

Universidade Federal do Rio de Janeiro

UM ESQUEMA EFICIENTE DE AMOSTRAGEM EM MODELOS DINÂMICOS  
GENERALIZADOS COM APLICAÇÕES EM FUNÇÕES DE TRANSFERÊNCIA

Romy Elena Rodríguez Ravines

2006



UFRJ

# UM ESQUEMA EFICIENTE DE AMOSTRAGEM EM MODELOS DINÂMICOS GENERALIZADOS COM APLICAÇÕES EM FUNÇÕES DE TRANSFERÊNCIA

Romy Elena Rodríguez Ravines

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.

Orientador: Prof. Helio S. Migon

Co-orientadora: Profa. Alexandra M. Schmidt

Rio de Janeiro, Dezembro de 2006

# UM ESQUEMA EFICIENTE DE AMOSTRAGEM EM MODELOS DINÂMICOS GENERALIZADOS COM APLICAÇÕES EM FUNÇÕES DE TRANSFERÊNCIA

Romy Elena Rodríguez Ravines

Orientador: Prof. Helio S. Migon

Co-orientadora: Profa. Alexandra M. Schmidt

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, Prof. Helio S. Migon  
IM-UFRJ

---

Profa. Alexandra M. Schmidt  
IM-UFRJ

---

Prof. Dani Gamerman  
IM-UFRJ

---

Prof. Hedibert F. Lopes  
GSB-UC, EUA

---

Prof. Nikolai V. Kolev  
IME-USP

---

Prof. Francisco Cribari Neto  
DE-UFPE

Rio de Janeiro, Dezembro de 2006

Ravines, Romy Elena Rodríguez

Um Esquema Eficiente de Amostragem em Modelos Dinâmicos Generalizados com Aplicações em Funções de Transferência/ Romy Elena Rodríguez Ravines.- Rio de Janeiro: UFRJ/IM, 2006.

xiii, 130f.: il.; 31cm.

Orientadores: Helio S. Migon, Alexandra M. Schmidt

Tese (doutorado) - UFRJ/IM/ Programa de Pós-graduação em Estatística, 2006.

Referências Bibliográficas: f.131-136.

1. MCMC. 2. Inferência Bayesiana. 3. Modelos Espaço-Temporais.  
I. Migon, Helio. II. Universidade Federal do Rio de Janeiro, Instituto de Matemática. III. Título.

### **Agradecimentos**

Esta tese foi possível graças ao apoio financeiro da CAPES, à valiosa colaboração e à atenção dedicada a este trabalho por meus orientadores Helio S. Migon e Alexandra M. Schmidt, ao incentivo de todos os meus amigos do DME, ao apoio incondicional de minha família e, em especial, de meu querido Ralph.

A todos, muito obrigada.

## RESUMO

### UM ESQUEMA EFICIENTE DE AMOSTRAGEM EM MODELOS DINÂMICOS GENERALIZADOS COM APLICAÇÕES EM FUNÇÕES DE TRANSFERÊNCIA

Romy Elena Rodríguez Ravines

Orientador: Prof. Helio S. Migon

Co-orientadora: Profa. Alexandra M. Schmidt

Resumo da 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.

Os principais objetivos desta tese foram: (i) propor um esquema de amostragem eficiente para fazer inferência em modelos dinâmicos não normais e não lineares, usando o enfoque bayesiano, e (ii) mostrar a sua aplicação em modelos de alta complexidade, em particular os modelos de função de transferência.

O esquema de amostragem aqui proposto é denominado *CUBS*, abreviação do inglês *Conjugate Updating Backward Sampling*. Este esquema combina duas propostas previamente estabelecidas na literatura: o *Conjugate Updating* ou *Linear Bayes* de West et al. (1985) e o *FFBS* de Frühwirth-Schnater (1994). Os resultados obtidos mostram que o esquema proposto é eficiente no sentido de reduzir significativamente o tempo computacional e ser de fácil implementação. Milhares de iterações do MCMC são realizadas em minutos e as cadeias geradas apresentam menos autocorrelação que as geradas com métodos de amostragem individual.

Utilizamos o esquema proposto na modelagem estocástica conjunta da chuva e vazão provenientes de múltiplas bacias. A modelagem utilizada é outra proposta desta tese. A classe dos modelos dinâmicos de função de transferência foi a escolhida para representar a relação entre ambas variáveis e um modelo espaço-temporal com troca de suporte foi escolhido para representar a chuva. Os resultados mostraram que nossa abordagem é bastante flexível e parcimoniosa no sentido de ter poucos parâmetros para representar (bem) todos os processos físicos envolvidos nessa relação.

*Palavras-chave:* Chuva-vazão; Defasagens Distribuídas; Inferência Bayesiana; Funções de Transferência; Linear Bayes; Markov chain Monte Carlo; Modelos Dinâmicos; Troca de Suporte; WinBUGS.

## ABSTRACT

### AN EFFICIENT SAMPLING SCHEME FOR DYNAMIC GENERALIZED MODELS WITH APPLICATIONS IN TRANSFER FUNCTIONS

Romy Elena Rodríguez Ravines

Orientador: Prof. Helio S. Migon

Co-orientadora: Profa. Alexandra M. Schmidt

*Abstract* da 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.

The main goals of this thesis were: (i) to propose an efficient sampling scheme to make inference on non-normal and non-linear dynamic models, under the Bayesian framework, and (ii) to show its application to complex models, in particular, the transfer function models.

The sampling scheme proposed here is called *CUBS*, short for *Conjugate Updating Backward Sampling*. This scheme combines two algorithms previously established in the literature: the *Conjugate Updating* or *Linear Bayes* from West et al. (1985) and, the *FFBS* from Frühwirth-Schnater (1994). The results showed that *CUBS* is efficient because its implementation is quite simple and has low demand of computational time. Thousands of iterations can be performed in seconds and the autocorrelation of its chains is lower than those from chains generated with simple move schemes.

*CUBS* was used in a particular approach, proposed here, for modelling rainfall and runoff jointly. The class of dynamic models of transfer functions was chosen to represent the relationship between both variables. A spatio-temporal model with change of support was used to model the rainfall. The results showed that our approach is flexible and parsimonious, this because it simplifies the physical processes involved in this relationship, in a function with few parameters all with a clear physical interpretation.

*Key-words*: Bayesian inference; Change of support; Distributed Lag; Dynamic Models; Linear Bayes; Transfer functions; Monte Carlo Markov chain, Rainfall-Runoff; WinBUGS.

## SUMÁRIO

<b>Lista de Tabelas</b>	<b>xi</b>
<b>Lista de Figuras</b>	<b>xii</b>
<b>Capítulo 1: Introdução</b>	<b>1</b>
1.1 Modelos de Função de Transferência . . . . .	1
1.2 Relação Chuva-Vazão . . . . .	2
1.3 Objetivos da Tese . . . . .	3
1.4 Organização da Tese . . . . .	4
<b>Capítulo 2: Inferência Bayesiana em Modelos de Defasagens Distribuídas</b>	<b>5</b>
2.1 Introdução . . . . .	5
2.2 Modelos de Defasagens Distribuídas . . . . .	7
2.2.1 Defasagens Distribuídas Infinitas . . . . .	7
2.2.2 Defasagens Distribuídas Finitas . . . . .	9
2.3 Modelos Lineares Dinâmicos . . . . .	9
2.3.1 Definição . . . . .	9
2.3.2 Modelos de Função de Transferência . . . . .	10
2.3.3 Funções de Transferência e Defasagens Distribuídas . . . . .	11
2.4 Procedimento de Inferência . . . . .	12
2.5 Aplicação: uma Função Consumo . . . . .	13
2.5.1 Dados e Modelos . . . . .	13
2.5.2 Resultados . . . . .	15
2.5.3 Seleção de Modelos . . . . .	17
2.5.4 Comparação com uma Abordagem Clássica . . . . .	19
2.5.5 Bondade de Ajuste e Previsão . . . . .	21
2.5.6 Tratamento da Endogeneidade . . . . .	22

2.6	Considerações Finais . . . . .	23
2.7	Apêndice: Exemplo de código usado em WinBUGS . . . . .	25
<b>Capítulo 3: Um Esquema de Amostragem Eficiente para Modelos Dinâmicos Generalizados</b>		<b>26</b>
3.1	Introdução . . . . .	26
3.2	Esquema de Amostragem Proposto . . . . .	28
3.3	Comparação de Esquemas de Amostragem . . . . .	33
3.3.1	Estudo de Monte Carlo . . . . .	33
3.3.2	Dados Reais: Precipitação em Tokyo . . . . .	39
3.3.3	Dados Reais: Chuva e Vazão no Rio Fartura (SP) . . . . .	43
3.4	Considerações Finais . . . . .	46
3.5	Apêndice: Algumas Equações para os Parâmetros a Priori e a Posteriori . . . . .	50
3.6	Apêndice: O Caso da Distribuição Gama . . . . .	51
<b>Capítulo 4: Uma Abordagem Bayesiana da Relação Chuva-Vazão</b>		<b>53</b>
4.1	Introdução . . . . .	53
4.2	Bacia do Rio Grande (BA), Brasil. . . . .	55
4.3	Abordagem Proposta . . . . .	56
4.3.1	Um Modelo Dinâmico de Função de Transferência . . . . .	56
4.3.2	Modelando a Chuva . . . . .	58
4.4	Procedimento de Inferência . . . . .	59
4.5	Modelagem na Prática . . . . .	61
4.5.1	Distribuições a Priori e Condicionais Completas . . . . .	62
4.5.2	Principais Aspectos Computacionais . . . . .	63
4.5.3	Resultados . . . . .	63
4.5.4	Interpolações Espaciais e Previsões Temporais . . . . .	70
4.6	Considerações Finais . . . . .	71
4.7	Apêndice: Verossimilhança - Mudança de Suporte . . . . .	72
4.8	Apêndice: Postos Fluviométricos e Pluviométricos . . . . .	74

<b>Capítulo 5:</b>	<b>Modelagem de Múltiplas Séries de Vazão</b>	<b>75</b>
5.1	Colocação do Problema . . . . .	75
5.2	Modelo conjunto proposto . . . . .	75
5.2.1	Especificação Condicional . . . . .	77
5.2.2	Efeito da Chuva: função de transferência . . . . .	77
5.3	Procedimento de Inferência . . . . .	79
5.4	Aplicação . . . . .	80
5.4.1	Modelando a Chuva . . . . .	80
5.4.2	Modelando a Vazão . . . . .	84
5.5	Considerações Finais . . . . .	90
5.6	Apêndice I: Algumas Estatísticas da Distribuição Log-Normal Multivariada . . . . .	91
<b>Capítulo 6:</b>	<b>Considerações Finais e Trabalhos Futuros</b>	<b>95</b>
6.1	Considerações Finais . . . . .	95
6.2	Trabalhos Futuros . . . . .	97
6.2.1	Sobre o Esquema Proposto para Modelos Generalizados Dinâmicos . . . . .	97
6.2.2	Sobre a Relação Chuva-Vazão . . . . .	101
<b>Apêndice A:</b>	<b>Abreviações e Definições</b>	<b>104</b>
<b>Apêndice B:</b>	<b>Distribuições Condicionais Completas e Métodos de Amostragem</b>	<b>109</b>
B.1	Modelo Linear Normal de Função de Transferência Não Estocástica . . . . .	109
B.2	Modelo Linear Normal de Função de Transferência Estocástica . . . . .	111
B.3	Modelo Linear Gama de Função de Transferência Não Estocástica . . . . .	115
B.3.1	A Distribuição Gama . . . . .	116
B.3.2	Condicionais Completas . . . . .	117
B.4	Modelo Linear Gama de Função de Transferência Estocástica . . . . .	120
B.4.1	Alternativa de Amostragem para $E_t, t = 1, \dots, T$ . . . . .	121
B.5	Modelo de Função de Transferência Não Estocástica com Estrutura Hierárquica . . . . .	124
<b>Apêndice C:</b>	<b>CUBS na Família Exponencial <math>K</math>-Paramétrica</b>	<b>128</b>

## LISTA DE TABELAS

2.1	Estatísticas a posteriori no modelo (2.15) com Erro I . . . . .	16
2.2	Estatísticas a posteriori no modelo (2.15) com Erro II . . . . .	18
2.3	Comparação de modelos: <i>DIC</i> e <i>EPD</i> . . . . .	20
2.4	Elasticidades de longo e curto prazo . . . . .	21
2.5	Estatísticas a posteriori no modelo (2.16) com Erro II e <i>AR(1)</i> para a renda . . . . .	23
3.1	RMSE, tempo computacional e taxas de aceitação no estudo de Monte Carlo . . . . .	38
3.2	RMSE <sub>Y</sub> e taxas de aceitação para a Chuva em Tokyo. . . . .	43
3.3	Estatísticas a posteriori de <i>W</i> para a chuva em Tokyo. . . . .	43
3.4	Comparação para a vazão no Rio Fartura . . . . .	45
3.5	PSRF e fatores de ineficiência para a Vazão no Rio Fartura . . . . .	49
3.6	Equações para calcular momentos a priori e a posteriori no <i>CUBS</i> . . . . .	50
4.1	Estatísticas a posteriori para o modelo de chuva . . . . .	65
4.2	Comparação de modelos: <i>DIC</i> , <i>EPD</i> , <i>MSE</i> , <i>MAE</i> . . . . .	68
4.3	Postos fluviométricos e pluviométricos . . . . .	74
5.1	Postos fluviométricos . . . . .	84
5.2	Estatísticas a posteriori do modelo conjunto . . . . .	86
5.3	Correlação no modelo conjunto . . . . .	87
5.4	Estatísticas a posteriori nos modelos univariados . . . . .	88
5.5	Comparação de estimativas médias: conjunto vs univariados . . . . .	89
5.6	Comparação de estimativas pontuais: conjunto vs univariados . . . . .	90
5.7	Comparação de modelos: conjunto vs univariados . . . . .	90

## LISTA DE FIGURAS

2.1	Dados de renda e consumo . . . . .	14
2.2	$k$ e $\lambda$ no o modelo (2.15) com Erro I. . . . .	17
2.3	$k$ e $\lambda$ no modelo (2.15) com Erro II . . . . .	19
2.4	Elasticidades de longo e curto prazo . . . . .	21
2.5	Valores replicados e previsão do consumo . . . . .	22
3.1	<i>Box plots</i> das ineficiências . . . . .	40
3.2	<i>Box plots</i> das medianas a posteriori de $W$ . . . . .	41
3.3	Autocorrelogramas para a Chuva em Tokyo. . . . .	42
3.4	<i>Box plots</i> das ineficiências e amostras de $W$ para a Chuva em Tokyo. . . . .	44
3.5	<i>Box plots</i> e autocorrelogramas para a Vazão no Rio Fartura . . . . .	47
3.6	Correlogramas cruzados para a Vazão no Rio Fartura . . . . .	48
4.1	Dados da bacia do Rio Grande . . . . .	55
4.2	Hipóteses nas funções de transferência . . . . .	57
4.3	Trajetória estimada para os parâmetros do modelo da chuva . . . . .	64
4.4	Médias a posteriori da chuva . . . . .	65
4.5	Chuva na bacia . . . . .	66
4.6	Modelo da vazão . . . . .	68
4.7	Função resposta-impulso da vazão . . . . .	69
4.8	Valores ajustados de vazão . . . . .	70
4.9	Valores previstos de chuva e vazão . . . . .	71
5.1	Hidrografia do Rio Grande . . . . .	76
5.2	Sub-bacias do Rio Grande . . . . .	81
5.3	Dados das sub-bacias do Rio Grande . . . . .	81
5.4	Chuva acumulada por sub-bacia . . . . .	83
5.5	Vazão no modelo condicional . . . . .	85

## Capítulo 1

### INTRODUÇÃO

A modelagem estocástica da relação entre chuva e vazão, duas das mais importantes variáveis hidrológicas, envolve alguns desafios estatísticos que motivaram a realização desta tese. Um deles é a modelagem conjunta de ambas variáveis utilizando modelos que representem adequadamente os processos físicos envolvidos. A classe dos modelos dinâmicos de funções de transferência foi escolhida para abordar esta relação. A inferência sobre os parâmetros dos modelos foi realizada sob o paradigma bayesiano e a ênfase foi dada à eficiência dos métodos de amostragem das distribuições a posteriori dos parâmetros de estado dos modelos dinâmicos.

Desta forma, esta tese teve como objetivos gerais: (i) propor um esquema de amostragem eficiente para fazer inferência em modelos dinâmicos não normais e não lineares, usando o enfoque bayesiano, e (ii) mostrar a sua aplicação em modelos de alta complexidade, em particular os modelos de função de transferência, aplicados à modelagem conjunta de chuva e vazão.

Esta introdução está organizada da seguinte forma. As seções 1.1 e 1.2 apresentam um resumo dos principais temas abordados nesta tese: modelos de função de transferência e a relação chuva-vazão. Na seção 1.3 são estabelecidos os objetivos gerais e específicos desta tese. A seção 1.4 descreve a organização dos outros capítulos desta tese.

#### 1.1 Modelos de Função de Transferência

Os modelos de função de transferência (FT) estão entre os mais utilizados na modelagem de séries temporais em diversas áreas como Engenharia, Economia, Meio Ambiente, etc. Seja  $\{Y_t, t = 1, 2, \dots\}$  a série de interesse (dependente, resposta ou *output*) e sejam  $\{X_{it}, t = 1, 2, \dots\}$   $k$  séries exógenas (*inputs*), onde  $i$  indexa a  $i$ -ésima série,  $i = 1, \dots, k$ . O modelo de FT (na representação polinomial) relaciona as séries através da seguinte equação:

$$Y_t = \sum_{i=1}^k E_{it} + v_t = \sum_{i=1}^k v_i(B)X_{it} + v_t, \quad (1.1)$$

onde  $E_{it}$  representa o efeito da FT da  $i$ -ésima variável exógena,  $v_i(B) = \sum_{j=-\infty}^{\infty} v_{ij}B^j$ ,  $B$  é o operador de defasagens, isto é,  $BY_t = Y_{t-1}$ , e  $v_t$  é uma série de ruídos independente de  $\{X_{it}\}$ . O termo  $v_t$  pode tomar diferentes especificações o que faz este tipo de modelagem muito flexível e versátil.

Nesta tese optamos por utilizar a representação de modelos de função de transferência apresentada em West & Harrison (1997, p. 281), mas para o caso mais geral em que a variável resposta,  $Y_t$ , segue uma distribuição na família exponencial; isto é,

$$Y_t \sim p(Y_t|\mu_t, \lambda_t) \quad (1.2a)$$

$$f(\mu_t) = \sum_{i=1}^k E_{it} \quad (1.2b)$$

$$E_{it} = g(E_{i,t-1}, \dots, E_{i0}, X_{it}, \dots, X_{i1}), \quad (1.2c)$$

onde  $p(Y_t|\mu_t, \lambda_t)$  denota uma distribuição normal, ou binomial, ou Poisson, ou gama, etc., e  $\mu_t$  denota a média de  $p(Y_t|\mu_t, \lambda_t)$ . A função  $f(\mu_t)$  é conhecida e, na linguagem de modelos generalizados, é denominada função de ligação. A função  $g(\cdot)$  também é conhecida e descreve os efeitos da covariável  $X_i$  sobre  $Y_t$ . O modelo em (1.2) é um modelo não-linear e inclui importantes casos particulares, por exemplo, os coeficientes da função  $g(\cdot)$  podem variar suavemente ao longo do tempo e/ou  $f(\mu_t)$  pode incluir um termo de erro aleatório.

Sob o ponto de vista bayesiano, o modelo em (1.2) fica completamente especificado depois da especificação da distribuição a priori dos parâmetros do modelo. A estimação de seus (hiper)parâmetros pode não ser uma tarefa fácil pois não é possível fazer inferência analítica. Alves (2005) estuda a inferência usando o método Monte Carlo via cadeias de Markov (MCMC), mostrando que esse método, embora computacionalmente intensivo, permite a obtenção de bons resultados. Neste sentido, um esquema que permita reduzir o tempo computacional e que seja de fácil implementação é um produto que pode auxiliar na modelagem de problemas cada vez mais complexos e/ou incentivar o uso destes modelos por parte de pesquisadores de outras áreas.

## 1.2 Relação Chuva-Vazão

Das fases do ciclo hidrológico, talvez a mais importante seja a do escoamento superficial que tem origem, fundamentalmente, nas precipitações. A vazão, ou volume escoado por unidade de tempo, é a principal grandeza que caracteriza um escoamento e normalmente é expressa em metros cúbicos por segundo. Sabe-se que a vazão do período atual depende da vazão do período anterior e do volume da chuva passada e atual, sendo modelos do tipo auto-regressivo (AR) ou média

móvel (*MA*), como os descritos em Sales (1989), comumente utilizados para a previsão de vazões. Contudo, estes tipos de modelos são ajustados sob as hipóteses de normalidade e estacionariedade. Além disso, a relação chuva-vazão é basicamente não linear. Nesse sentido, são necessários modelos estocásticos que descrevam essa relação e permitam o ajuste de séries não estacionárias com distribuição para variáveis reais positivas como, por exemplo, o modelo em (1.2).

Por outro lado, os modelos tradicionais de chuva-vazão consideram que a chuva é dada, isto é, que existe uma única medida de precipitação numa bacia. Porém, na maioria dos casos, a precipitação é medida em mais de uma estação meteorológica dentro de uma mesma bacia, portanto é necessário algum método para usar a informação de todos os postos e estimar a precipitação de toda a bacia. Existem métodos que propõem uso de polígonos para determinar a área de influência de cada posto na bacia e representar a precipitação total como uma média ponderada das medidas em cada posto. O principal problema com este tipo de método é que a incerteza associada com este processo não é considerada no processo de inferência.

Do descrito nesta seção nota-se a importância da modelagem conjunta da chuva e da vazão, considerando que a medição de ambas variáveis é realizada de forma diferente. Também, o modelo proposto deve ser capaz de fornecer boas previsões da quantidade d'água disponível, informação de vital importância para a gestão d'águas.

### 1.3 Objetivos da Tese

Com base na breve exposição realizada nas duas seções anteriores, formulamos como objetivos gerais desta tese: (1) propor um esquema de amostragem eficiente para obter amostras das distribuições a posteriori das quantidades desconhecidas nos modelos dinâmicos generalizados; (2) estudar os modelos de função de transferência dentro da classe dos modelos dinâmicos não normais, usando o enfoque bayesiano; e (3) aplicar o esquema proposto em modelos de alta complexidade com dados reais. Como objetivos específicos temos: (i) re-estimar os modelos de chuva-vazão estudados em Migon & Monteiro (1997), incluindo possíveis extensões dentro da especificação de modelos dinâmicos, (ii) implementar diversos algoritmos *MCMC* para amostrar da distribuição a posteriori e comparar os resultados obtidos, (iii) modelar dados de chuva considerando sua correlação espacial, e (iv) combinar a informação de vazões medidas em diferentes pontos de uma bacia ou de várias bacias num único modelo.

## 1.4 Organização da Tese

Esta tese está dividida em mais cinco capítulos. No capítulo 2 apresentamos os resultados de um exercício no qual foram explorados alguns modelos básicos de função de transferência, utilizando métodos de amostragem implementados no software `WinBUGS` (Spiegelhalter et al., 2003). No capítulo 3 apresentamos, em detalhe, o esquema de amostragem, via *MCMC*, proposto para a classe de modelos dinâmicos generalizados. Nos capítulos 4 e 5 apresentamos os resultados da modelagem proposta nesta tese para a relação chuva-vazão, aplicada a dados da bacia do Rio Grande (BA). No último capítulo apresentamos as conclusões e apontamos as possíveis extensões desta tese. Do segundo ao quinto, cada capítulo corresponde a um artigo preparado com resultados obtidos ao longo do doutorado. Alguns termos estatísticos e algumas das distribuições condicionais completas dos parâmetros envolvidos nos modelos estudados, necessárias para o procedimento de inferência via o método *MCMC* denominado amostrador de Gibbs, são apresentados como apêndices ao final deste documento.

## Capítulo 2

### INFERÊNCIA BAYESIANA EM MODELOS DE DEFASAGENS DISTRIBUÍDAS

O objetivo deste capítulo é mostrar a flexibilidade e a facilidade com que é possível implementar o paradigma bayesiano nos modelos de defasagens distribuídas quando considerados dentro da classe dos modelos dinâmicos. Os modelos de defasagens distribuídas são utilizados quando se acredita que, no tempo  $t$ , uma covariável, por exemplo  $X_t$ , causa um impacto sobre o valor médio de uma variável resposta, por exemplo,  $Y_t$ . Especificamente, quando se sabe que o efeito de  $X$  sobre  $Y$  persiste por um período e decai para zero ao longo do tempo. Na literatura existem vários modelos para lidar com este tipo de situação. Neste capítulo revisamos algumas dessas propostas e mostramos que sob uma re-parametrização relativamente simples, os modelos de defasagens distribuídas podem ser vistos como um caso particular dos modelos lineares dinâmicos (*DLM*), em particular, os modelos de função de transferência. Sendo assim, utilizamos inferência bayesiana e utilizamos o método Monte Carlo via cadeias de Markov (*MCMC*) para obter amostras das distribuições a posteriori de todos os parâmetros. A implementação desses métodos foi realizada no software *WinBugs* (Spiegelhalter et al., 2003).

Para ilustrar a abordagem proposta, analisamos uma função consumo usando a transformação de Koyck e um modelo de função de transferência. Comparamos os resultados com os obtidos mediante técnicas clássicas de co-integração.

#### 2.1 Introdução

As funções de transferência são freqüentemente utilizadas na área de econometria para que os coeficientes do modelo representem o comportamento de uma variável resposta  $Y$  sob o impacto de uma covariável  $X$ . Estas funções são necessárias quando a variável dependente responde a mudanças numa ou várias variáveis independentes somente depois de um período de tempo. Esta resposta retardada sugere a inclusão no modelo de variáveis explicativas defasadas (distribuindo o impacto da variável  $X$  em vários períodos de tempo), tendo como resultado um modelo dinâmico.

Um dos objetivos deste capítulo é ajustar modelos de função de transferência usando uma abordagem inteiramente bayesiana mediante o uso de métodos *MCMC*. Especificamente comparamos

diferentes modelos de defasagens distribuídas e mostramos que eles podem ser representados, depois de alguma re-parametrização, como uma classe particular dos modelos lineares dinâmicos.

De forma geral, um modelo é dinâmico toda vez que as variáveis estão indexadas pelo tempo e aparecem com diferentes defasagens. Por exemplo,  $Y_t = \beta_0 X_t + \beta_1 X_{t-1} + \epsilon_t$  é um modelo dinâmico simples conhecido como modelo de defasagens distribuídas. Nele a estrutura dinâmica aparece na variável exógena,  $X_t$ , porém também pode aparecer na variável endógena, por exemplo,  $Y_t = \alpha Y_{t-1} + \epsilon_t$ , que é o modelo conhecido como auto-regressivo de ordem 1,  $AR(1)$ . Os modelos com defasagens tanto nas variáveis exógenas como endógena são conhecidos como modelos auto-regressivos com defasagens distribuídas ( $ARDL$ ). Por outro lado, a estrutura dinâmica pode também aparecer no termo de erro ou nos parâmetros do modelo.

Os modelos de defasagens distribuídas têm um papel importante em numerosas aplicações na economia e na agricultura. Alguns dos primeiros exemplos incluem os estudos de Nerlove (1958) sobre a resposta da oferta agrícola ao preço, o estudo de apropriações de capital e despesas de Almon (1965) e a resposta do investimento de capital a vários aspectos do ambiente econômico (Koyck, 1954; Jorgenson, 1966). Mais recentemente, Frances & van Oest (2004) os utilizaram para estabelecer uma relação entre vendas e publicidade, enquanto Bentzen & Engsted (2001) estudaram equações de demanda de energia. Na verdade os modelos de defasagens distribuídas têm sido utilizados também nas áreas ambientais e epidemiológicas. Alguns exemplos de aplicações nessas áreas são apresentados em Huang et al. (2004) e Welty & Zeger (2004). Além disso, muitos livros texto de econometria incluem um capítulo sobre esta classe de modelos, entre eles Berndt (1991, Cap. 6), Greene (1999, Cap. 17), Gujarati (2000, Cap. 17) e Zellner (1971, Cap. 6).

Ao utilizar os modelos de defasagens distribuídas pode ser necessário assumir que o efeito dos valores passados diminui conforme a passagem do tempo. Essa forma funcional, denominada modelo de Koyck devido a Koyck (1954), comemorou o seu 50<sup>o</sup> aniversário em 2004 e, como mencionado por Frances & van Oest (2004), este modelo é um pouco mais complicado do que aparenta ser. Em primeiro lugar, a transformação de Koyck envolve uma restrição paramétrica que deve ser considerada por razões da sua eficiência e, em segundo lugar, a estatística  $t$  para o efeito da covariável não tem distribuição conhecida.

Além disso, os modelos de defasagens distribuídas são um caso particular dos  $ARDL$  e como Pesaran & Shin (1999) mostram que existe forte evidência a favor do uso dos tradicionais  $ARDL$  para a modelagem de séries não estacionárias, consideramos que é apropriado revisar a inferência dos modelos de defasagens distribuídas e compará-los com modelos de co-integração.

Este capítulo está organizado da seguinte maneira. A seção 2.2 apresenta alguns dos princi-

país modelos de defasagens distribuídas. A seção 2.3 apresenta a definição geral dos *DLM* e, em particular, as duas classes de modelos de função de transferência. A seção 2.4 discute o procedimento de inferência em *DLM*. Em particular, destacamos a importância dos métodos de simulação estocástica desenvolvidos nos últimos 15 anos. Na seção 2.5 apresentamos um exemplo que usa os dados analisados em Zellner & Geisel (1970), usando o modelo de Koyck e o modelo de função de transferência proposto. Finalmente, as conclusões e projetos futuros, relacionados com este capítulo, são apresentados na seção 2.6.

## 2.2 Modelos de Defasagens Distribuídas

A forma geral de um modelo de defasagens distribuídas *infinito* é

$$Y_t = \sum_{i=0}^{\infty} \beta_i X_{t-i} + \epsilon_t, \quad (2.1)$$

onde toda mudança em  $X_t$  afeta  $E[Y_t]$  em todos os períodos seguintes. O termo  $\beta_i$  em (2.1) é o  $i$ -ésimo coeficiente de reação. Usualmente assume-se que  $\lim_{i \rightarrow \infty} \beta_i = 0$  e  $\sum_{i=0}^{\infty} \beta_i = \beta < \infty$ . Supondo que as mudanças em  $X_t$  deixam de ter uma forte influência após alguns períodos de tempo, por exemplo,  $m$ , então o modelo se reduz a um modelo de defasagens distribuídas *finito*, e o limite superior do somatório em (2.1) é  $m$ .

Um aspecto importante a ser levado em conta é o número de parâmetros envolvidos neste tipo de modelo. Com a finalidade de ser parcimonioso, assume-se que os coeficientes das variáveis defasadas não são totalmente independentes e que estão relacionados mediante alguma função (Zellner, 1971). Existem diversas especificações para os modelos de defasagens infinitos e finitos. Algumas delas estão baseadas na teoria econômica, outras em hipóteses de natureza indutiva. Nas seções seguintes são apresentados os modelos de defasagens distribuídas mais conhecidos na literatura.

### 2.2.1 Defasagens Distribuídas Infinitas

Koyck (1954) propôs uma simplificação do modelo em (2.1). Assumiu que os  $\beta_i$ s decrescem exponencialmente com a passagem do tempo, isto é

$$\beta_i = \alpha \lambda^i, \quad \forall i, \text{ onde } 0 < \lambda < 1. \quad (2.2)$$

Logo,

$$Y_t = \alpha X_t + \alpha \lambda X_{t-1} + \alpha \lambda^2 X_{t-2} + \dots + \epsilon_t. \quad (2.3)$$

Um caso particular do modelo de defasagens distribuídas de Koyck (o geométrico) é o modelo econômico de ajustamento parcial. Por exemplo, seja  $Y_t$  o estoque de capital no instante  $t$  e  $Y_t^*$  o estoque de capital desejado no mesmo instante  $t$ . De acordo com o modelo de ajustamento parcial, a mudança no estoque de capital é proporcional à diferença entre o nível atual de capital desejado e o nível passado; isto é,

$$Y_t - Y_{t-1} = \lambda(Y_t^* - Y_{t-1}) + \epsilon_t, \quad \text{onde } 0 < \lambda < 1,$$

onde  $\epsilon_t$  é um erro aleatório. Se, adicionalmente, assume-se que o estoque de capital desejado é um múltiplo de  $X_t$ , isto é  $Y_t^* = \alpha X_t$ , então,

$$Y_t = (1 - \lambda)Y_{t-1} + \lambda\alpha X_t + \epsilon_t,$$

que é equivalente ao modelo de defasagens distribuídas:

$$Y_t = \alpha \sum_{i=0}^{\infty} \lambda(1 - \lambda)^i X_{t-i} + v_t,$$

onde  $v_t = (1 - \lambda)v_{t-1} + \epsilon_t$ . Este modelo mostra a dependência do estoque de capital atual nos níveis atual e passados de  $X_t$ .

Outro exemplo de modelo de defasagens infinitas é o modelo proposto por Solow (1960). A hipótese de Solow sobre os  $\beta_i$  é que eles são determinados por uma distribuição de Pascal, isto é,

$$\beta_i = \alpha \binom{r+i-1}{i} (1 - \lambda)^r \lambda^i, \quad 0 < \lambda < 1, \quad \forall i, r > 0. \quad (2.4)$$

Daí, se  $\alpha = 1$ , tem-se que

$$\beta_0 = (1 - \lambda)^r \quad \text{e} \quad \beta_i = \frac{r+i-1}{i} \lambda \beta_{i-1}.$$

O modelo de defasagens de Solow possui um esquema flexível de ponderação dos coeficientes  $\beta_i$ . Em particular, se  $r = 1$  tem-se o modelo de Koyck. Solow (1960) sugere que no contexto das defasagens é muito importante localizar a moda da distribuição dos  $\beta_i$ . Por esse motivo ele propôs o uso da distribuição Pascal, onde a moda é sempre menor que a média, portanto é uma distribuição assimétrica à direita. Quanto maior é  $\lambda$  e menor é  $r$ , maior é a assimetria dessa distribuição.

Um modelo mais geral, conhecido como Defasagens Racionais, foi apresentado por Jorgenson (1966). Ele considera que

$$Y_t = \mu + \lambda(L)X_t + V_t, \quad (2.5)$$

onde  $\lambda(L)$  é um polinômio infinito. Recentemente Carter & Zellner (2002) propuseram o seguinte modelo:

$$\phi(L)Y_t = \mu + \delta(L)X_t + U_t \quad (2.6a)$$

$$Y_t = \frac{\mu}{\phi(L)} + \frac{\delta(L)}{\phi(L)}X_t + \frac{1}{\phi(L)}U_t, \quad (2.6b)$$

cuja característica principal é a de manter a mesma estrutura de defasagens infinita de (2.5),  $\lambda(L) = \frac{\delta(L)}{\phi(L)}$ , mas com um processo mais geral para  $V_t$ . Carter & Zellner (2002) denominaram (2.6) como modelo *ARDLAR*.

## 2.2.2 Defasagens Distribuídas Finitas

Almon (1965) propôs uma aproximação conhecida como a distribuição de interpolação para os coeficientes de um modelo de defasagens finito. A hipótese de Almon é que os  $\beta_i$  são bem aproximados por polinômios de grau  $p < m$ , com  $p + 1$  parâmetros, isto é:

$$\beta_i = \sum_{k=0}^p \alpha_k i^k. \quad (2.7)$$

Esta aproximação polinomial fornece uma ampla variedade de formas para  $\beta_i$ , porém, na prática somente polinômios de grau baixo são utilizados.

Mais recentemente e no contexto bayesiano, Chotikapanich & Griffiths (1999) propuseram o modelo de defasagens distribuídas flexível que coloca menos estrutura nos pesos e representa as restrições dos coeficientes nas suas distribuições a priori.

## 2.3 Modelos Lineares Dinâmicos

### 2.3.1 Definição

Os modelos lineares dinâmicos (*DLM*) (West & Harrison, 1997) são uma classe ampla de modelos com parâmetros variando no tempo, muito úteis para modelar dados de séries temporais e problemas de regressão com parâmetros variando ao longo do tempo. Uma representação geral de um *DLM* é:

$$Y_t = F_t' \theta_t + \epsilon_t \quad \epsilon_t \sim N(0, V_t) \quad (2.8a)$$

$$\theta_t = G_t \theta_{t-1} + \omega_t \quad \omega_t \sim N(0, W_t), \quad (2.8b)$$

onde, para todo  $t$ ,  $Y_t$  é um vetor  $l \times 1$ ,  $F_t$  é uma matriz  $n \times l$  conhecida,  $G_t$  é uma matriz  $n \times n$  conhecida,  $\theta_t$  é o vetor de estados e tem dimensão  $n$ ,  $\epsilon_t$  é o termo de erro da observação e  $\omega_t$  é o termo de erro do sistema ou erro de evolução. Frequentemente  $\epsilon_t$  e  $\omega_t$  são interna e mutuamente independentes. A equação (2.8a) é conhecida como equação da observação e define a distribuição de  $Y_t$  condicional a  $\theta_t$ . A equação (2.8b) é conhecida como equação do sistema e define a evolução do vetor de estados. Neste capítulo estamos interessados na modelagem de séries de tempo univariadas, neste caso o modelo acima se reduz para o caso em que  $l = 1$ .

Os modelos de defasagens distribuídas podem ser expressos na forma de (2.8). O efeito de qualquer regressor no instante  $t$  sobre a média da resposta nos instantes futuros pode ser expresso como uma função de transferência. Na análise de funções de transferência, o objetivo é estudar como a trajetória de uma variável endógena é afetada pelos movimentos ou mudanças de uma variável exógena. Se esse efeito não tem uma forma estrutural ao longo do tempo, é dito que se trata de uma função de transferência de forma livre. Entretanto, se os efeitos (ou coeficientes do modelo de regressão) estão relacionados através de uma forma conhecida, é dito que se trata de uma função de transferência de forma funcional. Ambos casos são descritos na seção seguinte.

### 2.3.2 Modelos de Função de Transferência

Uma função de transferência de forma livre é um *DLM* de regressão sobre um número finito de valores defasados. De acordo com West & Harrison (1997, p. 281), se a resposta média no instante  $t$  é definida por

$$E(Y_t | \theta_t) = \sum_{i=0}^m \beta_i X_{t-i} = \beta_0 X_t + \beta_1 X_{t-1} + \dots + \beta_m X_{t-m}, \quad (2.9)$$

a matriz  $F_t$  em (2.8) é dada por  $F'_t = (X_t, X_{t-1}, \dots, X_{t-m})$  e  $\theta'_t = \theta' = (\beta_0, \beta_1, \dots, \beta_m)$ . Em (2.9), o efeito do valor corrente da regressora  $X_t = X$  sobre a resposta média no instante futuro  $i$ , com a condição  $X_{t+1} = \dots = X_{t+i} = 0$ , define a função

$$\begin{cases} \beta_i X & i = 0, 1, \dots, m; \\ 0 & i > m. \end{cases}$$

Neste caso, os coeficientes de regressão  $\beta_i$  não estão relacionados (entre eles) e tem-se, pelo menos,  $m$  quantidades desconhecidas. Uma maneira simples de adaptar a estrutura de regressão para incorporar uma relação funcional entre os  $\beta_i$  é considerar uma regressão, não diretamente em  $X$ , mas numa variável de efeito construída para medir o efeito combinado dos valores passados e corrente de  $X$ . Seja  $X_t$  o valor no instante  $t$  de uma variável escalar independente  $X$ . Como descrito

em West & Harrison (1997), um modelo de função de transferência geral para o efeito de  $X$  sobre  $Y$  é dado por

$$Y_t = F' \theta_t + \epsilon_t, \quad \epsilon_t \sim N(0, \sigma_\epsilon^2) \quad (2.10a)$$

$$\theta_t = G \theta_{t-1} + \psi_t X_t + \partial \theta_t \quad (2.10b)$$

$$\psi_t = \psi_{t-1} + \partial \psi_t, \quad (2.10c)$$

cujos termos foram definidos na seção 2.3.1. Em particular,  $\epsilon_t$  e  $\partial \theta_t$  são os termos de erro da observação e da evolução, respectivamente, e  $\psi_t$  é um vetor, de dimensão  $n$ , de parâmetros que evoluem com a adição do termo  $\partial \psi_t$ , que segue uma distribuição normal com média zero e é independente de  $\epsilon_t$  e  $\partial \theta_t$ .

O vetor de estados carrega o efeito dos valores passados e corrente de  $X$  para  $Y_t$  através da equação (2.10a); este efeito é formado em (2.10b) com a soma de uma função linear dos efeitos passados  $\theta_{t-1}$ , o efeito corrente  $\psi_t X_t$  e um termo de ruído. O modelo em (2.10) pode ser re-escrito na forma padrão dos *DLM* (West & Harrison, 1997, Cap. 9). A parametrização acima é muito mais flexível que (2.1), permitindo diferentes interpretações estocásticas para o efeito de  $X_t$  sobre  $Y_t$ .

### 2.3.3 Funções de Transferência e Defasagens Distribuídas

Seguindo as descrições acima apresentadas, nesta seção mostramos que os modelos de defasagens distribuídas da seção 2.2 podem ser representados como um *DLM*, na forma da equação (2.10). Por exemplo, o modelo de Koyck apresentado em (2.3) pode ser re-escrito como

$$Y_t = E_t + \epsilon_t \quad (2.11a)$$

$$E_t = \lambda E_{t-1} + \alpha X_t, \quad (2.11b)$$

onde  $E_t = \alpha X_t + \alpha \lambda X_{t-1} + \alpha \lambda^2 X_{t-2} + \dots$  e  $0 < \lambda < 1$ . Neste modelo, a função de transferência de  $X$  é  $\alpha \lambda^i X$ . Aqui, com base na parametrização em (2.10), temos  $n = 1$ ,  $\theta_t = E_t$ , o efeito total,  $\psi_t = \alpha$ , o efeito corrente para todo  $t$ ,  $F = 1$ ,  $G = \lambda$  e  $\partial \theta_t$  é igual a zero. Repare que este modelo pode ser facilmente estendido para o caso em que  $\alpha$  varia suavemente ao longo do tempo, isto é,  $\alpha_t = \alpha_{t-1} + \partial \alpha_t$  como em (2.10c) ou, mais geralmente, pode ser incluído um ruído para  $E_t$  em (2.11b). Por exemplo,  $E_t = \lambda E_{t-1} + \alpha_t X_t + \partial E_t$ , onde  $\partial E_t$  segue uma distribuição normal com média zero e variância  $\sigma_E^2$ . Note que a representação como *DLM* resulta numa estrutura muito geral para o termo de erro. Assim, assumindo independência entre os termos de erro da observação e da evolução, ao substituir (2.10b) em (2.10a) obtém-se um processo *ARMA*.

O modelo de defasagens distribuídas de Solow também pode ser escrito como um *DLM*. Neste caso, temos  $(1 - \lambda L)^r E_t = \alpha(1 - \lambda)^r X_t$  e uma equação de evolução pode ser considerada para  $\lambda$  e  $r$ . Também, o modelo de Almon é um caso particular de uma função de transferência de forma livre (2.9) já que é uma regressão num número finito e fixo de variáveis defasadas. Usando (2.7), a função de resposta de  $Y_t$  é

$$\begin{cases} \sum_{k=0}^p \alpha_k i^k X_{t-i} & i = 0, 1, \dots, m; \\ 0 & i > m. \end{cases}$$

Um exemplo econométrico é apresentado em Migon (2000), onde é desenvolvido um modelo para prever o valor das exportações do Brasil como uma função de uma tendência estocástica e da taxa de câmbio real. A natureza da dinâmica do modelo foi introduzida através da hipótese de expectativas adaptativas. O modelo resultante foi uma função de transferência de primeira ordem mais uma tendência.

## 2.4 Procedimento de Inferência

De acordo com o paradigma bayesiano, a especificação de um modelo fica completa depois da determinação da distribuição a priori de todos os seus parâmetros. De acordo com o teorema de Bayes, a distribuição a posteriori é proporcional ao produto dessa distribuição a priori vezes a verossimilhança.

No contexto dos *DLM*, a inferência pode ser realizada em forma seqüencial, como é descrito em West & Harrison (1997, Cap 4). Entretanto, nos modelos de função de transferência descritos na seção 2.3, a matriz de evolução  $G$  (em (2.8b)) pode depender de quantidades desconhecidas, como em (2.11). Por esta razão é necessário fazer inferência sobre  $\lambda$  simultaneamente e, portanto, outros algoritmos são necessários.

Até o final da década de 80, técnicas de integração numérica eram utilizadas para aproximar as constantes de normalização das distribuições a posteriori e obter estimativas para os parâmetros. As distribuições a posteriori dos modelos de defasagens distribuídas são difíceis de serem obtidas analiticamente. No exemplo em Zellner & Geisel (1970) foi utilizada a transformação de Koyck, que implica o ajuste da seguinte equação

$$Y_t = \lambda Y_{t-1} + k(1 - \lambda)X_t + \epsilon_t - \lambda \epsilon_{t-1}, \quad (2.12)$$

onde  $\lambda, k$  e  $\sigma_\epsilon^2$  são os parâmetros de interesse. Devido à complexidade das distribuições a posteriori marginais, eles tiveram que fazer uso de técnicas de integração numérica bivariada para obter

amostras das distribuições de interesse. Contudo, dado que as quantidades desconhecidas são três, os autores tiveram que integrar um dos parâmetros a fim de cair no caso bivariado.

Os métodos *MCMC*, como o amostrador de Gibbs (Gelfand & Smith, 1990), têm sido muito utilizados na inferência bayesiana desde o início dos anos 90. O software *BUGS* (*Bayesian Analysis using Gibbs Sampling*) é uma ferramenta muito conhecida que nos auxilia com esta tarefa. Este pacote foi desenvolvido por D. Spiegelhalter e alguns colegas do *MRC Biostatistics Unit* e está disponível de forma gratuita no sítio <http://www.mrc-bsu.cam.ac.uk/bugs>.

Neste capítulo mostramos que a inferência sobre os *DLM*, em particular os modelos de função de transferência, pode ser realizada via *BUGS*. O maior atrativo aí é a possibilidade de testar facilmente várias especificações de modelo, tais como estruturas auto-regressivas ou transições de estados polinomiais, assim como avaliar a sensibilidade à escolha da distribuição a priori.

O *BUGS* executa a estimação da distribuição a posteriori dos *DLM* de forma razoável. Contudo, um importante fato a ser levado em conta é que a convergência das cadeias pode ser muito lenta devido à alta correlação entre os elementos de  $\theta$ . Uma solução prática para este caso é guardar amostras a cada  $k$  iterações, utilizando um valor grande para  $k$ . Em resumo, *BUGS* é muito útil para explorar novos modelos, porém rotinas computacionais mais específicas podem ser necessárias para aplicações detalhadas e eficientes (Meyer & Yu, 2000).

## 2.5 Aplicação: uma Função Consumo

### 2.5.1 Dados e Modelos

De forma similar a Zellner & Geisel (1970), usamos o conjunto de dados apresentado em Griliches et al. (1962, pp. 499-500), que corresponde à renda disponível ( $X_t$ ) e ao gasto ( $Y_t$ ) per-capita trimestral dessazonalizados nos Estados Unidos no período 1947.I-1960.IV. A figura 4.1(b) mostra ambas séries temporais. Seguindo a análise de Zellner & Geisel (1970) consideramos a seguinte função consumo:

$$Y_t = kX_t^* + \epsilon_t, \quad (2.13)$$

onde, para o  $t$ -ésimo período,  $t = 1, 2, \dots, T$ ,  $Y_t$  é o consumo real observado,  $X_t^*$  é a renda real “normal”,  $k$  é desconhecido e  $\epsilon_t$  é um ruído ou consumo transitório. Como  $X_t^*$  não é observável, Zellner & Geisel (1970) assumiram que a renda real “normal” satisfaz a hipótese de expectativas adaptativas:

$$X_t^* - X_{t-1}^* = (1 - \lambda)(X_t - X_{t-1}^*), \quad (2.14)$$

onde o parâmetro  $\lambda$  é tal que  $0 < \lambda < 1$ . Combinando (2.13) e (2.14), obtemos

$$Y_t = k(1 - \lambda)(X_t + \lambda X_{t-1} + \lambda^2 X_{t-2} + \dots + \lambda^n X_{t-n} + \dots) + \epsilon_t, \quad (2.15)$$

ou, de forma equivalente,

$$Y_t = E_t + \epsilon_t, \quad t \geq 2 \quad (2.16a)$$

$$E_t = \lambda E_{t-1} + \psi X_t, \quad (2.16b)$$

onde  $\psi = k(1 - \lambda)$  e  $E_t = \psi X_t + \lambda \psi X_{t-1} + \lambda^2 \psi X_{t-2} + \dots$

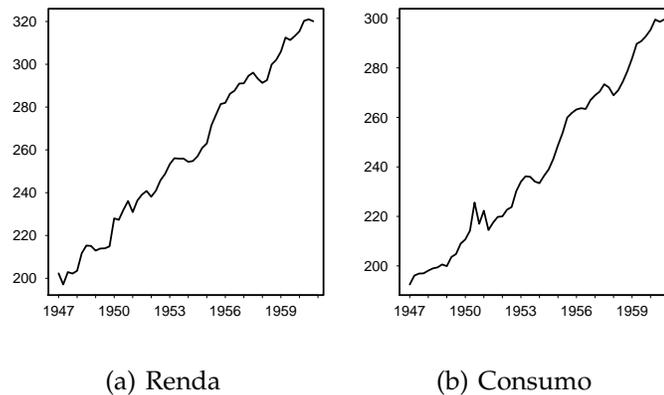


Figura 2.1: **Renda disponível ( $X_t$ ) e Gasto ( $Y_t$ ) per-capita trimestral dessazonalizados nos Estados Unidos no período 1947.I-1960.IV.**

A partir daqui desenvolvemos a análise de (2.15) assumindo que  $X_t$  é uma variável predeterminada, tal como na definição dada em Lancaster (2004, p.345):  $X$  é um vetor de variáveis predeterminadas que são dependentes de valores passados do vetor de erro mas não entre valores contemporâneos. Esta hipótese nos permite trabalhar com uma única equação; contudo, na seção 2.5.6 fazemos uso do método de variáveis instrumentais para eliminar a correlação entre a regressora e os erros.

Nosso objetivo é usar a abordagem bayesiana para fazer inferência no modelo em (2.15) sob as seguintes hipóteses de correlação serial do termo de erro  $\epsilon_t$ :

- Erro I:  $\epsilon_t \stackrel{\text{ind}}{\sim} N(0, \sigma_I^2)$ ;
- Erro II:  $\epsilon_t = \rho \epsilon_{t-1} + v_t$ , with  $v_t \stackrel{\text{ind}}{\sim} N(0, \sigma_{II}^2)$ .

Segundo o Erro I, os  $\epsilon_t$ 's são independentes e normalmente distribuídos com média zero e variância comum  $\sigma_I^2$ . Segundo o Erro II existe auto-correlação entre os  $\epsilon_t$ 's. Neste caso assumimos que  $\epsilon_t$  segue um processo auto-regressivo de primeira ordem com parâmetro desconhecido  $\rho$  e variância  $\sigma_{II}^2$ .

Sob a abordagem bayesiana, os modelos acima ficam completamente especificados após a atribuição da distribuição a priori dos parâmetros. Para  $\lambda$  e  $k$  usamos uma distribuição beta com parâmetros  $(1, 1)$ , isto é  $\lambda \sim \text{Beta}(1, 1)$  e  $k \sim \text{Beta}(1, 1)$ . Em outras palavras, assumimos que  $\lambda$  e  $k$  são independentes e estão uniformemente distribuídos entre zero e um, *a priori*. Dado que não temos informação sobre a variância do termo de erro, adotamos uma distribuição gama invertida pouco informativa como priori, isto é  $\sigma^2 \sim \text{IG}(0, 0001; 0, 0001)$ . Testamos também outras distribuições do tipo  $\text{IG}(\epsilon, \epsilon)$  mas os resultados não foram influenciados por esta escolha. Para o modelo II adicionamos uma priori para  $\rho$ ,  $\rho \sim \text{Normal}(0, 100)$ , isto é, utilizamos uma priori relativamente plana (*flat*) sem a restrição de que  $\rho$  deva estar na região de estacionariedade  $[-1, 1]$ . Em outras palavras, permitimos que os dados determinem a região com maior massa de probabilidade a posteriori. Além disso, assumimos que no primeiro instante de tempo,  $E_1 = y_1$ , pois  $E_t$  é o valor esperado de  $Y_t$  em (2.16). Na realidade esta não é mais do que uma aproximação. Em forma mais geral, poderíamos ter atribuído uma distribuição a priori para  $E_0$ , por exemplo,  $p(E_0)$ , o efeito anterior as observações, e utilizado a equação recursiva para obter uma distribuição a priori para  $E_1$ ,  $p(E_1|E_0, \lambda, \phi)$ .

## 2.5.2 Resultados

Para obter as principais estatísticas da distribuição a posteriori de interesse, usamos o software WinBUGS versão 1.4 (Spiegelhalter et al., 2003). No apêndice 2.7 apresentamos a rotina utilizada no WinBUGS para o modelo de Koyck. Para os modelos com estrutura de erros I e II, geramos duas cadeias paralelas começando de valores diferentes. Iteramos as cadeias 10000 vezes, eliminamos as primeiras 5000 iterações como período de aquecimento e guardamos as amostras a cada 10 iterações. Para verificar a convergência das cadeias, usamos a estatística de Gelman & Rubin (1992), modificada por Brooks & Gelman (1998). Finalmente, comparamos nossos resultados com os obtidos por Zellner & Geisel (1970), que usaram a transformação de Koyck (KT), apresentada em (2.12), e integração numérica para aproximar a distribuição a posteriori de interesse.

A média, moda e desvio padrão a posteriori de  $k$  e  $\lambda$  com o modelo com erro I são mostrados nas segunda e terceira linhas da tabela 2.1. Nela observamos que as duas abordagens fornecem resultados similares aos obtidos por Zellner & Geisel (1970). A figura 2.2 mostra as densidades a pos-

teriori de  $k$  e  $\lambda$  obtidas com a transformação de Koyck (painéis 2.2(a) e 2.2(b)) e com a representação de função de transferência (painéis 2.2(c) e 2.2(d)), respectivamente. Nesta figura fica evidente a similaridade dos resultados. Observe que as duas modas das distribuições a posteriori são bem identificadas.

Tabela 2.1: Estatísticas a posteriori associadas com os parâmetros no modelo (2.15) com Erro I

Parâmetro	média	d.p.	2,5%	50,0%	97,5%	$\hat{R}^b$
Publicado em Zellner & Geisel (1970) <sup>a</sup>						
$k$	0,948	0,020		0,940 - 1,000		–
$\lambda$	0,508	0,254		0,380 - 0,900		–
Usando a transformação de Koyck's no WinBUGS						
$k$	0,942	0,016	0,930	0,936	0,995	1,037
$\lambda$	0,419	0,219	0,074	0,385	0,895	1,019
Usando a função de transferência no WinBUGS						
$k$	0,942	0,017	0,930	0,936	0,996	1,351
$\lambda$	0,413	0,222	0,063	0,372	0,895	1,051

<sup>a</sup> os valores na coluna 50% correspondem às modas.

<sup>b</sup>  $\hat{R}$  é o fator de redução de escala (na convergência,  $\hat{R} = 1$ ).

A tabela 2.2 mostra estatísticas a posteriori de alguns parâmetros do modelo (2.15) com Erro II:  $k$ ,  $\lambda$  e  $\rho$ . Nessa tabela observamos uma vez mais que as duas representações, a transformação de Koyck e a função de transferência, são equivalentes e fornecem a mesma moda para  $k$ , porém as amostras das distribuições a posteriori obtidas com a representação de função de transferência estão mais concentradas em torno de suas médias. Contudo, os resultados para  $\lambda$  são muito diferentes, provavelmente porque Zellner & Geisel (1970) marginalizaram a distribuição a posteriori em  $\rho$ , enquanto, nós fazemos inferência de forma conjunta. Amostras das densidades a posteriori de  $k$ ,  $\lambda$  e  $\rho$  obtidas sob as duas representações paramétricas do modelo em (2.15) com Erro II (KT, TF) são apresentadas na figura 2.3. Fica evidente que ambas representações são equivalentes.

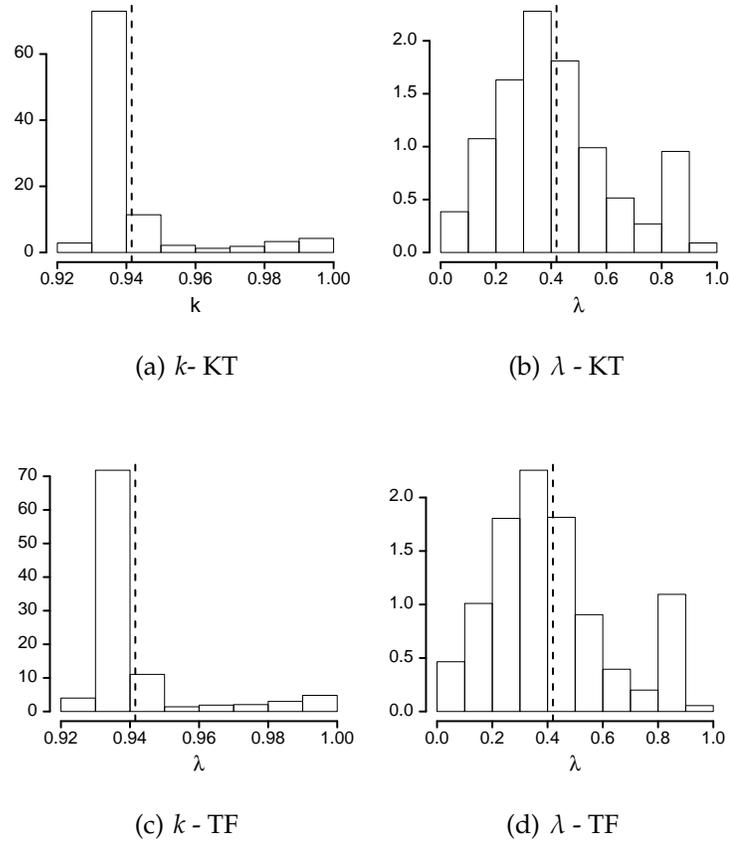


Figura 2.2: **Amostras a posteriori dos parâmetros  $k$  e  $\lambda$  sob o modelo (2.15) com Erro I.** A linhas pontilhadas correspondem à média a posteriori. (KT = usando a transformação de Koyck, TF usando a parametrização como função de transferência).

### 2.5.3 Seleção de Modelos

Dado que as estimativas sobre  $\lambda$  são sensíveis às hipóteses utilizadas para o termo de erro, é necessário selecionar o “melhor” modelo entre os que foram ajustados na seção anterior. Neste capítulo usamos dois critérios de seleção de modelos diferentes. O primeiro é o *DIC*, abreviação do inglês *Deviance Information Criterion* (Spiegelhalter et al., 2001). O *DIC* é dado por  $DIC = \bar{D} + p_D = D(\bar{\theta}) + 2p_D$ , onde  $\bar{D}$  é a média a posteriori da função desvio (*deviance*) dada por  $D = -2\log[f(y|\theta)]$ , onde  $f(y|\theta)$  é a verossimilhança condicionada avaliada nos valores amostrados dos parâmetros.  $D(\bar{\theta})$  é obtido ao substituir a média a posteriori de  $\theta$  na função desvio, de forma que  $D(\bar{\theta}) =$

Tabela 2.2: Estatísticas a posteriori associadas com os parâmetros no modelo (2.15) com Erro II

Parâmetro	média	d.p.	2,5%	50,0%	97,5%	$\hat{R}^b$
Publicado em Zellner & Geisel (1970) <sup>a</sup>						
$k$	0,878	0,201		0,940		–
$\lambda$	0,597	0,184		0,610		–
Usando a transformação de Koyck's no WinBUGS						
$k$	0,958	0,014	0,938	0,954	0,992	1,002
$\lambda$	0,754	0,087	0,562	0,764	0,893	1,001
$\rho$	0,790	0,141	0,454	0,811	0,991	1,002
Usando a função de transferência no WinBUGS						
$k$	0,959	0,018	0,933	0,956	0,998	1,001
$\lambda$	0,737	0,125	0,462	0,762	0,899	1,001
$\rho$	0,751	0,106	0,548	0,750	0,955	1,001

<sup>a</sup> os valores na coluna 50% correspondem às modas.

<sup>b</sup>  $\hat{R}$  é o fator de redução de escala (na convergência,  $\hat{R} = 1$ ).

$-2 \log[f(y|\bar{\theta})]$ , e  $p_D = \bar{D} - D(\bar{\theta})$ . A versão 1.4 do WinBUGS calcula o *DIC* de forma automática. O modelo com o menor *DIC* deve ser selecionado.

O segundo critério usado para comparar os modelos foi proposto por Gelfand & Ghosh (1998) e é denominado *EPD*, abreviação do inglês *Expected Predictive Deviance*. O *EPD* é obtido minimizando a perda a posteriori de um dado modelo. Quando se considera uma função de perda quadrática, o *EPD* pode ser calculado explicitamente. Nesse caso, se  $Y_{i,rep}$  representa valores replicados do  $i$ -ésimo dado observado,  $y_{i,obs}$ , temos  $EPD = \sum_{i=1}^n \sigma_i^2 + \frac{\kappa}{\kappa+1} \sum_{i=1}^n (\mu_i - y_{i,obs})^2$ , onde  $\mu_i = E[Y_{i,rep}|y_{i,obs}]$  e  $\sigma_i^2 = Var[Y_{i,rep}|y_{i,obs}]$ , são a média e a variância da distribuição preditiva, respectivamente. O modelo que minimiza o *EPD* é selecionado.

A tabela 2.3 mostra os valores do *DIC* e *EPD*, considerando uma perda quadrática, obtida para cada modelo. Observamos que os valores calculados não diferem muito sob as duas representações. Notamos também que os menores *DIC* e *EPD* foram obtidos com o modelo de Erro II, sugerindo que é mais apropriado assumir que os termos de erro  $\epsilon_t$  em (2.13) estão auto-correlacionados. Maior atenção deve ser prestada a este modelo, por exemplo deve-se pesquisar sua sensibilidade

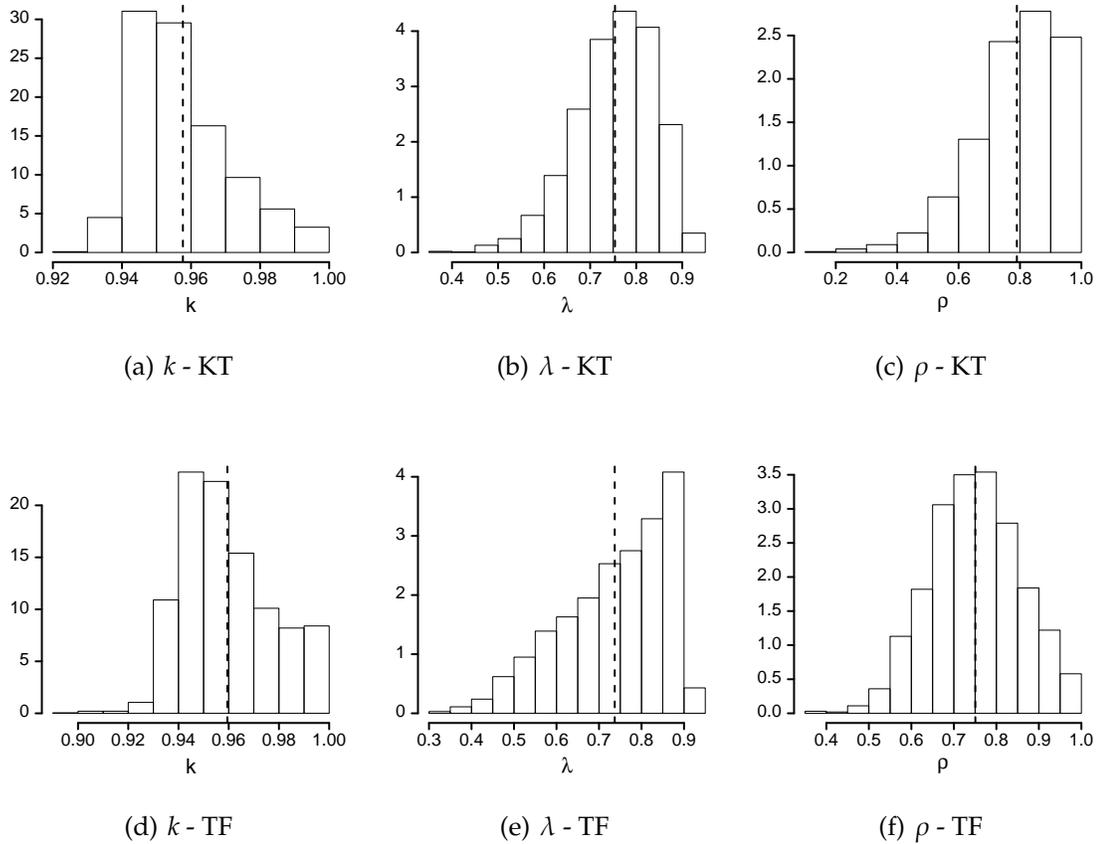


Figura 2.3: Amostras a posteriori dos parâmetros  $k$  e  $\lambda$  sob o modelo (2.15) com Erro II. A linhas pontilhadas correspondem à média a posteriori. (KT = usando a transformação de Koyck, TF usando a parametrização como função de transferência).

à escolha da distribuição a priori e testar a hipótese  $\rho = \lambda$ .

#### 2.5.4 Comparação com uma Abordagem Clássica

Com a finalidade de comparar os resultados das seções anteriores com os de uma abordagem clássica, seguimos uma estratégia bastante utilizada para trabalhar com séries de tempo não estacionárias. Especificamente seguimos Pesaran & Shin (1999) e Bentzen & Engsted (2001) que mostraram que a aproximação com modelos *ARDL* é perfeitamente válida quando (algumas das) as variáveis são não-estacionárias integradas de ordem 1,  $I(1)$ , com a única restrição de que as

Tabela 2.3: Critérios de Comparação de Modelos: *DIC* (*Deviance Information Criteria*) e *EPD* (*Expected Predictive Deviance*)

		Transformação de Koyck (KT)	Função de Transferência (FT)
Modelo (2.15)	<i>DIC</i>	312,49	311,40
com Erro I	<i>EPD</i>	1455,57	1453,29
Modelo (2.15)	<i>DIC</i>	280,37	279,20
com Erro II	<i>EPD</i>	760,16	767,91

variáveis  $I(1)$  sejam cointegradas e exista apenas uma equação de co-integração.

Usando a transformação de Koyck, o modelo em (2.15) pode ser visto como um modelo  $ARDL(1, 0)$ , assim  $k$  e  $k(1 - \lambda)$  podem ser interpretados como elasticidades de longo e curto prazos, respectivamente. No contexto bayesiano, a inferência sobre funções dos parâmetros pode ser realizada, de forma direta, a partir das amostras obtidas das distribuições a posteriori dos mesmos. Neste caso, basta usar as amostras do *MCMC* de  $k$  e  $\lambda$  para ter uma amostra a posteriori de  $k(1 - \lambda)$ . A primeira linha da tabela 2.4 mostra as elasticidades de longo prazo estimadas com o modelo selecionado na seção 2.5.3 (com Erro II). Ela também apresenta a elasticidade de longo prazo obtida com a representação de função de transferência (com Erro II) incluindo um termo constante (intercepto) no modelo (segunda linha), com o procedimento de mínimos quadrados ordinários (*OLS*)<sup>1</sup> e com o método de co-integração de Johansen<sup>2</sup>. Observamos que as estimativas pontuais para o parâmetro de longo prazo obtidas com todos os modelos são bastante similares. A figura 2.4(a) apresenta o histograma da amostra da posteriori de  $k$  obtida com o *MCMC* e a distribuição normal obtida para  $\hat{k}$  com o procedimento *OLS*. A figura 2.4(b) mostra as distribuições empírica e assintótica obtidas para  $k(1 - \lambda)$ . Embora as estimativas pontuais para  $k$  sob ambos procedimentos sejam similares, na figura observamos que a distribuição assintótica é muito diferente da distribuição a posteriori. Notamos que na distribuição assintótica  $P(\hat{k}(H) \geq 0.905) \geq 0.50$  enquanto que na distribuição a posteriori,  $P(k \geq 0.905|H) = 0.025$ , onde  $H$  representa a base de dados completa. Este fato pode ser explicado porque o último resultado é baseado unicamente na amostra observada.

<sup>1</sup>sugerido por Engle & Granger (1987)

<sup>2</sup>sugerido por Johansen (1991)

Tabela 2.4: Elasticidades de longo e curto prazo

Model	Longo prazo		Curto prazo	
	média	d.p.	média	d.p.
Função de Transferência <sup>a</sup>	0,959	0,018	0,251	0,116
Função de Transferência <sup>b</sup>	0,910	0,042	0,262	0,108
<i>OLS</i>	0,878	0,014	0,266	0,114
Johansen ( <i>VECM</i> ) <sup>c</sup>	0,884	0,022	–	–

<sup>a</sup> Modelo em (2.16).

<sup>b</sup> Modelo em (2.16) incluindo um termo constante ou intercepto.

<sup>c</sup> Considerando uma tendência linear nos dados e 0 defasagens no VAR, este modelo foi escolhido seguindo o critério de Schwartz

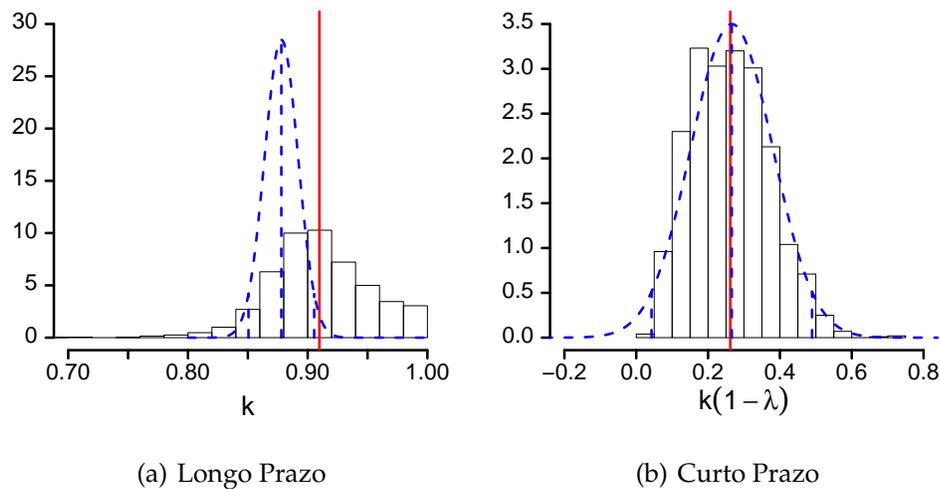


Figura 2.4: Elasticidades de longo e curto prazo estimadas com o modelo em (2.16) com Erro II e um intercepto na equação das observações. Em ambos painéis, os histogramas correspondem à amostra da distribuição a posteriori, a linha vertical corresponde à média a posteriori e as linhas pontilhadas a distribuição normal assintótica estimada via *OLS*.

### 2.5.5 Bondade de Ajuste e Previsão

Com o objetivo de examinar a bondade de ajuste e a capacidade preditiva de nossa abordagem, reestimamos os parâmetros do modelo em (2.16) com Erro II, incluindo um intercepto na primeira

equação, utilizando somente as primeiras 49 observações. A figura 2.5(a) mostra o intervalo de 95% de credibilidade estimado para a amostra entre 1947.I and 1959.II. Observamos um bom ajuste, pois somente duas das observações ficaram fora do intervalo obtido. A figura 2.5(b) apresenta o intervalo de 95% de credibilidade estimado para as observações deixadas fora da amostra. Observamos que todos os valores reais estão contidos no intervalo estimado.

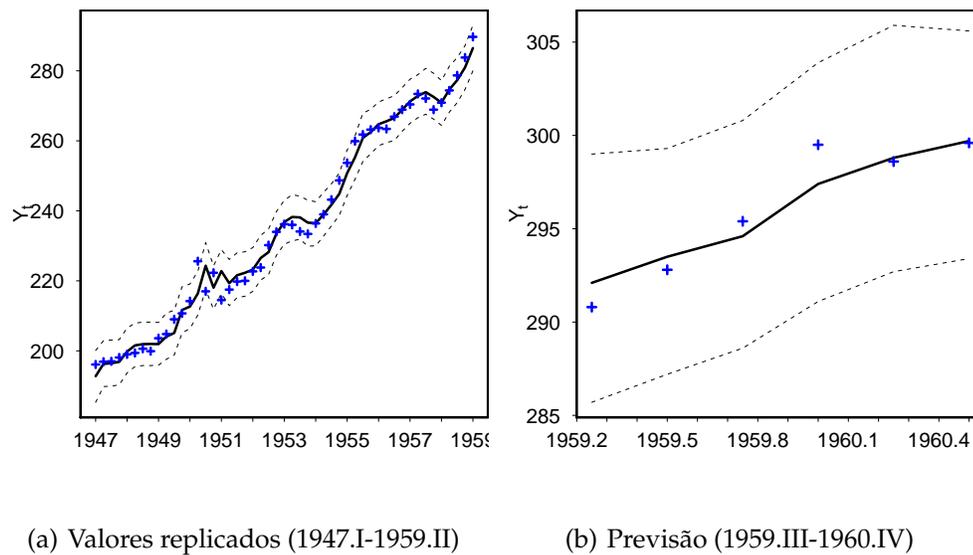


Figura 2.5: **Valores replicados e previsão do consumo obtido com o modelo em (2.16) com Erro II e incluindo um intercepto.** Em ambos painéis a linha sólida corresponde à média, as linhas pontilhadas ao primeiro e terceiro quartis da distribuição a posteriori. As cruzes correspondem aos valores observados.

### 2.5.6 Tratamento da Endogeneidade

Até agora temos considerado a renda como variável predeterminada somente com o objetivo de ilustrar nossa proposta e compará-la com Zellner & Geisel (1970); contudo, é bastante conhecida a interdependência presente num modelo macro-econômico de renda e consumo. Para lidar com esta situação, nesta seção adotamos a abordagem de Congdon (2001, exemplo 7.38). Primeiro ajustamos um modelo  $AR(1)$  para  $X$  com a finalidade de formar uma variável instrumental e depois estimamos o modelo em (2.16) com Erro II.

A tabela 2.5 apresenta os resultados obtidos, incluindo as estimativas para o coeficiente auto-

regressivo do modelo  $AR(1)$ ,  $\alpha$ . Basicamente, comparamos as estatísticas a posteriori para  $k$  e  $k(1 - \lambda)$  com as apresentadas na segunda linha da tabela 2.4. Observamos que, como era esperado, as médias obtidas são menores ao usar variáveis instrumentais enquanto que os desvios padrões são maiores.

Tabela 2.5: Estatísticas a posteriori associadas com os parâmetros do modelo em (2.16) com Erro II e modelo  $AR(1)$  para a renda

	média	d.p.	2,5%	50%	97,5%	$\hat{R}^b$
$k$	0,902	0,102	0,606	0,922	0,995	1,001
$\lambda$	0,768	0,145	0,407	0,808	0,952	1,003
$k(1 - \lambda)$	0,211	0,135	0,029	0,172	0,548	1,003
$\rho$	0,753	0,108	0,549	0,748	0,973	1,002
$\sigma^2$	9,819	2,011	6,587	9,602	14,33	1,001
$\alpha$	0,998	0,013	0,972	0,998	1,024	1,003

<sup>b</sup>  $\hat{R}$  é o fator de redução de escala (na convergência,  $\hat{R} = 1$ ).

## 2.6 Considerações Finais

Embora os primeiros modelos de defasagens distribuídas tenham sido propostos durante a década de 50, eles são bastante úteis e muito utilizados na atualidade. Certamente esta classe de modelos envolve vários tópicos que merecem atenção: a forma funcional da relação entre os seus coeficientes, a estrutura considerada para o termo de erro, a não estacionariedade das séries temporais, variáveis endógenas, dados de painel, etc.

Neste capítulo discutimos a inferência dos modelos de defasagens distribuídas do ponto de vista bayesiano. Especificamente, mostramos que estes modelos podem ser representados dentro de uma classe muito ampla: os modelos lineares dinâmicos. Este fato permite-nos usar todas as vantagens oferecidas pela abordagem bayesiana, como a incorporação de informação subjetiva e a descrição da incerteza, através da distribuição a posteriori dos parâmetros de interesse.

A inferência sobre modelos de defasagens distribuídas pode ser realizada com métodos *MCMC*. Neste capítulo mostramos que esta tarefa é bastante simplificada com o uso do *WinBUGS*, um pa-

cote de livre disponibilidade. Mostramos também que a comparação de modelos pode ser realizada utilizando o *DIC*, critério que está implementado no *WinBUGS*, e o *EPD*. Além disso, as cadeias obtidas através do *WinBUGS* podem ser exploradas com a biblioteca *CODA* (Best et al., 1997), que também é de livre disponibilidade, e é útil para obter indicadores (diagnósticos) de convergência das cadeias.

Na verdade, outros algoritmos particulares podem ser utilizados, mas o ganho na eficiência obtido com eles, é compensado pelo *WinBUGS* devido à facilidade de implementação dos códigos. Este fato permite-nos explorar diferentes alternativas de modelagem, avaliar o efeito da escolha de prioris, o uso de variáveis transformadas ou a inclusão de termos estocásticos na equação do sistema, como foi apresentado aqui.

Modelos mais complexos podem ser analisados do ponto de vista bayesiano. Por exemplo, Migon & Harrison (1985) mostram um caso não normal e não linear usando estimação *Linear Bayes*, e Alves (2005) propõe um esquema *MCMC* neste mesmo contexto.

Um tópico de importância teórica, que é parte do nosso trabalho atual, é a escolha da forma funcional da função de transferência e como isto afeta as previsões e incertezas associadas. Uma possibilidade é o uso de uma função de transferência de segunda ordem.

## 2.7 Apêndice: Exemplo de código usado em WinBUGS

Este é o código utilizado para ajustar o modelo de Koyck em (2.12) com erros não correlacionados e homoscedásticos, isto é ( $\epsilon_t \stackrel{\text{ind}}{\sim} N(0, \sigma_1^2)$ ).

```

model #(Ct=consumo, Yt=renda){
  e[1]    <- 0
  for(t in 2:T){
    mean.Y[t] <- lambda*Y[t-1] + k*(1-lambda)*X[t] - lambda*e[t-1]
    Y[t]      ~ dnorm(mean.Y[t], tau.y)
    Y.hat[t]  ~ dnorm(mean.Y[t], tau.y)
    e[t]      <- Y[t] - mean.Y[t]
  }
  # Prior
  lambda ~ dbeta(1,1)
  k      ~ dbeta(1,1)
  tau.y  ~ dgamma(0.0001,0.0001)
  var.y  <- 1/tau.y
  phi    <- k*(1-lambda)
}

```

## Capítulo 3

### UM ESQUEMA DE AMOSTRAGEM EFICIENTE PARA MODELOS DINÂMICOS GENERALIZADOS

Neste capítulo apresentamos um esquema de amostragem múltipla dos parâmetros de estado de modelos dinâmicos não lineares, dentro de um amostrador de Gibbs. O esquema proposto combina a abordagem do *Conjugate Updating* para modelos lineares generalizados dinâmicos com o *Backward Sampling* utilizado nos modelos dinâmicos normais para amostrar do vetor de estados. Denominamos este esquema de *Conjugate Updating Backward Sampling (CUBS)*. Observamos que o *CUBS* reduz significativamente o tempo computacional necessário para atingir a convergência de todas as cadeias dos modelos, sendo simples de implementar. As amostras das distribuições a posteriori dos outros parâmetros dos modelos podem ser obtidas com outros algoritmos. Apresentamos os resultados de um estudo de Monte Carlo extensivo realizado para comparar a eficiência do *CUBS* versus outros algoritmos previamente propostos na literatura bayesiana e, também, duas aplicações com dados reais.

#### 3.1 Introdução

Os modelos dinâmicos não lineares e não gaussianos são amplamente utilizados na estatística aplicada, em particular, sob a abordagem bayesiana. Neste capítulo consideramos que um modelo dinâmico de tempo discreto definido na família exponencial é tal que

$$y_t | \eta_t, \phi \sim \exp[\phi\{y_t \eta_t - a(\eta_t)\}] b(y_t, \phi), \quad t = 1, 2, \dots \quad (3.1a)$$

$$g(\mu_t) = F_t(\psi_1)' \theta_t \quad (3.1b)$$

$$\theta_t = G_t(\psi_2) \theta_{t-1} + w_t, \quad w_t \sim N(\mathbf{0}, \mathbf{W}) \quad (3.1c)$$

$$\theta_0 | \mathbf{Y}^0 \sim N(\mathbf{m}_0, \mathbf{C}_0),$$

onde os vetores  $\theta_t$  são denominados parâmetros de estado e estão relacionados através do tempo via (3.1c), a equação do sistema;  $\theta_0 | \mathbf{Y}^0$  é a informação disponível anterior ao período em que se tem informação através das observações;  $\eta_t$  é o parâmetro natural e  $E[y_t | \eta_t] = \mu_t = a(\eta_t)$  está relacionado a  $\theta_t$  via uma função de ligação conhecida,  $g(\cdot)$ ;  $\psi_1$  e  $\psi_2$  denotam todas as quantidades desconheci-

das envolvidas na definição de  $F_t(\psi_1)$  e  $G_t(\psi_2)$ , respectivamente. Também,  $w_t$  é o termo de erro do sistema com variância (de evolução)  $W$ . Frequentemente  $W$  é uma matriz diagonal e pode variar ao longo do tempo. Do ponto de vista bayesiano, a especificação do modelo fica completa depois da atribuição da distribuição a priori dos parâmetros. A distribuição a posteriori resultante para os parâmetros em (3.1) não tem uma forma fechada conhecida, por esta razão a inferência não pode ser realizada de forma analítica. Logo, métodos de simulação estocástica são necessários para gerar as amostras de interesse, mas a alta correlação existente entre os  $\theta_t$ s pode dificultar a geração de amostras independentes. No contexto dos métodos *MCMC*, nosso objetivo é propor um algoritmo com passo de Metropolis-Hastings independente e eficiente (pouca demanda de tempo computacional e fácil de implementar) para amostrar da distribuição condicional completa de  $\theta_t$ ,  $t = 1, \dots, T$ .

Antes do uso intensivo das técnicas *MCMC*, West et al. (1985), Kitagawa (1987) e Fahrmeir (1992), entre outros, propuseram diferentes métodos para obter aproximações da distribuição a posteriori em questão. Desde a década passada, como pode ser visto em Migon et al. (2005), vários algoritmos *MCMC* têm sido propostos para amostrar dos estados dos modelos dinâmicos. Em particular, Gamerman (1998) e Geweke & Tanizaki (2001) consideraram abordagens bayesianas com passos de Metropolis-Hastings dentro do amostrador de Gibbs. Dado que o ponto crítico de um algoritmo de Metropolis-Hastings é a escolha da distribuição proposta, eles estão interessados na construção de propostas eficientes. Por exemplo, Gamerman (1998) sugeriu o uso de um modelo dinâmico normal ajustado com a finalidade de construir uma boa densidade proposta. Esta idéia pode ser implementada de três maneiras diferentes: na amostragem individual ou em blocos dos estados, ou na amostragem individual dos termos de erro da equação de evolução. Gamerman concluiu que os movimentos individuais são preferíveis aos movimentos em blocos e que amostrar dos erros é mais eficiente porque as cadeias resultantes são menos autocorrelacionadas, o que ajuda a acelerar a convergência do algoritmo. Entretanto, esta proposta requer um esforço computacional bastante significativo pois o código, geralmente, é complicado de escrever e pode levar “muito” tempo para completar uma única iteração. Por outro lado, Geweke & Tanizaki (2001) propuseram várias formas de construir densidades propostas para amostrar os estados em passos individuais e realizaram um estudo de Monte Carlo para avaliar a sensibilidade dos resultados a esta escolha. Eles concluíram que os resultados são bastante robustos à escolha da densidade proposta para o passo de Metropolis-Hastings.

Nossa alternativa é amostrar do vetor de estados  $\{\theta_1, \dots, \theta_T\}$  de (3.1), em bloco, isto é, usando um movimento múltiplo num esquema análogo ao *FFBS* (do inglês *Forward Filtering and Backward*

*Sampling*), de Frühwirth-Schnater (1994) e de Carter & Kohn (1994) proposto para os modelos lineares dinâmicos normais. Ambos trabalhos concluíram que amostrar todos os estados em forma simultânea acelera a convergência das cadeias em comparação ao caso de movimentos individuais. A idéia principal do *FFBS* é o uso da decomposição da distribuição a posteriori conjunta dos estados, no produtório das distribuições retrospectivas (ou suavizadas). No caso gaussiano, as quantidades necessárias para a especificação dos momentos das distribuições suavizadas são obtidas a partir do Filtro de Kalman. Logo, para obter amostras da distribuição condicional completa conjunta, primeiro calculam-se os momentos “para frente” (*on-line*) ou passo de filtragem, e depois se amostra das distribuições normais suavizadas (que corresponde ao passo de *Backward Sampling*).

Neste capítulo, nossa proposta para modelos não normais é substituir o primeiro passo do *FFBS* (o passo de filtragem) pelo *Conjugate Updating* apresentado em West et al. (1985), para modelos dinâmicos na família exponencial. Em outras palavras, nossa proposta consiste em aproximar as distribuições *on-line* introduzindo uma priori conjugada para o parâmetro canônico da distribuição da observação, e depois usar um passo de amostragem retrospectiva como no caso gaussiano (o passo *Backward Sampling*). Denominamos este algoritmo *CUBS*, abreviação do inglês *Conjugate Updating Backward Sampling*. A grande força deste método reside em considerar a estrutura de correlação temporal dos  $\theta_t$ . As principais vantagens são: facilidade de implementação e redução do tempo e esforço computacional quando comparado com outros algoritmos. O *CUBS* pode ser usado dentro de outros algoritmos *MCMC* e, portanto, as distribuições a posteriori dos demais parâmetros do modelo em (3.1) podem ser amostradas de maneira usual.

Este capítulo está organizado da seguinte forma. Na seção 3.2 descrevemos o algoritmo *MCMC* proposto, destacando as vantagens de sua utilização. Na seção 3.3 comparamos os resultados obtidos com o *CUBS* com outros algoritmos previamente propostos na literatura bayesiana. A comparação é feita com base num estudo de simulação extensivo realizado com dados artificiais. Também apresentamos os resultados do uso do *CUBS* em duas aplicações a dados reais. Finalmente, apresentamos algumas conclusões e projetos futuros na seção 3.4.

### 3.2 Esquema de Amostragem Proposto

Nesta seção propomos um esquema de amostragem *MCMC* baseado na combinação da filtragem conjugada (*Conjugate Updating*) com a amostragem retrospectiva (*Backward Smoothing*). Iniciamos esta seção revisando os principais aspectos da filtragem conjugada e a amostragem retrospec-

tiva introduzidos, respectivamente, no contexto dos modelos lineares generalizados dinâmicos (*DGLM*) e modelos lineares dinâmicos (*DLM*) normais. Depois apresentamos nossa proposta, o algoritmo *CUBS*, que combina as idéias descritas anteriormente. Terminamos esta seção ressaltando as vantagens do *CUBS*. Para definir a nossa notação, daqui em diante  $\Theta = \{\theta_1, \dots, \theta_T\}$  representa o vetor de estados,  $\mathbf{Y}^t = \{y_1, \dots, y_t\}$  representa a informação até o tempo  $t$  e  $\Phi = (\phi, \psi_1, \psi_2, \mathbf{W})$  denota o vetor de todas as quantidades desconhecidas em (3.1), exceto  $\Theta$ . De acordo com a especificação em (3.1) e assumindo distribuições a priori independentes para  $\phi, \Psi_1, \Psi_2, \mathbf{W}$ , e para a informação inicial  $\theta_0$ , a distribuição a posteriori é

$$p(\theta_1, \dots, \theta_T, \theta_0, \psi_1, \psi_2, \mathbf{W}, \phi | y_1, \dots, y_T) \propto \prod_{t=1}^T p(y_t | \theta_t, \eta_t, \phi) \\ p(\theta_t | \theta_{t-1}, \mathbf{F}_t(\psi_1), \mathbf{G}_t(\psi_2), \mathbf{W}) p(\theta_0) p(\psi_1) p(\psi_2) p(\mathbf{W}) p(\phi). \quad (3.2)$$

Como a distribuição em (3.2) não tem forma fechada conhecida, propomos usar métodos *MCMC* para amostrá-la. Nosso interesse é amostrar da distribuição condicional completa de  $\{\theta_1, \dots, \theta_T\}$ . Especificamente, usamos um algoritmo híbrido, um amostrador de Gibbs com passos de Metropolis-Hastings. Nosso foco é a construção de uma distribuição proposta eficiente.

Nos anos 80, antes do uso dos métodos *MCMC*, West et al. (1985) propuseram um algoritmo seqüencial para aproximar a distribuição a posteriori de um modelo linear generalizado dinâmico (*DGLM*). Esse algoritmo é denominado *Conjugate Updating* e baseia-se nos procedimentos de *Linear Bayes*, uma aproximação realizada ao nível das distribuições a priori. Nos *DGLM*, a priori dos estados é especificada parcialmente através dos seus primeiro e segundo momentos, digamos  $\mathbf{a}_t$  e  $\mathbf{R}_t$ . Estes são utilizados para determinar  $r_t$  e  $s_t$ , os parâmetros de uma distribuição a priori conjugada apropriada para  $\eta_t$ , isto é,  $\eta_t | \mathbf{Y}^{t-1} \sim CP(r_t, s_t)$ . Estes momentos são determinados resolvendo um conjunto relativamente simples de duas equações tal como é mostrado no apêndice (ver tabela 3.6). Os parâmetros da posteriori conjugada para o parâmetro natural são obtidos facilmente depois que  $y_t$  é observado. A informação nesta posteriori é repassada para a distribuição a posteriori dos estados. Como esta distribuição não é completamente conhecida, West et al. (1985) propuseram o uso de estimação via *Linear Bayes*, considerando uma função perda quadrática para derivar uma aproximação para a distribuição a posteriori do vetor de estados. A aproximação consiste em

$$p(\theta_t | \mathbf{Y}^t, \Phi) \propto p(\theta_t | \mathbf{Y}^{t-1}, \Phi) p(y_t | \theta_t, \Phi) \\ = \int p(\theta_t | \eta_t, \mathbf{Y}^{t-1}, \Phi) p(\eta_t | \mathbf{Y}^{t-1}, \Phi) p(y_t | \eta_t, \Phi) d\eta_t,$$

onde  $\eta_t$  é o parâmetro canônico de  $p(y_t | \eta_t, \Phi)$  no tempo  $t$ . Como o primeiro termo na integral acima

é parcialmente conhecido, usamos o princípio de *Linear Bayes* para obter

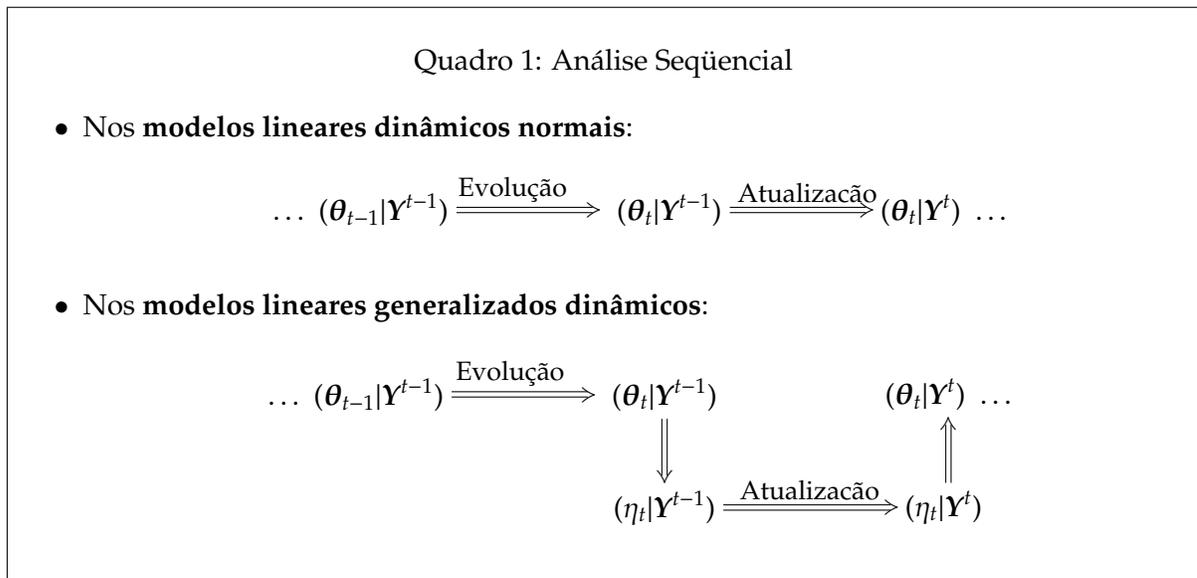
$$\begin{aligned} p(\boldsymbol{\theta}_t | \mathbf{Y}^t, \Phi) &\propto \int p(\boldsymbol{\theta}_t | \eta_t, \mathbf{Y}^{t-1}, \Phi) p(\eta_t | \mathbf{Y}^t, \Phi) d\eta_t \\ &\cong [\mathbf{m}_t, \mathbf{C}_t], \end{aligned}$$

onde  $[\mathbf{m}_t, \mathbf{C}_t]$  indica que  $p(\boldsymbol{\theta}_t | \mathbf{Y}^t, \Phi)$  é parcialmente conhecida através dos seus dois primeiros momentos, dados por

$$\begin{aligned} \mathbf{m}_t &= \mathbf{a}_t + \mathbf{R}_t \mathbf{F}_t (f_t^* - f_t) \frac{1}{q_t} \\ \mathbf{C}_t &= \mathbf{R}_t - \mathbf{R}_t \mathbf{F}_t \mathbf{F}_t' \mathbf{R}_t \left(1 - \frac{q_t^*}{q_t}\right) \frac{1}{q_t}, \end{aligned}$$

onde  $\mathbf{a}_t = \mathbf{G}_t \mathbf{m}_{t-1}$ ,  $\mathbf{R}_t = \mathbf{G}_t \mathbf{C}_{t-1} \mathbf{G}_t' + \mathbf{W}_t$  e  $(f_t, q_t)$  e  $(f_t^*, q_t^*)$  são, respectivamente, a média e variância, a priori e a posteriori de  $\eta_t$ . Vale a pena destacar que o conhecido Filtro de Kalman é recuperado sob a hipótese de normalidade para  $y_t$  e a função identidade como função de ligação.

O quadro 1 apresenta um esquema da análise seqüencial, via estimação *Linear Bayes*, para modelos lineares dinâmicos generalizados, em paralelo com a análise seqüencial, via Filtro de Kalman, para modelos lineares dinâmicos normais. Observamos que  $\eta_t$  faz o papel de uma variável latente que facilita a aproximação de  $p(\boldsymbol{\theta}_t | D_t)$ . Por conveniência assumimos que  $\eta_t | D_{t-1}$  segue uma distribuição conjugada com  $p(y_t | \eta_t)$ , mas esta hipótese pode ser relaxada.



No contexto dos modelos lineares dinâmicos normais é conhecido que as implementações de movimentos individuais nos métodos *MCMC* são bastante ineficientes devido à alta correlação

temporal entre os parâmetros de estado. Para contornar esta situação, Frühwirth-Schnater (1994) e Carter & Kohn (1994) propuseram amostrar todos os elementos de  $\Theta$  em (3.1) num passo de amostragem múltipla. A abordagem proposta por eles está baseada na conhecida decomposição da distribuição condicional completa conjunta dos estados:

$$p(\Theta|Y^T, \Phi) = p(\theta_T|Y^T, \Phi) \prod_{t=1}^{T-1} p(\theta_t|\theta_{t+1}, Y^t, \Phi). \quad (3.4)$$

Como  $p(\theta_t|\theta_{t+1}, Y^t, \Phi) \propto p(\theta_{t+1}|\theta_t, Y_t, \Phi)p(\theta_t|Y_t, \Phi)$ , é fácil mostrar que  $p(\theta_t|\theta_{t+1}, Y^t, \Phi)$  é  $N(m_t^s, C_t^s)$ , onde

$$\begin{aligned} m_t^s &= m_t + C_t G_t' (G_t C_t G_t' + W_t)^{-1} (\theta_{t+1} - G_t m_t) \\ C_t^s &= C_t - C_t G_t' (G_t C_t G_t' + W_t)^{-1} G_t C_t, \end{aligned}$$

onde  $m_t$  e  $C_t$  são os primeiros e segundos momentos obtidos fazendo uso do Filtro de Kalman. Daqui, o algoritmo *FFBS* consiste principalmente em amostrar os elementos de  $\Theta$  de forma seqüencial.

Com base nas idéias descritas acima, descrevemos agora como obter uma boa distribuição proposta para o passo de Metropolis-Hastings na amostragem de  $\Theta$  em (3.1). Um valor candidato para  $\Theta$ ,  $\Theta^*$ , será amostrado de uma distribuição multivariada obtida combinando as idéias de West et al. (1985) e Frühwirth-Schnater (1994). Especificamente, a distribuição normal multivariada  $N(\Theta^*|m^s, C^s)$ , com média e variância *on-line* -  $m, C$  - aproximadas pelo *Conjugate Updating*, ao invés do Filtro de Kalman. Cada  $\theta_t^*$  é amostrado seqüencialmente de  $t = T$  até  $t = 1$ , de suas distribuições retrospectivas, dadas pela fatorização de  $N(\Theta^*|m^s, C^s)$  em  $T$  densidades condicionais univariadas, de maneira análoga ao caso gaussiano.

A principal idéia de nosso esquema é descrita da seguinte maneira. Sejam  $m_t^{(i)}$  e  $C_t^{(i)}$ ,  $t = 1, \dots, T$ , os dois primeiros momentos das distribuições *on-line* obtidas via o *Conjugate Updating*, em (3.3), e sejam  $m_t^{s(i)}$  e  $C_t^{s(i)}$  os dois primeiros momentos das distribuições retrospectivas na  $i$ -ésima iteração; então, uma amostra de  $\Theta$  pode ser obtida nas seguintes três etapas:

(1). Amostre  $\Theta_T^*$  de  $N(m_T^{(i)}, C_T^{(i)})$ ;

(2). Amostre  $\Theta_t^*$  de  $N(m_t^{s(i)}, C_t^{s(i)})$ ,  $t = T - 1, \dots, 1$ ; onde

$$\begin{aligned} m_t^{s(i)} &= m_t^{(i)} + C_t^{(i)} G_t' (G_t C_t^{(i)} G_t' + W)^{-1} (\theta_{t+1}^* - G_t m_t^{(i)}) \\ C_t^{s(i)} &= C_t^{(i)} - C_t^{(i)} G_t' (G_t C_t^{(i)} G_t' + W)^{-1} G_t C_t^{(i)}. \end{aligned}$$

- (3). Faça  $\Theta^{(i)} = \Theta^*$  com probabilidade  $\pi$  e  $\Theta^{(i)} = \Theta^{(i-1)}$  com probabilidade  $1 - \pi$ , onde  $\pi = \min(1, A)$  e

$$A = \frac{p(\Theta^* | \mathbf{Y}^T, \Phi) q(\Theta | \mathbf{m}^{s(i)}, \mathbf{C}^{s(i)})}{p(\Theta | \mathbf{Y}^T, \Phi) q(\Theta^* | \mathbf{m}^{s(i)}, \mathbf{C}^{s(i)})}, \quad (3.5)$$

onde  $q(\cdot)$  é a densidade proposta com momentos  $\mathbf{m}$  e  $\mathbf{C}$ .

Em outras palavras, com o *CUBS* exploramos o uso de um algoritmo seqüencial, o *Conjugate Updating*, em forma conjunta com um método de amostragem, o *FFBS*. Basicamente, com o objetivo de construir nossa densidade proposta, introduzimos, a cada instante de tempo  $t$ , em (3.1), uma priori conjugada (*CP*) para o parâmetro natural da família exponencial. Daí que o *CUBS* é parecido com um algoritmo de dados aumentados (*data augmentation*) onde os parâmetros canônicos atuam como variáveis latentes que ajudam na simulação eficiente dos estados. Contudo, dado que várias aproximações são realizadas para tentar o uso de uma amostragem em bloco, este resultado é uma alternativa para a construção de uma densidade proposta independente para um passo do Metropolis-Hastings. Em particular, neste capítulo enfatizamos o uso de uma distribuição normal multivariada como densidade proposta, porém, outras distribuições, como a  $t$ -Student, podem ser utilizadas.

Vale a pena ressaltar que a distribuição proposta obtida via *CUBS* fornece uma boa aproximação para a distribuição objetivo  $p(\Theta | \cdot)$ , portanto, as taxas de aceitação são sempre bastante razoáveis. Além disso, o *CUBS* é fácil de implementar, característica que ajuda a diminuir o número de erros de programação involuntários. Finalmente, embora em alguns casos sejam necessários métodos numéricos para calcular os valores exatos de  $r_t$  and  $s_t$ , na prática podemos usar aproximações de funções que contribuam para diminuir o tempo computacional.

Dado o vetor de parâmetros,  $\Theta$ , amostras da distribuição a posteriori dos outros parâmetros do modelo podem ser obtidas seguindo as técnicas padrão como Metropolis-Hastings, *Slice Sampling* (Neal, 2003), *ARMS*, abreviação do inglês *Adaptive Rejection Metropolis Sampling* (Gilks & Wild, 1992; Gilks et al., 1995), etc. Estes métodos não afetam a convergência das cadeias geradas com nosso esquema de amostragem, mas podem afetar o tempo computacional necessário para a convergência de todas as cadeias (do algoritmo completo). Entretanto, pela nossa experiência, esse tempo será menor que o utilizado pelo algoritmo de Gamerman (1998). Para fins de exposição resumimos nosso esquema MCMC no quadro 2.

### Quadro 2: MCMC + CUBS

- (1). **Inicialização:** dar valores iniciais  $\theta^{(0)}, \psi^{(0)}$  e iniciar as iterações,  $i = 1$ ;
- (2). **Amostrar  $\Theta^{(i)}$**  usando *CUBS*:
  - (a) Calcular os momentos de  $p(\theta_t | D_t, \psi^{(i-1)})$ ,  $m^{(i)}$  e  $C^{(i)}$ , com o *Conjugate Updating*;
  - (b) Amostrar  $\theta^*$  com o *Backward Sampling*.
    - i. Amostrar  $\theta_T^*$  de  $\text{Normal}(m_T^{(i)}, C_T^{(i)})$
    - ii. Amostrar  $\theta_t^*, t = T - 1, \dots, 1$ , de  $p(\theta_t | \theta_{t+1}^*, \psi^{(i-1)})$
  - (c) Faça  $\Theta^{(i)} = \theta^*$  com probabilidade  $\pi_t$  e  $\Theta^{(i)} = \Theta^{(i-1)}$  com probabilidade  $1 - \pi_t$ , onde  $\pi_t = \min(1, A)$  e  $A$  é a razão de aceitação do Metropolis-Hastings:
 
$$A = \min\left\{1, \frac{\omega(\theta^*)}{\omega(\theta)}\right\}, \quad \omega(\theta^*) = \frac{p(\theta^*)}{q(\theta^*)};$$
- (3). **Amostrar  $\psi^{(i)}$**  usando, em geral, um passo de Metropolis-Hastings;
- (4). **Amostrar  $\phi^{(i)}$**  usando, em geral, um passo de Metropolis-Hastings;
- (5). **Atualização:** faça  $i = i + 1$  e volte para (2)c até a convergência.

## 3.3 Comparação de Esquemas de Amostragem

### 3.3.1 Estudo de Monte Carlo

Nesta seção descrevemos um estudo de Monte Carlo que teve como objetivo comparar o uso do *CUBS* dentro de um algoritmo *MCMC*, para fazer inferência sobre todos os parâmetros de estado de um particular modelo linear dinâmico, versus alguns dos esquemas de amostragem propostos por Gamerman (1998) e Geweke & Tanizaki (2001).

#### – Geração dos Dados Artificiais

Geramos dados artificiais segundo um modelo relativamente simples porque nosso interesse principal é avaliar a eficiência na geração de amostras somente dos parâmetros de estado. O modelo dinâmico Poisson de primeira ordem mostrado em (3.6), com  $\theta_0 = 0,50$  e  $W = 0,01$ , foi utilizado para gerar três conjuntos de dados, cada um com 100 séries temporais. A diferença entres os três

conjuntos foi o tamanho das séries temporais. Desta forma, assumimos que

$$y_t | \eta_t \sim \text{Poisson}(\eta_t), \quad t = 1, \dots, T \quad (3.6a)$$

$$\log(\eta_t) = \theta_t \quad (3.6b)$$

$$\theta_t = \theta_{t-1} + w_t \quad w_t \sim N(0, W) \quad (3.6c)$$

$$\theta_0 | \mathbf{Y}^0 \sim N(m_0, C_0). \quad (3.6d)$$

Em resumo, geramos valores aleatórios de  $\theta_t$  com base em (3.6c), a seguir geramos  $\mathbf{Y}^T$  com base em (3.6a) para  $T = 50, 100, 300$ . Depois disso, rodamos nossas rotinas MCMC para obter amostras da distribuição a posteriori de  $\Theta$  e  $W$ , e comparamos os resultados com os valores verdadeiros. As distribuições a priori de  $W$  e  $\theta_0$  foram, respectivamente, uma gama invertida com ambos parâmetros iguais a 0,001, isto é,  $IG(0,001; 0,001)$ , e uma normal com média zero e variância grande, isto é,  $N(0; 10^3)$ . Em outras palavras, utilizamos prioris não informativas.

#### – Esquemas de Amostragem Estudados

As quantidades desconhecidas em (3.6) são  $(\theta_0, \theta_1, \dots, \theta_T, W)$ . Como a distribuição condicional completa de  $W$  é gama invertida, amostramos diretamente dela, entretanto para o vetor de estados usamos passos de Metropolis-Hastings.

Como mencionamos na seção 3.1, Gamerman (1998) sugere a construção de um modelo linear dinâmico gaussiano ajustado (ou auxiliar) tal que

$$\tilde{y}_t = \mathbf{F}'_t \boldsymbol{\theta}_t + v_t, \quad v_t \sim N(0, \tilde{V}_t) \quad (3.7a)$$

$$\boldsymbol{\theta}_t = \mathbf{G}_t \boldsymbol{\theta}_{t-1} + \mathbf{w}_t, \quad \mathbf{w}_t \sim N(\mathbf{0}, \mathbf{W}_t), \quad (3.7b)$$

onde as observações ajustadas ou pseudo-observações  $\tilde{y}_t$  e suas variâncias associadas  $\tilde{V}_t$  são

$$\begin{aligned} \tilde{y}_t &= \eta_t + (y_t - \mu_t)g'(\mu_t) \\ \tilde{V}_t &= \ddot{a}(\eta_t)[g'(\mu_t)]^2. \end{aligned}$$

Com base nesse “novo” modelo, os seguintes esquemas de amostragem foram propostos:

- **Proposta I.** Amostrar  $w_t^*$ , um valor candidato para os erros do sistema,  $w_t$ , da distribuição condicional completa obtida a partir da seguinte re-parametrização de (3.7):

$$\begin{aligned} \tilde{y}_t &= \mathbf{F}_t \sum_{j=1}^t \mathbf{G}^{t-j} \mathbf{w}_j + v_t, \quad \mathbf{G}_t = \mathbf{G} \forall t, \quad t = 2, \dots, T, \\ v_t &\sim N(0, \tilde{V}_t), \quad \mathbf{w}_t \sim N(\mathbf{0}, \mathbf{W}), \quad \mathbf{w}_1 \sim N(\mathbf{m}_1, \mathbf{C}_1). \end{aligned}$$

- **Proposta II.** Amostrar  $\theta_t^*$  da condicional completa de  $\theta_t$  em (3.7) (densidade normal).

Na realidade há uma terceira proposta que consiste em usar o algoritmo *FFBS* em (3.7). Contudo, esta proposta não é recomendada por Gamerman pois conduz a taxas de aceitação muito baixas, por isso não é implementada neste capítulo.

Geweke & Tanizaki (2001) propõem diferentes distribuições propostas e recomendam a amostragem individual dos estados, isto é, amostrar os estados um a um e não em bloco. Três dos esquemas propostos são:

- **Proposta I.** Amostre  $\theta_t^*$  da densidade obtida a partir da equação do sistema. Em outras palavras, amostrar  $\theta_t^*$  de  $N(G_t(\psi_2)\theta_{t-1}, W_t)$ .
- **Proposta II.** Amostrar  $\theta_t^*$  da densidade normal com média e variância iguais às estimativas retrospectivas, no instante  $t$ , obtidas da aplicação do Filtro de Kalman Estendido (*EKS*). Esse filtro consiste na aplicação das recursões do Filtro de Kalman e recursões de suavização a um modelo dinâmico normal ajustado, como em (3.7). Em outras palavras, também utiliza pseudo-observações como em Gamerman (1998).
- **Proposta III.** Amostrar  $\theta_t^*$  de uma densidade normal com base num passeio aleatório, isto é, a média é o valor corrente da cadeia, e variância obtida como na Proposta II.

Neste capítulo utilizamos as duas alternativas de amostragem individual proposta por Gamerman (1998), e as três alternativas descritas acima de Geweke & Tanizaki (2001). Adicionalmente, consideramos os momentos retrospectivos obtidos depois de aplicar o *Conjugate Updating*, para construir densidades propostas individuais para  $\theta_t$ , em forma análoga à proposta II de Geweke & Tanizaki (2001), com este esquema avaliamos a amostragem de movimentos simples que decorre naturalmente de nossa proposta. Em resumo, os sete esquemas que comparamos são:

- I. *CUBS*: amostragem em bloco como apresentada na seção 3.2.
- II. *Conjugate Updating*: amostragem individual. A proposta é uma densidade normal com média e variância iguais aos momentos obtidos com a análise retrospectiva no instante  $t$ , após o uso do *Conjugate Updating* em toda a série.
- III. De Gamerman (1998) - Proposta I: amostragem individual dos erros do sistema;

- IV. De Gamerman (1998) - Proposta II: amostragem individual direta dos estados;
- V. De Geweke & Tanizaki (2001) - Proposta I: a proposta é obtida a partir da equação do sistema;
- VI. De Geweke & Tanizaki (2001) - Proposta II: a proposta é uma densidade normal com média e variância iguais aos momentos retrospectivos obtidos com o Filtro de Kalman Estendido.
- VII. De Geweke & Tanizaki (2001) - Proposta III: a proposta é uma densidade normal com média igual ao valor corrente da cadeia e variância igual ao segundo momento retrospectivo obtido com o Filtro de Kalman Estendido.

#### – Critérios de Comparação

Com a finalidade de comparar os esquemas de amostragem mencionados acima, calculamos o fator de ineficiência e a raiz quadrada do erro quadrático médio para  $\Theta$  e  $Y^T$ . Seja INEF o fator de ineficiência, dado por

$$\text{INEF} = 1 + 2 \sum_{j=1}^{n-1} \frac{n-j}{n} \rho_j,$$

onde  $\rho_j$  é a autocorrelação com defasagem  $j$  para os valores de uma cadeia de  $n$  valores ( $n \rightarrow \infty$ ). INEF é uma medida de quão longe está a cadeia para representar uma amostra independente (veja Gamerman & Lopes (2006) para maiores detalhes). Quanto maior é INEF, menor é a eficiência do esquema.

Seja  $\text{RMSE}_{\theta_t}$  a raiz quadrada do erro quadrático médio para  $\theta_{rep,t}$ , definida por

$$\text{RMSE}_{\theta_t} = \left[ \frac{1}{K} \sum_{k=1}^K (\hat{\theta}_{rep,t}^{(k)} - \theta_t^{(k)})^2 \right]^{1/2},$$

onde  $\hat{\theta}_{rep,t}^{(k)}$  é a média da amostra da distribuição a posteriori e  $\theta_t^{(k)}$  é o valor verdadeiro do  $t$ -ésimo estado do  $k$ -ésimo conjunto de dados artificiais ( $K = 100$ ). Então a média e desvio padrão de  $\text{RMSE}_{\theta_t}$  são dados, respectivamente, por

$$\text{média}(\text{RMSE}_{\theta}) = \frac{1}{T} \sum_{t=1}^T \text{RMSE}_{\theta_t}$$

e

$$\text{d.p.}(\text{RMSE}_{\theta}) = \left[ \frac{1}{T} \sum_{t=1}^T \left\{ \text{RMSE}_{\theta_t} - \text{média}(\text{RMSE}_{\theta}) \right\}^2 \right]^{1/2}.$$

Em outras palavras, média( $\text{RMSE}_\theta$ ) e d.p.( $\text{RMSE}_\theta$ ) são as medidas resumo do erro quadrático médio obtidos considerando as 100 replicações de cada conjunto de dados. Os valores de  $Y_{rep,t}$  e  $\text{RMSE}_Y$  são obtidos de forma análoga.

Os outros critérios de comparação apresentados são: tempo computacional e taxas de aceitação. Os dados foram gerados no pacote R versão 2.4 (R Development Core Team, 2005) e as rotinas MCMC foram escritas em  $\text{Ox}$  versão 3.2 (Doornik, 2002).

### – Resultados

Os resultados apresentados nesta seção provêm da análise de amostras de tamanho 10000 das cadeias que foram iteradas 50000 vezes, tendo as primeiras 40000 sido descartadas como período de aquecimento. A única exceção são os resultados do esquema **III** cujas cadeias foram iteradas somente por 15000 vezes, descartando-se as primeiras 10000 iterações. Adotamos esta estratégia com base nos resultados de Gamerman (1998) sobre a velocidade de convergência e a alta demanda de tempo computacional desse algoritmo. Além disso, não fazemos uso do espaçamento entre observações (*thinning interval*) para obter as estatísticas a posteriori, pois um dos nossos objetivos é medir a autocorrelação que cada método produz em iterações consecutivas.

Em forma global os resultados indicam que sob o esquema **I**, as ineficiências exibidas pelas cadeias do vetor de parâmetros são independentes da posição no tempo, isto é, as ineficiências das cadeias de  $\theta_t$  para  $t$  no início, meio e final da série de tempo são praticamente as mesmas. Outro resultado interessante é que não dependem do número de observações da série. Embora seja uma amostragem em bloco, as taxas de aceitação são sempre razoáveis, estando, em média, entre 26% e 46%. Os esquemas de amostragem individual (**II**, **V**, **VII** e **VII**) foram os menos eficientes. Adicionalmente, as funções de autocorrelação empíricas (que não são apresentadas aqui) do esquema **I** foram tão boas quanto as do esquema **III**.

A tabela 3.1 mostra o  $\text{RMSE}_{Y^T}$ ,  $\text{RMSE}_\Theta$ , a média e desvio padrão das taxas de aceitação, considerando todos os  $\theta_t$ , assim como a média e desvio padrão dos tempos computacionais utilizados com cada algoritmo para completar todas as iterações. Nessa tabela observamos que os valores médios dos RMSEs, (média( $\text{RMSE}$ )), de todos os esquemas são bastante similares, como esperado, tanto para a resposta ( $Y^T$ ) como para os estados ( $\Theta$ ). As taxas de aceitação dos esquemas **III** e **IV** são as maiores, mais de 90% em todos os casos, entretanto as taxas de aceitação de **V**, **VI** e **VII** estão entre 30% e 50%. As taxas de aceitação do esquema **I** são, em média, parecidas com as taxas dos esquemas individuais. Vale a pena mencionar que para alguns conjuntos de dados, a taxa de aceitação foi a mais baixa entre todos os esquemas, o que é esperado por se tratar de um

Tabela 3.1: Erro Quadrático Médio (RMSE), Taxa de Aceitação (média de  $T \times K$  taxas) e Tempo Computacional (média em segundo das  $K = 100$  amostras) para as séries temporais de tamanhos  $T = 50, 100, 300$ .

Esquema	RMSE $_{\mathbf{y}^T}$		RMSE $_{\Theta}$		Taxa de aceitação		Tempo(seg)
	média	d.p.	média	d.p.	média	d.p.	
$T = 50$							
I	1,252	0,109	0,236	0,036	42,629	–	280,30
II	1,276	0,116	0,234	0,026	33,525	7,157	214,98
III	1,259	0,110	0,249	0,051	97,252	1,228	629,08
IV	1,244	0,114	0,258	0,029	98,144	0,681	120,79
V	1,315	0,132	0,237	0,034	51,423	4,531	89,04
VI	1,240	0,113	0,261	0,028	37,796	6,301	248,25
VII	1,236	0,110	0,259	0,030	44,613	5,816	219,65
$T = 100$							
I	1,309	0,133	0,224	0,027	38,454	–	334,25
II	1,327	0,143	0,226	0,024	32,350	6,986	257,16
III	1,313	0,134	0,233	0,040	97,942	0,851	1533,09
IV	1,434	0,382	0,254	0,049	98,712	3,228	147,62
V	1,383	0,152	0,286	0,038	51,461	4,041	104,54
VI	1,312	0,142	0,235	0,019	34,271	6,864	301,94
VII	1,307	0,140	0,230	0,021	42,016	6,042	274,08
$T = 300$							
I	1,608	0,303	0,228	0,035	31,352	–	247,76
II	1,615	0,306	0,379	0,221	32,177	9,470	314,09
III	1,610	0,298	0,227	0,037	98,428	0,687	2158,93
IV	4,522	3,753	0,545	0,252	95,515	14,121	163,87
V	1,711	0,329	0,471	0,133	51,365	4,108	74,44
VI	1,607	0,301	0,236	0,034	32,881	8,985	371,31
VII	1,606	0,302	0,273	0,087	40,361	7,802	338,89

esquema em bloco. Contudo, é diferente de zero, resultado contrário ao obtido pela amostragem em bloco comentada em Gamerman (1998). Além disso, como esperado, observamos que a taxa de aceitação diminui conforme o tamanho da série temporal aumenta. Os tempos computacionais para os esquemas **I**, **II**, **V**, **VI** e **VII** foram muito parecidos (entre 200 e 300 segundos em média), porém, o esquema **III** utilizou o mesmo tempo para completar apenas 10% do número de iterações completadas pelos outros métodos.

Considerando as 100 séries temporais, a figura 3.1 mostra um resumo, através de um diagrama de caixa (*box plot*), dos logaritmos das ineficiências das cadeias amostradas dos estados  $\theta_t$ , nos instantes  $t = 5, 25, 45$  para  $T = 50$ ,  $t = 5, 45, 95$  para  $T = 100$ ,  $t = 5, 145, 295$  para  $T = 300$  e  $W$  para todos os casos. Nesta figura observamos que para as séries de tamanho  $T = 50$  e  $T = 100$ , os esquemas **V**, **VI** e **VII** exibem as maiores ineficiências. Para o esquema **III** notamos que as ineficiências diminuem quando o valor de  $t$  aumenta, ou seja, as cadeias são geradas de forma mais eficiente conforme a informação aumenta. Este resultado é esperado uma vez que a equação de estados é obtida em função dos termos de erro; logo, com o passar do tempo mais observações são utilizadas para a construção da distribuição proposta. Os diagramas de caixa do esquema **I** (*CUBS*) mostram que com este método obtivemos as menores ineficiências em  $t = 5$  e  $t = 25$ , e as segundas melhores em  $t = 45$ , quando  $T = 50$ . Para  $T = 300$  o esquema **III** resultou ser um pouco melhor (em termos de eficiência) que o esquema **I** no meio da série temporal, por exemplo em  $t = 145$ , e muito melhor no final da mesma, por exemplo em  $t = 295$ .

A figura 3.2 exhibe os diagramas de caixa dos logaritmos das medianas das amostras das distribuições a posteriori de  $W$ . As estimativas são razoáveis para todos os esquemas usados. Contudo, segundo estes resultados, o esquema **V** tende sempre a subestimar a variância do sistema. Este fato pode ser explicado pela pouca mistura das cadeias. Adicionalmente, vale a pena mencionar que alguns resultados, não apresentados aqui, mostraram que se o valor de  $W$  usado para gerar dados é “grande”, por exemplo  $W = 0, 1$ , os esquemas **III** e **IV** podem ter problemas para aproximar as distribuições a posteriori de interesse causando taxas de aceitação próximas de zero.

### 3.3.2 Dados Reais: Precipitação em Tokyo

Utilizamos todos os esquemas de amostragem descritos na seção anterior para ajustar um modelo dinâmico ao número de ocorrências de chuva acima de  $1\text{ mm}$  em Tokyo em cada dia do ano entre 1983-1984. Este exemplo foi analisado previamente por Kitagawa (1987) e Gamerman (1998), entre outros. O objetivo é estimar a probabilidade de sucesso (ocorrência de chuva) para cada dia calendário. O modelo considerado em (3.8) tem uma resposta binomial com probabilidade  $\pi_t$ , tal que

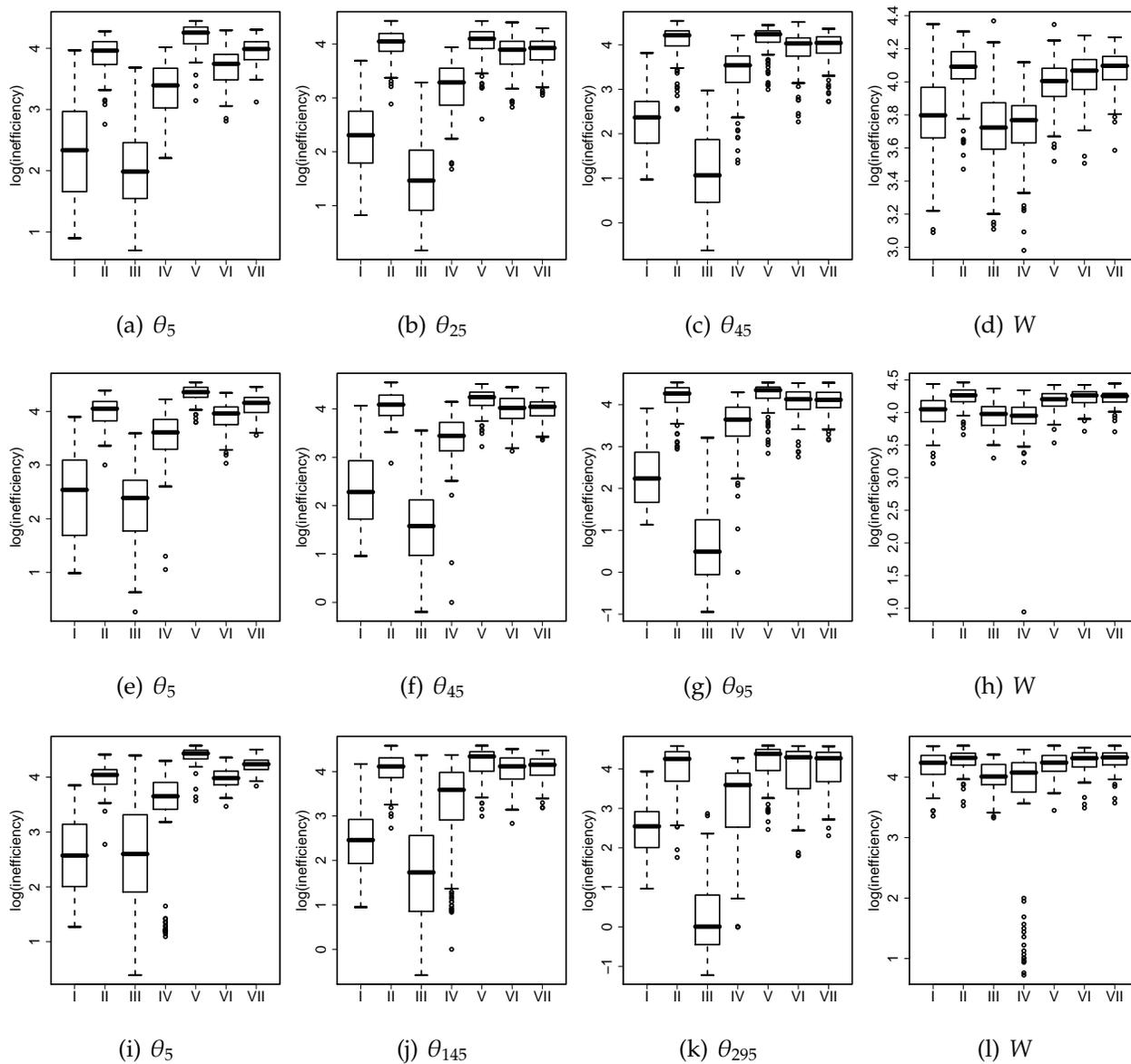


Figura 3.1: *Box plots* das ineficiências (em log) de  $\theta_t$ , em três instantes de tempo, e W: (a)–(d)  $T = 50$ , (e)–(h)  $T = 100$  e (i)–(l)  $T = 300$ .

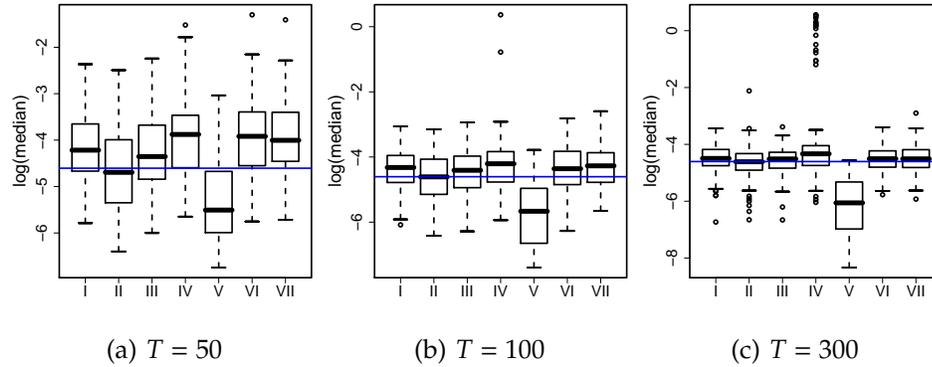


Figura 3.2: *Box plots dos logaritmos das medianas das amostras da posteriori de  $W$ .*

$\text{logit}(\pi_t)$  segue uma tendência de primeira ordem, isto é,

$$y_t | n, \pi_t \sim \text{Binomial}(n, \pi_t), \quad t = 1, \dots, 366; \quad n = 1 \text{ se } t = 60 \text{ e } n = 2 \text{ se } t \neq 60,$$

$$\text{logit}(\pi_t) = \theta_t \tag{3.8a}$$

$$\theta_t = \theta_{t-1} + w_t \quad w_t \sim N(0, W). \tag{3.8b}$$

Geramos uma cadeia com  $10^6$  iterações, e após de um aquecimento de tamanho 70000, guardamos as amostras a cada 10 iterações (exceto para o esquema **III**). Vale a pena mencionar que neste particular exemplo, devido aos valores pequenos da variável resposta, usamos métodos numéricos para calcular os momentos da priori conjugada para  $\pi_t$  no *CUBS*.

A tabela 3.2 mostra o  $\text{RMSE}_{\sqrt{T}}$ , a média e desvio padrão das taxas de aceitação, considerando todos os  $\theta_t$ . Observamos que a bondade de ajuste obtida com todos os esquemas é bastante similar, as menores taxas de aceitação correspondem a amostragem em blocos, e as maiores correspondem aos algoritmos **III** e **IV**. A figura 3.3 mostra as funções de autocorrelação das amostras de  $\theta_t$  em  $t = 50, 150, 250, 350$ . Notamos que as amostras com menor autocorrelação foram obtidas com os esquemas **I** e **III**, e que as maiores foram apresentadas pelas amostras obtidas para os esquemas **IV**, **V**, e **VI**. Os diagramas de caixa dos logaritmos das ineficiências calculados para todo o vetor de estados são apresentados na figura 3.4(a). Nela observamos que os melhores resultados (cadeias com melhores eficiência) foram obtidos com os esquemas **I** e **III**, sendo que o esquema **III** apresenta uma maior dispersão. Como visto no estudo de simulação, parâmetros de estado no início do tempo tendem a ser menos eficientes do que no final.

A tabela 3.3 apresenta algumas estatísticas associadas às amostras da distribuição a posteriori

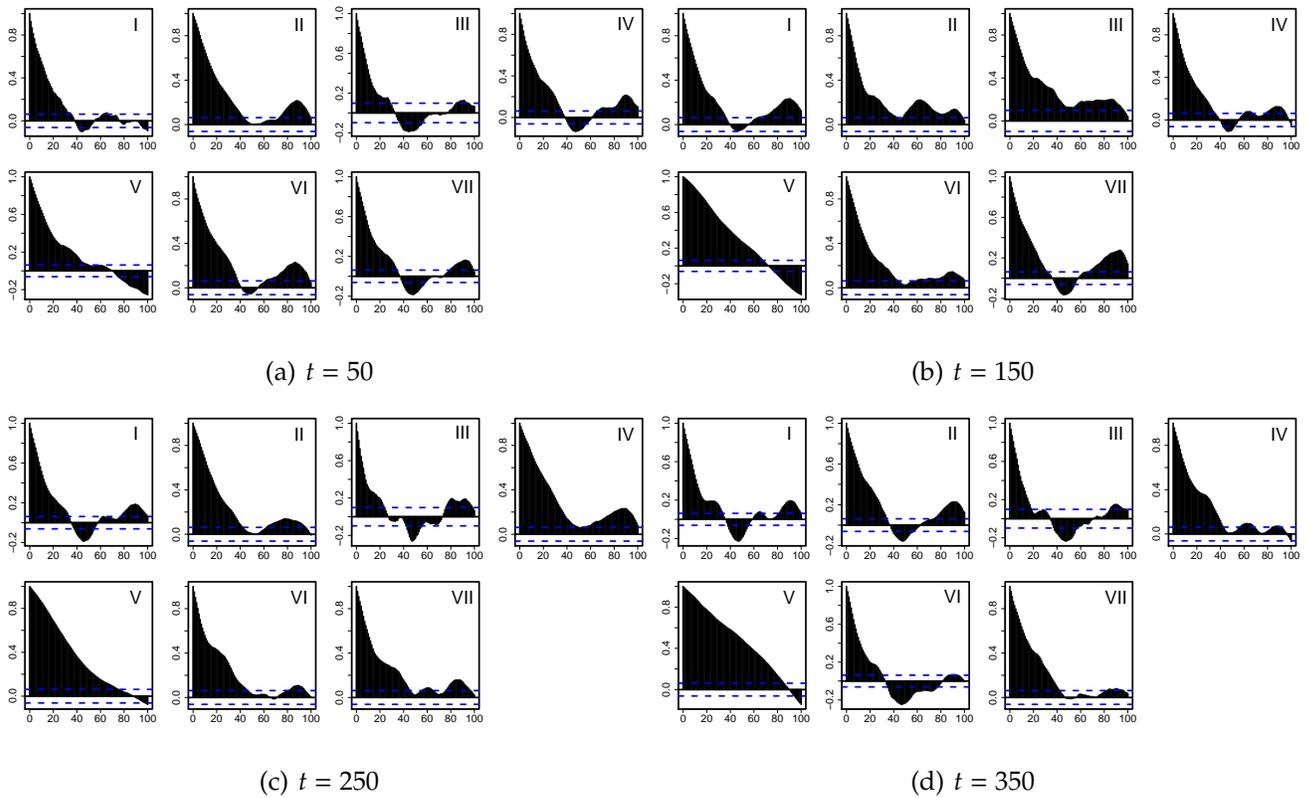


Figura 3.3: Funções de autocorrelação de  $\theta_t$ ,  $t = 50, 150, 250, 350$ . para o modelo da Chuva em Tokyo.

Tabela 3.2:  $RMSE_Y$  e Taxas de Aceitação para a Chuva em Tokyo.

Esquema	$RMSE_{Y^T}$		Taxa de aceitação	
	média	d.p.	média	d.p.
I	0,4698	0,2814	9,6400	–
II	0,4928	0,2825	29,9859	3,3905
III	0,4714	0,2836	98,0405	1,0536
IV	0,4772	0,2831	99,6911	4,9041
V	0,5771	0,2958	52,0549	1,8006
VI	0,4686	0,2891	26,6605	8,1579
VII	0,4793	0,2835	43,5408	4,0982

de  $W$  e a figura 3.4(b) mostra os diagramas de caixa das amostras de  $W$ . As medianas estimadas estão entre 0,04 e 0,10 (exceto para o esquema V, que, segundo nosso estudo de Monte Carlo, tende a subestimar a variância dos estados.)

Tabela 3.3: Estatísticas das amostras a posteriori de  $W$  para a chuva em Tokyo.

Esquema	média	d.p.	2.50%	mediana	97.50%	$\log(INEF)$
I	0,0919	0,0477	0,0235	0,0841	0,2187	3,5850
II	0,0492	0,0392	0,0079	0,0407	0,1496	3,7911
III	0,0903	0,0489	0,0254	0,0830	0,1984	2,8077
IV	0,0716	0,0443	0,0179	0,0618	0,1777	2,4307
V	0,0004	0,0003	0,0001	0,0003	0,0012	4,1382
VI	0,0838	0,0392	0,0326	0,0756	0,1787	4,1091
VII	0,0662	0,0380	0,0192	0,0570	0,1646	3,1724

### 3.3.3 Dados Reais: Chuva e Vazão no Rio Fartura (SP)

Baseados nas conclusões do experimento de simulação e da aplicação anterior, verificamos que os “melhores” esquemas de amostragem foram I e III. Assim, nesta seção comparamos as performances somente destes dois algoritmos num contexto mais complexo, o de um modelo dinâmico

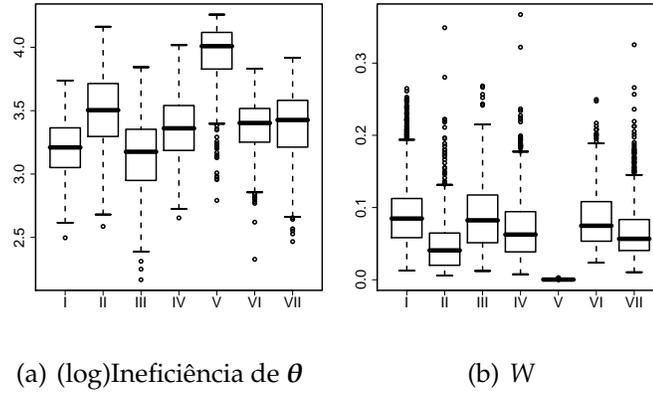


Figura 3.4: *Box plots* das (log)ineficiências de  $\theta_t, t = 1, \dots, 366$  e amostras da posteriori de  $W$  para a Chuva em Tokyo.

não linear. Especificamente, ajustamos um modelo de função de transferência com resposta gama a dados de precipitação e vazão mensal da bacia do Rio Fartura no estado de São Paulo (Brasil). O período observado vai de janeiro de 1960 a dezembro de 1964. Os dados estão publicados em Monteiro (1992).

Se  $y_t$  e  $x_t$  denotam, respectivamente, a vazão e a chuva no mês  $t$ , o modelo considerado é

$$y_t | \mu_t, \nu \sim \text{gama}(\mu_t, \nu), \quad t = 1, \dots, T, \quad (3.9a)$$

$$\log(\mu_t) = \alpha + E_t \quad (3.9b)$$

$$E_t = \rho E_{t-1} + \gamma x_t + \omega_t; \quad \omega_t \sim N(0, W), \quad (3.9c)$$

onde  $\mu_t$  é o valor esperado de  $y_t$  e  $\nu$  é o parâmetro de forma da distribuição gama,  $\alpha$  é o nível básico ou fluxo base de vazão,  $\rho$  é um fator de recarga ou taxa de memorização referente ao efeito  $E_t$ .  $\gamma$  é o efeito instantâneo da chuva e  $\omega_t$  é o erro do sistema que caracteriza nossa função de transferência como estocástica.

Em (3.9) temos que as distribuições condicionais completas de  $\gamma$  e  $W$  são normal e gama invertida, respectivamente, logo suas amostras são geradas de forma direta. As distribuições condicionais completas de  $\alpha$  e  $\rho$  não têm forma fechada conhecida e, portanto, usamos um passo de Metropolis-Hastings para gerar amostras de  $\alpha$  e o método *slice sampling* para gerar de  $\rho$ . Para gerar da distribuição a posteriori do vetor de estados  $E$  utilizamos os esquemas I e III em rotinas independentes.

Com o objetivo de fazer uma comparação “justa”, rodamos duas cadeias com ambos esquemas

durante 10 minutos cada uma, ou seja, fixamos o tempo computacional e deixamos livre o número de iterações. Na tabela 3.4 apresentamos o número de iterações que foi possível realizar com cada esquema. Observamos que com **I** realizamos, em média, 77477 iterações, entretanto, com **III** realizamos apenas 3946. Para determinar o período de aquecimento necessário para que em cada esquema, todas as cadeias (de todos os parâmetros deste exemplo) atingiram a distribuição estacionária, realizamos uma inspeção visual do gráfico do traço (*trace*) das cadeias geradas. Concluímos que o esquema **I** precisou de 2000 iterações, enquanto que o esquema **III** precisou de apenas 500 iterações.

Para comparar as autocorrelações e fatores de ineficiência das amostras geradas com ambos esquemas, construímos amostras de igual tamanho: 3363 elementos de cada cadeia. Para determinar esse tamanho, calculamos o máximo espaçamento (*thinning interval*) que poderia ser utilizado na amostra obtida com o esquema **I** para obter uma amostra de igual tamanho à obtida com o esquema **III**. O resultado, apresentado na tabela 3.4, indicou guardar as amostras do esquema **I** a cada 22 iterações. A motivação que está por trás deste exercício é que sabemos que as amostras obtidas com o esquema **I** podem apresentar maior autocorrelação que aquelas do **III**, mas o fato de poder realizar muitas iterações em pouco tempo, permite que o problema da autocorrelação seja contornado utilizando um espaçamento entre as amostras.

Tabela 3.4: Comparação dos esquemas I e III no caso da vazão do Rio Fartura

Critério	I	III
Tempo total por cadeia (em segundos)	600	600
Número de iterações da cadeia 1	78949	3961
Número de iterações da cadeia 2	76005	3931
Média do número de iterações por cadeia	77477	3946
Tempo para completar 100 iterações (em segundos)	0,77	15,20
Período de aquecimento (inspeção visual)	2000	500
Espaçamento ( <i>Thinning interval</i> )	22	1
Tamanho de amostra final de cada cadeia	3363	3363

Na figura 3.5 apresentamos os diagramas de caixa e os autocorrelogramas das amostras da distribuição a posteriori de alguns dos parâmetros em (3.9). Observamos que com ambos métodos

chegamos à mesma região de maior massa a posteriori e que as autocorrelações das cadeias geradas com o esquema **I** exibem valores menores. A figura 3.6 exibe os gráficos das correlações cruzadas das amostras geradas em cada esquema. Observamos que as amostras geradas para os diferentes parâmetros podem ser consideradas independentes, com exceção de  $\alpha$ .

Na tabela 3.5 temos os fatores potenciais de redução de escala, PSRF ou  $\hat{R}$  (Gelman & Rubin, 1992), e os fatores de ineficiência, INEF, calculados para as amostras da posteriori, de cada uma das cadeias, de alguns dos parâmetros em (3.9). Nesta tabela observamos que todos os valores do PSRF, estão próximos de 1, portanto, temos indícios que todas as cadeias geradas atingiram a convergência. Segundo o INEF quando a amostra é formada por todos os elementos gerados, isto é, considerando todas as iterações das cadeias, ( $Thin=1$ ), o esquema **I** exibe resultados com maior autocorrelação que aqueles do esquema **III**. Contudo, ao usar o espaçamento ( $Thin=22$ ) na amostra de **I**, os fatores de ineficiência resultam melhores que os do esquema **III**. Este resultado reforça a nossa premissa que com o esquema **I** podemos gerar amostras com maior autocorrelação, mas que isto, na prática, não constitui um problema porque podemos realizar muitas mais iterações no mesmo tempo computacional.

### 3.4 Considerações Finais

Neste capítulo mostramos que o algoritmo proposto por West et al. (1985), o *Conjugate Updating*, pode ser utilizado com resultados bastante satisfatórios para construir uma densidade proposta independente, para o passo de Metropolis-Hastings, para amostrar em bloco, todo o vetor de estados de um modelo dinâmico generalizado. Realizamos um estudo de Monte Carlo para comparar a performance de nossa proposta versus outras existentes na literatura. Os resultados mostraram que o *CUBS* trabalha bastante bem, fornece bons resultados e tem como vantagens: a simples implementação e o baixo custo computacional. Embora o modelo dinâmico utilizado para nossa simulação seja dos mais simples, este resultado é extensível para modelos mais complexos, pois alguns resultados não apresentados aqui mostraram que o *CUBS* funciona bem mesmo com modelos não lineares complexos, como funções de transferência, onde, por exemplo, o algoritmo **IV** não deve ser utilizado e o esquema **III** é difícil de implementar e demanda muito esforço computacional.

Como foi mostrado neste capítulo, a variável resposta pode ser uma série temporal, que é modelada segundo (3.1) onde os parâmetros variam ao longo do tempo. Mas o *CUBS* também pode ser utilizado quando a variável resposta é uma amostra i.i.d. que segue um modelo com pa-

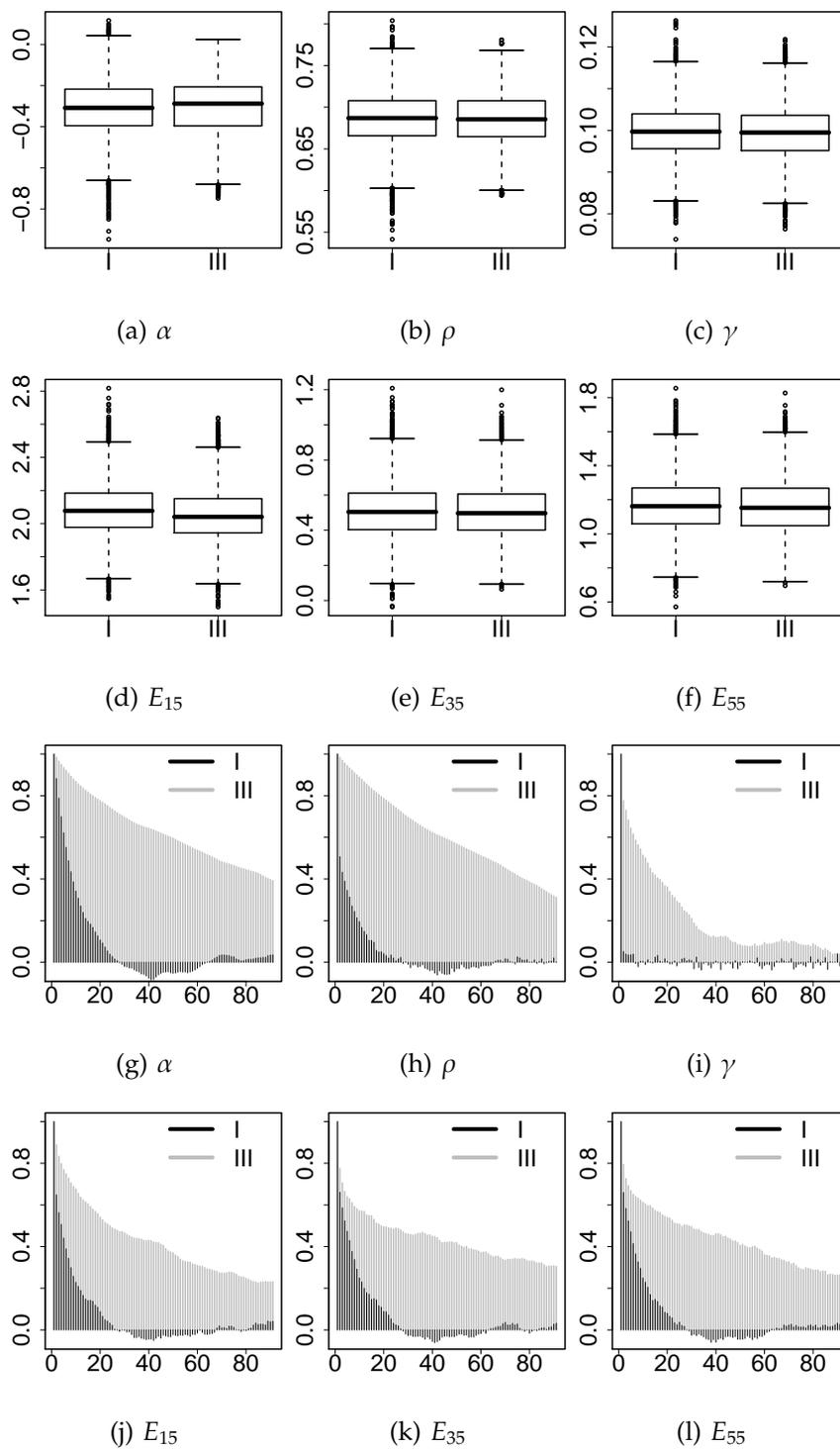


Figura 3.5: *Box plots* e Funções de Autocorrelação das amostras de alguns dos parâmetros em (3.9) para a Vazão no Rio Fartura (SP).

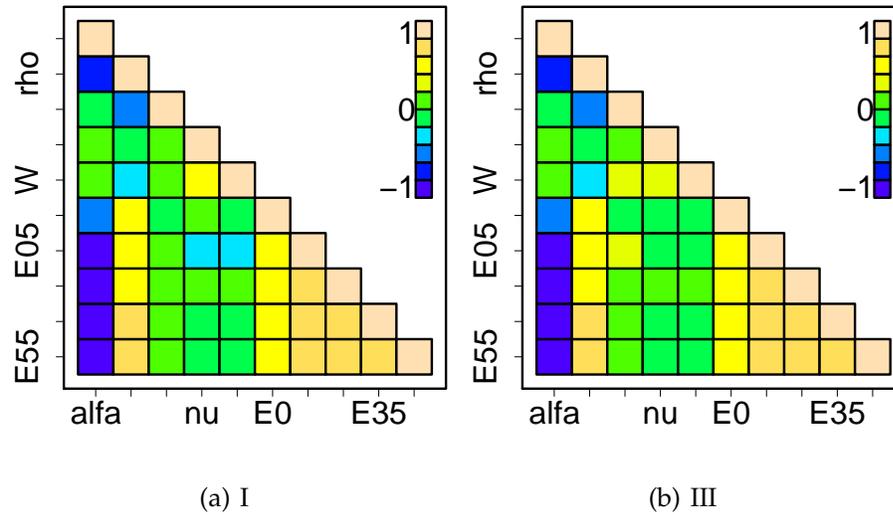


Figura 3.6: Funções de Correlação Cruzadas das amostras de alguns dos parâmetros em (3.9) para a Vazão no Rio Fartura (SP).

râmetros fixos. Neste caso, amostras da distribuição a posteriori podem ser geradas com um passo de Metropolis-Hastings onde a densidade proposta é construída usando os últimos dois momentos obtidos com o passo de filtragem. A idéia é estimar os parâmetros fixos de forma seqüencial, ou seja, supor uma ordem nos dados e incorporar as observações uma a uma no processo de atualização. Ao final, os momentos obtidos levarão em conta a amostra total e a moda da distribuição proposta estará próxima da moda da distribuição a posteriori de interesse.

Embora a ênfase da aplicação do *CUBS* resida nas distribuições na família exponencial, este algoritmo é mais geral porque o que realmente é necessário é uma distribuição a priori, de preferência conjugada, para o parâmetro de posição da distribuição da variável resposta. Por exemplo, o *CUBS* pode ser utilizado com a distribuição log-normal.

Um dos tópicos que atualmente estamos pesquisando é a aplicação do esquema proposto a distribuições  $k$ -paramétricas. A idéia é usar a distribuição a priori multivariada para a família exponencial (Bernardo & Smith, 1994, Cap. 5) para aproximar os primeiros momentos das distribuições *on-line* dos parâmetros de estado. Alguns detalhes desta abordagem são apresentados no apêndice C.

Tabela 3.5: Fatores Potenciais de Redução de Escala, PSRF, e Fatores de Ineficiência, INEF, para alguns dos parâmetros no modelo em (3.9) para a vazão do Rio Fartura

	I						III			
	PSRF <sup>a</sup>		INEF por cadeia				PSRF <sup>a</sup>		INEF por cadeia	
	$\hat{R}$	97.5%	1 <sup>b</sup>	2 <sup>b</sup>	1 <sup>c</sup>	2 <sup>c</sup>	$\hat{R}$	97.5%	1 <sup>b</sup>	2 <sup>b</sup>
$\alpha$	1,01	1,03	5,12	5,11	2,87	2,48	1,02	1,08	4,77	4,89
$\rho$	1,00	1,01	4,58	4,57	2,36	2,14	1,14	1,52	4,68	4,97
$\gamma$	1,00	1,00	2,75	2,70	1,13	1,22	1,07	1,29	3,59	3,49
$\nu$	1,00	1,00	3,91	3,85	1,43	1,47	1,08	1,31	3,61	3,70
$W$	1,00	1,00	3,40	3,40	1,29	1,34	1,04	1,18	3,06	3,32
$E_0$	1,00	1,00	3,91	3,86	1,88	1,55	1,00	1,01	3,53	3,86
$E_5$	1,00	1,02	4,83	4,81	2,67	2,27	1,02	1,10	4,28	4,55
$E_{15}$	1,00	1,01	4,80	4,77	2,65	2,31	1,00	1,00	4,37	4,56
$E_{35}$	1,01	1,03	4,85	4,83	2,63	2,37	1,01	1,04	4,41	4,70
$E_{55}$	1,00	1,01	4,86	4,84	2,63	2,42	1,01	1,07	4,41	4,61

<sup>a</sup> Abreviação do inglês *Potential scale reduction factors* (Gelman & Rubin, 1992).

<sup>b</sup> Amostras guardadas em todas as iterações (*Thin*=1).

<sup>c</sup> Amostras guardadas a cada 22 iterações (*Thin*=22).

### 3.5 Apêndice: Algumas Equações para os Parâmetros a Priori e a Posteriori

Sejam  $r_t$  e  $s_t$  os parâmetros da priori conjugada para  $\eta_t$ . Sejam  $f_t$  e  $q_t$  os momentos a priori obtidos do modelo linear e  $f_t^*$  e  $q_t^*$  os momentos a posteriori. A tabela 3.6 mostra as equações que devem ser resolvidas para calcular  $f_t^*$  e  $q_t^*$  nas distribuições utilizadas mais freqüentemente. Na tabela 3.6,  $\gamma(\cdot)$  e  $\gamma'(\cdot)$  denotam as funções digama e trigama, respectivamente. Na última coluna são utilizadas as seguintes aproximações:  $\gamma(x) \approx \log(x)$  e  $\gamma'(x) \approx 1/x$  (veja Abramovitch & Stegun (1965) para maiores detalhes).

Tabela 3.6: Exemplo de equações para calcular os momentos a priori e a posteriori envolvidos no passo de atualização do CUBS.

Distribuição	Ligação	Momentos a priori e a posteriori	Aproximações
Binomial	logit	$f_t^a = \gamma(r_t) - \gamma(s_t)$	$\approx \log(r_t) - \log(s_t)$
		$g_t^a = \gamma'(r_t) - \gamma'(s_t)$	$\approx 1/r_t - 1/s_t$
		$f_t^* = \gamma(r_t + y_t) - \gamma(s_t + n_t - y_t)$	
		$g_t^* = \gamma'(r_t + y_t) + \gamma'(s_t + n_t - y_t)$	
Binomial	identity	$r_t = f_t/n_t[f_t/q_t(n_t - f_t) - 1]$	
		$s_t = (1 - f_t/n_t)[f_t/q_t(n_t - f_t) - 1]$	
		$f_t^* = n_t(r_t + y_t)/(r_t + s_t + n_t)$	
		$g_t^* = [f_t^*(n_t - f_t^*)]/(r_t + s_t + n_t + 1)$	
Poisson	log	$f_t^a = \gamma(r_t) - \log(s_t)$	$\approx \log(r_t) - \log(s_t)$
		$g_t^a = \gamma'(r_t)$	$\approx 1/r_t$
		$f_t^* = \gamma(r_t + y_t) - \log(s_t + 1)$	
		$g_t^* = \gamma'(r_t + y_t)$	
Poisson	identity	$r_t = f_t^2/q_t$	
		$s_t = f_t/q_t$	
		$f_t^* = (r_t + y_t)/(s_t + 1)$	
		$g_t^* = (\alpha + y_t)/(s_t + 1)^2$	
Gama	log	$f_t^a = \log(r_t) - \gamma(s_t + 1)$	$\approx \log(r_t) - \log(s_t + 1)$
		$g_t^a = \gamma'(s_t + 1)$	$\approx 1/(s_t + 1)$
		$f_t^* = \log(r_t + \nu y_t) - \gamma(s_t + \nu + 1)$	
		$g_t^* = \gamma'(s_t + \nu + y_t)$	

<sup>a</sup> métodos numéricos são necessário para obter  $r_t$  e  $s_t$ .

### 3.6 Apêndice: O Caso da Distribuição Gama

Nesta seção apresentamos o *CUBS* para o modelo dinâmico gama:

$$y_t | \mu_t, \nu \sim \text{Gama}(\mu_t, \nu), \quad t = 1, \dots, N \quad (3.10a)$$

$$\log(\mu_t) = F_t' \theta_t \quad (3.10b)$$

$$\theta_t = G_t \theta_{t-1} + w_t, \quad w_t \sim N[0, W] \quad (3.10c)$$

$$\theta_0 | D_0 \sim N(m_0, C_0) \quad (3.10d)$$

onde  $\text{Gama}(\mu_t, \nu)$  denota a distribuição Gama com valor esperado  $\mu_t$  e parâmetro de forma  $\nu$ .

#### Momentos para $\log(\mu_t)$

Supondo  $\nu$  conhecido, pela análise conjugada temos:  $(-\eta_t | r_t, s_t) \sim \text{Gama}(s+1, r)$ . Dado que  $\log(\mu_t) = -\log(-\eta_t) = g(\eta_t)$ , temos:

$$\begin{aligned} E[g(\eta_t)] &= E[-\log(-\eta_t)] = -E[\log(-\eta_t)] = \log r_t - \gamma(s_t + 1). \\ \text{Var}[g(\eta_t)] &= \text{Var}[-\log(-\eta_t)] = \text{Var}[\log(-\eta_t)] = \gamma'(s_t + 1). \end{aligned}$$

onde  $\gamma(x) = d \log \Gamma(x) / dx = \Gamma'(x) / \Gamma(x) \approx \log(x)$  é a função digama e  $\gamma'(x) \approx 1/x$  é a função trigama.

#### *CUBS* para o modelo dinâmico gama em (3.10)

(1). Faça  $t = 1$

(2). Calcule os momentos a posteriori,  $m_t$  e  $C_t$ :

(a) Calcule os momentos a priori de  $\theta_t$  e  $g(\eta_t)$ , com base no modelo:

$$\begin{aligned} \theta_t | D_{t-1} &\sim [a_t, R_t] & \text{onde: } a_t &= G_t m_{t-1}, & R_t &= G_t C_{t-1} G_t' + W_t \\ g(\eta_t) | D_{t-1} &\sim [f_t, q_t] & \text{onde: } f_t &= F_t' a_t, & q_t &= F_t' R_t F_t \end{aligned}$$

(b) Determine os momentos da prior de  $g(\eta_t)$  com base na análise conjugada:

$$\begin{aligned} E[g(\eta_t) | D_{t-1}] &= f_t = \log r_t - \gamma(s_t + 1) \approx \log r_t - \log(s_t + 1) \\ \text{Var}[g(\eta_t) | D_{t-1}] &= q_t = \gamma'(s_t + 1) \approx 1/(s_t + 1) \end{aligned}$$

(c) Obtenha os momentos a posteriori de  $g(\eta_t)$ , com base na análise conjugada:

$$E[g(\eta_t)|D_t] = f_t^* = \log(r_t + \nu y_t) - \gamma(s_t + \nu + 1)$$

$$Var[g(\eta_t)|D_t] = q_t^* = \gamma'(s_t + \nu + 1)$$

(d) Calcule os momentos a posteriori de  $\theta_t$ , com base no modelo:

$$\theta_t|D_t \sim [\mathbf{m}_t, \mathbf{C}_t], \quad \text{onde:}$$

$$\mathbf{m}_t = \mathbf{a}_t + \mathbf{R}_t \mathbf{F}_t (f_t^* - f_t) \frac{1}{q_t}$$

$$\mathbf{C}_t = \mathbf{R}_t - \mathbf{R}_t \mathbf{F}_t \mathbf{F}_t' \mathbf{R}_t \left(1 - \frac{q_t^*}{q_t}\right) \frac{1}{q_t}.$$

(3). Faça  $t = t + 1$  e volte para (2)(c)(ii) se  $t < N$ ;

(4). Amostre  $\theta_N$  de  $N(m_N, C_N)$ ;

(5). Faça  $t = N - 1$  e amostre  $\theta_t$  de  $p(\theta_t | \theta_{t+1}, D_t, \theta) = N(\mathbf{m}_t^s, \mathbf{C}_t^s)$ ;

(6). Faça  $t = t - 1$  e volte para (5) se  $t > 1$ .

As etapas (2)b e (2)d acima, são específicas para cada família conjugada. As etapas (4)-(6) constituem o *backward sampling* (BS). Os momentos de  $p(\theta_t | \theta_{t+1}, D_t, \theta)$  são obtidos de

$$\theta_t|D_t \sim N[\mathbf{m}_t, \mathbf{C}_t]$$

$$\theta_{t+1}|\theta_t, D_t \sim N[\mathbf{G}_{t+1}\theta_t, \mathbf{G}_{t+1}\mathbf{C}_t\mathbf{G}_{t+1}' + \mathbf{W}_{t+1}]$$

$$\left( \begin{array}{c} \theta_{t+1} \\ \theta_t \end{array} \middle| D_t \right) \sim N \left[ \left[ \begin{array}{c} \mathbf{G}_{t+1}\mathbf{m}_t \\ \mathbf{m}_t \end{array} \right], \left[ \begin{array}{cc} \mathbf{G}_{t+1}\mathbf{C}_t\mathbf{G}_{t+1}' + \mathbf{W}_{t+1} & \mathbf{G}_{t+1}\mathbf{C}_t \\ \mathbf{C}_t\mathbf{G}_{t+1} & \mathbf{C}_t \end{array} \right] \right].$$

Seja  $\mathbf{R}_{t+1} = \mathbf{G}_{t+1}\mathbf{C}_t\mathbf{G}_{t+1}' + \mathbf{W}_{t+1}$  e  $\mathbf{B}_t = \mathbf{C}_t\mathbf{G}_{t+1}'\mathbf{R}_{t+1}^{-1}$ , logo:

$$\theta_t|\theta_{t+1}, D_t \sim N(\mathbf{m}_t^s, \mathbf{C}_t^s),$$

onde

$$\mathbf{m}_t^s = \mathbf{m}_t + \mathbf{B}_t(\theta_{t+1} - \mathbf{G}_{t+1}\mathbf{m}_t)$$

$$\mathbf{C}_t^s = \mathbf{C}_t - \mathbf{B}_t\mathbf{R}_{t+1}\mathbf{B}_t'.$$

## Capítulo 4

### UMA ABORDAGEM BAYESIANA DA RELAÇÃO CHUVA-VAZÃO

Neste capítulo apresentamos uma proposta de modelagem conjunta das variáveis chuva e vazão. Ilustramos nossa proposta com dados da bacia do Rio Grande, uma bacia de grandes proporções localizada no Nordeste do Brasil. Nossa abordagem está baseada nos modelos dinâmicos. Para a vazão consideramos funções de transferência sem impor restrições de normalidade e estacionariedade. Para a chuva consideramos o problema de mudança de suporte a partir de um modelo espaço-temporal. Utilizamos inferência bayesiana e métodos Monte Carlo via cadeias de Markov (*MCMC*) para obter amostras das distribuições a posteriori de todas os parâmetros desconhecidos.

#### 4.1 Introdução

Das fases do ciclo hidrológico, talvez a mais importante seja a do escoamento superficial, que tem origem, fundamentalmente, nas precipitações (Pinto et al., 1976). A vazão, ou volume escoado por unidade de tempo, é a principal grandeza que caracteriza um escoamento e é usualmente, expressa em metros cúbicos por segundo. Um dos desafios que os hidrólogos e operadores de recursos hidráulicos devem encarar é a previsão de vazão dada a chuva numa determinada região. A natureza incerta, de forma intrínseca, destes processos faz com que a abordagem bayesiana seja a mais apropriada sempre que problemas estatísticos sejam considerados (Rios-Insua et al., 2002).

Este capítulo apresenta uma estratégia alternativa para lidar com as características espaço temporais de duas das mais importantes variáveis hidrológicas: chuva e vazão. Nossos objetivos são: (i) modelar conjuntamente essas variáveis, levando em conta que são medidas em unidades de espaço diferentes e (ii) usar modelos estocásticos cujos parâmetros tenham interpretação física.

Várias classes de modelos estocásticos têm sido propostos para representar a relação chuva-vazão, seja com base nos modelos determinísticos ou na análise clássica de séries temporais. Duas das mais importantes classes de modelos estocásticos aplicados a modelagem da vazão dos rios são: funções de transferência e de mudança de regimes. A modelagem com funções de transferência é relativamente simples além de muito flexível e tem sido utilizada principalmente na

forma dos modelos *ARMAX* (Sales (1989) e Capkun et al. (2001), entre outros). Os modelos de Cadeias de Markov com mudanças de regime, como em Lu & Berliner (1999), por sua vez, têm sido incorporados recentemente. Estas abordagens nem sempre são adequadas porque podem requerer o uso de dados transformados e seus parâmetros não têm interpretação clara para os especialistas em hidrologia. Em todos os modelos propostos anteriormente, a incerteza associada à medição da chuva não foi levada em conta explicitamente. Além disso, a modelagem da relação chuva-vazão sempre foi realizada para bacias completas, considerando que ambas variáveis representam dados de áreas. Na prática, a precipitação é medida em mais de um ponto (pluviômetro) dentro de uma bacia, depois é utilizado algum procedimento para aproximar a precipitação correspondente a toda a região. Existem métodos que usam uma média ponderada das observações realizadas, onde os pesos são calculados com base no tamanho de polígonos que determinam a área de influência de cada posto. Entretanto, este tipo de procedimento não leva em conta a incerteza associada a essa medição.

Neste capítulo apresentamos uma alternativa para a modelagem conjunta de ambas variáveis. Para a chuva usamos modelos espaço-temporais, como aquele em Sansó & Guenni (2000). Para a vazão, usamos modelos dinâmicos não normais e não lineares, em particular estendemos os modelos apresentados em Migon & Monteiro (1997). Adicionalmente, para aproximar a quantidade de chuva da bacia, resolvemos o problema de suporte implícito (ver Gelfand et al. (2001) para detalhes sobre o problema de mudança de suporte). Os modelos apresentados neste capítulo constituem uma representação parcimoniosa de um sistema complexo de processos físicos e todos os parâmetros utilizados têm interpretação física (ou estão associados a uma característica física).

Os modelos formulados são analisados sob o paradigma bayesiano. Métodos Monte Carlo via cadeias de Markov (*MCMC*) são utilizados para avaliar as distribuições a posteriori das quantidades desconhecidas. Como as rotinas computacionais para os modelos propostos podem ser intensivas (demandar muito tempo computacional), consideramos o uso de métodos que executem milhares de iterações em poucos minutos. Em particular utilizamos a amostragem proposta no capítulo 3, para modelos dinâmicos generalizados.

O restante deste capítulo está organizado da seguinte forma. Na seção 4.2 descrevemos os dados da Bacia do Rio Grande (BA), utilizados para ilustrar nossa metodologia. A seção 4.3 é dedicada a uma discussão geral de nossa abordagem. Na seção 4.4 alguns detalhes a serem levados em conta no procedimento de inferência de nossos modelos são descritos. Na seção 4.5 discutimos os resultados da análise dos dados. Na seção 4.6 apresentamos algumas considerações finais e possíveis extensões desta modelagem.

## 4.2 Bacia do Rio Grande (BA), Brasil.

A bacia hidrográfica do Rio Grande está localizada dentro da bacia do Rio São Francisco, na região do médio São Francisco, no oeste baiano. A região a ser estudada encontra-se entre os paralelos  $11^{\circ}$  e  $13^{\circ}$  sul e os meridianos de  $43^{\circ}30'$  e  $46^{\circ}30'$  oeste. A bacia do Rio Grande ocupa uma extensão territorial da ordem de  $78900 \text{ km}^2$ , que representa aproximadamente 13% da bacia do São Francisco. A região tem clima tropical. O trimestre de maior estiagem ocorre em julho, agosto e setembro, enquanto que o trimestre mais chuvoso vai de janeiro a março. O mês de maior precipitação média, na bacia como um todo, é março e o de menor precipitação média é agosto (Di Bello, 2005).

Os dados utilizados neste trabalho foram fornecidos pelo Laboratório de Hidrologia da Universidade Federal do Rio de Janeiro (UFRJ). São séries históricas obtidas após procedimentos de consistência e sem observações faltantes, de dados mensais de vazão no posto Taguá (área de drenagem de  $37522,48 \text{ km}^2$ ) e chuva de 9 postos pluviométricos, espalhados irregularmente pela bacia, entre agosto de 1984 a setembro de 2004 (242 meses). A figura 4.1(a) mostra a localização dos postos e a figura 4.1(b) mostra 4 das 10 séries temporais disponíveis, as três primeiras de chuva e a quarta da vazão em Taguá.

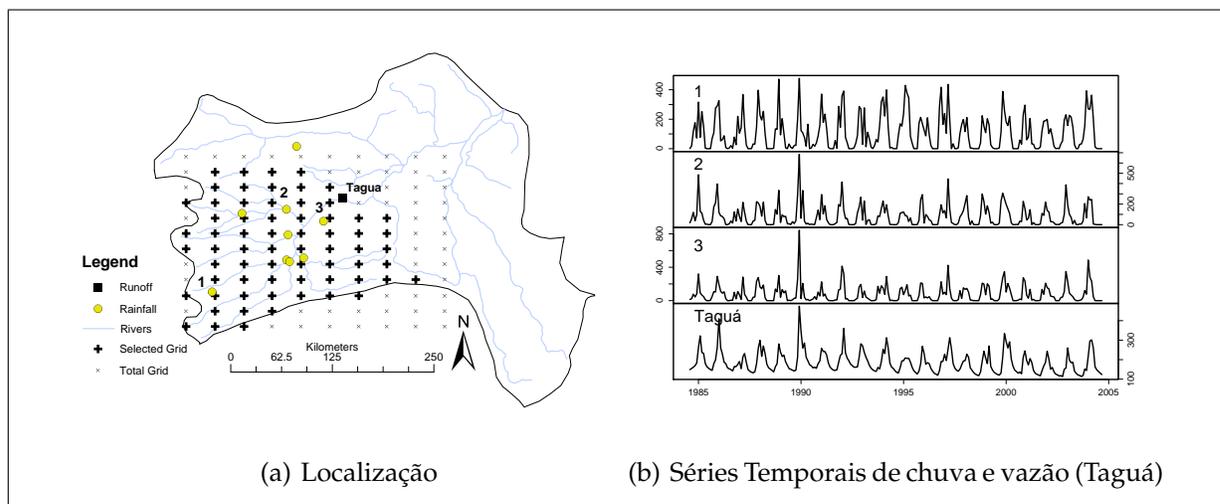


Figura 4.1: **Bacia do Rio Grande:** (a) Localização das estações monitoradoras; (b) Vazão mensal no posto Taguá e Chuva nos postos 2, 3 e 4 (loais marcados em (a)).

### 4.3 Abordagem Proposta

Duas das principais características da relação chuva-vazão são: a relação é não linear e a vazão do período atual depende da vazão do período anterior e da chuva passada e corrente. Pode-se assumir que não existe retroalimentação (*feedback*) entre vazão e chuva; logo, uma função de transferência é uma alternativa natural para ajustar e prever estes fenômenos. Além disso, a vazão é uma variável não negativa e suas séries temporais podem não ser estacionárias. Conseqüentemente, modelos dinâmicos não normais e não lineares parecem bastante apropriados para trabalhar com este tipo de dados.

Se  $Y_t$  é a vazão e  $X_t$  é a chuva acumulada no tempo  $t$ , a relação chuva-vazão pode ser representada por

$$Y_t \sim p(Y_t|\mu_t, \phi_t), \quad t = 1, 2, \dots \quad (4.1a)$$

$$g(\mu_t) = f_1(\alpha_t, E_t) \quad (4.1b)$$

$$E_t = f_2(E_{t-1}, \dots, E_0, X_t), \quad (4.1c)$$

onde  $p(Y_t|\mu_t, \phi_t)$  é uma função de densidade para variáveis aleatórias não negativas,  $\mu_t$  é o valor esperado de  $Y_t$ ,  $\phi_t$  denota os outros parâmetros de  $p(Y_t|\mu_t, \phi_t)$ ,  $\alpha_t$  é um nível básico e  $E_t$  é o efeito total da chuva no instante  $t$ ,  $g(\cdot)$ ,  $f_1(\cdot)$  e  $f_2(\cdot)$  são funções conhecidas que descrevem a dinâmica dos processos hidrológicos. Modelos com os parâmetros variando no tempo ou com variações estocásticas afetando  $E_t$  são casos particulares de (4.1).

#### 4.3.1 Um Modelo Dinâmico de Função de Transferência

De acordo com as hipóteses estabelecidas em Migon & Monteiro (1997), a relação entre a vazão e a chuva pode ser bem representada através de uma função de transferência, como descrevemos a seguir. O modelo em (4.1) admite que o valor esperado de vazão média gerada,  $\mu_t$ , ou uma função dela, digamos  $g(\mu_t)$ , pode ser representado como um nível básico,  $\alpha_t$ , mais o efeito da chuva passada e corrente,  $E_t$ , isto é,  $\mu_t = \alpha_t + E_t$ . Além disso, espera-se que o efeito da precipitação decaia entre os instantes  $t - 1$  e  $t$  à taxa  $\rho_t \in (0, 1)$ . Como  $E_{t-1}$  representa o efeito da precipitação anterior ao instante  $t$ , uma fração da chuva corrente, digamos  $\gamma_t X_t$ , pode ser acrescentada para estimar o efeito total da chuva no instante  $t$ . Adicionalmente, se  $\vartheta_t$  é o efeito máximo esperado da precipitação, então  $\vartheta_t > \mu_t$  e o restante de vazão possível é  $\vartheta_t - (\alpha_t + \rho_t E_{t-1})$ ; em conseqüência,  $E_t$ ,

em (4.1c), pode ser representado por alguma das seguinte alternativas:

$$E_t = \rho E_{t-1} + \gamma X_t \quad (4.2a)$$

$$E_t = \rho E_{t-1} + [1 - \exp(-\kappa X_t)][\vartheta_t - (\alpha_t + \rho E_{t-1})]. \quad (4.2b)$$

As equações (4.2a) e (4.2b) suportam a hipótese de que o efeito da precipitação decai exponencialmente com o tempo. Na equação (4.2a), quanto maior é a quantidade de chuva, maior é seu efeito sobre a vazão; esta hipótese é conhecida como retornos proporcionais às entradas. Por outro lado, na equação (4.2b), quanto maior é a quantidade de chuva, menor é seu efeito e também tem um limite superior; esta hipótese é conhecida como retornos decrescentes. A figura 4.2 ilustra as hipóteses aqui mencionadas.

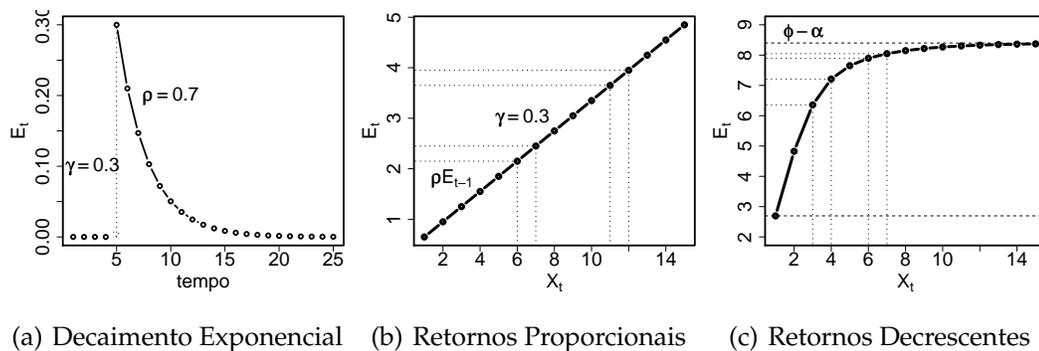


Figura 4.2: **Hipóteses sobre as Funções de Transferência**

### Interpretação dos Parâmetros

Nesta subseção associamos os parâmetros  $\alpha$ ,  $\rho$  e  $\gamma$  em (4.1) e (4.2a) a algumas das características físicas do processo chuva-vazão. O parâmetro  $\alpha$  representa o nível básico ou **fluxo base**, que depende do nível do lençol freático de cada bacia. Em geral, este fluxo base pode ser considerado constante. Já o parâmetro  $\rho$  pode ser considerado um **fator de recarga** ou também taxa de memorização do efeito da chuva. Este fator depende das características geológicas da bacia e, portanto, não deve variar muito ao longo do tempo. Variações neste parâmetro podem ser causadas por mudanças radicais no solo, como a redução significativa da vegetação.

Finalmente, o parâmetro  $\gamma$  é o **efeito instantâneo** da chuva e está associado principalmente à velocidade de escoamento superficial. Este parâmetro tem uma dinâmica particular. Está forte-

mente associado à taxa de infiltração do solo e/ou interceptação da chuva pela vegetação, a qual pode ser influenciada pelo regime de chuvas. Assim, após um período chuvoso, o solo estará saturado e o escoamento superficial será maior. Entretanto, após um período seco a tendência do solo será de absorver as águas da chuva. Este parâmetro também pode ser afetado pela presença ou não de vegetação na região. Quando a vegetação cresce passa a haver maior densidade de folhas e, portanto, maior interceptação da precipitação e menor efeito instantâneo na vazão.

### 4.3.2 Modelando a Chuva

Note que a entrada  $X_t$  no modelo (4.1) corresponde à precipitação na bacia toda, ou seja, uma única medida de chuva é necessária a cada instante do tempo. Entretanto, em muitas situações reais, a precipitação é medida em mais de uma estação localizada dentro de uma bacia. Portanto, a chuva total,  $X_t$ , deve ser obtida a partir da solução do problema de mudança de suporte. Use mudança de suporte quando se tem interesse na inferência sobre variáveis em unidades de área (*block data*) diferentes daquelas onde foi observada (Gelfand et al., 2001). A chuva de uma região (*Areal rainfall*) pode ser vista como uma média entre os dados de chuva pontuais observados porque é um processo espacial contínuo.

Seja  $\{X_t(s), s \in B \subset \mathbb{R}^2, t = 1, 2, \dots\}$  um campo espacial aleatório no instante  $t$  e na localização  $s$ . Aqui,  $X_t(s) \geq 0$  é uma variável aleatória que representa a quantidade de chuva no instante  $t$  na localização  $s$ . Então, a chuva numa bacia ou região  $B$ ,  $X_t$ , será descrita por

$$X_t = |B|^{-1} \int_B X_t(s) ds, \quad (4.3)$$

onde  $|B|$  é a área da bacia. Em particular,  $X_t(s)$  pode seguir uma distribuição normal truncada e, como sugerido em Sansó & Guenni (2000), pode ser representada pelo modelo espaço-temporal seguinte:

$$X_t(s_i) = \begin{cases} w_t(s_i)^\beta & \text{se } w_t(s_i) > 0, \\ 0 & \text{se } w_t(s_i) \leq 0 \end{cases} \quad s_i \in B \quad (4.4a)$$

$$w_t(s) = \boldsymbol{\theta}_t f(s) + Z_t(s) + \epsilon_t(s) \quad (4.4b)$$

$$Z_t(s) \sim GP(\mathbf{0}, \sigma^2 \rho(\|s_1, s_2\|, \lambda)), \quad (4.4c)$$

onde  $s$  é o vetor de localizações,  $w_t(s)$  é uma variável normal latente descrita pela soma de três componentes: (a)  $\boldsymbol{\theta}_t f(s)$ , uma tendência polinomial, (b)  $Z_t(s)$ , um processo com correlação espacial e (c)  $\epsilon_t(s)$ , um efeito aleatório com variância denotada por efeito pepita (*nugget effect*) (Cressie,

1993). O parâmetro  $\sigma^2$  é a variância do processo  $Z(s)$  e  $\varrho(\|s_1 - s_2\|, \lambda)$  representa uma função de correlação dependente do parâmetro  $\lambda$ . Uma forma alternativa de modelar a chuva é o uso de uma mistura de distribuições, como em Velarde et al. (2004) e Fernandes (2005), para levar em conta o excesso de zeros dos períodos secos.

Nossa proposta consiste em realizar o ajuste das equações (4.1) e (4.4) simultaneamente. Esta proposta pode ser vista como um sistema simples mas completo e eficiente para ajustar e prever duas das mais importantes variáveis hidrológicas. A especificação do modelo apresentado envolve várias subclasses de modelos. É bastante flexível e todos os parâmetros têm interpretação física clara. Além disso, toda a incerteza envolvida nos processos físicos é explicitamente levada em conta.

#### 4.4 Procedimento de Inferência

Suponha que dispomos de dados de vazão observados durante  $T$  períodos de tempo e de dados de chuva observados em  $S$  localizações (no espaço) na bacia  $B$  durante o mesmo período de tempo, isto é,  $\mathbf{Y} = (Y_1, \dots, Y_T)'$  e  $\mathbf{X}(s) = (X_1(s), \dots, X_T(s))'$ , onde  $s$  é o vetor de localizações  $(s_1, \dots, s_S)$ . Temos também  $\mathbf{X}_t(s) = (X_t(s_1), \dots, X_t(s_S))'$  e  $\mathbf{X}(s_i) = (X_1(s_i), \dots, X_T(s_i))'$ ,  $s_i \in B, i = 1, \dots, S$ . Além disso, suponha que  $\mathbf{X}$  denota a série temporal da chuva da bacia (de área), isto é,  $\mathbf{X} = (X_1, \dots, X_T)'$ . A distribuição conjunta de  $\mathbf{Y}, \mathbf{X}$  e  $\mathbf{X}(s)$  é dada por

$$p(\mathbf{Y}, \mathbf{X}, \mathbf{X}(s) | \Theta) = p(\mathbf{Y} | \mathbf{X}, \mathbf{X}(s), \Theta_Y) p(\mathbf{X}, \mathbf{X}(s) | \Theta_X), \quad (4.5)$$

onde  $\Theta = (\Theta_Y, \Theta_X)$ ,  $\Theta_Y$  denota os parâmetros em (4.1) e  $\Theta_X$  denota os parâmetros em (4.4). Também

$$\begin{aligned} p(\mathbf{Y}, \mathbf{X}, \mathbf{X}(s) | \Theta) &= \prod_{t=1}^T p(Y_t | X_t, \mathbf{X}_t(s), \Theta_Y) p(X_t, \mathbf{X}_t(s) | \Theta_X) \\ &= \prod_{t=1}^T p(Y_t | X_t, \mathbf{X}_t(s), \Theta_Y) p(X_t | X_t(s), \Theta_X) \prod_{i=1}^S p(X_t(s_i) | \Theta_X). \end{aligned} \quad (4.6)$$

Gelfand et al. (2001) propuseram aproximar  $p(X_t, \mathbf{X}_t(s) | \Theta_X)$  usando integração Monte Carlo. Para isso, eles propuseram amostrar um conjunto de observações em  $S_B$  localizações, independente e uniformemente distribuídas sob  $B$ , e calcular (amostrar)

$$\hat{X}_t = \frac{1}{S_B} \sum_{i=1}^{S_B} \hat{X}_t(s_i) \quad i = 1, \dots, S_B, \quad (4.7)$$

onde  $\hat{X}_t(s_i)$  é o valor predito de chuva para a  $i$ -ésima localização de uma grade regular de interpolação de  $S_B$  localizações, construída dentro dos limites da bacia  $B$ . Por sua vez (5.11) é a aproximação de Monte Carlo de (4.3).

A distribuição preditiva, necessária para interpolar a chuva,  $X_t(s_i)$ , num novo conjunto de localizações, por exemplo,  $X_t(s') = (X_t(s'_1), \dots, X_t(s'_N))'$ , é dada por:

$$p(X(s')|X(s)) = \int p(X(s')|X(s), \Theta_X) p(\Theta_X|X(s)) p(\Theta_X), \quad (4.8)$$

onde  $\Theta_X$  contém todos os parâmetros em (4.4).

De acordo com o paradigma bayesiano, a especificação do modelo fica completa somente depois da atribuição da distribuição a priori das quantidades desconhecidas. Após isso, amostras das distribuições a posteriori podem ser obtidas via métodos *MCMC* (Gamerman & Lopes, 2006). Dadas as expressões acima, o procedimento de inferência via *MCMC* pode ser realizado nas etapas seguintes:

- (1). Ajustar um modelo espaço-temporal para a chuva,  $X(s)$ , observada em  $S$  localizações da bacia  $B$ .
- (2). Obter as amostras da chuva na bacia,  $X$ , a partir da amostra preditiva de  $X(s)$ , gerada no passo (1), para cada ponto da grade de interpolação.
- (3). Ajustar um modelo chuva-vazão para  $Y$ , usando como  $X$  a cadeia gerada no passo (2).

Em particular, para os modelos dinâmicos de vazão considerando a distribuição gama, propomos o uso do esquema de amostragem chamado *CUBS*, abreviação do inglês *Conjugate Updating Backward Sampling* (apresentado no capítulo 3), que combina o *Conjugate Updating* de West et al. (1985) para modelos dinâmicos na família exponencial, com o *Backward Sampling* de Frühwirth-Schnater (1994). Observamos que para modelos de função de transferência não-normais, *CUBS* reduz significativamente o tempo computacional necessário para atingir a convergência das cadeias, além de ser de fácil implementação.

Sob esta abordagem, a inferência sobre eventos relacionados, por exemplo, a probabilidade de períodos secos ou a probabilidade de que a vazão exceda algum valor pre-fixado, pode ser realizada de maneira relativamente simples.

## 4.5 Modelagem na Prática

Aplicamos a abordagem descrita na seção 4.3 aos dados de chuva medidos nas 9 estações pluviométricas e à série de vazões da estação fluviométrica Taguá, todas pertencentes à bacia do Rio Grande. Especificamente, utilizamos a função em (4.2a) para o efeito  $E_t$  em (4.1c) e um modelo dinâmico normal linear multivariado (ver West & Harrison (1997, Cap. 16)) para a evolução temporal dos parâmetros em (4.4). Para uma melhor exposição, reproduzimos o nosso modelo completo em (4.9):

$$Y_t|X_t \sim p(\mu_t, \phi) \quad t = 1, \dots, T \quad (4.9a)$$

$$\log(\mu_t) = \alpha_t + E_t \quad (4.9b)$$

$$E_t = \rho E_{t-1} + \gamma_t X_t + w_t \quad w_t \sim N(0, \sigma_E^2) \quad (4.9c)$$

$$\alpha_t = \mathbf{G}_\alpha \alpha_{t-1} + w_{\alpha,t} \quad w_{\alpha,t} \sim N(0, \sigma_\alpha^2) \quad (4.9d)$$

$$\gamma_t = \mathbf{G}_\gamma \gamma_{t-1} + w_{\gamma,t} \quad w_{\gamma,t} \sim N(0, \sigma_\gamma^2) \quad (4.9e)$$

$$X_t = \frac{1}{S_B} \sum_{i=1}^{S_B} \hat{X}_t(s'_i) \quad i = 1, \dots, S_B \quad (4.9f)$$

$$X_t(s_i) = \begin{cases} w_t(s_i)^\beta & \text{se } w_t(s_i) > 0 \\ 0 & \text{se } w_t(s_i) \leq 0 \end{cases} \quad i = 1, \dots, N \quad (4.9g)$$

$$\mathbf{w}_t = \mathbf{z}_t + \mathbf{v}_t \quad \mathbf{v}_t \sim N_N(\mathbf{0}, \tau^2 \mathbf{I}) \quad (4.9h)$$

$$\mathbf{z}_t = \mathbf{F}' \boldsymbol{\theta}_t + \boldsymbol{\varepsilon}_t \quad \boldsymbol{\varepsilon}_t \sim N_N(\mathbf{0}, \sigma^2 \mathbf{V}_t) \quad (4.9i)$$

$$\boldsymbol{\theta}_t = \mathbf{G} \boldsymbol{\theta}_{t-1} + \boldsymbol{\varepsilon}_t \quad \boldsymbol{\varepsilon}_t \sim N_k(\mathbf{0}, \mathbf{W}_t), \quad (4.9j)$$

onde  $p(\mu_t, \phi)$  denota a distribuição log-normal ou gama e  $\phi$  corresponde ao parâmetro de precisão da log-normal ou o parâmetro de forma da gama. Em (4.9g)–(4.9i),  $N$  é o número de estações observadas e  $S$  é o número de pontos na grade de interpolação.  $X_t(s)$  denota a precipitação no instante  $t = 1, \dots, T$  na localização  $s = 1, \dots, N$ ,  $w_t(s)$  é uma variável gaussiana latente,  $\beta$  é uma potência desconhecida,  $\mathbf{w}_t$  é um vetor de dimensão  $N$  que aloca as observações das  $N$  localizações correspondentes ao tempo  $t$ ,  $\tau^2$  é o efeito pepita,  $\sigma^2 > 0$  e  $\mathbf{V}_t \in \mathbb{R}^{N \times N}$ .  $\mathbf{V}_t$  captura a correlação espacial; assumimos que  $V_{ij} = \exp(-\lambda d_{ij})$ , ou seja, uma correlação com decaimento exponencial onde  $\lambda$  controla a taxa de decaimento dessa correlação,  $\lambda > 0$  e  $d_{ij}$  é a distância euclidiana entre as localizações  $i$  e  $j$ ,  $i, j = 1, \dots, N$ . Em (4.9i)–(4.9j),  $\mathbf{F}' \in \mathbb{R}^{N \times k}$ ,  $\mathbf{G} \in \mathbb{R}^{k \times k}$  e  $\boldsymbol{\theta} \in \mathbb{R}^k$ . Os elementos de  $\boldsymbol{\theta}$  são tais que  $\boldsymbol{\theta}_t = (\boldsymbol{\theta}_{t1}, \boldsymbol{\theta}_{t2})'$ , onde  $\boldsymbol{\theta}_{t1}$  é um sub-vetor que descreve a tendência espacial e  $\boldsymbol{\theta}_{t2}$  descreve

os efeitos sazonais. As equações (4.9d) e (4.9e) representam possíveis evoluções no tempo dos parâmetros  $\alpha$  e  $\gamma$ , respectivamente. Na prática somente uma delas é ajustada e a escolha depende das características da região sob estudo.

#### 4.5.1 Distribuições a Priori e Condicionais Completas

Em geral, optamos pelo uso de distribuições a priori vagas. Contudo, dado que os parâmetros têm interpretação física clara, um processo de elicitação de prioris poderia ter sido utilizado. Para os parâmetros do modelo espaço-temporal em (4.9g)-(4.9j) consideramos todos os parâmetros independentes, a priori, isto é,  $p(\theta_0, \sigma^2, \zeta^2, \lambda, \beta) = p(\theta_0)p(\sigma^2)p(\zeta^2)p(\lambda)p(\beta)$ , onde  $\zeta^2 = \tau^2/\sigma^2$ ,  $p(\theta_0)$  é uma densidade normal  $N$ -variada com média  $\mathbf{0}$  e matriz variância identidade,  $N_N(\mathbf{0}, \mathbf{I})$  e  $p(\sigma^2)$  é uma distribuição imprópria igual a  $1/\sigma^2$ . Entretanto,  $p(\zeta^2)$ ,  $p(\lambda)$  e  $p(\beta)$  são densidades gama com parâmetros  $(0,001; 0,001)$ ,  $(2; 6/1, 86)$  e  $(12, 4)$ , respectivamente. Os hiperparâmetros para  $\lambda$  foram selecionados de acordo com a premissa que a correlação espacial seja praticamente zero a uma distância igual à metade da distância máxima entre os postos observados. Os hiperparâmetros para  $\beta$  foram escolhidos de forma que o seu valor esperado seja 3, representando a transformação por raiz cúbica amplamente recomendada na literatura hidrológica (Sansó & Guenni, 2000).

Seguindo o Teorema de Bayes, a distribuição a posteriori é proporcional à verossimilhança vezes a distribuição a priori. Para o modelo espaço-temporal em (4.9g)-(4.9j), a distribuição a posteriori é dada por

$$p(\sigma^2, \zeta^2, \lambda, \beta, \mathbf{z}, \boldsymbol{\theta}|\mathbf{x}) \propto \left(\frac{1}{\sigma^2}\right)^{NT} \left(\frac{1}{\zeta^2}\right)^{NT/2} |\mathbf{V}(\lambda)|^{-T/2} \exp\left(-\frac{1}{2\sigma^2} \sum_{t=1}^T \frac{1}{\zeta^2} \|\mathbf{w}_t - \mathbf{z}_t\|^2 + (\mathbf{z}_t - \mathbf{F}'\boldsymbol{\theta}_t)' \mathbf{V}(\lambda)^{-1} (\mathbf{z}_t - \mathbf{F}'\boldsymbol{\theta}_t) - \frac{1}{2} \sum_{t=1}^T (\boldsymbol{\theta}_t - \mathbf{G}\boldsymbol{\theta}_{t-1})' \mathbf{W}_t^{-1} (\boldsymbol{\theta}_t - \mathbf{G}\boldsymbol{\theta}_{t-1})\right) \left(\prod_{x_{it}>0} \frac{x_{it}^{1/\beta-1}}{\beta}\right) p(\theta_0, \sigma^2, \zeta^2, \lambda, \beta). \quad (4.10)$$

A partir de (4.10), as distribuições condicionais completas (d.c.c.) são:  $\sigma^2$  e  $\zeta^2$  são gamas invertidas,  $\mathbf{z}$  é normal multivariada,  $w_{ij} < 0$  é normal truncada univariada. As d.c.c. de  $\lambda$  e  $\beta$  não têm forma fechada conhecida. Como os  $\boldsymbol{\theta}_t$  são os estados de um modelo dinâmico normal, as suas d.c.c. também são normais multivariadas.

Para o modelo dinâmico em (4.9a)-(4.9c) atribuímos, também, distribuições a priori independentes. Em particular, consideramos distribuições normais com média zero e variância  $10^3$  para  $E_0$ ,  $\alpha_0$  e  $\gamma_0$  e uma distribuição uniforme entre  $[0,1]$  para  $\rho$ . Para todos os termos de variância,  $(\sigma_Y^2, \sigma_E^2, \sigma_W^2)$ , utilizamos distribuições gama invertidas com ambos parâmetros iguais a 0,01. Ao utilizar a distribuição gama para a vazão, consideramos uma distribuição gama com ambos parâmetros iguais a 0,01, como distribuição a priori para o parâmetro de forma, denotado por  $\phi$ . Neste

caso, as d.c.c. dos parâmetros em (4.9a)-(4.9c) dependem da distribuição assumida para  $Y_t$  e as hipóteses para  $\sigma_E^2, \sigma_\alpha^2$  e  $\sigma_\gamma^2$ . Em particular, se  $p(\mu_t, \phi)$  é uma distribuição gama e  $\sigma_\alpha^2 = \sigma_\gamma^2 = 0$ , as d.c.c de  $\gamma$  e  $\sigma_E^2$  são normal e gama invertida, respectivamente, e as d.c.c. de  $\alpha, \rho$  e  $\phi$  não têm forma fechada conhecida.

#### 4.5.2 Principais Aspectos Computacionais

Para obter amostras da distribuição a posteriori de interesse, usamos o amostrador de Gibbs (Gelfand & Smith, 1990). Para amostrar das d.c.c. de  $\lambda, \beta, \alpha$  e  $\rho$  utilizamos o método *slice sampling* (Neal, 2003). Fizemos uso de um passo de Metropolis-Hastings com uma proposta log-normal para amostrar o parâmetro de forma  $\phi$ . Obtivemos amostras de  $\theta_t$  usando o procedimento do *FFBS* (Frühwirth-Schnater, 1994) e de  $E_t$  usando o *CUBS*. Também, de forma análoga a Sansó & Guenni (2000), usamos fatores de desconto para  $W_t$ :  $\delta_T = 0,90$  para a tendência espacial e  $\delta_S = 0,95$  para os efeitos sazonais. Finalmente, para  $\sigma_\alpha^2$  e  $\sigma_\gamma^2$  usamos 0,95 como fator de desconto, quando as equações (4.9d) e (4.9d) foram incluídas nos modelos.

Realizamos 70000 iterações de duas cadeias paralelas do algoritmo *MCMC* para os modelos espaço-temporais, descartamos as primeiras 10000 iterações como período de aquecimento e, para quebrar possíveis auto-correlações, guardamos as amostras a cada 10 iterações. Para os modelos de vazão, rodamos duas cadeias paralelas com 60000 iterações depois do período de aquecimento de tamanho 10000. Neste caso, guardamos as amostras a cada 5 iterações. Escrevemos a rotinas computacionais em `ox` versão 3.20 (Doornik, 2002) e testamos a convergência das cadeias com a biblioteca *CODA*, desenvolvida por Plummer et al. (2005) para o software `R`.

#### 4.5.3 Resultados

Levando em conta a fatorização da verossimilhança, usamos os algoritmos para primeiro ajustar o modelo em (4.3) com várias especificações da tendência polinomial e depois ajustar vários casos particulares do modelo em (4.1).

O modelo espaço-temporal utilizado para a chuva tem um intercepto e um efeito linear da longitude na tendência espacial. Outros modelos testados indicaram que a latitude não tem um efeito significativo nessa região. O padrão sazonal está representado por dois harmônicos de Fourier que foram escolhidos depois de uma análise exploratória dos periodogramas das séries. Então,  $F_t$  em (4.9i) é um vetor com componentes:  $(1, \text{longitude}(s_i), 1, 0, 1, 0)'$  e  $G = \text{diag}(G_1, G_2)$ , onde  $G_1$  é uma

matriz identidade de ordem 2 e  $G_2$  tem blocos diagonais  $G_{2r} = \begin{pmatrix} \cos(2\pi r/12) & \sin(2\pi r/12) \\ -\sin(2\pi r/12) & \cos(2\pi r/12) \end{pmatrix}$ ,  $r = 1, 2$ .

A figura 4.3 apresenta as trajetórias estimadas para  $\theta_t$ . Observamos que o intercepto varia ao longo do tempo e aparentemente tem um ciclo inter-anual. O efeito da longitude é negativo e varia suavemente ao longo do tempo. O primeiro harmônico exibe um padrão bastante regular, porém, o efeito do segundo harmônico exibe dois períodos de comportamentos diferentes: antes e depois de 1992. A tabela 4.1 mostra as principais estatísticas das amostras das posteriores obtidas para os parâmetros estáticos nas equações (4.9g)-(4.9j). A média a posteriori estimada para  $\beta$  é 1,73, sugerindo que os dados são suavemente assimétricos, provavelmente porque os dados são mensais (dados semanais podem ser mais assimétricos). Na tabela 4.1 também observamos que a estatística  $\hat{R}$  (Gelman & Rubin, 1992) assume valores muito próximos de 1, o que sugere que a convergência de todas as cadeias foi atingida.

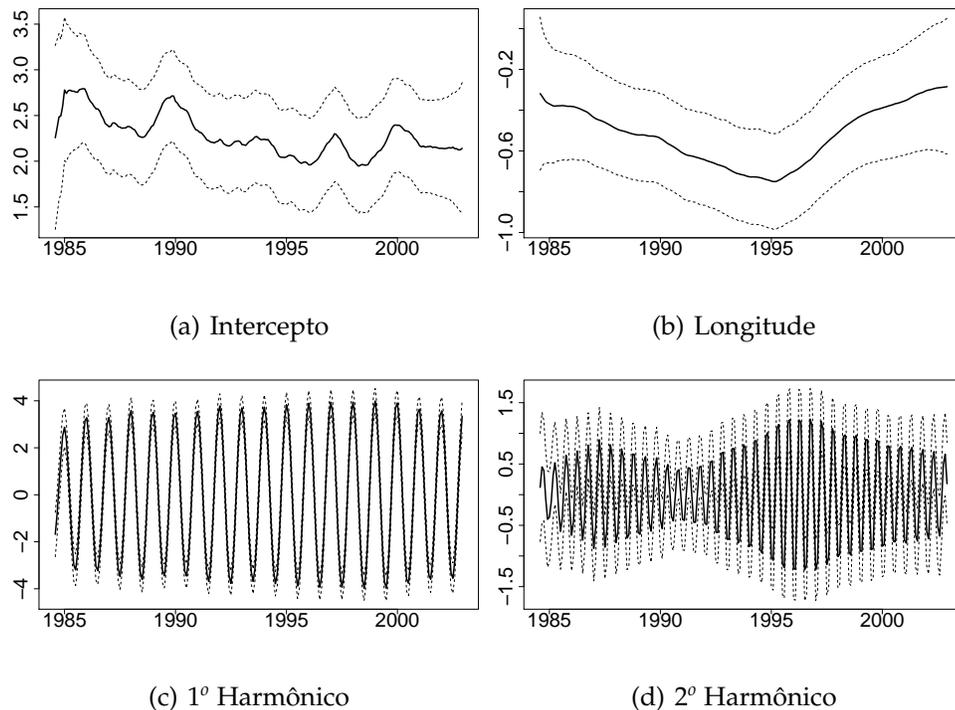


Figura 4.3: Trajetória estimada para os parâmetros dinâmicos do modelo da chuva em (4.9j). As linhas sólidas correspondem a média a posteriori e as linhas pontilhadas ao intervalo de 95% de credibilidade.

Tabela 4.1: Estatísticas a posteriori associadas com os parâmetros fixos em (4.9g)–(4.9j)

	média	d.p.	2,5%	25%	50%	75%	97,5%	$\hat{R}$
$\beta$	1,732	0,016	1,701	1,722	1,732	1,743	1,764	1,001
$\lambda$	0,045	0,007	0,033	0,040	0,044	0,050	0,061	1,001
$\zeta^2$	0,719	0,040	0,644	0,691	0,718	0,746	0,798	1,001
$\sigma^2$	1,100	0,044	1,015	1,070	1,098	1,128	1,191	1,003

Com a finalidade de ilustrar as interpolações obtidas com o modelo espaço-temporal ajustado, a figura 4.4 mostra a média a posteriori em dois meses selecionados. Observamos que o modelo reproduz padrões de chuva bastante diferentes, por exemplo, o padrão de um mês chuvoso como dezembro e de um mês seco, como junho.

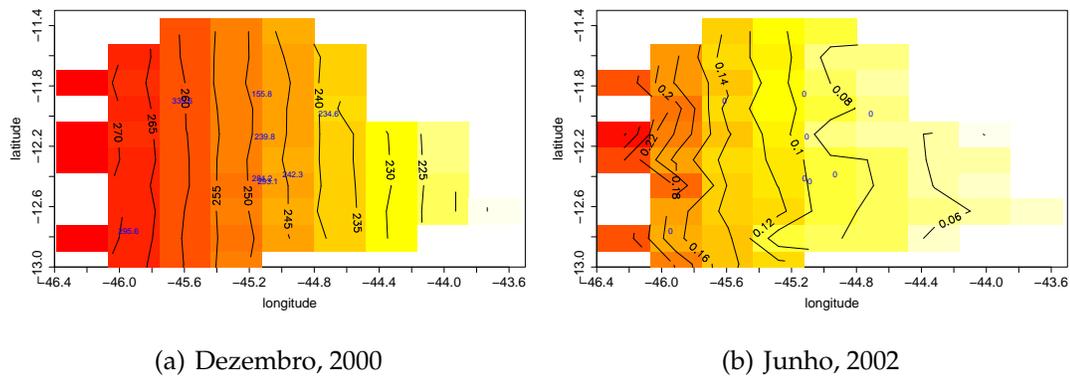
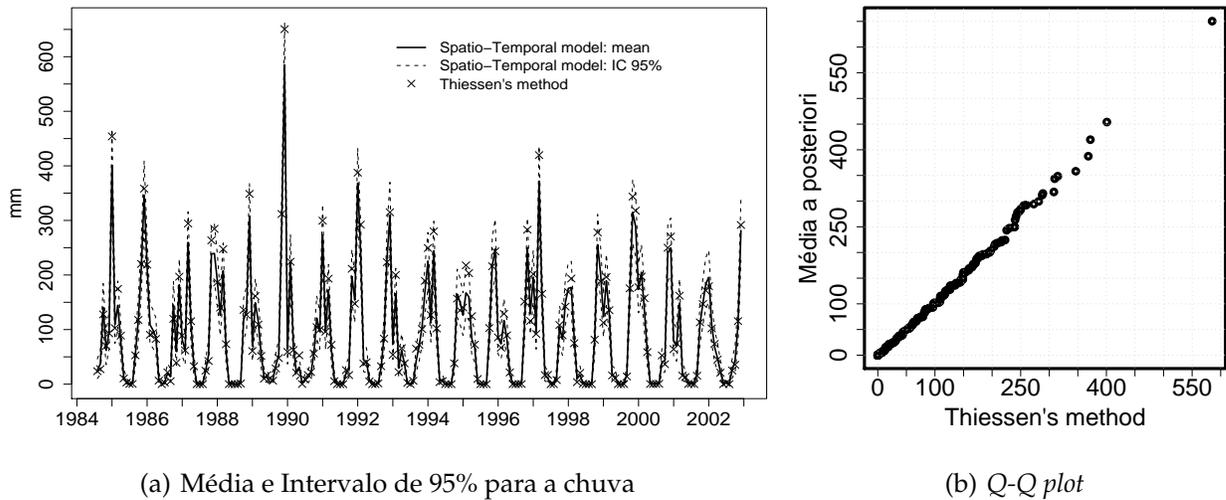


Figura 4.4: Médias a posteriori de chuva estimadas para dois meses diferentes. Os pontos sinalizam a localização das estações monitoradoras.

A precipitação para a bacia toda foi obtida por meio de interpolações espaciais numa grade de 63 pontos selecionados a partir de uma grade regular construída sobre a área de drenagem em estudo. Estas grades (regular e selecionada) são mostradas na figura 4.1(a). A integral em (4.9f) foi aproximada por Monte Carlo, pela média dos valores preditos nesses 63 pontos. A média e o intervalo de 95% de credibilidade para a chuva acumulada da região são apresentados na figura 4.5. Esta figura mostra também a precipitação estimada pelo método de Thiessen, método determinístico amplamente utilizado que consiste no cálculo de uma média ponderada

das precipitações observadas em cada posto. Vale a pena prestar atenção à figura 4.5 porque do ponto de vista bayesiano é possível levar em conta toda a incerteza associada aos fenômenos em estudo, o que nos permite construir um intervalo de credibilidade para cada instante do tempo e, mais ainda, considerar esta incerteza também durante o ajuste das equações para a vazão.



(a) Média e Intervalo de 95% para a chuva

(b) Q-Q plot

Figura 4.5: **Chuva na bacia.** A linha sólida corresponde a média a posteriori e as linhas pontilhadas ao intervalo de 95% de credibilidade. Os + correspondem as estimativas com o método de Thiessen.

Utilizamos a amostra a posteriori da chuva da bacia  $X_t$  para ajustar vários casos particulares das equações (4.9a)-(4.9c). Especificamente, consideramos duas alternativas para  $p(Y_t|X_t)$ : log-normal e gama, a função de transferência apresentada em (4.2a), e cinco alternativas para a dinâmica da vazão:

- (a) Fluxo base, função de transferência e efeito instantâneo da chuva estáticos:  $\sigma_\alpha^2 = \sigma_\gamma^2 = \sigma_E^2 = 0, \forall t$ ;
- (b) Fluxo base e efeito instantâneo da chuva estáticos, e função de transferência estocástica:  $\sigma_\alpha^2 = \sigma_\gamma^2 = 0$  e  $\sigma_E^2 > 0, \forall t$ ;
- (c) Fluxo base variando no tempo segundo um passeio aleatório, e função de transferência e efeito instantâneo da chuva estáticos:  $\alpha_t = \alpha_{t-1} + w_{\alpha,t}$ ,  $\sigma_\alpha^2 > 0$  e  $\sigma_\gamma^2 = \sigma_E^2 = 0, \forall t$ ;

- (d) Fluxo base estático, função de transferência estocástica e efeito instantâneo da chuva variando no tempo segundo um passeio aleatório:  $\gamma_t = \gamma_{t-1} + w_{\gamma,t}$ ,  $\sigma_\gamma^2 > 0$ ,  $\sigma_\alpha^2 = 0$  e  $\sigma_E^2 > 0$ ,  $\forall t$ ;
- (e) Fluxo base estático, função de transferência estocástica e efeito instantâneo da chuva variando no tempo segundo uma tendência constante e um padrão sazonal com dois harmônicos:

$$\gamma_t = \mathbf{G}_\gamma \gamma_{t-1} + w_{\gamma,t}, \sigma_\gamma^2 > 0, \sigma_\alpha^2 = 0 \text{ e } \sigma_E^2 > 0, \forall t. \mathbf{G}_\gamma = \text{diag}(1, \mathbf{G}_{2,\gamma}), \text{ onde}$$

$$\mathbf{G}_{2,\gamma} = \begin{pmatrix} \cos(2\pi r/12) & \sin(2\pi r/12) \\ -\sin(2\pi r/12) & \cos(2\pi r/12) \end{pmatrix}, r = 1, 2.$$

Vale a pena mencionar que também ajustamos a função em (4.2b), mas os resultados foram menos satisfatórios do que com a função em (4.2a), quando consideramos a bondade de ajuste de cada modelo. Realizamos a comparação dos modelos usando os seguintes critérios: (i) *DIC* (abreviação do inglês *Deviance Information Criterion*), proposto por Spiegelhalter et al. (2001); (ii) *EPD* (abreviação do inglês *Expected Predictive Deviation*), proposto por Gelfand & Ghosh (1998); (iii) erros quadráticos médios (*MSE*); e (iv) erros absolutos médios (*MAE*). Em todos os casos, quanto menor é o valor do critério, melhor é o ajuste do modelo. Ressaltamos que utilizamos a mesma amostra de chuva para toda a bacia ( $X$ ) para rodar os algoritmos *MCMC* para a vazão ( $Y$ ); logo, nossa seleção de modelos é realizada de forma condicional a  $X_t$ . A tabela 4.2 (colunas 4,7,8 e 9) mostra os valores do *DIC*, *EPD*, considerando uma perda quadrática, *MSE* e *MAE*, obtidos para cada uma das cinco especificações acima mencionadas utilizando as primeiras 221 observações. Duas conclusões podem ser extraídas: primeiro, todos os critérios sugerem a escolha da distribuição gama (esta conclusão não é válida quando consideramos as colunas 10 e 11); segundo, a especificação (e) fornece os melhores resultados em termos de ajuste.

Nosso modelo final para a vazão, então, utiliza uma distribuição gama para a resposta, com um fluxo base estático, uma função de transferência estocástica e um efeito instantâneo da chuva evoluindo no tempo seguindo uma tendência constante e um efeito sazonal representado por dois harmônicos. Na figuras 4.6(a) e 4.6(b) mostramos os histogramas das amostras das distribuições a posteriori de  $\alpha$  e  $\rho$ , respectivamente. Na figura 4.6(a) observamos que a média a posteriori de  $\alpha$  é 4,84, indicando que o fluxo base médio na região, durante o período observado, foi de 126,46  $m^3/s$ . Da figura 4.6(b) concluímos que o fator de recarga da região é em média igual a 0,64 e varia entre 0,57 e 0,71, que corresponde ao intervalo de 95% de credibilidade a posteriori. A figura 4.6(c) mostra a evolução no tempo do efeito instantâneo da chuva,  $\gamma_t X_t$ . Vale a pena lembrar que no modelo selecionado,  $\gamma_t$  é um vetor de dimensão 5, onde a primeira componente corresponde à tendência e as últimas 4 correspondem aos 2 harmônicos considerados. A figura 4.6(d) mostra a

Tabela 4.2: Critérios de comparação de modelos para 5 especificações alternativas de (4.9a)-(4.9c): DIC, EPD, Erro Quadrático Médio (MSE) e Erro Absoluto Médio (MAE).

Modelo	$\bar{D}$	$pd$	DIC	Ajuste	Penalidade	EPD	MSE <sup>a</sup>	MAE <sup>a</sup>	MSE <sup>b</sup>	MAE <sup>b</sup>
Distribuição log-normal para Vazão ( $Y_t$ )										
(a)	1917,6	27,7	1945,3	103475	94948	198423	429,1	14,9	1366,9	27,5
(b)	1733,0	99,9	1833,0	58577	33096	91674	149,9	7,4	1359,3	26,8
(c)	1799,8	50,0	1849,9	65485	64745	130231	292,1	10,3	2058,4	27,5
(d)	1757,5	86,3	1843,9	62068	40870	102939	186,5	8,4	1588,6	25,7
(e)	1887,2	120,1	2007,3	125683	60749	186433	274,2	9,7	1596,4	26,2
Distribuição gama para Vazão ( $Y_t$ )										
(a)	1917,8	16,8	1934,6	103157	94412	197570	427,5	14,8	1398,6	26,5
(b)	1718,6	110,5	1829,1	55038	29476	84514	134,3	7,0	1119,6	26,0
(c)	1810,7	27,3	1838,1	73879	61528	135407	280,2	10,0	2078,2	28,0
(d)	1722,9	108,6	1831,6	56077	30525	86602	139,0	7,2	1434,8	26,8
(e)	1679,5	138,3	1817,9	50692	19764	70457	88,8	5,6	1583,0	26,5

<sup>a</sup> Com valores ajustado (dentro da amostra: 221 meses),

<sup>b</sup> Com valores preditos (fora da amostra: últimos 21 meses).

trajetória da primeira componente de  $\gamma_t$ . Nela observamos que o efeito instantâneo da chuva (livre de efeitos sazonais) foi sempre significativo (diferente de zero) e seu valor variou entre 0,03 e 0,05 na maior parte do período. Observamos também que este efeito exibe uma tendência decrescente nos últimos meses.

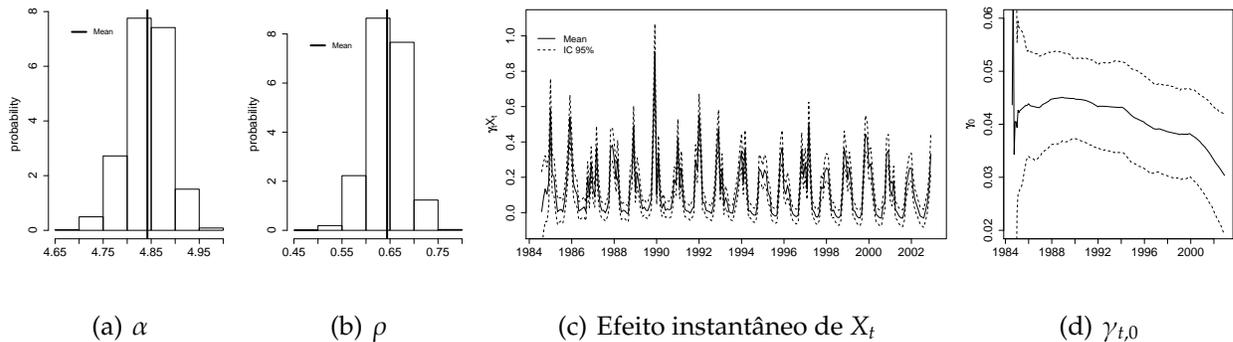


Figura 4.6: Parâmetros associados a (4.9a)-(4.9c), especificação (e) e resposta gama.

Uma das vantagens da abordagem bayesiana é que ao final do processo de inferência contamos com uma amostra de todas as distribuições a posteriori dos parâmetros dos modelos e,

portanto, a partir delas, é simples realizar inferência sobre funções desses parâmetros. A função resposta-impulso é um dos resultados mais importantes da classe de modelos que propomos ajustar à vazão. Esta função indica qual é a intensidade da resposta da vazão e o tempo de duração do efeito de um “impulso” (ocorrência) de chuva. Com base nas amostras a posteriori de  $\gamma$  e  $\rho$ , construímos a função resposta-impulso que é apresentada na figura 4.7(a). Nessa figura consideramos ausência de chuva ao longo de 26 períodos, exceto no instante  $t = 5$ . Já na figura 4.7(b) consideramos como entrada os primeiros valores observados de precipitação na bacia. Em ambas figuras, os pontos em cinza correspondem à amostra a posteriori de cada função estudada e as linhas pontilhadas representam o intervalo de 95% de credibilidade da função resposta-impulso obtida.

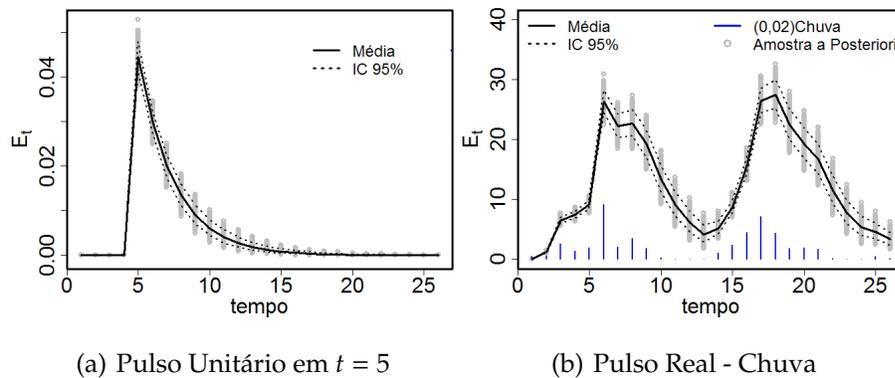


Figura 4.7: **Função Resposta-Impulso da Vazão.** A linha sólida representa a média e as linhas pontilhadas o intervalo de 95% de credibilidade a posteriori.

A figura 4.8(a) mostra os valores ajustados para os primeiros 221 meses da série temporal de vazão. Observamos que todos os valores reais estão dentro do intervalo de 95% de credibilidade, indicando boa qualidade de ajuste. Contudo, observamos também que as maiores vazões (acima de  $300 \text{ m}^3/\text{s}$ ) estão sempre muito próximas do limite superior do intervalo, o que sugere que para um melhor ajuste; esses valores podem ser modelados por uma distribuição de valores extremos. O problema do ajuste para esta cauda superior é observado também no  $Q-Q$  plot entre os valores reais e a média a posteriori da distribuição preditiva mostrado na figura 4.8(b).

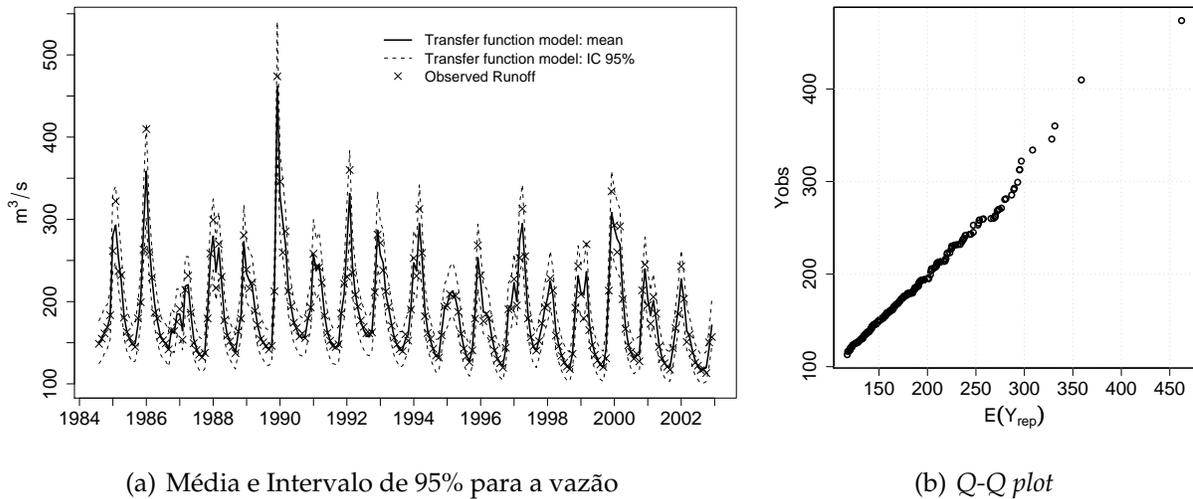


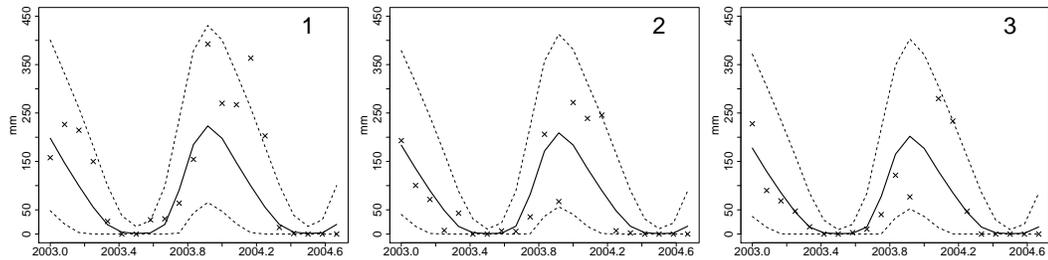
Figura 4.8: **Valores ajustados de vazão, usando o modelo (4.9a)-(4.9c).** A linha sólida representa a média e as linhas pontilhadas os limites do intervalo de 95% de credibilidade. Os + correspondem aos valores observados.

#### 4.5.4 Interpolações Espaciais e Previsões Temporais

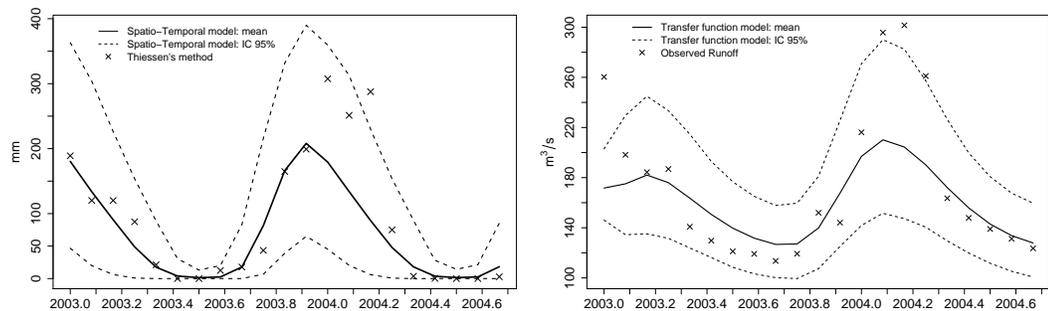
Um tópico importante relacionado a este trabalho é que os resultados do ajuste e as previsões dos modelos de chuva-vazão obtidos usando a abordagem bayesiana podem ser utilizados na área de hidrologia sintética. Como ressaltado por Rios-Insua et al. (2002), a amostra da distribuição preditiva pode ser utilizada para gerar seqüências de observações que imitem o comportamento de alguns fenômenos, o que é de grande utilidade para os engenheiros. Conseqüentemente, boas interpolações espaciais e previsões no tempo são importantes para auxiliar outras áreas da hidrologia.

Com a finalidade de avaliar as interpolações e previsões obtidas com nossos modelos, deixamos de fora da amostra as últimas 21 observações. Utilizamos a distribuição preditiva da chuva para estimá-la em cada estação monitoradora. Também, como mencionamos na seção 4.4, a cada iteração do *MCMC*, usamos a distribuição preditiva da chuva para calcular a chuva na bacia toda e com este resultado, prever a vazão. Como pode ser observado na tabela 4.2 (colunas 10 e 11), a especificação selecionada para o modelo de vazão (distribuição gama (e)) não necessariamente apresenta os menores erros quadrático e absoluto para as observações que ficaram fora da amostra. Contudo, utilizamos o mesmo modelo porque os valores do *MAE* são muito parecidos entre todos

os modelos considerados. A figura 4.9(a) mostra as previsões temporais obtidas para três estações pluviométricas, a previsão temporal para a chuva na região é apresentada na figura 4.9(b) e os valores preditos para a vazão estão representados na figura 4.9(c). Observamos que quase todos os valores reais estão dentro do intervalo de 95% de credibilidade obtido com nossa abordagem.



(a) Chuva predita em alguns postos



(b) Previsão da chuva

(c) Previsão da vazão

Figura 4.9: **Valores previstos de chuva e vazão.** Média (linha sólida) e intervalo 95% (linhas pontilhadas) das distribuições preditivas. +: em (a) e (c) correspondem aos valores observados, em (b) correspondem aos valores calculados com o método Thiessen.

#### 4.6 Considerações Finais

O desafio principal deste trabalho consistiu em realizar uma modelagem conjunta de duas variáveis hidrológicas: chuva e vazão, levando em conta toda a incerteza associada à ambos os processos e respeitando as formas de medição no espaço de cada uma. A abordagem aqui proposta é um marco bastante amplo e flexível para levar a cabo esta tarefa. Partimos de alguns modelos individuais (seja somente para chuva ou somente para vazão), previamente estabelecidos na literatura,

que permitem uma interpretação física dos parâmetros envolvidos e, também, o ajuste dos dados nas suas escalas originais. Além da atenção prestada à formulação do modelo conjunto, cuidamos da eficiência do algoritmo *MCMC*. Nossa proposta é ajustar modelos de função de transferência não normais (em particular, usar a distribuição gama para modelar dados de vazão) usando o esquema *CUBS* pois é fácil de implementar e contribui para diminuir o tempo computacional demandado. Embora não tenha sido mostrado aqui, vale a pena ressaltar que, sob o ponto de vista bayesiano, o tratamento dos dados faltantes é bastante simples pois eles podem ser considerados como parâmetros do modelo e estimados dentro do algoritmo *MCMC*. Conseqüentemente, nossa abordagem é uma ferramenta bastante promissória para a análise da relação chuva-vazão.

Uma extensão natural aos modelos aqui propostos é a inclusão no modelo de alguma variável que forneça informação sobre a vegetação na região sob estudo. A vegetação basicamente controla os processos de evaporação e interceptação, duas componentes importantes da equação de balanço hídrico.

Algumas das alternativas naturais aos modelos aqui utilizados são: ajustar outras funções de transferência para a vazão e outras funções de correlação espacial para a chuva. Uma extensão bastante interessante é a modelagem hierárquica (modelos dinâmicos hierárquicos como em Gaman & Migon (1993)) aplicada a várias séries temporais de vazão. Esta proposta pode ser utilizada em bacias com características geológicas e climatológicas bastante parecidas. Além disso, podem ser considerados modelos lineares para ambos parâmetros da distribuição gama, como em Capkun et al. (2001).

Finalmente, os resultados obtidos pela nossa abordagem fornecem uma das entradas mais importantes do problema de decisão de manejo de reservatórios (como em Rios-Insua et al. (1997)), que é um dos nossos tópicos de pesquisa atual.

#### 4.7 Apêndice: Verossimilhança - Mudança de Suporte

Sejam  $\mathbf{Y} = (Y_1, \dots, Y_T)'$ ,  $\mathbf{X} = (X_1, \dots, X_T)'$  e  $\mathbf{X}(\mathbf{s}) = (X_1(\mathbf{s}), \dots, X_T(\mathbf{s}))'$ , onde  $\mathbf{s} = (s_1, \dots, s_S)$ . Temos também  $\mathbf{X}_t(\mathbf{s}) = (X_t(s_1), \dots, X_t(s_S))'$  e  $\mathbf{X}(s_i) = (X_1(s_i), \dots, X_T(s_i))'$ ,  $s_i \in B, i = 1, \dots, S$ . A distribuição conjunta de  $\mathbf{Y}$  e  $\mathbf{X}(\mathbf{s})$  é dada por

$$\underbrace{p(\mathbf{Y}, \mathbf{X}(\mathbf{s})|\Theta)}_{\text{dados observados}} = \int \underbrace{p(\mathbf{Y}, \mathbf{X}|\mathbf{X}(\mathbf{s}), \Theta)}_{\text{não observado}} \underbrace{p(\mathbf{X}(\mathbf{s})|\Theta)}_{\text{aproximação MC}} d\mathbf{X},$$

onde  $\Theta = (\Theta_Y, \Theta_X)$ . Estritamente falando, a distribuição conjunta dos dados observados está dada por  $p(\mathbf{Y}, \mathbf{X}(\mathbf{s})|\Theta)$  e  $\mathbf{X}$  faz o papel de uma variável latente. Uma das vantagens do uso de métodos

MCMC é que podemos amostrar de  $p(\mathbf{Y}, \mathbf{X}, \mathbf{X}_t(\mathbf{s})|\Theta)$  e considerar que as amostras de  $\mathbf{Y}$  e  $\mathbf{X}_t(\mathbf{s})$  provêm da distribuição marginal conjunta de ambas variáveis. Logo, vamos nos preocupar com a distribuição  $p(\mathbf{Y}, \mathbf{X}, \mathbf{X}(\mathbf{s})|\Theta)$ , que para um  $t$  fixo é dada por

$$p(Y_t, X_t, \mathbf{X}_t(\mathbf{s})|\Theta) = p(Y_t|X_t, \mathbf{X}_t(\mathbf{s}), \Theta_Y)p(X_t, \mathbf{X}_t(\mathbf{s})|\Theta_X). \quad (4.11)$$

Voltamos agora nossa atenção para  $p(X_t, \mathbf{X}_t(\mathbf{s})|\Theta_X)$ . Lembrando que

$$X_t = |B|^{-1} \int_B \mathbf{X}_t(\mathbf{s})d\mathbf{s}, \quad (4.12)$$

ou seja,  $X_t$  é uma “função” de  $\mathbf{X}_t(\mathbf{s})$ , temos que a distribuição preditiva de  $X_t$  é

$$\underbrace{p(X_t|\mathbf{X}_t(\mathbf{s}))}_{\text{preditiva}} = \int p(X_t|\mathbf{X}_t(\mathbf{s}), \Theta_X) \underbrace{p(\Theta_X|\mathbf{X}_t(\mathbf{s}))}_{\text{posteriori}} d\Theta_X. \quad (4.13)$$

Os momentos de  $p(X_t|\mathbf{X}_t(\mathbf{s}))$  em (4.13) envolvem integrais com relação a  $\mathbf{s}$ . Por exemplo, supondo que a distribuição conjunta de  $X_t$  e  $\mathbf{X}_t(\mathbf{s})$  é normal, temos que

$$E(X_t|\Theta_X) = |B|^{-1} \int_B E(\mathbf{X}_t(\mathbf{s})|\Theta_X)d\mathbf{s}. \quad (4.14)$$

A proposta de Gelfand et al. (2001) é aproximar esses momentos por integração de Monte Carlo. Isto implica aproximar

$$\hat{p}((X_t, \mathbf{X}_t(\mathbf{s}))'|\Theta_X) = p((\hat{X}_t, \mathbf{X}_t(\mathbf{s}))'|\Theta_X), \quad (4.15)$$

onde o “chapéu” ( $\hat{\cdot}$ ) denota uma integração de Monte Carlo e

$$\hat{X}_t = \frac{1}{S_B} \sum_{i=1}^{S_B} \hat{X}_t(s_i) \quad i = 1, \dots, S_B. \quad (4.16)$$

Segundo Gelfand et al. (2001), (4.15) implica que a densidade conjunta aproximada de  $X_t$  e  $\mathbf{X}_t(\mathbf{s})$  é exatamente igual à densidade conjunta de  $\hat{X}_t$  e  $\mathbf{X}_t(\mathbf{s})$  e, na prática,  $\hat{X}_t$  é o valor amostrado. Os autores afirmam que  $\hat{X}_t \xrightarrow{P} X_t$  se  $\mathbf{X}_t(\mathbf{s})$  é um processo *almost surely continuous*.

#### 4.8 Apêndice: Postos Fluviométricos e Pluviométricos

Tabela 4.3: Postos Fluviométricos e Pluviométricos utilizados

Posto	Tipo	Código ANA	Latitude (graus)	Longitude (graus)	Altitude (m)
Taguá	Fluviométrica	46650000	-11,721	-44,502	459
São Sebastião	Pluviométrica	1144014	-11,984	-44,711	447
Fazenda Coqueiro	Pluviométrica	1244019	-12,389	-44,932	502
Barreiras	Pluviométrica	1244011	-11,154	-45,009	444
Nova Vida	Pluviométrica	1145014	-11,853	-45,122	458
Fazenda Redenção	Pluviométrica	1245004	-12,135	-45,104	490
Sítio Grande	Pluviométrica	1245007	-12,431	-45,086	512
Derocal	Pluviométrica	1245005	-12,411	-45,120	502
Ponte Serafim	Pluviométrica	1145013	-11,896	-45,612	713
Roda Velha	Pluviométrica	1245015	-12,765	-45,944	761

<sup>a</sup> ANA = Agência Nacional de Águas.

## Capítulo 5

### MODELAGEM DE MÚLTIPLAS SÉRIES DE VAZÃO

A estimação e previsão da vazão em vários pontos de uma bacia é importante na determinação das dimensões de trabalhos hidráulicos, na avaliação de riscos, na irrigação de campos, no abastecimento de água, na operação ótima de reservatórios, entre outros. Neste capítulo apresentamos uma estratégia de modelagem conjunta de várias séries temporais de vazão.

#### 5.1 Colocação do Problema

Em alguns casos, dispomos de dados de postos fluviométricos cujas áreas de drenagem estão *encaixadas*, isto é, os postos estão localizados no curso do mesmo rio e, portanto, as áreas de drenagem das sub-bacias correspondentes a cada um deles aumentam numa direção determinada (seguindo o fluxo das águas). Na bacia do Rio Grande (BA) temos um exemplo desta situação. A figura 5.1 mostra a localização de vários postos fluviométricos; em particular vamos trabalhar com: 46650000=Taguá, 46550000=Barreiras e 46543000=Redenção. Observamos que a área de drenagem de Redenção está totalmente contida na área de drenagem de Barreiras, que por sua vez está contida na área de drenagem de Taguá. Devido à localização destes postos, a vazão mensal medida em Redenção sempre será menor que a vazão medida em Barreiras, que por sua vez será menor que a medida em Taguá. Isto se deve ao fato do fluxo das águas nessa região mover-se de oeste a leste e, portanto, podemos assumir que a vazão observada no primeiro posto (Redenção) fornece informação sobre a vazão nos dois seguintes. Cabe ressaltar que esta informação dependerá da distância entre os postos e o tempo que demora (*lag*) o percurso das águas. Esta situação nos leva a pensar na modelagem conjunta condicional das séries temporais de vazão. A seção seguinte descreve esta abordagem.

#### 5.2 Modelo conjunto proposto

Suponha que dispomos de  $M$  séries temporais de vazão observadas durante  $T$  períodos de tempo. Se  $Y_t^m$  denota a vazão observada na bacia  $m$  e tempo  $t$ ,  $Y_t$  é um vetor de dimensão  $M$ , que denota todos os dados de vazão observados no tempo  $t$ , ou seja,  $Y_t = (Y_t^1, \dots, Y_t^M)'$ . O nosso objetivo é

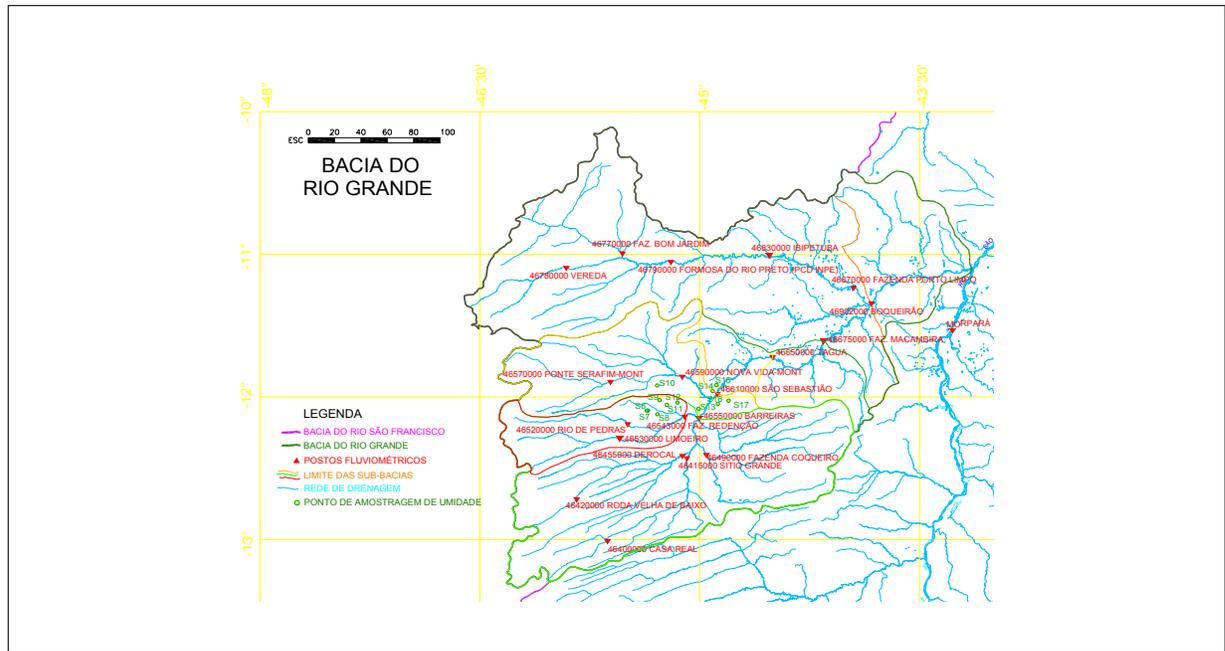


Figura 5.1: **Hidrografia e Sub-Bacias do Rio Grande.** Os postos fluviométricos utilizados são: A=Taguá, B=Barreiras e C=Faz. Redenção

modelar  $Y_t$  levando em conta as interações entre suas componentes e/ou explicar o movimento comum entre elas.

Suponha que  $Y_t \sim p(\mu_t, \Sigma)$ , onde  $\mu_t$  é um vetor  $M \times 1$  que define a posição e  $\Sigma$  a uma matriz  $M \times M$  cujos elementos são as variâncias e covariâncias de  $Y_t$ . Várias especificações podem ser dadas para  $\mu_t$  e  $\Sigma$ . Por exemplo, na análise de séries temporais multivariadas é bastante comum o uso de modelos auto-regressivos vetoriais (VAR), onde  $\mu_t$  depende dos valores passados de  $Y_t$ , ou de modelos de regressão onde  $\mu_t$  depende de covariáveis independentes. Mas também pode existir uma dependência entre os elementos de  $\mu_t$ , por exemplo,  $\mu_t^m$  pode depender de  $\mu_t^{m-1}$ . Neste último caso temos que  $p(Y)$  admite uma representação condicional. Em outras palavras, no tempo  $t$ , a distribuição conjunta  $p(Y_t)$ , condicionada a um vetor de parâmetros  $\theta$ , pode ser fatorada da maneira seguinte:

$$p(Y_t|\theta) = p(Y_t^M|Y_t^{M-1}, \theta)p(Y_t^{M-1}|Y_t^{M-2}, \theta) \cdots p(Y_t^2|Y_t^1, \theta)p(Y_t^1, \theta), \quad (5.1)$$

onde  $p(Y_t^m)$  é a distribuição de  $Y_t^m, i = 1, \dots, M$ . Esta representação condicional permite que o problema multivariado seja tratado como vários problemas univariados. Royle & Berliner (1999),

Berliner (2000) e Gelfand et al. (2004) discutem esta abordagem num contexto espaço-temporal bastante geral.

### 5.2.1 Especificação Condicional

A abordagem proposta na seção 4.3 pode ser naturalmente estendida para modelar múltiplas séries temporais de vazão conjuntamente. Sem perda de generalidade, suponha que temos duas bacias  $A$  e  $B$ , tais que  $A \supset B$  e  $Y_t^A > Y_t^B, \forall t$ . A distribuição conjunta de  $Y_t^A$  e  $Y_t^B$ , condicionadas a  $\theta$ , pode ser representada por

$$p(Y_t^A, Y_t^B | \theta) = p(Y_t^A | Y_t^B, \theta) p(Y_t^B | \theta). \quad (5.2)$$

Dado que a vazão é uma variável contínua definida nos reais positivos, é comum assumir que  $Y_t^m$  segue uma distribuição log-normal ou gama. É muito importante ressaltar que a ordem do condicionamento das variáveis deve respeitar o fluxo das águas na região sob estudo. Especificamente, nos modelos aqui apresentados, vamos considerar que o nível de vazão na bacia “menor”, neste caso  $B$ , fornece informação sobre o nível da bacia “maior”, neste caso  $A$  e, portanto,  $Y_t^B$ , ou uma função dela, deve entrar no modelo como regressor no valor esperado de  $Y_t^A$ .

### 5.2.2 Efeito da Chuva: função de transferência

Dado que a vazão depende diretamente da precipitação, o modelo proposto inclui também o volume acumulado de chuva que cai nas (sub-)bacias sob estudo. Sejam  $X_t^m$  e  $|m|$  o volume de chuva acumulado e a área (por exemplo, em  $km^2$ ,) da bacia  $m$ , respectivamente. O volume total de chuva que cai na bacia  $m$  no instante de tempo  $t$  é dado por  $|m|X_t^m$ . É importante observar que  $X_t^m$  representa a chuva média correspondente a toda a área de drenagem associada à  $m$ , ou seja é uma medida de área (conglomerado). Contudo, a chuva normalmente é medida em vários postos pluviométricos (localizações) dentro de uma bacia, isto é, é uma medida pontual e usualmente denotada por  $X_t(s)$ . Em conseqüência, para obter  $X_t^m$  é necessário realizar uma mudança de suporte (Gelfand et al., 2001). A mudança de suporte será descrita em detalhe na seção 5.4.1.

O volume de chuva que deve ser utilizado para modelar a vazão de cada sub-bacia, como em (5.1), é aquele que influencia diretamente somente o nível de vazão da respectiva sub-bacia. Tendo em conta este fato, a distribuição conjunta das duas séries de vazão,  $Y_t^A$  e  $Y_t^B$ , considerando a chuva, é dada por

$$p(Y_t^A | Y_t^B, X_t^{A|B}, \Theta) p(Y_t^B | X_t^B, \Theta), \quad (5.3)$$

onde  $X_t^{A|B} = X_t^A - X_t^B$  é a diferença entre a chuva total da sub-bacia  $A$  e da sub-bacia  $B$ . Em (5.3) as áreas das bacias não estão explicitadas somente para simplificar a notação. Em (5.3) observamos que  $Y_t^B$  é modelada de forma condicional a  $X_t^B$ , a chuva acumulada total de  $B$ . Mas a chuva utilizada para modelar  $Y_t^A$  é  $X_t^{A|B}$ , fato que se justifica facilmente porque a chuva que cai sobre  $B$  afeta  $Y_t^A$  através de  $Y_t^B$ . Verifica-se também que  $X_t^A = X_t^{A|B} + X_t^B$ .

Nossa proposta considera também realizar a modelagem conjunta das séries de vazão e chuva de várias bacias, usando a representação condicional que respeite a ordem “natural” das bacias. No caso de duas bacias, a nossa proposta é modelar

$$p(Y_t^A, Y_t^B, X_t^A, X_t^B | \Theta) = p(Y_t^A | Y_t^B, X_t^A, X_t^B, \Theta) p(Y_t^B | X_t^B, \Theta) p(X_t^A, X_t^B | \Theta), \quad (5.4)$$

onde  $p(X_t^m)$  é a distribuição preditiva de  $X_t^m$ . Duas das principais características do modelo em (5.4) são: (1) permite fazer inferência e previsão sobre a vazão (e eventos relacionados a ela) em várias localizações de uma mesma bacia, e (2) permite levar em conta a incerteza associada a todo o processo chuva-vazão.

O modelo formulado na seção 4.3 continua válido, mas agora o cálculo das chuvas de área (mudança de suporte) deve ser sobre grades que respeitem os limites das sub-bacias, em particular, se assumimos que  $p(Y_t^m)$  é uma distribuição log-normal com parâmetros  $\mu_t^m$  e  $\sigma_m^2$  e representamos a relação entre  $Y_t^m$  e  $X_t^m$  mediante uma função de transferência de primeira ordem, como a apresentada em (4.9c). Além disso, acreditamos que é possível utilizar um único modelo espaço-temporal para obter a distribuição preditiva de  $X^A$  e  $X^B$ . Dado a que  $A$  e  $B$  estão contidas numa mesma bacia (neste exemplo,  $A$ ), temos:

$$Y_t^A | Y_t^B \sim \text{log-normal}(\mu_t^{A|B}, \sigma_{A|B}^2), \quad t = 1, \dots, T \quad (5.5a)$$

$$Y_t^B \sim \text{log-normal}(\mu_t^B, \sigma_B^2), \quad (5.5b)$$

onde

$$\mu_t^{A|B} = \alpha^{A|B} + \eta^{A|B} \log(Y_t^B) + E_t^{A|B} \quad (5.6a)$$

$$\mu_t^B = \alpha^B + E_t^B \quad (5.6b)$$

$$E_t^m = \rho^m E_{t-1}^m + \gamma^m X_t^m |m| + w_t^m \quad w_t^m \sim N(0, W_m), \quad m = A|B, B, \quad (5.6c)$$

onde  $|m|$  denota a área da sub-bacia  $m$  e é multiplicada por  $X_t^m$  em (5.6c) para representar o volume acumulado de chuva na área de drenagem correspondente. Se  $m = A|B$ ,  $X_t^{A|B}$  e  $|A|B|$  denotam a chuva e é a área de drenagem de Taguá sem incluir a área de drenagem de Barreiras. Em (5.5)

assumimos a mesma forma funcional para a função de transferência nas duas sub-bacias. Porém, esta não é uma restrição do modelo, já que uma forma funcional diferente para cada sub-bacia pode ser adotada.

### 5.3 Procedimento de Inferência

A representação condicional em (5.5) tem a atrativa característica de produzir uma fatorização do modelo bivariado em 2 modelos univariados que podem ser ajustados separadamente. Este fato pode ser visto ao analisar a verossimilhança:

$$p(\mathbf{Y}|\boldsymbol{\theta}) = \prod_{t=1}^T p(\mathbf{Y}_t|\boldsymbol{\theta}) = \prod_{t=1}^T p(Y_t^A|Y_t^B, \boldsymbol{\theta}^{A|B}) \prod_{t=1}^T p(Y_t^B|\boldsymbol{\theta}^B), \quad (5.7)$$

onde  $\boldsymbol{\theta} = (\boldsymbol{\theta}^{A|B}, \boldsymbol{\theta}^B)$ . Em particular,  $\boldsymbol{\theta}^{A|B} = \{\alpha^{A|B}, \eta^{A|B}, \rho^{A|B}, \gamma^{A|B}, W_{A|B}, \sigma_{A|B}^2\}$  e  $\boldsymbol{\theta}^B = \{\alpha^B, \rho^B, \gamma^B, W_B, \sigma_B^2\}$ . Para completar o modelo, de acordo com o paradigma bayesiano, podemos assumir que, a priori, todos os parâmetros são independentes e que, para  $m = A|B, B$ :  $\alpha^m$  e  $\eta^{A|B}$  seguem uma distribuição normal com média zero e variância  $10^3$ ,  $N(0, 10^3)$ ,  $\rho^m$  segue uma distribuição uniforme entre 0 e 1,  $U[0, 1]$ , e,  $W_m$  e  $\sigma_m^2$  seguem uma distribuição gama invertida com ambos parâmetros iguais a 0,01,  $IG(10^{-2}; 10^{-2})$ .

Se  $p(\boldsymbol{\theta})$  denota a distribuição a priori de  $\boldsymbol{\theta}$ , vamos utilizar métodos MCMC (Gamerman & Lopes, 2006) para obter amostras da distribuição a posteriori:

$$p(\boldsymbol{\theta}|\mathbf{Y}) = \prod_{t=1}^T p(\mathbf{Y}_t|\boldsymbol{\theta})p(\boldsymbol{\theta}) = \prod_{t=1}^T p(Y_t^A|Y_t^B, \boldsymbol{\theta}^{A|B})p(\boldsymbol{\theta}^{A|B}) \prod_{t=1}^T p(Y_t^B|\boldsymbol{\theta}^B)p(\boldsymbol{\theta}^B). \quad (5.8)$$

A idéia é ajustar partes independentes do modelo, de forma seqüencial. A vantagem do uso de métodos MCMC é que em cada etapa, para cada parâmetro, obtemos uma amostra da posteriori que pode ser utilizada como “entrada” para modelos condicionados a eles.

No caso do modelo conjunto em (5.5), um procedimento a seguir é:

- (1). Ajustar um modelo espaço-temporal para a chuva,  $X(s)$ , observada em  $S$  localizações da bacia  $A$  ( $A \supset B$ ).
- (2). Obter as amostras da chuva em cada sub-bacia,  $X^A$  e  $X^B$ , e calcular  $X^{A|B} = X^A - X^B$ , a partir da amostra preditiva de  $X(s)$ , gerada no passo (1), para cada ponto das grades de interpolação.
- (3). Ajustar um modelo chuva-vazão para  $Y^B$ , usando como  $X^B$  a cadeia gerada no passo (2).

- (4). Ajustar um modelo chuva-vazão para  $Y^A$ , usando como  $Y^B$  a cadeia gerada no passo (3), e como  $X^{AB}$ , a cadeia gerada no passo (2).

O algoritmo apresentado na seção B.2 pode ser utilizado para o modelo de  $Y^B$ , mas para o modelos de  $Y^A|Y^B$  é necessário fazer a modificação a seguir. A geração de amostras de  $\alpha^{AB}$  e  $\eta^{AB}$  deve ser realizada em bloco diretamente da sua distribuição a posteriori conjunta, uma normal bivariada. Este passo é necessário pois ambos parâmetros são altamente correlacionados e passos individuais de amostragem são ineficientes para a identificação das regiões de maior massa de probabilidade.

## 5.4 Aplicação

Com a finalidade de ilustrar o modelo condicional descrito na seção anterior, utilizamos dados fornecidos pelo Laboratório de Hidrologia da Universidade Federal do Rio de Janeiro (UFRJ). Trata-se de séries históricas obtidas após procedimentos de consistência e sem observações faltantes, de dados mensais de vazão entre agosto de 1984 e setembro de 2004 (242 meses), de três (3) postos fluviométricos e nove (9) postos pluviométricos localizados na bacia do Rio Grande (BA). A figura 5.2 mostra a localização dos postos e a figura 5.3 mostra as 3 séries temporais de vazão (em  $mm^3/s$ ) e 3 séries de chuva (em  $mm$ ). A localização e área de drenagem correspondentes a cada posto fluviométrico são apresentados na tabela 5.1. Observamos que a área de drenagem total correspondente ao posto Taguá (A) é  $34250 km^2$  e que as sub-bacias Barreiras (B) e Redenção (C) têm áreas de  $23250$  e  $5185 km^2$ , respectivamente. Evidentemente, as áreas de drenagem são tais que  $Taguá \supset Barreiras \supset Redenção$  ( $A \supset B \supset C$ ).

### 5.4.1 Modelando a Chuva

Seja  $\{X_t(s_i) : s_i \in B \subset \mathbb{R}^2; i = 1, \dots, S; t = 1, \dots, T\}$  um realização do campo aleatório espacial  $\{X_t(s) : s \in B \subset \mathbb{R}^2; t = 1, 2, \dots\}$ . Neste capítulo,  $X_t(s_i)$  é a precipitação na localização  $s_i$  e tempo  $t$ ,  $S$  é o número de estações pluviométricas e  $T$  é o tamanho das séries temporais disponíveis. O modelo

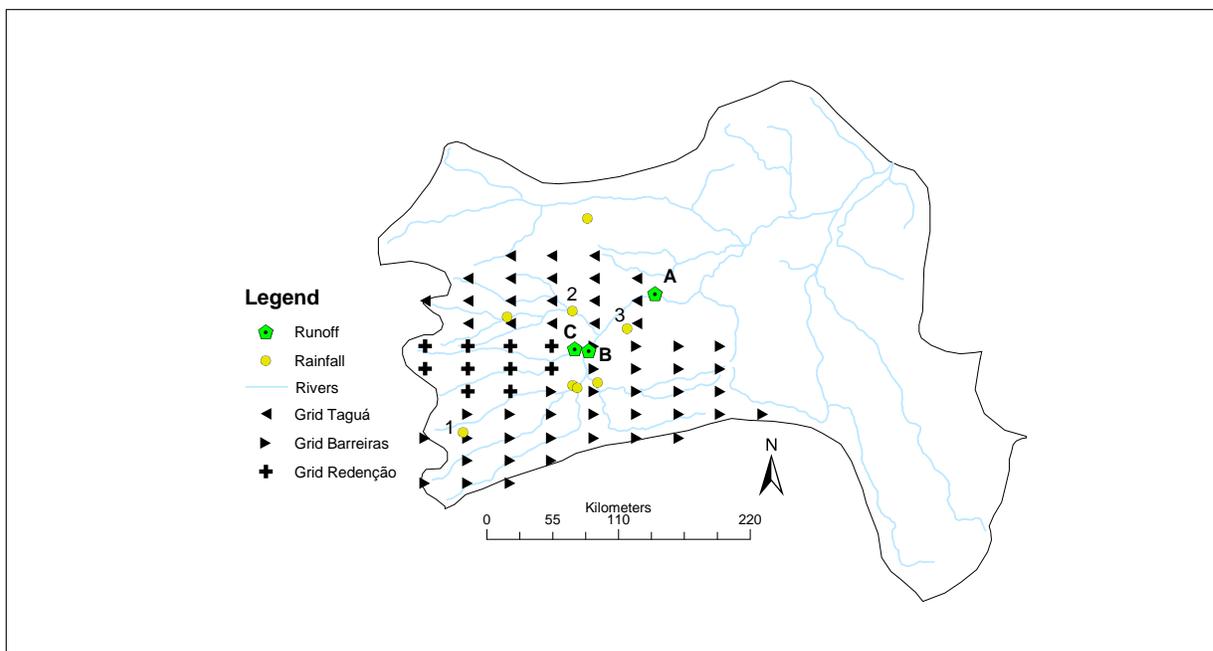


Figura 5.2: **Sub-Bacias do Rio Grande:** Localização das estações monitoradoras e Grade de Interpolação associada a cada sub-bacia. Os postos fluviométricos são: A=Taguá, B=Barreiras e C=Redenção

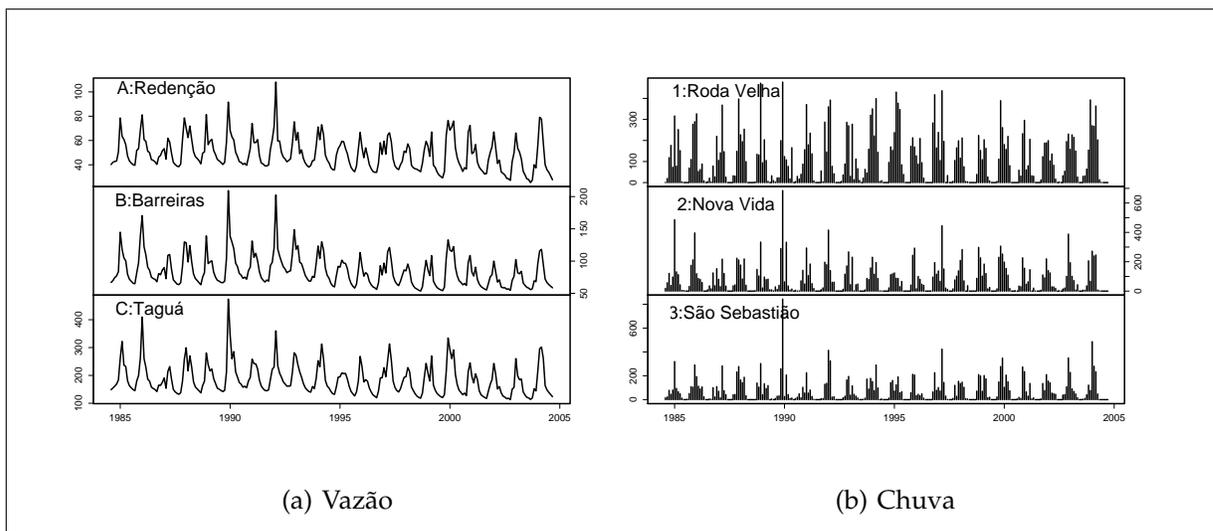


Figura 5.3: **Vazão nas sub-bacias do Rio Grande e Chuva em três estações pluviométricas** (estações marcadas em 5.2).

espaço-temporal sugerido em Sansó & Guenni (2000) pode ser ajustado para  $X_t(s_i)$ . Assim,

$$X_t(s_i) = \begin{cases} w_t(s_i)^\beta & \text{se } w_t(s_i) > 0 \\ 0 & \text{se } w_t(s_i) \leq 0 \end{cases} \quad i = 1, \dots, S \quad (5.9a)$$

$$\mathbf{w}_t = \mathbf{z}_t + \mathbf{v}_t \quad \mathbf{v}_t \sim N_S(\mathbf{0}, \tau^2 \mathbf{I}) \quad (5.9b)$$

$$\mathbf{z}_t = \mathbf{F}' \boldsymbol{\theta}_t + \boldsymbol{\epsilon}_t \quad \boldsymbol{\epsilon}_t \sim N_S(\mathbf{0}, \sigma^2 \mathbf{V}_t) \quad (5.9c)$$

$$\boldsymbol{\theta}_t = \mathbf{G} \boldsymbol{\theta}_{t-1} + \boldsymbol{\epsilon}_t \quad \boldsymbol{\epsilon}_t \sim N_k(\mathbf{0}, \mathbf{W}_t), \quad (5.9d)$$

onde  $w_t(s)$  é uma variável gaussiana latente associada a  $X_t(s)$ ,  $\beta$  é uma potência desconhecida,  $\mathbf{w}_t$  é um vetor de dimensão  $S$  que aloca as observações das  $S$  localizações correspondentes ao tempo  $t$ ,  $\tau^2 \mathbf{I}$  é o efeito pepita,  $\sigma^2 > 0$  e  $\mathbf{V}_t \in \mathbb{R}^{S \times S}$ .  $\mathbf{V}_t$  captura a correlação espacial e tem elementos:  $V_{ij} = \exp(-\lambda d_{ij})$ , ou seja assumimos uma correlação com decaimento exponencial onde  $\lambda$  controla a taxa de decaimento,  $\lambda > 0$  e  $d_{ij}$  é a distância euclidiana entre as localizações  $s_i$  e  $s_j$ . As equações (5.9c)–(5.9d) correspondem a um modelo dinâmico normal multivariado tal que  $\mathbf{F}' \in \mathbb{R}^{S \times k}$ ,  $\mathbf{G} \in \mathbb{R}^{k \times k}$  e  $\boldsymbol{\theta} \in \mathbb{R}^k$ . Os elementos de  $\boldsymbol{\theta}$  são tais que  $\boldsymbol{\theta}_t = (\boldsymbol{\theta}_{t1}, \boldsymbol{\theta}_{t2})'$ , onde  $\boldsymbol{\theta}_{t1}$  é um sub-vetor que descreve a tendência espacial e  $\boldsymbol{\theta}_{t2}$  descreve os efeitos sazonais. Se, em particular (com base nos resultados apresentados na seção 4.5.3), consideramos um intercepto e um efeito linear da longitude na tendência espacial e dois harmônicos de Fourier para representar o padrão sazonal, a dimensão de  $\boldsymbol{\theta}_t$  é  $k = 6$ ,  $\mathbf{F}_t$  é um vetor com componentes:  $(1, \text{longitude}(s_i), 1, 0, 1, 0)'$  e  $\mathbf{G} = \text{diag}(\mathbf{G}_1, \mathbf{G}_2)$ , onde  $\mathbf{G}_1$  é uma matriz identidade de ordem 2 e  $\mathbf{G}_2$  tem blocos diagonais  $\mathbf{G}_{2r} = \begin{pmatrix} \cos(2\pi r/12) & \sin(2\pi r/12) \\ -\sin(2\pi r/12) & \cos(2\pi r/12) \end{pmatrix}$ ,  $r = 1, 2$ .

Para estimar a chuva em cada sub-bacia ou área de drenagem associada a cada medição de vazão é preciso fazer a troca de suporte (Gelfand et al., 2001). Esta troca consiste em obter estimativas para áreas, a partir de dados pontuais, resolvendo a seguinte integral

$$X_t^m = \frac{1}{|m|} \int_m X_t(s) ds \quad \forall s \in m, \quad (5.10)$$

onde  $X_t^m$  denota a chuva no tempo  $t$  e  $|m|$  denota a área da bacia  $m$ . Na prática, a integral em (5.10) é aproximada por

$$X_t^m \approx \frac{1}{N_m} \sum_{i=1}^{N_m} \hat{X}_t(s_i) \quad i = 1, \dots, N_m, \quad (5.11)$$

onde  $N_m$  denota o número de pontos de uma grade de interpolação construída dentro dos limites da bacia  $m$  e  $\hat{X}_t(i)$  é o valor predito de chuva para a localização  $s_i$  dessa grade.

No caso da bacia do Rio Grande, construímos uma grade (regular) de 63 pontos localizados dentro dos limites da área de drenagem associada ao posto Taguá (A). Classificamos esses pontos

em três grupos segundo sua localização geográfica, isto é, dividimos a grade segundo os limites das sub-bacias da região. Ao final ficamos com 19 pontos dentro de Taguá,  $N_{A|B} = 19$ , 34 dentro de Barreiras,  $N_{B|C} = 34$ , e o restante (10) dentro de Redenção,  $N_C = 10$ . A figura 5.2 mostra a grade de interpolação e a divisão segundo as sub-bacias. Lembrando que as sub-bacias estão encaixadas ( $N_{A|B} + N_{B|C} + N_C = 63$ ), fica evidente que os pontos da grade de Redenção podem ser utilizados para uma análise individual de Barreiras ( $N_B = N_{B|C} + N_C$ ) e os pontos de Redenção e Barreiras podem ser utilizados para uma análise individual de Taguá ( $N_A = N_{A|B} + N_B$ ).

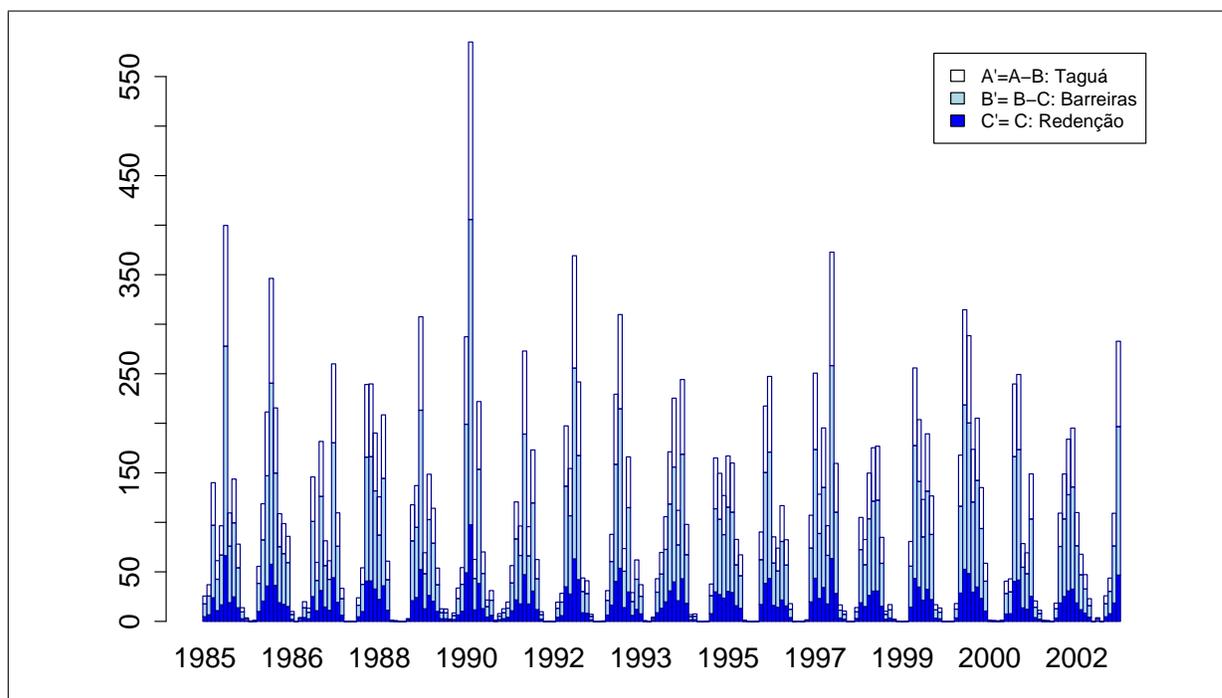


Figura 5.4: **Média a posteriori estimada para a chuva acumulada por sub-bacia.** Cada barra representa a chuva na bacia total. As cores representam a quantidade de chuva estimada para cada sub-bacia a partir das grades mostradas na figura 5.2

De forma similar ao capítulo 4, ajustamos o modelo espaço-temporal em (5.9) aos dados das 9 estações pluviométricas disponíveis e obtivemos amostras da distribuição preditiva da chuva para cada ponto da grade construída. A partir daí, utilizamos a aproximação em (5.11) nas três grades. O resultado deste processo é apresentado na figura 5.4. Nessa figura temos que cada barra corresponde à média (a posteriori) da precipitação num mês na bacia toda. Cada barra está subdividida em três partes, cada parte corresponde a chuva numa sub-bacia. Observamos que a parte

correspondente a Barreiras é a maior, como era esperado.

#### 5.4.2 Modelando a Vazão

Após obter as estimativas de área da chuva em cada sub-bacia, ajustamos o modelo conjunto das três séries temporais de vazão. O modelo em (5.5) estendido para o caso de três sub-bacias é

$$Y_t^A | Y_t^B \sim \text{log-normal}(\mu_t^{A|B}, \sigma_{A|B}^2) \quad t = 1, \dots, T \quad (5.12a)$$

$$Y_t^B | Y_t^C \sim \text{log-normal}(\mu_t^{B|C}, \sigma_{B|C}^2) \quad (5.12b)$$

$$Y_t^C \sim \text{log-normal}(\mu_t^C, \sigma_C^2) \quad (5.12c)$$

$$\mu_t^{A|B} = \alpha^{A|B} + \eta^{A|B} \log(Y_t^B) + E_t^{A|B} \quad (5.12d)$$

$$\mu_t^{B|C} = \alpha^{B|C} + \eta^{B|C} \log(Y_t^C) + E_t^{B|C} \quad (5.12e)$$

$$\mu_t^C = \alpha^C + E_t^C \quad (5.12f)$$

$$E_t^m = \rho^m E_{t-1}^m + \gamma^m X_t^m |m| + w_t^m \quad w_t^m \sim N(0, W_m) \quad m = A|B, B|C, C, \quad (5.12g)$$

onde, da tabela 5.1, temos as seguintes áreas:  $|A|B| = 11000 \text{ km}^2$ ,  $|B|C| = 18065 \text{ km}^2$  e  $|C| = 5185 \text{ km}^2$ .

Tabela 5.1: Postos Fluviométricos utilizados

Código ANA	Posto	AD <sup>b</sup> (km <sup>2</sup> )	Latitude (graus)	Longitude (graus)	Altitude (m)
46650000	Taguá	34250	-11,72	-44,50	459
46550000	Barreiras	23250	-12,15	-45,01	437
46543000	Redenção	5185	-12,14	-45,10	474

<sup>a</sup> ANA = Agência Nacional de Águas.

<sup>b</sup> AD = Area de drenagem.

Para ajustar o modelo seguimos o procedimento descrito na seção 5.3. Utilizamos o amostrador de Gibbs para gerar duas cadeias paralelas de 50000 iterações cada uma. Para formar a amostra da distribuição a posteriori, descartamos as primeiras 30000 iterações e guardamos valores a cada 10 iterações. Verificamos a convergência de todas as cadeias de forma visual e com auxílio da estatística  $\hat{R}$  (Gelman & Rubin, 1992).

Na tabela 5.2 mostramos algumas estatísticas calculadas a partir das amostras das distribuições a posteriori dos parâmetros do modelo em (5.5). Observamos que todos os  $\hat{R}$  estão próximos de

1, o que sugere que a convergência de todas as cadeias foi atingida. Observamos também que os fluxos base, denotados por  $\alpha^{A|B}$ ,  $\alpha^{B|C}$  e  $\alpha^C$ , diminuem conforme o tamanho da região considerada aumenta. Além disso, observamos que a média e o intervalo de 95% de credibilidade de  $\eta^{B|C}$  são 0,99 e (0,93; 1,00), respectivamente. Este resultado reflete bem a relação entre as vazões em C e B já que toda a vazão que passa por C deve passar por B. Entretanto, a média e o intervalo de 95% de credibilidade de  $\eta^{A|B}$  são 0,92 e (0,83; 0,99), respectivamente. Este resultado é explicado pela presença de pontos de captação de água ao longo do rio para agricultura e/ou outros usos. Vale a pena ressaltar que nenhuma restrição a priori foi utilizada para  $\eta^{A|B}$  nem  $\eta^{B|C}$ , porém os intervalos de credibilidade não superaram o valor 1, que é o limite natural a ser imposto. Ainda na tabela 5.2 observamos que os três efeitos instantâneos do volume de chuva,  $\gamma^m$ ,  $m = A|B, B|C, C$ , são significativos embora exista uma probabilidade de serem zero, e que o seu valor varia bastante de uma sub-bacia a outra. Em particular, observamos que o maior  $\gamma$  que corresponde à bacia C, que é a menor das três e não conta com informação de vazão de outras bacias.

Com a finalidade de ilustrar a bondade de ajuste de nosso modelo condicional, na figura (5.5) mostramos os *Q-Q plot* entre os valores reais de vazão e a média a posteriori da distribuição preditiva obtida. Nos três casos observamos um bom ajuste, mas com uma tendência a ser pior na sub-bacia maior ( $A=Taguá$ ).

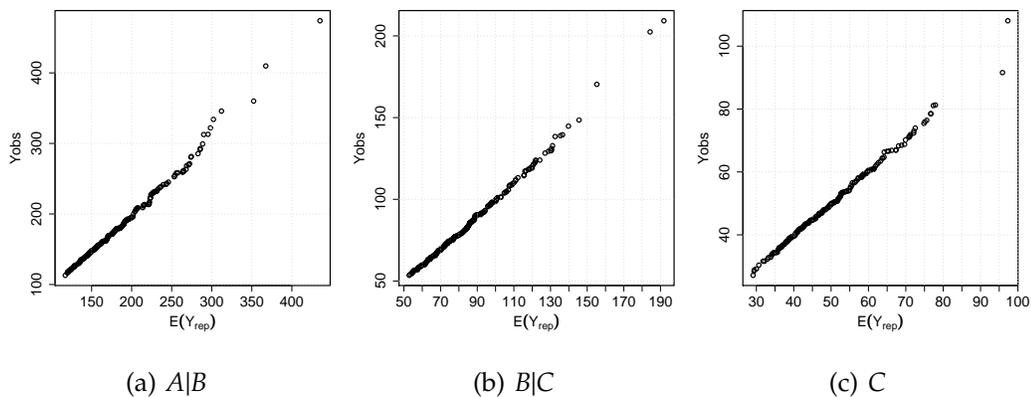


Figura 5.5: Valores ajustados de vazão, usando o modelo condicional.

Na tabela 5.3 apresentamos as variâncias (variância da distribuição normal e parâmetro de escala na distribuição log-normal), covariâncias e correlações incondicionais estimadas, entre as séries de vazão, obtidas a partir do modelo conjunto. As fórmulas utilizadas para o cálculo destes

Tabela 5.2: Estatísticas a posteriori associadas com os parâmetros no modelo conjunto condicional

Parâmetro	média	d.p.	2,5%	25%	50%	75%	97,5%	$\hat{R}$
<b>A B = Taguá   Barreiras</b>								
$\alpha^{A B}$	1,151	0,190	0,812	1,022	1,142	1,276	1,538	1,008
$\eta^{A B}$	0,915	0,043	0,827	0,887	0,916	0,945	0,992	1,010
$\rho^{A B}$	0,510	0,196	0,037	0,387	0,533	0,658	0,829	1,001
$\gamma^{A B}$	0,025	0,016	-0,005	0,014	0,024	0,035	0,059	1,001
$W_{A B}$	0,003	0,002	0,001	0,002	0,003	0,004	0,008	1,015
$\sigma_{A B}^2$	0,007	0,002	0,003	0,006	0,007	0,009	0,012	1,040
<b>B C = Barreiras   Redenção</b>								
$\alpha^{B C}$	0,587	0,111	0,388	0,507	0,581	0,654	0,814	1,006
$\eta^{B C}$	0,989	0,029	0,931	0,971	0,990	1,009	1,040	1,006
$\rho^{B C}$	0,743	0,092	0,526	0,692	0,752	0,806	0,900	1,004
$\gamma^{B C}$	0,005	0,004	-0,002	0,003	0,005	0,008	0,013	1,001
$W_{B C}$	0,003	0,001	0,001	0,002	0,002	0,003	0,005	1,005
$\sigma_{B C}^2$	0,004	0,001	0,001	0,003	0,004	0,005	0,007	1,068
<b>C = Redenção</b>								
$\alpha^C$	3,543	0,025	3,492	3,525	3,544	3,561	3,590	1,014
$\rho^C$	0,594	0,028	0,539	0,575	0,594	0,613	0,646	1,027
$\gamma^C$	1,645	0,074	1,496	1,597	1,645	1,696	1,786	1,001
$W_C$	0,005	0,001	0,003	0,005	0,005	0,006	0,008	1,010
$\sigma_C^2$	0,002	0,001	0,001	0,002	0,002	0,003	0,004	1,107

parâmetros são apresentadas na seção 5.6. Vale a pena ressaltar que estas estimativas foram obtidas a partir das amostras das distribuições a posterioris dos parâmetros, obtidas com o MCMC. Portanto, dispomos de uma amostra completa dessas quantidades e algumas estatísticas calculadas a partir dessas amostras são apresentadas na tabela 5.3. Observamos que as variâncias aumentam conforme aumenta o tamanho da bacia, resultado que era esperado pois cada bacia leva embutida a incerteza associada as bacias menores. Observamos, também, a alta correlação existente entre as séries: 0,75 entre  $Y^A$  e  $Y^B$ , 0,63 entre  $Y^B$  e  $Y^C$  e 0,47 entre  $Y^A$  e  $Y^B$ .

Para fins de comparação, na tabela 5.4 apresentamos o resultado do ajuste de modelos univariados independentes para as sub-bacias Taguá ( $A$ ) e Barreiras ( $B$ ). Os modelos ajustados consistem

Tabela 5.3: Variância, Covariância e Correlação entre as três séries de vazão, a partir do modelo conjunto

Parâmetro	média	d.p.	2,5%	25%	50%	75%	97,5%	$\hat{R}$
Parâmetros na escala logarítmica (na distribuição normal)								
- Variâncias								
$\sigma_A^2$	0,0391	0,0065	0,0304	0,0355	0,0386	0,0422	0,0498	1,0156
$\sigma_B^2$	0,0265	0,0070	0,0205	0,0239	0,0259	0,0283	0,0351	1,0194
$\sigma_C^2$	0,0107	0,0013	0,0085	0,0098	0,0106	0,0115	0,0137	1,0005
- Covariâncias								
$\sigma_{AB}$	0,0242	0,0059	0,0189	0,0218	0,0237	0,0259	0,0321	1,0170
$\sigma_{BC}$	0,0106	0,0014	0,0083	0,0097	0,0105	0,0114	0,0137	1,0008
$\sigma_{AC}$	0,0097	0,0014	0,0074	0,0087	0,0096	0,0106	0,0129	1,0028
Correlações na escala original (na distribuição log-normal)								
$Corr[Y^A, Y^B]$	0,7500	0,0508	0,6484	0,7156	0,7532	0,7847	0,8461	1,0145
$Corr[Y^B, Y^C]$	0,6303	0,0548	0,5266	0,5938	0,6305	0,6687	0,7327	1,0140
$Corr[Y^A, Y^C]$	0,4732	0,0545	0,3777	0,4342	0,4704	0,5087	0,5892	1,0173

em funções de transferência de primeira ordem como em (5.6c), mas as equações dos valores esperados não incluem a vazão da bacia menor. Em outras palavras ajustamos, independentemente,

$$\mu_t^m = \alpha^m + E_t^m, \quad m = A, B \quad (5.13a)$$

$$E_t^m = \rho^m E_{t-1}^m + \gamma^m X_t^m |m| + w_t^m, \quad w_t^m \sim N(0, W_m), \quad (5.13b)$$

onde, a diferença de (5.6c),  $X_t^m$  corresponde à chuva em toda a sub-bacia. As tabelas 5.2 e 5.4 devem ser comparadas com cuidado porque na primeira temos estimativas condicionais e na segunda temos estimativas independentes. Por isso, vamos somente comparar as variâncias marginais obtidas com o modelo conjunto versus as variâncias dos modelos univariados. Depois vamos comparar algumas previsões pontuais. As médias a posteriori de  $\sigma_A^2$  e  $\sigma_B^2$  apresentadas na tabela 5.4 são menores que aquelas da tabela 5.3. Este resultado era esperado porque mais incerteza é levada em conta no modelo conjunto.

Na tabela 5.5 mostramos as estimativas obtidas para o valor médio (valor esperado) da vazão em 4 instantes de tempo ( $t = 12, 37, 125, 200$ ), com o modelo multivariado e com os modelos individuais. Observamos que ambas alternativas fornecem boas estimativas, porém aquelas do modelo

Tabela 5.4: Estatísticas a posteriori associadas com os parâmetros nos modelos univariados independentes ajustados para a vazão em Taguá ( $Y^A$ ) e Barreiras ( $Y^B$ ).

Parâmetro	média	d.p.	2,5%	25%	50%	75%	97,5%	$\hat{R}$
<b>A = Taguá</b>								
$\alpha^A$	4,7760	0,0266	4,7220	4,7586	4,7766	4,7940	4,8271	1,0350
$\rho^A$	0,6766	0,0194	0,6368	0,6642	0,6779	0,6895	0,7138	1,0067
$\gamma^A$	0,0441	0,0015	0,0410	0,0431	0,0441	0,0451	0,0472	1,0037
$W_A$	0,0030	0,0008	0,0017	0,0024	0,0029	0,0034	0,0048	1,0030
$\sigma_A^2$	0,0048	0,0008	0,0033	0,0042	0,0048	0,0053	0,0066	1,0084
<b>B = Barreiras</b>								
$\alpha^B$	4,0279	0,0297	3,9647	4,0087	4,0305	4,0493	4,0801	1,0382
$\rho^B$	0,6533	0,0243	0,6051	0,6377	0,6531	0,6701	0,6983	1,0097
$\gamma^B$	0,0934	0,0038	0,0858	0,0908	0,0934	0,0958	0,1009	1,0208
$W_B$	0,0050	0,0008	0,0036	0,0045	0,0050	0,0055	0,0068	1,0035
$\sigma_B^2$	0,0023	0,0005	0,0014	0,0019	0,0023	0,0026	0,0035	1,0117

multivariado, na maioria dos casos, podem ser consideradas mais acuradas. Temos, por exemplo, que para  $t = 37$  os valores observados foram (135, 51; 64, 42; 38.87), as médias preditas com o modelo conjunto foram (136, 176; 60, 576; 34, 787), entretanto com os modelos individuais obtivemos (119, 353; 56, 510; 34, 787). Na tabela 5.6 mostramos algumas estimativas obtidas a partir das amostras a posteriori dos valores replicados da vazão em  $t = 35, 80, 140, 200$ . Observamos que todos os intervalos de 95% de credibilidade contêm o valor real. As médias obtidas com o modelo multivariado foram mais próximas dos valores reais do que as médias obtidas com os modelos univariados independentes. Uma conclusão mais geral neste sentido pode ser obtida da tabela 5.7. Nela mostramos os Erros Quadráticos Médios ( $MSE$ ) e Erros Absolutos Médios ( $MAE$ ) obtidos para cada uma das séries de vazão observada, com o modelo multivariado condicional e com os modelos univariados independentes. Na tabela estão as médias dos erros correspondentes aos primeiros 221 meses das séries, meses que foram utilizados para fazer inferência sobre os parâmetros (dentro da amostra). Observamos que o ajuste para Taguá ( $A$ ) ficou melhor com o modelo condicional segundo ambos critérios.

Tabela 5.5: Estimativas médias para a média da vazão (valor esperado) com o modelo conjunto e com os modelos univariados independentes em  $t = 12, 37, 125, 200$ .

	$Y_t$ real	média	d.p.	2,5%	25%	50%	75%	97,5%	$\hat{R}$
Modelo Multivariado									
$t = 12$									
$E[Y^A]$	155,67	136,415	3,611	129,264	134,031	136,422	138,848	143,569	1,00
$E[Y^B]$	68,74	60,688	1,555	57,584	59,641	60,712	61,776	63,611	1,01
$E[Y^C]$	41,45	34,850	0,876	33,121	34,233	34,869	35,457	36,505	1,01
$t = 37$									
$E[Y^A]$	135,51	136,176	3,616	129,019	133,779	136,183	138,616	143,334	1,00
$E[Y^B]$	64,42	60,576	1,557	57,462	59,524	60,604	61,670	63,496	1,00
$E[Y^C]$	38,87	34,787	0,878	33,054	34,169	34,805	35,396	36,447	1,01
$t = 125$									
$E[Y^A]$	193,74	242,012	15,793	225,148	235,053	240,879	247,438	262,534	1,06
$E[Y^B]$	91,91	110,046	9,484	103,152	107,071	109,528	112,158	118,702	1,12
$E[Y^C]$	53,29	61,578	1,500	58,779	60,553	61,517	62,554	64,635	1,01
$t = 200$									
$E[Y^A]$	205,73	233,219	15,437	217,550	226,704	232,131	238,145	252,687	1,07
$E[Y^B]$	90,54	105,759	9,363	99,390	103,032	105,266	107,722	113,848	1,12
$E[Y^C]$	56,71	59,107	1,312	56,663	58,209	59,050	59,940	61,796	1,00
Modelos Univariados Independentes									
$t = 12$									
$E[Y^A]$	155,67	119,548	3,156	113,297	117,442	119,545	121,639	125,755	1,03
$E[Y^B]$	68,74	56,599	1,658	53,154	55,498	56,722	57,776	59,599	1,04
$E[Y^C]$	41,45	34,926	0,886	33,222	34,324	34,919	35,503	36,800	1,01
$t = 37$									
$E[Y^A]$	135,51	119,353	3,160	113,092	117,244	119,348	121,444	125,569	1,03
$E[Y^B]$	64,42	56,510	1,660	53,057	55,406	56,631	57,690	59,512	1,04
$E[Y^C]$	38,87	34,787	0,878	33,054	34,169	34,805	35,396	36,447	1,01
$t = 125$									
$E[Y^A]$	193,74	240,679	5,615	230,447	236,891	240,397	244,288	251,946	1,00
$E[Y^B]$	91,91	108,106	2,831	102,811	106,229	108,021	109,885	113,943	1,01
$E[Y^C]$	53,29	61,521	1,494	58,860	60,457	61,434	62,511	64,750	1,00
$t = 200$									
$E[Y^A]$	205,73	239,240	5,516	229,196	235,489	238,944	242,822	250,235	1,00
$E[Y^B]$	90,54	107,909	2,815	102,651	106,038	107,817	109,677	113,717	1,01
$E[Y^C]$	56,71	59,065	1,306	56,706	58,154	58,987	59,914	61,875	1,00

Tabela 5.6: Estimativas pontuais para o valor da vazão com o modelo conjunto e com os modelos univariados independentes em  $t = 35, 80, 140, 200$ .

	$Y_t$ real	média	d.p.	2,5%	25%	50%	75%	97,5%	$\hat{R}$
<b>Modelo Multivariado</b>									
$t = 35$	149,37	154,80	17,25	125,11	142,69	153,95	164,67	194,17	1,00
$t = 80$	241,65	240,09	27,60	192,44	220,27	237,72	257,90	300,01	1,01
$t = 140$	184,90	190,56	22,03	153,09	175,58	189,39	203,63	239,06	1,00
$t = 200$	205,73	205,42	23,22	161,66	190,22	204,97	220,88	253,59	1,00
<b>Modelos Univariados Independentes</b>									
$t = 35$	149,37	156,42	12,96	132,63	147,18	155,95	164,33	183,32	1,00
$t = 80$	241,65	246,58	20,49	209,10	232,38	245,61	259,59	290,81	1,00
$t = 140$	184,90	189,07	15,20	161,72	178,25	187,89	198,56	221,65	1,00
$t = 200$	205,73	202,38	17,26	172,35	190,91	201,59	213,91	237,59	1,00

Tabela 5.7: Erro Quadrático Médio (*MSE*) e Erro Absoluto Médio (*MAE*) obtido com o modelo condicional e os modelos univariados independentes

		<i>MSE</i>	<i>MAE</i>
<b>Modelo Multivariado</b>			
A B	Taguá   Barreiras	120,684	7,025
B C	Barreiras   Redenção	9,060	1,820
C	Redenção	2,568	1,009
<b>Modelos Univariados Independentes</b>			
A	Taguá	162,991	7,756
B	Barreiras	8,047	1,652
C	Redenção	2,568	1,009

## 5.5 Considerações Finais

A abordagem conjunta com funções de transferência, apresentada neste capítulo, mostrou-se bastante adequada para modelar várias séries de vazão de bacias encaixadas. A idéia base é usar a representação condicional da distribuição conjunta. Mostramos que as estimativas pontuais

(média a posteriori) das séries de vazão ficam melhores com o modelo conjunto do que com modelos individuais para cada série. Mostramos também que a incorporação ao modelo, da vazão de uma bacia menor, permite uma melhor interpretação e discriminação dos efeitos de cada variável. Em particular, nesta abordagem temos um parâmetro capaz de representar e medir a captação ou retirada d'água ao longo dos rios.

Um modelo espaço-temporal foi utilizado para ajustar a chuva observada nos postos pluviométricos localizados dentro da bacia total. A distribuição preditiva desse modelo foi utilizada para interpolar a chuva em grades construídas dentro da área de drenagem de cada sub-bacia.

Utilizamos métodos *MCMC* para realizar a inferência em todos os modelos. Aproveitamos o fato de que ao final das iterações de uma rotina *MCMC* contamos com uma amostra da posteriori dos parâmetros do modelo assim como das distribuições preditivas, para levar em conta toda a incerteza associada a nosso procedimento. Conseguimos este objetivo porque os modelos podem ser estimados por partes (condicionalmente independentes) e ao invés de trabalharmos com estimativas pontuais para as covariáveis (por exemplo, o volume de chuva), trabalhamos com uma amostra de toda a sua distribuição a posteriori.

## 5.6 Apêndice I: Algumas Estatísticas da Distribuição Log-Normal Multivariada

Neste apêndice apresentamos algumas relações entre a distribuição conjunta e as distribuições marginais de  $(Y^A, Y^B, Y^C)$ , seguindo a especificação do modelo em (5.12). Fazemos uso de três fatos (propriedades) para, a partir da modelagem conjunta de variáveis com distribuição normal, definir os valores esperados e variâncias das variáveis log-normais de nosso interesse.

Neste apêndice,  $N(\mu, \sigma^2)$  denota uma distribuição Normal univariada com média  $\mu$  e variância  $\sigma^2$ . Valores esperados e variâncias, em geral, são denotados por  $E[\cdot]$  e  $Var[\cdot]$ , respectivamente.

**Fato I:**  $E_t$  é tendência-estacionário (*trend stationary*)

Dado que  $E_t = \rho E_{t-1} + \gamma X_t + w_t$ ,  $w_t$  é um ruído branco com variância  $W$  e  $|\rho| < 1$ , pois  $0 < \rho < 1$ ,  $E_t$  é tendência-estacionário com relação a  $\gamma$  e  $\mathbf{X} = \{X_1, \dots, X_T\}$  e, portanto,

$$E[E_t | \gamma, \mathbf{X}] = \frac{\gamma X_t}{1 - \rho} \quad \text{e} \quad Var[E_t | \gamma, \mathbf{X}] = \frac{W}{1 - \rho^2}, \quad (5.14)$$

pois  $E[E_t | \gamma, \mathbf{X}] = E[E_{t-1} | \gamma, \mathbf{X}]$  e  $Var[E_t | \gamma, \mathbf{X}] = Var[E_{t-1} | \gamma, \mathbf{X}]$  para  $t = 1, \dots, T$ .

**Fato II:** Se  $x \sim N(\mu, \sigma^2)$  e  $y = \exp(x)$  então  $y \sim \text{log-normal}(\mu, \sigma^2)$  e

$$E[y] = \exp(\mu + \sigma^2/2) \quad \text{e} \quad \text{Var}[y] = \exp(2\mu + \sigma^2)[\exp(\sigma^2) - 1], \quad (5.15)$$

**Fato III:** Se  $x_1$  e  $x_2$  têm uma distribuição normal bivariada, tal que  $\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \sim N \left[ \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix} \right]$ , e  $y_1 = \exp(x_1)$  e  $y_2 = \exp(x_2)$ , então  $y_1|y_2$  é log-normal com parâmetros:

$$\mu_1 + \rho \frac{\sigma_1}{\sigma_2} (\log(y_2) - \mu_2) \quad \text{e} \quad \sigma_1^2(1 - \rho^2),$$

e também,

$$\text{Cov}(y_1, y_2) = [\exp(\rho\sigma_1\sigma_2) - 1] \exp[\mu_1 + \mu_2 + (\sigma_1^2 + \sigma_2^2)/2],$$

$$\text{Corr}(y_1, y_2) = [\exp(\rho\sigma_1\sigma_2) - 1] / [(\exp(\sigma_1^2) - 1)(\exp(\sigma_2^2) - 1)]^{1/2}$$

– Sejam  $Z_t^A, Z_t^B$  e  $Z_t^C, t = 1, \dots, T$ , tais que:

$$Z_t^A|Z_t^B \sim N(\mu_t^{A|B}, \sigma_{A|B}^2) \quad \mu_t^{A|B} = \alpha^{A|B} + \eta^{A|B} Z_t^B + E_t^{A|B} \quad (5.16a)$$

$$Z_t^B|Z_t^C \sim N(\mu_t^{B|C}, \sigma_{B|C}^2) \quad \mu_t^{B|C} = \alpha^{B|C} + \eta^{B|C} Z_t^C + E_t^{B|C} \quad (5.16b)$$

$$Z_t^C \sim N(\mu_t^C, \sigma_C^2) \quad \mu_t^C = \alpha^C + E_t^C \quad (5.16c)$$

com  $E_t^{A|B}, E_t^{B|C}$  e  $E_t^C$ , definidos como em (5.12g). Aqui, as médias  $\mu_t^m$  estão condicionadas em  $\alpha^m, \eta^m, Z_t^{m-1}$  e  $E_t^m$ . Uma forma equivalente de representar (5.16) é:

$$Z_t^A|Z_t^B = \alpha^{A|B} + \eta^{A|B} Z_t^B + E_t^{A|B} + v_t^{A|B}, \quad E_t^{A|B} \sim N(\rho^{A|B} E_{t-1}^{A|B} + \gamma^{A|B} X_t^{A|B}, W_{A|B}), \quad v_t^{A|B} \sim N(0, \sigma_{A|B}^2)$$

$$Z_t^B|Z_t^C = \alpha^{B|C} + \eta^{B|C} Z_t^C + E_t^{B|C} + v_t^{B|C}, \quad E_t^{B|C} \sim N(\rho^{B|C} E_{t-1}^{B|C} + \gamma^{B|C} X_t^{B|C}, W_{B|C}), \quad v_t^{B|C} \sim N(0, \sigma_{B|C}^2)$$

$$Z_t^C = \alpha^C + E_t^C + v_t^C, \quad E_t^C \sim N(\rho^C E_{t-1}^C + \gamma^C X_t^C, W_C), \quad v_t^C \sim N(0, \sigma_C^2).$$

– Com base no Fato I e o resultado em (5.14), as médias, variâncias e covariâncias marginais incondicionais de  $Z^A, Z^B$  e  $Z^C$  estão dados por:

### Médias

$$\begin{aligned} \mu_t^C &= E[Z_t^C] = \alpha^C + E[E_t^C] = \alpha^C + \frac{\gamma^C X_t^C}{1 - \rho^C} \\ \mu_t^B &= E[Z_t^B] = E[E[Z_t^B|Z_t^C]] = \alpha^{B|C} + \eta^{B|C} E[Z_t^C] + E[E_t^{B|C}] = \alpha^{B|C} + \eta^{B|C} \mu_t^C + \frac{\gamma^{B|C} X_t^{B|C}}{1 - \rho^{B|C}} \\ \mu_t^A &= E[Z_t^A] = E[E[Z_t^A|Z_t^B]] = \alpha^{A|B} + \eta^{A|B} E[Z_t^B] + E[E_t^{A|B}] = \alpha^{A|B} + \eta^{A|B} \mu_t^B + \frac{\gamma^{A|B} X_t^{A|B}}{1 - \rho^{A|B}} \end{aligned} \quad (5.17)$$

### Variâncias

$$\begin{aligned}
\sigma_C^2 &= \text{Var}[Z^C] = \text{Var}[E_t^C] + \text{Var}[v_t^C] = \sigma_C^2 + \frac{W_C}{1 - \rho_C^2} \\
\sigma_B^2 &= \text{Var}[Z^B] = E[\text{Var}[Z^B|Z^C]] + \text{Var}[E[Z^B|Z^C]] \\
&= \frac{W_{B|C}}{1 - \rho_{B|C}^2} + \sigma_{B|C}^2 + \text{Var}[\alpha^{B|C} + \eta^{B|C}Z_t^C + E_t^{B|C}] = \sigma_{B|C}^2 + \eta_{B|C}^2 \sigma_C^2 + \frac{2W_{B|C}}{1 - \rho_{B|C}^2} \\
\sigma_A^2 &= \text{Var}[Z^A] = E[\text{Var}[Z^A|Z^B]] + \text{Var}[E[Z^A|Z^B]] \\
&= \frac{W_{A|B}}{1 - \rho_{A|B}^2} + \sigma_{A|B}^2 + \text{Var}[\alpha^{A|B} + \eta^{A|B}Z_t^B + E_t^{A|B}] = \sigma_{A|B}^2 + \eta_{A|B}^2 \sigma_B^2 + \frac{2W_{A|B}}{1 - \rho_{A|B}^2}
\end{aligned} \tag{5.18}$$

onde  $\rho_m^2$  e  $\eta_m^2$  denotam o quadrado de  $\eta^m$  e  $\rho^m$ , respectivamente.

### Covariâncias

$$\begin{aligned}
\sigma_{BC} &= \text{Cov}[Z^B, Z^C] = \text{Cov}[\alpha^{B|C} + \eta^{B|C}Z^C + E_t^{B|C} + v_t^{B|C}, Z^C] = \eta^{B|C} \text{Cov}[Z^C, Z^C] = \eta^{B|C} \sigma_C^2 \\
\sigma_{AB} &= \text{Cov}[Z^A, Z^B] = \text{Cov}[\alpha^{A|B} + \eta^{A|B}Z^B + E_t^{A|B} + v_t^{A|B}, Z^B] = \eta^{A|B} \text{Cov}[Z^B, Z^B] = \eta^{A|B} \sigma_B^2 \\
\sigma_{AC} &= \text{Cov}[Z^A, Z^C] = \text{Cov}[\alpha^{A|B} + \eta^{A|B}Z^B + E_t^{A|B} + v_t^{A|B}, Z^C] = \eta^{A|B} \text{Cov}[Z^B, Z^C] = \eta^{A|B} \sigma_{BC}
\end{aligned} \tag{5.19}$$

### Correlações

$$\begin{aligned}
\rho_{BC} &= \text{Corr}[Z^B, Z^C] = \frac{\sigma_{BC}}{\sqrt{\sigma_B^2 \sigma_C^2}} \\
\rho_{AB} &= \text{Corr}[Z^A, Z^B] = \frac{\sigma_{AB}}{\sqrt{\sigma_A^2 \sigma_B^2}} \\
\rho_{AC} &= \text{Corr}[Z^A, Z^C] = \frac{\sigma_{AC}}{\sqrt{\sigma_A^2 \sigma_C^2}}
\end{aligned} \tag{5.20}$$

– Com base no Fato II e assumindo que  $Y^A = \exp(Z^A)$ ,  $Y^B = \exp(Z^B)$  e  $Y^C = \exp(Z^C)$  temos:

$$\begin{aligned}
E[Y^A] &= \exp(\mu^A + 0,5\sigma_A^2) & e & \quad \text{Var}[Y^A] = \exp(2\mu^A + \sigma_A^2)[\exp(\sigma_A^2) - 1], \\
E[Y^B] &= \exp(\mu^B + 0,5\sigma_B^2) & e & \quad \text{Var}[Y^B] = \exp(2\mu^B + \sigma_B^2)[\exp(\sigma_B^2) - 1], \\
E[Y^C] &= \exp(\mu^C + 0,5\sigma_C^2) & e & \quad \text{Var}[Y^C] = \exp(2\mu^C + \sigma_C^2)[\exp(\sigma_C^2) - 1],
\end{aligned}$$

onde  $\mu^m$  e  $\sigma_m^2$  estão definidas em (5.17) e (5.18).

– Com base no Fato III e assumindo que  $Y^A = \exp(Z^A)$ ,  $Y^B = \exp(Z^B)$  e  $Y^C = \exp(Z^C)$  temos:

$$\begin{aligned}
Y^A|Y^B &\sim \text{log-normal}\left[\mu^A + \rho_{AB} \frac{\sigma_A}{\sigma_B} (\log(Y^B) - \mu^B), \sigma_A^2 (1 - \rho_{AB}^2)\right] \\
Y^B|Y^C &\sim \text{log-normal}\left[\mu^B + \rho_{BC} \frac{\sigma_B}{\sigma_C} (\log(Y^C) - \mu^C), \sigma_B^2 (1 - \rho_{BC}^2)\right] \\
Y^C &\sim \text{log-normal}\left[\mu^C, \sigma_C^2\right]
\end{aligned}$$

**Médias**

$$\begin{aligned}
E(Y^A|Y^B) &= \exp\left(\left[\mu^A + \rho_{AB} \frac{\sigma_A}{\sigma_B} (\log(Y^B) - \mu^B)\right] + \sigma_A^2/2(1 - \rho_{AB}^2)\right) \\
E(Y^B|Y^C) &= \exp\left(\left[\mu^B + \rho_{BC} \frac{\sigma_B}{\sigma_C} (\log(Y^C) - \mu^C)\right] + \sigma_B^2/2(1 - \rho_{BC}^2)\right) \\
E(Y^C) &= \exp(\mu^C + \sigma_C^2/2)
\end{aligned} \tag{5.21}$$

**Variâncias**

$$\begin{aligned}
Var(Y^A|Y^B) &= \exp\left(2\left[\mu^A + \rho_{AB} \frac{\sigma_A}{\sigma_B} (\log(Y^B) - \mu^B)\right] + \sigma_A^2(1 - \rho_{AB}^2)\right) \left(\exp(\sigma_A^2(1 - \rho_{AB}^2)) - 1\right) \\
Var(Y^B|Y^C) &= \exp\left(2\left[\mu^B + \rho_{BC} \frac{\sigma_B}{\sigma_C} (\log(Y^C) - \mu^C)\right] + \sigma_B^2(1 - \rho_{BC}^2)\right) \left(\exp(\sigma_B^2(1 - \rho_{BC}^2)) - 1\right) \\
Var(Y^C) &= \exp(2\mu^C + \sigma_C^2) \left[\exp(\sigma_C^2) - 1\right]
\end{aligned} \tag{5.22}$$

**Covariâncias**

$$\begin{aligned}
Cov(Y^A, Y^B) &= \left[\exp(\rho_{AB}\sigma_A\sigma_B) - 1\right] \exp\left[\mu_A + \mu_B + (\sigma_A^2 + \sigma_B^2)/2\right] \\
Cov(Y^A, Y^C) &= \left[\exp(\rho_{AC}\sigma_A\sigma_C) - 1\right] \exp\left[\mu_A + \mu_C + (\sigma_A^2 + \sigma_C^2)/2\right] \\
Cov(Y^B, Y^C) &= \left[\exp(\rho_{BC}\sigma_B\sigma_C) - 1\right] \exp\left[\mu_B + \mu_C + (\sigma_B^2 + \sigma_C^2)/2\right]
\end{aligned} \tag{5.23}$$

onde  $\mu^m$ ,  $\sigma_m^2$  e  $\rho_{m_1, m_2}$  estão definidas em (5.17), (5.18) e (5.20).

## Capítulo 6

### CONSIDERAÇÕES FINAIS E TRABALHOS FUTUROS

#### 6.1 Considerações Finais

Os objetivos formulados para esta tese foram satisfatoriamente atingidos e como principais produtos finais temos 1 artigo completo e 3 em fase de submissão.

O primeiro artigo, intitulado *“Revisiting Distributed Lag Models Through a Bayesian Perspective”*, foi publicado no *“Applied Stochastic Models to Business and Industry”*. Este artigo é o resultado de um exercício no qual foram explorados alguns modelos básicos de função de transferência, com resposta normal, utilizando-se métodos de amostragem implementados no software WinBUGS (Spiegelhalter et al., 2003). Os objetivos específicos deste artigo foram:

- estabelecer uma relação entre os modelos dinâmicos de função de transferência, como definidos em West & Harrison (1997), com os modelos econométricos de defasagens distribuídas; e
- mostrar que a inferência, sob o ponto de vista bayesiano, pode ser realizada com o software WinBUGS.

Com este artigo conseguiu-se fazer uma comparação entre resultados obtidos com modelos dinâmicos e métodos clássicos de co-integração para a análise de séries não estacionárias, assim como realizou-se a comparação entre os resultados obtidos com técnicas MCMC (do WinBUGS) e os obtidos com métodos de integração numérica utilizados antes do desenvolvimento dos métodos MCMC.

O segundo artigo, intitulado *“An Efficient Sampling Scheme for Generalized Dynamic Models,”* surgiu da nossa dificuldade na implementação de alguns algoritmos MCMC propostos na literatura, assim como do tempo computacional requerido para a obtenção dos resultados. Estes dois fatos são relevantes quando se tem interesse em explorar modelos de alta complexidade, onde a amostragem dos parâmetros que variam no tempo pode ser muito custosa e, ao mesmo tempo, pode afetar a convergência das cadeias dos outros (hiper)parâmetros do modelo. Este artigo tem como objetivos:

- mostrar uma forma de utilizar a proposta de West et al. (1985) dentro de um algoritmo *MCMC*,
- comparar a performance do esquema proposto com alguns previamente estabelecidos na literatura; e
- mostrar a aplicação do esquema proposto em dados reais.

O esquema aqui proposto foi denominado CUBS, do inglês *Conjugate Updating Backward Sampling*. Os resultados obtidos mostram que o esquema proposto é eficiente no sentido de reduzir significativamente o tempo computacional e ser de fácil implementação. Milhares de iterações do *MCMC* são realizadas em minutos e as cadeias geradas apresentam menos autocorrelação que as geradas no *WinBUGS*, por exemplo. Este fato permite explorar mais alternativas de modelos e, mais ainda, o esquema foi testado em várias sub-classes dos modelos generalizados dinâmicos com bons resultados.

O terceiro artigo, intitulado “*A Bayesian approach for the rainfall-runoff problem: The case of Rio Grande Basin,*” propõe uma forma muito geral para modelar conjuntamente chuva e vazão, levando em conta as diferenças na forma de medição de ambas variáveis. A principal contribuição deste artigo é a modelagem proposta, que combina um modelo de função de transferência para a vazão, condicionado na chuva da bacia, com um modelo espaço-temporal simultâneo para a chuva medida em várias localizações dentro da área de drenagem do posto de vazão em consideração. A chuva da bacia toda é obtida a partir da solução do problema de mudança de suporte. No artigo mostra-se que, aproveitando a decomposição da distribuição conjunta (em uma distribuição condicional (vazão dada chuva) e uma marginal (chuva)), é possível utilizar e explorar várias alternativas de modelos. Em particular, utilizam-se alguns critérios de comparação de modelos e mostra-se que os modelos fornecem boas previsões, tanto espaciais (para a chuva) quanto temporais (para chuva e vazão). A metodologia proposta foi aplicada em dados da bacia do Rio Grande (BA) cedidos pelo Laboratório de Hidrologia da Universidade Federal do Rio de Janeiro.

O quarto artigo, intitulado “*A Bayesian approach for modelling multiple runoff series,*” propõe uma alternativa de modelagem conjunta de várias séries temporais de vazão. Nossa proposta é útil principalmente para modelar vazões de áreas de drenagens encaixadas. A idéia é usar a representação condicional da distribuição conjunta das séries e incorporar o nível de vazão das bacias menores como covariáveis no valor esperado da vazão das bacias maiores. A principal contribuição deste artigo é a modelagem proposta, que estende a classe de modelos de função

de transferência, apresentada no artigo anterior, para uma versão multivariada. Esta abordagem permite também a modelagem conjunta com a chuva. Um modelo espaço-temporal para a chuva observada em várias localizações dentro da área de drenagem total pode ser ajustado. A chuva associada a cada sub-bacia pode ser obtida a partir da solução do problema de mudança de suporte. O fato importante a ser levado em conta aqui é a construção de grades de interpolação que respeitem os limites de cada sub-bacia. As propostas foram aplicadas aos dados da bacia do Rio Grande (BA).

## 6.2 Trabalhos Futuros

### 6.2.1 Sobre o Esquema Proposto para Modelos Generalizados Dinâmicos

Como trabalho futuro desta tese, pretendemos estender o esquema proposto no capítulo 3 para modelos lineares dinâmicos generalizados onde mais de um dos parâmetros da distribuição da observação tem algum tipo de evolução no tempo ou estão relacionados com outras variáveis através de relações funcionais conhecidas. Nesta seção apresentamos alguns detalhes sobre o desenvolvimento desta proposta.

#### Motivação

Fazendo algumas pequenas adaptações, o esquema de amostragem proposto no capítulo 3 pode ser utilizado no procedimento de inferência de classes de modelos que estão sendo discutidos recentemente. Nesta seção fazemos um resumo de trabalhos que servem de motivação para a extensão do *CUBS*.

A classe denominada como modelos lineares generalizados hierárquicos (*HGLM*) foi introduzida em Lee & Nelder (1996, 2001) e pode ser vista como uma síntese de três classes muito conhecidas: os modelos lineares generalizados, os modelos lineares mistos e os modelos com termos de dispersão estruturados. Se  $Y$  é uma variável resposta e  $u$  uma componente aleatória (não observável), no contexto dos modelos *HGLM* assume-se que  $p(Y|u)$  pertence à família exponencial com parâmetro canônico  $\eta$  e parâmetro de dispersão  $\phi$  e  $u$  segue  $p(u)$ , uma distribuição conjugada de  $p(Y|u)$ . Alguns casos particulares são: normal-normal, Poisson-gama, binomial-beta e gama-inversa-gama, onde a primeira distribuição refere-se a  $p(Y|u)$  e a segunda a  $p(u)$ . Nos *HGLM*, as covariáveis e os efeitos aleatórios estão relacionados com o predictor linear  $g(\eta)$ . Embora a análise bayesiana completa via métodos *MCMC* é factível de se realizar para os *HGLM*, levando em conta toda a incerteza associada e sem a especificação de distribuições conjugadas,

Lee & Nelder (1996, 2001) propuseram fazer uso da análise conjugada e da *h-likelihood* para obter estimativas dos parâmetros de interesse. Eles afirmam que a abordagem proposta pode ser vista como uma estatística a posteriori usando distribuições a posteriori impróprias. Contudo, algumas aproximações numéricas são necessárias e somente os dois primeiros momentos das distribuições dos parâmetros são estimados. Mais recentemente, Lee & Nelder (2006) propuseram a classe de modelos denominada *DHGLM*, abreviação do inglês *double hierarchical generalized linear models*, onde os efeitos aleatórios podem ser especificados para a média e a dispersão da distribuição da resposta e o mesmo procedimento de inferência dos *HGLM* pode ser utilizado.

A estrutura dos *HGLM* é similar à dos modelos lineares generalizados dinâmicos (*DGLM*) de West et al. (1985). A ligeira diferença é onde é considerada a distribuição (a priori) conjugada. Nos *DGLM* corresponde à distribuição do parâmetro canônico  $\eta$ , entretanto nos *HGLM* corresponde à distribuição do termo aleatório  $u$ . Levando em conta esta discussão, consideramos que, do ponto de vista bayesiano, o *Conjugate Updating* proposto por West et al. (1985) pode ser utilizado para fazer inferência em modelos da classe *HGLM*.

Por outro lado, outra extensão bastante útil dos modelos lineares generalizados são aqueles que consideram uma estrutura ou relação linear em alguma transformação de mais de um dos parâmetros da distribuição da variável resposta. Por exemplo, Cepeda & Gamerman (2003) apresentaram uma metodologia bayesiana para ajustar modelos lineares nos dois parâmetros das distribuições pertencentes a família exponencial bi-paramétrica, como a gama e a beta. A proposta baseia-se no trabalho de Dey et al. (1997), que estabeleceu a condição de ortogonalidade dos parâmetros desses tipos de distribuições quando é utilizada a parametrização na média. Esta característica permitiu aos autores implementar algoritmos *MCMC* que estimam os coeficientes de ambos modelos lineares de forma iterativa. Outro exemplo nesta classe de modelos é o trabalho de Vasconcellos & Cribari-Neto (2005), que propõe uma classe de modelos de regressão beta onde ambos parâmetros da distribuição estão relacionados com covariáveis. Este artigo é uma extensão de Ferrari & Cribari-Neto (2004) onde a verossimilhança é escrita em termos da sua média e precisão e a estimação via máxima verossimilhança é utilizada para fazer inferência sobre os coeficientes dos modelos lineares.

Adicionalmente, Rigby & Stasinopoulos (2005) apresentam uma classe de modelos de regressão denominada *GAMLSS*, abreviação do inglês *generalized additive model for location, scale and shape*. Eles assumiram que a distribuição da variável resposta  $p(Y|\theta)$ , pertence a uma classe de distribuições muito mais geral que a família exponencial e modelam todos os seus parâmetros via modelos aditivos. Estes autores maximizam a verossimilhança (penalizada) usando os algoritmos de

Newton-Raphson e *Fisher scoring*.

Dado que o *Conjugate Updating* de West et al. (1985) baseia-se na análise conjugada da família exponencial, fazendo uso das famílias conjugadas apresentadas em Bernardo & Smith (1994, p.266), este algoritmo pode ser facilmente estendido para o caso de distribuições  $K$ -paramétricas. Em outras palavras, o *Conjugate Updating* pode ser útil para a inferência em modelos lineares dinâmicos onde a distribuição da resposta pertence à família exponencial com  $K$  parâmetros e cada um deles está relacionado com covariáveis através de várias funções de ligação ou simplesmente varia ao longo do tempo. De fato esta classe de modelos envolve várias das classes descritas acima e esperamos que o esforço computacional com esta idéia deva ser muito menor que com alguma outra técnica padrão de *MCMC*.

### **CUBS na Família Exponencial $K$ -paramétrica**

Agora voltamos nossa atenção aos modelos lineares dinâmicos cuja variável resposta segue uma distribuição na família exponencial  $K$ -paramétrica e onde cada parâmetro está relacionado com covariáveis através de uma função (de ligação) conhecida. Seja  $Y_t$  uma observação no instante  $t$  de uma variável univariada com distribuição na família acima mencionada e seja  $\boldsymbol{\psi}_t$  o vetor de parâmetros. Supondo que existem  $J$  funções de  $\boldsymbol{\psi}_t$  denotadas por  $h_j(\boldsymbol{\psi}_t)$ ,  $j = 1, \dots, J$  ( $J < K$ ), então os modelos de nosso interesse são tais que

$$Y_t | \boldsymbol{\psi}_t \sim p(Y_t | \boldsymbol{\psi}_t) = \exp\{Y_t' \boldsymbol{\psi}_t - b(\boldsymbol{\psi}_t) + a(Y_t)\}, \quad t = 1, \dots, N \quad (6.1a)$$

$$h_j(\boldsymbol{\psi}_t) = F_{jt}' \boldsymbol{\theta}_{jt}, \quad j = 1, \dots, J \quad (6.1b)$$

$$\boldsymbol{\theta}_{jt} = G_{jt} \boldsymbol{\theta}_{j,t-1} + w_{jt}, \quad w_{jt} \sim [\mathbf{0}, \mathbf{W}_j] \quad (6.1c)$$

$$\boldsymbol{\theta}_j \sim [\mathbf{m}_{j0}, \mathbf{C}_{j0}],$$

onde  $\theta_{jt}$  são variáveis latentes ou parâmetros de estado. Repare que no modelo em (6.1) a dependência na informação passada não é explícita, ou seja  $D_t = \{D_{t-1}, Y_t\}$  não está denotada explicitamente na equação acima. Nosso objetivo é estender o esquema de amostragem *CUBS*, apresentado na seção 3.2, para ser utilizado no ajuste modelos dinâmicos lineares como em (6.1).

Lembrando que a idéia principal da primeira etapa do *CUBS* (o *Conjugate Updating*) é introduzir uma distribuição a priori conjugada para o parâmetro canônico da distribuição da resposta com a finalidade de aproximar os momentos *on-line*, a forma natural de estender esse algoritmo é fazendo uso da distribuição conjugada multivariada (a priori),  $p(\boldsymbol{\psi}_t | \phi_t, \boldsymbol{\tau}_t)$ , para a distribuição  $p(Y_t | \boldsymbol{\psi}_t)$  como

é apresentada em Bernardo & Smith (1994, Cap. 5):

$$\boldsymbol{\psi}_t \sim p(\boldsymbol{\psi}|\phi_t, \boldsymbol{\tau}_t) = \exp\{\phi_t[\boldsymbol{\tau}'_t \boldsymbol{\psi}_t - b(\boldsymbol{\psi}_t)] + c(\phi_t, \boldsymbol{\tau}_t)\}. \quad (6.2)$$

Lembrando também no *CUBS*, o *Conjugate Updating* é combinado com o *Backward Sampling* para obter amostras da distribuição a posteriori do vetor de estados, o novo *CUBS* (*ECUBS*) pode ser utilizado para obter amostras da distribuição a posteriori de  $\boldsymbol{\theta}_{jt}, j = 1, \dots, J; t = 1, \dots, T$  em bloco. Finalmente, amostras das outras distribuições a posteriori (dos outros parâmetros do modelo) podem ser obtidas com os métodos padrões do MCMC. O *ECUBS* é detalhado no Quadro 3.

### Quadro 3: MCMC + ECUBS

- (1). **Inicialize:** Atribua valores iniciais  $\boldsymbol{\theta}_1^{(0)}, \dots, \boldsymbol{\theta}_j^{(0)}, \boldsymbol{\psi}^{(0)}$  e faça  $i = 1$ , onde  $i$  é o contador de iterações;
- (2). **Amostre**  $\boldsymbol{\theta}_1^{(i)}, \dots, \boldsymbol{\theta}_j^{(i)}$  usando *CUBS*:
  - (a) Calcule os momentos de  $p(\boldsymbol{\theta}_{jt}|D_t, \boldsymbol{\psi}^{(i-1)})$ ,  $\mathbf{m}_j^{(i)}$  e  $\mathbf{C}_j^{(i)}$ , com o *Conjugate Updating*;
  - (b) Amostre  $\boldsymbol{\theta}_j^*$  com o *Backward Sampling*.
    - i. Amostre  $\boldsymbol{\theta}_{jN}^*$  de  $p(\mathbf{m}_{jN}^{(i)}, \mathbf{C}_{jN}^{(i)})$
    - ii. Amostre  $\boldsymbol{\theta}_{jt}^*, t = N - 1, \dots, 1$ , de  $p(\boldsymbol{\theta}_{jt}|\boldsymbol{\theta}_{j,t+1}^*, \boldsymbol{\psi}^{(i-1)})$
  - (c) Faça  $\boldsymbol{\theta}_j^{(i)} = \boldsymbol{\theta}_j^*$  com probabilidade  $p_t$  e  $\boldsymbol{\theta}_j^{(i)} = \boldsymbol{\theta}_j^{(i-1)}$  com probabilidade  $1 - p_t$ , onde  $p_t = \min(1, A)$  e  $A = \frac{\pi(\boldsymbol{\theta}_j^*)}{\pi(\boldsymbol{\theta}_j^{(i-1)})} \frac{q_t(\boldsymbol{\theta}_j^{(i-1)}|\boldsymbol{\theta}_j^*)}{q_t(\boldsymbol{\theta}_j^*|\boldsymbol{\theta}_j^{(i-1)})}$  onde  $\boldsymbol{\theta}^*$  é um valor candidato para  $\boldsymbol{\theta}$  e  $q(\cdot)$  e  $\pi(\cdot)$  denotam as densidades proposta e objetivo respectivamente.
- (3). **Amostre**  $\boldsymbol{\psi}^{(i)}$  usando, em geral, um passo Metropolis-Hastings ;
- (4). **Atualize:** faça  $i = i + 1$  e volte para (2) até a convergência.

Para obter os momentos do passo (2)a no quadro 3, os seguintes passos devem ser seguidos:

- (1). calcular a média e variância a priori de cada modelo linear
- (2). calcular os parâmetros da priori conjugada multivariada
  - (a) usando a média e variância da priori conjugada (quando e totalmente conhecida), ou
  - (b) usando a moda e curvatura da priori conjugada (quando somente os dois primeiros momentos são conhecidos)
- (3). calcular os parâmetros da posteriori conjugada
- (4). calcular a média e variância a posteriori de cada modelo linear.

No apêndice C apresentamos o algoritmo desenvolvido e um exemplo com dados artificiais para o caso em que a variável resposta segue uma distribuição normal.

### 6.2.2 Sobre a Relação Chuva-Vazão

Na hidrologia, o total de vazão é definido pela equação de balanço hídrico:

$$\text{vazão} = \text{chuva} + \underbrace{\text{evaporação} + \text{interceptação}} + \Delta\text{estoque.} \quad (6.3)$$

A evaporação e a interceptação são controlados principalmente pela vegetação da bacia, portanto a equação (6.3) pode ser reescrita como:

$$\text{vazão} = \text{chuva} + \text{vegetação} + \Delta\text{estoque.} \quad (6.4)$$

Como, em média, a variação do estoque,  $\Delta\text{estoque}$ , deve ser zero, ela pode ser considerada como um termo aleatório do modelo. Sendo assim, a incorporação de uma variável que represente a vegetação, nos modelos chuva-vazão apresentados nesta tese, seria bastante conveniente porque teríamos uma representação mais próxima do processo de formação de vazão. Neste sentido, pretendemos utilizar alguns índices de vegetação calculados a partir de imagens satelitais, como o *EVI* (abreviação do inglês: *Enhanced Vegetation Indice*) e *NDVI* (abreviação do inglês: *Normalized Difference Vegetation Indice*), como covariável. O efeito dessa covariável também pode ser representado por uma função de transferência.

Em outras situações dispõe-se de dados de vazão de várias bacias com características geológicas, climatológicas e de vegetação parecidas, mas que não necessariamente estão localizadas no curso

do mesmo rio e, portanto, as áreas de drenagem correspondentes são totalmente independentes. Contudo, as similaridades físicas fazem com que a resposta da vazão à chuva seja também parecida entre elas e portanto a mesma especificação do modelo, por exemplo, de função de transferência, poderia ser ajustada a cada bacia de forma independente. Contudo, nessa situação é bastante razoável pensar que os parâmetros desses modelos estejam relacionados (por exemplo, pertençam a mesma distribuição de probabilidade) e que a utilização da informação de todas as bacias simultaneamente resulte numa estimação mais robusta de cada parâmetro do modelo. Nesta situação leva a pensar na modelagem hierárquica das séries de vazão.

Por outro lado, como mencionamos na seção 5.1, às vezes é possível contar com dados de regiões hidrologicamente similares onde a mesma especificação de um modelo chuva-vazão pode ser utilizada. A nossa proposta para este caso é fazer a modelagem conjunta de todas as séries temporais de vazão mediante modelos hierárquicos. Uma das vantagens dos modelos hierárquicos é a utilização de toda a informação disponível, simultaneamente, para obter estimativas mais acuradas de alguns dos parâmetros de interesse. A idéia básica é “emprestar” informação entre as diferentes bacias, resultando num ganho na qualidade das estimativas dos parâmetros que suportam uma representação hierárquica. Vale a pena ressaltar que a principal hipótese de um modelo hierárquico é a de permutabilidade das unidades, daí a necessidade que as bacias sejam parecidas entre si. As similaridades devem ser, principalmente, de tipo e uso do solo, clima e vegetação.

Para diferenciar as modelagens propostas, agora a chuva e vazão na bacia  $k$  no tempo  $t$  são denotadas por  $Y_{k,t}$  e  $X_{k,t}$ , respectivamente, e  $K$  é o número de bacias sob estudo. Como caso particular temos que o modelo de função de transferência apresentado na seção 4.3, equação (4.1), pode ser facilmente estendido para uma modelagem conjunta hierárquica da vazão em  $K$  bacias da seguinte maneira:

$$Y_{k,t} \sim p(Y_{k,t}|\mu_{k,t}, \phi_{k,t}), \quad t = 1, 2, \dots; \quad k = 1, \dots, K, \quad (6.5a)$$

$$g(\mu_{k,t}) = f_1(\alpha_k, E_{k,t}) \quad (6.5b)$$

$$E_{k,t} = f_2(E_{k,t-1}, \gamma_k, X_{k,t}) \quad (6.5c)$$

$$\alpha_k = f_3(\alpha) \quad (6.5d)$$

$$\gamma_k = f_4(\gamma), \quad (6.5e)$$

onde  $p(Y_{k,t}|\mu_{k,t}, \phi_{k,t})$  é uma distribuição em  $\mathbb{R}^+$  e  $\mu_{k,t}$  é o valor esperado de  $Y_{k,t}$ . Sob este modelo, assumimos que alguns dos parâmetros do modelo, por exemplo,  $\alpha_k$  e  $\gamma_k$  das  $K$  bacias, são funções (têm algum valor em torno) de médias comuns denotadas por  $\alpha$  e  $\gamma$ , respectivamente. As funções

$f_j(\cdot)$ ,  $j = 1, \dots, 4$ , são conhecidas e descrevem os processos físicos envolvidos na relação chuva-vazão. De maneira análoga ao modelo condicional proposto em (5.4), o modelo em (6.5) pode ser ajustado conjuntamente com um modelo para  $X_{k,t}$ . Mas, neste caso, modelos diferentes para cada bacia podem ser necessários. Logo, entre as características do modelo em (6.5) estão: (i) Permite compartilhar informação entre as bacias e, em consequência, ter estimativas mais acuradas de  $\alpha$  e  $\gamma$ , e (ii) permite levar em conta toda a incerteza associada aos processos chuva-vazão.

Por outro lado, muitos reservatórios têm sido construídos e são operados sob diferentes condições climáticas e hidrológicas e com vários propósitos diferentes, incluindo o fornecimento d'água para agricultura e a indústria. Estes múltiplos objetivos não são a única dificuldade associada ao manejo de reservatórios. A vazão afluyente tem muita incerteza e normalmente varia ao longo do tempo. A demanda d'água e as prioridades de abastecimento podem mudar de um período a outro. As consequências das decisões tomadas podem afetar diferentes grupos em diferentes formas. Estas são algumas das características do problema de operação de reservatórios que fazem necessário a utilização de algum método eficiente e seguro para o planejamento da sua operação. Como trabalho futuro pretendemos utilizar o método bayesiano descrito em Rios-Insua et al. (1997) para resolver o problema de decisão que critérios múltiplos relacionado com a operação de reservatórios.

## Apêndice A

**ABREVIACÕES E DEFINIÇÕES****Abreviações**

<i>ARDL</i>	Auto-Regressive Distributed Lag Model
<i>ARMA</i>	Auto-Regressive Moving Average
<i>ARMS</i>	Adaptive Rejection Metropolis Sampling
<i>CUBS</i>	Conjugate Updating and Backward Sampling
<i>DIC</i>	Deviance Information Criterion
<i>DLM</i>	Dynamic Linear Model
<i>DGLM</i>	Dynamic Generalized Linear Model
<i>EKS</i>	Extended Kalman Smoother
<i>EPD</i>	Expected Predictive Deviance
<i>FFBS</i>	Forward Filtering Backward Sampling
<i>GAMLSS</i>	Generalized Additive Model for Location, Scale and Shape
<i>HGLM</i>	Hierarchical Generalized Linear Models
<i>MAE</i>	Mean Absolute Error
<i>MCMC</i>	Monte Carlo Markov Chain.
<i>MSE</i>	Mean Square Error
<i>OLS</i>	Ordinary Least Square
<i>PSRF</i>	Potential Scale Reduction Factors
<i>RMSE</i>	Root Mean Square Error
<i>TSSE</i>	Time series Standard Error
<i>VAR</i>	Vector Auto-Regression
<i>VECM</i>	Vector Error Correction Models
d.c.c	Distribuição Condicional Completa
FT	Função de Transferência

### Alguns Termos Estatísticos

**DIC** - É um critério para realizar seleção de modelos proposto por Spiegelhalter et al. (2001). O *DIC* é dado por  $DIC = \bar{D} + p_D = D(\bar{\theta}) + 2p_D$ , onde  $\bar{D}$  é a média a posteriori da função desvio (*deviance*) dada por  $D = -2 \log[f(y|\theta)]$ , onde  $f(y|\theta)$  é a verossimilhança condicionada avaliada nos valores amostrados dos parâmetros.  $D(\bar{\theta})$  é obtido ao substituir a média a posteriori de  $\theta$  na função desvio, de forma que  $D(\bar{\theta}) = -2 \log[f(y|\bar{\theta})]$ , e  $p_D = \bar{D} - D(\bar{\theta})$ . O modelo que minimiza o *DIC* é selecionado.

**EPD** - É um critério para realizar seleção de modelos proposto por Gelfand & Ghosh (1998). O *EPD* é obtido minimizando a perda a posteriori de um dado modelo. Quando se considera uma função de perda quadrática e a distribuição normal para os dados, o *EPD* pode ser calculado explicitamente. Nesse caso, se  $Y_{i,rep}$  representa valores replicados do  $i$ -ésimo dado observado,  $i = 1, \dots, n, y_{i,obs}$ , temos  $EPD = \sum_{i=1}^n \sigma_i^2 + \frac{\kappa}{\kappa+1} \sum_{i=1}^n (\mu_i - y_{i,obs})^2$ , onde  $\mu_i = E[Y_{i,rep}|y_{i,obs}]$  e  $\sigma_i^2 = Var[Y_{i,rep}|y_{i,obs}]$ , são a média e a variância da distribuição preditiva, respectivamente. O modelo que minimiza o *EPD* é selecionado.

**MCMC** - São uma classe de algoritmos utilizados geralmente para gerar amostras de distribuições complexas (Gelman & Lopes, 2006). Os métodos *MCMC* consistem na construção de cadeias de Markov que, sob certas condições de regularidade, tenham como distribuição estacionária a densidade de interesse. Em particular, na inferência bayesiana, os métodos *MCMC* são utilizados para gerar amostras da distribuição a posteriori dos parâmetros. Na prática, uma vez que as cadeias geradas atingem a convergência, considera-se que os valores gerados são amostras da distribuição de interesse.

**Amostrador de Gibbs** - É um algoritmo *MCMC* introduzido por Gelfand & Smith (1990). Consiste em:

- (1). Dar valores iniciais aos parâmetros desconhecidos.
- (2). Amostrar valores das densidades condicionais completas de cada parâmetro, respeitando uma ordem de amostragem e condicionando a amostragem no último valor gerado dos outros parâmetros.
- (3). Repetir o passo (2)  $N$  vezes.

**Metropolis-Hastings** - É um algoritmo *MCMC*. Utiliza-se quando as densidades condicionais completas não tem forma fechada conhecida. O algoritmo foi originalmente proposto por Metropolis et al. (1953) e modificado por Hastings (1970). Suponha que  $\pi(x)$  é a densidade de interesse e  $x$  é o valor corrente da cadeia de Markov. O algoritmo de Metropolis-Hastings consiste em:

- (1). Gerar um valor candidato,  $x^*$ , a partir de uma densidade proposta, ou núcleo de transição,  $q(x^*|x)$
- (2). Aceitar o valor gerado com probabilidade  $\min\left\{1, \frac{\pi(x^*)q(x|x^*)}{\pi(x)q(x^*|x)}\right\}$ .

Em geral, na prática, um passo de Metropolis-Hastings é inserido dentro do amostrador de Gibbs para gerar as amostras. A densidade  $q(\cdot)$  deve ser relativamente simples de amostrar. A taxa de convergência depende da proximidade entre  $\pi(\cdot)$  e  $q(\cdot)$ .

**Slice sampling** - É um método para gerar valores de distribuições que tomem valores num intervalo fechado e limitado. Suponha que  $\pi(x)$ ,  $x \in A \subseteq \mathbb{R}$  é a densidade de interesse. A idéia básica é gerar valores da distribuição uniforme definida pela região abaixo de  $\pi(x)$ , e considerar apenas as coordenadas horizontais. Especificamente, considere a região bi-dimensional abaixo de  $\pi(x)$  ou de  $g(x) = c\pi(x)$ , então:

- Seja  $z$  uma variável auxiliar tal que  $z|x \sim U(0, g(x))$ ;
- A distribuição conjunta de  $(z, x)$  é uniforme na região  $\{(z, x) : 0 \leq z \leq g(x)\}$  com densidade  $p(z, x) = \begin{cases} \frac{1}{c} & \text{se } 0 \leq z \leq g(x) \\ 0 & \text{caso contrário.} \end{cases}$
- A distribuição condicional de  $x|z$  é  $p(x|z) \propto p(z, x)$ , ou seja,  $(x|z) \sim U(S(z))$ , onde  $S(z) = \{x : g(x) \geq z\}$ . Logo,  $S(x)$  é a união dos intervalos que constituem a fatia (*slice*) através da densidade definida por  $z$ .

A estrutura acima nos conduz a simular  $x$  e  $z$  usando o amostrador de Gibbs:

- Gerar  $z^{(i)} \sim U(0, g(x^{(i-1)}))$
- Gerar  $x^{(i)} \sim U(S(z^{(i)}))$ ,

onde o índice  $(i)$  denota a  $i$ -ésima iteração e  $S(z^{(i)}) = \{x : g(x) \geq z^{(i)}\}$ . Dentre as vantagens desse método estão: aplica-se a várias distribuições, não necessita especificar uma densidade

proposta como no Metropolis-Hastings e só utiliza a distribuição uniforme para gerar os valores. A principal desvantagem é que a determinação de  $S(z)$  pode ser difícil.

**PSRF ou  $\hat{R}$**  - É um critério para verificar a convergência de duas ou mais cadeias paralelas geradas por um algoritmo *MCMC*, proposto em Gelman & Rubin (1992). A convergência é diagnosticada quando as cadeias estão em regiões distantes dos seus valores iniciais e as amostras de diferentes cadeias são indistinguíveis. Suponha que foram geradas  $m$  cadeias paralelas, cada uma com  $n$  elementos  $x$ , ou seja, temos:  $x_i^{(1)}, x_i^{(2)}, \dots, x_i^{(n)}$ ,  $i = 1, \dots, m$ . Este critério está baseado na comparação da variância entre as cadeias,  $V_E$ , e da variância dentro das cadeias,  $V_D$ , dadas por:

$$V_E = \frac{n}{m-1} \sum_{i=1}^m (\bar{x}_i - \bar{x})^2$$

$$V_D = \frac{1}{m(n-1)} \sum_{i=1}^m \sum_{j=1}^n (x_i^{(j)} - \bar{x}_i)^2,$$

onde  $\bar{x}_i$  é a média dos elementos da cadeia  $i$  e  $\bar{x}$  é a média dos elementos de todas as cadeias. Sob convergência, todos os  $mn$  elementos provém da mesma distribuição e, portanto, a variância de  $x$ ,  $V_x$  pode ser estimada de forma consistente por

$$\hat{V}_x = \left(1 - \frac{1}{n}\right)V_D + \frac{1}{n}V_E.$$

Se a convergência não for atingida, o fator potencial de escala (*PSRF*) definido por  $\hat{R} = \sqrt{\frac{\hat{V}_x}{V_D}}$  será sempre maior que 1. Segundo o teorema de ergodicidade, ver Gamerman & Lopes (2006, pag. 125), quando  $n \rightarrow \infty$ ,  $\hat{V}_x$  e  $\hat{V}_D$  convergem para  $V_x$  e  $\hat{R} \rightarrow 1$ . Logo, Se  $\hat{R} \approx 1$  concluímos que há indícios de convergência.

### Alguns Termos Hidrológicos

**Área de drenagem** - É a área do terreno que contribui para a vazão de um curso d'água, a montante de um determinado ponto. Usualmente é expressa em  $km^2$  ou em hectares.

**Bacia hidrográfica** - Conjunto de terras drenadas por um rio principal e seus afluentes. A noção de bacia hidrográfica inclui naturalmente a existência de cabeceiras ou nascentes, divisores d'água, cursos d'água principais, afluentes, subafluentes, etc. Área geográfica que drena suas águas para um mesmo local, geralmente um rio.

**Balço hídrico** - Equação que relaciona os diversos componentes do ciclo hidrológico em uma determinada área ou bacia, tais como precipitação, evapotranspiração, escoamento etc.

**Ciclo hidrológico** - É o processo da circulação das águas da Terra, que inclui os fenômenos de evaporação, precipitação, transporte, escoamento superficial, infiltração, retenção, evaporação do solo, do mar e das águas continentais; condensação para formar nuvens; acumulação no solo ou nas massas de água; escoamento direto ou retardado para o mar e reevaporação.

**Chuva** - Precipitação de água líquida seja sob a forma de gotas de diâmetro superior a 0.5mm, seja sob a forma de gotículas menores, largamente dispersas. É o resultado da condensação na atmosfera que caem em direção ao solo, quando as gotas superam as correntes verticais de ar. Normalmente é medida a partir da altura da precipitação em milímetros.

**Escoamento** - Parte da precipitação que escoar para um curso d'água pela superfície do solo (escoamento superficial) ou pelo interior do mesmo (escoamento subterrâneo).

**Hidrografia** - Conjunto de canais e rios que formam a drenagem de uma região ou bacia.

**Q** - em hidrologia, símbolo usado para vazão.

**Reservatório** - Massa de água formada por retenção; por exemplo, à montante de uma barragem.

**Vazão** - Volume d'água ou afluente que passa, na unidade de tempo, através de uma seção, pré-determinada. Pode ser expressa em milímetros de água por dia, mês ou ano, estendidos sobre a área da bacia hidrográfica.

**Vazão unitária** - ou vazão específica - Relação entre a vazão natural e a área de drenagem (da bacia hidrográfica) relativa a uma seção de um curso d'água. É expressa em litros por segundo por quilômetros quadrado ( $l/s/km^2$ ).

## Apêndice B

### DISTRIBUIÇÕES CONDICIONAIS COMPLETAS E MÉTODOS DE AMOSTRAGEM

Neste apêndice apresentamos as distribuições condicionais completas e os algoritmos amostradores de Gibbs utilizados para alguns dos modelos estudados nesta tese. Para simplificar a notação, apresentamos modelos para resposta com distribuição Normal ao invés de resposta LogNormal, pois sabemos que se  $Y \sim \text{LogNormal}(\mu, \sigma^2)$  então  $X = \log(Y) \sim \text{Normal}(\mu, \sigma^2)$ . Daí que os algoritmos aqui apresentados são aplicáveis os dados de vazão sob uma transformação logarítmica (quando a distribuição LogNormal é considerada),

Neste capítulo utilizamos a seguinte notação:

- $N(a, b)$  representa uma distribuição normal com média  $a$  e variância  $b$ .
- $U[a, b]$  representa uma distribuição uniforme no intervalo  $[a, b]$ .
- $G(a, b)$  representa uma distribuição gama com parâmetro de forma  $a$  e parâmetro de escala  $b$ .
- $IG(a, b)$  representa uma distribuição gama invertida com parâmetro de forma  $a$  e parâmetro de escala  $b$ .
- $\text{Gama}(\mu, \nu)$  representa uma distribuição gama com média  $\mu$  parâmetro de forma  $\nu$ .
- $NM_d(\mathbf{a}, \mathbf{b})$  representa uma distribuição normal  $d$ -variada com média  $\mathbf{a}$  e matriz variância-covariância  $\mathbf{b}$ .

#### B.1 Modelo Linear Normal de Função de Transferência Não Estocástica

Considere o seguinte modelo:

$$Y_t = \alpha + E_t + v_t, \quad v_t \sim N(0, V) \quad (\text{B.1a})$$

$$E_t = \rho E_{t-1} + \gamma X_t, \quad t = 1, \dots, T, \quad (\text{B.1b})$$

e a suposição que, a priori, os parâmetros são independentes e tais que:  $E_0 \sim N(a_0, b_0)$ ,  $\alpha \sim N(a_1, b_1)$ ,  $\rho \sim U[0, 1]$ ,  $\gamma \sim N(a_3, b_3)$  e  $V \sim IG(a_4, b_4)$ . O amostrador de Gibbs pode ser utilizado para amostrar das distribuições condicionais completas dos parâmetros do modelo em (B.1). Mas, como a condicional completa de  $\rho$  não tem forma fechada conhecida, é necessário utilizar algum método indireto de amostragem. As distribuições condicionais completas das quantidades desconhecidas em (B.1) e as alternativas de amostragem consideradas nesta tese, são apresentadas abaixo.

- Para  $E_0, \alpha, \gamma$ :

Sejam  $\boldsymbol{\theta} = (E_0, \alpha, \gamma)'$  e  $F$  uma matriz  $T \times 3$  tal que:

$$\begin{array}{ll} \text{Para } t = 1 : & \text{Para } t = 2, \dots, T : \\ F[1, 1] = 1, & F[1, t] = 1 \\ F[2, 1] = \rho, & F[2, t] = \rho \times F[2, t - 1] \\ F[3, 1] = X_1 & F[3, t] = \rho \times F[3, t - 1] + X_t. \end{array}$$

Então, o modelo em (B.1) pode ser re-escrito como:

$$\mathbf{Y} = F' \boldsymbol{\theta} + \mathbf{v}, \quad \mathbf{v} \sim NM_T(\mathbf{0}, V), \quad (\text{B.2})$$

onde  $NM_T$  denota uma distribuição Normal Multivariada de dimensão  $T$ . Se  $\boldsymbol{\theta} \sim NM_3(\mathbf{a}, \mathbf{R})$  onde  $\mathbf{a} = (a_1, a_0, a_3)'$  e  $\mathbf{R} = \text{diag}(b_1, b_0, b_3)$  é a distribuição a priori para  $\boldsymbol{\theta}$ , tem-se

$$\boldsymbol{\theta} | \mathbf{Y} \sim NM_3(\boldsymbol{\mu}_\theta, \boldsymbol{\Sigma}_\theta), \quad (\text{B.3})$$

onde:  $\boldsymbol{\mu}_\theta = \boldsymbol{\Sigma}_\theta (FV^{-1}\mathbf{Y} + \mathbf{R}^{-1}\mathbf{a})$  e  $\boldsymbol{\Sigma}_\theta = (\mathbf{R}^{-1} + FV^{-1}F')^{-1}$

- Para  $\rho$ :

A distribuição condicional completa para  $\rho$  é dada por:

$$p(\rho | \cdot) \propto \prod_{t=1}^T [p(Y_t | E_0, \alpha, \rho, \gamma, V)] p(\rho) = \exp\left\{-\frac{1}{2} \left[ \frac{1}{V} \sum_{t=1}^T (Y_t - \alpha - E_t)^2 \right]\right\} I_{[0,1]},$$

onde  $E_t = \rho^t E_0 + \gamma \sum_{j=0}^t \rho^j X_{t-j}$ . Daí

$$p(\rho | \cdot) \propto \exp\left\{-\frac{1}{2} \left[ \frac{1}{V} \sum_{t=1}^T (Y_t - \alpha - \rho^t E_0 - \gamma \sum_{j=0}^t \rho^j X_{t-j})^2 \right]\right\} I_{[0,1]}. \quad (\text{B.4})$$

Dado que o espaço paramétrico de  $\rho$  é fechado e limitado ( $\rho \in [0, 1]$ ), para gerar valores de sua distribuição a posteriori, é possível utilizar o algoritmo *slice sampling* proposto por Neal (2003). Outra alternativa para amostrar de  $p(\rho|\cdot)$  é utilizar uma passo de Metropolis-Hastings com uma distribuição proposta Normal Truncada ou Beta. Porém, nos exercícios com dados artificiais, o *slice sampling* mostrou-se mais eficiente.

- Para  $V$ :

A distribuição condicional completa para a variância de  $Y_t$ ,  $V$ , está dada por:

$$p(V|\cdot) \propto \prod_{t=1}^T [p(Y_t|E_0, \alpha, \rho, \gamma, V)] p(V) \propto (V)^{-T/2-a_4-1} \exp\left\{-\frac{1}{V} \left[ \frac{1}{2} \sum_{t=1}^T (Y_t - \alpha - E_t)^2 + b_4 \right]\right\}.$$

$$\therefore V|\cdot \sim IG\left(\frac{T}{2} + a_4, \frac{1}{2} \sum_{t=1}^T (Y_t - \alpha - E_t)^2 + b_4\right). \quad (\text{B.5})$$

#### Algoritmo para o modelo em (B.1)

- (1). Dar valores iniciais para  $E_0, \alpha, \rho, \gamma$  e  $V$
- (2). Amostrar na seguinte forma:
  - (a)  $E_0, \alpha, \gamma$  diretamente de (B.3)
  - (b)  $\rho$  de (B.4) usando *slice sampling*
  - (c)  $V$  diretamente de (B.5)
- (3). Voltar ao passo 2 e iterar  $N$  vezes.

## B.2 Modelo Linear Normal de Função de Transferência Estocástica

Considere o seguinte modelo:

$$Y_t = \alpha + E_t + v_t, \quad v_t \sim N(0, V) \quad (\text{B.6a})$$

$$E_t = \rho E_{t-1} + \gamma X_t + w_t, \quad w_t \sim N(0, W) \quad (\text{B.6b})$$

onde  $v_t$  é o erro de observação associado a  $Y_t$  e  $w_t$  é o erro (variação estocástica) associado a  $E_t$ . Suponha que, a priori, os parâmetros são independentes e tais que:  $E_0 \sim N(a_0, b_0)$ ,  $\alpha \sim N(a_1, b_1)$ ,  $\rho \sim U[0, 1]$ ,  $\gamma \sim N(a_3, b_3)$ ,  $V \sim IG(a_4, b_4)$  e  $W \sim IG(a_5, b_5)$ . A distribuição condicional completa conjunta é:

$$p(E, \alpha, \rho, \gamma, V, W | \mathbf{Y}) \propto \prod_{t=1}^T [p(Y_t | \alpha, E_t, V) p(E_t | E_{t-1}, \rho, \gamma, W)] \times p(E_0) p(\alpha) p(\rho) p(\gamma) p(V) p(W); \quad (\text{B.7})$$

logo, as distribuições condicionais completas são:

- Para  $E_0$ :

$$p(E_0 | \cdot) \propto p(E_1 | E_0, \rho, \gamma, W) p(E_0) = N(\rho E_0 + \gamma X_1, W) N(a_0, b_0).$$

$$\therefore E_0 | \cdot \sim N(\mu_{E_0}, \sigma_{E_0}^2) \quad (\text{B.8})$$

$$\text{onde: } \mu_{E_0} = \sigma_{E_0}^2 \left[ \frac{\rho}{W} (E_1 - \gamma X_1) + \frac{a_0}{b_0} \right] \text{ e } \sigma_{E_0}^2 = \left( \frac{\rho^2}{W} + \frac{1}{b_0} \right)^{-1}.$$

- Para  $E_1, \dots, E_T$ :

O modelo (B.6) pode ser re-escrito na forma do seguinte modelo espaço-estado:

$$Y_t - \alpha = E_t + v_t \quad \Rightarrow \quad Y_t^* = 1E_t + v_t, \quad v_t \sim N(0, V_t) \quad (\text{B.9a})$$

$$E_t = \rho E_{t-1} + w_t^* \quad \Rightarrow \quad E_t = \rho E_{t-1} + w_t^*, \quad w_t^* \sim N(\gamma X_t, W_t), \quad (\text{B.9b})$$

onde  $E_1, \dots, E_T$  correspondem aos estados do modelo, conseqüentemente podem ser gerados utilizando o algoritmo *FFBS* (Frühwirth-Schnater, 1994).

Os momentos das distribuições a serem utilizadas neste particular *FFBS* são obtidos a seguir:

- Temos a seguinte informação:

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

$$E_t = \rho E_{t-1} + w_t^*, \quad w_t^* \sim N(\gamma X_t, W_t).$$

– Evolução:

$$\begin{aligned}
 E(E_t|D_{t-1}) &= E[E(E_t|E_{t-1}, D_{t-1})] \\
 &= E(\rho E_{t-1} + \gamma X_t) \\
 &= \rho m_{t-1} + \gamma X_t. \\
 \text{Var}(E_t|D_{t-1}) &= \text{Var}[E(E_t|E_{t-1}, D_{t-1})] + E[\text{Var}(E_t|E_{t-1}, D_{t-1})] \\
 &= \text{Var}(\rho E_{t-1} + \gamma X_t) + W \\
 &= \rho^2 C_{t-1} + W. \\
 \therefore E_t|D_{t-1} &\sim N(\rho m_{t-1} + \gamma X_t, \rho^2 C_{t-1} + W)
 \end{aligned}$$

– Atualização:

$$\begin{aligned}
 p(E_t|D_t) &\propto p(Y_t|E_t)p(E_t|D_{t-1}) \\
 &\propto N(E_t, V)N(\rho m_{t-1} + \gamma X_t, \rho^2 C_{t-1} + W) \\
 \therefore E_t|D_t &\sim N(m_t, C_t) \quad \text{onde:} \\
 m_t &= C_t[V^{-1}Y_t + (\rho^2 C_{t-1} + W)^{-1}(\rho m_{t-1} + \gamma X_t)] \\
 C_t &= (V^{-1} + (\rho^2 C_{t-1} + W)^{-1})^{-1}
 \end{aligned}$$

– Suavização:

$$\begin{aligned}
 E_{t+1} &= \rho E_t + w_{t+1}^*, \quad w_{t+1}^* \sim N(\gamma X_{t+1}, W) \\
 E_{t+1}|E_t &\sim N[\rho E_t + \gamma X_{t+1}, W] \\
 \begin{pmatrix} E_{t+1} \\ E_t \end{pmatrix} | D_t &\sim N\left[\begin{pmatrix} \rho m_t + \gamma X_{t+1} \\ m_t \end{pmatrix}, \begin{pmatrix} \rho^2 C_t + W & \rho C_t \\ \rho C_t & C_t \end{pmatrix}\right] \\
 \therefore E_t|E_{t+1}, D_t &\sim N(m_t^s, C_t^s), \quad \text{onde:} \\
 m_t^s &= m_t + \rho C_t(\rho^2 C_t + W)^{-1}(E_{t+1} - \rho m_t - \gamma X_{t+1}) \\
 C_t^s &= C_t - \rho^2 C_t^2(\rho^2 C_t + W)^{-1}
 \end{aligned}$$

• Para  $\alpha$ :

$$p(\alpha|\cdot) \propto \prod_{t=1}^T [p(Y_t|E_t, \alpha, V)] p(\alpha) = \exp\left\{-\frac{1}{2}\left[\frac{1}{V} \sum_{t=1}^T (Y_t - \alpha - E_t)^2 + \frac{1}{b_1}(\alpha - a_1)^2\right]\right\}$$

$$\therefore \alpha|\cdot \sim N(\mu_\alpha, \sigma_\alpha^2) \quad (\text{B.10})$$

$$\text{onde: } \mu_\alpha = \sigma_\alpha^2 \left[ \frac{1}{V} \sum_{t=1}^T (Y_t - E_t) + \frac{a_1}{b_1} \right] \text{ e } \sigma_\alpha^2 = \left( \frac{T}{V} + \frac{1}{b_1} \right)^{-1}$$

- Para  $\gamma$ :

$$p(\gamma|\cdot) \propto \prod_{t=1}^T [p(E_t|E_{t-1}, \rho, \gamma, W)] p(\gamma) = \exp\left\{-\frac{1}{2} \left[ \frac{1}{W} \sum_{t=1}^T (E_t - \rho E_{t-1} - \gamma X_t)^2 + \frac{1}{b_3} (\gamma - a_3)^2 \right]\right\}$$

$$\therefore \gamma|\cdot \sim N(\mu_\gamma, \sigma_\gamma^2) \quad (\text{B.11})$$

$$\text{onde: } \mu_\gamma = \sigma_\gamma^2 \left[ \frac{1}{W} \sum_{t=1}^T X_t (E_t - \rho E_{t-1}) + \frac{a_3}{b_3} \right] \text{ e } \sigma_\gamma^2 = \left( \frac{1}{W} \sum_{t=1}^T X_t^2 + \frac{1}{b_3} \right)^{-1}$$

- Para  $\rho$ :

$$p(\rho|\cdot) \propto \prod_{t=1}^T [p(E_t|E_{t-1}, \rho, \gamma, W)] p(\rho) = \exp\left\{-\frac{1}{2} \left[ \frac{1}{W} \sum_{t=1}^T (E_t - \rho E_{t-1} - \gamma X_t)^2 \right]\right\} I_{[0,1]}$$

$$\therefore \rho|\cdot \sim NT(\mu_\rho, \sigma_\rho^2) \quad (\text{B.12})$$

$$\text{onde: } \mu_\rho = \sigma_\rho^2 \left[ \frac{1}{W} \sum_{t=1}^T E_{t-1} (E_t - \gamma X_t) \right] \text{ e } \sigma_\rho^2 = \left( \frac{1}{W} \sum_{t=1}^T E_{t-1}^2 \right)^{-1}$$

- Para  $V$ :

$$p(V|\cdot) \propto \prod_{t=1}^T [p(Y_t|E_t, \alpha, V)] p(V) = (V)^{-T/2 - a_4 - 1} \exp\left\{-\frac{1}{V} \left[ \frac{1}{2} \sum_{t=1}^T (Y_t - \alpha - E_t)^2 + b_4 \right]\right\}$$

$$\therefore V|\cdot \sim IG\left(\frac{T}{2} + a_4, \frac{1}{2} \sum_{t=1}^T (Y_t - \alpha - E_t)^2 + b_4\right) \quad (\text{B.13})$$

- Para  $W$ :

$$p(W|\cdot) \propto \prod_{t=1}^T [p(E_t|E_{t-1}, \rho, \gamma, W)] p(W) = (W)^{-T/2 - a_5 - 1} \exp\left\{-\frac{1}{W} \left[ \frac{1}{2} \sum_{t=1}^T (E_t - \rho E_{t-1} - \gamma X_t)^2 + b_5 \right]\right\}$$

$$\therefore W|\cdot \sim IG\left(\frac{T}{2} + a_5, \frac{1}{2} \sum_{t=1}^T (E_t - \rho E_{t-1} - \gamma X_t)^2 + b_5\right) \quad (\text{B.14})$$

### Algoritmo para o modelo em (B.6)

- (1). Dar valores iniciais para  $E_0, \alpha, \rho, \gamma, V$  e  $W$
- (2). Amostrar na seguinte forma:
  - (a)  $E_0$  diretamente de (B.8)
  - (b)  $E_1, \dots, E_T$  usando o algoritmo FFBS
  - (c)  $\alpha$  diretamente de (B.10)
  - (d)  $\gamma$  diretamente de (B.11)
  - (e)  $\rho$  diretamente de (B.12) ou usando *slice sampling*
  - (f)  $V$  diretamente de (B.13)
  - (g)  $W$  diretamente de (B.14)
- (3). Voltar ao passo 2 e iterar  $N$  vezes.

### B.3 Modelo Linear Gama de Função de Transferência Não Estocástica

Considere o seguinte modelo:

$$Y_t \sim \text{Gama}(\mu_t, \nu) \quad t = 1, \dots, T \quad (\text{B.15a})$$

$$\log(\mu_t) = \alpha + E_t \quad (\text{B.15b})$$

$$E_t = \rho E_{t-1} + \gamma X_t. \quad (\text{B.15c})$$

A equação (B.15c) é a mesma que (B.1b), mas, neste caso,  $E_t$  afeta o preditor linear da verossimilhança gama. Supondo que, a priori, os parâmetros são independentes, podemos assumir que  $E_0 \sim N(a_0, b_0)$ ,  $\alpha \sim N(a_1, b_1)$ ,  $\rho \sim U[0, 1]$ ,  $\gamma \sim N(a_3, b_3)$  e  $\nu \sim IG(a_4, b_4)$ . Antes de descrever as condicionais completas e métodos de amostragem, revisamos algumas das propriedades da distribuição gama que serão exploradas pelos métodos de amostragem utilizados nesta tese.

### B.3.1 A Distribuição Gama

Se  $Y \sim G(\nu, \lambda)$  a sua função de densidade  $p(y)$  é dada por:

$$p(y|\nu, \lambda) = \frac{\lambda}{\Gamma(\nu)} (\lambda y)^{\nu-1} \exp(-\lambda y), \quad y \geq 0, \quad (\text{B.16})$$

onde  $\nu$  e  $\lambda$  são denominados parâmetros de forma e de escala, respectivamente. Algumas das propriedades desta distribuição são:

- A média e variância estão dadas por:

$$E(Y) = \mu = \frac{\nu}{\lambda} \quad \text{Var}(Y) = \frac{\nu}{\lambda^2} = \mu^2 \frac{1}{\nu}. \quad (\text{B.17})$$

- Parametrização na média: de (B.17):  $\lambda = \frac{\nu}{\mu}$  logo, (B.16) pode ser re-escrita como

$$\begin{aligned} p(y|\mu, \nu) &= \frac{1}{\Gamma(\nu)} \left(\frac{\nu}{\mu}\right) \left(\frac{\nu}{\mu} y\right)^{\nu-1} \exp\left(-\frac{\nu}{\mu} y\right), \\ \therefore p(y|\mu, \nu) &= \frac{1}{\Gamma(\nu)} \left(\frac{\nu}{\mu}\right)^{\nu-1} \exp\left(-\frac{\nu}{\mu} y\right) \underbrace{\frac{1}{y} dy}_{d(\log y)}, \end{aligned} \quad (\text{B.18})$$

(B.18) é denotada por  $Y \sim \text{Gama}(\mu, \nu)$ .

- A forma da densidade (B.16) é determinada por  $\nu$ . Se  $0 < \nu < 1$ , a densidade tem um decaimento exponencial. Se  $\nu > 1$  a densidade é côncava. Se  $\nu = 1$  temos uma distribuição exponencial. Se  $\nu \rightarrow \infty$  temos uma distribuição normal.
- A distribuição gama pertence à família exponencial: se  $Y \sim \text{Gama}(\mu, \nu)$  com  $\nu$  conhecido, o logaritmo da verossimilhança,  $l(\mu, \nu|y)$ , é dado por:

$$l(\mu, \nu|y) = (\nu - 1) \log y - \frac{\nu}{\mu} y + \nu \log \nu - \nu \log \mu - \log \Gamma(\nu) = \nu \left( -\frac{y}{\mu} - \log \mu \right) + c(y, \nu).$$

Logo,

$$\begin{aligned} \theta &= -1/\mu \\ b(\theta) &= \log \mu = -\log(-\theta) \\ b'(\theta) &= -1/\theta = \mu && : \text{esperança de } Y \\ b''(\theta) &= 1/\theta^2 = \mu^2 && : \text{função variância} \\ a(\phi) &= \phi = 1/\nu && : \text{parâmetro de dispersão} \end{aligned}$$

- No contexto de modelos lineares generalizados:

Seja  $\mathbf{Y}$  uma amostra de  $n$  observações  $y_i$  de uma distribuição gama com função de densidade (B.18). Seja  $\mathbf{X}$  um conjunto de  $p$  variáveis independentes para cada observação,  $x_1, \dots, x_p$ . Seja  $\eta = g(\mu) = \mathbf{X}'\beta$ , o modelo linear de interesse,  $\eta$  é denominado preditor linear. Então, as possíveis função de ligação são:

$$\begin{aligned} \text{Logística} & : \eta = \log(\mu) \\ \text{Recíproca} & : \eta = 1/\mu \\ \text{Identidade} & : \eta = \mu \\ \text{Potência} & : \eta = \mu^a, \quad a \text{ constante.} \end{aligned}$$

### B.3.2 Condicionais Completas

Nesta seção são apresentadas as distribuições condicionais completas dos parâmetros em (B.15). Também, formas eficientes de amostra-las são descritas.

- Para  $\alpha, E_0, \gamma$

Sejam  $\boldsymbol{\theta} = (\alpha, E_0, \gamma)$  e  $\mathbf{F}$  uma matriz  $T \times 3$  tal que:

$$\begin{array}{ll} \text{Para } t = 1 : & \text{Para } t = 2, \dots, T : \\ F[1, 1] = 1, & F[1, t] = 1 \\ F[2, 1] = \rho, & F[2, t] = \rho \times F[2, t - 1] \\ F[3, 1] = X_1 & F[3, t] = \rho \times F[3, t - 1] + X_t. \end{array}$$

Então, o modelo em (B.15b) pode ser re-escrito como:

$$\log(\mathbf{u}) = \mathbf{F}'\boldsymbol{\theta} \quad (\text{B.19})$$

Seja  $\boldsymbol{\theta} \sim NM_3(\mathbf{a}, \mathbf{R})$  onde  $\mathbf{a} = (a_1, a_0, a_3)'$  e  $\mathbf{R} = \text{diag}(b_1, b_0, b_3)$  é a distribuição a priori para  $\boldsymbol{\theta}$ .

Então

$$p(\boldsymbol{\theta}|\mathbf{Y}) \propto \left\{ -\frac{1}{2}(\boldsymbol{\theta} - \mathbf{a})'\mathbf{R}^{-1}(\boldsymbol{\theta} - \mathbf{a}) + \sum_{i=1}^T \left[ v \left( -\frac{y_t}{\mu_t} - \log \mu_t \right) + c(y_t, v) \right] \right\} \quad (\text{B.20})$$

A distribuição em (B.20) não tem forma fechada conhecida, mas tem a forma (aspecto) da distribuição a posteriori dos parâmetros de efeitos fixos de um modelo linear generalizado, daí que para gerar valores dessa distribuição é possível utilizar o técnica de amostragem multivariada proposta em Gamerman (1997, pag.60). Em síntese, é um passo de Metropolis-Hastings onde a distribuição proposta é formada fazendo uma iteração, começando no estado corrente da cadeia, do método denominado Mínimos Quadrados Ponderados Iterativos ou *IWLS* (abreviação do inglês *Iterative Weighted Least Squares*).

Em particular, para gerar valores de (B.20), os passos a seguir são:

- (1). Calcular as pseudo-observações,  $\tilde{y}_t(\boldsymbol{\theta})$ , e os pesos,  $W_t$  :

$$\begin{aligned}\tilde{y}_t(\boldsymbol{\theta}) &= \eta_t + (y_t - \mu_t)g'(\mu_t) = \log(\mu_t) + (y_t - \mu_t)(1/\mu_t) \\ W_t^{-1}(\boldsymbol{\theta}) &= b''(\boldsymbol{\theta})[g'(\mu_t)]^2 = [1/\mu],\end{aligned}$$

onde  $\eta_t = \log \mu_t$  é a função de ligação logarítmica.

- (2). Amostrar  $\boldsymbol{\theta}^*$  da distribuição proposta  $N(\mathbf{m}^{(i)}, \mathbf{C}^{(j)})$ , onde:

$$\begin{aligned}\mathbf{m}^{(i)} &= \mathbf{C}^{(i)}(\mathbf{R}^{-1}\mathbf{a} + \mathbf{FW}(\boldsymbol{\theta}^{(i-1)})\tilde{\mathbf{y}}(\boldsymbol{\theta}^{(i-1)})) \\ \mathbf{C}^{(i)} &= (\mathbf{R}^{-1} + \mathbf{FW}(\boldsymbol{\theta}^{(i-1)})\mathbf{F}')^{-1},\end{aligned}\tag{B.21}$$

onde o índice ( $i$ ) denota a  $i$ -ésima iteração do *MCMC*.

- (3). Aceitar  $\boldsymbol{\theta}^*$  com probabilidade  $p_t = \min(1, A)$  onde  $A = \frac{\pi(\boldsymbol{\theta}^*)q(\boldsymbol{\theta}^*, \boldsymbol{\theta}^{(i-1)})}{\pi(\boldsymbol{\theta}^{(i-1)})q(\boldsymbol{\theta}^{(i-1)}, \boldsymbol{\theta}^*)}$ . Caso contrário fazer  $\boldsymbol{\theta}^{(i)} = \boldsymbol{\theta}^{(i-1)}$ . Em  $A$ ,  $\pi(\cdot)$  é a distribuição (B.20) e  $q(\cdot)$  é a distribuição Normal com momentos (B.21).

- Para  $\rho$ :

A distribuição condicional completa de  $\rho$  está dada por:

$$p(\rho|\cdot) \propto \prod_{t=1}^T [p(Y_t|\mu_t, \nu)]p(\rho) = \exp\left[\nu \sum_{t=1}^T (-Y_t/\mu_t - \log \mu_t)\right]I_{[0,1]},\tag{B.22}$$

onde  $\log \mu_t = \alpha + E_t$  e  $E_t = \rho^t E_0 + \gamma \sum_{j=0}^t \rho^j X_{t-j}$ . Logo  $p(\rho|\cdot)$  não tem forma fechada conhecida. Dado que o espaço paramétrico de  $\rho$  é fechado e limitado ( $\rho \in [0, 1]$ ), para gerar valores de sua distribuição a posteriori, é possível utilizar o algoritmo *slice sampling* proposto por Neal (2003).

- Para  $\nu$ :

Tem-se que  $\nu$  é o parâmetro de forma da distribuição gama portanto é uma valor positivo ( $\nu > 0$ ). Se a distribuição a priori para  $\nu$  é  $IG(a_4, b_4)$ , a sua distribuição a posteriori está dada por

$$p(\nu|\cdot) \propto \prod_{t=1}^T [p(Y_t|\mu_t, \nu)] p(\nu) = \prod_{t=1}^T \left[ \frac{1}{\Gamma(\nu)} \left( \frac{\nu}{\mu_t} y \right)^{\nu-1} \exp\left(-\frac{\nu}{\mu_t} y_t\right) \right] \nu^{-a_4-1} \exp(-b_4/\nu). \quad (\text{B.23})$$

Para amostrar de  $p(\nu|\cdot)$  pode ser utilizado um passo de Metropolis-Hastings com uma distribuição normal com média igual o valor corrente da cadeia e variância conhecida, como densidade proposta. Porém, dado que  $\nu$  é estritamente positivo, faz-se necessário usar o log do valor gerado como valor candidato e deve-se levar em conta o Jacobino da transformação ao calcular a probabilidade de aceitação.

#### Algoritmo para o modelo em (B.15)

- (1). Dar valores iniciais para  $E_0, \alpha, \rho, \gamma$  e  $\nu$ .
- (2). Amostrar na seguinte forma:
  - (a) Amostrar  $\alpha, E_0, \gamma$  em bloco, usando o passo de Metropolis-Hastings usando como proposta uma distribuição Normal com momentos (B.21);
  - (b) Amostrar  $\rho$  de (B.22) usando usando *slice sampling*;
  - (c) Amostrar  $\nu$  de (B.23) usando um passo de Metropolis-Hastings com proposta Normal e transformação log (ou proposta Log-Normal).
- (3). Voltar ao passo (2) e iterar  $N$  vezes.

#### B.4 Modelo Linear Gama de Função de Transferência Estocástica

Considere o seguinte modelo:

$$Y_t \sim \text{Gama}(\mu_t, \nu) \quad (\text{B.24a})$$

$$\log(\mu_t) = \alpha + E_t \quad (\text{B.24b})$$

$$E_t = \rho E_{t-1} + \gamma X_t + w_t, \quad w_t \sim N(0, W) \quad (\text{B.24c})$$

Supondo que, a priori, os parâmetros são independentes, podemos atribuir as seguintes distribuições:  $E_0 \sim N(a_0, b_0)$ ,  $\alpha \sim N(a_1, b_1)$ ,  $\rho \sim U[0, 1]$ ,  $\gamma \sim N(a_3, b_3)$ ,  $\nu \sim IG(a_4, b_4)$  e  $W \sim IG(a_5, b_5)$ . Conseqüentemente, as distribuições condicionais completas para os parâmetros do modelo em (B.24) são:

- Para  $\alpha$ :

$$p(\alpha|\cdot) \propto \prod_{t=1}^T [p(Y_t|E_t, \alpha, V)] = \prod_{t=1}^T \left[ \frac{1}{\Gamma(\nu)} \left( \frac{\nu}{\mu_t} y \right)^{\nu-1} \exp\left(-\frac{\nu}{\mu_t} y_t\right) \right] \exp\left(\frac{1}{b_3} (\gamma - a_3)^2\right) \quad (\text{B.25})$$

- Para  $\gamma$ :

$$p(\gamma|\cdot) \propto \prod_{t=1}^T [p(E_t|E_{t-1}, \rho, \gamma, W)] p(\gamma) = \exp\left\{-\frac{1}{2} \left[ \frac{1}{W} \sum_{t=1}^T (E_t - \rho E_{t-1} - \gamma X_t)^2 + \frac{1}{b_3} (\gamma - a_3)^2 \right]\right\}$$

$$\therefore \gamma|\cdot \sim N(\mu_\gamma, \sigma_\gamma^2) \quad (\text{B.26})$$

onde:  $\mu_\gamma = \sigma_\gamma^2 \left[ \frac{1}{W} \sum_{t=1}^T X_t (E_t - \rho E_{t-1}) + \frac{a_3}{b_3} \right]$  e  $\sigma_\gamma^2 = \left( \frac{1}{W} \sum_{t=1}^T X_t^2 + \frac{1}{b_3} \right)^{-1}$

- Para  $\rho$ :

$$p(\rho|\cdot) \propto \prod_{t=1}^T [p(E_t|E_{t-1}, \rho, \gamma, W)] p(\rho) = \exp\left\{-\frac{1}{2} \left[ \frac{1}{W} \sum_{t=1}^T (E_t - \rho E_{t-1} - \gamma X_t)^2 \right]\right\} I_{[0,1]} \quad (\text{B.27})$$

- Para  $\nu$ :

$$p(\nu|\cdot) \propto \prod_{t=1}^T [p(Y_t|\mu_t, \nu)] p(\nu) = \prod_{t=1}^T \left[ \frac{1}{\Gamma(\nu)} \left( \frac{\nu}{\mu_t} y \right)^{\nu-1} \exp\left(-\frac{\nu}{\mu_t} y_t\right) \right] \nu^{-a_4-1} \exp(-b_4/\nu) \quad (\text{B.28})$$

- Para  $W$ :

$$p(W|\cdot) \propto \prod_{t=1}^T [p(E_t|E_{t-1}, \rho, \gamma, W)] p(W) = W^{-T/2-a_5-1} \exp\left\{-\frac{1}{2W} \left(\sum_{i=1}^T (E_t - \rho E_{t-1} - \gamma X_t)^2 - b_5\right)\right\}$$

$$\therefore W|\cdot \sim IG(T/2 + a_5, \sum_{i=1}^T ((E_t - \rho E_{t-1} - \gamma X_t)^2) - b_5) \quad (\text{B.29})$$

- Para  $E_0$ :

$$p(E_0|\cdot) \propto p(E_1|E_0, \rho, \gamma, W) p(E_0) = N(\rho E_0 + \gamma X_1, W) N(a_0, b_0)$$

$$\therefore E_0|\cdot \sim N(\mu_{E_0}, \sigma_{E_0}^2) \quad (\text{B.30})$$

onde:  $\mu_{E_0} = \sigma_{E_0}^2 \left[ \frac{\rho}{W} (E_1 - \gamma X_1) + \frac{a_0}{b_0} \right]$  e  $\sigma_{E_0}^2 = \left( \frac{\rho^2}{W} + \frac{1}{b_0} \right)^{-1}$

- Para  $E_1, \dots, E_T$ :

$$p(E_t|\cdot) \propto p(Y_t|\alpha, E_t, \nu) p(E_t|E_{t-1}, \rho, \gamma, W) p(E_{t+1}|E_t, \rho, \gamma, W) \quad t = 1, \dots, T-1.$$

$$p(E_t|\cdot) \propto p(Y_t|\alpha, E_t, \nu) p(E_t|E_{t-1}, \rho, \gamma, W) \quad t = T.$$

$E = E_1, \dots, E_T$  corresponde ao vector de estados do modelo em (B.24).  $E$  pode ser amostrado em bloco usando *CUBS* ou em  $T$  passos individuais com o algoritmo descrito na seção seguinte.

#### B.4.1 Alternativa de Amostragem para $E_t$ , $t = 1, \dots, T$

No modelo (B.24)  $E_t$  pode ser re-escrito em função dos termos de erro  $w_j$  como em (B.31):

$$Y_t \sim \text{Gama}(\mu_t, \nu) \quad (\text{B.31a})$$

$$\log(\mu_t) = \alpha + E_t \quad (\text{B.31b})$$

$$E_t = \sum_{j=1}^t \rho^{t-j} (w_j + \gamma X_j), \quad w_j \sim N(0, W) \quad (\text{B.31c})$$

$$= \sum_{j=1}^t \rho^{t-j} w_j^*, \quad w_j^* \sim N(\gamma X_j, W) \quad (\text{B.31d})$$

A representação acima permite amostrar da distribuição a posteriori de  $E_t$  fazendo uso de um dos algoritmos propostos em (Gamerman (1998)). A idéia é construir a densidade proposta do passo de Metropolis-Hastings com base no modelo normal auxiliar:

$$\tilde{y}_t = \mathbf{F}'_t \boldsymbol{\theta}_t + v_t, \quad v_t \sim N(0, \tilde{V}_t), \quad (\text{B.32a})$$

$$E_t = \sum_{j=1}^t \rho^{t-j} w_j^*, \quad w_j^* \sim N(\gamma X_j, W), \quad (\text{B.32b})$$

onde:

$$\begin{aligned} \tilde{y}_t &= g(\mu_t) + g'(\mu_t)(y_t - \mu_t) = \log(\mu_t) + (1/\mu_t)(y_t - \mu_t) \\ \tilde{V} &= b''(\eta_t)[g'(\mu_t)]^2 = (1/\mu_t) \end{aligned}$$

A verossimilhança em (B.32) é dada por:

$$\begin{aligned} p(\tilde{y}_j | w_1^*, \dots, w_j^*, \tilde{V}) &\propto \exp\left\{-\frac{1}{2\tilde{V}}\left(\tilde{y}_j - \rho^{j-t} w_t^* - \sum_{k=0}^{t-1} \rho^{j-k} w_k^* - \sum_{k=t+1}^j \rho^{j-k} w_k^*\right)^2\right\} \\ &\propto \exp\left\{-\frac{1}{2\tilde{V}}\left(\tilde{y}_j - H_{tj} w_t^* - h_{tj}\right)^2\right\} \end{aligned} \quad (\text{B.33})$$

onde

$$H_{tj} = \rho^{j-t} \quad \text{e} \quad h_{tj} = \sum_{k=1}^{t-1} \rho^{j-k} w_k^* + \sum_{k=t+1}^j \rho^{j-k} w_k^*.$$

A distribuição condicional completa de  $w_t^*$ ,  $t = 1, \dots, N$  é

$$\pi(w_t^*) \propto \prod_{j=t}^N p(\tilde{y}_j | w_1^*, \dots, w_j^*, V) p(w_t^* | \gamma, \mathbf{X}, W) = N(b_t, B_t), \quad (\text{B.34})$$

onde

$$b_t = B_t \sum_{j=t}^N \tilde{V}^{-1} H_{tj} (\tilde{y}_j - h_{tj}) \quad \text{e} \quad B_t = \left(W^{-1} + \sum_{j=t}^N \tilde{V}^{-1} H_{tj}^2\right)^{-1}.$$

**Algoritmo para o modelo em (B.24)**

- (1). Dar valores iniciais para  $E_0, \alpha, \rho, \gamma, \nu$  e  $W$
- (2). Amostrar na seguinte forma:
  - (a)  $\alpha$  de (B.25) usando *slice sampling* ou um passo de Metropolis-Hastings
  - (b)  $E_0$  diretamente de (B.30)
  - (c)  $E_t$  usando *CUBS* ou usando o algoritmo descrito na seção B.4.1:
    - (i). Fazer  $t = 1$
    - (ii). Amostrar  $w_t^{new}$
    - (iii). Fazer  $w_t^* = w_t^{new}$  com probabilidade  $p_t = \min(1, A)$  onde  $A$  é
 
$$\frac{\pi(w_t^{new}) q_t(w_t^* | w_t^{new})}{\pi(w_t^*) q_t(w_t^{new} | w_t^*)}$$
    - (iv). Reconstruir  $w_t = w_t^* - \gamma X_j$  e  $E_t = \rho E_{t-1} + w_t$
    - (v). Fazer  $t = t + 1$  e volver a (2)(c)(ii) se  $t < T$
  - (d)  $\rho$  de (B.27) usando usando *slice sampling*
  - (e)  $\nu$  de (B.28) usando um passo de Metropolis-Hastings com distribuição proposta Normal e transformação log.
  - (f)  $W$  diretamente de (B.29)
- (3). Voltar ao passo (2) e iterar  $N$  vezes.

### B.5 Modelo de Função de Transferência Não Estocástica com Estrutura Hierárquica

Considere o seguinte modelo:

$$Y_{t,k} = \alpha_k + E_{t,k} + v_{t,k}, \quad v_{t,k} \sim N(0, V) \quad (\text{B.35a})$$

$$E_{t,k} = \rho E_{t-1,k} + \gamma_k X_{t,k} \quad (\text{B.35b})$$

$$\alpha_k = \alpha + w_\alpha, \quad w_\alpha \sim N(0, W_\alpha) \quad (\text{B.35c})$$

$$\gamma_k = \gamma + w_\gamma, \quad w_\gamma \sim N(0, W_\gamma) \quad (\text{B.35d})$$

onde  $t = 1, \dots, T$  e  $k = 1, \dots, K$ .  $T$  é o tamanho das séries temporais e  $K$  é o número de bacias em estudo. Suponha que, a priori, os parâmetros são independentes e tais que:  $E_{k,0} \sim N(a_0, b_0)$ ,  $\alpha \sim N(a_1, b_1)$ ,  $\rho \sim U[0, 1]$ ,  $\gamma \sim N(a_2, b_2)$ ,  $W_\alpha \sim IG(a_3, b_3)$ ,  $W_\gamma \sim IG(a_4, b_4)$  e  $V \sim IG(a_5, b_5)$ . O algoritmo apresentado na seção B.1, pode ser generalizado para o caso de  $K$  bacias, como o modelo em (B.35). A distribuição conjunta a posteriori das quantidades desconhecidas em (B.35) é:

$$p(\boldsymbol{\alpha}, \boldsymbol{\gamma}, \rho, \mathbf{E}_0, V, W_\alpha, W_\gamma | \mathbf{y}) \propto \prod_{k=1}^K \prod_{t=1}^T p(y_{t,k} | E_{0,k}, \alpha_k, \gamma_k, \rho, V) \\ \prod_{k=1}^K p(\alpha_k | \alpha, W_\alpha) \prod_{k=1}^K p(\gamma_k | \gamma, W_\gamma) \prod_{k=1}^K p(E_{0,k}) p(\alpha) p(\gamma) p(\rho) p(W_\alpha) p(W_\gamma) p(V)$$

Daí, as distribuições condicionais completas são:

- Para  $E_0, \boldsymbol{\alpha}, \boldsymbol{\gamma}$ :

As condicionais completas de  $E_{0,k}, \alpha_k, \gamma_k$ ,  $k = 1, \dots, K$  são distribuições normais, portanto, amostras desses parâmetros podem ser geradas em bloco, em forma análoga ao caso de uma única bacia (seção B.1).

Seja  $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_K)'$  onde  $\boldsymbol{\theta}_k = (\alpha_k, E_{0,k}, \gamma_k)$  e seja  $F$  uma matriz bloco-diagonal  $3K \times 3K$  tal que:

$$F = \begin{pmatrix} F_1 & 0 & 0 & \dots & 0 \\ 0 & F_2 & 0 & \dots & 0 \\ 0 & 0 & F_3 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & F_K \end{pmatrix}$$

onde  $F_1, \dots, F_K$  são matrizes  $3 \times T$  definidas como na seção B.1, isto é, com elementos dados por:

$$\begin{array}{ll} \text{Para } t = 1 : & \text{Para } t = 2, \dots, T : \\ F_k[1, 1] = 1, & F_k[1, t] = 1 \\ F_k[2, 1] = \rho, & F_k[2, t] = \rho \times F_k[2, t-1] \\ F_k[3, 1] = X_{k,1} & F_k[3, t] = \rho \times F_k[3, t-1] + X_{k,t}. \end{array}$$

Então, as equações (B.35a) e (B.35b) podem ser escritas de forma matricial como:

$$\mathbf{Y} = \mathbf{F}'\boldsymbol{\theta} + \boldsymbol{\nu}, \quad \boldsymbol{\nu} \sim NM_{TK}(\mathbf{0}, \mathbf{V}) \quad (\text{B.36})$$

onde  $NM_{TK}$  denota uma distribuição Normal Multivariada de dimensão  $TK$ .

Usando as equações (B.35c) e (B.35d) e a distribuição a priori para  $E_0$  define-se  $\boldsymbol{\theta}_k \sim NM_3(a_k, R_k)$  onde  $a_k = (\alpha, a_0, \gamma)'$  e  $R_k = \text{diag}(W_\alpha, b_0, W_\gamma)$  como distribuição a priori<sup>1</sup> para  $\boldsymbol{\theta}_k = (\alpha_k, E_{0,k}, \gamma_k)$ . Logo, a distribuição a priori para  $\boldsymbol{\theta}$  é  $NM_{3K}(\mathbf{a}, \mathbf{R})$  onde:

$$\mathbf{a} = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_K \end{pmatrix} \quad \text{e} \quad \mathbf{R} = \begin{pmatrix} R_1 & 0 & \dots & 0 \\ 0 & R_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & R_K \end{pmatrix}$$

Finalmente, a distribuição a posteriori condicional completa de  $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_K)'$  está dada por:

$$\boldsymbol{\theta} | \mathbf{Y} \sim NM_{3K}(\boldsymbol{\mu}_\boldsymbol{\theta}, \boldsymbol{\Sigma}_\boldsymbol{\theta}) \quad (\text{B.37})$$

onde:  $\boldsymbol{\mu}_\boldsymbol{\theta} = \boldsymbol{\Sigma}_\boldsymbol{\theta}(\mathbf{FV}^{-1}\mathbf{Y} + \mathbf{R}^{-1}\mathbf{a})$  e  $\boldsymbol{\Sigma}_\boldsymbol{\theta} = (\mathbf{R}^{-1} + \mathbf{FV}^{-1}\mathbf{F}')^{-1}$

- Para  $\rho$ :

$$p(\rho | \cdot) \propto \prod_{k=1}^K \prod_{t=1}^T [p(Y_{t,k} | E_{0,k}, \alpha, \rho, \gamma, V)] p(\rho) = \exp\left\{-\frac{1}{2} \left[ \frac{1}{V} \sum_{k=1}^K \sum_{t=1}^T (Y_{t,k} - \alpha_k - E_{t,k})^2 \right]\right\} I_{[0,1]}$$

onde  $E_{t,k} = \rho^t E_{0,k} + \gamma_k \sum_{j=0}^{t-1} \rho^j X_{t-j,k}$ . Dai

$$p(\rho | \cdot) \propto \exp\left\{-\frac{1}{2} \left[ \frac{1}{V} \sum_{k=1}^K \sum_{t=1}^T (Y_{t,k} - \alpha_k - \rho^t E_{0,k} - \gamma_k \sum_{j=0}^{t-1} \rho^j X_{t-j,k})^2 \right]\right\} I_{[0,1]} \quad (\text{B.38})$$

---

<sup>1</sup>poder ser considerada como priori ao nível de bacia

- Para  $V$ :

$$p(V|\cdot) \propto \prod_{k=1}^K \prod_{t=1}^T [p(Y_{t,k}|E_0, \alpha, \rho, \gamma, V)] p(V) = V^{-TK/2 - a_5 - 1} \exp\left\{-\frac{1}{V} \left[\frac{1}{2} \sum_{k=1}^K \sum_{t=1}^T (Y_{t,k} - \alpha_k - E_{t,k})^2 + b_5\right]\right\}$$

$$\therefore V|\cdot \sim IG\left(\frac{TK}{2} + a_5, \frac{1}{2} \sum_{k=1}^K \sum_{t=1}^T (Y_{t,k} - \alpha_k - E_{t,k})^2 + b_5\right) \quad (\text{B.39})$$

- Para  $\alpha$ :

$$p(\alpha|\cdot) \propto \prod_{k=1}^K p(\alpha_k|\alpha, W_\alpha) \times p(\alpha) = \prod_{k=1}^K N(\alpha_k, W_\alpha) \times N(a_1, b_1)$$

$$\therefore \alpha|\cdot \sim N(\mu_\alpha, \sigma_\alpha^2) \quad (\text{B.40})$$

onde  $\mu_\alpha = \frac{K\bar{\alpha}/W_1 + a_1/b_1}{K/W_1 + 1/b_1}$  e  $\sigma_\alpha^2 = \frac{1}{K/W_1 + 1/b_1}$

- Para  $\gamma$ :

$$p(\gamma|\cdot) \propto \prod_{k=1}^K p(\gamma_k|\gamma, W_2) \times p(\gamma) = \prod_{k=1}^K N(\gamma, W_2) \times N(a_2, b_2)$$

$$\therefore \gamma|\cdot \sim N(\mu_\gamma, \sigma_\gamma^2) \quad (\text{B.41})$$

onde  $\mu_\gamma = \frac{K\bar{\gamma}/W_2 + a_2/b_2}{K/W_2 + 1/b_2}$  e  $\sigma_\gamma^2 = \frac{1}{K/W_2 + 1/b_2}$

- Para  $W_\alpha$ :

$$p(W_1|\cdot) \propto \prod_{k=1}^K [p(\alpha_k|\alpha, W_\alpha)] p(W_\alpha) = W_\alpha^{-K/2 - a_3 - 1} \exp\left\{-\frac{1}{W_\alpha} \left[\frac{1}{2} \sum_{k=1}^K (\alpha_k - \alpha)^2 + b_3\right]\right\}$$

$$\therefore W_\alpha|\cdot \sim IG\left(\frac{K}{2} + a_3, \frac{1}{2} \sum_{k=1}^K (\alpha_k - \alpha)^2 + b_3\right) \quad (\text{B.42})$$

- Para  $W_\gamma$ :

$$p(W_2|\cdot) \propto \prod_{k=1}^K [p(\gamma_k|\gamma, W_\gamma)] p(W_\gamma) = W_\gamma^{-K/2 - a_4 - 1} \exp\left\{-\frac{1}{W_\gamma} \left[\frac{1}{2} \sum_{k=1}^K (\gamma_k - \gamma)^2 + b_4\right]\right\}$$

$$\therefore W_\gamma|\cdot \sim IG\left(\frac{K}{2} + a_4, \frac{1}{2} \sum_{k=1}^K (\gamma_k - \gamma)^2 + b_4\right) \quad (\text{B.43})$$

**Algoritmo para o modelo em (B.35)**

- (1). Dar valores iniciais para  $E_0, \alpha, \rho, \gamma, W_\alpha, W_\gamma$  e  $V$
- (2). Amostrar na seguinte forma:
  - (a)  $(E_0, \alpha, \gamma) = (E_{1,0}, \dots, E_{k,t}, \alpha_1, \dots, \alpha_k, \gamma_1, \dots, \gamma_k)$ , em bloco, diretamente de (B.37)
  - (b)  $\rho$  de (B.38) usando *slice sampling* ou um passo de Metropolis-Hastings
  - (c)  $V$  diretamente de (B.39)
  - (d)  $\alpha$  diretamente de (B.40)
  - (e)  $\gamma$  diretamente de (B.41)
  - (f)  $W_\alpha$  diretamente de (B.42)
  - (g)  $W_\gamma$  diretamente de (B.43)
- (3). Voltar ao passo 2 e iterar  $N$  vezes.

## Apêndice C

### CUBS NA FAMÍLIA EXPONENCIAL $K$ -PARAMÉTRICA

#### O Caso da Distribuição Normal

Considere que  $Y_t \sim \text{Normal}(\mu_t, \lambda_t^{-1})$ , onde  $\lambda_t$  denota precisão, o objetivo é ajustar

$$Y_t | \boldsymbol{\psi}_t \sim \text{Normal}(\mu_t, \lambda_t^{-1}), \quad \text{onde } \boldsymbol{\psi}_t = (\mu_t, \lambda_t)' \quad (\text{C.1a})$$

$$g_1(\boldsymbol{\psi}_t) = \mu_t = F'_1 \boldsymbol{\theta}_{1t} = \theta_{11} + \theta_{12} x_{1t} \quad (\text{C.1b})$$

$$g_2(\boldsymbol{\psi}_t) = \log(\lambda_t) = F'_2 \boldsymbol{\theta}_{2t} = \theta_{21} + \theta_{22} x_{2t} \quad (\text{C.1c})$$

$$\boldsymbol{\theta}_{1t} = G_1 \boldsymbol{\theta}_{1,t-1} + w_{1t}, \quad w_{1t} \sim [0, W_{1t}] \quad (\text{C.1d})$$

$$\boldsymbol{\theta}_{2t} = G_2 \boldsymbol{\theta}_{2,t-1} + w_{2t}, \quad w_{2t} \sim [0, W_{2t}], \quad (\text{C.1e})$$

onde  $\mu_t$  e  $\log(\lambda_t)$  estão relacionados com  $\boldsymbol{\theta}_t$  através de dois modelos lineares e  $\boldsymbol{\theta}_t$  evolui suavemente no tempo.

Como a verossimilhança do modelo acima tem a forma:

$$p(y|\mu, \lambda) = \left(\frac{\lambda}{2\pi}\right)^{1/2} \exp\left\{-\frac{\lambda}{2}(y - \mu)^2\right\}, \quad (\text{C.2})$$

a distribuição a priori conjugada bivariada, necessária para amostrar da condicional completa de  $\boldsymbol{\theta}_t$  usando o ECUBS, está dada por:

$$p(\mu, \lambda | \alpha, \beta, \gamma, \tau) = \frac{\lambda^{\alpha-1} \exp(-\frac{\lambda}{\beta})}{\Gamma(\alpha)\beta^\alpha} \times \left(\frac{\lambda\tau}{2\pi}\right)^{1/2} \exp\left\{-\frac{\lambda\tau}{2}(\mu - \gamma)^2\right\}. \quad (\text{C.3})$$

De (C.3), temos que a média e variância marginais a priori para  $\mu$  and  $\lambda$  são

$$\begin{aligned} E[\lambda] &= \alpha\beta & E[\mu] &= \gamma \\ \text{Var}[\lambda] &= \alpha\beta^2 & \text{Var}[\mu] &= \frac{1}{(\alpha-1)\tau\beta} \end{aligned}$$

e a distribuição a posteriori conjunta está especificada por

$$\begin{aligned} \alpha^* &= \alpha + 1/2 & \beta^* &= \left[\frac{1}{\beta} + \frac{\tau(y - \gamma)^2}{2(\tau + 1)}\right]^{-1} \\ \gamma^* &= \frac{\tau\gamma + y}{\tau + 1} & \tau^* &= \tau + 1. \end{aligned}$$

Conseqüentemente, temos que:

$$\begin{aligned}
E[g_1(\boldsymbol{\psi}_t)] &= E[\mu_t] &&= \gamma_t \\
Var[g_1(\boldsymbol{\psi}_t)] &= Var[\mu_t] &&= [(\alpha_t - 1)\tau_t\beta_t]^{-1} \\
E[g_2(\boldsymbol{\psi}_t)] &\approx \log(E[\lambda_t]) &&= \log(\alpha_t\beta_t) \\
Var[g_2(\boldsymbol{\psi}_t)] &\approx Var[\lambda_t] \left[ \frac{d}{d\lambda_t} \log \lambda_t \right]_{\lambda_t=E[\lambda_t]}^2 &&= Var[\lambda_t][E^2[\lambda_t]]^{-1} = \frac{1}{\alpha_t}.
\end{aligned} \tag{C.4}$$

Com estes resultados podemos escrever em detalhe primeira etapa do *ECUBS* assim:

(1). Distribuição a priori, dada pelo modelo linear, para  $\theta_1$ ,  $\theta_2$ ,  $g_1(\boldsymbol{\psi}_t)$  e  $g_2(\boldsymbol{\psi}_t)$ :

$$\begin{aligned}
(\theta_{jt}|D_{t-1}) &\sim [a_{jt}, R_{jt}], \quad j = 1, 2 \\
\text{onde: } a_{jt} &= G_{jt}m_{j,t-1}, \quad R_{jt} = G_{jt}C_{j,t-1}G'_{jt} + W_{jt} \\
(\mu_t|D_{t-1}) &\sim [f_{1t}, q_{1t}] \quad \text{and} \quad (\log(\lambda_t)|D_{t-1}) \sim [f_{2t}, q_{2t}] \\
\text{onde: } f_{jt} &= F'_{jt}a_{jt}, \quad q_{jt} = F'_{jt}R_{jt}F_{jt}, \quad j = 1, 2.
\end{aligned}$$

(2). Distribuição a priori conjugada para  $g_1(\boldsymbol{\psi}_t)$  e  $g_2(\boldsymbol{\psi}_t)$ :

$$\begin{aligned}
E[\mu_t|D_{t-1}] &= f_{1t} = \gamma_t \\
Var[\mu_t|D_{t-1}] &= q_{1t} = [\tau_t\beta_t(\alpha_t - 1)]^{-1} \\
E[\log(\lambda_t)|D_{t-1}] &= f_{2t} = \log(\alpha_t\beta_t) \\
Var[\log(\lambda_t)|D_{t-1}] &= q_{2t} = \frac{1}{\alpha_t}.
\end{aligned}$$

Logo, os parâmetros da priori conjugada são

$$\begin{aligned}
\alpha_t &= (q_{2t})^{-1} \\
\beta_t &= (q_{2t} \exp f_{2t})^{-1} \\
\gamma_t &= f_{1t} \\
\tau_t &= [q_{1t} \exp(f_{2t})(1 - q_{2t})]^{-1}.
\end{aligned}$$

Dai, os parâmetros da posteriori conjugada são:

$$\begin{aligned}
\alpha_t^* &= \alpha_t + 1/2 \\
\beta_t^* &= \frac{1}{\beta_t} + \frac{\tau_t(y_t - \gamma_t)^2}{2(\tau_t + 1)} \\
\gamma_t^* &= \frac{\tau_t\gamma_t + y_t}{\tau_t + 1} \\
\tau_t^* &= \tau_t + 1.
\end{aligned}$$

(3). Densidade a posteriori para  $g_1(\alpha_t)$  e  $g_2(\beta_t)$ , dada pelo modelo linear:

$$\begin{aligned} E[\mu_t|D_t] &= f_{1t}^* = \gamma_t^* \\ \text{Var}[\mu_t|D_t] &= q_{1t}^* = \frac{1}{\tau_t^* \beta_t^* (\alpha_t^* - 1)} \\ E[\log(\lambda_t)|D_t] &= f_{2t}^* = \log(\alpha_t^* \beta_t^*) \\ \text{Var}[\log(\lambda_t)|D_t] &= q_{2t}^* = \frac{1}{\alpha_t^*}. \end{aligned}$$

(4). Momentos da distribuição a posteriori de  $\theta_{1t}$  e  $\theta_{2t}$  (usando estimação *Linear Bayes*):

$$\begin{aligned} \theta_{jt} &\sim [m_{jt}, C_{jt}], \quad \text{onde:} \\ m_{jt} &= a_{jt} + R_{jt} F_{jt} (f_{jt}^* - f_{jt}) \frac{1}{q_{jt}} \\ C_{jt} &= R_{jt} - R_{jt} F_{jt} F_{jt}' R_{jt} \left(1 - \frac{q_{jt}^*}{q_{jt}}\right) \frac{1}{q_{jt}}, \quad j = 1, 2. \end{aligned}$$

### Exemplo com dados artificiais

Para ilustrar a aplicação do algoritmo descrito acima, geramos um conjunto de dados artificiais de tamanho 100, a partir do modelo em ():

$$Y_t | \psi_t \sim \text{Normal}(\mu_t, \lambda_t^{-1}) \quad (\text{C.5a})$$

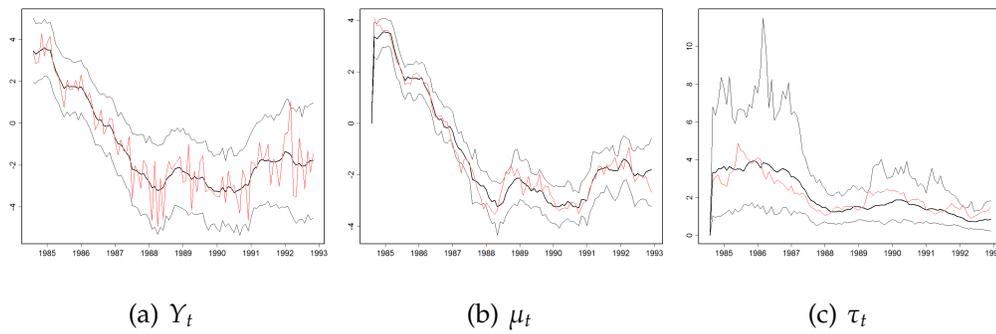
$$\mu_t = \theta_{1,t} \quad (\text{C.5b})$$

$$\log(\lambda_t) = \theta_{2,t} \quad (\text{C.5c})$$

$$\theta_{1,t} = \theta_{1,t-1} + w_{1,t}, w_{1,t} \sim N[0, W_1] \quad (\text{C.5d})$$

$$\theta_{2,t} = \theta_{2,t-1} + w_{2,t}, w_{2,t} \sim N[0, W_2]. \quad (\text{C.5e})$$

Ou seja, geramos uma série temporal que segue uma distribuição normal cujos parâmetros, média e (log)precisão, evoluem no tempo segundo um modelo dinâmico de primeira ordem. Para gerar os dados artificiais fixamos  $W_1$  em 0,30 e  $W_2$  em 0,05, e consideramos  $\theta_{1,0} = 4$  e  $\theta_{2,0} = 1$ . A figura abaixo mostra os resultados obtidos da utilização do *ECUBS* num algoritmo *MCMC* iterado por 10000 vezes. Embora as taxas de aceitação resultaram baixa, observamos que as amostras geradas aproximam bastante bem as séries geradas.



**Exemplo com dados artificiais da aplicação do *ECUBS* quando a resposta segue uma distribuição normal.** As linhas em vermelho correspondem às séries reais e as linhas em preto correspondem ao intervalo de 95% de credibilidade obtido.

## REFERÊNCIAS BIBLIOGRÁFICAS

- ABRAMOVITCH, M. & STEGUN, A. (1965). *Handbook of Mathematical Functions, with Formulas, Graphs and Mathematical Tables*. New York: Dover Publications.
- ALMON, S. (1965). The distributed lag between capital appropriations and expenditures. *Econometrica* 33 178–196.
- ALVES, M. (2005). *Funções de transferência em modelos dinâmicos generalizados*. Tese de Doutorado, Departamento de Estatística, IM - UFRJ, Rio de Janeiro, Brasil.
- BENTZEN, J. & ENGSTED, T. (2001). A revival of the autoregressive distributed lag model in estimating energy demand relationships. *Energy* 26 45–55.
- BERLINER, L. M. (2000). Hierarchical Bayesian modeling in the environmental sciences. *Allgemeine Statistische Archiv (Journal of the German Statistical Society)* 84 141–153.
- BERNARDO, J. & SMITH, A. (1994). *Bayesian Theory*. Chichester: John Wiley & Sons.
- BERNDT, E. (1991). *The Practice of Econometrics: Classic and Contemporary*. Reading: Addison Wesley.
- BEST, N., COWLES, M. & VINES, S. (1997). CODA: Convergence diagnosis and output analysis software for Gibbs sampling output, Version 0.4. URL <http://www.mrc-bsu.cam.ac.uk/bugs/classic/coda04/>.
- BROOKS, S. P. & GELMAN, A. (1998). Alternative methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics* 7 434–455.
- CAPKUN, G., DAVISON, A. C. & MUSY, A. (2001). A robust rainfall-runoff transfer model. *Water Resources Research* 37 3207–3216.
- CARTER, C. & KOHN, R. (1994). On Gibbs sampling for state space models. *Biometrika* 81 541–553.
- CARTER, R. & ZELLNER, A. (2002). The ARAR error model for univariate time series and distributed lag models. Relatório técnico, Department of Economics, University of Western Ontario.

- CEPEDA, E. & GAMERMAN, D. (2003). Bayesian methodology for modeling parameters in the two parameters exponential family. Relatório técnico, Instituto de Matemática/UFRJ.
- CHOTIKAPANICH, D. & GRIFFITHS, W. (1999). Flexible distributed lags. Relatório técnico, Department of Econometrics, University of New England.
- CONGDON, P. (2001). *Bayesian Statistical Modelling*. England: Wiley & sons.
- CRESSIE, N. (1993). *Statistics for Spatial Data*. New York: Wiley & sons.
- DEY, D., PENG, F. & GELFAND, A. (1997). Overdispersed generalized linear models. *Journal of Statistical Planning and Inference* 64 93–107.
- DI BELLO, R. (2005). *Análise do comportamento da umidade do solo no modelo chuva-vazão SMAP II - versão com suavização hiperbólica. Estudo de caso: região de Barreiras na bacia do Rio Grande - BA*. Dissertação de Mestrado, Engenharia Civil, COPPE - UFRJ, Rio de Janeiro, Brasil.
- DOORNIK, J. (2002). *Object-Oriented Matrix Programming Using Ox*. London: Timberlake Consultants Press and Oxford, 3rd ed. URL [www.nuff.ox.ac.uk/Users/Doornik](http://www.nuff.ox.ac.uk/Users/Doornik).
- ENGLE, R. F. & GRANGER, C. W. J. (1987). Cointegration and error correction: representation, estimation and testing. *Econometrica* 55 251–76.
- FAHRMEIR, L. (1992). Posterior mode estimation by extended Kalman filtering for multivariate dynamic linear models. *Journal of the American Statistical Association* 87 501–509.
- FERNANDES, M. V. (2005). *Modelos espaço-temporais inflacionados de zeros*. Dissertação de Mestrado, Instituto de Matemática - UFRJ, Rio de Janeiro, Brasil.
- FERRARI, S. & CRIBARI-NETO, F. (2004). Beta regression for modelling rates and proportions. *Journal of Applied Statistics* 31 799–815.
- FRANCES, P. H. & VAN OEST, R. (2004). On the econometrics of the Koyck model. Relatório técnico, Economic Institute, Erasmus University Rotterdam.
- FRÜHWIRTH-SCHNATER, S. (1994). Data augmentation and dynamic linear models. *Journal of Time Series Analysis* 15 183–202.
- GAMERMAN, D. (1997). Sampling from the posterior distribution in generalized linear mixed models. *Statistics and Computing* 7 57–68.

- GAMERMAN, D. (1998). Markov chain Monte Carlo for dynamic generalised linear models. *Biometrika* 85 215–227.
- GAMERMAN, D. & LOPES, H. F. (2006). *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*. New York: Chapman & Hall / CRC.
- GAMERMAN, D. & MIGON, H. (1993). Dynamic hierarchical models. *Journal of the Royal Statistical Society, B* 55 629–642.
- GELFAND, A. E. & GHOSH, S. (1998). Model choice: a minimum posterior predictive loss approach. *Biometrika* 85 1–11.
- GELFAND, A. E., SCHMIDT, A. M., BANERJEE, S. & SIRMANS, C. F. (2004). Nonstationary multivariate process modeling through spatially varying coregionalization. *Test* 13 263–312.
- GELFAND, A. E. & SMITH, A. (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association* 85 398–409.
- GELFAND, A. E., ZHU, L. & CARLIN, B. (2001). On the change of support problem for spatio-temporal data. *Biostatistics* 2 31–45.
- GELMAN, A. & RUBIN, D. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science* 7 457–511.
- GEWEKE, J. & TANIZAKI, H. (2001). Bayesian estimation of state space models using Metropolis-Hastings algorithm within Gibbs sampling. *Computational Statistics & Data Analysis* 37 151–170.
- GILKS, W., BEST, N. & TAN, K. (1995). Adaptive rejection metropolis sampling within Gibbs sampling. *Applied Statistics* 44 455–472.
- GILKS, W. & WILD, P. (1992). Adaptive rejection sampling for Gibbs sampling. *Applied Statistics* 41 337–348.
- GREENE, W. H. (1999). *Econometric Analysis*. Madrid: Prentice Hall, 3rd ed.
- GRILICHES, Z., MADDALA, G., LUCAS, R. & WALLACE, N. (1962). Notes on estimated aggregate quarterly consumption functions. *Econometrica* 30 491–500.
- GUJARATI, D. (2000). *Basic Econometrics*. New York: McGraw-Hill, 3rd ed.

- HASTINGS, W. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* 57 97–109.
- HUANG, Y., DOMINICI, F. & BELL, M. L. (2004). Bayesian hierarchical distributed lag models for summer ozone exposure and cardio-respiratory mortality. Relatório técnico, Johns Hopkins University, Dept. of Biostatistics.
- JOHANSEN, S. (1991). Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models. *Econometrica* 59 1551–1580.
- JORGENSON, D. (1966). Rational distributed lag functions. *Econometrica* 34 135–139.
- KITAGAWA, G. (1987). Non-gaussian state-space modeling of non-stationary time series. *Journal of the American Statistical Association* 82 1032–1041.
- KOYCK, L. M. (1954). *Distributed Lags models and Investment Analysis*. Amsterdam: North-Holland.
- LANCASTER, T. (2004). *An introduction to modern Bayesian Econometrics*. Oxford: Blackwell.
- LEE, Y. & NELDER, J. (1996). Hierarchical generalised linear models. *Journal of the Royal Statistical Society, B* 58 619–678.
- LEE, Y. & NELDER, J. (2001). Hierarchical generalised linear models: a synthesis of generalised linear models, random effects models and structured dispersions. *Biometrika* 88 987–1006.
- LEE, Y. & NELDER, J. (2006). Double hierarchical generalised linear models. *Applied Statistics* 55 1–29.
- LU, Z.-Q. & BERLINER, L. M. (1999). Markov switching time series models with application to a daily runoff series. *Water Resources Research* 35 523–534.
- METROPOLIS, N., ROSENBLUTH, M., TELLER, A. & TELLER, E. (1953). Equations of state calculations by fast computing machines. *Journal of Chemical Physics* 21 1087–1092.
- MEYER, R. & YU, J. (2000). BUGS for a Bayesian analysis of stochastic volatility models. *Econometrics Journal* 3 198–215.
- MIGON, H. (2000). The prediction of Brazilian exports using Bayesian forecasting. *Investigacion Operativa* 9 95–106.

- MIGON, H., GAMERMAN, D., LOPES, H. & FERREIRA, M. (2005). Bayesian dynamic models. Em D. Dey & C. Rao, eds., *Handbook of Statistic*, 25.
- MIGON, H. & HARRISON, J. (1985). An application of nonlinear Bayesian forecasting to television advertising. Em J. Bernardo, M. DeGroot, D. Lindley & A. Smith, eds., *Bayesian Statistics. Vol. 2*. North-Holland: Elsevier Science Publishers B.V., 681–696.
- MIGON, H. & MONTEIRO, A. B. (1997). Rain-fall modelling: An application of Bayesian forecasting. *Stochastic Hydrology and Hydraulics* 11 115–127.
- MONTEIRO, A. B. (1992). *Modelos dinâmicos aplicados a modelagem chuva-vazão*. Dissertação de Mestrado, Instituto de Matemática, UFRJ, Rio de Janeiro. Brasil.
- NEAL, R. (2003). Slice sampling (*with discussion*). *Annals of Statistics* 31 705–767.
- NERLOVE, M. (1958). Distributed lags and demand analysis. *Agricultural Handbook* 141.
- PESARAN, M. H. & SHIN, Y. (1999). An autoregressive distributed lag modeling approach to cointegration analysis. Em S. Strom, ed., *Econometrics and Economic Theory in the 20th century*. Cambridge University Press.
- PINTO, N., HOLTZ, A., MARTINS, J. & GOMIDE, F. (1976). *Hidrologia Básica*. São Paulo: Edgard Blucher.
- PLUMMER, M., BEST, N., COWLES, K. & VINES, K. (2005). *CODA: Output analysis and diagnostics for MCMC*. R package version 0.9-5, URL <http://www-fis.iarc.fr/coda/>.
- R DEVELOPMENT CORE TEAM (2005). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org>.
- RIGBY, R. & STASINOPOULOS, D. (2005). Generalized additive models for location, scale and shape. *Applied Statistics* 54 1–38.
- RIOS-INSUA, D., MONTES, R. & PALOMO, J. (2002). Bayesian methods in hydrology: a review. *Revista de la Real Academia de Ciencias. Serie A. Matemática* 96 461–479.
- RIOS-INSUA, D., SALEWIZ, K., MÜLLER, P. & BIELZA, C. (1997). Bayesian methods in reservoir operations: the Zambezi river case. Em S. French & J. Smith, eds., *The practice of Bayesian Analysis*. Arnold, 107–130.

- ROYLE, J. A. & BERLINER, L. M. (1999). A hierarchical approach to multivariate spatial modeling and prediction. *Journal of Agricultural, Biological and Environmental Statistics* 4 29–56.
- SALES, P. (1989). *Procedimentos lineares para identificação e estimação de parâmetros de modelos para séries temporais uni e multivariadas*. Tese de Doutorado, Engenharia de Produção, COPPE - UFRJ, Rio de Janeiro, Brasil.
- SANSÓ, B. & GUENNI, L. (2000). A non-stationary multi-site model for rainfall. *Journal of the American Statistical Association* 95 1089–1100.
- SOLOW, R. (1960). On a family of lag distributions. *Econometrica* 28 393–406.
- SPIEGELHALTER, D., BEST, N., CARLIN, B. & DER LINDE, A. V. (2001). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society, B* 64 583–639.
- SPIEGELHALTER, D., THOMAS, A. & BEST, N. (2003). winBUGS, Version 1.4. URL <http://www.mrc-bsu.cam.ac.uk/bugs>.
- VASCONCELLOS, K. & CRIBARI-NETO, F. (2005). Improved maximum likelihood estimation in a new class of beta regression models. *Brazilian Journal of Probability and Statistics* 19 13–31.
- VELARDE, L., MIGON, H. & PEREIRA, B. (2004). Space time modeling of non-negative variables with point of mass at zero. *Environmetrics* 15 561–576.
- WELTY, L. J. & ZEGER, S. L. (2004). Flexible distributed lag models: Are the acute effects of PM(10) on mortality the result of inadequate control for weather and season? Relatório técnico, Johns Hopkins University, Dept. of Biostatistics.
- WEST, M. & HARRISON, J. (1997). *Bayesian Forecasting and Dynamic Models*. New York: Springer-Verlag, 2nd ed.
- WEST, M., HARRISON, J. & MIGON, H. (1985). Dynamic generalized linear models and Bayesian forecasting. *Journal of the American Statistical Association* 80 73–83.
- ZELLNER, A. (1971). *An Introduction to Bayesian Inference in Econometrics*. New York: John Wiley & Sons.
- ZELLNER, A. & GEISEL, M. (1970). Analysis of distributed lag models with application to the consumption function. *Econometrica* 38 865–888.