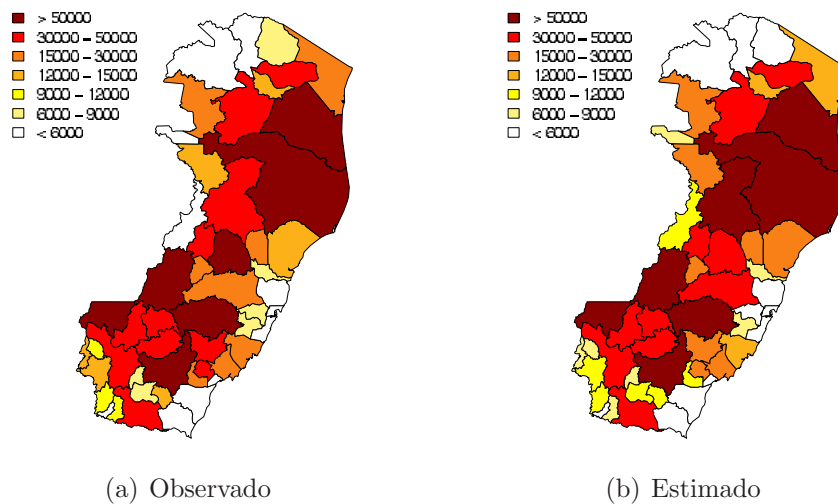


Figura 6.8: Valor de produção agrícola do E.S. (em milhares de reais) - 1995

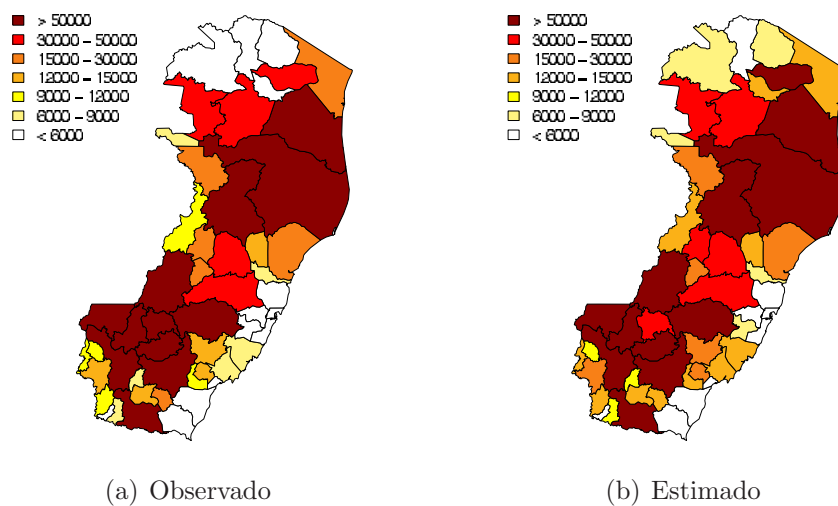


Figura 6.9: Valor de produção agrícola do E.S. (em milhares de reais) - 1997

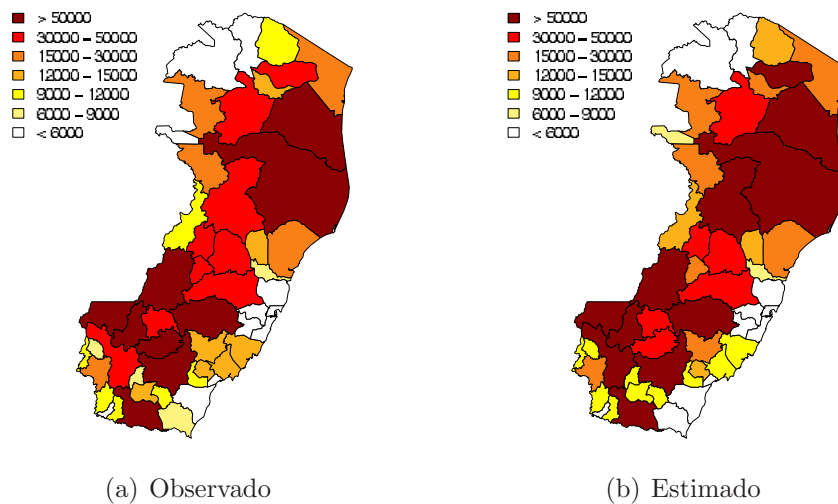


Figura 6.10: Valor de produção agrícola do E.S. (em milhares de reais) - 1999

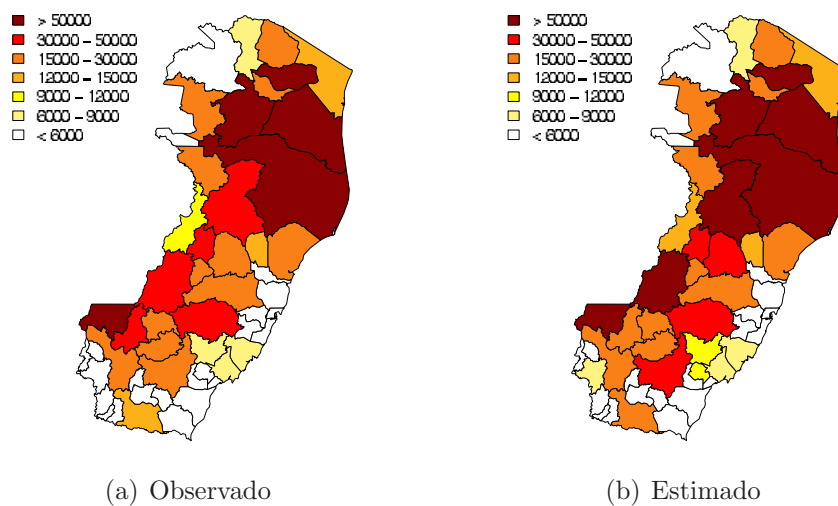


Figura 6.11: Valor de produção agrícola do E.S. (em milhares de reais) - 2001

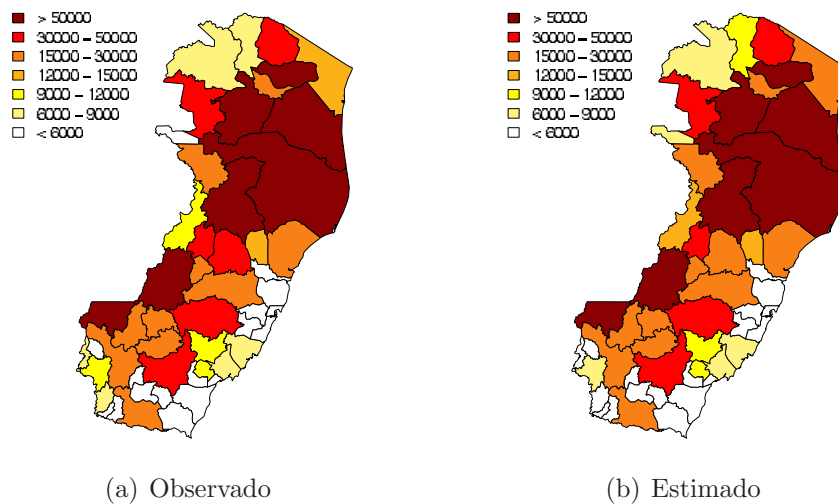


Figura 6.12: Valor de produção agrícola do E.S. (em milhares de reais) - 2003

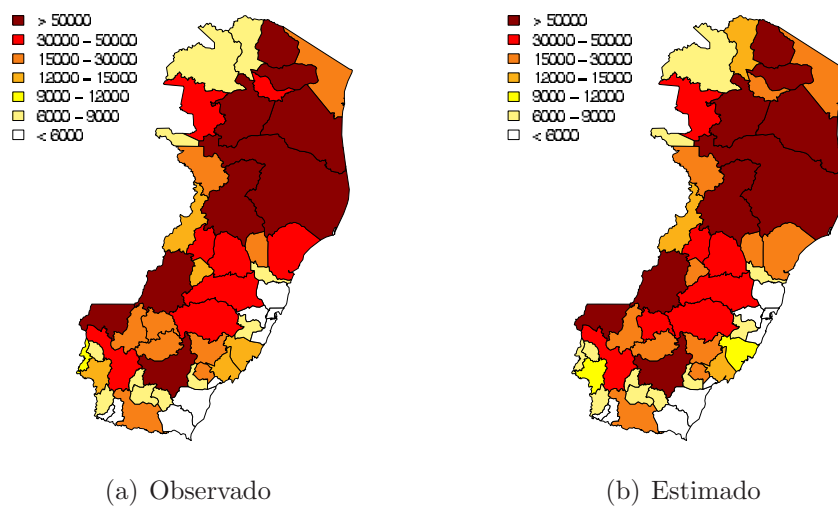


Figura 6.13: Valor de produção agrícola do E.S. (em milhares de reais) - 2005

Capítulo 7

Conclusões

Neste trabalho propomos um modelo espaço-temporal multi-escala para análise de dados de área. Fazemos uso da metodologia de construção multi-escala desenvolvida por Kolaczyk e Huang (2001), onde decomposmos o processo espacial a cada tempo e modelamos os coeficientes multi-escala no tempo através de modelos dinâmicos.

Os resultados obtidos nas aplicações a dados reais indicam que a modelagem proposta consegue captar os movimentos espaço-temporais do fator latente, estimando-os eficientemente e contribuindo para um melhor entendimento da evolução espaço-temporal do fenômeno em estudo.

Acreditamos que nos níveis mais grosseiros da agregação há pouca informação presente nos dados capaz de influenciar os resultados a posteriori para os parâmetros ψ_k , $\psi_{0,k}$ e $\psi_{1,k}$. Porém, o fato de o ajuste ter sido satisfatório para as médias dos fatores latentes nos níveis mais desagregados reforça o uso de uma forma alternativa de modelagem na estrutura de agregação mais grosseira, ou, até mesmo, o uso de outras distribuições a priori.

Potencialmente, pode-se incluir tendência e sazonalidade no modelo. Dificuldades técnicas desta extensão serão relacionadas às restrições de igualdade das somas de descendentes com respeito a ancestrais.

Vale enfatizar que o modelo apresentado nesse texto apresenta uma vantagem significativa em relação à metodologia proposta por Kolaczyk e Huang (2001) pois, possibilita a realizações de previsões. Além disso, a suavização obtida com o nosso modelo espaço-temporal considera tanto informações espaciais quanto temporais.

Finalmente, note que a decomposição multi-escala por nós utilizada neste trabalho reduz o problema de modelagem espaço-temporal de um vetor de observações em um nível mais desagregado a um problema de modelagem temporal dos coeficientes multi-escala. Como esses coeficientes são de dimensão muito menor do que o vetor de observações no nível mais desagregado, isto leva a um grande ganho computacional.

Reforçamos, portanto, que o nosso modelo espaço-temporal multi-escala aqui proposto considera os vários níveis de escalas espaciais do processo, inter-relacionando e atualizando os coeficientes multi-escala via modelagem dinâmica.

7.1 Meta

Com base nos resultados alcançados, planejamos explorar um pouco mais o assunto, pesquisando a aplicação do modelo para outras séries temporais multivariadas, estas não necessariamente observadas no contexto espacial.

Referências Bibliográficas

- Berliner, L.M., Wikle, C.K. e Milliff, R.F. (1999). Multiresolution wavelet analyses in hierarchical Bayesian turbulence models, 341-359. *Bayesian Inference in Wavelet Based Models*. Springer-Verlag.
- Carter, C. K. e Kohn, R. (1994). On Gibbs sampling for state space models. *Biometrika*, 81, 541-553.
- Ferreira, M. A. R. e Lee, H. K. (2007). *Multiscale Modeling: A Bayesian Perspective*. Springer.
- Ferreira, M. A. R., West, M., Lee, H. K. H. e Higdon, D. (2006). Multe-scale and hidden resolution time series models. *Bayesian Analysis* 1 (4), 947-968.
- Fruwirth-Schanatter, S. (1994). Data augmentation and dynamic linear models. *Journal Time Series*, 15 (2), 183-202.
- Gamerman, D. e Lopes, H. F. (2006). *Markov Chain Monte Carlo: stochastic simulation for Bayesian inference*. (2nd. ed.). Chapman end Hall.
- Gelman, A. e Rubin, D.B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science*, 7, 457-511.
- Geman, S. e Geman, D. (1984). Stochastic relaxation, Gibbs distribution and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6, 721-741.

- Geweke, J.. *Evaluating the accuracy of sampling-based approaches to calculating posterior moments*. In Bayesian Statistics 4 (ed JM Bernardo, JO Berger, AP Dawid and AFM Smith). Clarendon Press, Oxford, UK.
- Harrison, P. J. e Stevens, C. F. (1976). Bayesian forecasting (with discussion). *J. Roy. Statist. Soc. Ser. B*, 38, 205-247.
- Johannesson, G., Cressie, N. e Huang, H.C. (2007). Dynamic multi-Resolution spatial models. *Environmental and Ecological Statistics*, 14, 5-25.
- Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. *J. Basic Engineering*(82), 344-5.
- Kolaczyk E. D. e Huang, H. (2001). Multiscale Statistical Models for Hierarchical Spatial Aggregation. *Geographical Analysis*, Vol. 33, No. 2,95-118.
- Meinhold, R. J. e Singpurwalla, N. (1983). Understanding the Kalman filter. *American Statistician*, 37, 123-127.
- Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. e Teller, E. (1953). Equation of state calculations by fast computing machines. *Journal of Chemical Physics*, 21, 1087-1092.
- Migon, H. e Gamerman, D.(1999). *Statistical Inference: an Integrated approach*. Editora Arnold, London.
- Migon, H., D. Gamerman, Lopes, H. F. e Ferreira, M. (2005). Dynamic Models. In D. Dey and C. Rao (Eds.), *Handbook fo Statistics, Volume 25: Bayesian Thinking, Modeling and Computation*, Chapter 19, pp. 553-588.
- Plummer, M., Best, N., Cowles, K. e Vines, K.. (2007) CODA: Convergence Diagnosis and Output Analysis for MCMC. Version 0.11-3. Cran R.

West, M. e Harrison, P. J. (1997). *Bayesian Forecasting and Dynamic Models* (2nd. ed.). New York: Springer.

Apêndice A

Algumas Propriedades da Distribuição Normal Multivariada

Seja uma variável aleatória Z de dimensão p normalmente distribuída com média μ e variância Σ , denotada por $N(\mu, \Sigma)$. As seguintes afirmações são verdadeiras:

- A combinação linear de quaisquer componentes de Z segue uma distribuição Normal. Ou seja, se A é uma matriz $r \times p$ e b é um vetor r -dimensional então a transformação linear $Y = (AZ + b) \sim N(A\mu + b, A\Sigma A')$
- Qualquer subconjunto de componentes de Z segue uma distribuição normal. Isto significa que se dividirmos o vetor Z em dois blocos Z_1 e Z_2 , o primeiro com k componentes e o segundo com $p - k$, podemos também

particionar μ e Σ tal como

$$\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} \text{ e } \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$$

- A distribuição condicional dos componentes são normalmente distribuídos. Em outras palavras, $Z_1|Z_2$ é normalmente distribuída, analogamente valendo para $Z_2|Z_1$. Mantendo-se as mesmas partições anteriores para Z , μ e Σ obtém-se que:

$$Z_1|Z_2 \sim N(\mu_{1.2}, \Sigma_{11.2}),$$

onde $\mu_{1.2} = \mu_1 + \Sigma_{12}\Sigma_{22}^{-1}(Z_2 - \mu_2)$ e $\Sigma_{11.2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$.

- Pode-se reconstruir a distribuição conjunta da seguinte maneira: Se $Z_1|Z_2 \sim N(\mu_1 + A_1(Z_2 - \mu_2), A_2)$ e $Z_2 \sim N(\mu_2, \Sigma_{22})$, tem-se que:

$$Z = \begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix} \sim N(\mu, \Sigma)$$

com

$$\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} \text{ e } \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$$

onde $\Sigma_{11} = A_2 + A_1\Sigma_{22}A_1'$ e $\Sigma'_{21} = \Sigma_{12} = A_1\Sigma_{22}$.

Apêndice B

Descrição Multi-Escala Para o Caso Normal

A demonstração feita aqui não considera o índice de tempo t pois o resultado é válido para qualquer tempo, e por uma questão de simplificação de notação, omitiremos tal índice na demonstração.

Seja a variável aleatória $\mathbf{y}_{J-1,t} = (y_{J-1,1,t}, \dots, y_{J-1,n_{J-1},t})'$, expressa segundo uma distribuição $N(\boldsymbol{\mu}_{J-1,t}, \mathbf{V}_{J-1,t})$. Considere $\mathbf{y}_{D_{j,k}^*}$ o vetor de medidas, para o tempo t , dos descendentes do elemento $B_{j,k}$ excluindo um deles. Seja a distribuição conjunta de $y_{j,k}$ e $\mathbf{y}_{D_{j,k}^*}$ da pela expressão abaixo:

$$\begin{pmatrix} y_{j,k} \\ \mathbf{y}_{D_{j,k}^*} \end{pmatrix} \Big| \mu_{j,k} \sim N \left[\begin{pmatrix} \mu_{j,k} \\ \boldsymbol{\mu}_{D_{j,k}^*} \end{pmatrix}, \begin{pmatrix} \sigma_{j,k}^2 & \boldsymbol{\sigma}_{D_{j,k}^*}^2 \\ (\boldsymbol{\sigma}_{D_{j,k}^*}^2)' & \mathbf{V}_{D_{j,k}^*} \end{pmatrix} \right],$$

onde $\mathbf{V}_{D_{j,k}^*}$ é a matriz de covariância de dimensão $n_{D_{j,k}^*}$.

Recorde que $y_{j,k} = \sum_{(j+1,k') \in D_{j,k}^*} y_{j+1,k'}$. Sendo assim, tomando um par-

particular descendente de $y_{j,k}$ pertencente a $\mathbf{y}_{D_{j,k}^*}$, identificado por y_{j+1,k^*} , tem-se a seguinte expressão para a covariância entre $y_{j,k}$ e y_{j+1,k^*} :

$$Cov(y_{j,k}, y_{j+1,k^*}) = Cov\left(\sum_{(j+1,k') \in D_{j,k}^*} y_{j+1,k'}, y_{j+1,k^*}\right).$$

Observe que, pelo fato de as variáveis y 's serem condicionalmente independentes dado μ , todas as covariâncias do lado direito da expressão anterior serão iguais a 0, exceto uma delas, que é exatamente a variância relativa ao particular elemento y_{j+1,k^*} . Daí conclui-se que:

$$Cov(y_{j,k}, y_{j+1,k^*}) = \sigma_{j+1,k^*}^2,$$

e portanto,

$$Cov(y_{j,k}, \mathbf{y}_{D_{j,k}^*}) = \boldsymbol{\sigma}_{D_{j,k}^*}^2.$$

Segue, dos resultados já conhecidos da distribuição normal multivariada (ver Anexo A), que

$$\left(\mathbf{y}_{D_{j,k}^*} | y_{j,k}, \boldsymbol{\theta}_{j,k}, \boldsymbol{\Omega}_{j,k}\right) \sim N(\boldsymbol{\nu}_{j,k} y_{j,k} + \boldsymbol{\theta}_{j,k}, \boldsymbol{\Omega}_{j,k}),$$

onde:

$$\begin{aligned}
E\left(\mathbf{y}_{D_{j,k}^*} | y_{j,k}, \boldsymbol{\theta}_{j,k}, \boldsymbol{\Omega}_{j,k}\right) &= \boldsymbol{\mu}_{D_{j,k}^*} + \frac{\boldsymbol{\sigma}_{D_{j,k}^*}^2}{\sigma_{j,k}^2} (y_{j,k} - \mu_{j,k}) \\
E\left(\mathbf{y}_{D_{j,k}^*} | y_{j,k}, \boldsymbol{\theta}_{j,k}, \boldsymbol{\Omega}_{j,k}\right) &= \underbrace{\frac{\boldsymbol{\sigma}_{D_{j,k}^*}^2}{\sigma_{j,k}^2} y_{j,k}}_{\boldsymbol{\nu}_{j,k}} + \underbrace{\boldsymbol{\mu}_{D_{j,k}^*} - \frac{\boldsymbol{\sigma}_{D_{j,k}^*}^2}{\sigma_{j,k}^2} y_{j,k}}_{\boldsymbol{\theta}_{j,k}} \\
E\left(\mathbf{y}_{D_{j,k}^*} | y_{j,k}, \boldsymbol{\theta}_{j,k}, \boldsymbol{\Omega}_{j,k}\right) &= \boldsymbol{\nu}_{j,k} y_{j,k} + \boldsymbol{\theta}_{j,k}
\end{aligned}$$

e

$$\begin{aligned}
V\left(\mathbf{y}_{D_{j,k}^*} | y_{j,k}, \boldsymbol{\theta}_{j,k}, \boldsymbol{\Omega}_{j,k}\right) &= \mathbf{V}_{D_{j,k}^*} - \left(\frac{\boldsymbol{\sigma}_{D_{j,k}^*}^2 (\boldsymbol{\sigma}_{D_{j,k}^*}^2)'}{\sigma_{j,k}^2}\right) \\
V\left(\mathbf{y}_{D_{j,k}^*} | y_{j,k}, \boldsymbol{\theta}_{j,k}, \boldsymbol{\Omega}_{j,k}\right) &= \boldsymbol{\Omega}_{j,k}.
\end{aligned}$$

Apêndice C

Banco de Dados

C.1 Dados Sócio-Econômicos

Nas aplicações deste texto foram utilizados dois conjuntos de dados sócio-econômicos relativos ao estado do Espírito Santo, a saber:

- Um primeiro é o PIB *per capita* municipal para os anos de 1970, 1980, 1985, 1996, 1999, 2000, 2001, 2002 e 2003, com base em valores expressos em reais de 1º de agosto de 2000. Os dados populacionais e de totais do PIB foram obtidos em <http://www.ipeadata.gov.br> e <http://www.ibge.gov.br>. Segundo o IBGE, para os anos censitários de 1970 a 1996, as estimativas do Produto Interno Bruto Municipal (PIBM) foram desagregadas por classe de atividade econômica, baseadas na classificação das Contas Nacionais do Brasil. A partir de 1999, a metodologia de cálculo do PIB dos Municípios baseia-se na distribuição pelos municípios do valor adicionado das atividades econômicas das Contas Regionais do Brasil. A partir desse procedimento, estima-se o valor adicionado VA das atividades econômicas Agropecuária, Indústria e Serviços o *dummy*

financeiro, os impostos e o PIB, medido a preço corrente por município. Essa metodologia identifica as variáveis e fontes de informação que permitem distribuir o VA estadual (valor adicionado da unidade da federação) das 15 atividades econômicas pelos respectivos municípios do estado (ver em <http://www.ibge.gov.br> para mais detalhes).

- O segundo são os valores monetários de produção agrícola municipal no período de 1990 a 2005, deflacionados pelo IGP-M ano base 2005. Nesse caso, foi excluído o município de Vitória, por se tratar de área tipicamente urbana. Os dados foram obtidos em <http://www.ibge.gov.br>, onde consta a seguinte nota metodológica: "A pesquisa Produção Agrícola Municipal - PAM investiga um elenco de mais de 60 produtos, que são divididos em produtos de lavouras temporárias e produtos de lavouras permanentes. Os dados são obtidos pela rede de coleta do IBGE, mediante consulta a entidades públicas e privadas, a produtores, técnicos e órgãos ligados direta ou indiretamente aos setores da produção, comercialização, industrialização e fiscalização de produtos agrícolas. A coleta de dados baseia-se num sistema de fontes de informação, representativo de cada município, gerenciado pelo agente de coleta do IBGE que, acionando-o periodicamente, obtém os informes e subsídios para a consolidação das estimativas finais da produção. A unidade de investigação na pesquisa Produção Agrícola Municipal é o município". Como ponderador, tomamos a área (em hectare) média municipal de agricultura permanente cultivada ao longo do período de estudo.

C.2 Estrutura de Partição

A estrutura hierárquica multi-escala do estado do E.S. é formada por 3 níveis, a saber: 52 municípios, 12 micro-regiões e 4 macro-regiões. Face aos dados

utilizados, foi tomada como referência a divisão político administrativa do ano de 1979. Os quadros a seguir apresentam a estrutura multi-escala do estado do E.S. para o ano de 1979.

Estrutura Multi-Escala do estado do Espírito Santo/1979

Macro-regiões	Micro-regiões	Municípios
1	1	Cariacica
		Fundão
		Guarapari
		Serra
		Viana
		Vila Velha
		Vitória
	2	Aracruz
		Ibiraçu
		Linhares
	3	Alfredo Chaves
		Anchieta
		Iconha
		Itapemirim
		Piúma
	4	Afonso Cláudio
		Conceição do Castelo
		Domingos Martins
	5	Itaguaçu
		Itarana
Santa Leopoldina		
Santa Tereza		

Macro-regiões	Micro-regiões	Municípios
2	6	Conceição da Barra
		São Mateus
	7	Montanha
		Mucurici
		Pinheiros
3	8	Baixo Guandu
		Colatina
		Pancas
	9	Barra de São Francisco
		Ecoporanga
		Mantenópolis
	10	Boa Esperança
		Nova Venécia
4	11	Apiacá
		Aflio Vivácqua
		Bom Jesus do Norte
		Cachoeiro de Itapemirim
		Castelo
		Jerônimo Monteiro
		Mimoso do Sul
		Muqui
		Presidente Kennedy
		Rio Novo do Sul
	12	Alegre
		Divino de São Lourenço
		Dores do Rio Preto
		Guaçuí
		Iúna
		Muniz Freire
		São José do Calçado

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](#)