

Universidade Federal do Rio de Janeiro

UMA CLASSE DE MODELOS ESPAÇO-TEMPORAIS BASEADA
EM PROCESSOS AUTOREGRESSIVOS MULTIVARIADOS

Aline Araújo Nobre

2007

Livros Grátis

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

Milhares de livros grátis para download.

UMA CLASSE DE MODELOS ESPAÇO-TEMPORAIS BASEADA EM PROCESSOS AUTOREGRESSIVOS MULTIVARIADOS

Aline Araújo Nobre

Tese de Doutorado submetida ao Programa de Pós-Graduação em Estatística do Instituto de Matemática da Universidade Federal do Rio de Janeiro, como parte dos requisitos necessários para obtenção do grau de Doutor em Ciências Estatísticas.

Orientadora: Profa. Alexandra M. Schmidt

Co-orientador: Prof. Bruno Sansó

Rio de Janeiro, 5 de julho de 2007

UMA CLASSE DE MODELOS ESPAÇO-TEMPORAIS BASEADA EM PROCESSOS AUTOREGRESSIVOS MULTIVARIADOS

Aline Araújo Nobre

Orientadora: Alexandra M. Schmidt

Co-orientador: Bruno Sansó

Tese de Doutorado submetida ao Programa de Pós-Graduação em Estatística do Instituto de Matemática da Universidade Federal do Rio de Janeiro, como parte dos requisitos necessários para obtenção do grau de Doutor em Ciências Estatísticas.

Presidente, Alexandra M. Schmidt

IM-UFRJ

Prof. Bruno Sansó

UCSC-USA

Prof. Dani Gamerman

IM-UFRJ

Prof. Helio S. Migon

IM-UFRJ

Prof. Carlos Alberto de B. Pereira

IME-USP

Prof. Valderio A. Reisen

CCE-UFES

Rio de Janeiro, 5 de Julho de 2007

”Viver e não ter a vergonha de ser feliz Cantar.. (E cantar e cantar...) A beleza de ser um eterno aprendiz Ah meu Deus! Eu sei... (Eu sei...) Que a vida devia ser bem melhor e será Mas isso não impede que eu repita É bonita, é bonita e é bonita... ”

Gonzaguinha

Agradecimentos

As diversas experiências vividas, juntamente com as diferentes pessoas que conviveram comigo durante estes 4 anos e 4 meses de doutorado me trouxeram um grande enriquecimento profissional e pessoal. Os agradecimentos são inúmeros, mas para mim é um grande prazer lembrar estes momentos agora e agradecer a cada uma destas pessoas que direta ou indiretamente me ajudaram na realização deste sonho.

Em primeiro lugar agradeço a Deus pelo dom da vida e por me dar forças para chegar até aqui. Ao meu pai que, onde quer que esteja, sempre esteve ao meu lado como um anjo da guarda.

Sem dúvida nenhuma, a pessoa que mais me ajudou e esteve presente comigo durante todo este período, me apoiando de forma incondicional e torcendo por cada vitória conquistada, foi a minha querida mãe, para quem vai meu agradecimento especial. Agradeço a minha família, em especial aos tios e tias, pelo apoio, incentivo e carinho. Ao meu padrasto, Valdir, por me lembrar que no final tudo dá certo.

A minha orientadora Alexandra pelo incentivo, confiança e oportunidade durante todo o período. Ao meu co-orientador Bruno pela atenção, experiência transmitida, pelo bom humor e pelas oportunidades oferecidas durante a minha estada na Universidade da Califórnia, Santa Cruz.

Gostaria de agradecer aos professores do departamento de Métodos Estatísticos pelo ensinamento ministrado, em especial ao professor Dani pelo incentivo e pela ajuda durante o pedido da minha bolsa *sandwich*.

Ao prof. Ulisses Confalonieri pela bolsa concedida no início do doutorado e ao CNPq pelo suporte financeiro durante o restante do período.

Ao prof. Luiz Maia do Departamento de Meteorologia da UFRJ e ao Franz Munoz da Escola Nacional de Saúde Pública da Fiocruz pela disponibilização dos dados.

Amigo é coisa para se guardar do lado esquerdo do peito, dentro do coração.

O que seria de mim sem os meus amigos...

Aos amigos Helena, Heloísa e Paulo por representarem o meu referencial de família no Rio.

Aos amigos do miolo: Cris, Leo, Thaís, Carla e Gustavo. A Cris, minha amiga-irmã, que compartilhou de pertinho os bons e os maus momentos desta longa e árdua caminhada, pelo ombro amigo

e pelas longas conversas ao telefone. Ao Leo pelo apoio computacional online e por me fazer rir com suas piadas. A Thaís por sempre trazer palavras positivas quando a gente mais precisa. A Carla e ao Gustavo pelos papos no *msn* e pelo companheirismo durante todo esse período.

A Madalena e ao Rafael por serem os responsáveis pela minha fácil adaptação ao Rio e por me apresentarem ao samba, à Lapa, etc., ainda que a aluna tenha superado os mestres. A Karina, minha inesquecível ex-companheira de casa, pelas palavras certas nas horas certas. A Penelope por me lembrar que existe vida além do doutorado. As amigas Schirley e Marlene pelo apoio nos momentos difíceis, realçando sempre o verdadeiro valor da amizade. A Heide por me fazer sentir a pessoa mais inteligente do mundo. A Gecy por provar que a distância não abala uma verdadeira amizade. Vocês não imaginam o quanto são importante para mim.

Aos colegas do doutorado, em especial ao Luís por compartilhar comigo as dificuldades do curso de Probabilidade Avançada. Aos colegas da pós graduação: Mariane, Marcelo, Valmária e Adelmo pelos congressos, pelas festas e pelas conversas no laboratório.

Durante a minha estada em Santa Cruz convivi com diversas pessoas que foram fundamentais para que a minha estada fosse bastante tranquila e acolhedora. Aos colegas da UCSC: Yuzheng, Angela e Peter pela companhia no laboratório durante os finais de semana e por tornarem o ambiente mais caloroso. A Cíntia, que compartilhou comigo as experiências e angústias de se fazer um doutorado. A Silvia, Margherita e Serginho pelos muitos momentos intensos e inesquecíveis vividos. Às amigas Vanessa e Valéria, e ao pequeno Caio, por me fazerem sentir mais próxima do Brasil. Quanta saudade de vocês.

Aos meus mais novos amigos da Fiocruz pelo apoio emocional nesta tão difícil reta final. Em especial, a Camila e ao Ernesto por suportarem meu choro e por me ouvirem pacientemente falar sobre a minha tese. Ao Oswaldo por compartilhar comigo sua experiência e me incentivar a terminar. A Marília pela compreensão. A Taynãna pela paciência em tentar achar o erro no meu programa. Finalmente, ao Cosme por me lembrar que o aumento de salário pode ser um grande incentivo para terminar o doutorado.

RESUMO

UMA CLASSE DE MODELOS ESPAÇO-TEMPORAIS BASEADA EM PROCESSOS AUTOREGRESSIVOS MULTIVARIADOS

Aline Araújo Nobre

Orientadora: Profa. Alexandra M. Schmidt

Co-orientador: Prof. Bruno Sansó

Processos espaço-temporais vêm ganhando grande popularidade nos últimos anos. Uma razão para isto é o crescimento de aplicações surgindo das ciências ambientais e de saúde. O foco principal dos modelos espaço-temporais está na especificação da função de covariância do processo e, conseqüentemente, nas hipóteses usualmente feitas. Em geral, a estrutura de covariância é modelada em função da distância entre as estações monitoradoras. É comum assumir que o processo espacial é estacionário e isotrópico. Outra suposição bastante comum em modelos espaço-temporais é assumir uma estrutura de covariância separável, isto é, a função de covariância do processo é construída pelo produto entre as covariâncias espacial e temporal. No caso de problemas ambientais, esta hipótese não parece refletir realisticamente os processos sob estudo já que, implicitamente, assume-se que a função de covariância é simétrica. Com o objetivo de relaxar estas suposições usualmente feitas, propomos um modelo espaço-temporal baseado nos modelos de convoluções espaciais discreta, que além de contemplar uma estrutura de covariância flexível modela conjuntamente a média do processo.

A segunda parte deste trabalho investiga os processos autoregressivos de ordem p , $AR(p)$, que podem ser utilizados na modelagem de dados ambientais para detecção de tendência e sazonalidade. As raízes reais positivas fornecem persistência e as raízes complexas capturam periodicidades. Desta forma, a segunda proposta considera que os coeficientes do modelo $AR(p)$ variam espacialmente, implicando que para cada localização temos uma série temporal, modelada através de um processo autoregressivo com a ordem comum para todas as localizações mas com coeficientes diferentes, que variam suavemente ao longo do espaço. Prioris adequadas para os coeficientes do modelo $AR(p)$ são consideradas para garantir a estacionariedade do processo.

Palavras-chave: convolução espacial; covariância não- separável; não estacionariedade; previsão.

ABSTRACT

A CLASS OF SPATIO-TEMPORAL MODELS BASED ON MULTIVARIATE AUTOREGRESSIVE PROCESSES

Aline Araújo Nobre

Orientadora: Profa. Alexandra M. Schmidt

Co-orientador: Prof. Bruno Sansó

Spatio-temporal models have had been widely used in the last years. This is due to the large number of applications coming from the environmental and health sciences. The main focus of a spatio-temporal model is on the covariance structure of the process and its usual assumptions. In general, the covariance structure is modelled as a function of the distance between the monitoring stations. This implies on assuming stationarity and isotropy of the process. Another usual assumption is to suppose a separable covariance structure, in other words, the covariance function of the process is built by taking the product between the spatial and temporal covariances. In environmental problems these assumptions do not reflect realistically the processes since we assume that the covariance function is symmetric. We propose a spatio-temporal model based on discrete spatial convolutions to provide a flexible covariance structure. We also consider the joint modelling of the mean of the process.

The second part of this thesis studies autoregressive processes of order p ($AR(p)$) as they are useful in detecting trend and seasonality, usually present in environmental data. The positive real roots provide some persistence and the complex roots provide periodicities. Therefore, the other proposed model considers that the coefficients of an $AR(p)$ vary spatially, implying that for each location we have a time series modelled through an autoregressive process with a common order for all locations, but with different coefficients which vary smoothly over the region of interest. We impose a prior structure that guarantees that the resulting coefficients of the $AR(p)$ are in the stationarity region.

Key words: spatial convolution; non-separable covariance, non stationarity, prediction.

Sumário

1	Introdução	1
2	Definições e Revisão Bibliográfica	4
2.1	Geoestatística	4
2.2	Séries Temporais	7
2.2.1	Modelos Lineares Dinâmicos	9
2.3	Revisão Bibliográfica	10
3	Processos de Convolução	16
3.1	Modelos de Coregionalização Linear	16
3.2	Modelo Espaço-Temporal baseado em Convolução Discreta	18
3.2.1	Modelo de Convolução	18
3.2.2	Modelo de Convolução Temporal	19
3.2.3	Modelo de Convolução Espacial	21
3.3	Modelagem da Estrutura de Média	23
3.4	Especificação das Distribuições a Priori	25
3.5	Procedimento de Inferência	26
3.5.1	Distribuições Condicionais Completas a Posteriori	27
3.6	Previsão Temporal e Interpolação Espacial	33
3.7	Ajustando o Modelo de Convolução Espacial	34
3.7.1	Dados Artificiais com Covariáveis	35
3.7.2	Dados Artificiais com Estrutura Sazonal	45

3.7.3	Análise dos níveis de PM_{10} na Região Metropolitana do Rio de Janeiro	53
3.7.4	Análise da Temperatura na Costa Central da Califórnia	68
4	Processos Autoregressivos com coeficientes variando espacialmente	80
4.1	Modelos Autoregressivos	80
4.2	Generalização do Modelo de Convolução Espacial Discreta	83
4.3	Distribuição a priori para os coeficientes do SVAR(p)	85
4.4	Procedimento de Inferência	87
4.5	Previsão Temporal e Interpolação Espacial	91
4.6	SVAR(p) com a ordem p desconhecida	92
4.6.1	Priori de mistura com massa de probabilidade no zero	93
4.6.2	Mistura baseada em duas distribuições contínuas	97
5	Análise de Dados	104
5.1	Ajustando o modelo proposto considerando a ordem conhecida	104
5.1.1	Dados Artificiais - SVAR(2)	104
5.1.2	Análise dos níveis de PM_{10} em Santiago do Chile, Chile	114
5.2	Ajustando o modelo proposto considerando a ordem desconhecida	122
5.2.1	Priori de mistura com massa de probabilidade em zero	122
5.2.2	Priori de mistura baseada em duas distribuições contínuas	123
6	Considerações Finais e Trabalhos Futuros	127
6.1	Considerações Finais	127
6.2	Trabalhos Futuros	129
6.2.1	Problema de Redução de Dimensão	129
6.2.2	Distribuição a priori beta para as raízes recíprocas	131
A	Critério de Comparação de Modelos	132
B	Relação entre as raízes recíprocas e os coeficientes de um modelo AR(p)	133

Capítulo 1

Introdução

Processos espaço-temporais vêm ganhando grande popularidade nos últimos anos. Uma razão para isto é o crescimento de aplicações surgindo das ciências ambientais e de saúde. Exemplos comuns incluem o número de casos de doenças respiratórias numa cidade, a modelagem de poluentes do ar num grande centro urbano, velocidade do vento, etc. É usual que estes tipos de dados sejam observados em diferentes pontos da região de interesse e em diferentes instantes do tempo. Paralelamente, o avanço de técnicas de simulação estocástica e, também, de ferramentas computacionais vêm permitindo o uso de modelos altamente estruturados que procuram refletir melhor a realidade dos processos sob estudo nas diversas áreas da Estatística. Em particular, tem havido um grande desenvolvimento de novas técnicas para o estudo de dados observados ao longo de uma região geográfica. Schmidt, Nobre, e Ferreira (2002) apresentam alguns aspectos relevantes na modelagem de dados espacialmente referenciados.

A estatística espacial está dividida em três grandes áreas de acordo com os tipos de observações associadas ao espaço em que elas são observadas, são elas: dados de área, geoestatística e processos pontuais (Cressie 1993). Dados de área constituem dados observados em uma sub-região regular ou em uma sub-região irregular como, por exemplo, municípios de um particular estado. Dados geoestatísticos ocorrem quando as observações são realizações de um processo espacial contínuo. Finalmente, os processos pontuais ocorrem quando as localizações das observações são as quantidades aleatórias de interesse. Neste trabalho trataremos de dados geoestatísticos. O foco principal dos modelos para dados geoestatísticos está na especificação da função de covariância do processo e, conseqüentemente, nas

suposições usualmente feitas sobre esses processos. Em geral, a estrutura de covariância é modelada em função da distância entre as estações monitoradoras. Em geoestatística é comum assumir que o processo espacial é estacionário, ou seja, a distribuição espacial é invariável quando a origem do sistema de coordenadas é transladada, e isotrópico quando o processo é estacionário sob rotações em torno da origem (Schmidt, Nobre, e Ferreira 2002). Outra suposição bastante comum em modelos espaço-temporais é assumir uma estrutura de covariância separável, isto é, a função de covariância do processo é construída pelo produto entre as covariâncias espaciais e temporais. No caso de problemas ambientais, esta hipótese não parece refletir realisticamente os processos sob estudo já que, implicitamente, assume-se que a função de covariância é simétrica, isto é, $cov(y(s, t), y(s', t')) = cov(y(s, t'), y(s', t))$ (Schmidt e Sansó 2006). Em geral, para o ajuste de modelos espaço-temporais é necessário a retirada de qualquer estrutura temporal presente nos dados. Assim, o principal objetivo deste trabalho é modelar a média e a covariância conjuntamente de um processo espaço-temporal, utilizando estruturas flexíveis de covariância. A primeira proposta deste trabalho é baseada no modelo de convolução espacial discreta.

Os processos autoregressivos são bastante utilizados na literatura na modelagem de observações temporais. Os processos autoregressivos de ordem p , $AR(p)$, podem ser utilizados na modelagem deste tipo de dados para detecção de tendência e sazonalidade. As raízes reais positivas fornecem persistência e as raízes complexas capturam periodicidades da série sob estudo. Estas características são comumente encontradas em dados ambientais. Os processos autoregressivos podem ser vistos dentro de uma estrutura multivariada, por exemplo, no espaço. A idéia é combinar a flexibilidade destes processos na detecção de componentes temporais com uma estrutura espacial. A segunda proposta deste trabalho é considerar que os coeficientes de um processo $AR(p)$ variam espacialmente, implicando que para cada localização, temos uma série temporal modelada através de um processo autoregressivo, com uma ordem comum para todas as localizações mas, com diferentes coeficientes para cada uma das localizações.

Este trabalho está organizado como segue: O Capítulo 2 faz uma revisão bibliográfica citando os recentes trabalhos desenvolvidos nesta área, enfatizando as principais suposições usualmente consideradas e os esforços dos pesquisadores em propor modelos mais realistas. O Capítulo 3 faz uma revisão do modelo de coregionalização linear e do modelo proposto por Sansó, Schmidt, e Nobre (2006) para

obtenção de funções de covariância espaço-temporais não separáveis. Em seguida, propomos uma extensão do modelo inicialmente proposto por Sansó, Schmidt, e Nobre (2006) para modelar a média e a covariância conjuntamente. O Capítulo 3 também ilustra o modelo proposto em quatro conjuntos de dados, dois conjuntos de dados artificiais gerados a partir do próprio modelo, um conjunto de dados de material particulado (PM_{10}) em estações monitoradoras na região metropolitana do Rio de Janeiro e um conjunto de dados de temperatura medidos na costa central da Califórnia. O Capítulo 4 apresenta os processos autoregressivos variando espacialmente. Inicialmente, consideramos que a ordem da componente $AR(p)$ é conhecida. Em seguida, generalizamos o modelo para o caso em que a ordem p é desconhecida. No Capítulo 5 o modelo proposto no capítulo anterior é verificado em dados simulados e, em seguida, é aplicado a um conjunto de dados de PM_{10} em Santiago do Chile, Chile. Finalmente, o Capítulo 6 apresenta algumas conclusões e projetos de trabalhos futuros.

Capítulo 2

Definições e Revisão Bibliográfica

Uma das principais suposições feitas na análise espacial é a de que dados coletados em uma região do espaço que estão próximos entre si possuem uma correlação maior do que dados mais distantes entre si. Na análise de séries temporais, os dados são correlacionados no tempo. A diferença básica entre estes dois tipos de dados é que em séries temporais os dados tem uma ordenação natural, enquanto que dados espaciais exibem correlações em todas as direções. Além disso, as séries temporais estão definidas em um espaço unidimensional, enquanto os dados espaciais estão definidos em um espaço d -dimensional positivo.

Neste Capítulo fazemos uma revisão da modelagem de dados espacialmente referenciados. Mais especificamente sobre geoestatística, com a discussão das principais suposições usualmente feitas. Em seguida, fazemos uma revisão de séries temporais. Finalmente, fazemos uma revisão de alguns trabalhos encontrados na literatura para modelagem conjunta de dados espaço-temporais.

2.1 Geoestatística

Dados geoestatísticos são observações obtidas a partir de um processo contínuo no espaço. Este tipo de dados são amplamente utilizados em ciências ambientais e é o foco da abordagem deste trabalho. Seja $\{Y(s) : s \in D\}$, um processo aleatório contínuo, onde D é um subconjunto fixo do \mathbb{R}^d com volume d -dimensional positivo (Cressie 1993). Usualmente, assume-se $d = 1, 2$ ou 3 .

O objetivo principal da análise de dados espaciais é a identificação das variações de primeira

ordem ou de grande escala, isto é, $E(Y(s))$, e as variações de segunda ordem ou de pequena escala que são as dependências entre as diferentes localizações, isto é, $cov(Y(s), Y(s'))$, para $s, s' \in D$. Para a modelagem de dados espaciais, temos que considerar os conceitos de estacionariedade e isotropia que estão diretamente ligados a especificação destas variações.

Suponha que $\mu(s) = E(Y(s))$ e $Var(Y(s))$ existem para todo $s \in D$, onde $\mu(s)$ é a tendência do processo espacial $Y(s)$, e $Var(Y(s))$ é a variância de $Y(s)$.

- *Estacionariedade de 2ª ordem*

O processo $Y(s)$ é estacionário de segunda ordem, ou fracamente estacionário, se

$$\mu(s) = \mu \quad \text{e} \quad cov\{Y(s), Y(s')\} = c(s - s') \quad s, s' \in D,$$

isto é, $\mu(s)$ é constante para todo $s \in D$ e a covariância entre dois pontos quaisquer em D é função apenas da diferença entre as duas localizações. A função $c(s - s')$ é chamada de covariograma.

- *Estacionariedade Intrínseca*

O processo espacial $Y(s)$ é dito intrinsecamente estacionário se

$$E(Y(s+h) - Y(s)) = 0 \quad \text{e} \quad Var(Y(s+h) - Y(s)) = 2\gamma(h), \quad \forall s, s+h \in D,$$

onde a quantidade $2\gamma(\cdot)$, é conhecida como variograma e é um dos parâmetros mais importantes na modelagem de geoestatística (Cressie 1993).

- *Isotropia*

Um processo é isotrópico quando a covariância depende apenas da distância entre as localizações, isto é, $2\gamma(s - s')$ é função da distância euclideana, $\|s - s'\|$.

A estacionariedade de 2ª ordem implica estacionariedade intrínseca. Um processo intrinsecamente estacionário e isotrópico é chamado de homogêneo (Smith 1996). Se uma dessas condições não se aplica, o processo é heterogêneo. Quando um processo é homogêneo, sua variância é constante e a função de covariância pode ser escrita como

$$c(s, s') = \sigma^2 \rho(\|s - s'\|; \Lambda)$$

onde $\rho(\cdot; \Lambda)$ representa uma função de correlação válida, isto é, positiva definida, dependendo de um vetor paramétrico Λ . Este é um dos grande atrativos para os processos homogêneos, pois a estrutura de covariância do processo pode ser modelada apenas através dos parâmetros σ^2 e Λ . Em geral, assume-se que o processo espacial $Y(s)$ segue um processo gaussiano. Mais especificamente, a variável aleatória $Y(s)$, que assume valores $y(s)$ para $s \in D$, segue um processo gaussiano (PG) com média $\mu(s)$ e função de covariância $c(s_i, s_j)$, denotado por $Y(s) \sim PG(\mu(s), c(s_i, s_j))$, se para quaisquer $s_1, \dots, s_n \in D$, e qualquer $n = 1, 2, \dots$, a distribuição conjunta de $Y(s_1), \dots, Y(s_n)$ é uma normal multivariada com parâmetros dados por $\mu(s_i)$ e $c(s_i, s_j)$ (O'Hagan 1994). Sendo assim, precisamos especificar apenas os momentos de primeira e segunda ordens.

Schmidt, Nobre, e Ferreira (2002) apresentam alguns dos principais modelos paramétricos de funções de correlação, a saber:

1. Família exponencial potência

$$\rho(d; \Lambda) = \exp \left\{ - \left(\frac{d}{\lambda} \right)^\kappa \right\},$$

onde $\Lambda = (\lambda, \kappa)$ com $\lambda > 0$ e $\kappa \in (0, 2]$. O parâmetro λ é o parâmetro de escala e d é a distância euclidiana entre dois pontos quaisquer em D . Quando $\kappa = 1$ temos o caso particular da função de correlação exponencial e $\kappa = 2$ corresponde à função de correlação gaussiana;

2. Família Matérn

$$\rho(d; \Lambda) = \frac{1}{2^{\kappa-1} \Gamma(\kappa)} \left(\frac{d}{\lambda} \right)^\kappa K_\kappa$$

onde $\Lambda = (\lambda, \kappa)$, $\lambda > 0$ é o parâmetro de escala e $\kappa > 0$ é o parâmetro de forma. A função $\Gamma(\cdot)$ é a função gama usual e K_κ é a função modificada de Bessel do terceiro tipo de ordem κ .

Um aspecto importante de superfícies espaciais é o grau de suavidade. Matematicamente, esta propriedade é descrita através do grau de diferenciabilidade do processo. A especificação da família de função de correlação é de grande importância, pois em processos gaussianos a suavidade do processo está diretamente relacionada à diferenciabilidade da sua estrutura de covariância. Por exemplo, a função de correlação gaussiana resulta em processos que são infinitamente diferenciáveis, em outras palavras, em processos extremamente suaves. A função de correlação Matérn tem sido a mais usada na literatura, pois além das funções exponencial ($\kappa = 0,5$) e gaussiana ($\kappa \rightarrow \infty$) serem seus casos par-

ticulares, o parâmetro κ controla o grau de diferenciabilidade do processo (Schmidt, Nobre, e Ferreira 2002).

2.2 Séries Temporais

Uma série temporal é uma realização de um processo estocástico. A principal característica de uma série temporal é a dependência entre as observações. A análise de séries temporais consiste de técnicas para analisar esta dependência (Box, Jenkins, e Reinsel 1994).

Uma classe importante de modelos estocásticos para descrever séries temporais são os modelos estacionários. Note que é um caso análogo ao processo espacial, com a diferença que no caso espacial a dimensão de onde o processo está definido é geralmente 2 ou 3. Suponha que $\mu_t = E(Y_t)$ e $Var(Y_t)$ são a média e a variância do processo temporal Y_t e a função de covariância entre as observações Y_t e Y_{t+k} é denotada por $cov(Y_t, Y_{t+k})$.

- *Estacionariedade de 2ª ordem*

Um processo estocástico Y_t , $t \in \mathbb{R}$, é dito ser estacionário de segunda ordem ou fracamente estacionário, se a função de média é constante e a função de covariância depende apenas da defasagem, isto é,

$$E(Y_t) = \mu \quad \text{e} \quad cov(Y_t, Y_{t+k}) = \gamma_k.$$

γ_k , é chamada de função de autocovariância com defasagem k . Fazendo $k = 0$, temos que $Var(Y_t) = \gamma_0 = \sigma^2$, ou seja, a variância do processo assim como a média é constante. Similarmente, a função de autocorrelação com defasagem k é

$$\rho_k = \frac{cov(Y_t, Y_{t+k})}{\sqrt{Var(Y_t)Var(Y_{t+k})}} = \frac{\gamma_k}{\sigma^2}.$$

- *Estacionariedade estrita*

Uma série temporal é dita ser estritamente estacionária se a distribuição de probabilidade conjunta de $Y(t_1), \dots, Y(t_n)$ é a mesma de $Y(t_1 + k), \dots, Y(t_n + k)$. O deslocamento da origem dos tempos por uma quantidade k não afeta a distribuição conjunta.

Na prática é muito difícil usar a definição de estacionariedade estrita. A estacionariedade de 2ª ordem é mais utilizada já que muitas propriedades dos processos estacionários dependem apenas dos

momentos de primeira e segunda ordem. Se a distribuição de probabilidade de um processo temporal é um processo estacionário gaussiano, o processo está completamente especificado através de sua média, μ , variância, σ^2 , e função de autocorrelação, ρ_k .

A função de autocorrelação é uma das principais ferramentas de análise exploratória e modelagem de séries temporais, pois indica como cada valor em um dado instante de tempo t se relaciona com os valores em $t+1, \dots, t+k$. As matrizes de autocovariância e autocorrelação de um processo estacionário são ambas simétricas e positivas-definidas.

O modelo Autoregressivo de ordem p , $AR(p)$, o modelo de Médias Móveis de ordem q , $MA(q)$ e o modelo Autoregressivo e de Médias Móveis, $ARMA(p, q)$ descrevem procedimentos capazes de representar as observações de uma série temporal estacionária. Os modelos propostos aqui são baseados em processos autoregressivos e uma revisão destes processos é brevemente apresentada no Capítulo 4.

Em muitas aplicações práticas encontramos séries temporais que são não estacionárias e, portanto, precisamos de modelos que sejam capazes de relaxar tais suposições. Os modelos autoregressivos podem ser não estacionários. Neste caso, as séries temporais possuem um comportamento explosivo e não refletem a maioria dos problemas encontrados na prática. O método de Box e Jenkins consiste na busca de um modelo ARIMA (Autoregressivo Integrado de Médias Móveis) que represente o processo estocástico gerador da série temporal, a partir de um modelo ARMA aplicável na descrição de séries temporais estacionárias, estendendo este conceito para séries temporais não-estacionárias.

Vamos supor agora que $Y_t = \mu_t + \epsilon_t$, onde μ_t é a média do processo dependendo de t e ϵ_t são processos aleatórios com média zero. Quando a dependência temporal não é totalmente explicada por μ_t , faz-se necessário tratar a autocorrelação remanescente nos erros através de um modelo apropriado. Tomando o modelo de regressão como base, pode-se fazer uma extensão introduzindo uma equação que governe a evolução dos coeficientes de regressão ao longo do tempo (Paez e Gamerman 2005). Estes modelos são conhecidos como modelos dinâmicos lineares e são discutidos na Subseção 2.2.1. Estes modelos também são exemplos de modelos que tratam a não estacionariedade geralmente presente em séries temporais reais.

2.2.1 Modelos Lineares Dinâmicos

Um amplo estudo da classe de modelos lineares dinâmicos (MLD) pode ser encontrado em West e Harrison (1997), também chamados de modelos de espaço de estado. Nesta Seção é feita uma breve revisão sobre esta classe de modelos para o caso multivariado, já que é o caso utilizado neste trabalho. Suponha Y_t representando um vetor coluna de n observações para cada tempo $t = 1, \dots, T$. O modelo é definido pela quádrupla $\{F_t, G_t, V_t, W_t\}$ para cada tempo t tal que

$$\begin{aligned} Y_t &= F_t' \theta_t + v_t, & v_t &\sim N(0, V_t), \\ \theta_t &= G_t \theta_{t-1} + \omega_t, & \omega_t &\sim N(0, W_t), \end{aligned}$$

sendo que F_t é a matriz de regressão dinâmica $r \times n$ e G_t é a matriz de evolução do vetor de estados $r \times r$. A primeira equação é chamada equação de observação, na qual v_t é uma sequência de erros gaussianos independentes com média zero e matriz de covariância V_t , $n \times n$. A segunda é chamada de equação do sistema na qual ω_t é uma sequência de erros gaussianos independentes com média zero e covariância W_t . As sequências de erros v_t e ω_t são mutuamente independentes. Para cada tempo t o vetor de estados, θ_t , tem dimensão $r \times 1$.

Quando as matrizes F_t , G_t , V_t e W_t são conhecidas, o Filtro de Kalman é o procedimento utilizado para fazer inferência em modelos de espaço de estados. O filtro de Kalman fornece a distribuição condicional de θ_t , dada a informação disponível até o tempo t , D_t , de uma maneira computacionalmente eficiente. As equações do filtro de Kalman são obtidas utilizando o aspecto sequencial da inferência bayesiana. Inicialmente supõe-se que a priori inicial em $t = 0$ é normal multivariada, $\theta_0 \sim N(m_0, C_0)$, para algum vetor de médias m_0 e matriz de covariâncias C_0 , conhecidas. Assim, para cada tempo t as distribuições a priori, preditiva e posteriori são atualizadas utilizando as seguintes equações:

- Posteriori em t-1:

$$(\theta_{t-1} \mid D_{t-1}) \sim N(m_{t-1}, C_{t-1}),$$

- Priori em t:

$$(\theta_t \mid D_{t-1}) \sim N(a_t, R_t),$$

$$\text{com } a_t = G_t m_{t-1} \quad e \quad R_t = G_t C_{t-1} G_t' + W_t.$$

- Previsão um passo a frente:

$$(Y_t | D_{t-1}) \sim N(f_t, Q_t),$$

$$\text{com } f_t = F_t' a_t \text{ e } Q_t = F_t' R_t F_t + V_t.$$

- Posteriori em t:

$$(\theta_t | D_t) \sim N(m_t, C_t),$$

$$\text{com } m_t = a_t + A_t u_t \text{ e } C_t = R_t - A_t A_t' Q_t,$$

$$\text{onde } A_t = R_t F_t Q_t^{-1} \quad \text{e} \quad u_t = Y_t - f_t.$$

Quando as matrizes F_t , G_t , V_t e W_t não são conhecidas não é possível utilizar o Filtro de Kalman e a inferência é feita através de métodos numéricos como, por exemplo, os métodos de Monte Carlo via Cadeias de Markov (MCMC), mais particularmente, utilizamos o amostrador de Gibbs e Metropolis-Hastings (Gamerman e Lopes (2006)). Neste caso, a simulação do vetor de estados, θ_t , é realizada através do algoritmo *Forward Filtering Backward Smoothing* (FFBS). O algoritmo FFBS foi proposto simultaneamente por Fruhwirth-Schnater (1994) e Carter e Kohn (1994). O algoritmo segue os seguintes passos:

1. Calcula a média e a variância das distribuições $p(\theta_t | D_t)$, $t = 1, \dots, T$ usando as equações do filtro de Kalman;
2. Gera θ_T da distribuição $p(\theta_T | D_T)$;
3. Calcula recursivamente os momentos da distribuição $p(\theta_{T-j} | \theta_{T-j+1}, D_T)$, $j = 1, \dots, T - 1$, e gera θ_{T-j} desta distribuição.

Assim, cada θ_t gerado é condicionado ao conjunto completo das observações. O FFBS é um algoritmo eficiente, pois leva em consideração a estrutura de correlação entre os parâmetros e, além disso, é de implementação simples.

2.3 Revisão Bibliográfica

Dados ambientais e de saúde são geralmente observados em pontos (estações monitoradoras) fixos ao longo de uma determinada região geográfica e estão tipicamente disponíveis na forma de séries

temporais e, portanto, é clara a necessidade da consideração de uma componente que considera a interação entre espaço e tempo no modelo. Existem vários trabalhos na literatura para tratar este problema. Entre eles, podemos citar, Huang e Cressie (1996) que apresentam um modelo para fazer previsão espaço-temporal. O modelo proposto é dado pela soma de um processo autoregressivo de ordem p , uma componente espacial e um ruído branco. Diferente da krigagem ordinária, que apenas os dados do tempo corrente são utilizados, nesta abordagem a previsão é feita usando os dados dos tempos anteriores, bem como os dados correntes. Por outro lado, Sansó e Guenni (2000) propõem um modelo espaço-temporal para modelar a quantidade de chuva medida em 80 estações monitoradoras, localizadas na Venezuela, no período de janeiro de 1968 à dezembro de 1983. Dados de chuva geralmente possuem muitos problemas de qualidade, e conseqüentemente, há grande presença de dados faltantes. Assim, os dados de chuva foram agrupados em períodos de 10 dias, pois além de reduzir a proporção de dados faltantes, diminuí também a quantidade de zeros, facilitando a obtenção da suposição de normalidade. Os dados faltantes são tratados como variáveis latentes no processo de estimação. Sansó e Guenni (2000) propõem um modelo espaço-temporal dinâmico, usando uma distribuição normal multivariada truncada e transformada, sendo que a transformação para atingir normalidade é estimada pelo próprio modelo. O modelo inclui um termo para tratar a tendência espacial, através da localização das estações, bem como um termo para capturar a sazonalidade presente nos dados. Para acomodar a dependência espacial, a estrutura de correlação é modelada usando a função de correlação exponencial e o modelo ainda inclui uma fonte extra de variabilidade. Entretanto, o modelo é estacionário no espaço. Paez e Gamerman (2003) fazem uma aplicação de modelos espaço-temporais para dados de poluentes. A média do processo é dada pela soma de componentes espaciais e temporais, mais o efeito da temperatura máxima e de variáveis indicadoras do dia da semana. Diferentes especificações para as componentes espaciais e temporais são investigadas pelos autores. Para a componente temporal é considerado um passeio aleatório, um processo autoregressivo de ordem 1, $AR(1)$, entre outros. Para a componente espacial, as especificações incluem um processo autoregressivo condicional (CAR) e um processo gaussiano com função de correlação exponencial. Vários modelos são ajustados combinando estas diferentes especificações. Huerta, Sansó, e Stroud (2004) propõem um modelo espaço-temporal para os níveis de ozônio no México. O modelo incorpora uma variável meteorológica explicativa que é temperatura, cujo efeito varia no tempo, e componentes harmônicos com o objetivo de incorporar

a periodicidade presente nos dados de ozônio. O modelo é escrito na forma dos modelos lineares dinâmicos e os erros das equações de observação e sistema possuem estrutura espacial com função de covariância exponencial. Entretanto, os parâmetros da função de covariância dos erros de evolução não são estimados simultaneamente com os outros parâmetros do modelo e ele é estacionário no espaço.

Como vimos, em geoestatística é comum assumir estacionariedade e isotropia do processo espacial. Entretanto, em muitos problemas ambientais esta suposição não é realista pois acredita-se que existem influências locais na estrutura de correlação do processo espacial aleatório. Existem vários modelos propostos na literatura que não assumem isotropia do processo espacial sob estudo. Stroud, Müller, e Sansó (2001) propõem um modelo para analisar dados espaço-temporais não estacionários. Para capturar a variabilidade espacial, a média do processo é descrita através de uma mistura ponderada de superfícies de regressão linear e para capturar a variabilidade temporal as superfícies de regressão variam ao longo do tempo. O número de componentes de mistura bem como os centros, não são estimados e são escolhidos a partir de uma análise preliminar dos dados. Embora a suposição de isotropia não seja exigida, os dois exemplos apresentados usam núcleo gaussiano. Uma das vantagens do método proposto por Stroud, Müller, e Sansó (2001) é ser computacionalmente simples pois utiliza o filtro de Kalman e, portanto, as distribuições a posteriori dos parâmetros e, preditiva, têm forma fechada. Sampson e Guttorp (1992) propõem um método não paramétrico para tratar a não estacionariedade e a anisotropia espaciais. Eles usam uma técnica de deformação espacial para construção de um espaço latente. Schmidt e O'Hagan (2003) propõem uma aproximação bayesiana para o modelo proposto por Sampson e Guttorp (1992) baseado numa transformação não-linear de um espaço observado para um espaço latente, no qual a estrutura espacial é isotrópica e estacionária. Entretanto, os autores não consideram a estrutura temporal do processo e tratam da estrutura espacial após remover a dependência temporal dos dados. Higdon, Swall, e Kern (1999) propõem um modelo baseado em convoluções espaciais, permitindo que o núcleo varie suavemente ao longo do espaço. Desta maneira, obtém-se uma função de covariância que não é função da distância, garantindo a não estacionariedade do processo. O modelo proposto por Fuentes (2002) é também baseado em convoluções espaciais. Neste caso, a convolução se dá entre um núcleo e um processo gaussiano estacionário com média zero e função de covariância estacionária, resultando em um processo não estacionário. Paciorek e Schervish (2006) estendem as funções de covariância não estacionárias introduzida por Higdon, Swall, e Kern (1999)

utilizando uma versão não estacionária da função de correlação Matérn.

Como foi descrito anteriormente, fenômenos naturais, em geral, requerem modelos estatísticos capazes de descrever sua variabilidade no espaço e no tempo. O desafio é propor uma estrutura de covariância que descreva a correlação espacial entre as observações para cada instante do tempo de forma realista. É importante propor uma estrutura de covariância que seja válida (positiva definida). Uma aproximação simples para modelar processos espaço-temporais é considerar uma estrutura de covariância separável. Em geral, modelos separáveis são escolhidos por conveniência e não por sua capacidade de se ajustar bem aos dados. Cressie e Huang (1999) introduzem novas famílias paramétricas de funções de covariância espaço-temporais estacionárias e não-separáveis. Estas classes são baseadas em funções de covariância e funções de densidade espectral. Vários exemplos são construídos e alguns são utilizados para modelar dados de velocidade do vento. Entretanto, a escolha das funções de covariância ainda é um problema devido a dificuldade de interpretá-las. Brown, Karesen, Roberts, e Tonellato (2000) propõem um modelo baseado em equações diferenciais estocásticas que resulta em uma estrutura de covariância não-separável. Storvik, Frigessi, e Hirst (2002) comparam duas diferentes estratégias de modelagem, discutindo suas interpretações, inferências e vantagens computacionais de cada uma. A primeira estratégia especifica uma forma paramétrica para função de covariância. Na segunda, a função de covariância não-separável é construída a partir de modelos hierárquicos dinâmicos, usando a convolução de processos auto-regressivos temporais e perturbações aleatórias gaussianas, espacialmente estruturadas. Outro trabalho que trata do problema da especificação da função de covariância é Gneiting (2002). A proposta é similar a de Cressie e Huang (1999), mas sem as limitações de usar a transformação de Fourier de forma fechada. Assim, classes mais gerais de funções de covariância não-separáveis são construídas. O artigo também mostra que algumas das funções de covariância propostas por Cressie e Huang (1999) não são válidas. Stein (2005) discute algumas propriedades da função de covariância espaço-temporal, tais como suavização da origem. Uma classe paramétrica de densidades espectrais com funções de covariância espaço-temporais infinitamente diferenciáveis na origem e com diferentes graus de suavização para o processo no espaço e no tempo é construída. Paez, Gamerman, e Oliveira (2005) propõem um modelo de regressão com coeficientes variando no espaço e no tempo, estendendo o modelo inicialmente proposto por Paez e Gamerman (2003). Neste caso, a estrutura de média é decomposta em uma componente temporal seguindo um processo autoregressivo

de primeira ordem e covariáveis cujos coeficientes seguem processos gaussianos com função de correlação espacial, implicando em uma estrutura de covariância não estacionária. Fuentes, Chen, Davis, e Lackmann (2005) desenvolvem um modelo para mapear a velocidade do vento. A velocidade do vento é modelada como uma função de tendência espaço-temporal e um processo com média zero e covariância não separável e não estacionária. Os efeitos das covariáveis são permitidos variarem no tempo e no espaço de maneira similar a abordada por Gelfand, Banerjee, e Gamerman (2005). A estrutura de covariância não separável é obtida através de uma mistura finita de processos espaço-temporais localmente estacionários. Os parâmetros da função de correlação são diferentes para cada componente da mistura.

Um outro problema importante no contexto de processos ambientais é a modelagem de processos espaciais multivariados. Por exemplo, em geral, devido ao alto custo de instalação das estações monitoradoras, temos observações de diferentes poluentes em cada ponto. Portanto, neste caso precisamos considerar não só a estrutura de covariância espacial e temporal mas, também, a estrutura de covariância entre os diversos processos observados no mesmo ponto. Mardia e Goodall (1993) fazem uma análise de dados espaço-temporais multivariados. Após remover a tendência temporal em cada localização, os dados são considerados como medidas multivariadas repetidas no espaço. A matriz de covariância é dada pelo produto de Kronecker entre as covariâncias temporal, espacial e os diferentes processos em cada localização. A inferência destes modelos é feita usando métodos de máxima verossimilhança. Com o objetivo de obter normalidade, as observações são transformadas considerando a aproximação de Box-Cox. Além disso, Mardia e Goodall (1993) usam a técnica de deformação espacial introduzida por Sampson e Guttorp (1992) para obter isotropia. Brown, Le, e Zidek (1994) estendem a teoria de Le e Zidek (1992) para o caso multivariado, empilhando o vetor resposta obtendo para cada tempo, um vetor com as observações das diferentes medidas nas diferentes estações. Os autores assumem que a matriz de covariância tem uma distribuição Wishart invertida e que a média do processo é função de covariáveis. Assim, não é necessário remover tendência e sazonalidade. Uma restrição do modelo é não permitir a inclusão de um parâmetro para periodicidade, já que resultaria em um modelo não linear. A grande vantagem do artigo é usar uma estrutura de covariância bem geral, relaxando as suposições de estacionariedade e isotropia. Schmidt e Gelfand (2003) propõem um modelo bayesiano baseado no modelo de coregionalização linear, onde o processo

multivariado é modelado a partir da combinação linear de processos espaciais univariados. Neste trabalho as observações multivariadas são provenientes de diferentes medições ao longo do espaço. Por exemplo, temos vários poluentes medidos em diferentes estações monitoradoras. O modelo proposto pelos autores tem a vantagem de obter uma estrutura de covariância flexível, isto é, não separável e não-estacionária. A aproximação adotada por Gelfand, Banerjee, e Gamerman (2005) segue a classe de modelos dinâmicos. A variável resposta é definida pela soma de uma média variando no tempo e no espaço e erros de medida não correlacionados. A evolução dos coeficientes da estrutura de média é escrito pela soma entre uma componente puramente temporal e uma componente espaço-temporal. Ambas evoluem no tempo através de um processo autoregressivo de ordem $p = 1$. A diferença é que os erros da componente puramente temporal são independentes, enquanto que os erros da componente espaço-temporal segue um processo espacial gaussiano. Resultando assim, numa estrutura de covariância não separável. Esta aproximação também é baseada nos modelos de coregionalização. A proposta é estendida para o caso multivariado, com a simplificação de apenas o intercepto ter uma estrutura espaço-temporal. A proposta é ilustrada em dado artificiais, bem como em dados de temperatura e precipitação.

O principal objetivo deste trabalho é propor um modelo espaço temporal contemplando uma estrutura de covariância flexível, utilizando modelos alternativos aos mencionados anteriormente. Para isso, utilizamos duas propostas:

- Modelo baseado no modelo de convolução espacial discreta que modela conjuntamente a média do processo;
- Modelo considerando que os coeficientes do processo autoregressivo variam espacialmente.

Capítulo 3

Processos de Convolução

Neste capítulo é feito, inicialmente, uma revisão dos Modelos de Coregionalização Linear (MCL). Em seguida, apresentamos o modelo proposto como parte da proposta deste trabalho (Seções 3.2 e 3.3). O procedimento de inferência é baseado no Paradigma de Bayes. Na Seção 3.4 discutimos a escolha das distribuições a priori para os parâmetros do modelo e o esquema de amostragem utilizado para obtenção das amostras da distribuição a posteriori encontram-se na Seção 3.5. Na Seção 3.6 calculamos a distribuição preditiva do modelo usada para fazer previsão temporal e interpolação espacial; e, finalmente, na Seção 3.7 aplicamos o modelo proposto a dois conjuntos de dados simulados e a dois conjuntos de dados reais.

3.1 Modelos de Coregionalização Linear

Suponha que $Y(s)$, seja um vetor $p \times 1$ de observações na localização s em uma região $D \subset \mathbb{R}^2$. Sendo assim, devemos usar um modelo capaz de capturar a estrutura de associação existente entre as localizações espaciais, bem como a estrutura de correlação existente dentro do vetor de observações em cada localização. O MCL foi introduzido por Goulard e Voltz (1992) para o problema de redução de dimensão. Gelfand, Schmidt, Banerjee, e Sirmans (2004) propõem uma generalização deste modelo. O modelo mais simples chamado de modelo de coregionalização linear intrínseco é dado por $Y(s) = Kw(s)$ onde K é uma matriz $p \times p$ de posto completo e $w(s) = w_1(s), \dots, w_p(s)$ são processos espaciais independentes e identicamente distribuídos. Se $w_j(s)$, $j = 1, \dots, p$, tem média zero, variância 1 e

função de correlação $\rho(s-s')$, a matriz de covariância entre $Y(s)$ e $Y(s')$ é dada por $cov(Y(s), Y(s')) = \rho(s-s')KK^t$, com $KK^t = \Sigma$, $p \times p$, representando a matriz de covariância entre as componentes de Y , na localização s , e K uma matriz triangular inferior. Neste caso, a estrutura de covariância é separável, ou seja, é dada pelo produto entre a covariância do processo espacial, $\rho(s-s')$, e a covariância entre as componentes de Y , Σ . O termo intrínseco é usado pois o momento de primeira ordem é zero e o momento de segunda ordem depende das localizações apenas através da diferença, $h = s - s'$.

Um modelo de coregionalização mais geral pode ser considerado se os $w_j(s)$, $j = 1, \dots, p$ são independentes mas não identicamente distribuídos. Por exemplo, suponha que $w_j(s)$ tenha média μ_j , variância 1 e função de correlação $\rho_j(h)$. Assim, $E(Y(s)) = K\boldsymbol{\mu}$ onde $\boldsymbol{\mu} = (\mu_1, \dots, \mu_p)$ e a matriz de covariância será dada por

$$cov(Y(s), Y(s')) = \sum_{j=1}^p \rho_j(s-s')\Sigma_j,$$

onde $\Sigma_j = k_j k_j'$ e k_j é a j -ésima coluna da matriz K . Assim, a estrutura de covariância agora é não separável. Sendo assim, os modelos de coregionalização linear fornecem uma maneira de obter estruturas de covariâncias bem gerais, contemplando possíveis interações entre as observações ao longo do espaço e as diferentes variáveis medidas.

Uma função de distribuição de probabilidade conjunta pode ser construída através do produto de distribuições de probabilidade condicionais. Nos modelos de coregionalização, escrever a distribuição conjunta como produto das distribuições condicionais traz muitas vantagens computacionais e de interpretação. Por exemplo, suponha que $f(Y_1(s), \dots, Y_p(s)) = f(Y_1(s))f(Y_2(s)|Y_1(s)) \cdots f(Y_p(s)|Y_1(s), \dots, Y_{p-1}(s))$. Fazendo $p = 2$ e K uma matriz triangular inferior com elementos k_{ij} , pode-se escrever $Y_1(s) = k_{11}w_1(s)$ e $Y_2(s)|Y_1(s) = \frac{k_{21}}{k_{11}}Y_1(s) + k_{22}w_2(s)$. Quando trabalhamos com a forma conjunta obtemos uma matriz de covariância $np \times np$. Já na forma condicional obtemos p matrizes de covariância de dimensão $n \times n$, atingindo assim, uma redução significativa no custo computacional. Vantagens na interpretação surgem quando existe uma ordem natural no condicionamento das variáveis.

Gelfand, Schmidt, Banerjee, e Sirmans (2004) permitem que a matriz K varie espacialmente, ou seja, $Y(s) = K(s)w(s)$. Desta forma, obtém-se um modelo ainda mais flexível, pois além da estrutura de covariância ser não separável, o processo agora é também não estacionário. Assim, a covariância é tal que $cov(Y(s), Y(s')) = \sum_{j=1}^p \rho_j(s-s')k_j(s)k_j'(s')$.

3.2 Modelo Espaço-Temporal baseado em Convolução Discreta

Sansó, Schmidt, e Nobre (2006) consideram uma classe de modelos espaço-temporais baseados em processos de convoluções independentes com um núcleo discreto que é representado por uma matriz triangular inferior. Com o objetivo de obter um processo espaço-temporal são consideradas duas possibilidades: (a) localização fixa e convolução temporal, e (b) tempo fixo e convolução espacial. Essas abordagens são baseadas no modelo de coregionalização apresentado por Gelfand, Schmidt, Banerjee, e Sirmans (2004) discutido na seção anterior. A diferença é que $Y(s)$ é um vetor de observações temporais de dimensão T . Seja $y(t, s)$, $t = 1, \dots, T$ e $s \in D$ um processo espaço-temporal. Neste caso, precisamos considerar tanto a estrutura de correlação espacial quanto a estrutura de correlação temporal. Como foi discutido no Capítulo 1, uma maneira simples de construir a função de correlação conjunta é considerar o produto das funções de correlação espacial e temporal. O problema é que esta não é uma suposição realista para a maioria dos problemas encontrados na prática, pois estamos assumindo que a covariância do processo é simétrica, isto é, $cov(y(t, s), y(t', s')) = cov(y(t', s), y(t, s'))$ (Schmidt e Sansó 2006). Esta simetria implica que a covariância entre o processo no tempo t e localização s e o processo no tempo t' e localização s' é a mesma que a covariância entre o processo no tempo t' e localização s e o processo no tempo t e localização s' . Portanto, o desafio é propor um modelo que considere possíveis interações entre espaço e tempo. Vimos na seção anterior que a partir do modelo de coregionalização linear podemos obter estruturas de covariâncias bem gerais. Uma amostra de um processo espaço-temporal pode ser vista de duas maneiras: observações temporais em uma região (denotado aqui por modelo de convolução temporal) ou observações espaciais ao longo do tempo (chamado aqui de modelo de convolução espacial). Em outras palavras, a especificação desses modelos está relacionada com a forma do empilhamento das observações. A seguir discutimos as duas abordagens.

3.2.1 Modelo de Convolução

Os processos de convolução são uma aproximação para construir um processo espacial. Existem vários trabalhos na literatura que utilizam os processos de convolução. (Thiébaux e Pedder 1987) e (Barry e Ver Hoef 1996) introduzem a aproximação como um processo média móvel espacial. (Higdon 1998) e (Higdon 2002) utilizam uma aproximação discreta para um processo contínuo produzindo um processo

espaacial contínuo. (Higdon 1998) considera o modelo de convolução espacial definido pela convolução entre um ruído branco gaussiano e um núcleo suave que depende da localização e do tempo. O modelo é aplicado a um conjunto de dados de temperatura no oceano Atlântico.

Uma representação flexível de um processo gaussiano é dada através da convolução de um ruído branco e um núcleo. Sejam $w(s)$ ruído branco e $k(\cdot; \theta)$ um núcleo suave, então

$$y(s) = \int_D k(u - s; \theta)w(u)du$$

A função de covariância do processo gaussiano resultante é

$$c(d) = \text{cov}(y(s), y(s')) = \int_D k(u - d; \theta)k(u; \theta)du .$$

Isto significa que se variarmos a escolha do núcleo, obtemos uma variedade de funções de covariância.

3.2.2 Modelo de Convolução Temporal

O processo espaço-temporal $y(t, s)$, $t = 1, \dots, T$ e $s \in D$ do modelo de convolução temporal proposto por Sansó, Schmidt, e Nobre (2006) pode ser escrito como

$$y(t, s) = \int_{\tau} k(u - t; \Phi)x(u, s)du, \quad (3.1)$$

onde $x(u, s)$ é um processo gaussiano, e $k(\cdot; \Phi)$ é um núcleo, fornecendo a estrutura temporal e é função de um vetor paramétrico Φ . A discretização do modelo de convolução temporal é

$$Y(s) = KX(s), \quad s \in D,$$

onde $Y(s) = (y(1, s), \dots, y(T, s))'$, $X(s) = (x(1, s), \dots, x(T, s))'$ e $KK' = \Sigma$, $x(t, \cdot)$ é um processo Gaussiano com média zero e função de covariância $\eta(s - s'; \sigma^2, \lambda_t)$ e Σ corresponde a matriz de covariância de um processo AR(p) especificada pelos coeficientes, $\Phi = (\phi_1, \dots, \phi_p)$ e por sua variância, τ^2 . Assim, temos uma série temporal para cada localização s_i , $i = 1, \dots, n$, no espaço, expressa como uma combinação linear de processos espaciais não correlacionados no tempo.

Lema 1: A covariância resultante é

$$\text{cov}(y(t, s), y(t', s')) = \sum_{m=1}^{\min(t, t')} k_{tm}k_{t'm} \eta(s - s'; \sigma^2, \lambda_m),$$

Prova: Seja $y(t, s) = \sum_{m=1}^t k_{tm}x(m, s)$

$$\begin{aligned} \text{cov}(y(t, s), y(t', s')) &= \text{cov}\left(\sum_{m=1}^t k_{tm}x(m, s), \sum_{m=1}^{t'} k_{t'm}x(m, s')\right) \\ &= \sum_{m=1}^{\min\{t, t'\}} k_{tm}k_{t'm} \text{cov}(x(m, s), x(m, s')) \\ &= \sum_{m=1}^{\min\{t, t'\}} k_{tm}k_{t'm} \rho(s - s'; \sigma^2, \lambda_m). \end{aligned}$$

onde k_{tm} denota o elemento (t, m) da matriz K e $\text{cov}(x(m, s), x(m', s')) = 0$ se $m \neq m'$.

A estrutura de covariância do processo é estacionária no espaço, pois depende apenas de $s - s'$, mas não no tempo, e não separável. Se $\lambda_t = \lambda$ para todo tempo t , a estrutura de covariância do processo será dada pelo produto entre a covariância de um processo $\text{AR}(p)$ e a covariância do processo espacial (separável). Esta abordagem considera que a estrutura de correlação espacial pode ser diferente para cada tempo, quando permite que o parâmetro da função de correlação, λ_t , que representa o decaimento da correlação espacial, varie ao longo do tempo.

Outra possibilidade ainda mais geral é a utilização de um núcleo variando espacialmente como proposto em Gelfand, Schmidt, Banerjee, e Sirmans (2004),

$$Y(s) = K(s)X(s),$$

ou seja, os parâmetros do processo $\text{AR}(p)$ mudam para cada localização.

Lema 2: A covariância resultante é

$$\text{cov}(y(t, s), y(t', s')) = \sum_{m=1}^{\min(t, t')} k_{tm}(s)k_{t'm}(s')\eta(s - s'; \sigma^2, \lambda_m).$$

Prova: Seja $y(t, s) = \sum_{m=1}^t k_{tm}(s)x(m, s)$ e $\text{cov}(x(m, s), x(m', s')) = 0$ se $m \neq m'$.

$$\begin{aligned} \text{cov}(y(t, s), y(t', s')) &= \text{cov}\left(\sum_{m=1}^t k_{tm}(s)x(m, s), \sum_{m=1}^{t'} k_{t'm}(s')x(m, s')\right) \\ &= \sum_{m=1}^{\min\{t, t'\}} k_{tm}(s)k_{t'm}(s') \text{cov}(x(m, s), x(m, s')) \\ &= \sum_{m=1}^{\min\{t, t'\}} k_{tm}(s)k_{t'm}(s') \rho(s - s'; \sigma^2, \lambda_m). \end{aligned}$$

Neste caso, além de obter uma estrutura de covariância não separável, o processo também será não estacionário no espaço.

A seguir, discutimos os modelos de convolução espacial onde temos superfícies espaciais evoluindo no tempo.

3.2.3 Modelo de Convolução Espacial

Considerando agora o modelo de convolução espacial, o processo espaço-temporal $y(t, s), t = 1, \dots, T$ e $s \in D$ é dado por

$$y(t, s) = \int_S k(v - s; \Lambda) x(t, v) dv, \quad (3.2)$$

onde $k(\cdot; \Lambda)$ é um núcleo representando a correlação espacial do processo em cada tempo e $x(\cdot, s)$ é um processo $AR(p)$ com função de correlação $\rho(\cdot; \tau^2, \Phi)$, onde $\Phi = (\phi_1, \dots, \phi_p)$. A discretização do modelo de convolução espacial é

$$Y_t = KX_t, t = 1, \dots, T,$$

onde $Y_t = (y(t, s_1), \dots, y(t, s_n))'$, $X_t = (x(t, s_1), \dots, x(t, s_n))'$ com s_1, \dots, s_n representando as n localizações espaciais e $\Sigma = KK'$ representa a matriz de correlação espacial. Uma representação alternativa do modelo pode ser obtida considerando que $X_t = \sum_{j=1}^p \phi_j X_{t-j} + \epsilon_t$, $\epsilon_t \sim N(0, \tau^2 I)$. Assim,

$$Y_t = KX_t = \sum_{j=1}^p \phi_j KX_{t-j} + K\epsilon_t = \sum_{j=1}^p \phi_j Y_{t-j} + v_t,$$

onde $v_t \sim N(0, \tau^2 \Sigma)$. Assim, o modelo é equivalente a um $AR(p)$ multivariado com erros espacialmente correlacionados.

Lema 3: A estrutura de covariância é dada por

$$cov(y(t, s_i), y(t', s_j)) = [\Sigma]_{ij} \rho(t - t'; \tau^2, \Phi).$$

Prova: Seja $y(t, s_i) = \sum_{m=1}^i k_{im} x(t, s_m)$ e $cov(x(t, s_m), x(t', s_{m'})) = 0$ se $m \neq m'$.

$$\begin{aligned} cov(y(t, s_i), y(t', s_j)) &= cov\left(\sum_{m=1}^i k_{im} x(t, s_m), \sum_{m=1}^j k_{jm} x(t', s_m)\right) \\ &= \sum_{m=1}^{\min\{i, j\}} k_{im} k_{jm} cov(x(t, s_m), x(t', s_m)) \end{aligned}$$

$$\begin{aligned}
&= \sum_{m=1}^{\min\{i,j\}} k_{im}k_{jm}\rho(t-t';\tau^2,\Phi) \\
&= [\Sigma]_{ij}\rho(t-t';\tau^2,\Phi).
\end{aligned}$$

A covariância é descrita pelo produto entre a correlação espacial e temporal e será, portanto, separável, e estacionária no espaço se K é a decomposição de Cholesky de uma matriz de covariância estacionária. Uma estrutura de covariância não separável é obtida quando o núcleo varia no tempo, isto é, quando

$$Y_t = K_t X_t. \quad (3.3)$$

Lema 4: A estrutura de covariância é dada por

$$\text{cov}(y(t, s_i), y(t', s_j)) = \sum_{m=1}^{\min(i,j)} k_{im}(t)k_{jm}(t')\rho(t-t';\tau^2,\Phi).$$

Prova: Seja $y(t, s_i) = \sum_{m=1}^i k_{im}(t)x(t, s_m)$ e $\text{cov}(x(t, s_m), x(t', s_{m'})) = 0$ se $m \neq m'$.

$$\begin{aligned}
\text{cov}(y(t, s_i), y(t', s_j)) &= \text{cov}\left(\sum_{m=1}^i k_{im}(t)x(t, s_m), \sum_{m=1}^j k_{jm}(t')x(t', s_j)\right) \\
&= \sum_{m=1}^{\min\{i,j\}} k_{im}(t)k_{jm}(t')\text{cov}(x(t, s_m), x(t', s_m)) \\
&= \sum_{m=1}^{\min\{i,j\}} k_{im}(t)k_{jm}(t')\rho(t-t';\tau^2,\Phi).
\end{aligned}$$

A covariância do processo é não separável e não estacionária no tempo e no espaço. Como na convolução temporal, a estrutura de covariância não separável é obtida quando permitimos que os parâmetros da função de correlação espacial varie ao longo do tempo. Por enquanto, estamos considerando o caso em que o decaimento da função de correlação pode variar. Uma outra alternativa para obter não separabilidade, seria permitir a variação temporal da variância do processo espacial. Esta abordagem é discutida na próxima seção.

Assumindo X_t um $\text{AR}(p)$, podemos obter uma representação alternativa do modelo em (3.3). Isto é, $X_t = \sum_{j=1}^p \phi_j X_{t-j} + \epsilon_t$, $\epsilon_t \sim N(0, \tau^2 I)$, e

$$Y_t = K_t X_t = \sum_{j=1}^p \phi_j K_t K_{t-j}^{-1} Y_{t-j} + K_t \epsilon_t = \sum_{j=1}^p \nu_{jt} Y_{t-j} + v_t, \quad t = p+1, \dots, T,$$

onde $v_t \sim N(0, \tau^2 \Sigma_t)$ e $\nu_{jt} = \phi_j K_t K_{t-j}^{-1}$, $i = 1, \dots, p$. Portanto, obtemos um modelo AR(p) multivariado com coeficientes variando no tempo e no espaço. Assim, condicional a Y_1, \dots, Y_p a verossimilhança é proporcional a

$$\left(\prod_{t=p+1}^T \frac{|\Sigma_t|^{-1/2}}{(\tau^2)^{n/2}} \right) \exp \left\{ -\frac{1}{2\tau^2} \sum_{t=p+1}^T \left(Y_t - \sum_{j=1}^p \nu_{jt} Y_{t-j} \right)' \Sigma_t^{-1} \left(Y_t - \sum_{j=1}^p \nu_{jt} Y_{t-j} \right) \right\}.$$

Se Σ_t é uma função de covariância isotrópica, para cada tempo t , teremos $\Sigma_t = \sigma^2 \eta(s - s'; \lambda_t)$. Assim, a verossimilhança é uma função do produto $\sigma^2 \tau^2$. Como apenas o produto é identificável, decidimos por fixar $\tau^2 = 1$.

Em resumo, o modelo de convolução temporal pode levar a estruturas de covariância estacionárias no espaço, mas não no tempo, e não separáveis. Já o modelo de convolução espacial, pode levar a uma estrutura de covariância separável ou não-separável e não estacionária no tempo e no espaço. Desta forma, este modelo fornece uma classe rica de estruturas de covariância espaço-temporal.

É difícil visualizar situações realísticas em que temos séries temporais evoluindo no espaço, como é o caso da abordagem do modelo de convolução temporal. Este tipo de modelo pode ser conveniente estatisticamente quando o número de localizações espaciais é grande comparado ao número de observações no tempo. Neste trabalho iremos nos concentrar nos modelos de convolução espacial, onde temos superfícies espaciais evoluindo no tempo.

3.3 Modelagem da Estrutura de Média

O modelo em (3.3) não incorpora na modelagem a média do processo e, portanto, para utilização desses modelos é necessário inicialmente retirar qualquer tipo de estrutura temporal, por exemplo, tendência e sazonalidade, presente nos dados. Portanto, uma extensão natural é a incorporação desta componente no modelo. O objetivo é tratar, conjuntamente, os momentos de primeira e segunda ordens do processo espaço-temporal. A proposta aqui é modelar a média do processo através de covariáveis cujos efeitos variam ao longo do tempo e de componentes sazonais, se necessárias, utilizando o modelo de convolução espacial, descrito na Subseção 3.2.3.

Seguindo o modelo de convolução espacial em (3.3) e usando os modelos lineares dinâmicos (MLD)

proposto por West e Harrison (1997) propomos a seguinte estrutura

$$\begin{aligned}
Y_t &= F_t' \theta_t + K_t X_t \\
X_t &= \sum_{j=1}^p \phi_j X_{t-j} + \epsilon_t, \quad \epsilon_t \sim N(0, \tau^2 \mathbf{I}_n) \\
\theta_t &= G \theta_{t-1} + \omega_t, \quad \omega_t \sim N(0, W_t),
\end{aligned} \tag{3.4}$$

onde F_t representa r possíveis covariáveis e/ou fatores sazonais que possam influenciar a média do processo Y_t , com os respectivos coeficientes variando ao longo do tempo θ_t , K_t é a decomposição de Cholesky de uma matriz de correlação espacial, X_t um processo autoregressivo de ordem p , $AR(p)$, tal que $cov(X(t, s), X(t', s')) = 0$ se $s \neq s'$. Assumimos uma estrutura dinâmica para os coeficientes θ_t , onde G é uma matriz ($r \times r$) representando a evolução dinâmica destes coeficientes e W é uma matriz de covariância diagonal representando a matriz de evolução, $W = \text{diag}\{\tau_1^2, \dots, \tau_r^2\}$. Por exemplo, suponha que F_t represente covariáveis, então podemos pensar na matriz de evolução como sendo $G = \mathbf{I}_r$, isto é, uma matriz identidade de ordem r . Se F_t representar fatores sazonais, a matriz de evolução pode ter uma representação de Fourier, isto é, G é uma matriz bloco diagonal, onde cada bloco corresponde a um harmônico e o bloco associado ao r -ésimo harmônico é dado por

$$G_r = \begin{pmatrix} \cos(2\pi r/f) & \text{sen}(2\pi r/f) \\ -\text{sen}(2\pi r/f) & \cos(2\pi r/f) \end{pmatrix},$$

onde f denota o período e $r = 1, \dots, m$.

Como discutido na Seção anterior, podemos escrever Y_t como

$$\begin{aligned}
Y_t &= F_t' \theta_t + \sum_{j=1}^p \phi_j K_t X_{t-j} + K_t \epsilon_t \\
&= F_t' \theta_t + \sum_{j=1}^p \phi_j K_t K_{t-j}^{-1} (Y_{t-j} - F_{t-j}' \theta_{t-j}) + v_t \\
&= F_t' \theta_t + \sum_{j=1}^p \nu_{jt} (Y_{t-j} - F_{t-j}' \theta_{t-j}) + v_t,
\end{aligned}$$

onde $\nu_{jt} = \phi_j K_t K_{t-j}^{-1}$, $v_t = K_t \epsilon_t$ e $v_t \sim N(0, \tau^2 \Sigma_t)$ com $\Sigma_t = K_t' K_t$. Então condicional às p primeiras observações, temos que $Y_t | Y_1, \dots, Y_p \sim N_n \left(F_t' \theta_t + \sum_{j=1}^p \nu_{jt} (Y_{t-j} - F_{t-j}' \theta_{t-j}), \tau^2 \Sigma_t \right)$, $t = p+1, \dots, T$. Observe que a média do processo no tempo t é dada pela soma dos efeitos das covariáveis e/ou fatores sazonais no tempo t , mais a soma das diferenças entre as observações e os efeitos das covariáveis

e/ou fatores sazonais observados em p tempos anteriores. Já a variância do processo, é dada pelo produto entre o parâmetro do processo $\text{AR}(p)$, τ^2 , e a função de covariância do processo espacial, isto é, $\Sigma_t = \sigma^2 \rho(s - s'; \lambda_t)$.

Condicional a Y_1, \dots, Y_p , a verossimilhança é dada por

$$\left(\prod_{t=p+1}^T \frac{|\Sigma_t|^{-1/2}}{(\tau^2)^{n/2}} \right) \exp \left\{ -\frac{1}{2\tau^2} \sum_{t=p+1}^T \left(Y_t^* - \sum_{j=1}^p \nu_{jt} Y_{t-j}^* \right)' \Sigma_t^{-1} \left(Y_t^* - \sum_{j=1}^p \nu_{jt} Y_{t-j}^* \right) \right\}, \quad (3.5)$$

onde $Y_t^* = Y_t - F_t' \theta_t$. Como discutimos na Seção 3.2 a estrutura de covariância do processo é não separável.

Várias funções de correlação espaciais podem ser utilizadas, tais como as funções de correlação Exponencial, Gaussiana e a Matérn. Consideramos quatro diferentes estruturas para a função de covariância

- $\Sigma_t = \sigma^2 \rho(s - s'; \lambda_t)$, ou seja, diferentes decaimentos para diferentes tempos;
- $\Sigma_t = \sigma_t^2 \rho(s - s'; \lambda)$, isto é, a variância do processo evolui suavemente ao longo do tempo;
- $\Sigma_t = \sigma_t^2 \rho(s - s'; \lambda_t)$, ambos, variância e decaimento, variando ao longo do tempo.
- $\Sigma = \sigma^2 \rho(s - s'; \lambda)$, ambos, variância e decaimento, fixos no tempo.

Note que cada uma dessas hipóteses resulta em diferentes estruturas de covariância. As três primeiras resultam em estruturas de covariância não separável, e a última em uma estrutura separável. As quatro abordagens serão consideradas nas aplicações analisadas na Seção 3.7.

3.4 Especificação das Distribuições a Priori

O modelo em (3.4) contém os parâmetros da estrutura de média, os coeficientes $\theta = (\theta_1, \dots, \theta_T)$, os parâmetros do processo $\text{AR}(p)$, $\Phi = (\phi_1, \dots, \phi_p)$ e τ^2 e, finalmente, os parâmetros da função de covariância espacial, σ^2 e $\Lambda = (\lambda_1, \dots, \lambda_T)$. Como a verossimilhança é função do produto $\sigma^2 \tau^2$ e como mencionado anteriormente apenas o produto é identificável, decidimos por fixar $\tau^2 = 1$. O procedimento de inferência destes modelos é feito sob o enfoque bayesiano. Portanto, o modelo estará completamente especificado após atribuir distribuições a priori para os parâmetros do modelo. O vetor paramétrico deste modelo pode ser escrito como $\Theta = (\Phi, \sigma^2, \Lambda, \theta)$.

Inicialmente precisamos atribuir distribuições a priori para os parâmetros da componente $AR(p)$. Com o objetivo de garantir estacionariedade do processo, os coeficientes ϕ_1, \dots, ϕ_p devem ser restritos a um espaço que é muito difícil especificar. Prioris adequadas para os coeficientes do $AR(p)$ garantem a estacionariedade do processo. Neste capítulo trataremos apenas do caso em que $p = 1$. Neste caso, atribuímos uma priori uniforme, isto é, $\phi \sim U(-1, 1)$, garantindo a estacionariedade do processo. No capítulo seguinte, discutimos com detalhes a especificação da priori para o caso de um $AR(p)$.

Considerando que o decaimento da função de correlação espacial muda suavemente ao longo do tempo, podemos usar uma evolução para λ_t da seguinte maneira

$$\log(\lambda_t) = \log(\lambda_{t-1}) + \eta_t, \quad \eta_t \sim N(0, \delta\sigma_\lambda^2), \quad (3.6)$$

com, $\lambda_1 \sim LN(\mu_\lambda, \sigma_\lambda^2)$ e $\lambda_t | \lambda_{t-1} \sim LN(\lambda_{t-1}, \delta\sigma_\lambda^2)$, $t = 2, \dots, T$ com $LN(a, b)$ representando a distribuição lognormal, cuja normal associada tem média a e variância b .

Ou ainda, uma evolução baseada na conjugação entre a beta e a gama, de modo que se

$$\lambda_t = \frac{\eta_t}{\delta} \lambda_{t-1} \quad \text{e} \quad \eta_t \sim Be\left(\frac{\delta a}{2}, \frac{(1-\delta)a}{2}\right), \quad a > 0 \quad \text{e} \quad 0 < \delta < 1. \quad (3.7)$$

Note que nas duas distribuições δ representa o fator de desconto que mede a quantidade de informação perdida entre os tempos t e $t - 1$. Assumindo $\lambda_{t-1} \sim Ga(\frac{a}{2}, \frac{b}{2})$ temos que $\lambda_t \sim Ga(\frac{\delta a}{2}, \frac{\delta b}{2})$.

No caso em que consideramos a variância do processo espacial variando ao longo do tempo, podemos atribuir essas mesmas prioris para evolução, já que a lognormal e a gama são distribuições definidas num espaço paramétrico positivo. Quando o parâmetro λ é fixo podemos atribuir $\lambda \sim LN(\mu_\lambda, \sigma_\lambda^2)$ ou $\lambda \sim Ga(\frac{a}{2}, \frac{b}{2})$. Já quando σ^2 é considerado fixo, atribuímos uma priori inversa gama com parâmetros conhecidos a, b , $\sigma^2 \sim IG(a, b)$.

Seguindo os MLD's devemos especificar uma priori para θ_0 , que assumimos seguir uma distribuição normal, com vetor de médias m_0 e matriz de covariância C_0 conhecidas.

3.5 Procedimento de Inferência

Assumindo independência a priori dos parâmetros do modelo e considerando a função de verossimilhança em (3.5); segundo o teorema de Bayes a posteriori é proporcional ao produto da verossimilhança pela priori, assim a distribuição a posteriori conjunta dos parâmetros do modelo é proporcional

a

$$p(\Theta|y) \propto \left[\prod_{t=p+1}^T p(y_t|y_{t-p}, \dots, y_{t-1}, \Theta) \right] \left[\prod_{t=1}^T p(\theta_t|\theta_{t-1}) \right] p(\theta_0) \left[\prod_{t=2}^T p(\lambda_t|\lambda_{t-1}) \right] p(\lambda_1) \times p(\Phi) p(\sigma^2). \quad (3.8)$$

Como podemos ver, a distribuição a posteriori dos parâmetros do modelo não tem forma analítica fechada e, portanto, precisamos de métodos de simulação estocástica para obter amostras desta distribuição. Para isso são utilizados os métodos de Monte Carlo via Cadeias de Markov (MCMC), tais como o amostrador de Gibbs e Metropolis-Hastings (Gamerman e Lopes 2006). Mais especificamente, utilizaremos o amostrador de Gibbs com alguns passos do algoritmo Metropolis-Hastings. Para o uso do amostrador de Gibbs precisamos calcular as condicionais completas para as componentes do vetor paramétrico Θ . Quando as condicionais completas têm forma conhecida sorteamos da respectiva distribuição. Caso contrário, fazemos uso do algoritmo Metropolis-Hastings. Além disso, podemos reescrever o modelo na forma dos modelos lineares dinâmicos, com o objetivo de utilizar métodos eficientes para obter amostras da distribuição a posteriori dos coeficientes $(\theta_1, \dots, \theta_T)$. Na subseção seguinte mostraremos as condicionais completas resultantes e o respectivo esquema de amostragem para obtenção de amostras da distribuição a posteriori, $p(\Theta|y)$.

3.5.1 Distribuições Condicionais Completas a Posteriori

Para que o algoritmo MCMC seja implementado, precisamos calcular as distribuições condicionais completas a posteriori de cada um dos parâmetros. A seguir mostraremos essas distribuições. Denotemos por $p(\psi|.)$ a distribuição condicional completa do parâmetro ψ .

Inicialmente, vamos supor θ_t conhecido para cada tempo t e definir $Y_t^* = Y_t - F_t'\theta_t$. A distribuição condicional completa de Φ , σ^2 e Λ é proporcional a verossimilhança em (3.5) multiplicada pela respectiva priori.

Amostrando ϕ

Assumindo $p = 1$, a distribuição condicional completa para ϕ é uma distribuição normal truncada, isto é

$$\phi|. \sim N \left[\left(\sum_{t=2}^T Y_t^* \Sigma_t^{-1} m_t \right) \left(\sum_{t=2}^T m_t \Sigma_t^{-1} m_t \right)^{-1}, \left(\sum_{t=2}^T m_t \Sigma_t^{-1} m_t \right)^{-1} \right] I(\phi)_{(-1,1)},$$

onde $m_t = K_t K_{t-1}^{-1} Y_{t-1}^*$ e $I(\phi)_{(-1,1)}$ é a variável indicadora, tal que,

$$I(\phi)_{(-1,1)} = \begin{cases} 1, & -1 < \phi < 1, \\ 0, & \text{caso contrário.} \end{cases}$$

Amostrando σ^2

A distribuição condicional completa para σ^2 quando considerada fixa tem forma conhecida, é uma distribuição gama invertida, isto é

$$\sigma^2|. \sim IG \left[\frac{n(T-p)}{2} + a, \frac{1}{2} \sum_{t=p+1}^T \left(Y_t^* - \sum_{i=1}^p \nu_{it} Y_{t-i}^* \right)' V_t^{-1} \left(Y_t^* - \sum_{i=1}^p \nu_{it} Y_{t-i}^* \right) + b \right]$$

onde $V_t = \eta(s - s'; \lambda_t)$.

Amostrando Λ

Observamos que a verossimilhança é função de λ_t , $t = 1, \dots, T$. A distribuição condicional completa para λ_t não corresponde a nenhuma forma de densidade conhecida. Assim, o algoritmo de Metropolis-Hastings será utilizado para obter amostras da posteriori. No algoritmo de Metropolis-Hastings precisamos definir a proposta de transição e a probabilidade de aceitação. O algoritmo tem os seguintes passos:

1. Sorteia um valor proposto λ_t^p , da distribuição proposta $q(\lambda_t^p | \lambda_t^c)$ centrada no valor corrente λ_t^c ;
2. Avalia a probabilidade de transição dada por

$$A = \min \left\{ 1, \frac{p(\lambda_t^p | \cdot) q(\lambda_t^c | \lambda_t^p)}{p(\lambda_t^c | \cdot) q(\lambda_t^p | \lambda_t^c)} \right\},$$

onde $p(\cdot|\cdot)$ representa a distribuição condicional completa e $q(\cdot|\cdot)$ a distribuição proposta de transição;

3. Gera uma quantidade uniforme u . Se $A < u$, aceitamos o valor proposto.

Utilizamos como proposta $\lambda_t^p \sim LN(\lambda_t^c, U)$, isto é uma distribuição lognormal cuja normal associada está centrada no valor corrente com variância U . A razão de Metropolis-Hastings envolvida na probabilidade de transição é calculada na escala log para evitar problemas numéricos. A seguir apresentamos a distribuição condicional completa, bem como a razão de Metropolis-Hastings usando as distribuições a priori lognormal e gama. As distribuições condicionais completas apresentadas aqui foram calculadas para λ_t . Quando usamos a variância do processo espacial variando suavemente ao longo do tempo, σ_t^2 , podemos atribuir as mesmas prioris. Assim, resultados similares são obtidos, trocando apenas λ_t por σ_t^2 nas respectivas distribuições condicionais completas.

1. Priori Lognormal

A distribuição condicional completa para λ_t usando a priori lognormal está dividida em três diferentes estruturas: a condicional completa para λ_1 , para $\lambda_2, \dots, \lambda_{T-1}$ e, finalmente, para λ_T , pois

$$\lambda_1 \sim LN(\mu_\lambda, \sigma_\lambda^2), \quad \log(\lambda_t) = \log(\lambda_{t-1}) + \eta_t, \quad \eta_t \sim N(0, \delta\sigma_\lambda^2), \quad t = 2, \dots, T.$$

• Distribuição condicional completa para λ_1

$$p(\lambda_1|\cdot) \propto |\Sigma_{p+1}|^{-1/2} \exp \left\{ -\frac{1}{2} \left(Y_{p+1}^* - \sum_{i=1}^p \nu_{ip+1} Y_{p+1-i}^* \right)' \Sigma_{p+1}^{-1} \left(Y_{p+1}^* - \sum_{i=1}^p \nu_{ip+1} Y_{p+1-i}^* \right) \right\} \\ \exp \left\{ -\frac{1}{2\sigma_\lambda^2} (\log \lambda_1 - \mu_\lambda)^2 \right\} \frac{1}{\lambda_1} \exp \left\{ -\frac{1}{2\delta\sigma_\lambda^2} (\log \lambda_2 - \log \lambda_1)^2 \right\}.$$

Após algumas simplificações, a razão de Metropolis-Hastings é igual a,

$$\frac{p(\lambda_1^p|\cdot)q(\lambda_1^c|\lambda_1^p)}{p(\lambda_1^c|\cdot)q(\lambda_1^p|\lambda_1^c)} = \frac{L(\lambda_1^p) \exp \left\{ -\frac{1}{2\sigma_\lambda^2} (\log \lambda_1^p - \mu_\lambda)^2 \right\} \exp \left\{ -\frac{1}{2\delta\sigma_\lambda^2} (\log \lambda_2 - \log \lambda_1^p)^2 \right\}}{L(\lambda_1^c) \exp \left\{ -\frac{1}{2\sigma_\lambda^2} (\log \lambda_1^c - \mu_\lambda)^2 \right\} \exp \left\{ -\frac{1}{2\delta\sigma_\lambda^2} (\log \lambda_2 - \log \lambda_1^c)^2 \right\}},$$

$$\text{onde } L(\lambda_1) = |\Sigma_{p+1}|^{-1/2} \exp \left\{ -\frac{1}{2} \left(Y_{p+1}^* - \sum_{j=1}^p \nu_{jp+1} Y_{p+1-j}^* \right)' \Sigma_{p+1}^{-1} \left(Y_{p+1}^* - \sum_{j=1}^p \nu_{jp+1} Y_{p+1-j}^* \right) \right\}.$$

- **Distribuição condicional completa para λ_t , $t \in \{2, \dots, T-1\}$**

$$p(\lambda_t | \cdot) \propto |\Sigma_t|^{-1/2} \exp \left\{ -\frac{1}{2} \left(Y_t^* - \sum_{i=1}^p \nu_{it} Y_{t-i}^* \right)' \Sigma_t^{-1} \left(Y_t^* - \sum_{i=1}^p \nu_{it} Y_{t-i}^* \right) \right\} \\ \exp \left\{ -\frac{1}{2\delta\sigma_\lambda^2} (\log \lambda_t - \log \lambda_{t-1})^2 \right\} \frac{1}{\lambda_t} \exp \left\{ -\frac{1}{2\delta\sigma_\lambda^2} (\log \lambda_{t+1} - \log \lambda_t)^2 \right\}.$$

Após algumas simplificações, a razão de Metropolis-Hastings é igual a,

$$\frac{p(\lambda_t^p | \cdot) q(\lambda_t^c | \lambda_t^p)}{p(\lambda_t^c | \cdot) q(\lambda_t^p | \lambda_t^c)} = \frac{L(\lambda_t^p) \exp \left\{ -\frac{1}{2\delta\sigma_\lambda^2} (\log \lambda_t^p - \log \lambda_{t-1})^2 \right\} \exp \left\{ -\frac{1}{2\delta\sigma_\lambda^2} (\log \lambda_{t+1} - \log \lambda_t^p)^2 \right\}}{L(\lambda_t^c) \exp \left\{ -\frac{1}{2\delta\sigma_\lambda^2} (\log \lambda_t^c - \log \lambda_{t-1})^2 \right\} \exp \left\{ -\frac{1}{2\delta\sigma_\lambda^2} (\log \lambda_{t+1} - \log \lambda_t^c)^2 \right\}},$$

onde $L(\lambda_t) = |\Sigma_t|^{-1/2} \exp \left\{ -\frac{1}{2} \left(Y_t^* - \sum_{i=1}^p \nu_{it} Y_{t-i}^* \right)' \Sigma_t^{-1} \left(Y_t^* - \sum_{i=1}^p \nu_{it} Y_{t-i}^* \right) \right\}$.

- **Distribuição condicional completa para λ_T**

$$p(\lambda_T | \cdot) \propto |\Sigma_T|^{-1/2} \exp \left\{ -\frac{1}{2} \left(Y_T^* - \sum_{i=1}^p \nu_{iT} Y_{T-i}^* \right)' \Sigma_T^{-1} \left(Y_T^* - \sum_{i=1}^p \nu_{iT} Y_{T-i}^* \right) \right\} \\ \exp \left\{ -\frac{1}{2\delta\sigma_\lambda^2} (\log \lambda_T - \log \lambda_{T-1})^2 \right\} \frac{1}{\lambda_T}.$$

Após algumas simplificações, a razão de Metropolis-Hastings é igual a,

$$\frac{p(\lambda_T^p | \cdot) q(\lambda_T^c | \lambda_T^p)}{p(\lambda_T^c | \cdot) q(\lambda_T^p | \lambda_T^c)} = \frac{L(\lambda_T^p) \exp \left\{ -\frac{1}{2\delta\sigma_\lambda^2} (\log \lambda_T^p - \log \lambda_{T-1})^2 \right\}}{L(\lambda_T^c) \exp \left\{ -\frac{1}{2\delta\sigma_\lambda^2} (\log \lambda_T^c - \log \lambda_{T-1})^2 \right\}},$$

onde $L(\lambda_T) = |\Sigma_T|^{-1/2} \exp \left\{ -\frac{1}{2} \left(Y_T^* - \sum_{i=1}^p \nu_{iT} Y_{T-i}^* \right)' \Sigma_T^{-1} \left(Y_T^* - \sum_{i=1}^p \nu_{iT} Y_{T-i}^* \right) \right\}$.

2. Priori Gama

A distribuição condicional completa para λ_t usando a priori gama está dividida em duas diferentes estruturas: a condicional completa para λ_1 e para $\lambda_2, \dots, \lambda_T$, já que,

$$\lambda_1 \sim Ga \left(\frac{a_\lambda}{2}, \frac{b_\lambda}{2} \right), \quad \lambda_t \sim Ga \left(\frac{\delta a_\lambda}{2}, \frac{\delta b_\lambda}{2} \right), \quad t = 2, \dots, T.$$

- **Distribuição condicional completa para λ_1**

$$p(\lambda_1|\cdot) \propto |\Sigma_{p+1}|^{-1/2} \exp \left\{ -\frac{1}{2} \left(Y_{p+1}^* - \sum_{i=1}^p \nu_{ip+1} Y_{p+1-i}^* \right)' \Sigma_{p+1}^{-1} \left(Y_{p+1}^* - \sum_{i=1}^p \nu_{ip+1} Y_{p+1-i}^* \right) \right\} \\ \lambda_1^{\frac{a\lambda}{2}-1} \exp \left\{ -\frac{b\lambda}{2} \lambda_1 \right\}.$$

Após algumas simplificações, a razão de Metropolis-Hastings é igual a,

$$\frac{p(\lambda_1^p|\cdot)q(\lambda_1^c|\lambda_1^p)}{p(\lambda_1^c|\cdot)q(\lambda_1^p|\lambda_1^c)} = \frac{L(\lambda_1^p)\lambda_1^{p\frac{a\lambda}{2}-1} \exp \left\{ -\frac{b\lambda}{2} \lambda_1^p \right\} \frac{1}{\lambda_1^c}}{L(\lambda_1^c)\lambda_1^{c\frac{a\lambda}{2}-1} \exp \left\{ -\frac{b\lambda}{2} \lambda_1^c \right\} \frac{1}{\lambda_1^p}}.$$

- **Distribuição condicional completa para $\lambda_2, \dots, \lambda_T$**

$$p(\lambda_t|\cdot) \propto |\Sigma_t|^{-1/2} \exp \left\{ -\frac{1}{2} \left(Y_t^* - \sum_{i=1}^p \nu_{it} Y_{t-i}^* \right)' \Sigma_t^{-1} \left(Y_t^* - \sum_{i=1}^p \nu_{it} Y_{t-i}^* \right) \right\} \\ \lambda_t^{\frac{\delta a\lambda}{2}-1} \exp \left\{ -\frac{\delta b\lambda}{2} \lambda_t \right\}.$$

Após algumas simplificações, a razão de Metropolis-Hastings é igual a,

$$\frac{p(\lambda_t^p|\cdot)q(\lambda_t^c|\lambda_t^p)}{p(\lambda_t^c|\cdot)q(\lambda_t^p|\lambda_t^c)} = \frac{L(\lambda_t^p)\lambda_t^{p\frac{\delta a\lambda}{2}-1} \exp \left\{ -\frac{\delta b\lambda}{2} \lambda_t^p \right\} \frac{1}{\lambda_t^c}}{L(\lambda_t^c)\lambda_t^{c\frac{\delta a\lambda}{2}-1} \exp \left\{ -\frac{\delta b\lambda}{2} \lambda_t^c \right\} \frac{1}{\lambda_t^p}}.$$

Amostrando o vetor de estados $\theta_1, \dots, \theta_T$

Para amostrar o vetor de estados, reescrevemos o modelo proposto na forma dos Modelos Lineares Dinâmicos (MLD), com o objetivo de utilizar métodos eficientes de amostragem. Assim, condicional a Φ , σ^2 e Λ , e definindo $Z_t = Y_t - \sum_{j=1}^p \nu_{jt} Y_{t-j}$, temos que

$$Z_t = \sum_{j=0}^p -\nu_{jt} F'_{t-j} \theta_{t-j} + v_t, \quad v_t \sim N(0, \Sigma_t),$$

onde $\nu_{0t} = -I$. Reescrevendo o modelo na classe dos MLD's, temos

$$Z_t = F_t^* \theta_t^* + v_t, \quad v_t \sim N(0, \Sigma_t) \\ \theta_t^* = G^* \theta_{t-1}^* + \omega_t, \quad \omega_t \sim N(0, W_t^*)$$

onde o vetor de estados a cada tempo t é $\theta_t^* = (\theta_t, \dots, \theta_{t-p})'$, com dimensão $r(p+1)$. Lembrando que p denota a ordem do modelo, r o número de covariáveis e fatores sazonais e n o número de estações monitoradoras. Como mostrado na Seção 2.2.1, devemos especificar a quádrupla $\{F_t^*, G_t^*, V_t^*, W_t^*\}$. A matriz de regressão dinâmica F_t^* é a concatenação de $(p+1)$ matrizes de dimensão $(n \times r)$. Cada matriz é dada por $-\nu_{jt}F'_{t-j}$ com $j = 0, \dots, p$. A variância do sistema é a matriz de covariância espacial, isto é, $V_t^* = \Sigma_t$, com dimensão $n \times n$. A matriz de evolução G^* e a variância de evolução W_t^* tem dimensão $(r(p+1) \times r(p+1))$. Observe que neste caso, o vetor de estados a cada tempo t , θ_t^* engloba também o vetor de estados originais, θ_t , em p tempos anteriores. Portanto não é necessário amostrar o vetor de estados θ_t^* inteiro a cada iteração. Por isso, as matrizes G^* e W_t^* devem ser definidas, de maneira a considerar esta parametrização. Para exemplificar, vamos supor $p = 1$. Assim,

$$Z_t = \begin{pmatrix} F_t' & -\nu_{1t}F'_{t-1} \end{pmatrix} \begin{pmatrix} \theta_t \\ \theta_{t-1} \end{pmatrix} + v_t, \quad v_t \sim N(0, \Sigma_t),$$

$$\begin{pmatrix} \theta_t \\ \theta_{t-1} \end{pmatrix} = \begin{pmatrix} G & 0_r \\ I_r & 0_r \end{pmatrix} \begin{pmatrix} \theta_{t-1} \\ \theta_{t-2} \end{pmatrix} + \begin{pmatrix} \eta_{t1} \\ \eta_{t2} \end{pmatrix}, \quad \begin{pmatrix} \eta_{t1} \\ \eta_{t2} \end{pmatrix} \sim N \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} W_t & 0 \\ 0 & 0 \end{pmatrix} \right].$$

Após especificar as componentes dos MLD's, utilizaremos o algoritmo FFBS descrito na Seção 2.2.1 para sortear os θ_t 's de maneira eficiente.

Amostrando as observações faltantes

Do ponto de vista bayesiano, as observações faltantes são consideradas parâmetros do modelo e, portanto, também precisam ser estimadas. Seja $Y_t = (Y_t^u, Y_t^g)'$ a partição do vetor Y para cada tempo t , onde Y_t^u é um vetor de dimensão n_{ut} que contém as observações faltantes no tempo t e Y_t^g um vetor de dimensão n_{gt} que contém as observações no tempo t , tal que $n_{ut} + n_{gt} = n, \forall t$. Assim,

$$\begin{pmatrix} Y_t^u \\ Y_t^g \end{pmatrix} \sim N_n \left[\begin{pmatrix} \mu_t^u \\ \mu_t^g \end{pmatrix}; \begin{pmatrix} \Sigma_t^u & \Sigma_t^{ug} \\ \Sigma_t^{gu} & \Sigma_t^g \end{pmatrix} \right]$$

onde μ_t^u e μ_t^g denotam as médias dos vetores Y_t^u e Y_t^g , respectivamente. A matriz de covariância Σ_t^u representa a covariância entre os elementos de Y_t^u , Σ_t^g a covariância entre os elementos de Y_t^g . Finalmente, Σ_t^{ug} e Σ_t^{gu} denotam as covariâncias entre os elementos de Y_t^u e Y_t^g . A partir da partição

da normal multivariada (Anderson 1994), temos:

$$Y_t^u | Y_t^g \sim N_{n_{ut}}(\mu_t^u + \Sigma_t^{ug}(\Sigma_t^g)^{-1}(Y_t^g - \mu_t^g); \Sigma_t^u - \Sigma_t^{ug}(\Sigma_t^g)^{-1}\Sigma_t^{gu}).$$

Agora vamos calcular a distribuição condicional completa que está dividida em três estruturas: Y_1 , (Y_2, \dots, Y_{T-1}) e Y_T . Para simplificar a notação vamos denotar $\mu_t = F_t' \theta_t$ e $M_t = K_t K_{t-1}^{-1}$, $t = 1, \dots, T$, assim obtemos:

- Distribuição condicional completa para Y_1 :

$$Y_1 | \cdot \sim N_n \left[\left(\phi M_2' \Sigma_2^{-1} (Y_2 - \mu_2) + \phi^2 M_2' \Sigma_2^{-1} M_2 \mu_1 \right) \left(\phi^2 M_2' \Sigma_2^{-1} M_2 \right)^{-1}; \left(\phi^2 M_2' \Sigma_2^{-1} M_2 \right)^{-1} \right].$$

- Distribuição condicional completa para Y_t , $t \in \{2, \dots, T-1\}$:

$$Y_t | \cdot \sim N_n \left[\left(\Sigma_t^{-1} \left(\phi M_t' (Y_{t-1} - \mu_{t-1}) + \mu_t \right) + \Sigma_{t+1}^{-1} \left(\phi M_{t+1}' (Y_{t+1} - \mu_{t+1}) + \phi^2 M_{t+1}' M_{t+1} \mu_t \right) \right) \left(\Sigma_t^{-1} + \phi^2 M_{t+1}' \Sigma_{t+1}^{-1} M_{t+1} \right)^{-1}; \left(\Sigma_t^{-1} + \phi^2 M_{t+1}' \Sigma_{t+1}^{-1} M_{t+1} \right)^{-1} \right].$$

- Distribuição condicional completa para Y_T :

$$Y_T | \cdot \sim N_n \left[\mu_T + \phi M_T (Y_{T-1} - \mu_{T-1}); \Sigma_T^{-1} \right].$$

3.6 Previsão Temporal e Interpolação Espacial

Previsões temporais para k passos a frente e interpolações espaciais para localizações não medidas são de grande interesse na prática e podem ser obtidas a partir da distribuição preditiva.

Suponha que Y_t^u denote o vetor das observações em localizações não medidas e Y_t o vetor das observações em localizações medidas a cada tempo t e seja $Y_{1:t} = (Y_1, \dots, Y_t)$. Seguindo o modelo em (3.4), a distribuição conjunta de (Y_t, Y_t^u) , condicional ao vetor paramétrico Θ , e às p observações anteriores, é uma distribuição normal multivariada com vetor de médias igual a $(F_t^y \theta_t + \sum_{j=1}^p \nu_{jt} (Y_{t-j} - F_{t-j}^y \theta_{t-j}), F_t^u \theta_t + \sum_{j=1}^p \nu_{jt} (Y_{t-j}^u - F_{t-j}^u \theta_{t-j}))'$ correspondendo as médias do processo para as observações medidas e não medidas, onde F_t^y e F_t^u são as matrizes de regressão dinâmica das observações medidas e não medidas, respectivamente. A matriz de covariância é dividida em 4 blocos, Σ_t^y representando as covariâncias entre as observações medidas, Σ_t^{yu} e Σ_t^{uy} , as covariâncias entre as observações medidas e não medidas, e Σ_t^u a covariância entre as observações não medidas.

A distribuição preditiva é calculada pela seguinte integral

$$p(Y_t^u|Y_t) = \int p(Y_t^u|\Theta, Y_t)p(\Theta|Y_t)d\Theta, \quad (3.9)$$

onde $p(\Theta|Y_t)$ é a distribuição a posteriori especificada em (3.8) e $p(Y_t^u|\Theta, Y_t)$ é calculada a partir da partição da normal multivariada (Anderson 1994), isto é,

$$Y_t^u|Y_t, \Theta \sim N[\mu_t^u + \Sigma_t^{uy}(\Sigma_t^y)^{-1}(Y_t - \mu_t^y), (\Sigma_t^u - \Sigma_t^{uy}(\Sigma_t^y)^{-1}\Sigma_t^{yu})], \quad (3.10)$$

onde $\mu_t^y = F_t^y\theta_t + \sum_{j=1}^p \nu_{jt}(Y_{t-j} - F_{t-j}^y\theta_{t-j})$ e $\mu_t^u = F_t^u\theta_t + \sum_{j=1}^p \nu_{jt}(Y_{t-j}^u - F_{t-j}^u\theta_{t-j})$ e a integral em (3.9) pode ser aproximada por Monte Carlo (Gamerman e Lopes (2006)), isto é,

$$p(Y_t^u|Y_t) \approx \frac{1}{M} \sum_{m=1}^M p(Y_t^u|\Theta^m, Y_t),$$

onde M denota o total de amostras da distribuição a posteriori do vetor paramétrico Θ , e Θ^m denota a m -ésima amostra da distribuição a posteriori do vetor paramétrico Θ .

Por outro lado, para fazer previsão temporal para k passos a frente é necessário avaliar a seguinte distribuição preditiva.

$$\begin{aligned} p(Y_{T+k}|Y_{1:T}) &= \int p(Y_{T+k}|\mathbf{F}_{T+k}, Y_{T+k-1}, \theta_{T+k}, \phi, \sigma^2, \Sigma_{T+k})p(\theta_{T+k} | \theta_{T+k-1}, W) \\ & p(\Sigma_{T+k}|\Sigma_{T+k-1})p(Y_{T+k-1}|\mathbf{F}_{T+k-1}, Y_{T+k-2}, \phi, \sigma^2, \theta_{T+k-1}, \Sigma_{T+k-1})p(\theta_{T+k-1} | \theta_{T+k-2}, W) \\ & p(\Sigma_{T+k-1}|\Sigma_{T+k-2}) \dots p(\phi, \sigma^2, \theta_1, \dots, \theta_T, \Sigma_1, \dots, \Sigma_T|Y_{1:T})d(\phi, \sigma^2, \theta_1, \dots, \theta_T, W, \Sigma_1, \dots, \Sigma_T), \end{aligned}$$

onde, para simplificar notação, assumimos $p = 1$.

A integral acima pode ser aproximada por

$$p(Y_{T+k}|Y_{1:T}) \approx \frac{1}{M} \sum_{m=1}^M N_n(Y_{T+k}|\mu_{T+k}^{(m)}, \Sigma_{T+k}^{(m)}).$$

onde $\mu_{T+k} = \mathbf{F}_{T+k}\theta_{T+k} + \nu_{1T+k}(Y_{T+k-1} - \mathbf{F}_{T+k-1}\theta_{T+k-1})$. Observe que $Y_{T+i}^{(m)}$ e $\Sigma_{T+i}^{(m)}$, $i > 0$ são obtidos através das equações de evolução de Y_t e Σ_t , respectivamente, descritas através do modelo em (3.4).

3.7 Ajustando o Modelo de Convolução Espacial

Nesta seção ajustamos o modelo de convolução espacial a quatro conjuntos de dados. Inicialmente, em dois conjuntos de dados artificiais, obtidos a partir do modelo, onde os verdadeiros valores dos

parâmetros são conhecidos. Portanto, podemos verificar o comportamento do modelo e do algoritmo de simulação proposto. No primeiro, simulamos um modelo cuja estrutura de média possui covariáveis e no segundo, um modelo cuja média é descrita por fatores sazonais. Em seguida, o modelo é ajustado a um conjunto de dados de concentrações diárias de material particulado (PM_{10}) em 18 estações monitoradoras localizadas na região metropolitana do Rio de Janeiro ao longo de 46 observações no ano de 2002. Finalmente, ajustamos o modelo proposto a séries temporais de temperatura em 23 estações no estado da Califórnia medidas entre janeiro de 1992 e dezembro de 2002.

Diferentes especificações para variância das observações, Σ_t , são investigadas. Assim como, diferentes priors para seus parâmetros. Seja V a função de correlação com parâmetro λ , tal que $V = \eta(\|s - s'\|; \lambda)$, onde $\|s - s'\|$ representa a distância euclidiana entre as localizações s e s' . A Tabela 3.1 a seguir mostra o resumo dos modelos que são ajustados para os quatro conjuntos de dados:

Modelos	Σ_t	Priori	
		Lognormal	Gama
Modelo 1	$\sigma^2 V$	$\lambda \sim LN(\mu_\lambda, \sigma_\lambda^2)$	$\lambda \sim Ga(\frac{a_\lambda}{2}, \frac{b_\lambda}{2})$
Modelo 2	$\sigma^2 V_t$	$\lambda_1 \sim LN(\mu_\lambda, \sigma_\lambda^2)$ $\log(\lambda_t) = \log(\lambda_{t-1}) + \eta_t, \eta_t \sim N(0, \delta\sigma_\lambda^2)$	$\lambda_1 \sim Ga(\frac{a_\lambda}{2}, \frac{b_\lambda}{2})$ $\lambda_t \sim Ga(\frac{\delta a_\lambda}{2}, \frac{\delta b_\lambda}{2})$
Modelo 3	$\sigma_t^2 V$	$\sigma_1^2 \sim LN(\mu_\sigma, \tau_\sigma^2)$ $\log(\sigma_t^2) = \log(\sigma_{t-1}^2) + \eta_t, \eta_t \sim N(0, \delta\tau_\sigma^2)$	$\sigma_1^2 \sim Ga(\frac{a_\sigma}{2}, \frac{b_\sigma}{2})$ $\sigma_t^2 \sim Ga(\frac{\delta a_\sigma}{2}, \frac{\delta b_\sigma}{2})$
Modelo 4	$\sigma_t^2 V_t$	$\lambda_1 \sim LN(\mu_\lambda, \sigma_\lambda^2)$ $\log(\lambda_t) = \log(\lambda_{t-1}) + \eta_t, \eta_t \sim N(0, \delta\sigma_\lambda^2)$ $\sigma_1^2 \sim LN(\mu_\sigma, \tau_\sigma^2)$ $\log(\sigma_t^2) = \log(\sigma_{t-1}^2) + \eta_t, \eta_t \sim N(0, \delta\tau_\sigma^2)$	$\lambda_1 \sim Ga(\frac{a_\lambda}{2}, \frac{b_\lambda}{2})$ $\lambda_t \sim Ga(\frac{\delta a_\lambda}{2}, \frac{\delta b_\lambda}{2})$ $\sigma_1^2 \sim Ga(\frac{a_\sigma}{2}, \frac{b_\sigma}{2})$ $\sigma_t^2 \sim Ga(\frac{\delta a_\sigma}{2}, \frac{\delta b_\sigma}{2})$

Tabela 3.1: Especificação dos modelos ajustados ao longo deste capítulo.

3.7.1 Dados Artificiais com Covariáveis

Geramos um conjunto de observações para 30 localizações ao longo de 50 instantes de tempo, utilizando o modelo proposto em (3.4) com a presença de duas covariáveis, X_{1t} e X_{2t} , respectivamente, e com $p = 1$. As covariáveis foram geradas usando uma distribuição normal multivariada, com vetor de

médias igual a uma contante e matriz de covariância espacialmente correlacionada. Reescrevemos o modelo da seguinte maneira

$$Y_t = \beta_t X_{1t} + \gamma_t X_{2t} + \phi K_t K_{t-1}^{-1} (Y_{t-1} - \beta_{t-1} X_{1t-1} - \gamma_{t-1} X_{2t-1}) + v_t, \quad v_t \sim N(0, \Sigma_t)$$

$$\beta_t = \beta_{t-1} + \omega_{1t}, \quad \omega_{1t} \sim N(0, \tau_\beta^2), \quad \gamma_t = \gamma_{t-1} + \omega_{2t}, \quad \omega_{2t} \sim N(0, \tau_\gamma^2).$$

Os valores iniciais dos coeficientes das covariáveis foram $\beta_0 = 2$ e $\gamma_0 = -3$ com variâncias $\tau_\beta^2 = 0,1$ e $\tau_\gamma^2 = 0,25$, respectivamente. O parâmetro autoregressivo foi fixado em $\phi = 0,75$. Utilizamos a função de correlação exponencial e os parâmetros σ^2 e λ foram fixados em 1 e 5,5, respectivamente, no Modelo 1. Para os Modelos 2, 3 e 4, estes valores foram usados como valores iniciais. Os dados foram gerados usando a evolução lognormal descrita na Tabela 3.1. A Figura 3.1 mostra a localização das 30 estações, destacando as 2 estações (denotadas, respectivamente, por 1 e 2), utilizadas para fazer a interpolação espacial. A Figura 3.2 apresenta os dados simulados com estes parâmetros para as 30 localizações geradas em um quadrado unitário, ao longo dos 50 instantes de tempo para os quatro modelos.

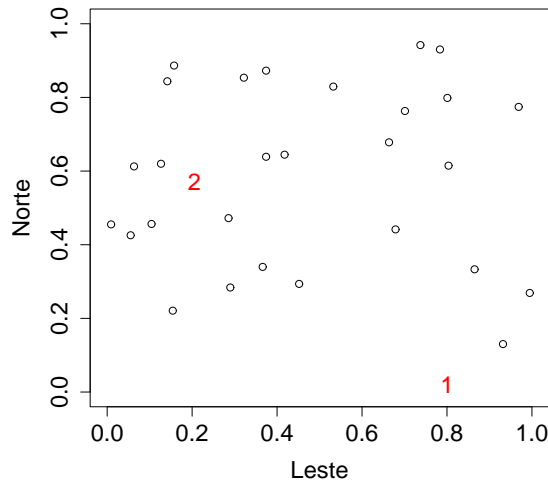


Figura 3.1: Localização das estações monitoradoras para os dados artificiais.

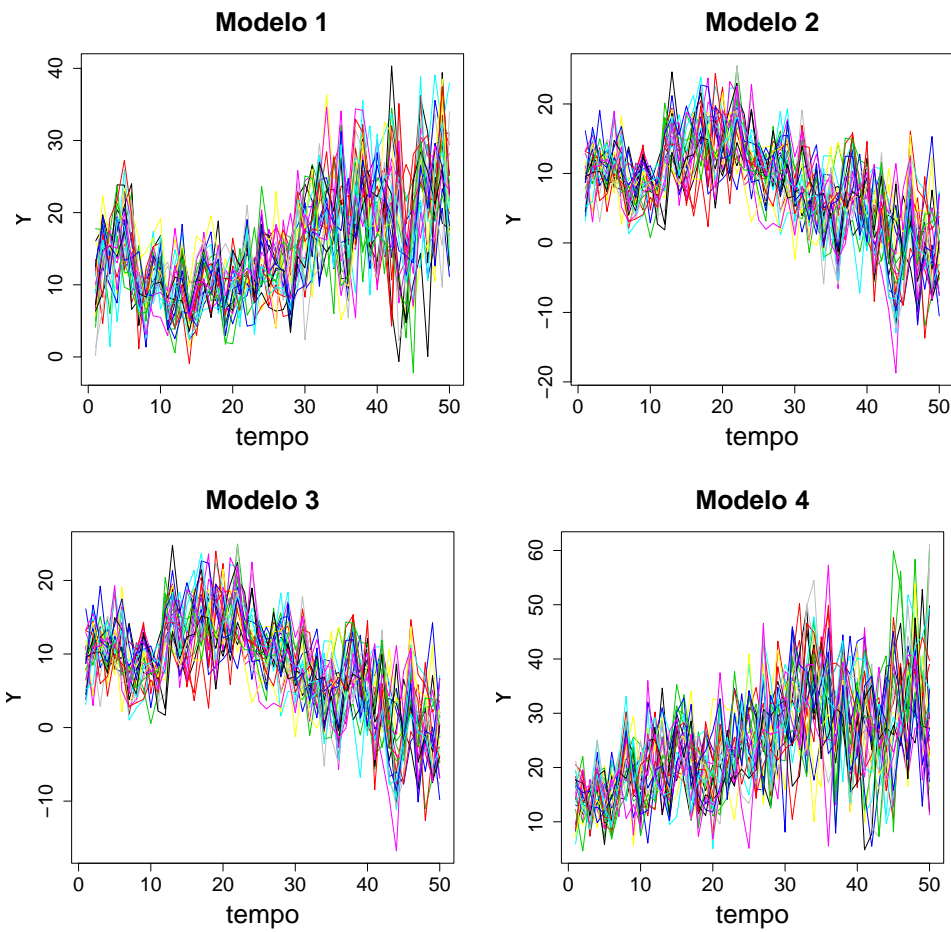


Figura 3.2: Dados simulados para 30 localizações ao longo de 50 instantes de tempo segundo os Modelos 1, 2, 3 e 4, descritos na Tabela 3.1.

Uma vez que as observações foram geradas propusemos o ajuste dos quatro modelos descritos na Tabela 3.1. Ajustamos o Modelo 1, utilizando a priori lognormal, para os dados gerados a partir do Modelo 1. O mesmo procedimento foi adotado para os Modelos 2, 3 e 4. As prioris utilizadas foram as seguintes: $\tau_\beta^2 \sim IG(0.1, 0.1)$, $\tau_\gamma^2 \sim IG(0.1, 0.1)$, $\phi \sim U(-1, 1)$ e $(\beta_0, \gamma_0)' \sim N_2(\mathbf{0}, 100\mathbf{I})$. Para os Modelos 1 e 2, para os quais a variância é fixa, utilizamos uma priori gama invertida, isto é, $\sigma^2 \sim IG(1, 1)$. A função de correlação exponencial foi utilizada e para o parâmetro de decaimento da função adotamos uma priori centrada em um valor cuja função de correlação exponencial é pequena (igual a 0,05) quando a distância euclidiana é maior ou igual a metade da distância máxima observada entre as estações. A distância euclidiana máxima encontrada entre as estações foi 1,08. A variância da distribuição a priori foi escolhida tal que $P(\lambda > 10\mu_\lambda) = 0,10$. Então, por exemplo, os hiperparâmetros da normal associada foram, $\mu_\lambda = 1,57$ e $\sigma_\lambda^2 = 0,27$. As distribuições a priori utilizadas para σ_1^2 nos Modelos 3 e 4, foram escolhidas de maneira que a média fosse igual a 1 (valor verdadeiro) com variância tal que $P(\sigma^2 > 10\mu_\sigma) = 0,10$. Os hiperparâmetros da normal associada foram $\mu_\sigma = -0,72$ e $\tau_\sigma^2 = 1,44$.

Um programa em 0x (Doornik e Ooms 2001) foi escrito para obter uma amostra da distribuição a posteriori de interesse. Uma amostra final de tamanho 1.000 foi utilizada, após a retirada das 10.000 iterações iniciais e a coleta de elementos da amostra a cada 20ª iteração do algoritmo. Além disso, duas cadeias foram utilizadas. Além da inspeção visual, verificamos a convergência dos parâmetros através do pacote BOA do R, usando o critério de diagnóstico de convergência proposto por Gelman e Rubin (1992). A variância da proposta foi escolhida usando 5.000 iterações, calculando a média das variâncias das 100 amostras de tamanho 50, de modo que a taxa de aceitação ficasse entre 15% e 50% (Roberts 1996).

A Tabela 3.2 apresenta os resultados da estimação para os parâmetros fixos no tempo para os quatro modelos. Notamos que o coeficiente autoregressivo, ϕ , e as variâncias de evolução, τ_β^2 e τ_γ^2 são relativamente bem estimados para os quatro modelos. Na Figura 3.3 apresentamos os coeficientes das covariáveis, β_t e γ_t , para o Modelo 3. Observamos que estes coeficientes também são bem estimados, pois observamos que a média e o intervalo de 95% de credibilidade a posteriori acompanham o comportamento do valor verdadeiro ao longo do tempo. Os resultados obtidos para outros modelos foram similares e, portanto, não apresentamos aqui.

	$\phi = 0,75$		$\tau_{\beta}^2 = 0,10$		$\tau_{\gamma}^2 = 0,25$	
Modelo	Média	IC(95%)	Média	IC(95%)	Média	IC(95%)
M1	0,765	(0,732; 0,797)	0,124	(0,077; 0,198)	0,249	(0,159; 0,369)
M2	0,733	(0,697; 0,766)	0,108	(0,065; 0,172)	0,231	(0,141; 0,355)
M3	0,766	(0,731; 0,801)	0,109	(0,065; 0,169)	0,290	(0,180; 0,455)
M4	0,765	(0,730; 0,799)	0,129	(0,081; 0,204)	0,347	(0,223; 0,532)

Tabela 3.2: Média e intervalo de 95% de credibilidade a posteriori para os parâmetros fixos no tempo para os dados artificiais com covariáveis.

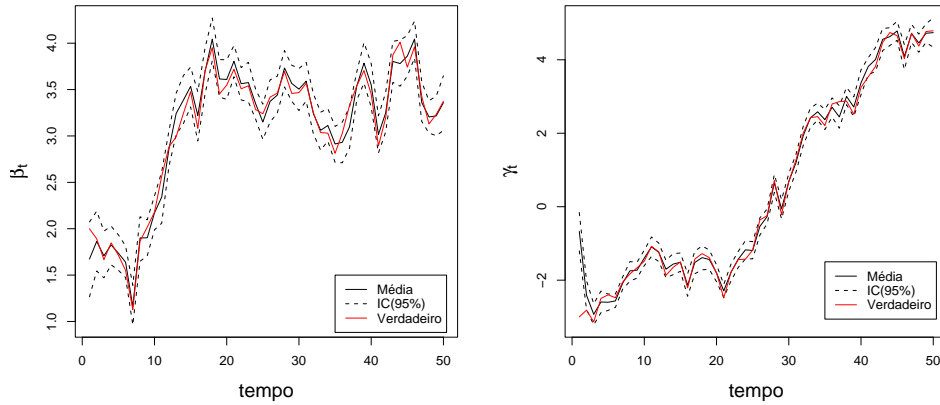


Figura 3.3: Média e Intervalo de Credibilidade (95%) para os coeficientes das variáveis X_1 e X_2 ao longo do tempo, $t = 1, \dots, 50$ segundo o Modelos 3 para os dados artificiais com covariáveis.

O mesmo ocorre com os parâmetros da estrutura de covariância, σ^2 e λ , os intervalos de credibilidade contém os verdadeiros valores. A Figura 3.4 apresenta o histograma da amostra da distribuição a posteriori para os parâmetros da estrutura de covariância, usando o Modelo 1. Já a Figura 3.5 mostra os resultados para o Modelo 2, no qual o parâmetro da função de correlação varia ao longo do tempo. A média e o intervalo de 95% de credibilidade a posteriori para cada λ_t é apresentado, assim como o histograma da amostra da posteriori para a variância σ^2 . Podemos observar que as estimativas para λ_t são menos suaves que o valor verdadeiro. O mesmo comportamento é observado para σ_t^2 no Modelo 3 e para λ_t e σ_t^2 no Modelo 4 (Figuras 3.6 e 3.7, respectivamente). Ainda assim, o valor verdadeiro está dentro do intervalo de credibilidade para quase todos os tempos. Observamos ainda que, a evolução para σ_t^2 parece relativamente melhor estimada do que para λ_t , quando comparamos todos os modelos ajustados.

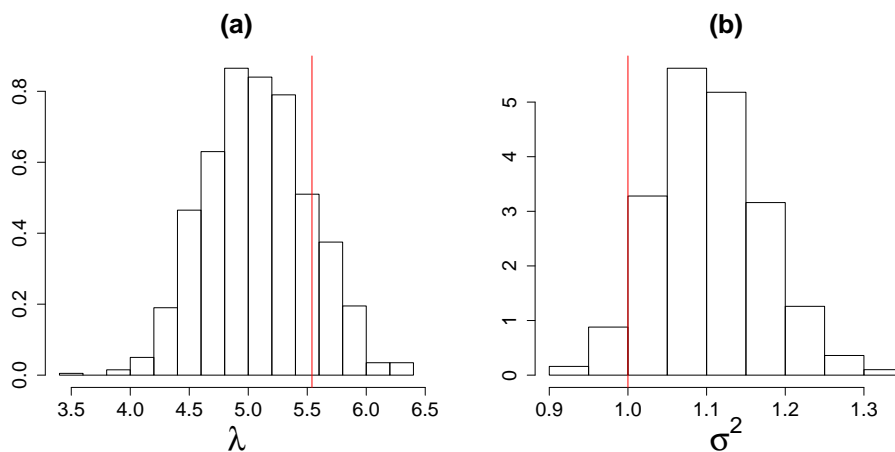


Figura 3.4: Histograma da amostra da distribuição a posteriori: (a) parâmetro da função de correlação, λ , e (b) parâmetro da variância, σ^2 , segundo o Modelo 1 para os dados artificiais com covariáveis.

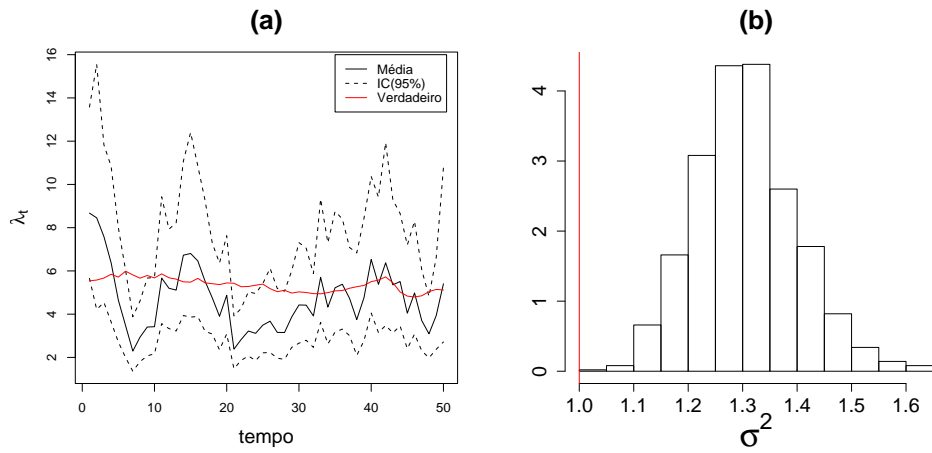


Figura 3.5: (a) Média e intervalo de 95% de credibilidade a posteriori para o parâmetro da função de correlação, λ_t e (b) Histograma da amostra da distribuição a posteriori para variância, σ^2 , segundo o Modelo 2 para os dados artificiais com covariáveis.

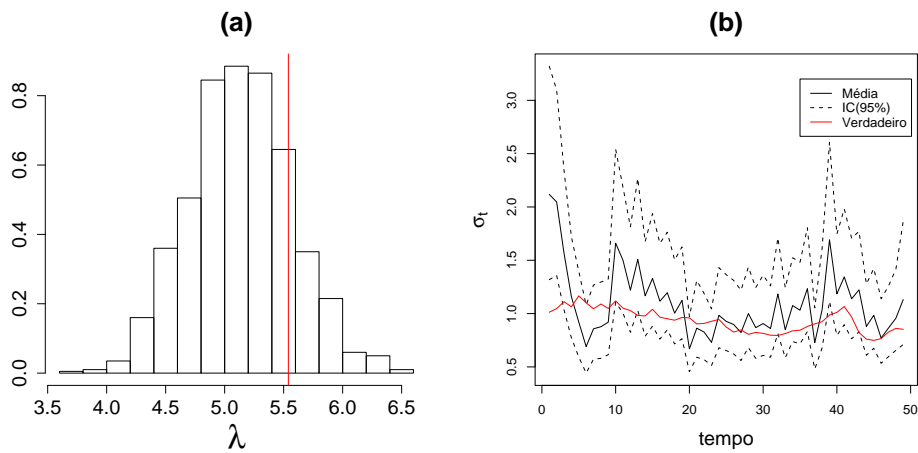


Figura 3.6: (a) Histograma da amostra da distribuição a posteriori para o parâmetro da função de correlação, λ e (b) Média e intervalo de 95% de credibilidade a posteriori para variância, σ_t^2 , segundo o Modelo 3 para os dados artificiais com covariáveis.

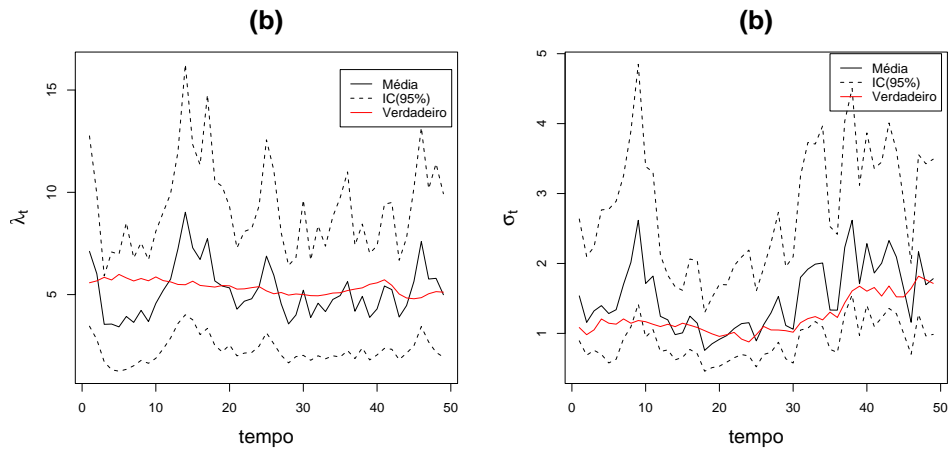


Figura 3.7: Média e intervalo de 95% de credibilidade a posteriori: (a) parâmetro da função de correlação λ_t e (b) variância, σ_t^2 , segundo o Modelo 4 para os dados artificiais com covariáveis.

A interpolação espacial foi realizada a partir da equação (3.10) mostrada na Seção 3.6, onde Y_t^u é um vetor referente as 2 localizações retiradas para fazer a interpolação. Os modelos foram ajustados sem as 2 estações, e as amostras da distribuição a posteriori foram encontradas, e utilizadas na obtenção da interpolação espacial. A Figura 3.8 mostra o valor verdadeiro, bem como a média e o intervalo de 95% de credibilidade para previsão ao longo do tempo para o Modelo 3. Observamos que o modelo está conseguindo captar bem o comportamento da série, pois a média encontra-se bem próxima ao valor verdadeiro e o intervalo de credibilidade é bem concentrado, embora no início das séries os intervalos de 95% de credibilidade não contenham os verdadeiros valores. Resultados similares foram obtidos para os outros modelos.

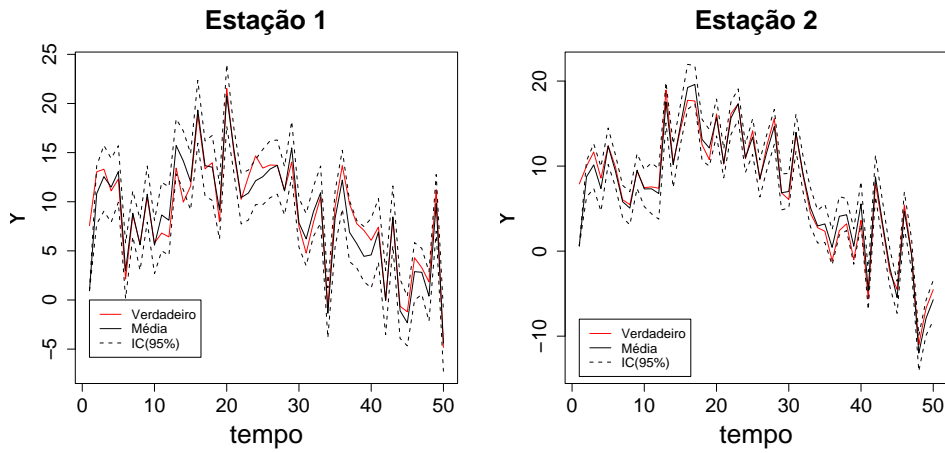


Figura 3.8: Previsão para 2 localizações segundo o Modelo 3 para os dados artificiais com covariáveis.

Com o objetivo de verificar a capacidade preditiva do modelo, uma grade regular de 400 pontos foi criada, e foi feita uma interpolação espacial nessa grade para 8 tempos, $t = 30, \dots, 37$ para cada um dos modelos. Agora Y_t^u representa o vetor de observações nos 400 pontos não medidos. Em seguida, as superfícies foram construídas usando os valores verdadeiros dos parâmetros. Para isso, substituímos estes valores na equação (3.10). As superfícies usando os valores verdadeiros dos parâmetros para o Modelo 3 são mostradas na Figura 3.9, enquanto que as superfícies estimadas pelo modelo usando a mediana a posteriori, encontram-se na Figura 3.10. Podemos observar que as superfícies são bem similares entre si. O intervalo interquartil a posteriori é apresentado na Figura 3.11. Como esperado, o intervalo interquartil apresenta valores menores em regiões próximas dos valores observados. As superfícies para os outros modelos apresentaram resultados similares.

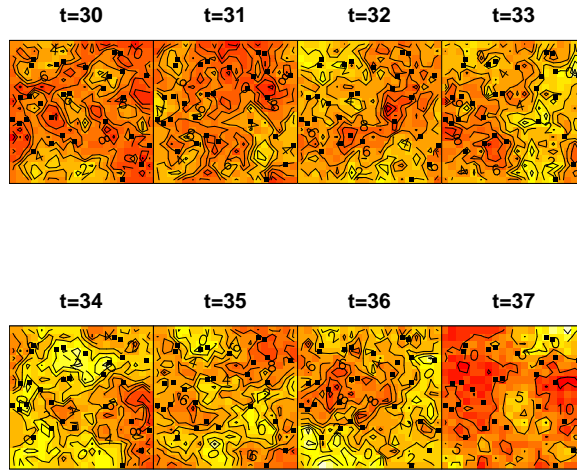


Figura 3.9: Interpolação espacial usando os valores verdadeiros dos parâmetros em 8 tempos consecutivos, segundo o Modelo 3 para os dados artificiais com covariáveis.

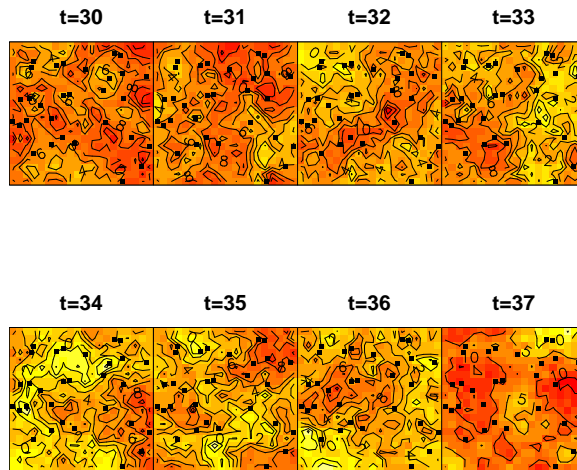


Figura 3.10: Mediana a posteriori da distribuição preditiva em 8 tempos consecutivos, segundo o Modelo 3 para os dados artificiais com covariáveis.

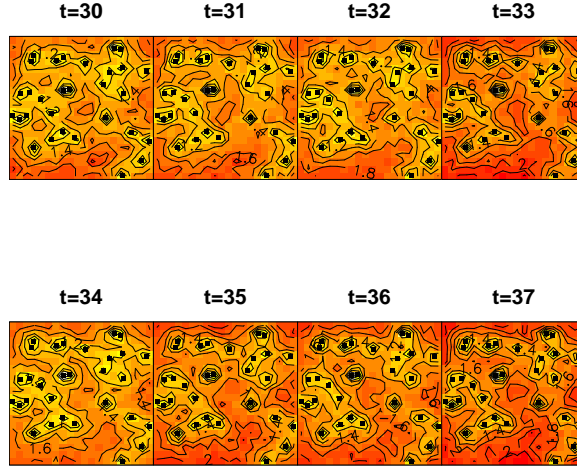


Figura 3.11: Intervalo interquartilico da distribuição preditiva em 8 tempos consecutivos, segundo o Modelo 3 para os dados artificiais com covariáveis.

3.7.2 Dados Artificiais com Estrutura Sazonal

Nesta Seção desejamos investigar a capacidade do modelo de convolução espacial em estimar estruturas sazonais no processo sob estudo. Para isso, geramos um conjunto de observações para as mesmas 30 localizações mostradas na Seção 3.7.1 ao longo de 60 instantes de tempo, utilizando o modelo proposto com a presença de 1 componente harmônico e $p = 1$. Reescrevemos o modelo da seguinte maneira

$$\begin{aligned}
 Y_t &= \alpha_t E + \phi K_t K_{t-1}^{-1} (Y_{t-1} - \alpha_{t-1} E) + v_t, & v_t &\sim N(0, \Sigma_t) \\
 \alpha_t &= G \alpha_{t-1} + \omega_t, & \omega_t &\sim N(0, \tau_\alpha^2),
 \end{aligned} \tag{3.11}$$

onde $\alpha_t = (\alpha_{1t}, \alpha_{2t})$ são os coeficientes do componente harmônico e E é uma matriz $n \times 2$ que tem a primeira coluna de 1's e a segunda coluna de zeros. Os valores iniciais foram $\alpha_{10} = 2$, $\alpha_{20} = -3$, $\tau_\alpha^2 = 0,1$ e $\phi=0,75$. Os valores fixos e/ou os valores iniciais dos parâmetros da função de correlação exponencial σ^2 e λ foram iguais a 1 e 5,54, respectivamente. Consideramos a representação de Fourier para sazonalidade. Assim, seguindo West e Harrison (1997), a matriz de evolução G com periodicidade

igual a 6 é tal que,

$$G = \begin{pmatrix} \cos(2\pi/6) & \text{sen}(2\pi/6) \\ -\text{sen}(2\pi/6) & \cos(2\pi/6) \end{pmatrix}.$$

A Figura 3.12 apresenta os dados simulados a partir do modelo (3.11) usando as diferentes especificações para variância apresentadas na Tabela 3.1. Os dados foram gerados usando a evolução lognormal para λ_t e σ_t^2 nos Modelo 2, 3 e 4.

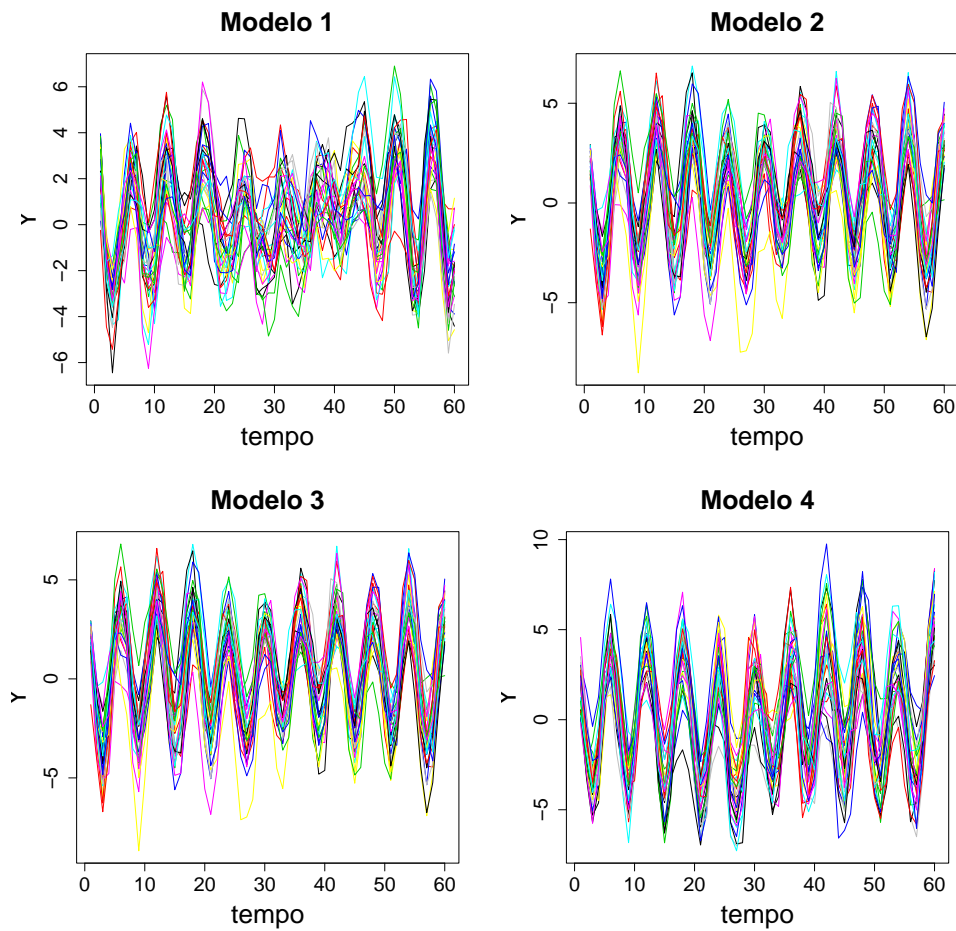


Figura 3.12: Dados simulados para 30 localizações ao longo de 60 instantes de tempo para os Modelos 1, 2, 3 e 4 utilizando o modelo (3.11).

Para verificar o impacto da especificação da evolução dos parâmetros da função de covariância, utilizamos a evolução baseada na priori gama para ajustar os modelos. Ajustamos cada modelo utilizando os dados gerados a partir dos respectivos modelos. A Tabela 3.3 apresenta os resultados para os parâmetros fixos, ϕ e τ_α^2 para todos os modelos ajustados. A variância de evolução τ_α^2 é estimada em um valor superior ao valor gerado em todos os modelos, indicando a dificuldade em identificar a variância de evolução. Observamos que o valor verdadeiro do parâmetro autoregressivo, encontra-se deslocado em relação a média a posteriori, mas está dentro do intervalo de 95% de credibilidade a posteriori. A Figura 3.13 apresenta as estimativas para os coeficientes α_{1t} e α_{2t} , da componente harmônica para o Modelo 2. Aparentemente, o modelo ajusta-se bem aos dados visto que o verdadeiro valor encontra-se dentro do intervalo de 95% de credibilidade a posteriori. O intervalo de credibilidade do coeficiente α_{1t} está bem concentrado. Resultados similares foram observados para os outros modelos ajustados.

	$\phi = 0,75$		$\tau_\alpha^2 = 0,10$	
Modelo	Média	IC(95%)	Média	IC(95%)
M1	0,767	(0,735; 0,798)	1,935	(1,083; 3,163)
M2	0,771	(0,743; 0,800)	7,442	(5,090; 10,437)
M3	0,770	(0,742; 0,799)	3,158	(1,624; 5,425)
M4	0,771	(0,742; 0,800)	4,386	(2,346; 7,857)

Tabela 3.3: Média e intervalo de 95% de credibilidade a posteriori para os parâmetros fixos no tempo para os dados artificialmente gerados com estrutura sazonal.

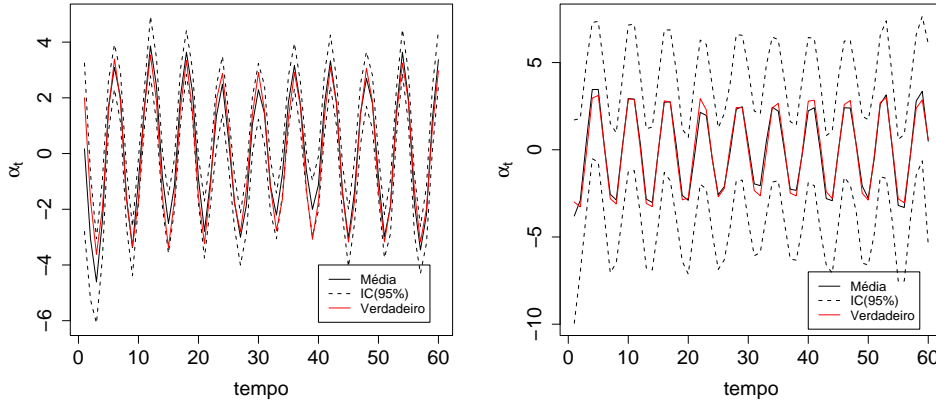


Figura 3.13: Média e Intervalo de Credibilidade (95%) a posteriori para os coeficientes da componente sazonal, α_{1t} (esquerda) e α_{2t} (direita), $t = 1, \dots, 60$, segundo o Modelo 2 para os dados com estrutura sazonal.

As Figuras 3.14, 3.15, 3.16 e 3.17 mostram as estimativas para os parâmetros λ_t e σ_t^2 para os modelos 1, 2, 3 e 4, respectivamente. Observamos que as médias a posteriori destes parâmetros quando eles são fixos no tempo, encontram-se bem próximas do verdadeiro valor. Quando estes parâmetros variam ao longo do tempo, o intervalo de 95% de credibilidade a posteriori contém o verdadeiro valor para todos os tempos. A convergência dos parâmetros também foi verificada através do critério proposto por Gelman e Rubin (1992).

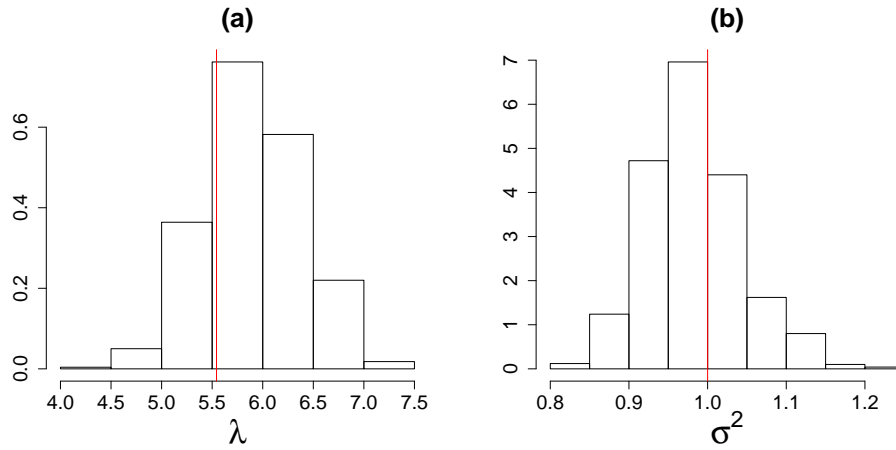


Figura 3.14: Histograma da amostra da distribuição a posteriori: (a) parâmetro da função de correlação, λ , e (b) parâmetro da variância, σ^2 , segundo o Modelo 1 para os dados com estrutura sazonal.

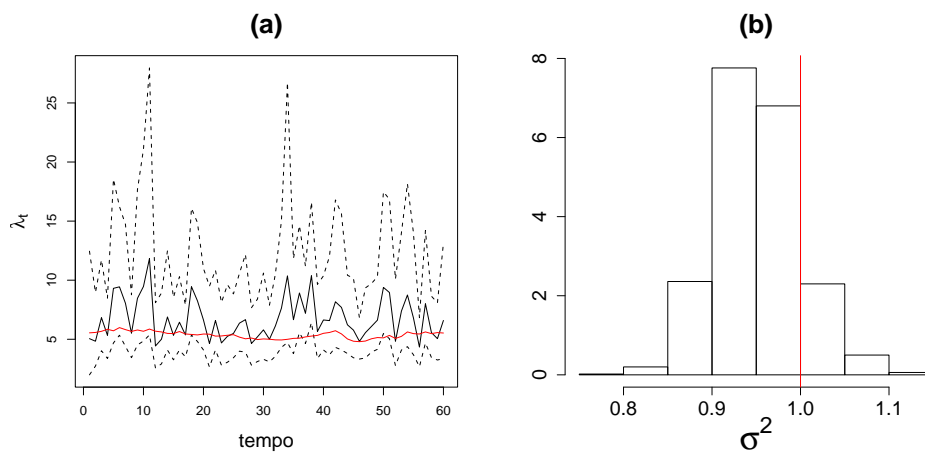


Figura 3.15: (a) Média e intervalo de 95% de credibilidade a posteriori para o parâmetro da função de correlação, λ_t e (b) Histograma da amostra da distribuição a posteriori para variância, σ^2 , segundo o Modelo 2 para os dados com estrutura sazonal.

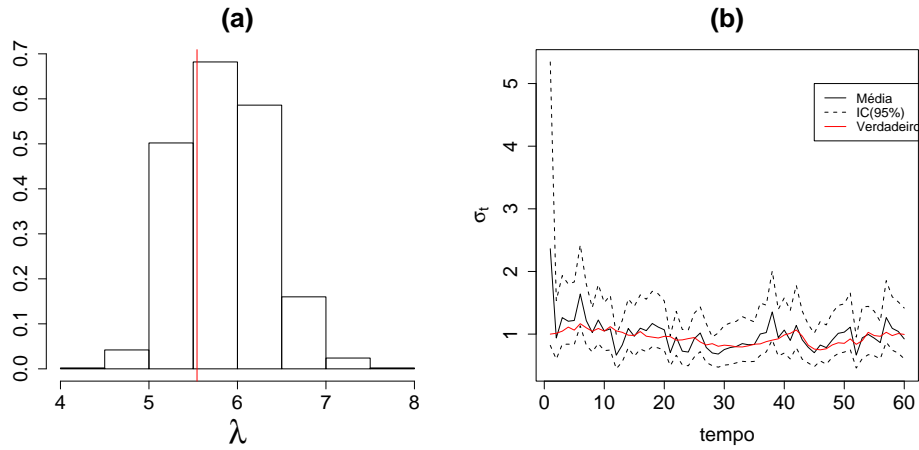


Figura 3.16: (a) Histograma da amostra da distribuição a posteriori para o parâmetro da função de correlação, λ e (b) Média e intervalo de 95% de credibilidade a posteriori para variância, σ_t^2 , segundo o Modelo 3 para os dados com estrutura sazonal.

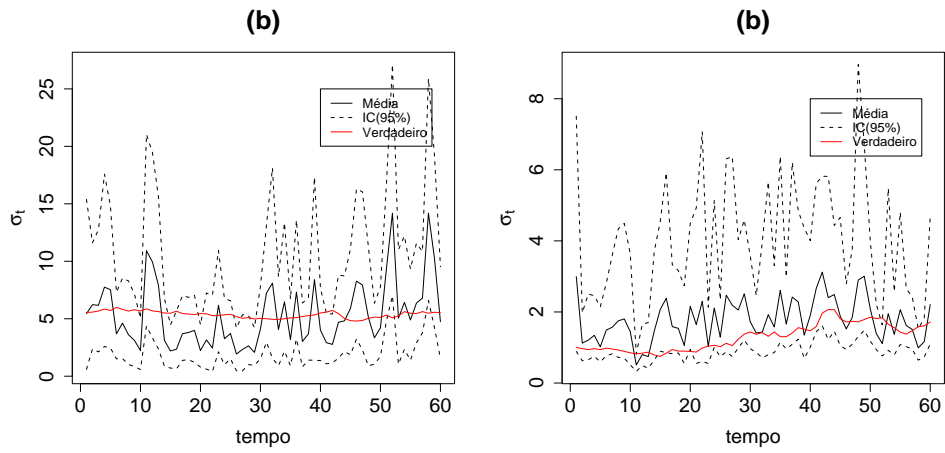


Figura 3.17: Média e intervalo de 95% de credibilidade a posteriori (a) parâmetro da função de correlação λ_t e (b) variância, σ_t^2 , segundo o Modelo 4 para os dados com estrutura sazonal.

A equação (3.10) exibida na Seção 3.6 foi utilizada para fazer a interpolação espacial, utilizando o mesmo procedimento descrito na Seção anterior. Inicialmente, retiramos as 2 estações exibidas na Seção 3.7.1 para verificar a capacidade preditiva dos modelos usando essa estrutura de sazonalidade. O modelo consegue captar razoavelmente bem a estrutura das séries. Para ambas as estações retiradas da análise, a média a posteriori encontra-se bem próxima do valor verdadeiro, principalmente para a estação 2, como pode ser visto na Figura 3.18. Podemos ver na Figura 3.1 que a estação 2 possui mais estações ao seu redor do que a estação 1 e, talvez por isso, a previsão para esta estação tenha sido melhor. Os Modelos 1, 3 e 4 apresentaram resultados semelhantes e não apresentamos aqui.

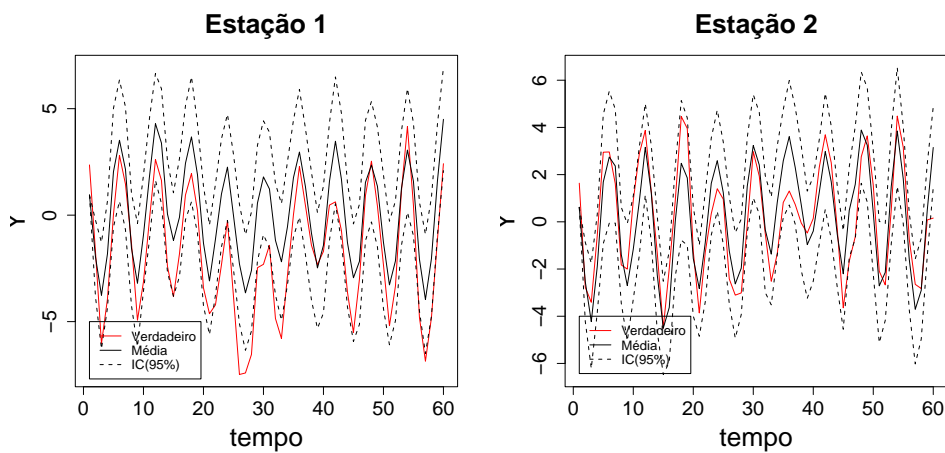


Figura 3.18: Previsão para 2 localizações, segundo o Modelo 2 para os dados com estrutura sazonal.

Finalmente, a superfície utilizando os verdadeiros valores dos parâmetros para a grade de 400 pontos em 8 tempos, sob o Modelo 2 encontra-se na Figura 3.19. A superfície verdadeira é calculada utilizando a equação (3.10) substituindo θ pelos valores verdadeiros utilizados para simular os dados. A superfície construída com a amostra da distribuição a posteriori dos parâmetros encontra-se na Figura 3.20. Comparando as duas superfícies, verificamos que elas são semelhantes, ou seja, valores mais altos são observados na mesma região, embora a superfície prevista pelo modelo é bem mais suave que a superfície verdadeira. Como esperado, na Figura 3.21 vemos que o intervalo interquartilício apresenta valores maiores em regiões distantes dos valores observados.

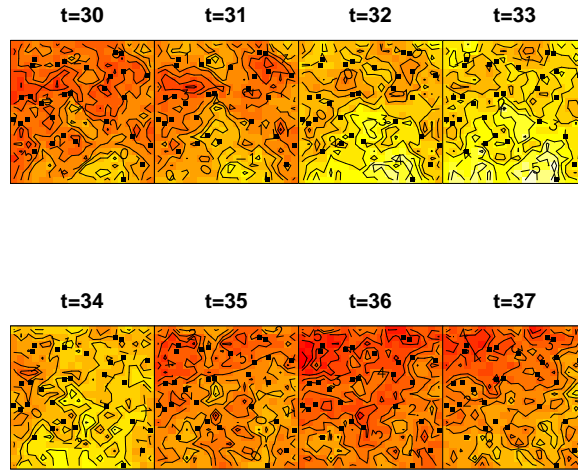


Figura 3.19: Interpolação espacial usando os valores verdadeiros dos parâmetros em 8 tempos consecutivos, segundo o Modelo 2 para os dados com estrutura sazonal.

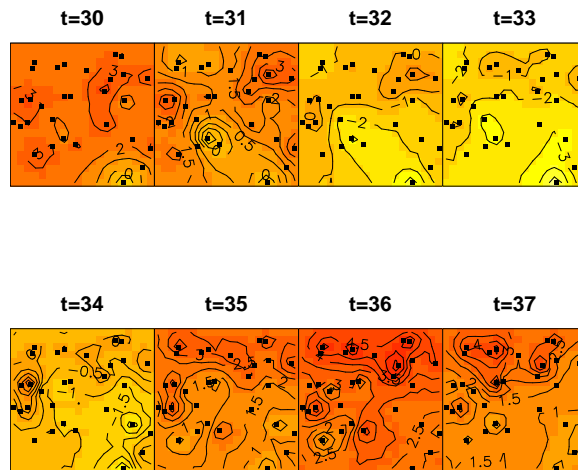


Figura 3.20: Mediana a posteriori da distribuição preditiva em 8 tempos consecutivos, segundo o Modelo 2 para os dados com estrutura sazonal.

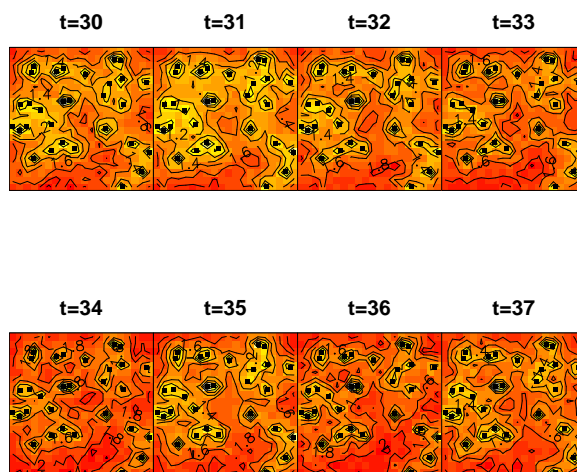


Figura 3.21: Intervalo interquartílico da distribuição preditiva em 8 tempos consecutivos, segundo o Modelo 2 para os dados com estrutura sazonal.

3.7.3 Análise dos níveis de PM_{10} na Região Metropolitana do Rio de Janeiro

O poluente PM_{10} corresponde a partículas inaláveis que ficam suspensas no ar na forma de poeira, neblina, fumaça, fuligem, etc. A faixa de tamanho dessas partículas é menor que 10 micra. O monitoramento destas partículas na atmosfera é de grande importância pois elas podem causar problemas à saúde humana, principalmente no aparelho respiratório. A cidade do Rio de Janeiro caracteriza-se por uma topografia peculiar com montanhas relativamente altas, presença do mar e de mudanças bruscas no clima que influenciam a dispersão dos poluentes em suspensão na atmosfera. As principais fontes de poluição do ar da cidade são os veículos automotores, as indústrias e a queima de lixo. A capacidade de dispersão dos poluentes ao longo da região depende das características topográficas locais, da direção do vento, de fatores meteorológicos, entre outros. Devido a estas diversidades encontradas na região observamos a necessidade de buscar um modelo que seja capaz de capturar essas mudanças no espaço e no tempo. Paez e Gamerman (2003) fazem um estudo dos níveis de PM_{10} na região metropolitana do Rio de Janeiro durante o ano de 1999. Vários modelos espaço-temporais

são ajustados considerando diferentes especificações para as componentes espaciais e temporais. Os modelos ajustados supõem que a estrutura de covariância é separável, isto é, não existe interação entre espaço e tempo. Entretanto, esta hipótese pode não ser razoável, de acordo com as características da cidade.

Nossa proposta é ajustar o modelo proposto na Seção 3.3 aos dados de PM_{10} observados numa região da cidade do Rio de Janeiro. Mais especificamente, os dados disponíveis foram coletados durante o ano de 2002, em 18 estações monitoradoras localizadas na região metropolitana do Rio de Janeiro. Estas estações são gerenciadas por diversos órgãos públicos. Neste trabalho usamos 13 estações gerenciadas pela Fundação Estadual de Engenharia de Meio Ambiente (FEEMA) das quais 12 possuem coleta manual e 1 automática, e 5 estações da Secretaria do Meio Ambiente da Prefeitura do Rio de Janeiro (SMAC) onde a coleta é automática. As medidas foram feitas de 6 em 6 dias entre abril e dezembro de 2002, totalizando 46 observações. A Figura 3.22 apresenta o mapa da cidade do Rio de Janeiro com a localização das 18 estações monitoradoras de PM_{10} . As estações que compõem o estudo, bem como o número de identificação são mostrados na Tabela 3.4.

Os níveis de PM_{10} para cada uma das estações ao longo dos 46 dias podem ser vistos na Figura 3.23. Observamos a ocorrência de dados faltantes em todas as estações. As observações faltantes foram estimadas a partir do próprio modelo, já que segundo o enfoque bayesiano elas são vistas como parâmetros do modelo.

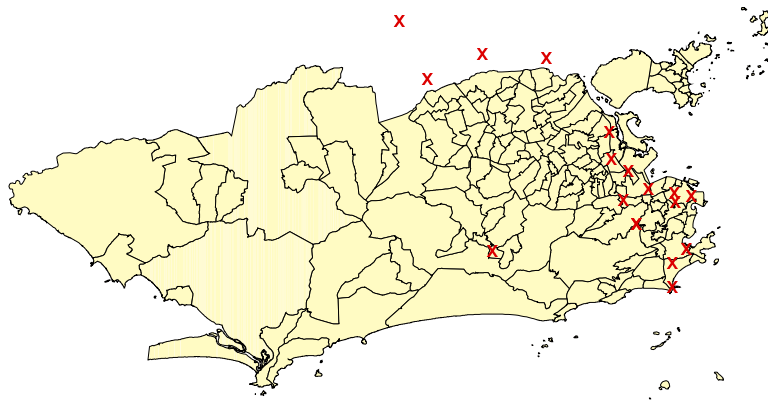


Figura 3.22: Localização das 18 estações monitoradoras de PM_{10} na região Metropolitana do Rio de Janeiro.

ID	Estação	ID	Estação
1	Sumaré	10	Duque de Caxias
2	São João de Meriti	11	Botafogo
3	São Cristovão	12	Centro 1
4	Centro 2	13	Manguinhos
5	Bonsucesso	14	São Cristovão 2
6	Copacabana	15	Saens Pena
7	Maracanã	16	Carioca
8	Nilópolis	17	Arcoverde
9	Jacarepaguá 2	18	Nova Iguaçu

Tabela 3.4: Identificação das estações monitoradoras de PM_{10}

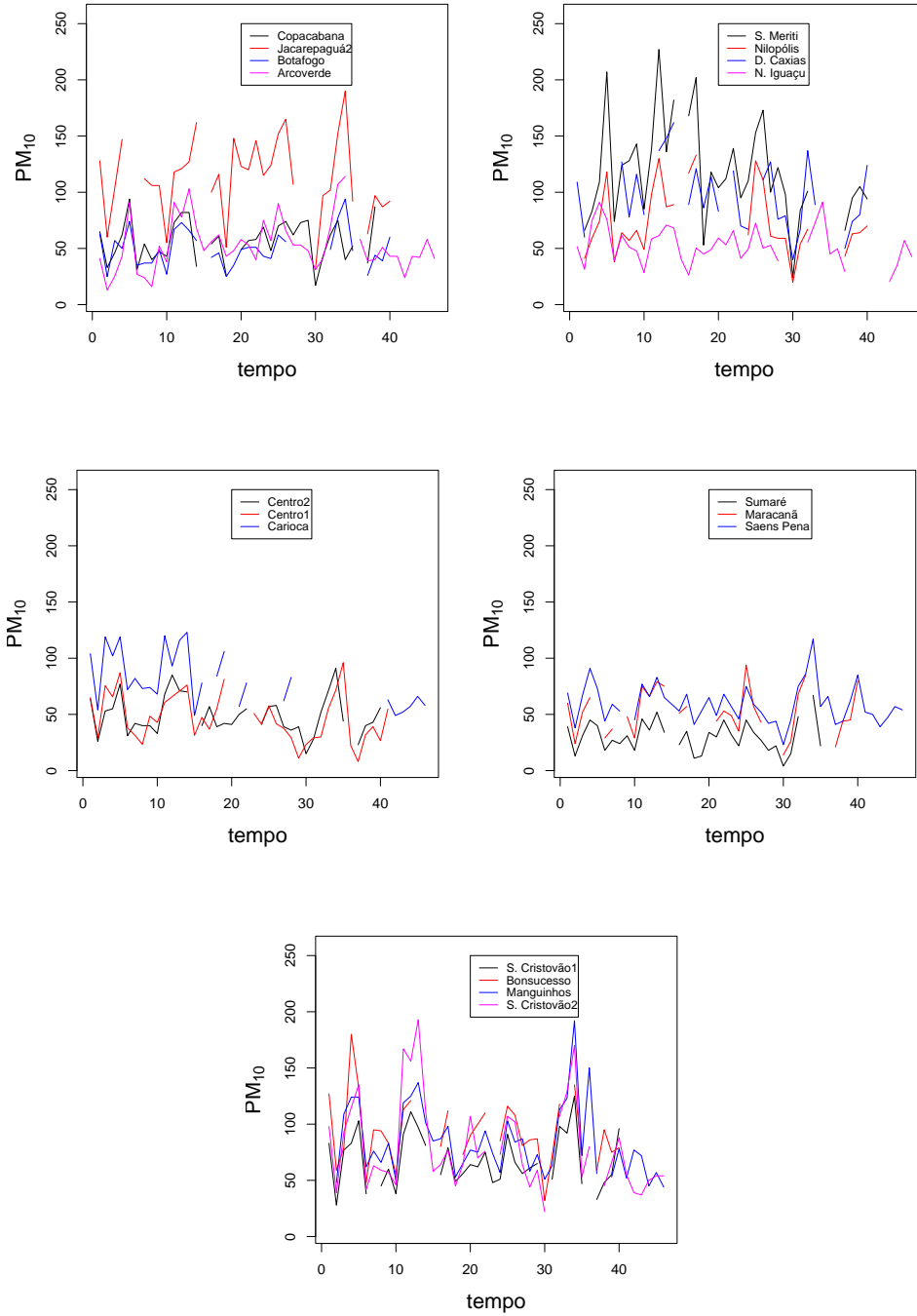


Figura 3.23: Nível de PM_{10} observado nas 18 estações monitoradoras ao longo dos 46 instantes de tempo.

Krigagem Bayesiana para Temperatura e Umidade Relativa

Vários estudos relacionam fatores climatológicos com a poluição do ar. Por isso, neste trabalho investigamos o efeito de fatores meteorológicos tais como temperatura máxima e umidade relativa, nos níveis de PM_{10} . Paez e Gamerman (2003) utilizaram a temperatura máxima diária medida em uma única estação como covariável. Aqui, os dados de temperatura e umidade foram coletados em 9 estações monitoradoras no mesmo período do PM_{10} , sendo que apenas 5 destas estações coincidem com as estações medidoras de PM_{10} . As 4 estações restantes são localizadas em Santa Cruz, Campo dos Afonsos e nos aeroportos do Galeão e Santos Dumont. Portanto, inicialmente foi utilizado o método de krigagem bayesiana para prever a temperatura e a umidade para as 13 estações de PM_{10} restantes. Krigagem significa fazer inferência sobre valores não observados do processo aleatório condicional aos valores observados. Do ponto de vista Bayesiano, é atribuído uma distribuição a priori para todos os parâmetros envolvidos no modelo.

Seja Y o vetor de observações nas n localizações s_1, \dots, s_n , e Z um processo Gaussiano estacionário com $E(Z(s)) = 0$, $Var(Z(s)) = \sigma^2$ e função de correlação $\rho(\lambda, \|s - s'\|)$. A distribuição condicional de $Y(s)$ dado $Z(s)$ é normal com média $\mu(s) + Z(s)$ e variância τ^2 , onde $\mu(s) = \sum_{j=1}^p \beta_j X_j(s)$ com X_j representando a j -ésima variável explicativa, e τ^2 representando o erro de medida chamado de efeito pepita. Então a distribuição conjunta de Y é normal multivariada tal que

$$Y|\boldsymbol{\theta} \sim N(X\boldsymbol{\beta}, \sigma^2 V + \tau^2 I). \quad (3.12)$$

onde $\boldsymbol{\theta} = (\beta, \sigma^2, \tau^2, \lambda)$ denota o vetor paramétrico, V a matriz de correlação $n \times n$ com elementos $\rho(\lambda; |s_i - s_j|)$ e λ o parâmetro da função de correlação. Seguindo o teorema de Bayes e supondo independência a priori, a distribuição conjunta a posteriori dos parâmetros é

$$p(\boldsymbol{\theta}|Y) \propto |\Sigma|^{-1/2} \exp \left\{ -\frac{1}{2} (Y - X\boldsymbol{\beta})' \Sigma^{-1} (Y - X\boldsymbol{\beta}) \right\} p(\boldsymbol{\beta}) p(\sigma^2) p(\tau^2) p(\lambda),$$

onde $\Sigma = \sigma^2 V + \tau^2 I$. Em seguida, a distribuição preditiva das localizações não medidas é obtida. A distribuição preditiva é baseada na distribuição do processo nas localizações não medidas condicional às observações das localizações medidas, e este resultado é obtido através da propriedade da partição da normal multivariada.

O modelo (3.12) foi ajustado para cada tempo, separadamente, para os dados de temperatura e umidade utilizando apenas um nível. A função de correlação utilizada foi a exponencial. As seguintes

prioris foram utilizadas: $\beta \sim N(0, 100)$, λ uma gama com média calculada usando a idéia de distância máxima discutida na Seção 3.7.1, e para σ^2 e τ^2 , uma priori gama invertida com média igual a variância observada e coeficiente de variação em torno de 0,3. A Figura 3.24 apresenta as séries de temperatura máxima e umidade relativa para cada uma das 18 estações, incluindo as 13 séries previstas pelo procedimento de krigagem. A temperatura e umidade prevista para estas 13 estações correspondem a média a posteriori da distribuição preditiva correspondente.

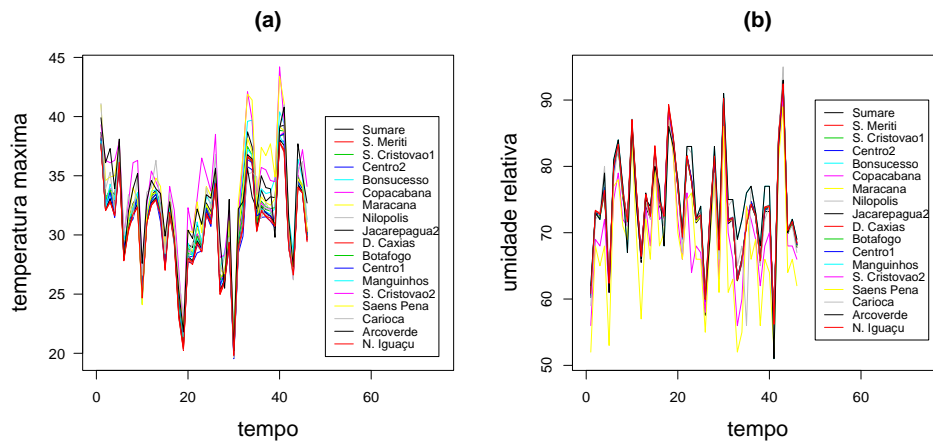


Figura 3.24: (a) Temperatura máxima e (b) umidade relativa preenchidas para cada tempo t , após a krigagem bayesiana, para as 18 estações ao longo dos 46 instantes de tempo.

Análise Preliminar

Existem padrões internacionais estabelecidos para verificar se uma região está com níveis de particulados em suspensão na atmosfera prejudiciais à saúde. O padrão médio anual aceitável é de $50\mu g/m^3$ e para as médias diárias o padrão é de $150\mu g/m^3$. A Figura 3.25 mostra o gráfico de caixa dos níveis de PM_{10} para cada tempo, indicando a média diária padrão e o nível de atenção ($250\mu g/m^3$) fornecido pelo CONAMA (Conselho Nacional do Meio Ambiente). Podemos observar que as estações de São João do Meriti (2) e Jacarepaguá 2 (9) apresentam níveis acima da média diária padrão para vários tempos. A partir da Figura 3.26 verificamos o comportamento dos níveis de PM_{10} nos diferentes dias da semana. Verificamos níveis mais altos na sexta, quinta e sábado e níveis mais baixos na terça e no

domingo.

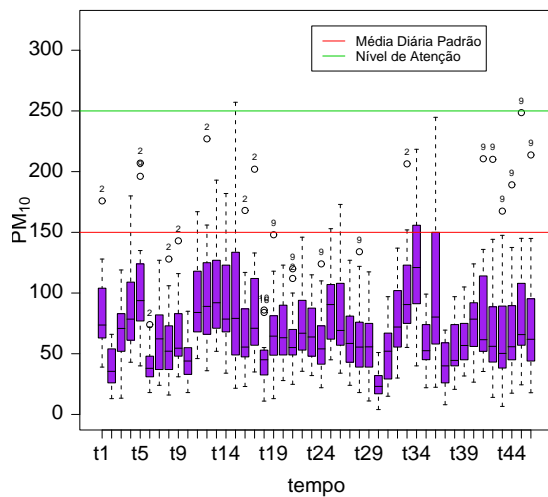


Figura 3.25: Distribuição dos níveis de PM_{10} para cada instante de tempo.

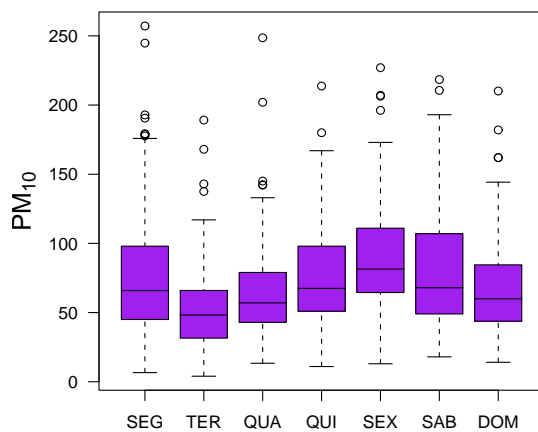


Figura 3.26: Distribuição dos níveis de PM_{10} para cada um dos dias da semana.

A Figura 3.27 apresenta as séries de PM_{10} , temperatura máxima e umidade relativa para as estações Sumaré e Jacarepaguá 2, que apresentam o menor e o maior nível de PM_{10} , respectivamente. Com o objetivo de obter normalidade dos dados, utilizamos a raiz quadrada de PM_{10} . A Figura 3.28 apresenta as variâncias observadas ao longo do tempo. Para cada tempo, foi calculada a variância entre as 18 estações monitoradoras. Observamos um pico na variância observada nos dias 24/06 (t=15) e 28/10 (t=36), que foram os dias em que os níveis de PM_{10} estiveram bem próximos do nível de atenção para algumas estações.

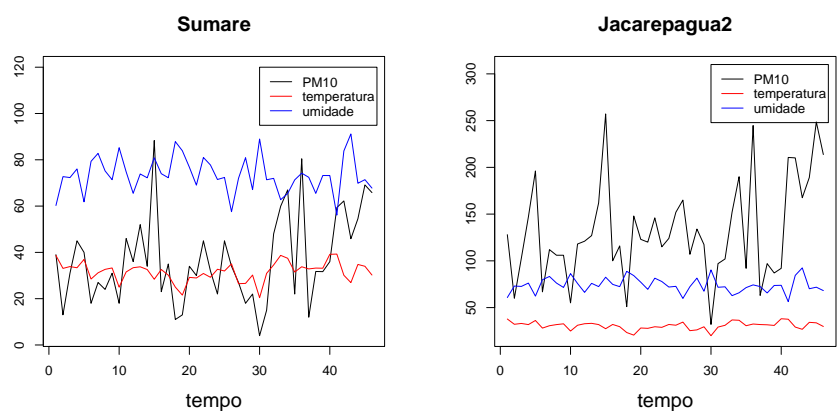


Figura 3.27: Nível de PM_{10} , temperatura máxima e umidade relativa para as estações Sumaré e Jacarepaguá 2 ao longo dos 46 instantes de tempo.

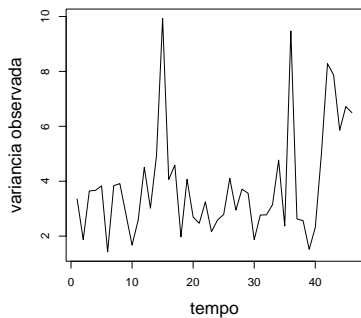


Figura 3.28: Variância observada da $\sqrt{PM_{10}}$ ao longo dos 46 instantes de tempo.

Modelo Proposto

Com base nos resultados obtidos na análise exploratória dos níveis de PM_{10} , propomos o ajuste do seguinte modelo

$$\sqrt{PM_{10t}} = F_t\theta_t + \phi K_t K_{t-1}^{-1}(\sqrt{PM_{10t-1}} - F_{t-1}\theta_{t-1}) + v_t, \quad v_t \sim N(0, \Sigma_t),$$

$$F_t\theta_t = \beta_t temp_t + \gamma_t umid_t + \alpha_1 seg_t + \alpha_2 ter_t + \alpha_3 qua_t + \alpha_4 qui_t + \alpha_5 sex_t + \alpha_6 sab_t,$$

$$\beta_t = \beta_{t-1} + \omega_{1t}, \quad \omega_{1t} \sim N(0, \tau_\beta^2), \quad \gamma_t = \gamma_{t-1} + \omega_{2t}, \quad \omega_{2t} \sim N(0, \tau_\gamma^2), \quad \alpha_k \sim N(0, \tau_\alpha),$$

onde $temp_t$ e $umid_t$, representam a média a posteriori da temperatura máxima padronizada e da umidade relativa padronizada para cada tempo e $(seg_t, ter_t, qua_t, qui_t, sex_t, sab_t)$ são as variáveis indicadoras do dia da semana. Não conseguimos ajustar os Modelos 2 e 4, onde o parâmetro de decaimento da função de correlação exponencial varia ao longo do tempo para estes dados. Uma possível explicação para isso é o fato de que existe uma grande quantidade de dados faltantes e o parâmetro λ_t é um parâmetro de difícil estimação. Para algumas estações, a proporção de dados faltantes chega a 37% das observações. Assim, apresentaremos os resultados obtidos com o ajuste dos Modelos 1 e 3 e iremos focar sobre como o nosso modelo é capaz de prever os dados faltantes.

As priors utilizadas foram $\sigma^2 \sim IG(1, 1)$, $\alpha_k \sim N(0, 10)$, $k = 1, \dots, 6$, $(\beta_0, \gamma_0)' \sim N_2(\mathbf{0}, 10\mathbf{I})$. Os hiperparâmetros da normal associada para λ foram $\mu_\lambda = -1,84$ e $\tau_\lambda^2 = 1,11$, escolhidos de maneira que a partir de 11 km (metade da distância máxima) a correlação seja menor que 0,05, e $P(\lambda > 6\mu_\lambda) = 0,10$. Já para o Modelo 3, os hiperparâmetros foram $\mu_\sigma = 2,25$ e $\tau_\sigma^2 = 0,10$, usando a priori lognormal. Os hiperparâmetros foram escolhidos de maneira que a média fosse igual a variância observada com variância tal que $P(\sigma_2^2 > 10\mu_\sigma) = 0,90$.

Foram geradas 30.000 iterações para 2 cadeias. Após 5.000 iterações os parâmetros convergiram e foi considerada uma amostra final de tamanho 1.000. Através do critério de Gelman e Rubin (1992) verificamos a convergência dos parâmetros. A variância da proposta de transição foi encontrada após 5000 iterações do MCMC. Essas iterações foram utilizadas apenas para encontrar a variância da proposta. A taxa de aceitação para λ foi aproximadamente igual a 25% no Modelo 1. Enquanto para σ_t^2 no Modelo 3, a taxa ficou entre 25% e 34%.

Os resultados do ajuste dos dois modelos para os parâmetros fixos, encontram-se na Tabela 3.5. Os coeficientes dos efeitos dos dias da semana de segunda e quinta são positivos e significativos, para

os dois modelos. Os intervalos de 95% de credibilidade dos coeficientes α_1 e α_4 não contêm o zero. Já o coeficiente de sexta (α_5), é negativo e significativo para o Modelo 1, enquanto para o Modelo 3 não encontramos significância estatística. A média a posteriori do coeficiente autoregressivo ϕ variou entre 0,979 e 0,992, indicando a forte dependência entre os níveis de PM_{10} corrigidos pela estrutura de média, no tempo t e $t - 1$. As variâncias de evolução τ_β^2 e τ_γ^2 foram estimadas usando fator de desconto $\delta = 0,95$, devido a dificuldade de estimação para este parâmetro como visto na Seção 3.7.2.

Parâmetro	Modelo 1		Modelo 3	
	Média	IC(95%)	Média	IC(95%)
ϕ	0,979	(0,964; 0,994)	0,992	(0,981; 0,999)
α_1	0,184	(0,002; 0,356)	0,363	(0,051; 0,656)
α_2	-0,089	(-0,259; 0,076)	0,127	(-0,195; 0,446)
α_3	-0,118	(-0,284; 0,032)	0,075	(-0,240; 0,448)
α_4	0,171	(0,008; 0,331)	0,460	(0,083; 0,873)
α_5	-0,201	(-0,367; -0,027)	0,002	(-0,400; 0,421)
α_6	0,133	(-0,024; 0,282)	0,273	(-0,080; 0,699)

Tabela 3.5: Média e intervalo de 95% de credibilidade a posteriori para os parâmetros fixos no tempo relativos aos modelos 1 e 3 ajustados para os dados de PM_{10} .

Com o objetivo de verificar a capacidade de ajuste dos modelos aos níveis observados de PM_{10} , calculamos uma estatística sugerida por Gelfand e Ghosh (1998) (ver apêndice A). A estatística $D(m)$ foi calculada para cada uma das estações e para cada tempo, ou seja, calculamos $D(m)_i$, $i = 1, \dots, 18$ e $L = 46$ e $D(m)_i$, $i = 1, \dots, 46$ e $L = 18$, respectivamente, na equação (A.1). A Figura 3.29 e a Tabela 3.6 apresentam os resultados para os Modelos 1 e 3. Verificamos valores menores da estatística D para o Modelo 3. Além disso, observamos que apesar da complexidade do Modelo 3, a componente P , que penaliza a complexidade do modelo, é menor para o Modelo 3 quando comparada ao Modelo 1. Este resultado é confirmado quando observamos este valor para cada estação e para cada tempo, sugerindo então que segundo este critério, o Modelo 3 é mais adequado que o Modelo 1 para o ajuste dos dados de PM_{10} . Assim, os resultados que iremos apresentar daqui em diante são provenientes do Modelo 3.

Modelo	P	G	D
1	27112	12771	33497
3	22953	11213	28559

Tabela 3.6: Critério de Gelfand e Gosh e suas componentes, para os modelos ajustados aos dados de PM_{10} .

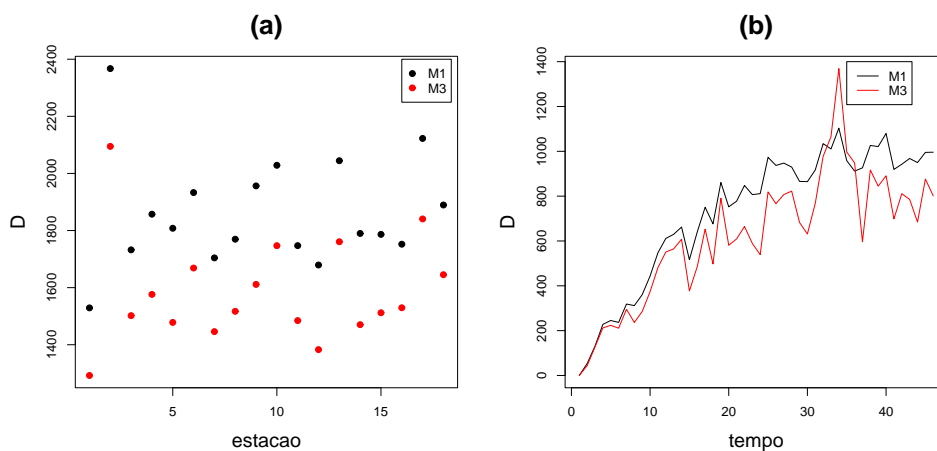


Figura 3.29: Critério de Gelfand e Gosh para cada estação (a) e tempo (b), para os modelos ajustados aos dados de PM_{10} .

O coeficiente da temperatura é positivo e significativo para todos os tempos, indicando que com o aumento da temperatura, observamos um aumento nos níveis de PM_{10} . Enquanto o coeficiente da umidade, muda de sinal ao longo do tempo e não é significativo. Os intervalos de 95% de credibilidade a posteriori incluem o zero (linha vermelha horizontal), sugerindo que a umidade relativa não ajuda a explicar a $\sqrt{PM_{10}}$. Podemos observar também que estes coeficientes realmente variam ao longo do tempo, como pode ser visto na Figura 3.30. Na Figura 3.31 encontram-se o histograma da amostra da distribuição a posteriori para o parâmetro λ usando a priori lognormal, bem como a média e o intervalo de credibilidade de 95% a posteriori para o parâmetro σ_t^2 . Comparando com a Figura 3.28 podemos notar que as estimativas para σ_t^2 conseguem captar relativamente bem a estrutura da variância observada ao longo do tempo, isto é, parece fazer sentido dizer que a variância evolui

temporalmente.

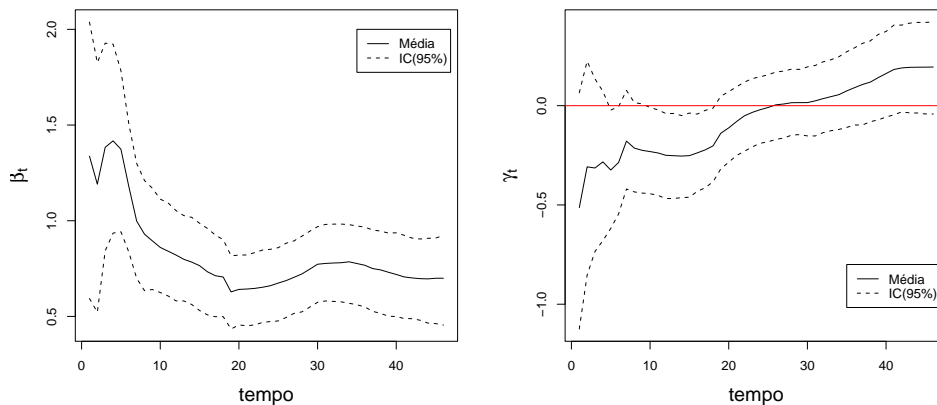


Figura 3.30: Média (linha cheia) e os respectivos limites de 95% do intervalo de credibilidade (linha tracejada) a posteriori para β_t e γ_t , segundo o Modelo 3 para os dados de PM_{10} .

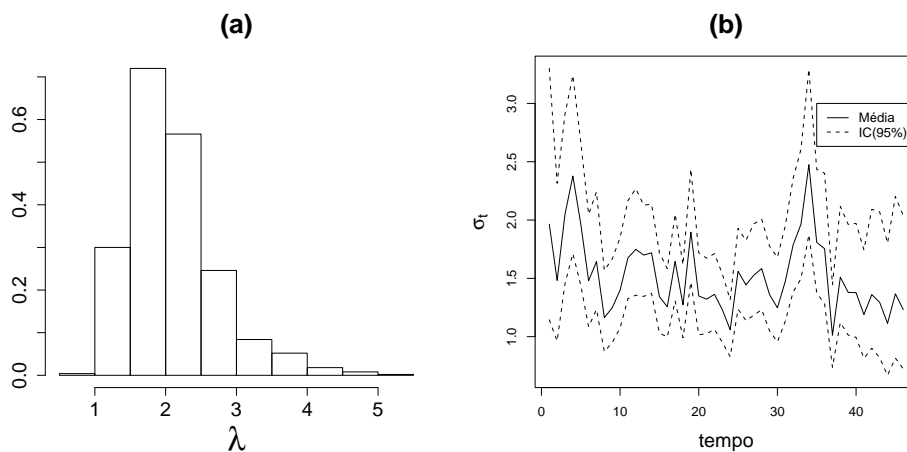


Figura 3.31: (a) Histograma da amostra da distribuição a posteriori para o parâmetro da função de correlação, λ . (b) Média e intervalo de 95% de credibilidade a posteriori para variância, σ_t^2 , segundo o Modelo 3 para os dados de PM_{10} .

A Figura 3.32 mostra a média e o intervalo de 95% de credibilidade a posteriori para os dados faltantes em duas estações Sumaré e Jacarepaguá. As estimativas parecem bem razoáveis e no caso de Jacarepaguá, que é uma estação localizada mais distante das outras, observamos um intervalo de credibilidade crescente no final da série. Este resultado provavelmente deve-se a quantidade de dados faltantes no final da série.

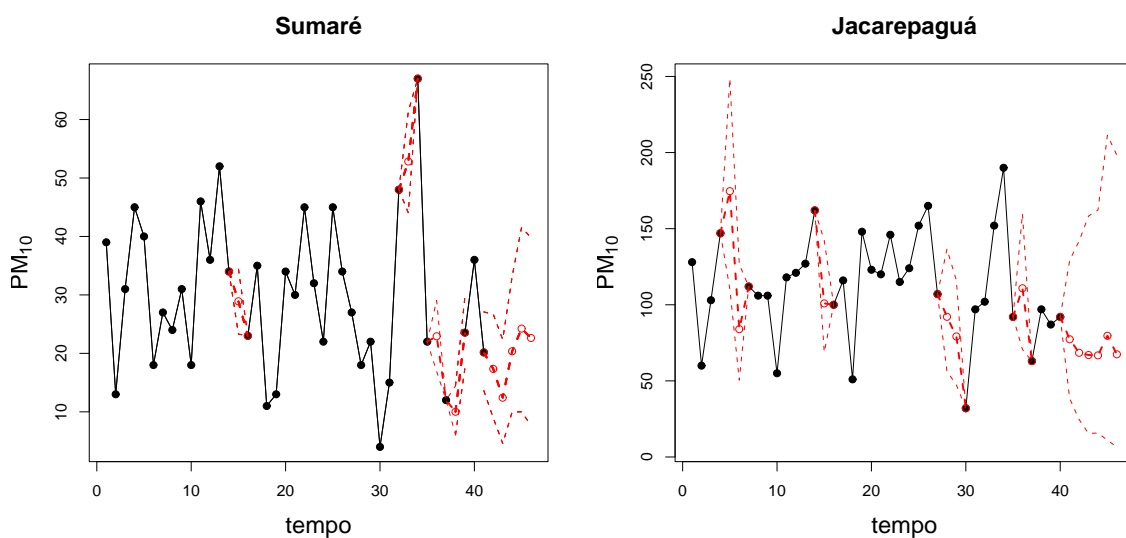


Figura 3.32: Média e o respectivo intervalo de 95% de credibilidade a posteriori para os dados faltantes nas estações Sumaré e Jacarepaguá, segundo o Modelo 3 para os dados de PM_{10} .

Usando o mesmo critério utilizado para os dados simulados, ajustamos o modelo retirando duas estações, Sumaré e Arcoverde, e em seguida realizamos uma interpolação espacial para estas estações. Estas estações foram escolhidas por serem estações com poucos dados faltantes. A previsão ao longo do tempo encontra-se na Figura 3.33. A média a posteriori para Sumaré encontra-se bem próxima do valor verdadeiro. As estimativas para Arcoverde estão mais distantes, entretanto, ainda dentro do intervalo de 95% de credibilidade a posteriori para a maioria dos tempos.

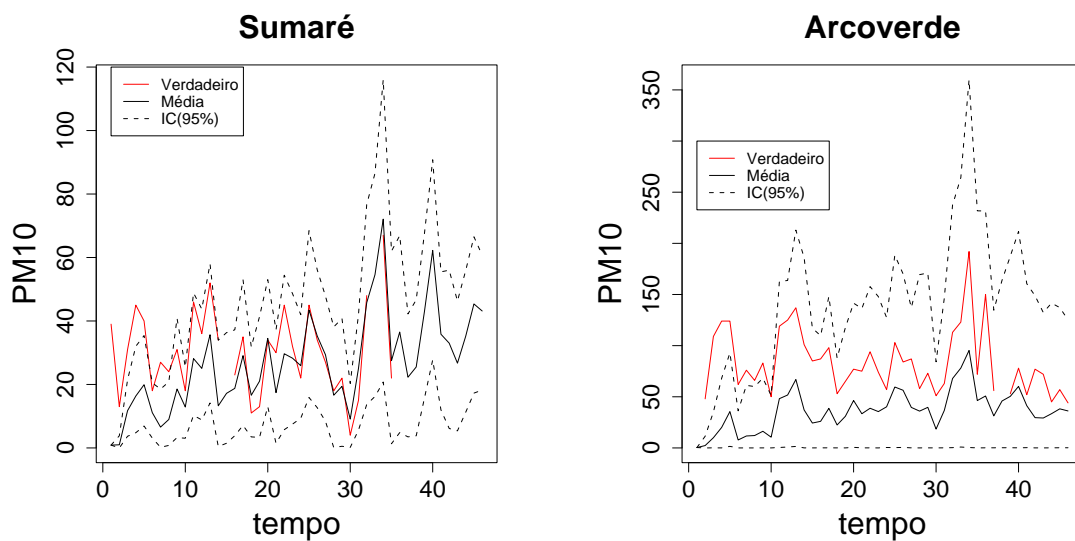


Figura 3.33: Média e intervalo de 95% de credibilidade a posteriori da distribuição preditiva de PM_{10} para as estações Sumaré e Arcoverde, segundo o Modelo 3.

Finalmente, uma grade regular de 400 pontos é criada e é feita uma interpolação espacial nesta grade para 8 tempos, 25/04/02 a 06/06/02 para o Modelo 3. As superfícies estimadas pelo modelo usando a mediana a posteriori encontram-se na Figura 3.34. Como esperado, observamos níveis de PM_{10} mais altos na segunda e quinta-feira e mais baixos no domingo. O intervalo interquartílico a posteriori para o mesmo período é apresentado na Figura 3.35. Observamos que o intervalo interquartílico parece aumentar ao longo do tempo. Talvez isto deva-se ao fato de que a maioria dos dados faltantes esteja no final das séries.

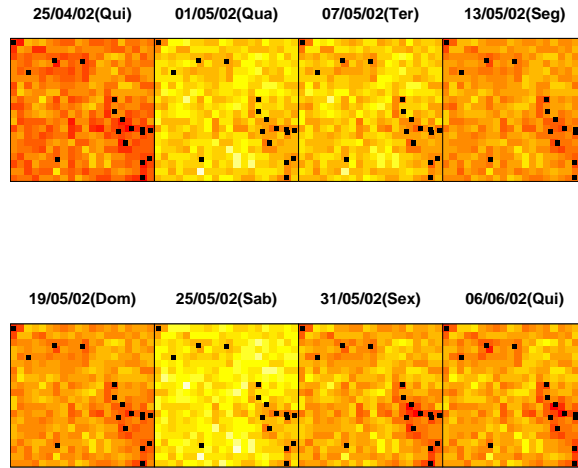


Figura 3.34: Mediana a posteriori da distribuição preditiva dos níveis de PM_{10} entre 25/04/02 e 06/06/02, segundo o Modelo 3.

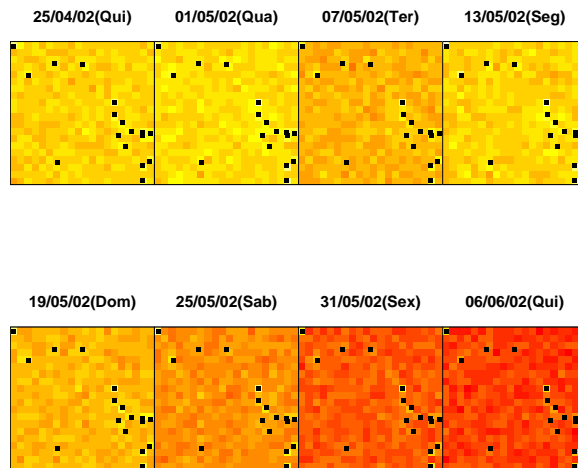


Figura 3.35: Intervalo interquartilico a posteriori da distribuição preditiva dos níveis de PM_{10} entre 25/04/02 e 06/06/02, segundo o Modelo 3.

3.7.4 Análise da Temperatura na Costa Central da Califórnia

Os dados consistem de medidas mensais de temperatura em graus centígrados em 23 estações monitoradoras localizadas na 5ª Região Climática da Califórnia como definida pelo *National Oceanic and Atmospheric Administration* (NOAA). As séries temporais correspondem ao período de janeiro de 1992 a dezembro de 2002. Os dados foram obtidos através do site do *Western Regional Climate Center* (WRCC), www.wrcc.dri.edu. A Figura 3.36 mostra a localização das estações monitoradoras bem como a superfície da altitude. As séries temporais para as 23 estações são apresentadas na Figura 3.37. Em todas as séries observamos um claro padrão sazonal anual com amplitudes e níveis variando suavemente. Temperaturas mais baixas são observadas em estações com altitudes mais altas. Parece existir uma relação linear entre temperatura e altitude.

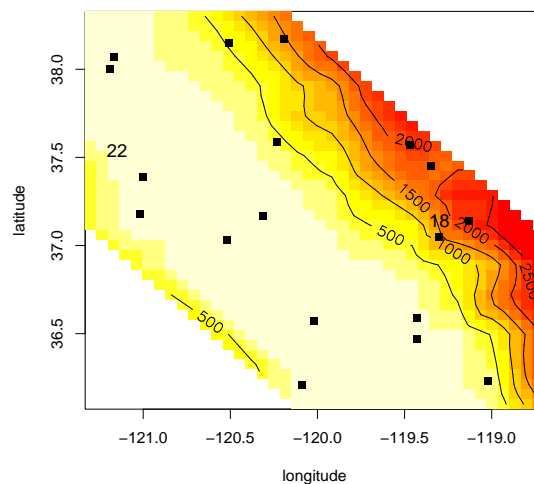


Figura 3.36: Localização das estações monitoradoras no estado da Califórnia. A parte central da área corresponde ao *Central Valley*. A serra corresponde a fronteira norte-leste, enquanto a costa corresponde a fronteira sul-oeste.

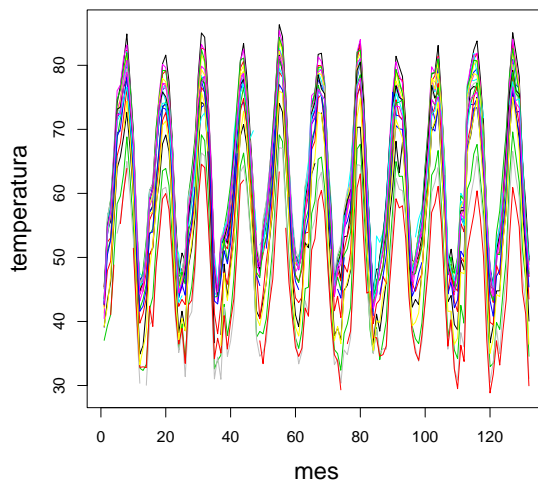


Figura 3.37: Temperatura observada em graus centígrados para as 23 estações monitoradoras no período de janeiro de 1992 a dezembro de 2002.

Baseados em uma análise preliminar dos dados, modelamos a estrutura de média usando um nível variando ao longo do tempo, β_{0t} , uma componente sazonal com um ciclo anual, α_{1t}, α_{2t} e um termo linear para a altitude, β_1 . Temos assim, uma estrutura de média variando no tempo com o objetivo de capturar possíveis tendências globais na temperatura. O modelo proposto é especificado da seguinte maneira

$$\begin{aligned}
 temp_t &= F_t \theta_t + \phi K_t K_{t-1}^{-1} (temp_{t-1} - F_{t-1} \theta_{t-1}) + v_t, \quad v_t \sim N(0, \Sigma_t), \\
 F_t \theta_t &= \beta_{0t} + E \alpha_t + \beta_1 alt, \\
 \theta_t &= G \theta_{t-1} + \omega_t, \quad \omega_t \sim N(0, W_t),
 \end{aligned}$$

onde $\theta_t = (\beta_{0t}, \alpha_{1t}, \alpha_{2t})$, $F_t = (1, 1, 0)$. O vetor $(\alpha_{1t}, \alpha_{2t})$ corresponde ao primeiro harmônico da representação de Fourier da componente sazonal. Neste caso, a matriz de evolução G é dada por

$$G = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos(2\pi/12) & \sin(2\pi/12) \\ 0 & -\sin(2\pi/12) & \cos(2\pi/12) \end{pmatrix}$$

A correlação espacial é descrita através da função de correlação Matérn com parâmetros, λ , representando o decaimento da função de correlação e, κ , representando a suavidade do processo. Quatro modelos foram ajustados considerando as diferentes especificações na estrutura de covariância como especificada na Tabela 3.1, usando a priori lognormal.

As seguintes distribuições a priori foram consideradas. Para κ usamos $N(1, 0.3^2)$ seguindo Whittle (1954) que sugere fixar $\kappa = 1$. Para a variância nos Modelos 1 e 2, usamos uma distribuição gama cujos parâmetro de forma igual a 1 e de escala é igual a $1/10$, isto é, uma distribuição com média 10, igual a variância observada. Para obter uma distribuição a priori para σ_1^2 nos modelos 3 e 4, usamos a informação da distribuição a posteriori obtida a partir do Modelo 1. Os parâmetros da normal associada resultantes foram $\mu_\sigma = 0,82$ e $\tau_\sigma^2 = 0,009$. Para obter uma distribuição a priori para λ , usamos a suposição de que a correlação entre duas estações é igual a 0.05 quando a distância entre elas é igual a metade da distância máxima observada, para uma função de correlação Matérn com $\kappa = 1$. Assim, os parâmetros da distribuição normal associada foram $\mu_\lambda = 0,052$ e $\sigma_\lambda^2 = 0,36$. Os parâmetros da distribuição a priori para λ_1 nos Modelos 2 e 4 foram obtidos a partir da posteriori de λ no Modelo 1. O fator de desconto δ (ver equação (3.6)) foi fixado em 0.995 para todos os modelos. As prioris para os parâmetros da estrutura de média foram $\beta_{00} \sim N(0, 10^2)$, α_{10} e α_{20} ambas $N(0, 10^2)$ e para $\beta_1 \sim N(0, 10^2)$. A variância de evolução, neste caso foi estimada usando a idéia de fator de desconto. Para ϕ atribuímos uma distribuição a priori uniforme em $(-1, 1)$.

Foram geradas 5000 iterações. Após 1000 iterações os traços dos parâmetros estabilizaram para diferentes valores iniciais e essas iterações foram descartadas. A Tabela 3.7 mostra a média e o intervalo de 95% de credibilidade a posteriori para os parâmetros fixos no tempo para todos os modelos ajustados. Notamos que em todos os casos o efeito estimado da altitude é negativo e significativo. A cada 100 metros de altitude, observamos um decréscimo de aproximadamente 0,5 graus centígrados de temperatura. O parâmetro que controla a suavidade na função de correlação Matérn é bem menor do que 1, fornecendo evidência de que o processo é não diferenciável. O coeficiente do processo autoregressivo, ϕ foi estimado em torno de 0,77 para todos os modelos, indicando que mesmo após a inclusão da estrutura de média, ainda resta uma estrutura autoregressiva. Assim, um MLD com erros independentes provavelmente não seria apropriado.

Modelo	β_1 (Deg. Cel./100 mts.)		ϕ		κ	
	Média	IC(95%)	Média	IC(95%)	Média	IC(95%)
M1	-0.548	(-0.570; -0.527)	0.753	(0.729; 0.776)	0.209	(0.186; 0.233)
M2	-0.544	(-0.567; -0.521)	0.770	(0.746; 0.796)	0.175	(0.164; 0.193)
M3	-0.590	(-0.620; -0.562)	0.788	(0.762; 0.815)	0.210	(0.187; 0.231)
M4	-0.591	(-0.619; -0.562)	0.792	(0.767; 0.820)	0.190	(0.179; 0.208)

Tabela 3.7: Média e intervalo de 95% de credibilidade a posteriori para os parâmetros fixos no tempo para os dados de temperatura.

A Tabela 3.8 apresenta os valores da estatística proposta por Gelfand e Ghosh (1998) para cada um dos modelos ajustados. O Modelo 3 tem o menor valor de D . Como esse resultado consiste de uma soma dupla no tempo e no espaço, calculamos as componentes P , G e D para cada tempo e cada observação no espaço. Observe que o Modelo 3 tem os menores valores de D para a maioria das observações no tempo e no espaço (Figura 3.38). Assim as estimativas que mostraremos a partir de agora são baseadas no Modelo 3.

Model	P	G	D
1	23621	12562	29902
2	44214	12639	50534
3	22982	13573	29769
4	26781	13536	33549

Tabela 3.8: Critério de Gelfand e Gosh e suas componentes, para os modelos ajustados aos dados de temperatura.

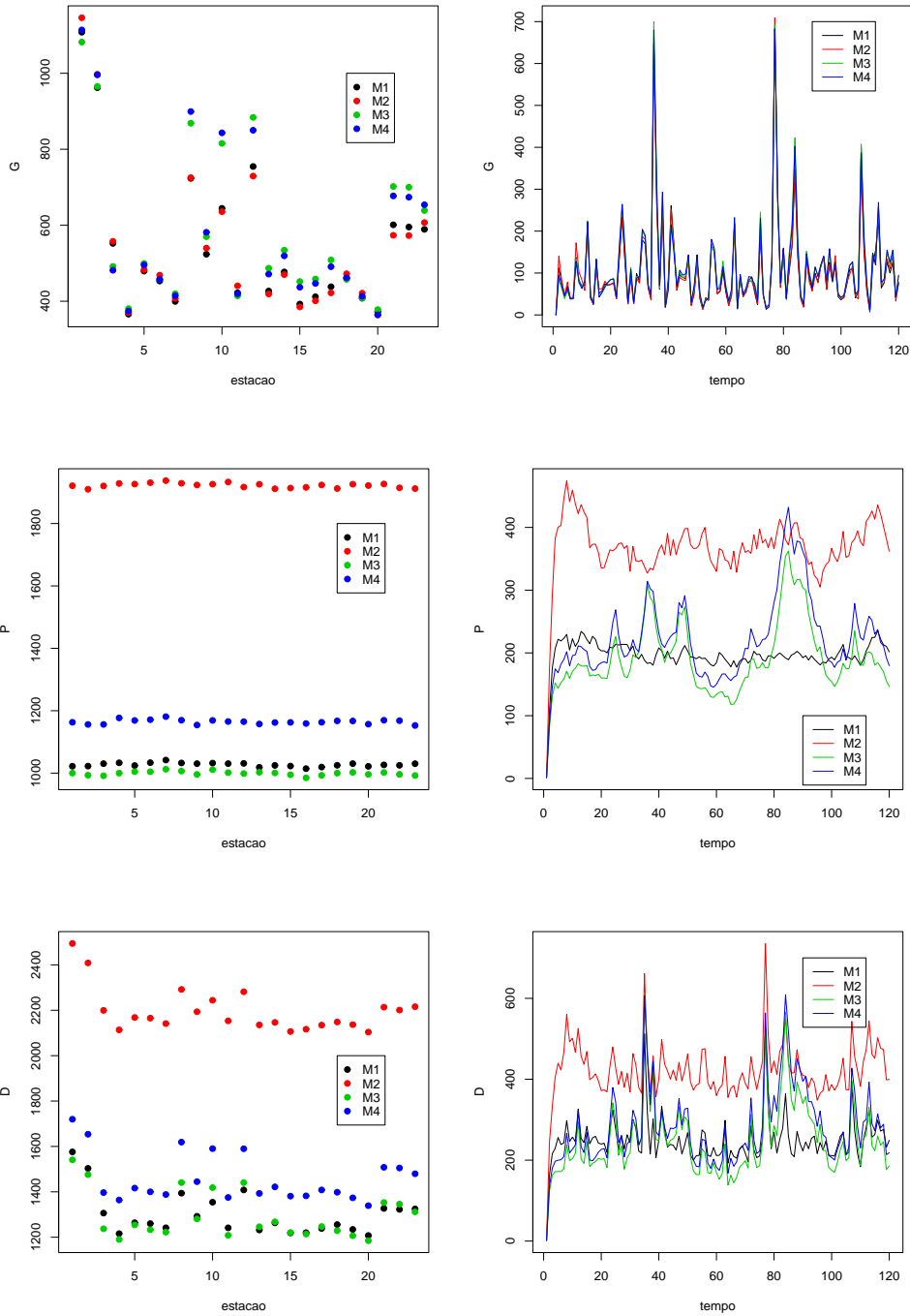


Figura 3.38: Valores das componentes G , P e D para cada estação (esquerda) e tempo (direita) para os dados de temperatura.

Na Figura 3.39 mostramos a média e o intervalo de 95% de credibilidade a posteriori para o nível β_{0t} e para a amplitude, $\sqrt{\alpha_{1t}^2 + \alpha_{2t}^2}$. Aparentemente não existe tendência para o nível, mas um leve crescimento é observado para a amplitude. Isto sugere que a diferença entre meses mais quentes e mais frios são de aproximadamente um grau centígrado durante os 10 anos observados. A Figura 3.40 mostra a distribuição a posteriori para os parâmetros da estrutura de covariância, λ e σ_t^2 . Os valores estimados para λ variam entre 10 e 50, indicando uma correlação significativa entre as estações. As medianas a posteriori para λ e κ fornecem correlações que variam entre 0,59 e 0,92 entre as estações. A média e o intervalo de credibilidade a posteriori para σ_t^2 suporta a suposição de que o parâmetro de escala realmente deve variar no tempo, já que observamos intervalos bem diferentes ao longo do tempo. Este resultado confirma a análise preliminar onde as variâncias observadas são diferentes para diferentes pontos no tempo.

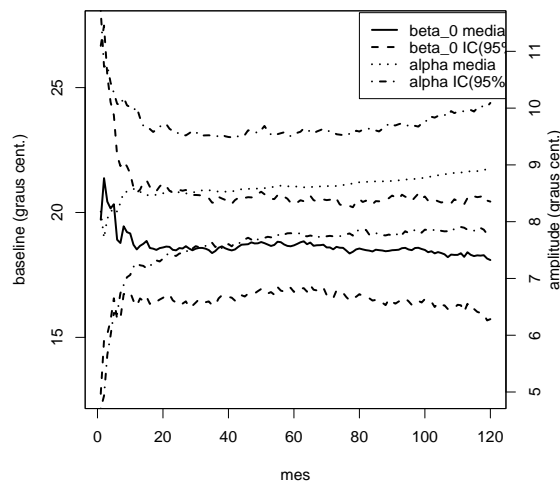


Figura 3.39: Média e os respectivos limites de 95% do intervalo de credibilidade a posteriori para β_{0t} e α_t , segundo o Modelo 3.

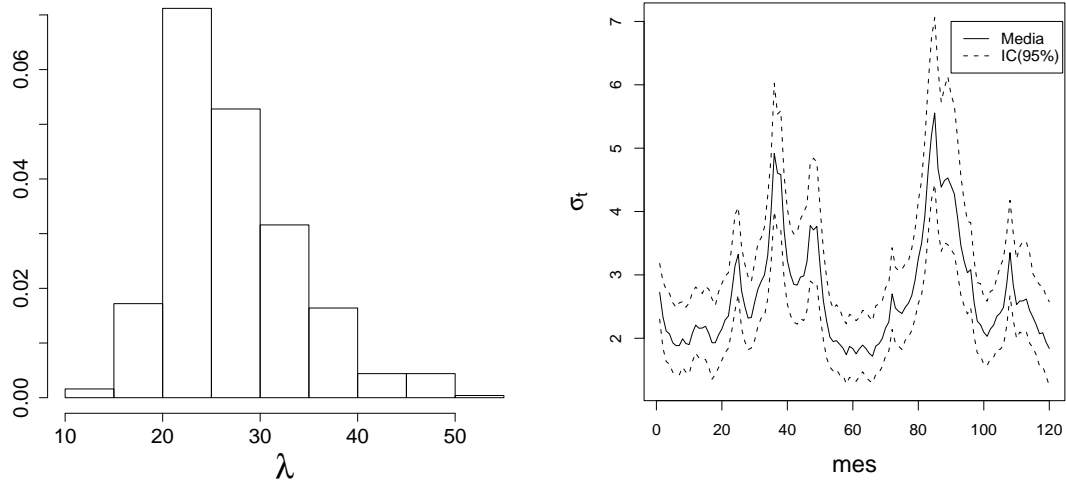


Figura 3.40: Painel esquerdo: Histograma da amostra da distribuição a posteriori para o parâmetro da função de correlação, λ . Painel direito: Média e intervalo de 95% de credibilidade a posteriori para variância, σ_t^2 , segundo o Modelo 3 para os dados de temperatura.

Com o objetivo de checar a capacidade preditiva do modelo, ajustamos o Modelo 3 deixando de fora 2 estações e calculamos a distribuição preditiva a posteriori para cada tempo t como descrito na Seção 3.6. As estações 18 e 22 foram escolhidas e suas altitudes são respectivamente 801,624 e 6,096 metros. Estas estações estão marcadas na Figura 3.36. A comparação entre os valores observados e os valores estimados pelo modelo indicam um bom ajuste (Figura 3.41).

A evolução da superfície de temperatura estimada entre o período de janeiro e agosto de 1992 é apresentado na Figura 3.42. Os resultados são compatíveis com as temperaturas médias durante estes meses. Como esperado, observamos temperaturas mais baixas na serra, onde observamos as maiores altitudes e, temperaturas mais altas no vale central, onde a altitude é menor.

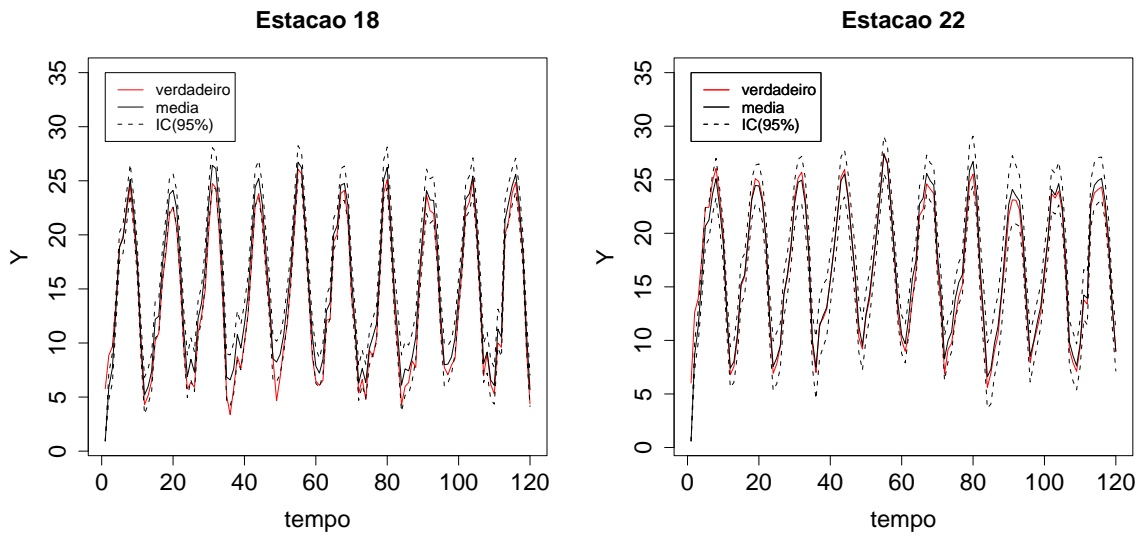


Figura 3.41: Média da distribuição preditiva a posteriori, respectivos intervalos de 95% de credibilidade e valores observados para as estações, 18 e 22, segundo o Modelo 3 para os dados de temperatura.

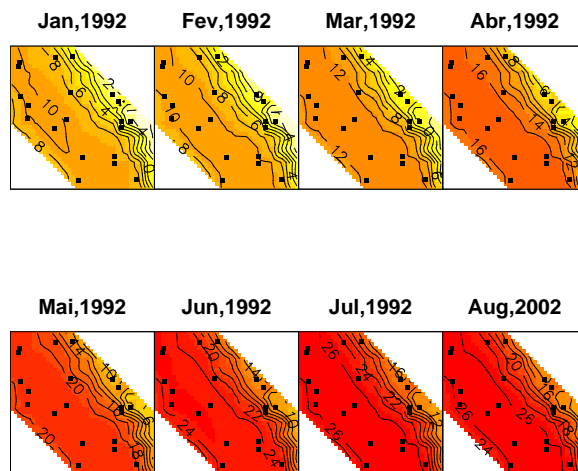


Figura 3.42: Mediana a posteriori da distribuição preditiva das temperaturas estimadas na Califórnia em graus centígrados, segundo o Modelo 3.

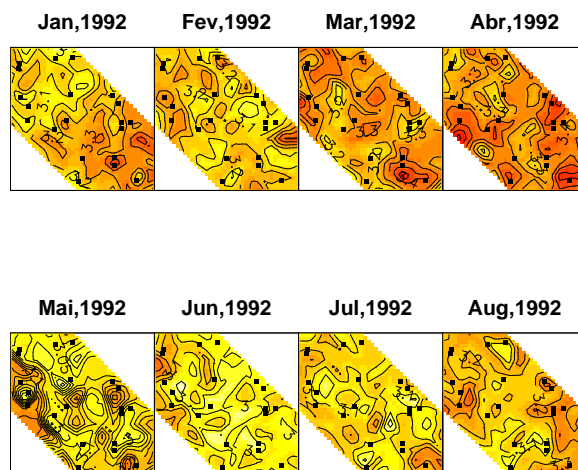


Figura 3.43: Intervalo interquartílico a posteriori da distribuição preditiva das temperaturas estimadas na Califórnia em graus centígrados, segundo o Modelo 3.

Com o objetivo de verificar a capacidade de fazer previsão temporal, o modelo foi ajustado deixando de fora um ano. Em seguida, calculamos a previsão temporal 12 passos a frente, conforme descrito na Seção 3.6. Os resultados para quatro estações encontram-se na Figura 3.44. Os resultados são bastante satisfatórios, indicando a capacidade do nosso modelo em fazer previsão temporal.

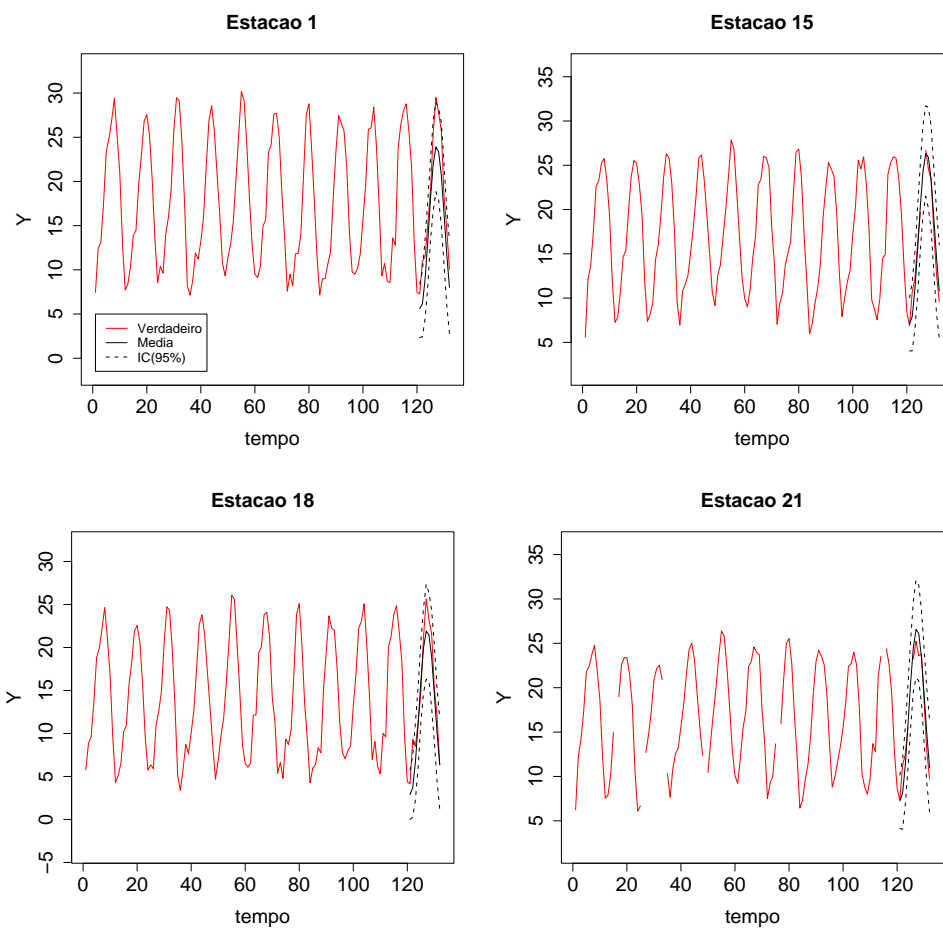


Figura 3.44: Previsão temporal 12 passos a frente para quatro estações, segundo o Modelo 3 para os dados de temperatura.

Para verificar o ajuste do modelo, calculamos a distribuição preditiva um passo a frente para cada localização. Seja $u_t(s)$ uma variável aleatória para cada localização s definida como

$$u_t(s) = P(X_{t+1} \leq y_{t+1}(s) | y_i(s), i = 1, \dots, t),$$

onde X_{t+1} denota uma variável aleatória cuja distribuição é a distribuição preditiva um passo a frente da temperatura no tempo t , localização s , condicional a todos os parâmetros e $y_{t+1}(s)$ denota a observação verdadeira. De acordo com Rosenblatt (1952), $u_t(s)$, $t = 1, \dots, T - 1$ são independentes e uniformemente distribuídos. Note que a distribuição preditiva um passo a frente da temperatura para qualquer localização é normal. Assim, para cada iteração do MCMC após a convergência, obtemos uma coleção de amostras de variáveis aleatórias que devem ser independentes e uniformemente distribuídas. Consideramos uma transformação para normalidade dada por $\Phi^{-1}(u_t(s))$ (Kim, Shepard, e Chib 1998). Para checar que estas variáveis são independentes e normalmente distribuídas, calculamos os gráficos dos quantis para cada localização e observamos que os resultados foram satisfatórios. A Figura 3.45 mostra o gráfico dos quantis teóricos e estimados da distribuição normal para a estação 1.

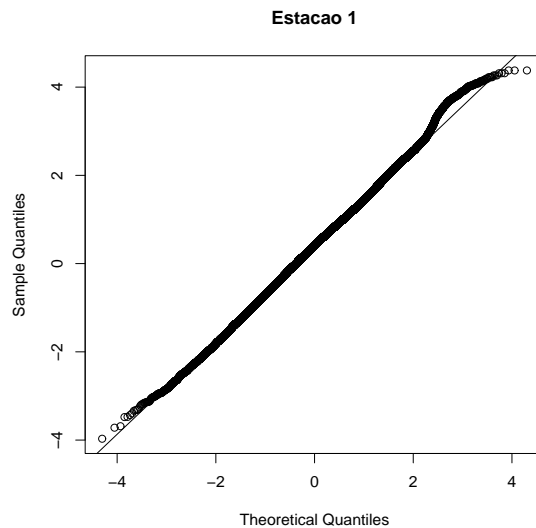


Figura 3.45: Gráfico dos quantis teóricos da distribuição normal para a estação 1, segundo o Modelo 3 para os dados de temperatura.

Neste capítulo apresentamos uma classe flexível de modelos espaço-temporais com funções de covariância não separáveis, e não estacionárias no tempo e no espaço. Essas características são obtidas através da evolução temporal dos parâmetros da função de covariância espacial. O modelo proposto pode ser escrito sob a forma dos modelos dinâmicos, e os momentos de primeira e segunda ordem são modelados conjuntamente. Os parâmetros da estrutura de média são bem estimados e os resultados da interpolação espacial e previsão temporal k passos a frente são satisfatórios. O ajuste do modelo proposto aos conjuntos de dados reais, demonstram que a evolução da variância temporal é mais intuitiva do que a evolução temporal do parâmetro que controla o decaimento da função de correlação exponencial .

Capítulo 4

Processos Autoregressivos com coeficientes variando espacialmente

Neste Capítulo é feito inicialmente uma revisão de Modelos Autoregressivos para séries temporais. Na Seção seguinte apresentamos o modelo proposto, que pode ser visto como uma extensão do modelo apresentado no capítulo anterior. Como permitimos que os coeficientes autoregressivos variem ao longo da região D , chamamos este modelo de processos autoregressivos variando espacialmente. Na Seção 4.3, discutimos as *prioris* atribuídas para os coeficientes do modelo autoregressivo. O procedimento de inferência utilizado para obtenção das amostras da distribuição a posteriori encontra-se na Seção 4.4. Na Seção 4.5 calculamos a distribuição preditiva do modelo usada para fazer previsão temporal e interpolação espacial. Finalmente, na Seção 4.6 generalizamos o modelo proposto, considerando a ordem dos processos autoregressivos desconhecida.

4.1 Modelos Autoregressivos

O modelo autoregressivo é um modelo de regressão linear no qual as variáveis regressoras são os valores do processo em p instantes anteriores (Box, Jenkins, e Reinsel 1994). Vamos denotar os valores do processo no tempo $t, t - 1, t - 2, \dots$ por $X_t, X_{t-1}, X_{t-2}, \dots$. Assim,

$$X_t = \sum_{j=1}^p \phi_j X_{t-j} + \epsilon_t, \quad (4.1)$$

é chamado modelo autoregressivo de ordem p , ou $AR(p)$, onde (ϕ_1, \dots, ϕ_p) é um vetor de coeficientes e ϵ_t é um processo aleatório normal com média zero e variância τ^2 . Seja B o operador *backward* definido por $BX_t = X_{t-1}$. O modelo $AR(p)$ pode ser escrito através do operador autoregressivo de ordem p , $\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$ da seguinte forma,

$$\phi(B)X_t = \epsilon_t,$$

ou ainda,

$$X_t = \psi(B)\epsilon_t,$$

onde $\psi(B) = \phi(B)^{-1}$. Os processos autoregressivos podem ser estacionários ou não estacionários. Para que o processo seja estacionário, os coeficientes ϕ_1, \dots, ϕ_p devem ser escolhidos de maneira que a série $\psi(B)$ seja convergente, isto é, finita.

Vamos considerar, por exemplo, o processo autoregressivo de primeira ordem, $AR(1)$, isto é,

$$X_t = \phi X_{t-1} + \epsilon_t,$$

ou ainda, usando o operador diferença

$$(1 - \phi B)X_t = \epsilon_t.$$

Assim,

$$X_t = \frac{1}{(1 - \phi B)}\epsilon_t = (1 + \phi_1 B + \phi^2 B^2 + \dots)\epsilon_t = \sum_{j=0}^{\infty} \phi^j \epsilon_{t-j}.$$

Logo, $\psi(B) = (1 - \phi B)^{-1} = \sum_{j=0}^{\infty} \phi^j B^j$. Assim, $\psi(B)$ converge se $|\phi| < 1$, garantindo assim a estacionariedade do processo. Esta condição é equivalente a dizer que a raiz do polinômio $\phi(B) = 0$ está fora do círculo unitário. O polinômio $\phi(B) = 0$ é conhecido também como polinômio característico.

Voltando ao caso geral de um $AR(p)$, temos que o polinômio característico é dado por

$$\phi(B) = 1 - \sum_{j=1}^p \phi_j B^j = \prod_{j=1}^p (1 - \alpha_j B),$$

onde $(\alpha_1, \dots, \alpha_p)$ são as raízes recíprocas de $\phi(B) = 0$. Usando o operador diferença podemos reescrever a equação (4.1) como

$$\prod_{j=1}^p (1 - \alpha_j B)X_t = \epsilon_t.$$

De acordo com esta parametrização o processo é estacionário se e somente se $|\alpha_j| < 1, \forall j$, ou seja, se as raízes de $\phi(B) = 0$ estiverem fora do círculo unitário.

A função de autocorrelação de um processo $AR(p)$ estacionário com média $E(X_t) = 0$ e covariância $\gamma(k) = E(X_t, X_{t+k})$ é dada por

$$\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2} + \dots + \phi_k \rho_{k-p}, \quad k > 0,$$

onde $\gamma(0) = Var(X_t)$ e $\rho_k = \frac{\gamma(k)}{\gamma(0)}$. Para maiores detalhes ver Box, Jenkins, e Reinsel (1994). A equação acima é chamada de equações de Yule-Walker e elas são utilizadas para obter estimativas dos coeficientes ϕ_1, \dots, ϕ_p através das autocorrelações estimadas. A variância de um processo $AR(p)$ como função das autocorrelações é dada por

$$Var(X_t) = \frac{\tau^2}{1 - \rho_1 \phi_1 - \rho_2 \phi_2 - \dots - \rho_p \phi_p}.$$

O processo $AR(p)$ pode ser escrito na forma dos Modelos Lineares Dinâmicos através das seguintes equações

$$\begin{aligned} X_t &= F' \theta_t, \\ \theta_t &= G \theta_{t-1} + \omega_t, \end{aligned}$$

onde $F = (1, 0, \dots, 0)'$, $\theta_t = (x_t, x_{t-1}, \dots, x_{t-p+1})'$, $\omega_t = (\epsilon_t, 0, \dots, 0)$ e

$$G = \begin{pmatrix} \phi_1 & \phi_2 & \phi_3 & \dots & \phi_p \\ 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ \vdots & & \ddots & \dots & \vdots \\ 0 & 0 & \dots & 1 & 0 \end{pmatrix}.$$

Os autovalores da matriz G correspondem às raízes recíprocas do polinômio característico $\phi(B) = 0$. As raízes recíprocas ou, autovalores, α_j , podem ocorrer através de C pares de raízes conjugadas e complexas e R raízes reais, levando a $p = R + 2C$. Vamos denotar as raízes reais por a_j e os pares de raízes complexas por $r_j e^{\pm i \omega_j}$, a_j , r_j e ω_j são quantidades reais com $\omega_j > 0$. Segundo West (1997) podemos decompor os processos autoregressivos da seguinte forma

$$X_t = \sum_{j=1}^C z_{tj} + \sum_{j=1}^R y_{tj},$$

onde z_{tj} e y_{tj} são processos latentes correspondentes as raízes complexas e reais, respectivamente. Cada y_{tj} segue um processo autoregressivo de primeira ordem AR(1) com coeficiente a_j , e cada z_{tj} segue processos autoregressivos médias móveis ARMA(2, 1), cujo componente AR(2) representa o comportamento quase periódico, com período $2\pi/\omega_j$ e módulo r_j .

Os processos autoregressivos de ordem p podem ser utilizados na modelagem de dados para detecção de tendência e sazonalidade. As raízes reais positivas fornecem persistência e as raízes complexas capturam periodicidades. Estas características são comumente encontradas em dados ambientais. Quando as raízes encontram-se fora da região de estacionariedade, o processo apresenta um comportamento explosivo. Suponha agora que observamos séries temporais em diferentes localizações. Neste caso, temos um AR(p) para cada localização s , isto é,

$$X_t(s) = \sum_{j=1}^p \phi_j(s) X_{t-j}(s) + \epsilon_t(s),$$

ou

$$\prod_{j=1}^p (1 - \alpha_j(s)B) X_t(s) = \epsilon_t(s).$$

Para obter interpolações espaciais não explosivas, devemos garantir que os coeficientes para cada localização s estejam na região de estacionariedade. Se impomos distribuições a priori adequadas para as raízes $\alpha_j(s)$ obtemos estacionariedade temporal, resultando em superfícies espaciais suaves. A seção seguinte apresenta uma generalização do modelo espaço-temporal proposto no Capítulo 3, ao considerar que os coeficientes do AR(p) variam suavemente ao longo do espaço. Distribuições a priori adequadas são utilizadas, de maneira a garantir interpolações espaciais coerentes.

4.2 Generalização do Modelo de Convolução Espacial Discreta

Seja $Y_t = (Y_t(s_1), \dots, Y_t(s_n))$ o vetor de observações de um processo espaço temporal na região D . Assim, propomos o seguinte modelo

$$\begin{aligned} Y_t &= F_t' \theta_t + K X_t, \\ X_t &= \sum_{j=1}^p \Phi_j X_{t-j} + \epsilon_t, \quad \epsilon_t \sim N(0, \tau^2 I_n) \\ \theta_t &= G \theta_{t-1} + w_t, \quad w_t \sim N(0, W_t) \end{aligned} \tag{4.2}$$

onde $F_t'\theta_t$ representa a estrutura de média evoluindo no tempo e X_t , que captura as estruturais locais, é um processo autoregressivo multivariado de ordem p com coeficientes $\Phi_j = \text{diag}(\phi_j(s_1), \dots, \phi_j(s_n))$, $j = 1, \dots, p$. Seguindo (4.2) podemos reescrever o modelo da seguinte forma

$$\begin{aligned} Y_t &= F_t'\theta_t + \sum_{j=1}^p \Phi_j K X_{t-j} + K \epsilon_t \\ Y_t &= F_t'\theta_t + \sum_{j=1}^p \Phi_j (Y_{t-j} - F_{t-j}'\theta_{t-j}) + v_t \\ Y_t^* &= \sum_{j=1}^p \Phi_j Y_{t-j}^* + v_t, \quad v_t \sim N(0, \tau^2 \Sigma), \end{aligned}$$

onde $Y_t^* = Y_t - F_t'\theta_t$ e $\Sigma = KK' = \sigma^2 \rho(\|\cdot\|; \Lambda, \kappa)$ representando a matriz de covariância espacial, com λ denotando o parâmetro que controla o decaimento e κ o parâmetro que controla a suavidade do processo. Assim, Y_t^* é um AR(p) multivariado com coeficientes Φ_j , variando no espaço, e covariância correlacionada espacialmente, Σ . Vamos denotar o nosso modelo como SVAR(p). O modelo proposto considera que os coeficientes do AR(p) multivariado variam espacialmente, implicando que para cada localização temos uma série temporal modelada através de um processo autoregressivo com uma ordem comum para todas as localizações, entretanto cujos coeficientes mudam suavemente ao longo da região. Como pode ser visto, o modelo proposto aqui é uma extensão do modelo em (3.4), onde os coeficientes da componente autoregressiva variam espacialmente. Sendo assim, neste capítulo vamos focar principalmente nas prioris para estes coeficientes, já que particularidades tais como modelagem da média e estrutura de covariância já foram abordadas e discutidas no capítulo anterior.

A estrutura de covariância é dada por

$$\text{cov}(y(t, s_i), y(t', s_j)) = \sum_{m=1}^{\min(i,j)} k_{im} k_{jm} \rho(t - t'; \tau^2, \Phi(s_m)),$$

onde $\Phi(s_m) = (\phi_1(s_m), \dots, \phi_p(s_m))$. Note que neste caso a covariância é não separável mesmo considerando a matriz de covariância espacial Σ , fixa no tempo. O processo é estacionário no tempo, e não estacionário no espaço. Condicional às p primeiras observações, a verossimilhança é dada por

$$\left(\prod_{t=p+1}^T \frac{|\Sigma|^{-1/2}}{(\tau^2)^{n/2}} \right) \exp \left\{ -\frac{1}{2\tau^2} \sum_{t=p+1}^T \left(Y_t^* - \sum_{j=1}^p \Phi_j Y_{t-j}^* \right)' \Sigma^{-1} \left(Y_t^* - \sum_{j=1}^p \Phi_j Y_{t-j}^* \right) \right\}. \quad (4.3)$$

Vamos fixar $\tau^2 = 1$, já que como antes apenas o produto $\sigma^2 \tau^2$ é identificável.

Como discutido na Seção anterior, os coeficientes Φ_j , $j = 1, \dots, p$ de um $\text{AR}(p)$ estão definidos em um espaço que é muito difícil de especificar. Assim, vamos representar o modelo proposto em função das raízes recíprocas. Para uma localização s podemos representar o modelo usando as raízes recíprocas e o operador diferença da seguinte forma

$$\prod_{j=1}^p (1 - \alpha_j(s)B)y_t^*(s) = v_t(s),$$

com $\alpha_j(s)$, $j = 1, \dots, p$, denotando as raízes recíprocas. Suponha que $p = R + 2C$, isto é, as raízes recíprocas ocorrem como R raízes reais e C pares de raízes complexas e conjugadas. Vamos denotar as raízes recíprocas reais por a_j e os pares de raízes complexas por $r_j e^{\pm i\omega_j}$; a_j , r_j e ω_j são quantidades reais com $\omega_j > 0$. Assim, podemos escrever a série temporal decomposta como

$$\prod_{j=1}^R (1 - a_j(s)B) \prod_{j=1}^C (1 - r_j(s)e^{i\omega_j(s)}B)(1 - r_j(s)e^{-i\omega_j(s)}B)y_t^*(s) = v_t(s),$$

ou ainda,

$$\prod_{j=1}^R (1 - a_j(s)B) \prod_{j=1}^C (1 - 2r_j(s)\cos(\omega_j(s))B + r_j^2(s)B^2)y_t^*(s) = v_t(s).$$

A relação entre os coeficientes do processo autoregressivo, Φ_j e as raízes recíprocas reais e complexas, a_j , r_j e ω_j é mostrada no apêndice B.

Seguindo a função de verossimilhança em (4.3), os parâmetros do modelo são $(\sigma^2, \lambda, \kappa, \theta_t; t = 1, \dots, T, a_j, j = 1, \dots, R, r_j, \omega_j, j = 1, \dots, C)$, onde $a_j = \text{diag}(a_j(s_1), \dots, a_j(s_n))$, $r_j = \text{diag}(r_j(s_1), \dots, r_j(s_n))$ e $\omega_j = \text{diag}(\omega_j(s_1), \dots, \omega_j(s_n))$.

4.3 Distribuição a priori para os coeficientes do $\text{SVAR}(p)$

Com o objetivo de garantir estacionariedade do processo, os coeficientes da componente $\text{AR}(p)$ (ϕ_1, \dots, ϕ_p) devem ser restritos a um espaço que é muito difícil especificar. Valores no limite da região de estacionariedade produzem séries temporais com padrões cíclicos, e valores fora da região de estacionariedade, produzem processos com comportamento explosivo que pode trazer problemas para fazer previsões (Box, Jenkins, e Reinsel 1994). Do ponto de vista bayesiano, prioris adequadas para os coeficientes do $\text{AR}(p)$ garantem a estacionariedade do processo. Então, se impomos uma priori para as raízes do polinômio característico $\alpha_j, j = 1, \dots, p$, com a restrição $|\alpha_j| < 1$ induzimos uma priori para

(ϕ_1, \dots, ϕ_p) que garante a estacionariedade do processo $AR(p)$. Huerta e West (1999) propõem uma classe de prioris considerando essa parametrização. No caso do $SVAR(p)$, precisamos garantir que o vetor de coeficientes seja restrito a um determinado espaço, de maneira que satisfaça as condições de estacionariedade.

Condicional a C e R , assumimos prioris independentes para as raízes recíprocas α_j , distinguindo entre os casos real e complexo. Inicialmente vamos descrever as distribuições a priori para as raízes reais.

Raízes reais

Para garantir que os coeficientes resultantes estejam na região de estacionariedade, impomos uma estrutura a priori para as raízes reais tal que,

$$a_j(s) = \frac{2 \exp(\Psi_j(s))}{1 + \exp(\Psi_j(s))} - 1, \quad (4.4)$$

onde Ψ_j é um processo Gaussiano com média zero e matriz de covariância espacial, Σ_j conhecida. Essa transformação garante que, $-1 < a_j(s) < 1$, para todo s e $j = 1, \dots, R$.

Raízes complexas

Seja $(r_j(s), \omega_j(s))$ os parâmetros referentes ao par de raízes complexas e conjugadas, $r_j e^{\pm i\omega_j}$, para localização s com $j = 1, \dots, C$, onde $r_j(s)$ representa o módulo e $\lambda_j(s) = 2\pi/\omega_j(s)$, o período da componente quase periódica z_{jt} correspondente a este par de raiz complexa. Para garantir estacionariedade, definimos $0 < r_j(s) < 1$ e $2 < \lambda_j(s) < \lambda_u$, com $\lambda_u = T/2$. Vamos denotar $\phi_{1j}(s) = 2r_j(s) \cos \omega_j(s)$ e $\phi_{2j}(s) = -r_j(s)^2$. Atribuímos prioris para ϕ_{1j} e ϕ_{2j} ao invés de prioris para r_j e ω_j . Temos as seguintes restrições a priori e, conseqüentemente, a posteriori para ϕ_{1j} e ϕ_{2j}

$$\phi_{1j}(s) \in (l_{1j}(s), l_{2j}(s)) \quad \text{e} \quad \phi_{2j}(s) \in (-1, 0),$$

onde $l_{1j}(s) = -2(-\phi_{2j}(s))^{1/2}$ e $l_{2j}(s) = 2 \cos(4\pi/t)(-\phi_{2j}(s))^{1/2}$.

Desta forma, atribuímos as seguintes prioris para ϕ_{1j} e ϕ_{2j} :

$$\phi_{1j}(s) = \frac{(l_{2j}(s) - l_{1j}(s)) \exp(\Psi_{1j}(s))}{1 + \exp(\Psi_{1j}(s))} + l_{1j} \quad \text{e} \quad \phi_{2j}(s) = \frac{\exp(\Psi_{2j}(s))}{1 + \exp(\Psi_{2j}(s))} - 1, \quad (4.5)$$

onde $\Psi_{1j} \sim GP(0, \Sigma_{1j})$ e $\Psi_{2j} \sim GP(0, \Sigma_{2j})$, com Σ_{1j} e Σ_{2j} conhecidas.

A escolha de Σ_j , Σ_{1j} e Σ_{2j} é um passo importante para o ajuste destes modelos. Vamos supor $\Sigma_j = \sigma_j^2 \rho(\|\cdot\|; \Lambda_j)$. Inicialmente, devemos escolher uma função de correlação que seja suave já que estamos supondo que os coeficientes variam muito suavemente no espaço. A escolha da variância σ_j^2 também é de suma importância, pois se σ_j^2 for grande, obtemos uma priori para a_j que dá peso maior para os valores extremos, assim como se σ_j^2 for pequeno estaremos atribuindo uma priori muito concentrada em torno do zero. A Figura 4.1 mostra o histograma da distribuição a priori de uma raiz real a_j em uma determinada localização s_i , ilustrando estas duas situações. Quando $\sigma_j^2 = 0,5$, o histograma é bem concentrado em torno do zero, dando alta probabilidade a priori de a_j ser zero. Quando $\sigma_j^2 = 10$, observamos uma alta probabilidade a priori nos valores extremos. Assim, escolhamos σ_j^2 de maneira que obtemos uma distribuição aproximadamente uniforme para os a_j 's, como ilustrado na Figura 4.1 quando $\sigma_j^2 = 2,5$. O parâmetro que controla o decaimento da função de correlação é escolhido baseado na idéia de distância máxima discutida no Capítulo anterior.

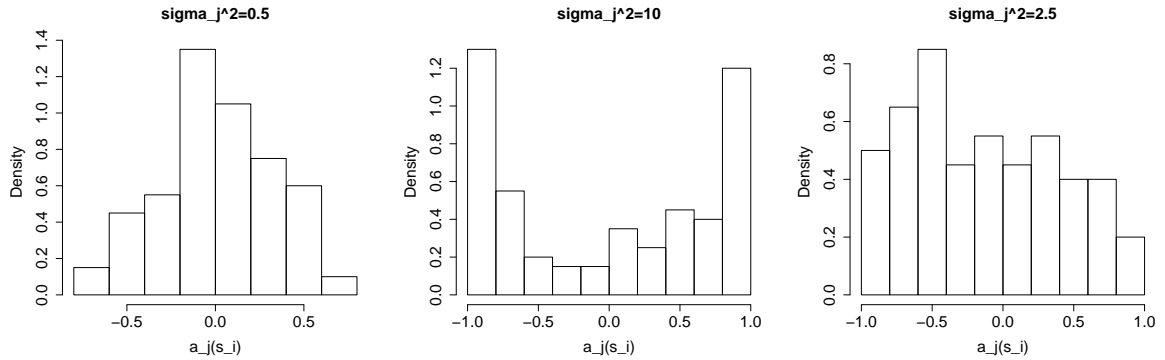


Figura 4.1: Comparação da distribuição a priori para $a_j(s_i)$, utilizando $\sigma_j^2 = 0,5$, $\sigma_j^2 = 10$ e $\sigma_j^2 = 2,5$.

4.4 Procedimento de Inferência

O procedimento de inferência para o modelo $SVAR(p)$ proposto usando a estrutura de prioris descrita na seção anterior é desenvolvido usando métodos de simulação estocástica. As amostras da distribuição a posteriori são obtidas através de métodos MCMC, tais como o amostrador de Gibbs e

Metropolis-Hastings (Gamerman e Lopes 2006).

Considerando a função de verossimilhança em (4.3) e supondo independência a priori dos parâmetros do modelo, a distribuição a posteriori conjunta é proporcional a

$$p(\Theta|y) \propto \left[\prod_{t=p+1}^T p(y_t|y_{t-p}, \dots, y_{t-1}, \Theta) \right] \left[\prod_{t=1}^T p(\theta_t|\theta_{t-1}) \right] p(\theta_0)p(\sigma^2)p(\lambda)p(\kappa) \left[\prod_{j=1}^R p(a_j) \right] \left[\prod_{j=1}^C p(\phi_{1j})p(\phi_{2j}) \right] \quad (4.6)$$

Para que o algoritmo de MCMC seja implementado, devemos calcular a condicional completa para os parâmetros do modelo. A seguir, apresentamos, o esquema de amostragem para obtenção de amostras a posteriori para cada um dos parâmetros do modelo.

Amostrando as raízes reais

Condicional as raízes reais, seja $z_{jt} = \prod_{i \neq j}^p (1 - \alpha_i B) y_t^*$ a série temporal sem a j -ésima raiz real, pode-se mostrar que z_{jt} é um AR(1) multivariado com coeficiente a_j . Assim,

$$z_{jt} = a_j z_{jt-1} + v_t,$$

onde $a_j = \text{diag}(a_j(s_1), \dots, a_j(s_n))$ com $-1 < a_j(s) < 1, \forall s$ e $v_t \sim N(0, \Sigma)$ com $\Sigma = KK'$. Como atribuímos priori para Ψ_j ao invés de a_j , vamos calcular a condicional completa para Ψ_j :

$$p(\Psi_j|\cdot) \propto \prod_{t=p+1}^T p(z_{jt}|g(\Psi_j)) \times |J_{\Psi_j}| \times p(\Psi_j),$$

onde $|J_{\Psi_j}| = \prod_{i=1}^n \frac{2 \exp(\Psi_j(s_i))}{(1 + \exp(\Psi_j(s_i)))^2}$ é o jacobiano da transformação, por atribuir priori para Ψ_j e não a_j . A distribuição condicional completa para Ψ_j não corresponde a nenhuma forma de densidade conhecida e, portanto, o algoritmo de Metropolis-Hastings será utilizado para obter amostras desta distribuição. Usaremos como distribuição proposta, $\Psi_j^p \sim GP(\Psi_j^c, U)$, isto é, um processo gaussiano centrado no valor corrente com variância U . A probabilidade de transição é dada por

$$\min \left\{ 1, \frac{p(\Psi_j^p|\cdot)q(\Psi_j^c|\Psi_j^p)}{p(\Psi_j^c|\cdot)q(\Psi_j^p|\Psi_j^c)} \right\}.$$

Assim,

$$\frac{p(\Psi_j^p|\cdot)q(\Psi_j^c|\Psi_j^p)}{p(\Psi_j^c|\cdot)q(\Psi_j^p|\Psi_j^c)} = \frac{\prod_{t=p+1}^T f_N(z_{jt}|g(\Psi_j^p), \Sigma) \times |J_{\Psi_j^p}| \times f_N(\Psi_j^p|0, \Sigma_j)}{\prod_{t=p+1}^T f_N(z_{jt}|g(\Psi_j^c), \Sigma) \times |J_{\Psi_j^c}| \times f_N(\Psi_j^c|0, \Sigma_j)},$$

onde $f_N(\cdot|\mu, \theta)$ denota a densidade de uma distribuição normal multivariada com média μ e variância θ .

Amostrando as raízes complexas

Condicional as raízes complexas, seja $z_{jt} = \prod_{i \in A_j}^p (1 - \alpha_i B) y_t^*$ a série temporal sem o j -ésimo par de raiz complexa e conjugada. Pode ser mostrado que z_{jt} é um AR(2) multivariado com coeficientes $\phi_{1j} = 2r_j \cos \omega_j$ e $\phi_{2j} = -r_j^2$ e variância Σ onde A_j representa o conjunto de todas as outras raízes. Assim,

$$z_{jt} = \phi_{1j} z_{jt-1} + \phi_{2j} z_{jt-2} + v_t,$$

onde $\phi_{1j} = \text{diag}(\phi_{1j}(s_1), \dots, \phi_{1j}(s_n))$ e $\phi_{2j} = \text{diag}(\phi_{2j}(s_1), \dots, \phi_{2j}(s_n))$. Similar ao caso anterior, calculamos a condicional completa para Ψ_{ij} , $i = 1, 2$:

$$p(\Psi_{ij} | \cdot) \propto \prod_{t=p}^T p(z_{jt} | g(\Psi_{ij})) \times |J_{\Psi_{ij}}| \times p(\Psi_{ij}),$$

onde $|J_{\Psi_{1j}}| = \prod_{i=1}^n \frac{(l2-l1) \exp(\Psi_{1j}(s_i))}{(1+\exp(\Psi_{1j}(s_i)))^2}$ e $|J_{\Psi_{2j}}| = \prod_{i=1}^n \frac{\exp(\Psi_{2j}(s_i))}{(1+\exp(\Psi_{2j}(s_i)))^2}$ são os jacobianos da transformação, por atribuir prioris para Ψ_{1j} e Ψ_{2j} ao invés de ϕ_{1j} e ϕ_{2j} , respectivamente. Para simular Ψ_{ij} , usaremos o algoritmo de Metropolis- Hastings com distribuição proposta: $\Psi_{ij}^p \sim GP(\Psi_{ij}^c, U_i)$, onde Ψ_{ij}^c é o valor corrente e U_i a variância da proposta. A razão de Metropolis-Hastings utilizada no cálculo da probabilidade de transição é dada por

$$\frac{p(\Psi_{ij}^p | \cdot) q(\Psi_{ij}^c | \Psi_{ij}^p)}{p(\Psi_{ij}^c | \cdot) q(\Psi_{ij}^p | \Psi_{ij}^c)} = \frac{\prod_{t=p+1}^T f_N(z_{jt} | g(\Psi_{ij}^p), \Sigma) \times |J_{\Psi_{ij}^p}| \times f_N(\Psi_{ij}^p | 0, \Sigma_{ij})}{\prod_{t=p+1}^T f_N(z_{jt} | g(\Psi_{ij}^c), \Sigma) \times |J_{\Psi_{ij}^c}| \times f_N(\Psi_{ij}^c | 0, \Sigma_{ij})}.$$

Amostrando os parâmetros da função de covariância

Seja $\Sigma = \sigma^2 \rho(\|\cdot\|; \lambda, \kappa)$, a função de covariância do SVAR(p), temos que

$$v_t = \prod_{j=1}^p (1 - \alpha_j B) y_t^* = y_t^* - \sum_{j=1}^p \Phi_j y_{t-j}^*,$$

onde $v_t \sim N(0, \sigma^2 V)$ com V denotando a função de correlação. Se atribuímos uma distribuição a priori gama invertida para σ^2 , isto é, $\sigma^2 \sim IG(a, b)$, a condicional completa para σ^2 é

$$\sigma^2 | \cdot \sim IG \left(\frac{n(T-p-1)}{2} + a; \frac{1}{2} \sum_{t=p+1}^T v_t' V^{-1} v_t + b \right).$$

Já para o parâmetro que controla o decaimento da função de correlação, podemos atribuir a seguinte priori, $\lambda \sim Ga \left(\frac{a\lambda}{2}, \frac{b\lambda}{2} \right)$. A distribuição condicional completa de λ não possui nenhuma forma

conhecida e, portanto, precisamos usar o algoritmo de Metropolis-Hastings para obter amostras da distribuição a posteriori. Utilizamos como distribuição proposta uma distribuição lognormal cuja normal associada é centrada no valor corrente com variância U , isto é, $\lambda^p \sim LN(\lambda^c, U)$. A razão de Metropolis-Hastings é dada por

$$\frac{p(\lambda^p|\cdot)q(\lambda^c|\lambda^p)}{p(\lambda^c|\cdot)q(\lambda^p|\lambda^c)} = \frac{\prod_{t=p+1}^T f_N(v_t|0, \Sigma^p) \times f_{Ga}\left(\frac{a_\lambda}{2}, \frac{b_\lambda}{2}\right)^p \times f_{LN}(\lambda^c|\lambda^p, U)}{\prod_{t=p+1}^T f_N(v_t|0, \Sigma^c) \times f_{Ga}\left(\frac{a_\lambda}{2}, \frac{b_\lambda}{2}\right)^c \times f_{LN}(\lambda^p|\lambda^c, U)},$$

onde $f_{Ga}(a, b)$ é a densidade da distribuição gama com parâmetros a e b , e $f_{LN}(\cdot|\mu, \theta)$ é a densidade da distribuição lognormal cuja normal associada tem média μ e variância θ .

Para o parâmetro que controla a suavidade da função de correlação, atribuímos uma priori, $\kappa \sim N(\mu_\kappa, \sigma_\kappa^2)$. A distribuição condicional completa de κ também não possui nenhuma forma conhecida e, portanto, usamos o algoritmo de Metropolis-Hastings para obter amostras da distribuição a posteriori. Como $\kappa > 0$, utilizamos como distribuição proposta uma distribuição normal truncada centrada no valor corrente com variância U , isto é, $\kappa^p \sim NT_{[0, \infty)}(\kappa^c, U)$. Dessa forma, a razão de Metropolis-Hastings é dada por

$$\frac{p(\kappa^p|\cdot)q(\kappa^c|\kappa^p)}{p(\kappa^c|\cdot)q(\kappa^p|\kappa^c)} = \frac{\prod_{t=p+1}^T f_N(v_t|0, \Sigma^p) \times f_N(\kappa^p|\mu_\kappa, \sigma_\kappa^2) \times f_{NT_{[0, \infty)}}(\kappa^c|\kappa^p, U)}{\prod_{t=p+1}^T f_N(v_t|0, \Sigma^c) \times f_N(\kappa^c|\mu_\kappa, \sigma_\kappa^2) \times f_{NT_{[0, \infty)}}(\kappa^p|\kappa^c, U)},$$

onde $f_{NT_{[a, b]}}(\cdot|\mu, \theta)$ é a densidade da distribuição normal truncada no intervalo $[a, b]$ com média μ e variância θ .

Amostrando o vetor de estados

Vamos inicialmente definir, $Z_t = Y_t - \sum_{j=1}^p \Phi_j Y_{t-j}$, condicional a (Φ_1, \dots, Φ_p) e a Σ , temos que

$$Z_t = \sum_{j=0}^p -\Phi_j F_{t-j} \theta_{t-j} + v_t, \quad v_t \sim N(0, \Sigma),$$

onde $\Phi_0 = -I$. Reescrevendo na classe dos MLD's, temos

$$\begin{aligned} Z_t &= F_t^* \theta_t^* + v_t, & v_t &\sim N(0, \Sigma_t) \\ \theta_t^* &= G^* \theta_{t-1}^* + \omega_t, & \omega_t &\sim N(0, W_t^*). \end{aligned}$$

Como visto no Capítulo 2, devemos especificar a quádrupla $\{F_t^*, G_t^*, V_t^*, W_t^*\}$ dos MLD's. A matriz de regressão dinâmica F_t^* é a concatenação de $(p+1)$ matrizes de dimensão $(n \times r)$. Cada matriz é

dada por $-\Phi_j F'_{t-j}$ com $j = 0, \dots, p$. A variância do sistema é a matriz de covariância espacial, isto é, $V_t^* = \Sigma$, com dimensão $n \times n$. Finalmente, a matriz e a variância de evolução são exatamente as mesmas discutidas na Seção 3.5.1. O algoritmo FFBS descrito na Seção 2.2.1, será utilizado para obter amostras da distribuição a posteriori para $\theta_1, \dots, \theta_T$.

Como podemos observar, precisamos especificar Φ_j , $j = 1, \dots, p$, para amostrar os parâmetros da função de covariância e o vetor de estados. O apêndice B apresenta como calcular os Φ_j 's como função das raízes recíprocas reais e complexas.

4.5 Previsão Temporal e Interpolação Espacial

Suponha agora que Y_t^u represente o vetor das observações em localizações não medidas e Y_t o vetor das observações em localizações medidas a cada tempo t . A distribuição conjunta de (Y_t, Y_t^u) , condicional ao vetor paramétrico, e às p observações anteriores é uma distribuição normal multivariada de modo que

$$\begin{pmatrix} Y_t \\ Y_t^u \end{pmatrix} \sim N \left[\begin{pmatrix} F_t^y \theta_t + \sum_{j=1}^p \Phi_j^y (Y_{t-j} - F_{t-j}^y \theta_{t-j}) \\ F_t^u \theta_t + \sum_{j=1}^p \Phi_j^u (Y_{t-j}^u - F_{t-j}^u \theta_{t-j}) \end{pmatrix}; \begin{pmatrix} \Sigma^y & \Sigma^{yu} \\ \Sigma^{uy} & \Sigma^u \end{pmatrix} \right]$$

onde F_t^y e F_t^u correspondem as matrizes de regressão dinâmica nas localizações medidas e não medidas, respectivamente. Analogamente, Φ_j^y e Φ_j^u correspondem ao vetor de coeficientes do SVAR(p) para as localizações medidas e não medidas. A matriz de covariância é dividida em 4 blocos, Σ_y é a matriz de covariância para as localizações medidas, Σ_u a matriz de covariância para as localizações não medidas e Σ_{yu} e Σ_{uy} são as matrizes de covariância entre as localizações medidas e não medidas.

A distribuição preditiva é calculada pela seguinte integral

$$p(Y_t^u | Y_t) = \int p(Y_t^u | \Theta, Y_t) p(\Theta | Y_t) d\Theta, \quad (4.7)$$

onde $p(\Theta | Y_t)$ é a distribuição a posteriori especificada em (4.6). Utilizando o resultado da partição da distribuição normal multivariada como descrito em Anderson (1994), obtemos

$$Y_t^u | Y_t, \Theta \sim N[\mu_t^u + \Sigma^{uy} (\Sigma^y)^{-1} (Y_t - \mu_t^y), (\Sigma^u - \Sigma^{uy} (\Sigma^y)^{-1} \Sigma^{yu})], \quad (4.8)$$

onde $\mu_t^y = F_t^y \theta_t + \sum_{j=1}^p \Phi_j^y (Y_{t-j} - F_{t-j}^y \theta_{t-j})$ e $\mu_t^u = F_t^u \theta_t + \sum_{j=1}^p \Phi_j^u (Y_{t-j}^u - F_{t-j}^u \theta_{t-j})$. Como podemos observar, a quantidade Φ_j^u é desconhecida e, portanto, precisaremos calcular esta quantidade antes do

cálculo da distribuição preditiva. Para obtenção de Φ_j^u utilizaremos o mesmo resultado da partição da normal multivariada, através de Ψ_j , Ψ_{1j} e Ψ_{2j} que seguem processos gaussianos com hiperparâmetros conhecidos.

A integral em (4.7) pode ser calculada usando a aproximação de Monte Carlo

$$p(Y_t^u | Y_t) \approx \frac{1}{M} \sum_{m=1}^M p(Z_t | \Theta^m, Y_t)$$

onde Θ^m denota a m -ésima amostra da distribuição a posteriori do vetor paramétrico Θ e M denota o total de amostras da distribuição a posteriori do vetor paramétrico Θ .

Para calcular a previsão temporal para k passos a frente, é necessário calcular a seguinte distribuição preditiva. Para simplificar a notação, vamos assumir $p = 1$. Assim,

$$\begin{aligned} p(Y_{T+k} | Y_{1:T}) &= \int p(Y_{T+k} | \mathbf{F}_{T+k}, Y_{T+k-1}, \theta_{T+k}, \Phi, \sigma^2, \Lambda) p(\theta_{T+k} | \theta_{T+k-1}, W) \\ & p(Y_{T+k-1} | \mathbf{F}_{T+k-1}, Y_{T+k-2}, \theta_{T+k-1}, \Phi, \sigma^2, \Lambda) p(\theta_{T+k-1} | \theta_{T+k-2}, W) \dots \\ & \dots p(\Phi, \sigma^2, \Lambda, \theta_1, \dots, \theta_T | Y_{1:T}) d(\Phi, \sigma^2, \Lambda, \theta_1, \dots, \theta_T, W). \end{aligned}$$

A integral acima pode ser aproximada por

$$p(Y_{T+k} | Y_{1:T}) \approx \frac{1}{M} \sum_{m=1}^M p(Y_{T+k} | \mathbf{F}_{T+k} \theta_{T+k}^{(m)} + \Phi^{(m)} (Y_{T+k-1} - \mathbf{F}_{T+k-1} \theta_{T+k-1}^{(m)}), \Sigma^{(m)}).$$

4.6 SVAR(p) com a ordem p desconhecida

Geralmente para encontrar a ordem de um AR(p), ajustamos modelos com diferentes ordens e decidimos qual o valor adequado de p baseado nos resíduos. Os métodos AIC (Critério de Informação de Akaike) e BIC (Critério de Informação Bayesiana) propostos por Akaike (1974) e Schwarz (1978), respectivamente, seguem esta idéia e podem ser utilizados para decidir a ordem do modelo. Do ponto de vista Bayesiano, p é visto como um parâmetro do modelo e uma priori deve ser atribuída a p . A distribuição a posteriori para p é geralmente intratável e técnicas de simulação tais como MCMC baseado em saltos reversíveis (Green 1995) podem ser utilizadas. Prado e Huerta (2002) consideram um modelo, onde os coeficientes da componente autoregressiva evoluem temporalmente através de um passeio aleatório. Além disso, a incerteza sobre a ordem do modelo é incorporada usando um passeio aleatório discreto de primeira ordem.

Uma extensão natural do modelo proposto é considerar que a ordem do $\text{SVAR}(p)$ é desconhecida. Discutimos aqui as diferentes tentativas de prioris para a modelagem da incerteza da ordem do modelo e os diversos problemas encontrados. Na Subseção 4.6.1 apresentamos uma priori de mistura com massa de probabilidade no zero e na Subseção 4.6.2 apresentamos um priori de mistura baseada em duas distribuições contínuas.

4.6.1 Priori de mistura com massa de probabilidade no zero

Raízes Reais

Inicialmente, consideramos a seguinte priori de mistura

$$p(a_j|\pi_R) = \pi_R f(a_j) + (1 - \pi_R)\delta_0, \quad j = 1, \dots, R_+,$$

onde δ_0 é uma função degenerada em zero. A nossa priori faz $a_j = 0, \forall s$, com probabilidade $1 - \pi_R$, ou $a_j \neq 0$ e densidade $f(a_j)$, com probabilidade π_R . O número de raízes reais pode ser menor do que a ordem máxima especificada, R_+ . Assim, a priori de mistura com ponto de massa em $a_j = 0$, implica em incerteza na ordem do modelo. Observamos que a ordem do modelo muda se qualquer uma das raízes for igual a zero. Utilizamos a mesma transformação utilizada anteriormente, ou seja, assumiremos que $a_j = \frac{2 \exp(\Psi_j)}{1 + \exp(\Psi_j)} - 1$ com $\Psi_j \sim GP(0, \Sigma_j)$, $a_j = \text{diag}(a_j(s_1), \dots, a_j(s_n))$ tal que $-1 < a_j(s) < 1$, $j = 1, \dots, R_+$. Assim, a densidade $f(a_j)$ resultante é dada por

$$f(a_j) = (2\pi)^{-n/2} |\Sigma_j|^{-1/2} \exp \left\{ -\frac{1}{2} \left[\log \frac{1+a_j}{1-a_j} \right]' \Sigma_j^{-1} \left[\log \frac{1+a_j}{1-a_j} \right] \right\} \prod_{i=1}^n \frac{2}{1-a_j(s_i)^2}.$$

Seja $I(a_j)$ uma variável indicadora tal que

$$I(a_j) = \begin{cases} 1, & \text{se } a_j \neq 0, \\ 0, & \text{se } a_j = 0. \end{cases}$$

Assim, $I(a_j) \sim \text{Bernoulli}(\pi_R)$ tal que $p(I(a_j) = 1) = \pi_R$, ou seja π_R é a probabilidade a priori da j -ésima raiz real estar presente. A maneira usual de tratar modelos de mistura é através da introdução de variáveis aleatórias indicadoras não observáveis. Estas variáveis especificam de qual componente da mistura cada observação é amostrada (Gelman, Carlin, Stern, e Rubin 2004). Neste caso, a variável indicadora $I(a_j)$ indica se a_j vem de uma densidade $f(a_j)$, ou se $a_j = 0$.

Huerta e West (1999) propõem uma classe de prioris considerando uma priori de mistura para as raízes recíprocas de um $AR(p)$. Esta classe de prioris, além de considerar a incerteza na ordem do modelo, considera também as raízes unitárias e permite a incorporação de valores iniciais como variáveis latentes. A priori proposta por Huerta e Prado (2006) estende a priori introduzida por Huerta e West (1999) para séries temporais multivariadas que incorpora a incerteza na ordem do modelo. Diferente do caso em que estamos considerando aqui, as séries temporais não possuem nenhuma estrutura espacial. Nossa proposta também pode ser vista como uma extensão multivariada da priori de mistura proposta por Huerta e West (1999) para o caso espacial.

Amostrando as raízes reais

Condicional as raízes reais, $z_{jt} = \prod_{i=1, i \neq j}^p (1 - \alpha_i B)y_t^*$ é um $AR(1)$ multivariado com parâmetro a_j , $-1 < a_j < 1$, e variância Σ . Assim,

$$z_{jt} = a_j z_{jt-1} + v_t.$$

Desta forma, a condicional completa para a_j é dada por

$$\begin{aligned} p(a_j | \cdot) &\propto \prod_{t=p+1}^T p(z_{jt} | a_j z_{jt-1}, \Sigma) \times p(a_j | \pi_R) \\ &= \pi_R \times \prod_{t=p+1}^T f_N(z_{jt} | a_j z_{jt-1}, \Sigma) \times f(a_j) + (1 - \pi_R) \times \prod_{t=p+1}^T f_N(z_{jt} | 0, \Sigma). \end{aligned}$$

Inicialmente devemos calcular a condicional completa de $p(I(a_j) = 1)$, isto é, a probabilidade condicional a posteriori de a_j ser gerado de $f(a_j)$,

$$p(I(a_j) = 1 | \cdot) = \frac{\pi_R \times \prod_{t=p+1}^T f_N(z_{jt} | a_j z_{jt-1}, \Sigma) \times f(a_j)}{\pi_R \times \prod_{t=p+1}^T f_N(z_{jt} | a_j z_{jt-1}, \Sigma) \times f(a_j) + (1 - \pi_R) \times \prod_{t=p+1}^T f_N(z_{jt} | 0, \Sigma)}. \quad (4.9)$$

Em seguida, geramos $u \sim U(0, 1)$. Se $u < p(I(a_j) = 1)$, fazemos $I(a_j) = 1$ e amostramos a_j através da seguinte distribuição

$$p(a_j | \cdot) \propto \prod_{t=p+1}^T p(z_{jt} | a_j z_{jt-1}, \Sigma) \times f(a_j),$$

caso contrário, $I(a_j) = 0$ e fazemos $a_j(s) = 0 \forall s$.

Dividindo o numerador e o denominador da equação (4.9), pelo numerador, simplificamos a expressão acima e obtemos:

$$p(I(a_j) = 1 | \cdot) = \frac{1}{\frac{1 - \pi_R}{\pi_R} L + 1},$$

onde $L = \frac{\prod_{t=p+1}^T f_N(z_{jt}|0, \Sigma)}{\prod_{t=p+1}^T f_N(z_{jt}|a_j z_{j,t-1}, \Sigma) \times f(a_j)}$. Se $a_j(s) = 0, \forall s$, temos que

$$\log L = -\log(f(a_j)) = \frac{n}{2} \log \frac{\pi}{2} + \frac{1}{2} \log |\Sigma_j|.$$

Como podemos observar, o cálculo da probabilidade da indicadora depende fortemente da escolha dos valores de Σ_j . Calculamos essa probabilidade para diferentes funções de covariância e diferentes escolhas de σ_j^2 , fixando $\pi_R = 0,5$. Em seguida, escolhemos duas funções de correlação: (i) a gaussiana, que é infinitamente diferenciável, ou seja, produz correlações extremamente suaves e (ii) a Matérn, com grau de suavidade igual a um, isto é, uma vez diferenciável. A escolha de σ_j^2 foi baseada na idéia discutida na Seção 4.3 de modo a obter prioris para as raízes reais, com as três características ilustradas na Figura 4.1. Se $a_j(s) = 0, \forall s$, esperamos que $p(I(a_j) = 1|.)$ seja pequena. A Tabela 4.1 apresenta o cálculo de $p(I(a_j) = 1|.)$ para as diferentes especificações de variância e função de correlação. Baseado nestes resultados, observamos que a correlação Matérn deve ser escolhida já que a correlação gaussiana pode produzir uma probabilidade muito alta indevidamente. Se compararmos as diferentes escolhas de σ_j^2 , observamos que quanto maior o valor de σ_j^2 , menor a probabilidade a posteriori de a_j ser gerado de $f(a_j)$, indicando que σ_j^2 não deve ser grande. Podemos então calcular σ_j^2 como função de $p(I(a_j) = 1|.)$ da seguinte maneira,

$$\sigma_j^2 = \exp \left\{ \frac{2}{n} \log L - \log \frac{\pi}{2} - \frac{1}{n} \log |V_j| \right\}, \quad (4.10)$$

onde $L = \frac{1-p(I(a_j)=1)}{p(I(a_j)=1)} \frac{\pi_R}{1-\pi_R}$ e V_j é a matriz de correlação espacial. Podemos então fixar essas quantidades para obter σ_j^2 .

σ_j^2	Gaussiana	Matérn
1	0,97	0,15
2,5	$3,91 \times 10^{-5}$	$1,87 \times 10^{-7}$
10	$3,64 \times 10^{-14}$	$1,74 \times 10^{-16}$

Tabela 4.1: Probabilidade a posteriori da variável indicadora, $p(I(a_j) = 1)$, usando as funções de correlação Gaussiana e Matérn, e variâncias $\sigma_j^2 = 1, 2,5$ e 10 , e fixando $\pi_R = 0,5$.

Raízes complexas

Como $\phi_{1j}(s) = 2r_j(s) \cos \omega_j(s)$ e $\phi_{2j}(s) = -r_j(s)^2$, $\phi_{2j} = 0$, implica em $\phi_{1j} = 0$. Assim, atribuímos uma priori de mistura com ponto de massa no zero, apenas para ϕ_{2j} . A distribuição a priori para ϕ_{1j} é como antes, dada por

$$\phi_{1j} = \frac{(l_{2j} - l_{1j}) \exp(\Psi_{1j})}{1 + \exp(\Psi_{1j})} + l_{1j},$$

onde $\Psi_{1j} \sim GP(0, \Sigma_{1j})$, $\phi_{1j} = (\phi_{1j}(s_1), \dots, \phi_{1j}(s_n))'$ tal que $l_{1j}(s) < \phi_{2j}(s) < l_{2j}(s)$. Como antes, $l_{1j}(s) = -2(-\phi_{2j}(s))^{1/2}$ e $l_{2j}(s) = 2 \cos(4\pi/t)(-\phi_{2j}(s))^{1/2}$. A distribuição a priori de mistura para ϕ_{2j} é definida como

$$p(\phi_{2j} | \pi_C) = \pi_C f_2(\phi_{2j}) + (1 - \pi_C) \delta_0,$$

onde $\phi_{2j} = \frac{\exp(\Psi_{2j})}{1 + \exp(\Psi_{2j})} - 1$ e $\Psi_{2j} \sim GP(0, \Sigma_{2j})$, $\phi_{2j} = (\phi_{2j}(s_1), \dots, \phi_{2j}(s_n))'$, tal que $-1 < \phi_{2j}(s) < 0$. Então,

$$f_2(\phi_{2j}) = (2\pi)^{-n/2} |\Sigma_{2j}|^{-1/2} \exp \left\{ -\frac{1}{2} \left[\log \frac{-1 - \phi_{2j}}{\phi_{2j}} \right]' \Sigma_{2j}^{-1} \left[\log \frac{-1 - \phi_{2j}}{\phi_{2j}} \right] \right\} \prod_{i=1}^n \frac{1}{\phi_{2j}(1 + \phi_{2j})}.$$

Seja $I(\phi_{2j})$ uma variável indicadora com

$$I(\phi_{2j}) = \begin{cases} 1, & \text{se } \phi_{2j} \neq 0, \\ 0, & \text{se } \phi_{2j} = 0. \end{cases}$$

Assim, $I(\phi_{2j}) \sim Bernoulli(\pi_C)$ tal que $P(I(\phi_{2j}) = 1) = \pi_C$, onde π_C é a probabilidade a priori do par de raiz complexa estar presente. Assim, o número de raízes complexas pode ser menor que a ordem máxima especificada, C_+ .

Amostrando as raízes complexas

Condicional as raízes complexas, seja $z_{jt} = \prod_{i \in A_j}^p (1 - \alpha_i B) y_t^*$ a série temporal sem o j -ésimo par de raízes complexas e conjugadas, onde A_j representa o conjunto de todas as outras raízes. Assim, z_{jt} é um AR(2) multivariado com parâmetros $\phi_{1j} = 2r_j \cos \omega_j$, $\phi_{2j} = -r_j^2$ e variância Σ . Logo,

$$z_{jt} = \phi_{1j} z_{jt-1} + \phi_{2j} z_{jt-2} + v_t.$$

A condicional completa para ϕ_{2j} é dada por

$$\begin{aligned} p(\phi_{2j}|\cdot) &\propto \prod_{t=p+1}^T p(z_{jt}^*|\phi_{2j}z_{jt-2}, \Sigma) \times p(\phi_{2j}|\pi_C) \\ &= \pi_C \times \prod_{t=p+1}^T f_N(z_{jt}^*|\phi_{2j}z_{jt-2}, \Sigma) \times f_2(\phi_{2j}) + (1 - \pi_C) \times \prod_{t=p+1}^T f_N(z_{jt}^*|0, \Sigma), \end{aligned}$$

onde $z_{jt}^* = z_{jt} - \phi_{1j}z_{jt-1}$. Inicialmente devemos calcular a condicional completa de $p(I(\phi_{2j}) = 1)$, isto é,

$$p(I(\phi_{2j}) = 1|\cdot) = \frac{\pi_C \times \prod_{t=p+1}^T f_N(z_{jt}^*|\phi_{2j}z_{jt-2}, \Sigma) \times f_2(\phi_{2j})}{\pi_C \times \prod_{t=p+1}^T f_N(z_{jt}^*|\phi_{2j}z_{jt-2}, \Sigma) \times f_2(\phi_{2j}) + (1 - \pi_C) \times \prod_{t=p+1}^T f_N(z_{jt}^*|0, \Sigma)}. \quad (4.11)$$

Em seguida, geramos $u \sim U(0, 1)$. Se $u < p(I(\phi_{2j}) = 1|\cdot)$, fazemos $I(\phi_{2j}) = 1$ e amostramos ϕ_{2j} da seguinte distribuição

$$p(\phi_{2j}|\cdot) \propto \prod_{t=p+1}^T p(z_{jt}^*|\phi_{2j}z_{jt-2}, \Sigma) \times f_2(\phi_{2j}).$$

Caso contrário, $I(\phi_{2j}) = 0$ e fazemos $\phi_{2j} = 0$, $\forall s$.

Dividindo o numerador e o denominador da equação (4.11), pelo numerador, simplificamos a expressão acima e obtemos:

$$p(I(\phi_{2j}) = 1|\cdot) = \frac{1}{\frac{1-\pi_C}{\pi_C}L + 1},$$

onde $L = \frac{\prod_{t=p+1}^T f_N(z_{jt}^*|0, \Sigma)}{\prod_{t=p+1}^T f_N(z_{jt}^*|\phi_{2j}z_{jt-2}, \Sigma) \times f_2(\phi_{2j})}$. Se $\phi_{2j} = 0$, temos um problema numérico, já que $f_2(\phi_{2j})$ depende do cálculo do $\log \frac{-1-\phi_{2j}}{\phi_{2j}}$. Para tentar evitar este problema, adicionamos uma constante no numerador e denominador da transformação. Os resultados obtidos não foram satisfatórios, já que estávamos deslocando o espaço paramétrico de ϕ_{2j} .

4.6.2 Mistura baseada em duas distribuições contínuas

Raízes Reais

Vamos considerar agora uma priori de mistura de duas distribuições contínuas,

$$p(a_j|\pi_R) = \pi_R f(a_j) + (1 - \pi_R)g(a_j),$$

onde $a_j = \frac{2 \exp(\Psi_j)}{1 + \exp(\Psi_j)} - 1$, $f(a_j)$ tal que $\Psi_j \sim GP(0, \Sigma_j)$ e $g(a_j)$ tal que $\Psi_j \sim GP(0, \Sigma_j^*)$. Neste caso, a_j pode vir de uma distribuição cuja função de densidade de probabilidade é $f(a_j)$ com probabilidade π_R

ou de uma distribuição cuja função de densidade de probabilidade é $g(a_j)$ com probabilidade $1 - \pi_R$. A idéia aqui, é escolher Σ_j^* de maneira que obtenhamos uma distribuição mais concentrada em torno do zero para $g(a_j)$ do que para $f(a_j)$, mas agora a_j nunca será exatamente igual a zero. Desta forma, o modelo considerado é uma ponderação de diversos modelos cujas ordens são sempre a ordem máxima especificada, $p_+ = R_+ + 2C_+$, mas com uma ordem $k < p_+$, em que as raízes recíprocas fornecem coeficientes Φ_1, \dots, Φ_k significativamente diferentes de zero.

Seja $I(a_j)$ uma variável indicadora tal que

$$I(a_j) = \begin{cases} 1, & \text{se } a_j \sim f(a_j), \\ 0, & \text{se } a_j \sim g(a_j). \end{cases}$$

Assim, $I(a_j) \sim \text{Bernoulli}(\pi_R)$ tal que $P(I(a_j) = 1) = \pi_R$. A variável indicadora $I(a_j)$ indica se a_j vêm de uma densidade $f(a_j)$ ou $g(a_j)$.

Raízes Complexas

Como comentado anteriormente, quando $\phi_{2j} = 0$, resulta que $\phi_{1j} = 0$. Assim, atribuímos uma priori de mistura, apenas para ϕ_{2j} . A priori para ϕ_{1j} é dada por

$$\phi_{1j} = \frac{(l_{2j} - l_{1j}) \exp(\Psi_{1j})}{1 + \exp(\Psi_{1j})} + l_{1j},$$

onde $\Psi_{1j} \sim GP(0, \Sigma_{1j})$. Assim,

$$f_1(\phi_{1j}) = (2\pi)^{-n/2} |\Sigma_{1j}|^{-1/2} \exp \left\{ -\frac{1}{2} \left[\log \frac{l_{1j} - \phi_{1j}}{\phi_{1j} - l_{2j}} \right]' \Sigma_{1j}^{-1} \left[\log \frac{l_{1j} - \phi_{1j}}{\phi_{1j} - l_{2j}} \right] \right\} \prod_{i=1}^n \frac{l_{2j} - l_{1j}}{(l_{1j} - \phi_{1j})(\phi_{1j} - l_{2j})}.$$

Para ϕ_{2j} , consideramos uma priori de mistura de duas distribuições contínuas, isto é,

$$p(\phi_{2j} | \pi_R) = \pi_R f_2(\phi_{2j}) + (1 - \pi_R) g_2(\phi_{2j}),$$

onde $\phi_{2j} = \frac{\exp(\Psi_{2j})}{1 + \exp(\Psi_{2j})} - 1$, $f_2(\phi_{2j})$ é tal que $\Psi_{2j} \sim GP(0, \Sigma_{2j})$ e $g_2(\phi_{2j})$ tal que $\Psi_{2j} \sim GP(0, \Sigma_{2j}^*)$. Neste caso, ϕ_{2j} pode vir de uma distribuição cuja função de densidade de probabilidade é $f_2(\phi_{2j})$ com probabilidade π_C ou vir de uma distribuição cuja função de densidade de probabilidade é $g_2(\phi_{2j})$ com probabilidade $1 - \pi_C$. A escolha de Σ_{2j}^* é feita seguindo a mesma idéia discutida para o caso das raízes reais.

Como antes, seja $I(\phi_{2j})$ uma variável indicadora com

$$I(\phi_{2j}) = \begin{cases} 1, & \text{se } \phi_{2j} \sim f_2(\phi_{2j}), \\ 0, & \text{se } \phi_{2j} \sim g_2(\phi_{2j}). \end{cases}$$

Assim, $I(\phi_{2j}) \sim \text{Bernoulli}(\pi_C)$ tal que $P(I(\phi_{2j}) = 1) = \pi_C$, onde π_C é a probabilidade a priori do par de raiz complexa vir de $f_2(\phi_{2j})$.

Procedimento de Inferência

Agora, os parâmetros do modelo são: $\Theta = (\theta_t; t = 1, \dots, T, \sigma^2, \lambda, \kappa, \pi_R, \pi_C, a_j; j = 1, \dots, R_+, \phi_{1j}, \phi_{2j}; j = 1, \dots, C_+)$ e pelo teorema de Bayes, a distribuição a posteriori é proporcional a,

$$p(\Theta|y) \propto \left(\prod_{t=p+1}^T p(y_t|\Theta) \right) p(\theta_1, \dots, \theta_p) p(\sigma^2) p(\lambda) p(\kappa) \left(\prod_{j=1}^{R_+} p(a_j|\pi_R) \right) p(\pi_R) \left(\prod_{j=1}^{C_+} p(\phi_{1j}) p(\phi_{2j}|\pi_C) \right) p(\pi_C).$$

A distribuição a posteriori dos parâmetros do modelo não tem forma analítica fechada e, precisamos de métodos de MCMC para obtenção das amostras da distribuição a posteriori dos parâmetros do modelo. Mostraremos a seguir, o esquema de amostragem das raízes reais e complexas, assim como dos parâmetros π_R e π_C . Para os demais parâmetros do modelo o esquema de amostragem utilizado será o mesmo apresentado quando consideramos a ordem do SVAR(p) conhecida. Neste caso, $p = R + 2C$, onde R e C são respectivamente, o número de raízes reais e complexas estimadas pelo modelo.

Amostrando as raízes reais

Condicional as raízes reais, $z_{jt} = \prod_{\substack{i=1 \\ i \neq j}}^p (1 - \alpha_i B) y_t^*$ é um AR(1) multivariado com parâmetro a_j , $-1 < a_j < 1$, e variância Σ .

$$z_{jt} = a_j z_{jt-1} + v_t.$$

A condicional completa para a_j é dada por

$$\begin{aligned} p(a_j|\cdot) &\propto \prod_{t=p+1}^T p(z_{jt}|a_j z_{jt-1}, \Sigma) \times p(a_j|\pi_R) \\ &= \pi_R \times \prod_{t=p+1}^T f_N(z_{jt}|a_j z_{jt-1}, \Sigma) \times f(a_j) + (1 - \pi_R) \times \prod_{t=p+1}^T f_N(z_{jt}|a_j z_{jt-1}, \Sigma) \times g(a_j). \end{aligned}$$

A condicional completa de $p(I(a_j) = 1)$ é dada por

$$\frac{\pi_R \times \prod_{t=p+1}^T f_N(z_{jt}|a_j z_{jt-1}, \Sigma) \times f(a_j)}{\pi_R \times \prod_{t=p+1}^T f_N(z_{jt}|a_j z_{jt-1}, \Sigma) \times f(a_j) + (1 - \pi_R) \times \prod_{t=p+1}^T f_N(z_{jt}|a_j z_{jt-1}, \Sigma) \times g(a_j)}.$$

Dividindo o numerador e o denominador pelo numerador, simplificamos a expressão acima e obtemos:

$$p(I(a_j) = 1|\cdot) = \frac{1}{\frac{1-\pi_R}{\pi_R} \frac{g(a_j)}{f(a_j)} + 1},$$

onde

$$f(a_j) = (2\pi)^{-n/2} |\Sigma_j|^{-1/2} \exp \left\{ -\frac{1}{2} \left[\log \frac{1+a_j}{1-a_j} \right]' \Sigma_j^{-1} \left[\log \frac{1+a_j}{1-a_j} \right] \right\} \prod_{i=1}^n \frac{2}{1-a_j(s_i)^2},$$

$$g(a_j) = (2\pi)^{-n/2} |\Sigma_j^*|^{-1/2} \exp \left\{ -\frac{1}{2} \left[\log \frac{1+a_j}{1-a_j} \right]' \Sigma_j^{*-1} \left[\log \frac{1+a_j}{1-a_j} \right] \right\} \prod_{i=1}^n \frac{2}{1-a_j(s_i)^2}.$$

Em seguida, geramos $u \sim U(0, 1)$. Se $u < p(I(a_j) = 1)$, fazemos $I(a_j) = 1$ e amostramos a_j da seguinte distribuição

$$p(a_j|\cdot) \propto \prod_{t=p+1}^T p(z_{jt}|a_j z_{jt-1}, \Sigma) \times f(a_j),$$

caso contrário, $I(a_j) = 0$ e amostramos a_j de

$$p(a_j|\cdot) \propto \prod_{t=p+1}^T p(z_{jt}|a_j z_{jt-1}, \Sigma) \times g(a_j).$$

Como pode ser visto, $p(a_j|\cdot)$ não é uma distribuição conhecida e, portanto utilizamos o algoritmo de Metropolis-Hastings, para obtenção das amostras da distribuição a posteriori de interesse. Utilizamos como proposta a seguinte distribuição $a_j^p = \frac{2 \exp(\Psi_j^p)}{1 + \exp(\Psi_j^p)} - 1$ onde $\Psi_j^p \sim GP(\Psi_j^c, U)$ e Ψ_j^c é o valor corrente de Ψ_j . A razão de Metropolis-Hastings após simplificações é dada por

$$\frac{p(a_j^p|\cdot)q(a_j^c|a_j^p)}{p(a_j^c|\cdot)q(a_j^p|a_j^c)} = \frac{\prod_{t=p+1}^T f_N(z_{jt}|a_j^p z_{jt-1}, \Sigma) \times f(a_j^p)}{\prod_{t=p+1}^T f_N(z_{jt}|a_j^c z_{jt-1}, \Sigma) \times f(a_j^c)}, \quad \text{se } I(a_j) = 1,$$

ou,

$$\frac{p(a_j^p|\cdot)q(a_j^c|a_j^p)}{p(a_j^c|\cdot)q(a_j^p|a_j^c)} = \frac{\prod_{t=p+1}^T f_N(z_{jt}|a_j^p z_{jt-1}, \Sigma) \times g(a_j^p)}{\prod_{t=p+1}^T f_N(z_{jt}|a_j^c z_{jt-1}, \Sigma) \times g(a_j^c)}, \quad \text{se } I(a_j) = 0.$$

Como podemos observar o cálculo da probabilidade indicadora depende de $f(a_j)$ e $g(a_j)$ e, conseqüentemente, dos hiperparâmetros de Σ_j e Σ_j^* . Inicialmente, escolhemos Σ_j e Σ_j^* , de maneira a obter

uma função mais suave, mais correlacionada e mais concentrada em torno do zero para $g(a_j)$, isto é, $\kappa_j < \kappa_j^*$, $\lambda_j > \lambda_j^*$ e $\sigma_j^2 > \sigma_j^{2*}$. Após algumas comparações, observamos que as covariâncias não podem diferir muito, já que o cálculo da probabilidade indicadora envolve o cálculo do determinante destas matrizes, tornando o cálculo da probabilidade indicadora bastante sensível a esta escolha. Concluimos então que a função de correlação deve ser a mesma, e escolhemos as variâncias σ_j^2 e σ_j^{2*} como função de $p(I(a_j) = 1)$ e π_R . Vamos supor por exemplo que $a_j(s) = 0, \forall s$. Assim, $\frac{g(a_j)}{f(a_j)} = \frac{|\Sigma_j^*|^{-1/2}}{|\Sigma_j|^{-1/2}}$. Considerando a mesma função de correlação e fazendo $\lambda_j = \lambda_j^*$ e $\kappa_j = \kappa_j^*$, obtemos $\frac{g(a_j)}{f(a_j)} = \left(\frac{\sigma_j}{\sigma_j^*}\right)^n$, já que $|\Sigma_j|^{-1/2} = (\sigma_j^2)^{-n/2}|V_j|^{-1/2}$. Finalmente,

$$\sigma_j = \sigma_j^* \sqrt[n]{\frac{1 - p(I(a_j) = 1)}{p(I(a_j) = 1)} \frac{\pi_R}{1 - \pi_R}}. \quad (4.12)$$

Fixando $p(I(a_j) = 1)$, π_R e σ_j^* , obtemos σ_j . Dessa forma, a equação (4.12) pode ser usada para fixar as densidades envolvidas na mistura.

Amostrando as raízes complexas

Condicional as raízes complexas, seja $z_{jt} = \prod_{i \in A_j} (1 - \alpha_i B) y_t^*$ a série temporal sem o j -ésimo par de raízes complexas e conjugadas, onde A_j representa o conjunto de todas as outras raízes. Assim, z_{jt} é um AR(2) multivariado com parâmetros $\phi_{1j} = 2r_j \cos \omega_j$, $\phi_{2j} = -r_j^2$ e variância Σ . Logo,

$$z_{jt} = \phi_{1j} z_{jt-1} + \phi_{2j} z_{jt-2} + v_t.$$

A condicional completa para ϕ_{1j} é dada por

$$p(\phi_{1j} | \cdot) \propto p(z_{jt}^* | \phi_{1j} z_{jt-1}, \Sigma) p(\phi_{1j}),$$

onde $z_{jt}^* = z_{jt} - \phi_{2j} z_{jt-2}$. A razão de Metropolis-Hastings é dada por

$$\frac{p(\phi_{1j}^p | \cdot) q(\phi_{1j}^c | \phi_{1j}^p)}{p(\phi_{1j}^c | \cdot) q(\phi_{1j}^p | \phi_{1j}^c)} = \frac{\prod_{t=p+1}^T f_N(z_{jt}^* | \phi_{1j}^p z_{jt-1}, \Sigma) \times p(\phi_{1j}^p)}{\prod_{t=p+1}^T f_N(z_{jt}^* | \phi_{1j}^c z_{jt-1}, \Sigma) \times p(\phi_{1j}^c)}.$$

Já a condicional completa para ϕ_{2j} é dada por

$$\begin{aligned} p(\phi_{2j} | \cdot) &\propto \prod_{t=p+1}^T p(z_{jt}^* | \phi_{2j} z_{jt-2}, \Sigma) \times p(\phi_{2j} | \pi_C) \\ &= \pi_C \times \prod_{t=p+1}^T f_N(z_{jt}^* | \phi_{2j} z_{jt-2}, \Sigma) \times f_2(\phi_{2j}) + (1 - \pi_C) \times \prod_{t=p+1}^T f_N(z_{jt}^* | \phi_{2j} z_{jt-2}, \Sigma) \times g_2(\phi_{2j}), \end{aligned}$$

onde $z_{jt}^* = z_{jt} - \phi_{1j}z_{jt-1}$. Inicialmente devemos calcular a condicional completa da indicadora de ϕ_{2j} ser 1, isto é, $p(I(\phi_{2j}) = 1)$,

$$\frac{\pi_C \times \prod_{t=p+1}^T f_N(z_{jt}^* | \phi_{2j}z_{jt-2}, \Sigma) \times f_2(\phi_{2j})}{\pi_C \times \prod_{t=p+1}^T f_N(z_{jt}^* | \phi_{2j}z_{jt-2}, \Sigma) \times f_2(\phi_{2j}) + (1 - \pi_R) \times \prod_{t=p+1}^T f_N(z_{jt}^* | \phi_{2j}z_{jt-2}, \Sigma) \times g_2(\phi_{2j})}.$$

Dividindo o numerador e o denominador pelo numerador, simplificamos a expressão acima e obtemos:

$$p(I(\phi_{2j}) = 1 | \cdot) = \frac{1}{\frac{1 - \pi_C}{\pi_C} \frac{g_2(\phi_{2j})}{f_2(\phi_{2j})} + 1}.$$

Em seguida, geramos $u \sim U(0, 1)$. Se $u < p(I(\phi_{2j}) = 1)$, fazemos $I(\phi_{2j}) = 1$ e amostramos ϕ_{2j} da seguinte distribuição

$$p(\phi_{2j} | \cdot) \propto \prod_{t=p+1}^T p(z_{jt}^* | \phi_{2j}z_{jt-2}, \Sigma) \times f_2(\phi_{2j}),$$

caso contrário, $I(\phi_{2j}) = 0$ e amostramos ϕ_{2j} de

$$p(\phi_{2j} | \cdot) \propto \prod_{t=p+1}^T p(z_{jt}^* | \phi_{2j}z_{jt-2}, \Sigma) \times g_2(\phi_{2j}).$$

De maneira similar ao caso anterior, utilizamos o algoritmo de Metropolis-Hastings para obter as amostras da distribuição a posteriori do parâmetro ϕ_{2j} . Utilizamos como distribuição proposta $\phi_{2j}^p = \frac{\exp(\Psi_{2j}^p)}{1 + \exp(\Psi_{2j}^p)} - 1$ onde $\Psi_{2j}^p \sim GP(\Psi_{2j}^c, U)$ e Ψ_{2j}^c é o valor corrente de Ψ_{2j} . Após simplificações, a razão de Metropolis-Hastings para os dois casos, $I(\phi_{2j}) = 1$ e $I(\phi_{2j}) = 0$, respectivamente, são dadas por

$$\frac{p(\phi_{2j}^p | \cdot) q(\phi_{2j}^c | \phi_{2j}^p)}{p(\phi_{2j}^c | \cdot) q(\phi_{2j}^p | \phi_{2j}^c)} = \frac{\prod_{t=p+1}^T f_N(z_{jt}^* | \phi_{2j}^p z_{jt-2}, \Sigma) \times f_2(\phi_{2j}^p)}{\prod_{t=p+1}^T f_N(z_{jt}^* | \phi_{2j}^c z_{jt-2}, \Sigma) \times f_2(\phi_{2j}^c)},$$

e,

$$\frac{p(\phi_{2j}^p | \cdot) q(\phi_{2j}^c | \phi_{2j}^p)}{p(\phi_{2j}^c | \cdot) q(\phi_{2j}^p | \phi_{2j}^c)} = \frac{\prod_{t=p+1}^T f_N(z_{jt}^* | \phi_{2j}^p z_{jt-2}, \Sigma) \times g_2(\phi_{2j}^p)}{\prod_{t=p+1}^T f_N(z_{jt}^* | \phi_{2j}^c z_{jt-2}, \Sigma) \times g_2(\phi_{2j}^c)}.$$

Amostrando π_R

A distribuição conjunta de $(a_j, I(a_j))$ condicional a π_R é igual a

$$p(a_j, I(a_j) | \pi_R) = p(I(a_j) | \pi_R) p(a_j | \pi_R) = (\pi_R f(a_j))^{I(a_j)} [(1 - \pi_R) g(a_j)]^{1 - I(a_j)}.$$

Assumindo uma distribuição a priori beta para π_R , $\pi_R \sim \text{Beta}(a_{\pi_R}, b_{\pi_R})$ e $\sum_{j=1}^{R_+} I(a_j) = R$. A distribuição condicional completa de π_R é dada por

$$\pi_R | \cdot \sim \text{Beta}(R + a_{\pi_R}, R_+ - R + b_{\pi_R})$$

Amostrando π_C

A distribuição conjunta de $(\phi_{2j}, I(\phi_{2j}))$ condicional a π_C é dada por

$$p(\phi_{2j}, I(\phi_{2j}) | \pi) = p(I(\phi_{2j}) | \pi_C) p(\phi_{2j} | \pi_C) = (\pi_C f_2(\phi_{2j}))^{I(\phi_{2j})} [(1 - \pi_C) g_2(\phi_{2j})]^{1 - I(\phi_{2j})}.$$

Assumindo uma distribuição a priori beta para π_C , $\pi_C \sim \text{Beta}(a_{\pi_C}, b_{\pi_C})$ e $\sum_{j=1}^{C_+} I(\phi_{2j}) = C$. A distribuição condicional completa de π_C é dada por

$$\pi_C | \cdot \sim \text{Beta}(C + a_{\pi_C}, C_+ - C + b_{\pi_C}).$$

Neste capítulo generalizamos o modelo proposto no Capítulo 3, considerando que os coeficientes da componente autoregressiva variam suavemente ao longo do espaço. Em seguida, consideramos que a ordem do modelo é desconhecida e fixada num valor máximo. A distribuição a priori inicialmente proposta é uma extensão multivariada da priori proposta por Huerta e West (1999). Observamos que no caso multivariado, essa priori apresenta alguns problemas. Quando as raízes recíprocas são complexas, observamos um problema numérico. Ainda para o caso onde as raízes são reais, observamos uma grande sensibilidade no cálculo da probabilidade da variável indicadora, possivelmente isto deve-se ao fato de que é muito difícil fazer um vetor aleatório exatamente igual a zero, ou seja, dizer que as raízes para todas as estações são exatamente iguais a zero. Assim, propomos uma mistura de duas distribuições contínuas, considerando uma delas bastante concentrada em torno do zero, o suficiente para que possamos considerar o vetor igual a zero. Além disso, podemos generalizar a priori para o caso em que as raízes são complexas.

No capítulo seguinte ajustamos o modelo proposto neste capítulo, em dois conjuntos de dados simulados e em um conjunto de dados reais, utilizando o esquema de inferência apresentado aqui.

Capítulo 5

Análise de Dados

Neste capítulo ajustamos o modelo baseado em processos autoregressivos com coeficientes variando espacialmente, proposto no capítulo anterior. Na Seção 5.1 ajustamos o modelo proposto em (4.2) a dois conjunto de dados, um conjunto de dados artificiais gerados a partir do modelo e um conjunto de dados reais de material particulado, PM_{10} , observados em 8 estações monitoradoras localizadas em Santiago do Chile, Chile. Na Seção 5.2 ajustamos o modelo proposto em (4.2) a um conjunto de dados simulados usando as duas propostas de prioris de mistura apresentadas no capítulo anterior.

5.1 Ajustando o modelo proposto considerando a ordem conhecida

5.1.1 Dados Artificiais - SVAR(2)

Geração dos Dados Artificiais

Simulamos dados do modelo como especificado em (4.2), com $p = 2$ e $C = 1$, isto é, um par de raiz complexa e conjugada, para $n = 30$ localizações e $T = 50$ instantes de tempo. Além disso, assumimos a presença de duas covariáveis, X_{1t} e X_{2t} geradas usando uma distribuição normal multivariada com vetor de médias igual a uma constante e matriz de covariância espacialmente correlacionada. Reescrevemos o modelo da seguinte maneira

$$Y_t = F_t' \theta_t + \Phi_1(Y_{t-1} - F_{t-1}' \theta_{t-1}) + \Phi_2(Y_{t-2} - F_{t-2}' \theta_{t-2}) + v_t, \quad v_t \sim N(0, \Sigma)$$

$$\beta_t = \beta_{t-1} + w_{t1}, \quad w_{t1} \sim N(0, \tau_\beta^2), \quad \gamma_t = \gamma_{t-1} + w_{t2}, \quad w_{t2} \sim N(0, \tau_\gamma^2)$$

onde $F_t' = (X_{1t}, X_{2t})$, $\theta_t = (\beta_t, \gamma_t)'$ e $\Sigma = \sigma^2 \rho(\|\cdot\|, \lambda)$. Utilizamos a função de correlação exponencial com parâmetro de decaimento λ . Os coeficientes da componente autoregressiva são $\Phi_1 = \phi_{11} = \text{diag}(\phi_{11}(s_1), \dots, \phi_{11}(s_{30}))$ com $\phi_{11}(s_i) = 2r_1(s_i) \cos \omega_1(s_i)$ e $\Phi_2 = \phi_{21} = \text{diag}(\phi_{21}(s_1), \dots, \phi_{21}(s_{30}))$ com $\phi_{21}(s_i) = -r_1(s_i)^2$, $i = 1, \dots, 30$. Geramos Ψ_{11} e Ψ_{21} através de um processo gaussiano com vetor de média igual a zero e matriz de covariância Σ_{11} e Σ_{21} , respectivamente. Utilizamos uma função de correlação Matérn com parâmetro $\lambda_{11} = \lambda_{21} = 0,01$, calculado baseado na idéia de distância máxima e $\kappa_{11} = \kappa_{21} = 2$ para obter um processo suave, e variâncias $\sigma_{11}^2 = \sigma_{21}^2 = 1,5$. Em seguida, obtemos ϕ_{11} e ϕ_{21} através das transformações definidas em (4.5). Os valores iniciais dos coeficientes das covariáveis foram $\beta_0 = 5$ e $\gamma_0 = -3$ com variâncias fixadas em $\tau_\beta^2 = \tau_\gamma^2 = 0,1$. A Figura 5.1 apresenta os dados simulados com estes parâmetros para as 30 estações, ao longo dos 50 instantes de tempo, bem como suas respectivas localizações.

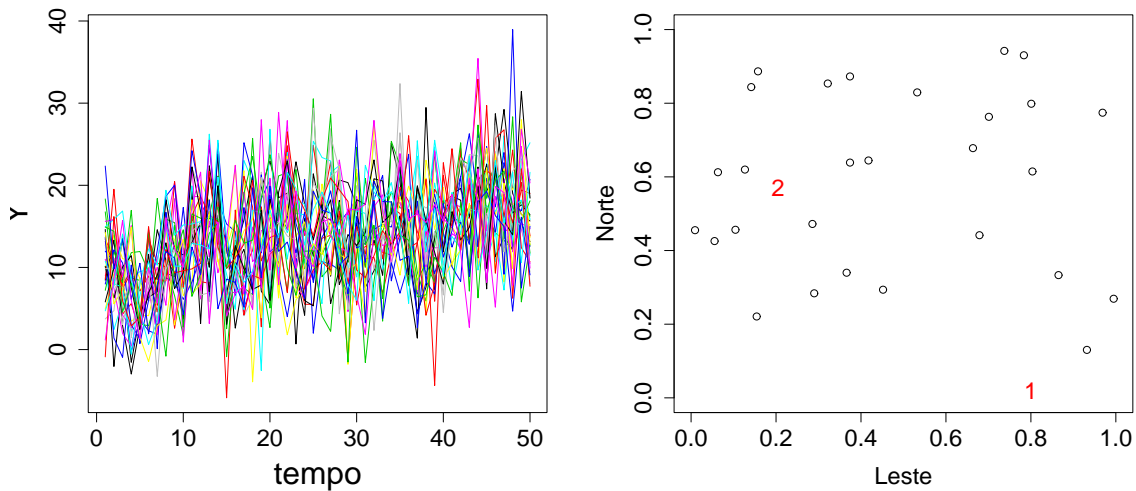


Figura 5.1: Dados artificialmente gerados ao longo de 30 localizações e 50 instantes de tempo segundo o modelo SVAR(2), e as respectivas localizações das estações monitoradoras.

Ajuste do Modelo

Ajustamos o modelo utilizando as seguintes priors: $\sigma^2 \sim IG(1, 1)$, $\lambda \sim Ga(1, \frac{\max(dist)}{6})$, $\tau_\beta^2 \sim IG(0, 1; 0, 1)$, $\tau_\gamma^2 \sim IG(0, 1; 0, 1)$, $(\beta_0, \gamma_0)' \sim N_2(\mathbf{0}, 100\mathbf{I})$, $\Psi_{11} \sim GP(0, \Sigma_{11})$ e $\Psi_{21} \sim GP(0, \Sigma_{21})$ com Σ_{11} e Σ_{21} fixadas nos mesmos parâmetros utilizados para gerar os dados, onde $\max(dist)$ denota a distância máxima. A Tabela 5.1 apresenta a média e o intervalo de 95% de credibilidade a posteriori para os parâmetros fixos no tempo, λ , σ^2 , τ_β^2 e τ_γ^2 . A média a posteriori encontra-se bem próxima do valor verdadeiro, exceto para o parâmetro λ . A Figura 5.2 apresenta a média e o intervalo de credibilidade a posteriori para os parâmetros da estrutura de média que variam no tempo. Observamos mais uma vez que estes parâmetros são bem estimados pelo nosso modelo. A Figura 5.3 mostra o boxplot da distribuição a posteriori para $\Phi_1(s)$ e $\Phi_2(s)$ para cada estação, bem como o valor verdadeiro indicado com o símbolo "x". Observamos que o modelo consegue estimar relativamente bem estes parâmetros, pois para a maioria das estações a mediana a posteriori encontra-se próxima ao valor verdadeiro.

	Verdadeiro	Média	IC(95%)
λ	5,55	7,648	(6,410; 9,166)
σ^2	1	1,007	(0,901; 1,124)
τ_β^2	0,10	0,111	(0,068; 0,174)
τ_γ^2	0,10	0,112	(0,065; 0,180)

Tabela 5.1: Média e Intervalo de 95% de Credibilidade a posteriori para os parâmetros fixos no tempo, segundo o modelo SVAR(2) para os dados artificiais.

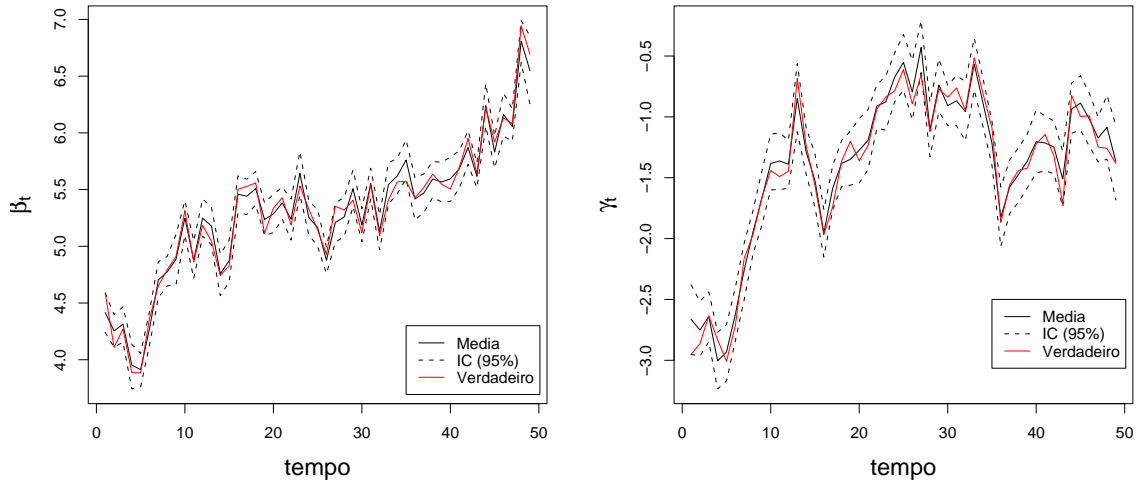


Figura 5.2: Média e Intervalo de 95% de Credibilidade a posteriori para os coeficientes β_t e γ_t ao longo do tempo, segundo o modelo SVAR(2) para os dados artificiais.

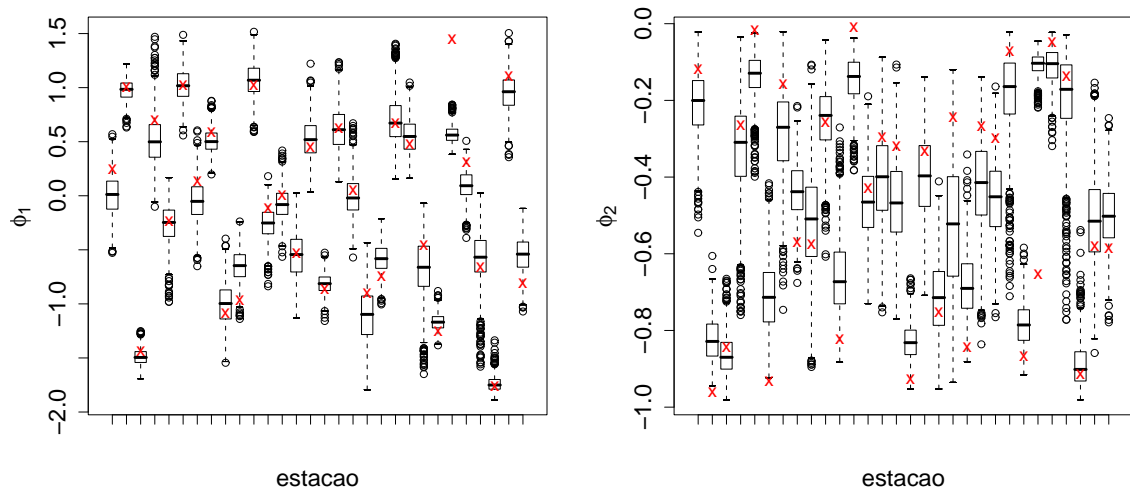


Figura 5.3: Boxplot da distribuição a posteriori para os parâmetros $\Phi_1(s)$ (direita) e $\Phi_2(s)$ (esquerda) para cada estação, segundo o modelo SVAR(2) para os dados artificiais.

Seguindo o mesmo procedimento utilizado no Capítulo 3, ajustamos o modelo retirando 2 estações, e as amostras da distribuição a posteriori foram encontradas, e utilizadas para fazer interpolação espacial para as 2 estações retiradas. A Figura 5.4 mostra o valor verdadeiro, bem como a média e o intervalo de 95% de credibilidade para previsão ao longo do tempo. O modelo consegue captar bem o comportamento da série pois o intervalo de credibilidade é bem concentrado e contém o valor verdadeiro.

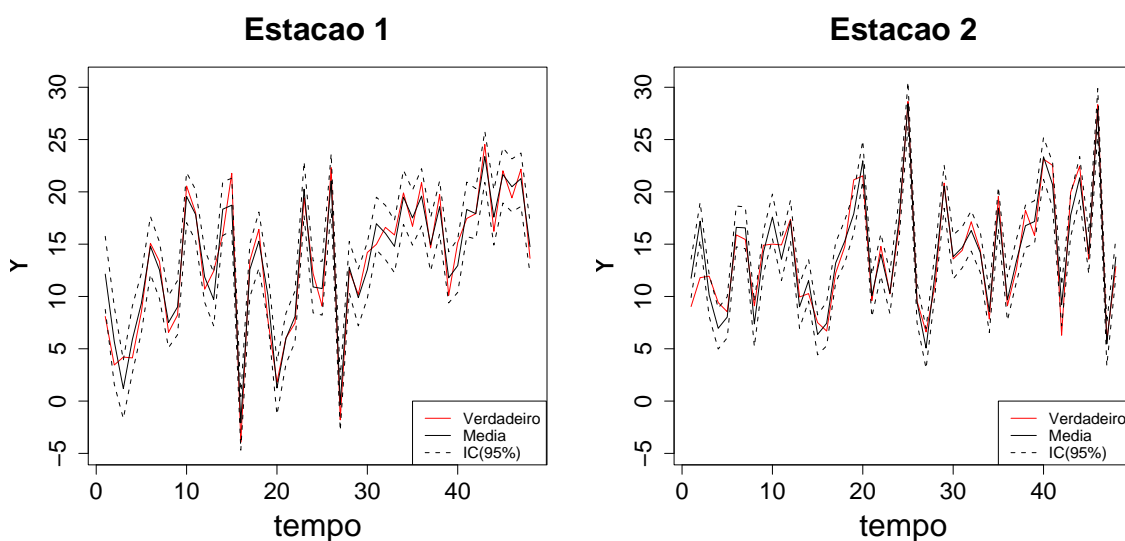


Figura 5.4: Previsão para 2 localizações ausentes no processo de inferência, segundo o modelo SVAR(2) para os dados artificiais.

Com o objetivo de verificar a capacidade preditiva do modelo, uma grade regular de 400 pontos foi construída, e foi feita uma interpolação espacial para esta grade. Em seguida, calculamos a interpolação espacial usando os valores verdadeiros dos parâmetros. As superfícies utilizando os valores verdadeiros para 8 tempos são apresentadas na Figura 5.5. As Figuras 5.6 e 5.7 apresentam, respectivamente, as superfícies estimadas pelo modelo usando a mediana e o intervalo interquartilico a posteriori. Podemos observar que as superfícies são bem similares entre si.

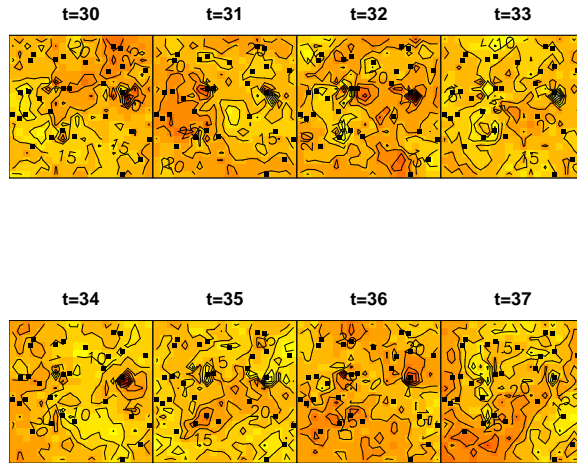


Figura 5.5: Interpolação espacial usando os valores verdadeiros dos parâmetros em 8 tempos consecutivos, segundo modelo SVAR(2) para os dados artificiais.

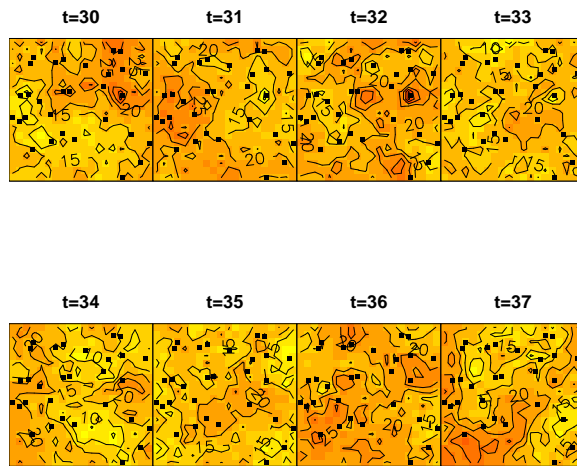


Figura 5.6: Mediana a posteriori da distribuição preditiva em 8 tempos consecutivos, segundo o modelo SVAR(2) para os dados artificiais.

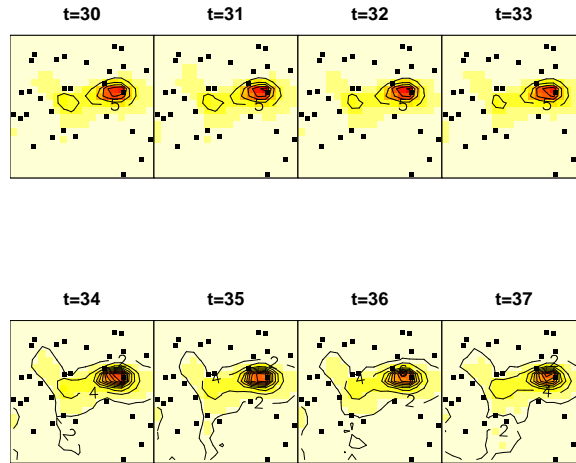


Figura 5.7: Intervalo interquartílico da distribuição preditiva em 8 tempos consecutivos, segundo o modelo SVAR(2) para os dados artificiais.

Para esta mesma grade de pontos, construímos também as superfícies para os parâmetros $\Phi_1(s)$, $\Phi_2(s)$, $r_1(s)$ e $\omega_1(s)$. As Figuras 5.8 e 5.9 apresentam as superfícies estimadas (esquerda) e usando os valores verdadeiros dos parâmetros (direita). Observamos que as superfícies estimadas pelo modelo são ligeiramente mais suaves do que as superfícies verdadeiras, mas elas são razoavelmente semelhantes entre si.

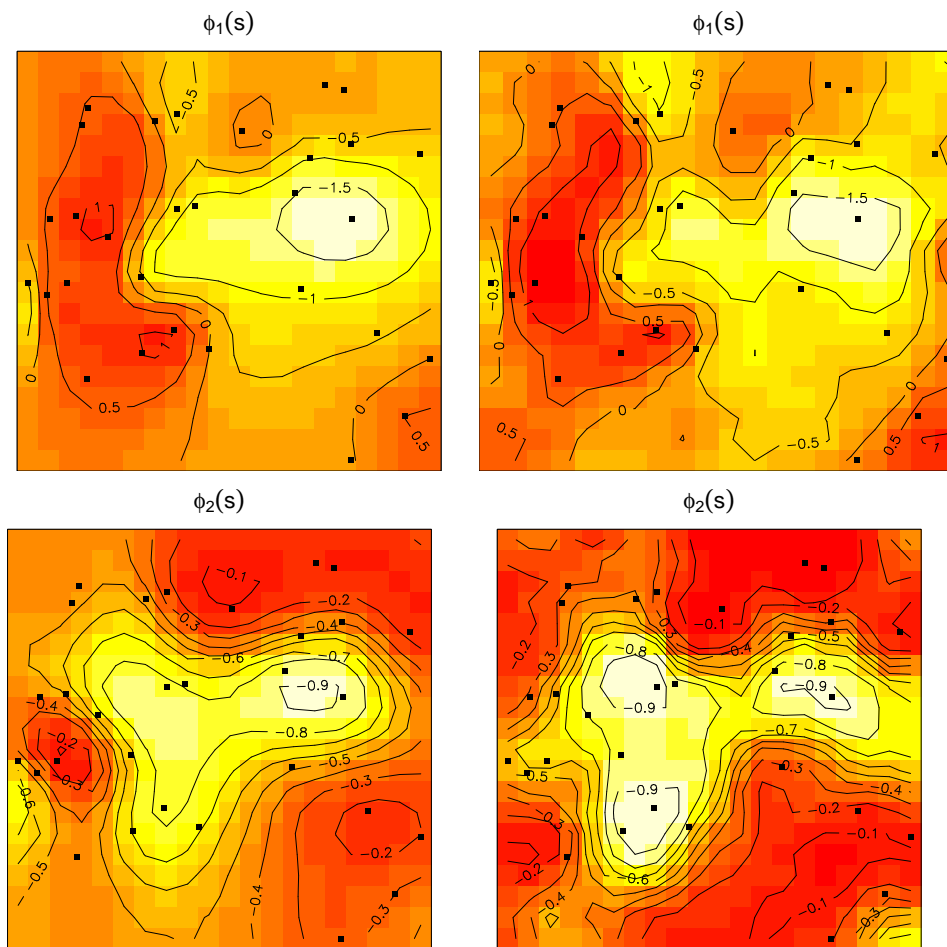


Figura 5.8: Superfície prevista (esquerda) e verdadeira (direita) para os parâmetros $\Phi_1(s)$ e $\Phi_2(s)$ para os dados artificiais.

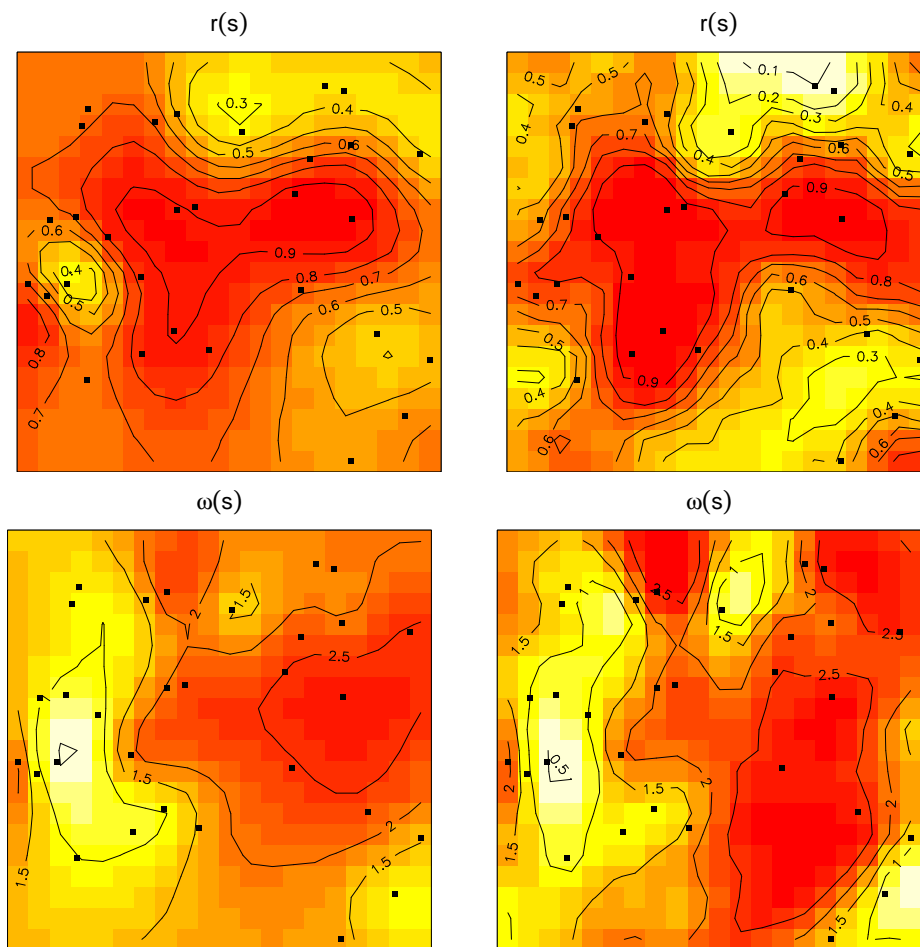


Figura 5.9: Superfície prevista (esquerda) e verdadeira (direita) para os parâmetros $r_1(s)$ e $\omega_1(s)$ - Dados Artificiais.

A previsão temporal 10 passos a frente, também foi calculada e o resultado para 4 das 30 estações, é mostrado na Figura 5.10. O valor verdadeiro está sempre dentro do intervalo de 95% de credibilidade a posteriori para todos os tempos, indicando que o modelo está prevendo bem a estrutura dos dados.

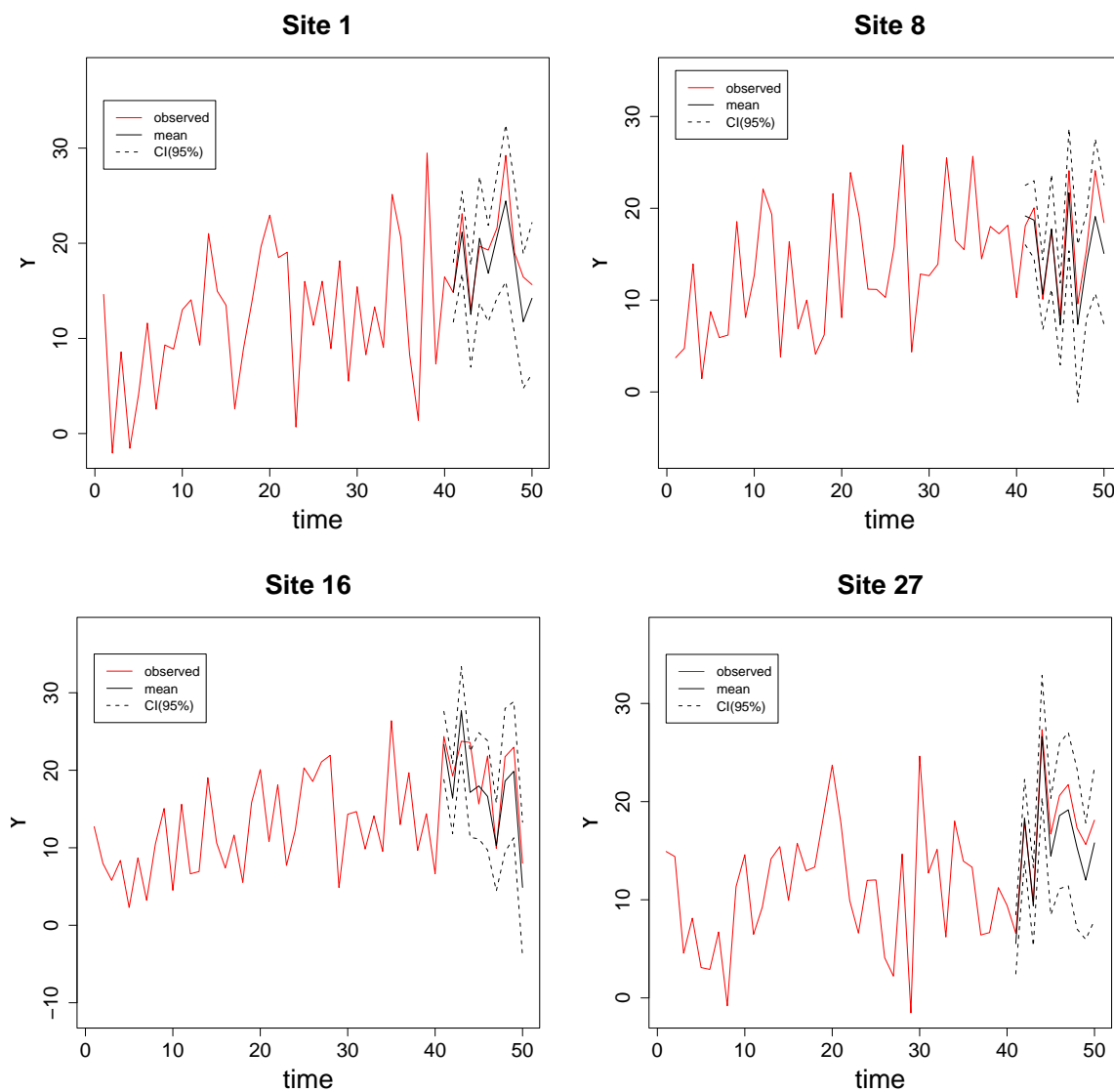


Figura 5.10: Previsão temporal 10 passos a frente para 4 estações segundo o modelo SVAR(2) para os dados artificiais.

5.1.2 Análise dos níveis de PM_{10} em Santiago do Chile, Chile

Nesta Subsecção analisamos a concentração mensal de material particulado PM_{10} em 8 estações monitoradoras em Santiago do Chile, Chile, no período de 1998 a 2003. O PM_{10} constitui o principal contaminante do ar em Santiago, e representa o objeto de análise da maioria dos estudos epidemiológicos. O problema da poluição é grave, especialmente na região metropolitana, onde as médias mensais para os meses de inverno ultrapassam o padrão aceitável. A cidade de Santiago do Chile possui um clima temperado com chuvas durante o inverno, sendo maio e agosto, os meses mais frios. A Figura 5.11 apresenta as séries temporais originais para cada estação, bem como as respectivas medidas de anomalias. A anomalia é calculada pela diferença entre o valor observado em um determinado mês e a média do mês. A anomalia é uma medida comumente utilizada em ciências ambientais. Podemos observar, que parece existir uma tendência decrescente nos dados. A Figura 5.12 mostra a localização das 8 estações monitoradoras, bem como a superfície de altitude e a temperatura observada em 3 pontos da cidade. A temperatura observada é praticamente a mesma nas 3 estações e, portanto, usaremos a temperatura média como a temperatura observada para todas as estações. Notamos um aumento gradativo na altitude no sentido leste-oeste.

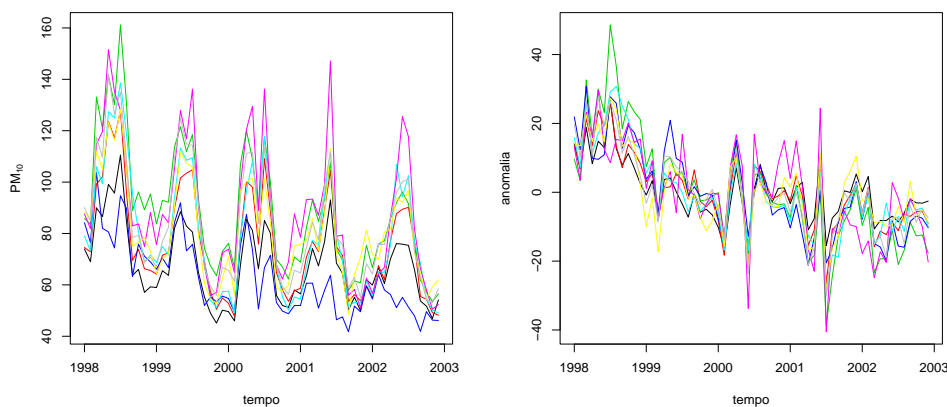


Figura 5.11: Níveis mensais de PM_{10} (direita) e anomalias mensais (esquerda) em Santiago do Chile, Chile no período de 1998 a 2003.

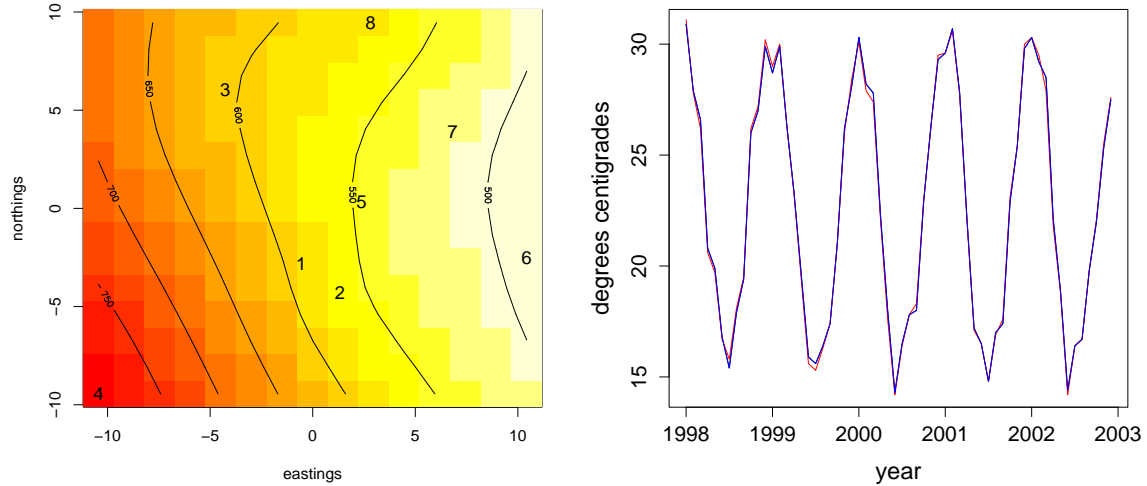


Figura 5.12: Localização das estações monitoradoras e superfície de altitude em metros (direita) e temperatura mensal observada (esquerda) em 3 estações no período de 1998 à 2003, em Santiago do Chile, Chile.

Inicialmente, realizamos uma análise exploratória dos dados originais, ajustamos um modelo de regressão linear simples para cada estação, considerando tendência e as covariáveis altitude e temperatura. Um processo autoregressivo foi ajustado para os resíduos deste modelo. Usando o Critério de Informação Bayesiana (BIC), proposto por Schwarz (1978), a ordem estimada para cada série foi superior a 10. Em seguida, utilizamos o mesmo procedimento com as séries de anomalias e usando o mesmo critério BIC, a ordem estimada para cada série variou entre 1, 2 e 5. Sendo assim, ajustamos o modelo proposto para as séries de anomalias. Baseado na análise exploratória preliminar, ajustamos um SVAR(p) com $p = 1, 2$ e 5 . Uma das implicações do nosso modelo é especificar o número de raízes reais (R) e complexas (C). Neste caso, quando $p = 1$ especificamos, $R = 1$, quando $p = 2$, especificamos, $C = 1$, e, finalmente, $p = 5$, fixamos $C = 2$ e $R = 1$. Com base nos resultados obtidos na análise exploratória, propomos o ajuste do seguinte modelo

$$anom_t = F_t' \theta_t + \sum_{j=1}^p \Phi_j (anom_{t-j} - F_{t-j}' \theta_{t-j}) + v_t, \quad v_t \sim N(0, \Sigma)$$

$$F_t \theta_t = \beta_0 + \beta_1 t + \beta_2 h(s) + \beta_3 T_t(s), \quad \beta_k \sim N(0, \tau_\beta^2),$$

onde $anom_t$ denota o vetor de anomalias no tempo t . A estrutura de média é modelada através de um nível β_0 , tendência β_1 , e as covariáveis altitude ($h(s)$) e temperatura ($T_t(s)$) com coeficientes, β_2 e β_3 , respectivamente.

As priors utilizadas foram $p(\beta) \propto 1$ com $\beta = (\beta_0, \beta_1, \beta_2, \beta_3)$, $\sigma^2 \sim IG(1, 1)$, $\lambda \sim Ga(1, \frac{\max(dist)}{6})$ com $\max(dist)$ denotando a distância máxima, $\Psi_j \sim GP(0, \Sigma_j)$, $j = 1, \dots, R$, $\Psi_{1j} \sim GP(0, \Sigma_{1j})$ e $\Psi_{2j} \sim GP(0, \Sigma_{2j})$, $j = 1, \dots, C$, tal que $\Sigma_j = \Sigma_{1j} = \Sigma_{2j}$, utilizando uma função de correlação Matérn, com variância igual a 1,5, parâmetro de decaimento igual a 0,3 e de suavização igual a 1. Essa escolha fornece uma correlação de 5% quando a distância entre as estações é igual a metade da distância máxima. Estamos considerando que o processo é uma vez diferenciável e a variância consegue cobrir bem o espaço paramétrico das raízes, isto é, obtemos uma priori aproximadamente uniforme para as raízes, como discutido na Seção 4.3. Utilizamos a função de correlação exponencial para os dados, e a função de correlação Matérn para as raízes do SVAR(p). Essa escolha foi motivada pela idéia de que as superfícies das raízes devem ser mais suaves que a superfície dos dados.

Com objetivo de comparar os modelos ajustados, utilizamos o critério proposto por Gelfand e Ghosh (1998). A Figura 5.13 apresenta o resultado para cada componente no tempo e no espaço, e a Tabela 5.2 mostra o resultado geral. Segundo esse critério, o modelo com o melhor ajuste foi o SVAR(2).

p	P	G	D
1	66040.23	47710.39	89895.42
2	54656.05	46449.84	77880.97
5	59547.68	56352.16	87723.77

Tabela 5.2: Critério de Gelfand e Ghosh (1998) e seus componentes, para os modelos ajustados aos dados de anomalia.

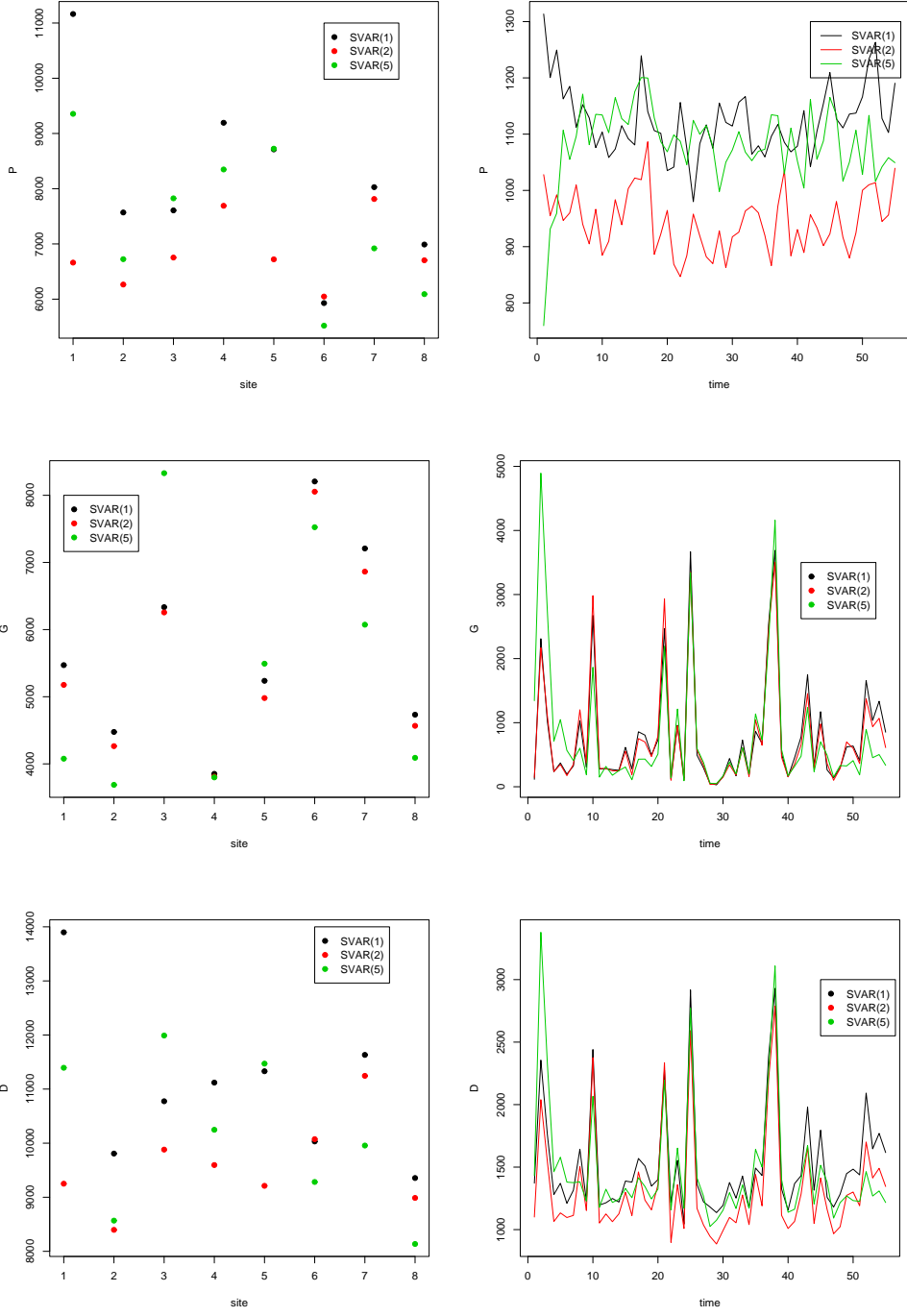


Figura 5.13: Critério de Gelfand e Gosh (1998) e seus componentes, para cada estação (à direita) e cada tempo (à esquerda) para os modelos ajustados aos dados de anomalia.

Apresentamos os resultados para $p = 2$ com $C = 1$, isto é, um par de raiz complexa e conjugada. A Tabela 5.3 mostra a média e o intervalo de credibilidade a posteriori para os parâmetros da estrutura de média para o SVAR(2) com um par de raízes complexas e conjugadas. Observamos que a inclinação é negativa e significativa, embora tanto a altitude quanto a temperatura não são significativas. A média a posteriori dos parâmetros da função de correlação dos erros fornece uma correlação variando de 0,479 a 0,927, entre as estações.

Coeficiente	Média	IC(95%)
Intercepto	18,69	(12,56; 24,83)
Inclinação	-0,63	(-0,83; -0,43)
Altitude	-0,11	(-1,02; 0,79)
Temperatura	0,69	(-1,08; 2,71)

Tabela 5.3: Média e Intervalo de 95% de Credibilidade a posteriori para os parâmetros da estrutura de média, segundo o modelo SVAR(2) para os dados de anomalia.

A Figura 5.14 apresenta uma interpolação espacial para os parâmetros $r_1(s)$ e $\omega_1(s)$ que representam respectivamente o módulo e o período da série. Observamos um padrão similar entre as superfícies, ou seja, tanto o módulo quanto o período são menores na região central da cidade.

A evolução da superfície de anomalia estimada para uma grade de 225 pontos entre o período de julho a dezembro de 2000 é apresentado na Figura 5.15.

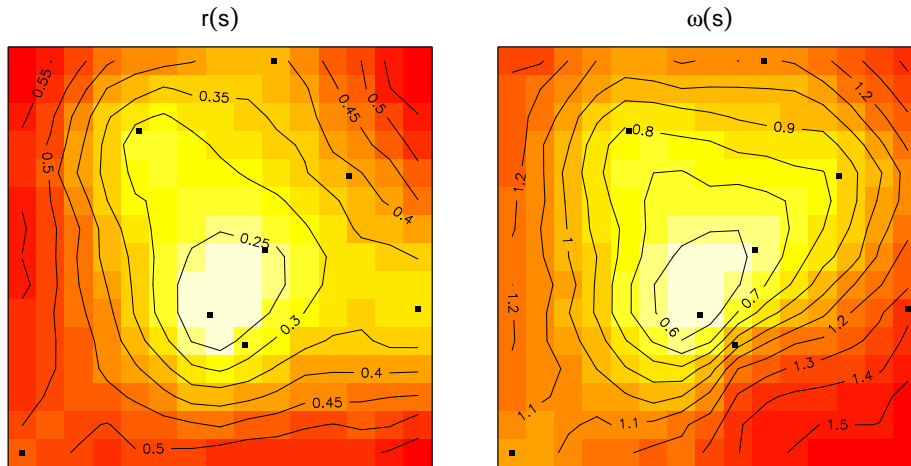


Figura 5.14: Superfície estimada para os parâmetros $r_1(s)$ e $\omega_1(s)$ representando o módulo e o período, respectivamente, para os dados de anomalia.

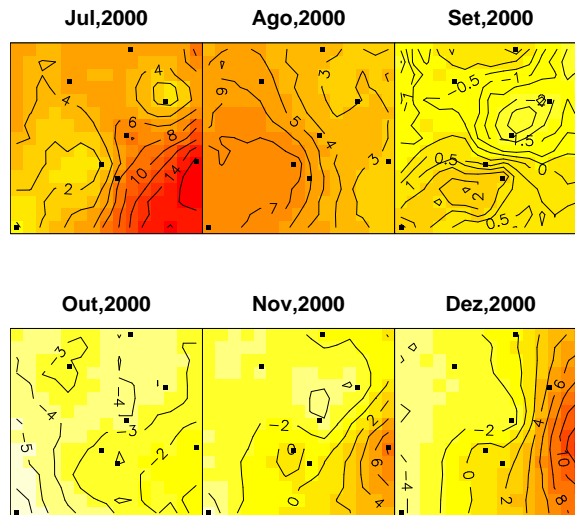


Figura 5.15: Mediana a posteriori da distribuição preditiva das anomalias de PM_{10} em Santiago do Chile, Chile, no período de junho 2000 à janeiro de 2001, segundo o modelo SVAR(2).

Com o objetivo de verificar a capacidade de fazer previsão temporal, o modelo foi ajustado deixando de fora 20 meses. Em seguida, calculamos a previsão temporal 20 passos a frente, conforme descrito na Seção 4.5. O resultado para as 8 estações encontram-se nas Figuras 5.16 e 5.17. O resultado é razoavelmente satisfatório, pois o intervalo de credibilidade de 95% a posteriori contém o verdadeiro valor, embora a amplitude dos intervalos seja grande. Além disso, observamos que no final da série, a média a posteriori decresce e o verdadeiro valor cresce, para algumas estações.

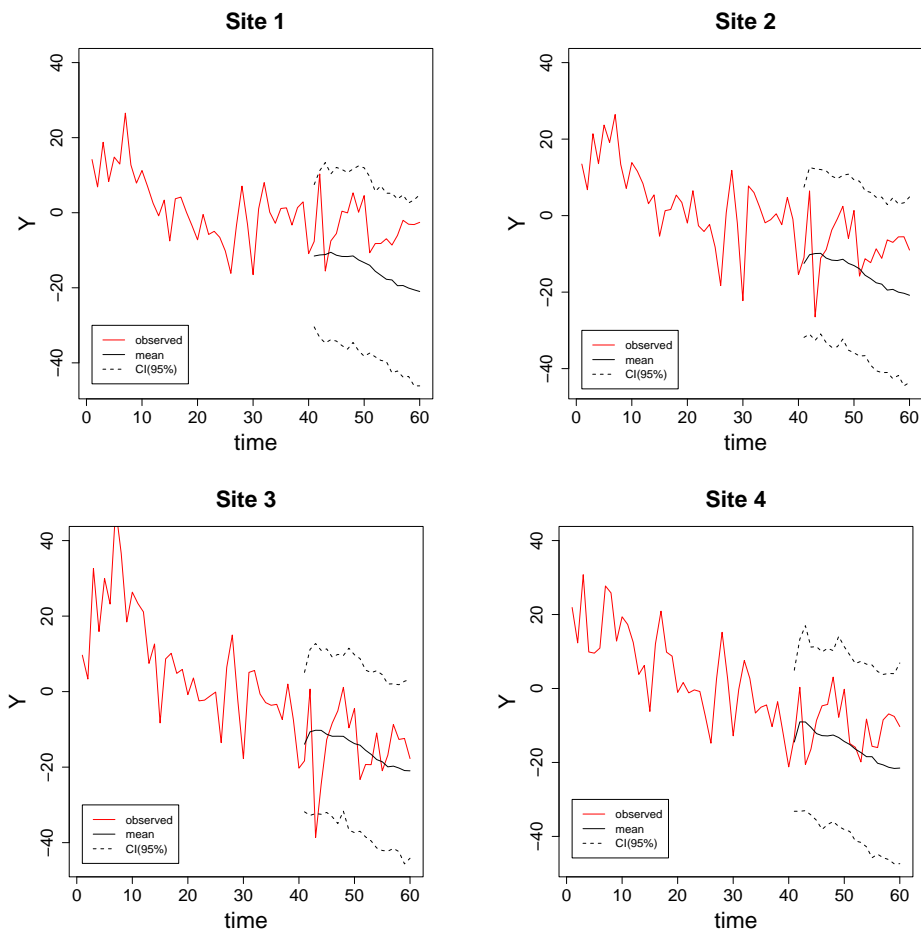


Figura 5.16: Previsão Temporal 20 meses a frente das estações 1 a 4, para as anomalias de PM_{10} em Santiago do Chile, Chile, segundo o modelo SVAR(2).

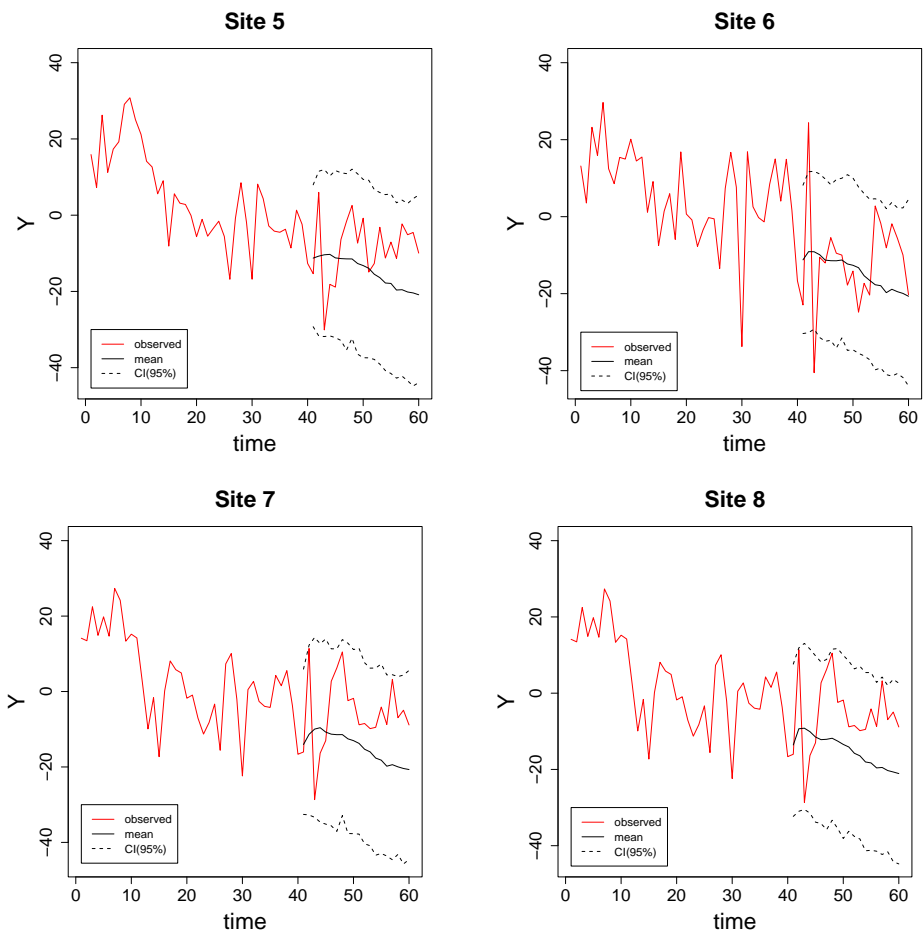


Figura 5.17: Previsão Temporal 20 meses a frente das estações 5 a 8, para as anomalias de PM_{10} em Santiago do Chile, Chile, segundo o modelo SVAR(2).

5.2 Ajustando o modelo proposto considerando a ordem desconhecida

Simulamos um conjunto de observações para 30 localizações ao longo de 50 instantes de tempo, de acordo com o modelo proposto em (4.2) com $p = 3$ raízes reais. Os dados foram gerados sem a presença de covariáveis e/ou fatores sazonais. Utilizamos a função de correlação Matérn para a covariância do processo, Σ , e para a variância das raízes reais, Σ_j , $j = 1, 2$ e 3 . O parâmetro que controla o decaimento foi fixado em 0,17, escolhido com base na idéia da distância máxima e, o parâmetro que controla a suavidade do processo foi fixado em 1, para ambas as correlações. A variância do processo foi $\sigma^2 = 1$, enquanto a variância das raízes recíprocas foi $\sigma_j^2 = 2, 5$. Os dados simulados com estas especificações, bem como a localização das 30 estações encontram-se na Figura 5.18. Em seguida, ajustamos o modelo proposto utilizando as duas prioris de mistura propostas.

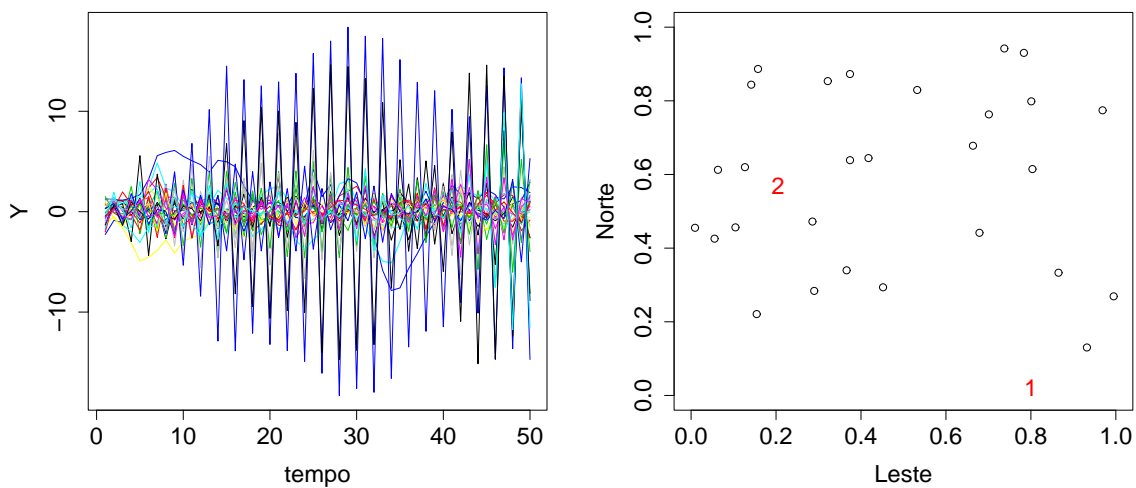


Figura 5.18: Dados simulados para 30 localizações ao longo de 50 instantes de tempo utilizando o modelo em (4.2) com $p=3$ raízes reais e a respectiva localização das estações monitoradoras.

5.2.1 Priori de mistura com massa de probabilidade em zero

Ajustamos o modelo fixando o número máximo de raízes reais em $R_+ = 5$, utilizando diferentes variâncias escolhidas de acordo com a expressão (4.10), fixando $p(I(a_j) = 1)$ em diferentes valores e

$\pi_R = 0, 5$. Observamos que a probabilidade da variável indicadora é bastante sensível a escolha de σ_j^2 . Para 5 escolhas de σ_j^2 , a ordem estimada foi igual a ordem máxima do modelo. Quando aumentamos a variância para $\sigma_j^2 = 2, 5$, a ordem estimada foi $R = 4$ e quando $\sigma_j^2 = 5$, a ordem estimada foi $R = 3$. Uma observação importante aqui é que a cadeia não passeia muito pelas diferentes ordens do modelo. Por exemplo, quando a ordem estimada foi $p = 4$, a cadeia visitou a ordem 4, 98% das vezes.

5.2.2 Priori de mistura baseada em duas distribuições contínuas

Inicialmente, ajustamos o modelo utilizando diferentes especificações escolhidas baseados na expressão (4.12). Fixando $p(I(a_j) = 1)$ em 5 valores diferentes e $\sigma_j^* = 1$, obtemos diferentes valores para σ_j . Após 30000 iterações, a ordem estimada pelo modelo utilizando as diferentes prioris é apresentada na Figura 5.19. Como esperado, quanto menor a probabilidade $p(I(a_j) = 1)$ maior a probabilidade das raízes virem da densidade $g(a_j)$.

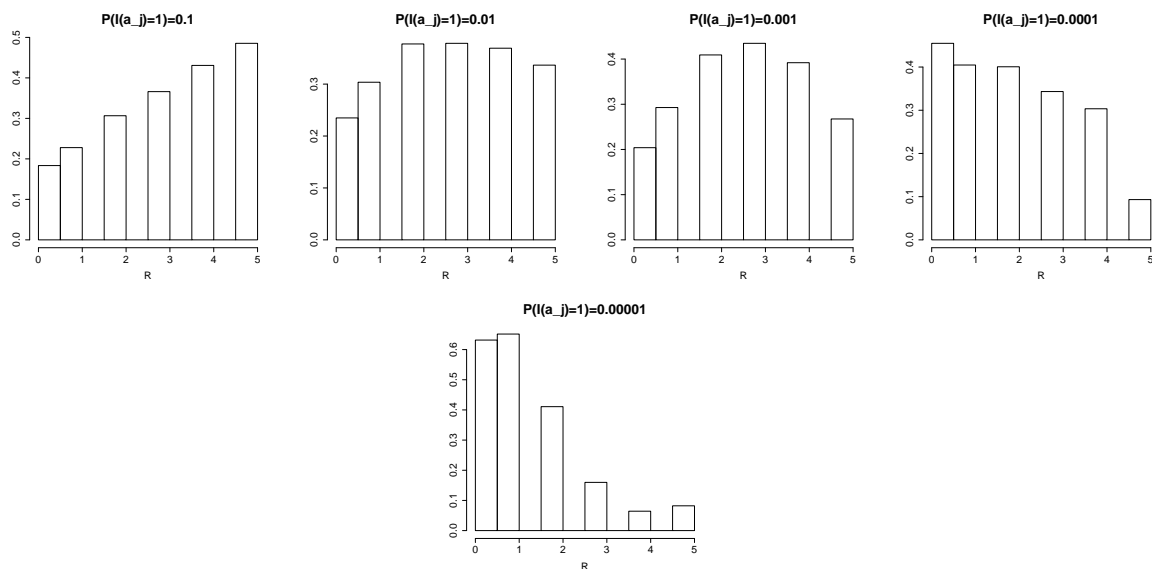


Figura 5.19: Ordem estimada pelo modelo utilizando diferentes valores de σ_j^2 e $p(I(a_j) = 1)$.

Escolhemos a priori tal que $p(I(a_j) = 1) = 0,001$ e geramos 100000 iterações. Descartamos as 50000 iterações iniciais e coletamos a amostra gerada a cada 50 iterações, obtendo uma amostra final de tamanho 1000. A Tabela 5.4 apresenta os resultados do ajuste do modelo para os parâmetros λ , σ^2 , κ e π_R . A média a posteriori para λ e κ encontram-se bem próxima do valor verdadeiro. Já para variância σ^2 , o intervalo de credibilidade possui um comprimento grande, incluindo o verdadeiro valor.

	Verdadeiro	Média	IC(95%)
λ	0,14	0,102	(0,049; 0,184)
σ^2	1	4,058	(0,734; 14,764)
κ	1	0,936	(0,385; 1,613)
π_R	-	0,5	(0,032; 0,963)

Tabela 5.4: Média e Intervalo de 95% de Credibilidade a posteriori para os parâmetros fixos no tempo, segundo o modelo SVAR(p) com ordem desconhecida.

A Figura 5.20 apresenta o histograma da distribuição a posteriori para a ordem do modelo, R . Notamos que a ordem sugerida pelo modelo de acordo com o histograma é 2 ou 3.

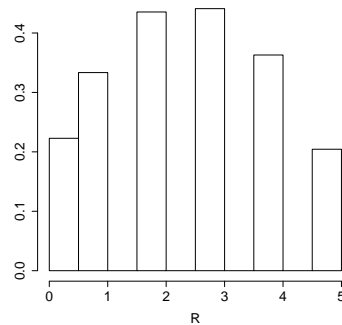


Figura 5.20: Ordem estimada pelo modelo fixando $p(I(a_j) = 1) = 0,001$.

O boxplot da distribuição a posteriori para os coeficientes do SVAR(p) para cada estação, isto é, $\phi_j(s)$, $j = 1, \dots, 5$ encontra-se na figura 5.21. O valor verdadeiro está marcado com o símbolo "x". Podemos observar que o valor verdadeiro encontra-se bem próximo a mediana a posteriori para $\phi_1(s)$ e $\phi_2(s)$, $\forall s$. Os valores verdadeiros do coeficiente $\phi_3(s)$ encontram-se distantes da distribuição a posteriori para algumas estações. A mediana a posteriori para algumas estações estão longe do zero, indicando que a ordem estimada do modelo pode ser igual a 3. Já os coeficientes $\phi_4(s)$ e $\phi_5(s)$ encontram-se bem próximos do zero, $\forall s$.

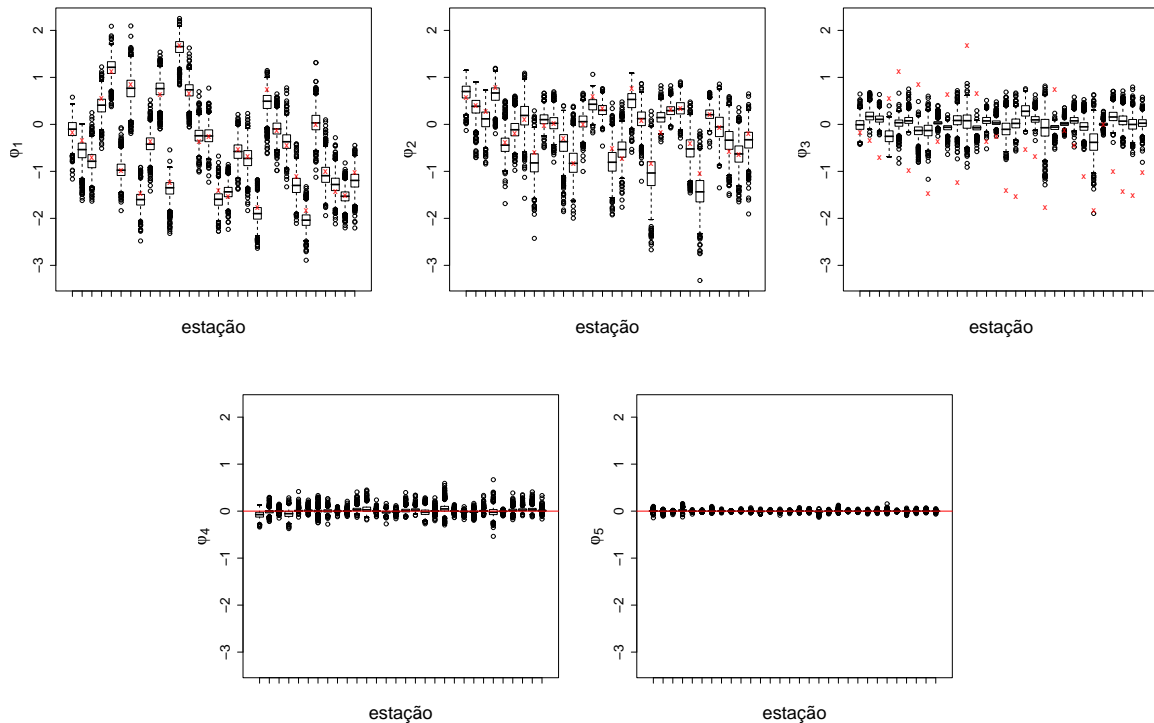


Figura 5.21: Boxplot da distribuição a posteriori para os parâmetros $\phi_j(s)$, $j = 1, \dots, 5$ para cada estação.

Neste capítulo ajustamos o modelo proposto no Capítulo 4 para diferentes situações. Inicialmente, consideramos a ordem do modelo conhecida e ajustamos o modelo a um conjunto de dados artificiais, gerados a partir do próprio modelo. Os resultados do ajuste do modelo para os dados artificiais foram satisfatórios. O modelo consegue estimar bem os parâmetros, inclusive os coeficientes da componente autoregressiva que variam no espaço. Em seguida, aplicamos o modelo aos dados de anomalias mensais de PM_{10} . Consideramos três diferentes ordens para o modelo, com a presença de raízes reais e complexas. O modelo escolhido segundo o critério de Gelfand e Ghosh (1998) foi o SVAR(2), com um par de raízes complexas e conjugadas. As superfícies estimadas indicam que parece fazer sentido considerar uma variação espacial para os coeficientes da componente autoregressiva.

Finalmente, consideramos a ordem do modelo desconhecida e ajustamos o modelo a um conjunto de dados simulados. Ajustamos o modelo utilizando as duas priors de mistura discutidas no Capítulo anterior. Observamos uma sensibilidade no cálculo da probabilidade da variável indicadora, com relação a escolha dos hiperparâmetros da priori. Quando consideramos a priori de mistura com massa de probabilidade no zero, parece existir um problema de identificabilidade. O ajuste utilizando a priori de mistura baseada em duas distribuições contínuas encontra algumas restrições com relação a escolha dos seus hiperparâmetros mas os resultados parecem mais promissores.

Capítulo 6

Considerações Finais e Trabalhos

Futuros

Neste capítulo apresentamos as considerações finais e discutimos possíveis extensões do trabalho. Dados espaço-temporais costumam ter alta dimensão e um dos problemas que encontramos para fazer inferência utilizando os modelos propostos na literatura é o alto custo computacional. Assim, discutimos que os modelos aqui propostos podem ser utilizados para a construção de estruturas que sejam capazes de reduzir a dimensão do problema. Outra extensão é o uso de distribuições a priori beta para as raízes dos processos autoregressivos.

6.1 Considerações Finais

A classe de modelos espaço-temporais proposta neste trabalho, baseada em convoluções discretas e processos autoregressivos, fornece uma estrutura de covariância bastante flexível. A estrutura mais simples é obtida considerando o produto entre uma covariância estacionária no tempo e uma covariância isotrópica no espaço, produzindo um processo espaço-temporal separável. Quando consideramos a evolução temporal dos parâmetros da função de covariância espacial, obtemos uma estrutura de covariância não separável e não estacionária, no tempo e no espaço. Essa evolução temporal é induzida através da variação da variância e/ou do parâmetro que controla o decaimento da função de correlação espacial. Os resultados encontrados revelam que faz mais sentido considerar a variância do

processo espacial evoluindo ao longo do tempo, do que o parâmetro da função de correlação.

Duas aproximações foram discutidas, o empilhamento de observações temporais ao longo de uma região (modelo de convolução temporal) e o empilhamento de observações espaciais ao longo do tempo (modelo de convolução espacial). Neste trabalho nos concentramos nos modelos de convolução espacial, já que superfícies espaciais evoluindo no tempo é fisicamente mais realista do que processos temporais evoluindo no espaço.

A estrutura de média do processo é modelada conjuntamente e, condicionalmente, o modelo pode ser escrito na forma dos modelos lineares dinâmicos, permitindo uma maneira eficiente de amostrar os coeficientes da estrutura de média.

O critério de comparação de modelos para processos espaço-temporais ainda é um desafio, já que temos que considerar as dimensões espacial e temporal. Como uma tentativa de levar em conta estas dimensões, utilizamos o critério proposto por Gelfand e Ghosh (1998) calculando suas componentes no tempo e no espaço.

Os modelos propostos fornecem uma maneira coerente de usar os processos autoregressivos para dados coletados no espaço e tempo. Os processos autoregressivos são incorporados no modelo para capturar estruturas locais tais como tendência ou ciclos que ainda persistem, mesmo após a remoção da estrutura de média.

O segundo modelo proposto, considera que os coeficientes da componente autoregressiva varia suavemente ao longo do espaço. A estrutura espacial do processo é obtida através de processos gaussianos transformados para as raízes do polinômio característico. As prioris impostas para as raízes garantem estacionariedade temporal para todas as localizações, evitando comportamentos explosivos.

Segundo esta abordagem, obtemos uma estrutura de covariância não separável mesmo considerando a função de correlação fixa no tempo. Essa característica é atingida por considerar a evolução espacial dos coeficientes da componente autoregressiva.

Uma extensão natural deste modelo é considerar que a ordem do $AR(p)$ é desconhecida. Para isso, propomos uma priori de mistura para as raízes recíprocas com massa de probabilidade no zero. Ao considerar esta aproximação, encontramos alguns problemas. Primeiro, a sensibilidade no sorteio da variável indicadora. Isto possivelmente deve-se ao fato de que é muito difícil dizer que as raízes para todas as estações é exatamente igual a zero. Além disso, quando as raízes são complexas observamos um

problema numérico. A segunda aproximação, considera uma priori de mistura de duas distribuições contínuas, de maneira que uma delas seja suficientemente concentrada em torno do zero, para que possamos obter coeficientes significativamente iguais a zero. Esta aproximação ainda encontra algumas restrições com relação a escolha dos hiperparâmetros da priori, mas os resultados obtidos parecem mais promissores. Além disso, podemos generalizar para o caso em que as raízes são complexas.

Como trabalho futuro, propomos o ajuste deste modelo a um conjunto de dados reais que possuam as características descritas pelo modelo.

6.2 Trabalhos Futuros

6.2.1 Problema de Redução de Dimensão

Dados ambientais são comumente medidos em n estações monitoradoras, ao longo de T observações no tempo. Em geral, tanto o número de estações (n) quanto o número de observações no tempo (T) são muito grandes. Portanto, um problema comum encontrado na modelagem de dados espaço-temporais é que estes modelos são computacionalmente intensivos se o número de estações monitoradoras, n , cresce. Quanto maior o número de estações monitoradoras e o número de observações temporais, T , maior é a dimensão da matriz de covariância. Portanto, obtemos matrizes de covariância de dimensão nT . Utilizando a aproximação proposta no Capítulo 3, temos n matrizes de dimensão $T \times T$ (convolução temporal) ou T matrizes de dimensão $n \times n$ (convolução espacial) e, ainda assim, obtemos matrizes de covariância de alta dimensão quando n é grande. O algoritmo de MCMC exige o armazenamento e a decomposição, de T matrizes de dimensão n , em cada iteração, demandando um alto tempo computacional. Assim, a redução na dimensão é um problema bastante relevante.

Existem alguns artigos na literatura que tratam deste problema de redução de dimensão. Mardia, Goodall, Redfern, e Alonso (1998) apresentam uma aproximação do filtro de Kalman espaço-temporal de dimensão reduzida. Essa abordagem combina duas técnicas conhecidas, a krigagem e o filtro de Kalman. Além disso, permite a modelagem de processos espaciais não estacionários e não separáveis. Wikle e Cressie (1999) consideram a redução de dimensão do vetor de estado de um modelo dinâmico fazendo projeções sobre uma base de polinômios ortogonais empíricos. Este é um método comumente utilizado em ciências atmosféricas que requer dados observados sobre uma grade regular e é similar a

análise de componentes principais. Nesta aproximação eles consideram a presença de um termo não dinâmico para capturar a variabilidade espacial de pequena escala. Outra aproximação é proposta por Higdon (2002). Esta aproximação usa modelos de convolução para construir processos espaciais e espaço-temporais flexíveis, que são capazes de acomodar uma grande quantidade de observações. Nesta abordagem uma forma paramétrica é considerada para o kernel. Assim, a escolha do kernel e do ruído branco permitem características desejáveis tais como não estacionariedade, não-separabilidade, campos não Gaussianos e redução de dimensão. Calder (2003) apresenta uma outra versão do filtro de kalman espaço-temporal de dimensão reduzida incorporando uma componente dinâmica no modelo de convolução discreta discutido anteriormente. Além disso, o modelo é estendido para o caso de processos espaço-temporais multivariados. A dimensão do modelo é reduzida através da construção de um modelo fatorial, onde cada fator evolui no tempo e os dados são modelados como uma média ponderada destes fatores. Lopes, Salazar, e Gamerman (2007) propõem um modelo fatorial espacial dinâmico para dados espaço temporais. A idéia é reduzir a dimensão do vetor de observações através de fatores latentes, e desta forma medir as correlações existentes entre as observações no espaço. O comportamento dinâmico do modelo é obtido através dos fatores e das suas matrizes de cargas que seguem processos autoregressivos de ordem um.

O modelo de convolução espacial proposto no Capítulo 3 pode ser utilizado na redução de dimensão. Assumindo que obtemos observações de um processo espacial ao longo de T instantes de tempo, Y_t , o processo em cada instante de tempo $t = 1, 2, \dots, T$ é descrito pela combinação linear de processos espaciais independentes. Mais especificamente, $Y_t = K\omega_t$ onde K é uma matriz, $n \times n$ triangular inferior e t é um vetor de dimensão n , cujas componentes são independentes entre si mas, cada uma, é espacialmente correlacionada. Então uma possibilidade para a redução de dimensão, é truncar o número de linhas da matriz K . Para isso pode ser utilizada uma decomposição de Cholesky com pivot, que corresponde a uma rotação de linhas e colunas de Σ , isto é, uma permutação na ordem dos locais, de maneira que a inferência do processo espacial não seja afetada. Suponha que um limiar C seja fixado e que apenas os m primeiros elementos da diagonal estejam acima de C . Se as colunas de K são truncadas nos elementos remanescentes da diagonal então todas as entradas truncadas de K serão menores que k_{mm} em valor absoluto. Seja \widetilde{K} uma matriz $n \times m$ dos elementos restantes, então

o modelo pode ser escrito como

$$\begin{aligned} Y_t &= \widetilde{K}\widetilde{\omega}_t + \eta_t, \quad \eta_t \sim N_n(0, \xi^2 I) \\ \widetilde{\omega}_t &= \sum_{i=1}^p \phi_i \widetilde{\omega}_{t-i} + \epsilon_t, \quad \epsilon_t \sim N_k(0, \sigma^2 I), \end{aligned}$$

tal que o modelo truncado pode também ser considerado na forma de um modelo linear dinâmico condicional.

6.2.2 Distribuição a priori beta para as raízes recíprocas

A priori proposta no Capítulo 4 para as raízes recíprocas dos processos autoregressivos é baseada em transformações de processos gaussianos, de maneira que as raízes recíprocas estejam definidas em um espaço paramétrico que garanta a estacionariedade do processo. Entretanto, observamos que esta priori pode apresentar um comportamento não desejado como discutido na Seção 4.3. Em outras palavras, a escolha da variância do processo gaussiano pode levar a prioris que dão alta probabilidade para os valores extremos. Com o objetivo de evitar este tipo de problema, pode ser utilizada uma priori beta para as raízes recíprocas. Por exemplo, para as raízes reais

$$a_j(s) = \left(2 \sum_{u=1}^U K_j(s-u)x_j(u) - 1\right),$$

onde K_j é um núcleo espacial e $x_j(u) \sim Beta(m_j, n_j)$. Observe que deslocamos a distribuição beta, com o objetivo de obter raízes definidas em um espaço paramétrico, de maneira a garantir a estacionariedade do processo. Prioris similares são obtidas para as raízes complexas, considerando o espaço paramétrico em que elas estão definidas. Podemos também considerar prioris de mistura baseadas nesta aproximação.

Apêndice A

Critério de Comparação de Modelos

Com o objetivo de comparar os modelos propostos com as diversas parametrizações da estrutura de covariância, calculamos o critério sugerido por Gelfand e Ghosh (1998) denotado aqui por D . Este critério é baseado na esperada perda preditiva a posteriori, considerando a bondade de ajuste e o poder preditivo dos modelos. A escolha de um critério de comparação para modelos espaço-temporais é ainda um desafio. Assim, inicialmente calculamos a estatística D para cada estação e para cada tempo, e finalmente uma quantidade D geral.

Esta medida é baseada na replicação dos dados observados. Seja $Y_{l,rep}^m$, $l = 1, \dots, L$, um vetor de replicações dos dados observados onde m representa o modelo. A estatística $D(m)$ é calculada pela combinação linear de duas quantidades: $P(m)$ e $G(m)$, mais especificamente,

$$\begin{aligned} P(m) &= \sum_{l=1}^L \text{Var}(Y_{l,rep}^m | \mathbf{y}) \\ G(m) &= \sum_{l=1}^L \left(E(Y_{l,rep}^m | \mathbf{y}) - \mathbf{Y}_{1,obs} \right)^2 \\ D(m) &= P(m) + \frac{1}{2}G(m). \end{aligned} \tag{A.1}$$

onde G representa a bondade de ajuste e P a penalização pela complexidade do modelo. A estatística $D(m)$ é calculada para cada uma das estações, ou seja, calculamos $D(m)_i$, $i = 1, \dots, n$ e $L = T$ na equação (A.1). Da mesma forma, calculamos $D(m)$ para cada tempo, $D(m)_i$, $i = 1, \dots, T$ e $L = n$. Finalmente, usando um duplo somatório, calculamos a estatística $D(m)$ geral.

Apêndice B

Relação entre as raízes recíprocas e os coeficientes de um modelo $AR(p)$

Este apêndice apresenta a relação entre as raízes recíprocas reais (a_j) e complexas ($r_j e^{\pm i\omega_j}$) com os coeficientes Φ_j do $AR(p)$.

Raízes reais

O polinômio característico considerando apenas as raízes reais é dado por

$$\prod_{j=1}^R (1 - a_j B) = (1 + A_1 B + A_2 B^2 + \dots + A_R B^R),$$

onde a_j denota as raízes reais conhecidas. Supomos inicialmente $R = 3$, $A_1 = -(a_1 + a_2)$ e $A_2 = a_1 \times a_2$.

Assim,

$$(1 + A_1 B + A_2 B^2)(1 - a_3 B) = 1 + (A_1 - a_3)B + (A_2 - a_3 A_1)B^2 - a_3 A_2 B^3.$$

Obtemos então, $A_1 = A_1 - a_3$, $A_2 = A_2 - a_3 A_1$ e $A_3 = -a_3 A_2$. Vamos supor agora $R = k + 1$. Assim,

$$\begin{aligned} \prod_{j=1}^k (1 - a_j B)(1 - a_{k+1} B) &= (1 + A_1 B + A_2 B^2 + \dots + A_k B^k)(1 - a_{k+1} B) \\ &= 1 + (A_1 - a_{k+1})B + (A_2 - A_1 a_{k+1})B^2 + \dots + (-A_k a_{k+1})B^{k+1}. \end{aligned}$$

Generalizando obtemos, $A_1 = A_1 - a_{k+1}$, $A_j = A_j - A_{j-1} a_{k+1}$ e $A_{k+1} = -A_k a_{k+1}$, $k = 2, \dots, R - 1$ e $j = 2, \dots, k$.

Raízes complexas

O polinômio característico considerando apenas as raízes complexas é dado por

$$\prod_{j=1}^C (1 - \phi_{1j}B - \phi_{2j}B^2) = 1 + C_1B + C_2B^2 + \dots + C_{2C}B^{2C}$$

onde $\phi_{1j} = 2r_j \cos(\omega_j)$ e $\phi_{2j} = -r_j^2$ são quantidades conhecidas. Supomos inicialmente $C = 2$. Assim,

$$(1 + C_1B + C_2B^2)(1 - \phi_{12}B - \phi_{22}B^2) = (1 + (C_1 - \phi_{12})B + (C_2 - C_1\phi_{12} - \phi_{22})B^2 - (C_1\phi_{22} - C_2\phi_{12})B^3 - C_2\phi_{22}B^4)$$

Obtemos então, $C_1 = C_1 - \phi_{12}$, $C_2 = C_2 - C_1\phi_{12} - \phi_{22}$, $C_3 = -C_1\phi_{22} - C_2\phi_{12}$ e $C_4 = -C_2\phi_{22}$. Vamos supor agora $C = k + 2$. Assim,

$$\begin{aligned} & \prod_{j=1}^k (1 - \phi_{1j}B - \phi_{2j}B^2)(1 - \phi_{1k+1}B - \phi_{2k+1}B^2) \\ &= (1 + C_1B + C_2B^2 + \dots + C_kB^k)(1 - \phi_{1k+1}B - \phi_{2k+1}B^2) \\ &= 1 + (C_1 - \phi_{1k+1})B + (C_2 - \phi_{2k+1} - C_1\phi_{1k+1})B^2 + (C_3 - C_2\phi_{1k+1} - C_1\phi_{2k+1})B^3 + \\ & \quad \dots + (-C_{k+1}\phi_{2k+1} - C_k\phi_{1k+1})B^{k+1} + (-C_k\phi_{2k+1})B^{k+2}. \end{aligned}$$

Generalizando obtemos, $C_1 = C_1 - \phi_{1k+1}$, $C_2 = C_2 - \phi_{2k+1} - C_1\phi_{1k+1}$, $C_j = C_j - C_{j-1}\phi_{1k+1} - C_{j-2}\phi_{2k+1}$, $C_{k+1} = -C_{k+1}\phi_{2k+1} - C_k\phi_{1k+1}$ e $C_{k+2} = -C_k\phi_{2k+1}$, $k = 3, \dots, C$ e $j = 3, \dots, k + 1$.

Caso geral

Finalmente, calculando o produto entre os polinômios para os casos reais e complexos, obtemos:

$$\begin{aligned} & \prod_{j=1}^R (1 - a_jB) \prod_{j=1}^C (1 - \phi_{1j}B - \phi_{2j}B^2) \\ &= (1 + A_1B + A_2B^2 + \dots, A_RB^R)(1 + C_1B + C_2B^2 + \dots, C_{2C}B^{2C}) \\ &= 1 + (A_1 + C_1)B + (A_2 + C_2 + A_1C_1)B^2 + (A_3 + C_3 + A_1C_2 + A_2C_1)B^3 + \dots + A_RC_{2C}B^{R+2C} \\ &= 1 + \Phi_1B + \Phi_2B^2 + \dots + \Phi_pB^p. \end{aligned}$$

Assim,

$$\Phi_i = A_i + C_i + \sum_{\substack{j,k=1 \\ j+k=i}}^{\max(R,2C)} A_jC_k$$

onde $A_j = 0$ para $j > R$ e $C_k = 0$ para $k > 2C$.

Referências

- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control* 19, 716–723.
- Anderson, T. (1994). *An Introduction to Multivariate Statistical Analysis*. John Wiley & Sons, Inc.
- Barry, R. P. e J. M. Ver Hoef (1996). Blackbox kriging: Spatial prediction without specifying variogram models. *Journal of Agricultural, Biological and Environmental Statistics* 1, 297–322.
- Box, G., G. Jenkins, e G. Reinsel (1994). *Time Series Analysis*. New Jersey: Prentice Hall.
- Brown, P. E., K. F. Karesen, G. O. Roberts, e S. Tonellato (2000). Blur-generated non-separable space-time models. *Journal of the Royal Statistical Society B* 62, 847–860.
- Brown, P. J., N. D. Le, e J. V. Zidek (1994). Multivariate spatial interpolation and exposure to air pollutants. *The Canadian Journal of Statistics* 22, 489–509.
- Calder, C. A. (2003). Exploring latent structure in spatial temporal processes using process convolutions. Tese de doutorado, Instituto de Estatística e Ciências da Decisão, Universidade Duke.
- Carter, C. K. e R. Kohn (1994). On Gibbs sampling for state space models. *Biometrika* 81, 541–553.
- Cressie, N. e H. Huang (1999). Classes of nonseparable, spatio-temporal stationary covariance functions. *Journal of the American Statistical Association* 94, 1330–1340.
- Cressie, N. A. C. (1993). *Statistical for Spatial Data*. New York: Wiley.
- Doornik, J. A. e M. Ooms (2001). Introduction to Ox. Ox version 3.
- Fruhworth-Schnater, S. (1994). Data augmentation and dynamic linear models. *Journal of Time Series Analysis* 15, 183–202.

- Fuentes, M. (2002). Spectral methods for nonstationary spatial processes. *Biometrika* 89, 197–210.
- Fuentes, M., L. Chen, J. M. Davis, e G. M. Lackmann (2005). Modeling and predicting complex space-time structures and patterns of coastal wind fields. *Environmetrics* 16, 449–464.
- Gamerman, D. e H. F. Lopes (2006). *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*. Chapman & Hall.
- Gelfand, A., A. Schmidt, S. Banerjee, e C. Sirmans (2004). Nonstationary multivariate process modeling through spatially varying coregionalization (with discussion). *TEST* 13, 1–50.
- Gelfand, A. E., S. Banerjee, e D. Gamerman (2005). Spatial process modelling for univariate and multivariate dynamic spatial data. *Environmetrics* 16, 465–479.
- Gelfand, A. E. e S. K. Ghosh (1998). Model choice: a minimum posterior predictive loss approach. *Biometrika* 85, 1–11.
- Gelman, A., J. B. Carlin, H. S. Stern, e D. B. Rubin (2004). *Bayesian Data Analysis*. Chapman & Hall - Second Edition.
- Gelman, A. e D. B. Rubin (1992). Inference from iterative simulation using multiple sequences. *Statistical Science* 7, 457–472.
- Gneiting, T. (2002). Nonseparable, stationary covariance functions for space-time data. *Journal of the American Statistical Association* 97, 590–600.
- Goulard, M. e M. Voltz (1992). Linear Coregionalization Model: Tools for estimation and choice of cross-variogram matrix. *Mathematical Geology* 24, 269–286.
- Green, P. (1995). Reversible jump markov chain monte carlo computation and bayesian model determination. *Biometrika* 82, 711–732.
- Higdon, D. (1998). A process-convolution approach to modelling temperatures in the north atlantic ocean. *Environmental and Ecological Statistics* 5, 173–190.
- Higdon, D. (2002). Space and space-time modeling using process convolutions. In C. Anderson, V. Barnett, P. C. Chatwin, e A. H. El-Shaarawi (Eds.), *Quantitative Methods for Current Environmental Issues*, pp. 37–56. Springer Verlag.

- Higdon, D., J. Swall, e J. Kern (1999). Non-stationary spatial modelling. In A. P. D. J. M. Bernardo, J. O. Berger e A. F. M. Smith (Eds.), *Bayesian Statistics 6 - Proceedings of the Sixth Valencia Meeting*, pp. 761–768. Claredon Press - Oxford.
- Huang, H. e N. Cressie (1996). Spatio-temporal prediction of snow water equivalent using the Kalman filter. *Computational Statistics and Data Analysis* 22, 159–175.
- Huerta, G. e R. Prado (2006). Structured priors for multivariate time series. *Journal of Statistical Planning and Inference* 136, 3802–3821.
- Huerta, G., B. Sansó, e J. R. Stroud (2004). A spatio temporal model for mexico city ozone levels. *Applied Statistics* 53, 231–248.
- Huerta, G. e M. West (1999). Priors and component structures in autoregressive time series models. *Journal of the Royal Statistical Society, Series B* 61, 881–899.
- Kim, S., N. Shepard, e S. Chib (1998). Stochastic volatility: likelihood inference and comparison with arch models. *Rev. Fin. Stud.* 65, 361–393.
- Le, N. D. e J. V. Zidek (1992). Interpolation with uncertain spatial covariances: A Bayesian alternative to kriging. *Journal Multivariate Analysis* 43, 351–374.
- Lopes, H. F., E. Salazar, e D. Gamerman (2007). Dynamic spatial factor model. Relatório técnico, DME, UFRJ.
- Mardia, K. V., C. Goodall, E. J. Redfern, e F. J. Alonso (1998). The kriged kalman filter. *Test* 7, 217–285.
- Mardia, K. V. e C. R. Goodall (1993). Spatial-temporal analysis of multivariate environmental monitoring data. In G. P. Patil e C. R. Rao (Eds.), *Multivariate Environmental Statistics*, pp. 347–386. Elsevier Science Publishers B. V.
- O’Hagan, A. (1994). *Kendall’s Advanced Theory of Statistics, Volume 2B, Bayesian Inference*. John Wiley & Sons, Inc.
- Paciorek, C. e M. Schervish (2006). Spatial modelling using a new class of nonstationary covariance functions. *Environmetrics* 17, 483–506.

- Paez, M., D. Gamerman, e V. d. Oliveira (2005). Interpolation performance of a spatio-temporal model with spatially varying coefficients: application to pm_{10} concentrations in rio de janeiro. *Environmental and Ecological Statistics* 12, 169–193.
- Paez, M. S. e D. Gamerman (2003). Study of the space-time effects in the concentration of airborne pollutants in the metropolitan region of Rio de Janeiro. *Environmetrics* 14, 1–22.
- Paez, M. S. e D. Gamerman (2005). *Modelagem de Processos espaço-temporais*. Minicurso 11a Escola de Séries Temporais, ABE. Vila Velha, ES, Brasil.
- Prado, R. e G. Huerta (2002). Time-varying autoregressions with model order uncertainty. *Journal of Time Series Analysis* 23, 599–618.
- Roberts, G. O. (1996). Markov chain concepts related to sampling algorithms. In W. R. Gilks, S. Richardson, e D. J. Spiegelhalter (Eds.), *Markov Chain Monte Carlo in Practice*, pp. 45–57. Chapman & Hall.
- Rosenblatt, M. (1952). Remarks on a multivariate transformation. *Annals of Mathematical Statistics* 23, 470–472.
- Sampson, P. e P. Guttorp (1992). Nonparametric estimation of nonstationary spatial covariance structure. *Journal of the American Statistical Association* 87, 108–119.
- Sansó, B. e L. Guenni (2000). A non-stationary multisite model for rainfall. *Journal of the American Statistical Association* 95, 1089–1100.
- Sansó, B., A. M. Schmidt, e A. A. Nobre (2006). Bayesian spatio-temporal models based on discrete convolutions. Relatório técnico, DME-UFRJ.
- Schmidt, A. M. e A. E. Gelfand (2003). A Bayesian coregionalization approach for multivariate pollutant data. *Journal of Geophysical Research-Atmospheres* 108.
- Schmidt, A. M., A. A. Nobre, e G. S. Ferreira (2002). Alguns aspectos da modelagem de dados espacialmente referenciados. *Revista Brasileira de Estatística* 63, 59–88.
- Schmidt, A. M. e A. O’Hagan (2003). Bayesian inference for nonstationary spatial covariance structures via spatial deformations. *Journal of the Royal Statistical Society Series B* 65, 743–758.

- Schmidt, A. M. e B. Sansó (2006). *Modelagem Bayesiana da Estrutura de Covariância de Processos Espaciais e Espaço-Temporais*. Minicurso 17o Simpósio Nacional de Probabilidade e Estatística, ABE. Caxambu, MG, Brasil.
- Schwarz, G. (1978). Estimating the dimension of a model. *Annals of Statistics* 6, 461–464.
- Smith, R. L. (1996). Estimating nonstationary spatial correlations. Relatório técnico, Cambridge University, UK.
- Stein, M. L. (2005). Space-time covariance functions. *Journal of the American Statistical Association* 100, 310–321.
- Storvik, G., A. Frigessi, e D. Hirst (2002). Stationary space-time gaussian fields and their time autoregressive representation. *Statistical Modelling* 2, 139–161.
- Stroud, J. R., P. Müller, e B. Sansó (2001). Dynamic model for spatio-temporal data. *Journal of the Royal Statistical Society, B*, 63, 673–689.
- Thiébaux, H. J. e M. A. Pedder (1987). Spatial objective analysis with applications in atmospheric science. Technical report, London: Academic Press.
- West, M. (1997). Time series decomposition. *Biometrika* 84, 489–494.
- West, M. e J. Harrison (1997). *Bayesian Forecasting and Dynamic Models*. New York: Springer - Second Edition.
- Whittle, P. (1954). On stationary processes in the plane. *Biometrika* 41, 434–449.
- Wikle, C. K. e N. Cressie (1999). A dimension-reduced approach to space-time kalman filtering. *Biometrika* 86, 815–829.

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](#)