

# **Modelagem espaço-temporal do desmatamento na Amazônia brasileira: Uma abordagem via modelos de regressão com coeficientes variando no espaço e no tempo**

Maritza Mendoza Sánchez



Universidade Federal do Rio de Janeiro  
Instituto de Matemática  
Departamento de Métodos Estatísticos

2025

**Modelagem espaço-temporal do desmatamento na Amazônia  
brasileira: Uma abordagem via modelos de regressão com  
coeficientes variando no espaço e no tempo**

Maritza Mendoza Sánchez

Dissertação de Mestrado submetida ao  
Programa de Pós-Graduação em Estatística  
do Instituto de Matemática da Universidade  
Federal do Rio de Janeiro-UFRJ, como parte  
dos requisitos necessários à obtenção do Título  
de Mestre em Estatística.

Orientador: João Batista de Moraes Pereira

Rio de Janeiro, RJ – Brasil

2025

**Modelagem espaço-temporal do desmatamento na Amazônia  
brasileira: Uma abordagem via modelos de regressão com  
coeficientes variando no espaço e no tempo**

Maritza Mendoza Sánchez

Dissertação de Mestrado submetida ao Programa de Pós-Graduação em Estatística  
do Instituto de Matemática da Universidade Federal do Rio de Janeiro-UFRJ, como  
parte dos requisitos necessários à obtenção do Título de Mestre em Estatística.

Aprovada por :

João Batista de Moraes Pereira

Prof. João Batista de Moraes Pereira, D.Sc, UFRJ

Marina S. Paez

Profª. Marina Silva Paez, D.Sc, UFRJ

Rafael Santos Erbisti

Prof. Rafael Santos Erbisti, D.Sc, UFF

Rio de Janeiro, RJ – Brasil

2025

## CIP - Catalogação na Publicação

M211m Mendoza Sanchez, Maritza  
Modelagem espaço-temporal do desmatamento na Amazônia brasileira: Uma abordagem via modelos de regressão com coeficientes variando no espaço e no tempo / Maritza Mendoza Sanchez. -- Rio de Janeiro, 2025.  
75 f.

Orientador: João Batista de Moraes Pereira.  
Dissertação (mestrado) - Universidade Federal do Rio de Janeiro, Instituto de Matemática, Programa de Pós-Graduação em Estatística, 2025.

1. Desmatamento. 2. Modelos espaço-temporais. 3. Modelos dinâmicos. 4. Modelo de regressão beta. I. de Moraes Pereira, João Batista, orient. II. Título.

*Dedico este trabalho ao meu pai,  
cuja presença permanece viva em  
cada conquista.*

*“O coração do homem pode fazer planos, mas a resposta certa  
vem dos lábios de Deus.”*

**Provérbios 16:1**

## Agradecimentos

Quero expressar, em primeiro lugar, minha gratidão a Deus, cuja misericórdia são novas cada manhã. Agradeço pela força, sabedoria e pelas oportunidades concedidas ao longo deste caminho desafiador, mas gratificante.

Agradeço de forma especial ao meu orientador, o professor João, cuja paciência, dedicação e orientação foram essenciais para o desenvolvimento deste trabalho. Seu compromisso acadêmico e sua capacidade de compartilhar conhecimento foram fundamentais para superar as dificuldades e transformar desafios em aprendizado. Sua disposição em orientar cada etapa deste processo fez toda a diferença, e por isso, fico muito grata.

Agradeço aos meus pais, Edelmira e Roque, pela educação que me proporcionaram, pelo amor incondicional e pelos inúmeros sacrifícios que fizeram por mim. Obrigada por cada ensinamento, pelas exortações e pelo afeto constante.

Ao meu esposo, Dany, pelo amor e companheirismo.

Sou grata aos professores Giulio, Guillerme Ost e Carlos Tadeu que contribuíram para a minha formação. Em especial, ao professor Carlos Abanto, pelo apoio desde a minha chegada ao Brasil. Sua exigência me motivaram a superar meus próprios limites.

Ao programa de Pós-Graduação em Estatística da UFRJ, pela oportunidade concedida

Ao professor Glauco Taranto, do LASPOT, por me acolher e disponibilizar um espaço de estudo, mesmo não sendo oficialmente parte do seu grupo. Fico grata.

Agradeço à Secretaria de Pós Graduação do IM-UFRJ, especialmente ao Claudio, pela ajuda contínua e pela disposição em explicar as coisas de forma clara e atenciosa.

À CAPES, pelo financiamento dos meus estudos, os quais seriam impossíveis sem este apoio financeiro.

## Resumo

O desmatamento na Amazônia brasileira é um problema ambiental crítico, com impactos significativos no clima, na biodiversidade e no ciclo do carbono. Modelar a dinâmica espaço-temporal desse processo pode ser de extrema importância para compreender seus padrões e identificar fatores associados. Assim, neste trabalho, propõe-se uma modelagem espaço-temporal para a proporção da área desmatada no bioma Amazônia, explorando diferentes abordagens metodológicas. Inicialmente, é utilizado o modelo de regressão geograficamente ponderada (GWR, na sigla em inglês) para analisar a variabilidade espacial e o modelo de regressão geograficamente e temporalmente ponderada (GTWR, na sigla em inglês) para analisar a variabilidade espaço-temporal dos efeitos das variáveis associadas ao desmatamento.

A partir dessas análises e da limitação dos modelos GWR e GTWR, identificou-se a necessidade de uma abordagem mais robusta, levando à adoção de um modelo de regressão espaço-temporal com coeficientes variando no espaço e no tempo. Especificamente, considerou-se um modelo normal para o logit da proporção de área desmatada, em que a variabilidade espacial dos coeficientes é capturada por modelos autorregressivos condicionais (CAR) e a variabilidade temporal por modelos lineares dinâmicos. Como alternativa ao modelo normal, também se propõe um modelo de regressão beta, que modela diretamente a proporção de área desmatada em sua escala original.

A avaliação desses modelos foi realizada por meio de aplicações em dados simulados e em dados reais para municípios do estado do Pará. O modelo normal foi aplicado também para dados reais de municípios de todo o Bioma Amazônico. A inferência seguiu o enfoque bayesiano, utilizando métodos de Monte Carlo via cadeias de Markov para estimar os parâmetros. Os resultados obtidos proporcionam uma compreensão mais aprofundada dos fatores que influenciam o desmatamento e apresentam ferramentas metodológicas que podem contribuir para estudos ambientais e a formulação de políticas públicas.

*Palavras-chave:* desmatamento, modelos espaço-temporais, modelos dinâmicos, modelo de regressão beta.

## Abstract

Deforestation in the Brazilian Amazon is a critical environmental issue with significant impacts on climate, biodiversity, and the carbon cycle. Modeling the spatio-temporal dynamics of this process is crucial for understanding its patterns and identifying associated factors. Therefore, this study proposes a spatiotemporal modeling approach for the proportion of deforested area in the Amazon biome, exploring different methodological approaches. Initially, the Geographically Weighted Regression (GWR) model is employed to analyze spatial variability, and the Geographically and Temporally Weighted Regression (GTWR) model is used to assess the spatiotemporal variability of the effects of variables associated with deforestation.

Based on these analyses and the limitations of GWR and GTWR models, the need for a more robust approach was identified, leading to the adoption of a spatiotemporal regression model with coefficients varying across space and time. Specifically, a normal model is considered for the logit of the proportion of deforested area, where the spatial variability of the coefficients is captured by Conditional Autoregressive (CAR) models and the temporal variability by dynamic linear models. As an alternative to the normal model, a beta regression model is also proposed, which directly models the proportion of deforested area in its original scale.

The evaluation of these models was performed using applications on both simulated and real data for municipalities in the state of Pará. The normal model was also applied to real data from municipalities across the entire Amazon biome. The inference followed a Bayesian approach, employing Markov Chain Monte Carlo methods to estimate the parameters. The results provide a deeper understanding of the factors influencing deforestation and present methodological tools that can contribute to environmental studies and the formulation of public policies.

*Keywords:* deforestation, spatiotemporal models, dynamic models, beta regression model.

# Sumário

<b>Lista de Figuras</b>	xii
<b>Lista de Tabelas</b>	xvii
<b>1 Introdução</b>	1
1.1 Organização do documento . . . . .	6
<b>2 Análise preliminar do desmatamento via modelos GWR e GTWR</b>	7
<b>3 Modelo de regressão espaço-temporal com coeficientes variando no espaço e no tempo</b>	15
3.1 Modelo proposto . . . . .	16
3.2 Procedimento de inferência . . . . .	19
3.2.1 Sorteando $\beta_t$ . . . . .	21
3.2.2 Sorteando os demais parâmetros . . . . .	22
3.3 Aplicação: dados artificiais . . . . .	25
3.3.1 Modelo considerando $0 < \alpha < 1$ . . . . .	25
3.3.2 Modelo considerando $\alpha = 1$ . . . . .	30
<b>4 Modelagem espaço-temporal do desmatamento</b>	34
4.1 Análise para o estado do Pará . . . . .	34
4.1.1 Modelo considerando $0 < \alpha < 1$ . . . . .	35
4.1.2 Modelo considerando $\alpha = 1$ . . . . .	38
4.2 Análise para o Bioma Amazônico . . . . .	43

<b>5</b>	<b>Modelo de regressão beta espaço-temporal com coeficientes variando no espaço e no tempo</b>	<b>46</b>
5.1	Modelo proposto . . . . .	47
5.2	Procedimento de inferência . . . . .	47
5.2.1	Distribuições condicionais completas . . . . .	49
5.3	Aplicação: dados artificiais . . . . .	50
5.3.1	Modelo considerando $0 < \alpha < 1$ . . . . .	51
5.3.2	Modelo considerando $\alpha = 1$ . . . . .	54
<b>6</b>	<b>Modelagem espaço-temporal do desmatamento com regressão beta</b>	<b>58</b>
6.1	Análise para o estado do Pará . . . . .	58
6.1.1	Modelo considerando $\alpha \in (0, 1)$ . . . . .	59
6.1.2	Modelo considerando $\alpha = 1$ . . . . .	61
<b>7</b>	<b>Conclusões e trabalhos futuros</b>	<b>66</b>
7.1	Trabalhos futuros . . . . .	67
<b>A</b>	<b>Traços das cadeias dos coeficientes temporais e espaciais do modelo normal</b>	<b>68</b>
<b>B</b>	<b>Critério de informação Watanabi Akaike</b>	<b>70</b>
<b>C</b>	<b>Traços das cadeias dos coeficientes temporais e espaciais do modelo de regressão beta</b>	<b>71</b>
	<b>Referências Bibliográficas</b>	<b>73</b>

# **Lista de Figuras**

1.1	Mapa do Bioma Amazônico. . . . .	2
1.2	Desmatamento anual no Bioma Amazônico entre 2002 e 2022. . . . .	3
1.3	Evolução anual da quantidade de área desmatada no período de 2002 a 2022 para os estados do bioma. . . . .	3
1.4	Mapas coropléticos da proporção de área desmatada por município nos anos 2004, 2010, 2016 e 2022. . . . .	4
1.5	Evolução da área desmatada ao longo dos anos em alguns municípios selecionados. . . . .	5
2.1	Coeficientes estimados através do modelo GWR nos anos 2003, 2009, 2015 e 2022. . . . .	12
2.2	Coeficientes estimados através do modelo GTWR nos anos 2003, 2009, 2015 e 2022. . . . .	13
2.3	Estimativas do intercepto e dos coeficientes das covariáveis, obtidas por meio do modelo GWR, com base nas médias do desmatamento e das variáveis explicativas entre 2002 e 2022. . . . .	14
3.1	Mapa do estado do Pará . . . . .	25
3.2	Dados artificiais simulado do modelo. . . . .	26
3.3	Valores estimados (linha preta) e valores verdadeiros (linha vermelha) com respectivos intervalos de 95% de credibilidade <i>a posteriori</i> (área hachurada) dos coeficientes temporais ao longo do tempo, para o modelo normal com $\alpha \in (0, 1)$ . . . . .	26

3.4	Valores estimados (pontos) e valores reais (cruzes) do intercepto e coeficientes espaciais das variáveis preditoras com respectivos intervalos de 95% de credibilidade (linhas verticais) para $i = 1, \dots, 144$ , no modelo normal com $\alpha \in (0, 1)$ . . . . .	27
3.5	Valores verdadeiros e valores estimados do intercepto e dos coeficientes espaciais das variáveis preditoras para todos os municípios, no modelo normal com $\alpha \in (0, 1)$ . . . . .	28
3.6	Traços das cadeias e histogramas das distribuições <i>a posteriori</i> de $\sigma_\eta^2$ , $\sigma_\varepsilon^2$ , $\sigma_\beta^2$ e $\alpha$ do modelo normal. As linhas vermelhas horizontal e vertical indicam o valor verdadeiro. . . . .	29
3.7	Dados artificiais simulado do modelo. . . . .	30
3.8	Valores estimados através da média <i>a posteriori</i> (linha preta) e valores verdadeiros (linha vermelha) com respectivos intervalos de 95% de credibilidade (área hachurada) dos coeficientes temporais ao longo do tempo, para o modelo normal com $\alpha = 1$ . . . . .	31
3.9	Valores estimados (pontos) e valores reais (cruzes) do intercepto e coeficientes espaciais das variáveis preditoras com respectivos intervalos de 95% de credibilidade (linhas verticais) para $i = 1, \dots, 144$ , no modelo normal com $\alpha = 1$ . . . . .	31
3.10	Valores verdadeiros simulados e valores estimados através da média <i>a posteriori</i> para o intercepto e os coeficientes espaciais das variáveis preditoras para todos os municípios, no modelo normal com $\alpha = 1$ . . . . .	32
3.11	Traços das cadeias e histogramas das distribuições <i>a posteriori</i> de $\sigma_\eta^2$ , $\sigma_\varepsilon^2$ , $\sigma_\beta^2$ no modelo normal. As linhas vermelhas horizontal e vertical indicam o valor verdadeiro. . . . .	33
4.1	Valores estimados através da média <i>a posteriori</i> (linha preta) com respectivos intervalos de 95% de credibilidade (área hachurada) dos coeficientes temporais ao longo do tempo, para o modelo normal com $\alpha \in (0, 1)$ . . . . .	35
4.2	Valores estimados através da média <i>a posteriori</i> para o intercepto e os coeficientes espaciais das variáveis preditoras para todos os municípios no modelo normal com $\alpha \in (0, 1)$ . . . . .	36
4.3	Traços das cadeias e histogramas das distribuições <i>a posteriori</i> de $\sigma_\eta^2$ , $\sigma_\varepsilon^2$ , $\sigma_\beta^2$ e $\alpha$ no modelo normal para dados reais. . . . .	37

4.4	Valores estimados através da média <i>a posteriori</i> (linha preta) com respectivos intervalos de 95% de credibilidade (área hachurada) dos coeficientes temporais ao longo do tempo, para o modelo normal com $\alpha = 1$ .	38
4.5	Valores estimados através da média <i>a posteriori</i> para o intercepto e os coeficientes espaciais das variáveis preditoras para todos os municípios no modelo normal com $\alpha = 1$	39
4.6	Traços das cadeias e histogramas das distribuições <i>a posteriori</i> de $\sigma_\eta^2$ , $\sigma_\epsilon^2$ , $\sigma_\beta^2$ no modelo normal para dados reais.	40
4.7	Gráfico dos valores replicados versus os valores observados.	41
4.8	Mapas dos $\tilde{\beta}(s, t)$ para o modelo normal com $\alpha = 1$ com efeitos significativamente positivos (em vermelho), não significativos (em branco) e negativos (em azul) para alguns anos.	42
4.9	Média <i>a posteriori</i> (linha preta) com respectivos intervalos 95% credibilidade (área hachurada) dos coeficientes temporais ao longo dos anos, para o Bioma Amazônico.	44
4.10	Valores estimados do intercepto e dos coeficientes espaciais das variáveis para o Bioma Amazônico.	44
4.11	Traços das cadeias e histogramas das distribuições <i>a posteriori</i> de $\sigma_\eta^2$ , $\sigma_\epsilon^2$ , $\sigma_\beta^2$ no modelo normal para o bioma.	45
5.1	Valores estimados através da média <i>a posteriori</i> (linha preta) e valores verdadeiros (linha vermelha) com respectivos intervalos de 95% de credibilidade (área hachurada) dos coeficientes temporais ao longo do tempo, para o modelo beta com $\alpha \in (0, 1)$ .	51
5.2	Valores estimados (pontos) e valores reais (cruzes) do intercepto e coeficientes espaciais das variáveis preditoras com respectivos intervalos de 95% de credibilidade (linhas verticais) para $i = 1, \dots, 144$ , no modelo beta com $\alpha \in (0, 1)$ .	52
5.3	Valores verdadeiros do intercepto e dos coeficientes espaciais das variáveis preditoras para o modelo beta com $\alpha \in (0, 1)$ .	53
5.4	Traços das cadeias e histogramas das distribuições <i>a posteriori</i> de $\sigma_\eta^2$ , $\sigma_\epsilon^2$ , $\sigma_\beta^2$ , $\alpha$ e $\phi$ do modelo beta. As linhas vermelhas horizontal e vertical indicam o valor verdadeiro.	54

5.5	Valores estimados através da média <i>a posteriori</i> (linha preta) e valores verdadeiros (linha vermelha) com respectivos intervalos de 95% de credibilidade (área hachurada) dos coeficientes temporais ao longo do tempo, para o modelo beta com $\alpha = 1$ . . . . .	55
5.6	Valores estimados (pontos) e valores reais (cruzes) do intercepto e coeficientes espaciais das variáveis preditoras com respectivos intervalos de 95% de credibilidade (linhas verticais) para $i = 1, \dots, 144$ , para o modelo normal com $\alpha = 1$ . . . . .	56
5.7	Valores verdadeiros e estimados do intercepto e dos coeficientes espaciais das variáveis preditoras para todos os municípios, para o modelo normal com $\alpha = 1$ . . . . .	56
5.8	Traços das cadeias e histogramas das distribuições <i>a posteriori</i> de $\sigma_\eta^2$ , $\sigma_\varepsilon^2$ , $\sigma_\beta^2$ e $\phi$ para o modelo beta. . . . .	57
6.1	Valores estimados através da média <i>a posteriori</i> (linha preta) com respectivos intervalos 95% credibilidade (área hachurada) dos coeficientes temporais ao longo dos anos, para o modelo beta com $\alpha \in (0, 1)$ . . . . .	59
6.2	Valores estimados através da média <i>a posteriori</i> para o intercepto e os coeficientes espaciais das variáveis preditoras para todos os municípios no modelo beta com $\alpha \in (0, 1)$ . . . . .	60
6.3	Traços das cadeias e histogramas das distribuições <i>a posteriori</i> de $\sigma_\eta^2$ , $\sigma_\varepsilon^2$ , $\sigma_\beta^2$ e $\phi$ no modelo beta em dados reais. . . . .	61
6.4	Valores estimados através da média <i>a posteriori</i> (linha preta) com respectivos intervalos 95% credibilidade (área hachurada) dos coeficientes temporais ao longo dos anos, para o modelo beta com $\alpha = 1$ . . . . .	62
6.5	Valores estimados através da média <i>a posteriori</i> para o intercepto e os coeficientes espaciais das variáveis preditoras para todos os municípios no modelo beta com $\alpha = 1$ . . . . .	62
6.6	Traços das cadeias e histogramas das distribuições <i>a posteriori</i> de $\sigma_\eta^2$ , $\sigma_\varepsilon^2$ , $\sigma_\beta^2$ e $\phi$ no modelo beta em dados reais. . . . .	63
6.7	Gráfico dos valores replicados versus os valores observados para o modelo beta. . . . .	64

6.8	Mapas dos $\tilde{\beta}(s, t)$ para o modelo beta com $\alpha = 1$ com efeitos significativamente positivos (em vermelho), não significativos (em branco) e negativos (em azul) para alguns anos. . . . .	65
A.1	Traços das cadeias de alguns coeficientes temporais $\beta_t$ e espaciais $\beta$ para o modelo normal com $\alpha = 1$ em dados artificiais. . . . .	68
A.2	Traços das cadeias de alguns coeficientes temporais $\beta_t$ e espaciais $\beta$ para o modelo normal com $\alpha = 1$ para dados reais. . . . .	69
C.1	Traços das cadeias de alguns coeficientes temporais $\beta_t$ e espaciais $\beta$ para o modelo beta com $\alpha \in (0, 1)$ em dados artificiais. . . . .	71
C.2	Traços das cadeias de alguns coeficientes temporais $\beta_t$ e espaciais $\beta$ para o modelo beta com $\alpha = 1$ em dados artificiais. . . . .	72
C.3	Traços das cadeias de alguns coeficientes temporais $\beta_t$ e espaciais $\beta$ para o modelo beta com $\alpha = 1$ para dados reais. . . . .	72

# Lista de Tabelas

2.1	Matriz de correlação . . . . .	8
2.2	Ajuste do modelo de regressão . . . . .	8
2.3	Descrição das variáveis com suas respectivas unidades de medida e fontes de informação, referentes ao período de 2002 a 2022. . . . .	9
3.1	Resumo das distribuições <i>a posteriori</i> de $\sigma_\eta^2$ , $\sigma_\varepsilon^2$ , $\sigma_\beta^2$ e $\alpha$ no modelo normal. . . . .	29
3.2	Resumo das distribuições <i>a posteriori</i> de $\sigma_\eta^2$ , $\sigma_\varepsilon^2$ , $\sigma_\beta^2$ do modelo normal. . . . .	32
4.1	Resumo da distribuição <i>a posteriori</i> de $\sigma_\eta^2$ , $\sigma_\varepsilon^2$ , $\sigma_\beta^2$ e $\alpha$ do modelo normal aplicado aos dados reais. . . . .	37
4.2	Resumo das distribuições <i>a posteriori</i> de $\sigma_\eta^2$ , $\sigma_\varepsilon^2$ , $\sigma_\beta^2$ do modelo normal em dados reais. . . . .	39
4.3	Critério de comparação de modelos WAIC. . . . .	41
4.4	Resumo da distribuição <i>a posteriori</i> dos parâmetros $\sigma_\eta^2$ , $\sigma_\varepsilon^2$ e $\sigma_\beta^2$ do modelo normal no Bioma Amazônico. . . . .	45
5.1	Resumo das distribuições <i>a posteriori</i> de $\sigma_\eta^2$ , $\sigma_\varepsilon^2$ , $\sigma_\beta^2$ , $\alpha$ e $\phi$ no modelo beta. . . . .	53
5.2	Resumo das distribuições <i>a posteriori</i> de $\sigma_\eta^2$ , $\sigma_\varepsilon^2$ , $\sigma_\beta^2$ e $\phi$ do modelo beta. . . . .	57
6.1	Resumo da distribuição <i>a posteriori</i> dos parâmetros $\sigma_\eta^2$ , $\sigma_\varepsilon^2$ , $\sigma_\beta^2$ , $\alpha$ e $\phi$ no modelo beta. . . . .	60
6.2	Resumo das distribuições <i>a posteriori</i> de $\sigma_\eta^2$ , $\sigma_\varepsilon^2$ , $\sigma_\beta^2$ e $\phi$ para o modelo beta com $\alpha = 1$ . . . . .	63
6.3	Critério de comparação de modelos WAIC para o modelo de regressão beta. . . . .	64

# Capítulo 1

## Introdução

A Amazônia, vasta e exuberante floresta tropical que abrange grande parte do noroeste do Brasil, Colômbia, Peru e outros países da América do Sul, desempenha um papel vital na preservação da saúde do planeta, especialmente através de sua contribuição fundamental para a regulação climática (Malhi; Roberts; Betts, 2008). Além disso, a rica biodiversidade e os ecossistemas complexos da Amazônia são indispensáveis para sustentar o equilíbrio ecológico e apoiar diversas formas de vida, tornando-a uma pedra angular dos esforços globais de conservação da biodiversidade. No entanto, no transcorrer das últimas décadas, a Amazônia tem sido alvo de intensa exploração e degradação ambiental, ocasionada principalmente pela atuação humana (Ellwanger; Kulmann-Leal; Kaminski, 2020).

O desmatamento é um processo devastador que ganhou atenção global devido aos seus efeitos ambientais, sociais e econômicos. Segundo (Fearnside, 2005), esse fenômeno contribui para a perda de produtividade, a alteração do regime hidrológico, a perda de biodiversidade e o aumento das emissões de gases de efeito estufa. Estudos como o de Correia, Manzi e Cândido (2007) destacam seu impacto no ciclo hidrológico, evidenciando a sensibilidade do balanço de umidade às mudanças na cobertura vegetal. Todas essas repercussões evidenciam a complexidade do problema e a urgência de ações para enfrentá-lo.

Diversos estudos apontam que o desmatamento é impulsionado por diferentes fatores, tais como: expansão agrícola, exploração madeireira, pecuária, crédito rural, entre outros. Nesse contexto, os autores Barona, Navin e Glenn (2010) investigaram o papel do pasto e da soja na expansão da fronteira agrícola, destacando sua relação com o aumento da destruição florestal. Domingues e Bermann (2012) argumentam que a expansão de novas áreas para a pecuária tem sido uma das principais causas

da diminuição da vegetação, resultando na conversão das florestas em pastagens para a criação de bovinos. No entanto, segundo [Assunção, Clarissa e Romero \(2020\)](#), o aumento do desmatamento se deve, em grande medida, aos incentivos proporcionados pelo crédito rural destinado aos produtores, e também destacam que a mudança nas políticas públicas promoveu uma redução significativa do desmatamento, especialmente em municípios onde a pecuária é a principal atividade econômica. Esses e outros fatores que conduzem ao desmatamento estão constantemente presentes. No entanto, eles interagem de maneiras distintas, dependendo da dinâmica espacial e temporal de contextos regionais ou locais específicas ([Geist; Lambin, 2001](#)).

No Brasil, o Bioma Amazônico tem uma extensão de 4.212.472 km<sup>2</sup>, correspondendo a cerca de 49,5% do território nacional ([IBGE, 2019](#)). Essa imensa área abrange integralmente os estados do Acre, Amazonas, Pará, Rondônia, Roraima e, parcialmente, o Maranhão, Tocantins e Mato Grosso, totalizando 559 municípios como mostra a Figura 1.1.

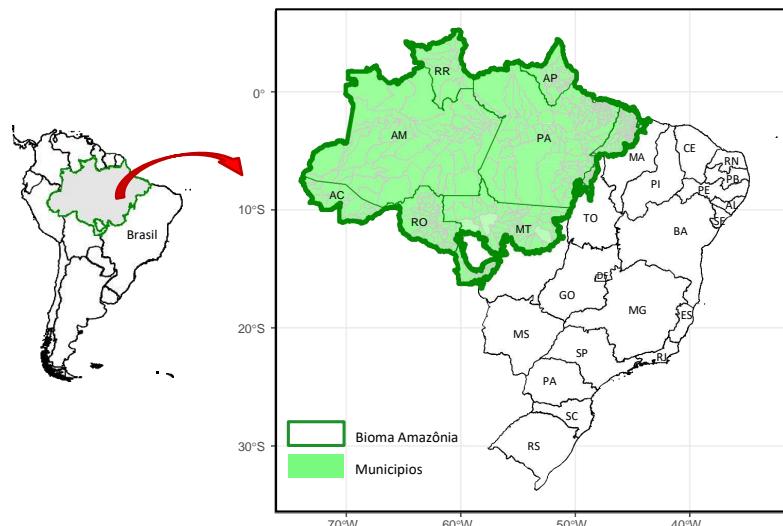


Figura 1.1: Mapa do Bioma Amazônico.

Por outro lado, os índices de desmatamento no bioma entre 2002 e 2022, com base nos dados fornecidos pelo [INPE \(2023\)](#) através do PRODES (Projeto de Monitoramento do Desmatamento na Amazônia Legal por Satélite), revelam que a área desmatada em 2003 atingiu 24.700 km<sup>2</sup>. Em menos de uma década, a taxa de desmatamento despencou para 4.127 km<sup>2</sup> em 2012. Essa redução significativa foi impulsionada, em grande parte, por políticas públicas de prevenção e controle do desmatamento ([Nepstad; Mccgrath; Stickler, 2014](#)). No entanto, devido ao enfraquecimento das ações de combate, as taxas de desmatamento voltaram a subir a partir de 2015, com aumentos

expressivos em 2021 e 2022: 12.192 km<sup>2</sup> e 12.48 km<sup>2</sup>, respectivamente, como mostra a Figura 1.2.

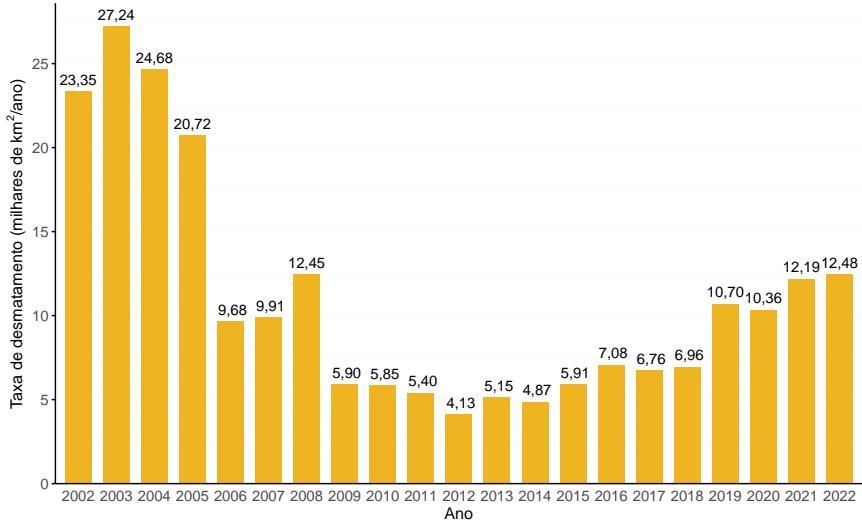


Figura 1.2: Desmatamento anual no Bioma Amazônico entre 2002 e 2022.

Entre 2002 e 2005, o Mato Grosso destacou-se com a maior área desmatada entre os estados do bioma, e a partir do 2006 o Pará passou a liderar essa taxa. Em 2022, o Pará registrou 4.528 km<sup>2</sup> de área desmatada; enquanto os demais estados apresentaram as seguintes taxas: Amazonas 3.047 km<sup>2</sup>, Mato Grosso 1.980 km<sup>2</sup>, Rondônia 1.448 km<sup>2</sup>, Acre 1.005 km<sup>2</sup>, Roraima 293 km<sup>2</sup>, Maranhão 144 km<sup>2</sup> e Amapá 14 km<sup>2</sup> (ver Figura 1.3).

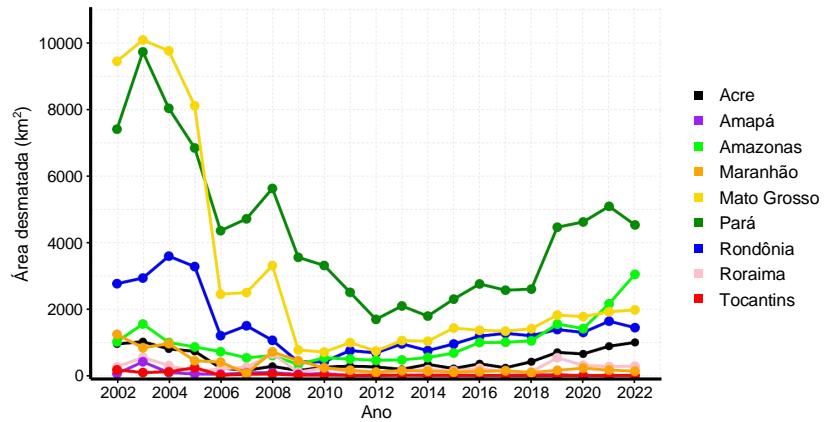


Figura 1.3: Evolução anual da quantidade de área desmatada no período de 2002 a 2022 para os estados do bioma.

A Figura 1.4 apresenta mapas que ilustram a distribuição geográfica da proporção de área desmatada em cada município para anos específicos. Em 2004, o município de

Buritis registrou 237,1 km<sup>2</sup> de desmatamento, correspondendo a 7,3% de sua extensão territorial, enquanto o município de Itanhangá atingiu 185,2 km<sup>2</sup>, representando 6,4% de sua área total. Esses valores configuraram as maiores proporções de desmatamento registradas naquele período. Nos anos seguintes, observou-se uma tendência de redução em diversos municípios. No entanto, a partir de 2016, houve um aumento em algumas regiões, como Buritis, Cujubim e entre outros, indicando uma mudança na dinâmica do desmatamento em determinadas áreas. Cada mapa foi elaborado com uma escala de cores específica para cada um dos anos selecionados, de modo a evidenciar com maior clareza as variações espaciais observadas entre os municípios.

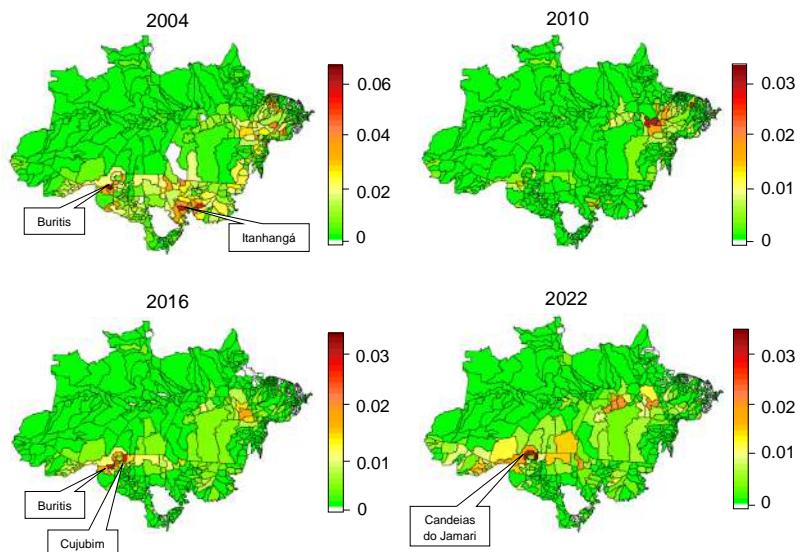


Figura 1.4: Mapas coropléticos da proporção de área desmatada por município nos anos 2004, 2010, 2016 e 2022.

A Figura 1.5 apresenta informações dos municípios onde o desmatamento tem ocorrido de forma mais acentuada ao longo do tempo, evidenciando os índices de perda florestal e a necessidade de ações estratégicas para a formulação de políticas de mitigação.

A análise exploratória da área desmatada nos municípios e estados do bioma Amazônia ao longo dos anos evidencia uma expressiva heterogeneidade espacial, acompanhada de variações temporais significativas. Embora esses dados não indiquem diretamente os fatores responsáveis, eles ilustram a dinâmica do desmatamento na região e reforçam a necessidade de abordagens analíticas mais abrangentes.

Nesse contexto, este trabalho é motivado pelo interesse na modelagem dinâmica espaço-temporal dos processos que influenciam o desmatamento no Bioma Amazônico. Especificamente, tem-se por objetivo estudar e entender a relação entre as variáveis

preditoras e a taxa de desmatamento nos municípios da região, considerando sua variação no espaço e no tempo.

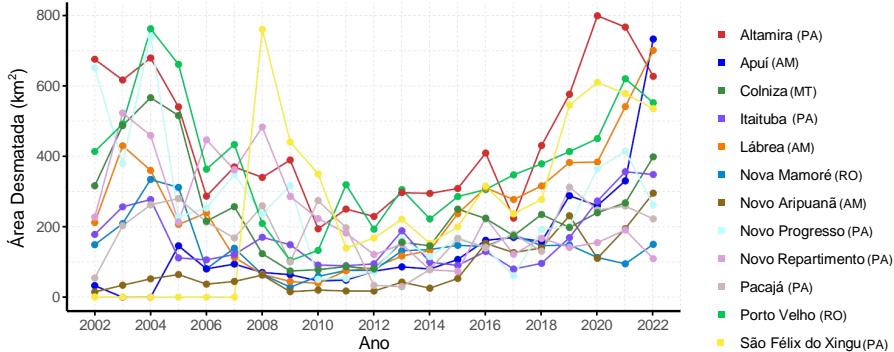


Figura 1.5: Evolução da área desmatada ao longo dos anos em alguns municípios selecionados.

Como etapa inicial da investigação, realizaram-se análises exploratórias com o objetivo de obter uma visão preliminar da dinâmica do desmatamento na região. Para isso, analisou-se a variabilidade espacial dos efeitos das variáveis associadas ao desmatamento com base no trabalho de [Santos, Assunção e Monteiro \(2021\)](#), que empregaram o modelo GWR para analisar essa relação no espaço. Com o aprofundamento da análise, considerou-se também o modelo GTWR desenvolvido por [Fotheringham, Crespo e Yao \(2015\)](#), que permite incorporar simultaneamente variações espaciais e temporais. Embora esses modelos ofereçam uma base útil para compreender padrões iniciais, apresentam limitações metodológicas: ambos exigem a escolha arbitrária da largura de banda, e o GTWR, em particular, não modela explicitamente a estrutura de dependência espaço-temporal.

Diante da necessidade de uma abordagem que integrasse melhor os aspectos espaciais e temporais, adotou-se um modelo espaço-temporal univariado, baseado no trabalho de [Gelfand, Banerjee e Gamerman \(2005\)](#), o qual permite que os coeficientes de regressão variem no espaço e no tempo. Nesse contexto, uma diferença fundamental entre o modelo proposto neste trabalho e dos autores mencionados é a abordagem adotada, pois enquanto eles utilizam uma abordagem geoestatística baseada em processos gaussianos, este trabalho opta por uma modelagem baseada em modelos autorregressivos condicionais (CAR) para modelar a dependência espacial.

A suposição de normalidade da variável resposta pode levar a previsões fora do intervalo  $(0, 1)$ , além de afetar a representação da dispersão dos dados. Para contornar esse problema, adotaram-se duas estratégias: transformar a variável resposta por meio do logit, que mapeia valores de  $(0, 1)$  para todo o conjunto de números reais, permitindo

aplicá-la no modelo normal com dados reais; e utilizar o modelo de regressão beta, que se ajusta diretamente à natureza dos dados.

Além disso, este trabalho se justifica pela necessidade de uma abordagem mais robusta para analisar o desmatamento, em comparação com abordagens convencionais, oferecendo uma maior compreensão do problema, o que pode subsidiar na formulação de políticas públicas mais eficientes para mitigar seu avanço.

Ainda, o estudo alinha-se aos objetivos da Agenda 2030 da Organização das Nações Unidas (ONU) para o Brasil, especialmente com o ODS 13 (Ação contra a mudança global do clima) e o ODS 15 (Vida terrestre), pois a análise do desmatamento está relacionada à mitigação das mudanças climáticas e à conservação da biodiversidade.

Para todas as análises, assim como para a implementação dos modelos, empregou-se o software R ([R Core Team, 2022](#)), que possibilitou a geração dos resultados apresentados neste estudo.

## 1.1 Organização do documento

No Capítulo 2, analisa-se a variabilidade espacial dos efeitos das variáveis associadas ao desmatamento utilizando o modelo GWR. Para capturar também a variabilidade temporal, a análise é aprofundada com a aplicação do modelo GTWR.

No Capítulo 3, apresenta-se o modelo proposto para a modelagem espaço-temporal do desmatamento. Especificamente, é descrito o modelo espaço-temporal e sua aplicação a dados artificiais. Em seguida, no Capítulo 4, esse modelo é aplicado aos dados reais do desmatamento.

O Capítulo 5 introduz uma mudança metodológica ao apresentar o modelo de regressão beta. Nele, são discutidos os principais conceitos do modelo, sua extensão para o contexto espaço-temporal e sua aplicação a dados simulados no estudo do desmatamento. Na sequência, o Capítulo 6 aborda a aplicação do modelo de regressão beta aos dados reais do desmatamento.

Por fim, no Capítulo 7, são apresentadas as conclusões do estudo, destacando os principais achados da pesquisa e sua relevância para a compreensão dos padrões de desmatamento na região analisada. Além disso, são discutidas as possíveis melhorias e trabalhos futuros.

# Capítulo 2

## Análise preliminar do desmatamento via modelos GWR e GTWR

Neste capítulo, são conduzidas análises exploratórias e descritivas da variabilidade espacial e temporal das variáveis que podem influenciar o desmatamento no Bioma Amazônico. Para isso, utilizam-se os modelos de regressão geograficamente ponderada (GWR) e regressão geograficamente e temporalmente ponderada (GTWR), com o objetivo de identificar possíveis padrões nos dados. Ressalta-se que os ajustes são realizados sob a perspectiva da inferência estatística clássica, sem caráter preditivo ou inferencial formal nesta etapa da análise.

Com base nos estudos de [Araujo, Bonjean e Combes \(2009\)](#), [Santos, Assunção e Monteiro \(2021\)](#) e [Guimarães, Santos e Rudke \(2023\)](#), que utilizam modelos estatísticos para identificar as principais causas do desmatamento, selecionou-se um conjunto de variáveis potencialmente associadas a esse fenômeno, tais como: área de vegetação nativa remanescente (Floresta), número de cabeças de gado bovino (Gado), área total com culturas temporárias cultivadas (Lavoura), quantidade de madeira extraída (Madeira), estimativa populacional (População), crédito rural (CrediRural), e PIB per capita (PIBpc). Essas variáveis, cujas descrições, unidades de medida e fontes de dados estão detalhadas na Tabela 2.3, abrangem o período de 2002 a 2022.

A seleção das variáveis preditoras foi realizada utilizando o método stepwise, o qual é uma técnica estatística utilizada para identificar o subconjunto de variáveis preditoras que melhor explicam a variável resposta em um modelo. Para isso, utilizaram-se as médias das covariáveis ao longo do período analisado, de 2002 a 2022, a fim de

identificar o subconjunto mais relevante para o modelo.

A Tabela 2.1 apresenta a correlação entre as variáveis preditoras selecionadas. Observa-se que a média da variável Gado apresenta a maior correlação com o desmatamento, enquanto as demais covariáveis não apresentam correlações fortes entre si. Como destacado por [Vatcheva, Lee e McCormick \(2016\)](#), ignorar a multicolinearidade pode comprometer tanto os resultados quanto a interpretação das análises de regressão.

Considerando esse aspecto, foi aplicado o modelo de regressão por Mínimos Quadrados Ordinários com o objetivo de identificar associações estatisticamente significativas entre a variável resposta (desmatamento) e as variáveis explicativas. A Tabela 2.2 apresenta as estimativas dos coeficientes de regressão, os p-valores e os fatores de inflação da variância (VIF), os quais foram calculados com o objetivo de identificar a presença de multicolinearidade.

Os resultados indicam que todas as variáveis analisadas (floresta, gado, lavoura e madeira) possuem uma associação estatisticamente significativa com o desmatamento. Dentre elas, gado e lavoura se destacam como os fatores com maior influência positiva no desmatamento. Além disso, os baixos valores de VIF sugerem ausência de multicolinearidade entre as variáveis explicativas.

Tabela 2.1: Matriz de correlação

	Desmatamento	Floresta	Gado	Lavoura	Madeira
Desmatamento	1	0,44	0,59	0,09	0,50
Floresta	0,44	1	0,13	-0,01	0,23
Gado	0,59	0,13	1	0,05	0,21
Lavoura	0,09	-0,01	0,05	1	0,03
Madeira	0,50	0,23	0,21	0,03	1

Tabela 2.2: Ajuste do modelo de regressão

Variável	Estimativa	Valor t	p-valor	VIF
Intercepto	-5.038	-3.312	0.00099	–
Floresta	0.000875	10.824	$< 2e^{-16}$	1.064
Gado	0.000115	16.801	$< 2e^{-16}$	1.056
Lavoura	0.003369	2.164	0.0309	1.003
Madeira	0.000216	11.262	$< 2e^{-16}$	1.098

Tabela 2.3: Descrição das variáveis com suas respectivas unidades de medida e fontes de informação, referentes ao período de 2002 a 2022.

Variável	Descrição	Fonte
<b>Variável resposta</b>	Área desmatada por município ( $\text{km}^2$ ), padronizada pela área de cada município.	Terrabrasilis <a href="#">INPE (2023)</a>
	Área de vegetação nativa remanescente ( $\text{km}^2$ ), padronizada pela área de cada município.	<a href="#">MapBiomas (2023)</a>
<b>Variáveis preditoras selecionadas</b>	Número de cabeças bovinas (Und.), padronizada pela área de cada município.	<a href="#">IBGE (2023)</a>
	Madeira extraída ( $\text{m}^3$ )	<a href="#">IBGE (2023)</a>
	Estimativa populacional	
	Crédito Rural (Cada R\$ 1000) padronizada pela área de cada município.	<a href="#">BCB (2023)</a>
<b>Variáveis preditoras não selecionadas</b>	Culturas temporárias ( $\text{km}^2$ ), padronizada pela área de cada município.	<a href="#">IBGE (2023)</a>
	PIB per capita (Cada R\$ 1000) padronizada pela área de cada	<a href="#">IBGE (2023)</a>

A obtenção desses dados envolveu a consulta a diversas fontes, como MapBiomas Brasil, Banco Central do Brasil (BCB) e Instituto Brasileiro de Geografia e Estatística (IBGE). Coletar e consolidar essas informações foi um processo desafiador, dada a necessidade de compatibilizar diferentes formatos, períodos e escalas espaciais, garantindo a qualidade e a coerência das variáveis utilizadas no modelo.

Além disso, os dados da variável resposta foram obtidos do Instituto Nacional de Pesquisa Espacial (INPE) por meio do programa PRODES Amazônia, que utiliza imagens do satélite Landsat, lançado desde 1972 pela NASA e pelo Serviço Geológico dos Estados Unidos (USGS). Essas imagens são utilizadas para detectar e quantificar áreas de desmatamento tipo corte raso superiores a 6,25 hectares, oferecendo alta resolução espacial para o monitoramento contínuo da superfície terrestre. No entanto, é importante destacar que existem outras fontes que apresentam estimativas divergentes do desmatamento, possivelmente devido a variações na metodologia de coleta, na resolução espacial das imagens, nos critérios de classificação e em outros fatores. Contudo, a ausência de informações detalhadas sobre essas discrepâncias impede uma

análise conclusiva sobre suas causas.

Em uma primeira análise, desconsiderou-se a natureza temporal dos dados e eles foram analisados para cada ano separadamente. Em seguida, ajustou-se o modelo GWR para a proporção de área desmatada, conforme utilizado por [Santos, Assunção e Monteiro \(2021\)](#) na modelagem do desmatamento. Esse modelo permite que os coeficientes de regressão, em vez de serem fixos para todas as localizações, variem de acordo com elas. Seja  $y_i$  a proporção de área desmatada do município  $i = 1, \dots, N$ . Então, a equação de regressão adotada é a seguinte:

$$y_i = \beta_0(s_i) + \sum_{j=1}^{p-1} \beta_j(s_i)x_{ij} + \varepsilon_i, \quad (2.1)$$

em que

- $\beta_j(s_i)$  é o coeficiente de regressão da  $j$ -ésima variável no ponto  $i$ ,
- $\varepsilon_i$  é o termo de erro.

Assim, empregando a equação (2.1) é possível estimar os coeficientes da seguinte forma:

$$\hat{\beta}(s_i) = (X'W(s_i)X)^{-1}X'W(s_i)\mathbf{y}, \quad (2.2)$$

em que  $\mathbf{y} = (y_1, y_2, \dots, y_N)'$  é um vetor  $(N \times 1)$  de observações da variável dependente obtida em cada um dos  $N$  locais,  $X$  é a matriz de desenho de dimensão  $(N \times p)$ ,  $\hat{\beta}(s_i)$  é um vetor  $(p \times 1)$ , e  $W(s_i)$  é uma matriz diagonal  $(N \times N)$  cuja diagonal principal contém os pesos que são obtidos por meio de uma função *kernel*. Essa matriz diagonal utiliza a distância  $d_{ij}$  entre dois localidades e um parâmetro que define a largura de banda que está associada ao grau de dependência espacial entre os  $\beta_j(s_i)$ . A largura de banda é um valor específico que define a extensão da interação espacial entre localizações. Mais detalhes do modelo em [Brunsdon, Fotheringham e Charlton \(1998\)](#).

Ao analisar os dados do desmatamento para cada ano separadamente, desconsiderar-se seu aspecto temporal. Levar em conta essa característica é fundamental para compreender melhor as dinâmicas do desmatamento. Nesse contexto, [Fotheringham, Crespo e Yao \(2015\)](#) propuseram o modelo GTWR, o qual é uma extensão do modelo GWR que incorpora a variação espacial e temporal dos coeficientes. Assim, o modelo

pode ser representado pela seguinte equação

$$y_i = \beta_0(s_i, t_i) + \sum_{j=1}^{p-1} \beta_j(s_i, t_i)x_{ij} + \varepsilon_i, \quad (2.3)$$

A estimativa dos coeficientes pode ser obtida por meio da seguinte expressão

$$\hat{\beta}(s_i, t_i) = (X'W(s_i, t_i)X)^{-1}X'W(s_i, t_i)\mathbf{y}. \quad (2.4)$$

A principal diferença em relação ao GWR está na construção da matriz de pesos, que considera simultaneamente a distância espacial e a distância temporal entre as localizações.

Com base nos modelos apresentados acima, foram obtidas as estimativas do intercepto e dos coeficientes das variáveis vegetação nativa remanescente, cabeças bovinas, lavoura e madeira extraída, para o período de 2002 a 2022. Essas estimativas foram representadas por meio de mapas coropléticos, apresentados nas Figuras 2.1 e 2.2, correspondentes aos modelos GWR e GTWR, respectivamente. Para esta análise preliminar, apresenta-se os mapas apenas para os anos de 2003, 2009, 2015 e 2022.

A Figura 2.1 apresenta os coeficientes locais estimados para cada covariável ao longo do tempo, com base no modelo GWR. Observa-se que, na maioria dos municípios, os coeficientes da variável vegetação nativa remanescente são negativos, enquanto em outros, com coloração branca, os coeficientes aparecem estar próximos de zero, o que pode indicar uma influência local reduzida ou nula. Em 2003, a variável número de cabeças bovinas mostra grande variabilidade espacial nos coeficientes, com valores distintos entre os municípios; nos anos seguintes, essa variabilidade parece diminuir. No caso da madeira extraída, nota-se que em 2003 alguns municípios apresentam coeficientes relativamente altos, mas ao longo do tempo esses valores tendem a diminuir, sugerindo um possível enfraquecimento do efeito espacial dessa atividade. Por fim, no caso da lavoura temporária, observa-se que diversos municípios do estado de Mato Grosso apresentam coeficientes positivos, o que pode sugerir uma associação local positiva com o desmatamento nesse território.

A Figura 2.2 apresenta os mapas dos coeficientes locais estimados pelo modelo GTWR ao longo do tempo. Observa-se que os padrões espaciais obtidos são bastante semelhantes aos apresentados na Figura 2.1. Apesar de o GTWR incorporar a dimensão temporal de forma mais explícita, as diferenças entre os dois modelos não são muito marcantes. Isso pode indicar que, no contexto analisado, a variabilidade espacial tem um papel mais relevante do que a variação temporal na associação entre as

covariáveis e o desmatamento.

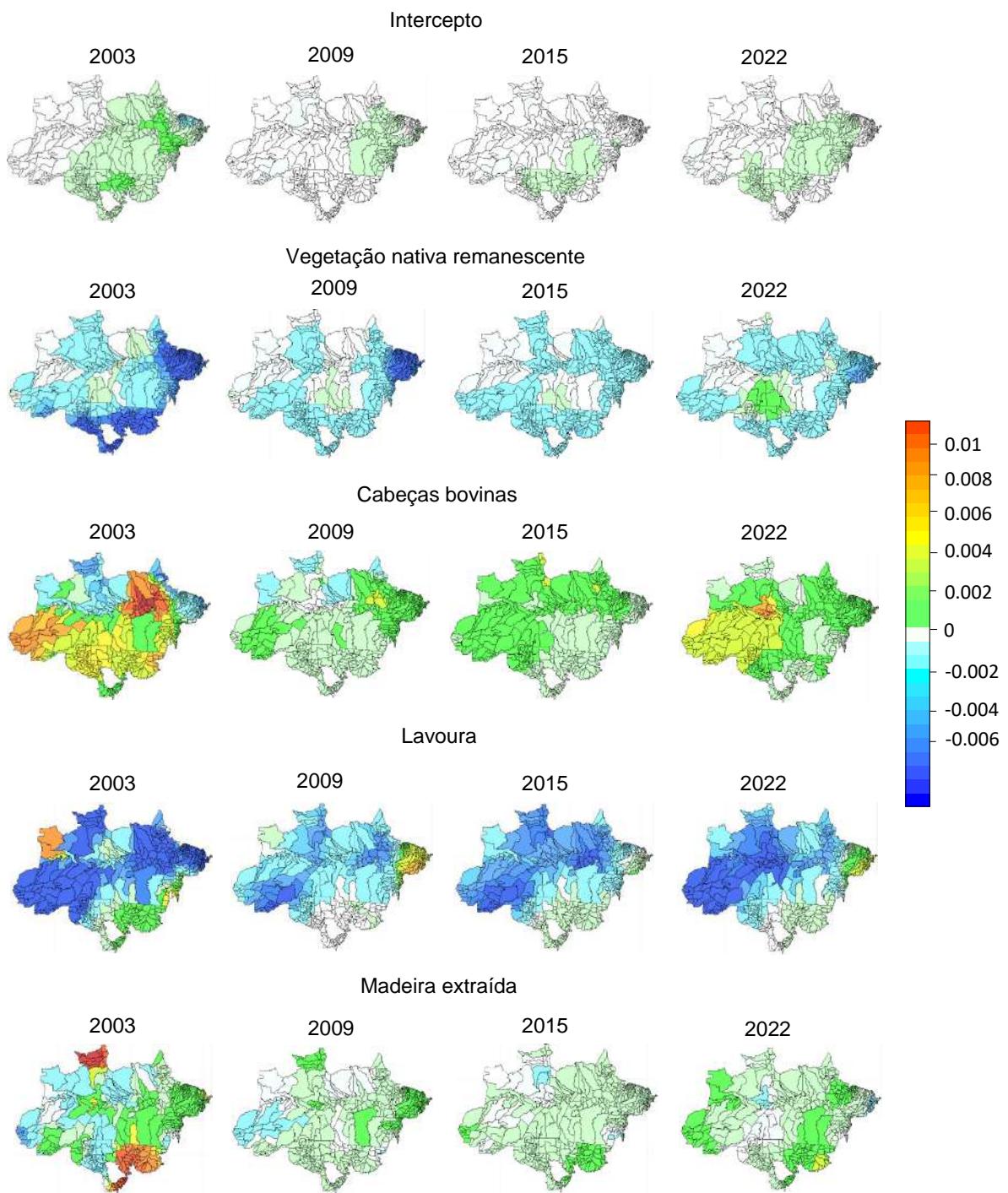


Figura 2.1: Coeficientes estimados através do modelo GWR nos anos 2003, 2009, 2015 e 2022.

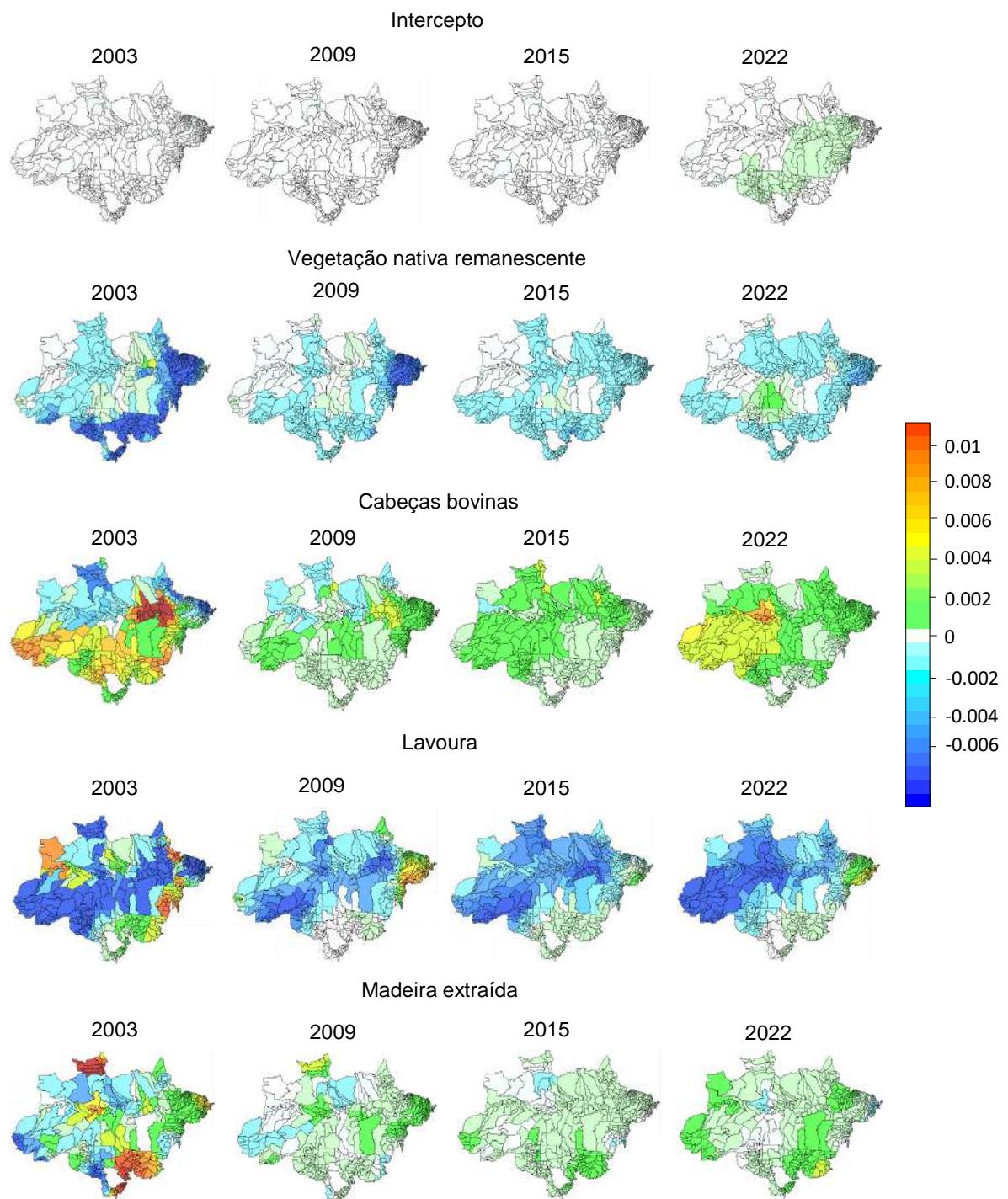


Figura 2.2: Coeficientes estimados através do modelo GTWR nos anos 2003, 2009, 2015 e 2022.

A Figura 2.3 apresenta as estimativas do intercepto e dos coeficientes das covariáveis obtidas por meio do modelo GWR, utilizando as médias do desmatamento e das variáveis explicativas no período de 2002 a 2022. Observa-se que as variáveis cabeças bovinas e madeira extraída exibem coeficientes positivos na maioria dos municípios. A variável lavoura, por sua vez, apresenta coeficientes predominantemente negativos na maior parte do municípios do bioma; no entanto, em municípios localizados no estado de Mato Grosso e no norte do Pará, os coeficientes são positivos, sugerindo que, nessas regiões, a atividade agrícola pode estar associada a um aumento do desmatamento.

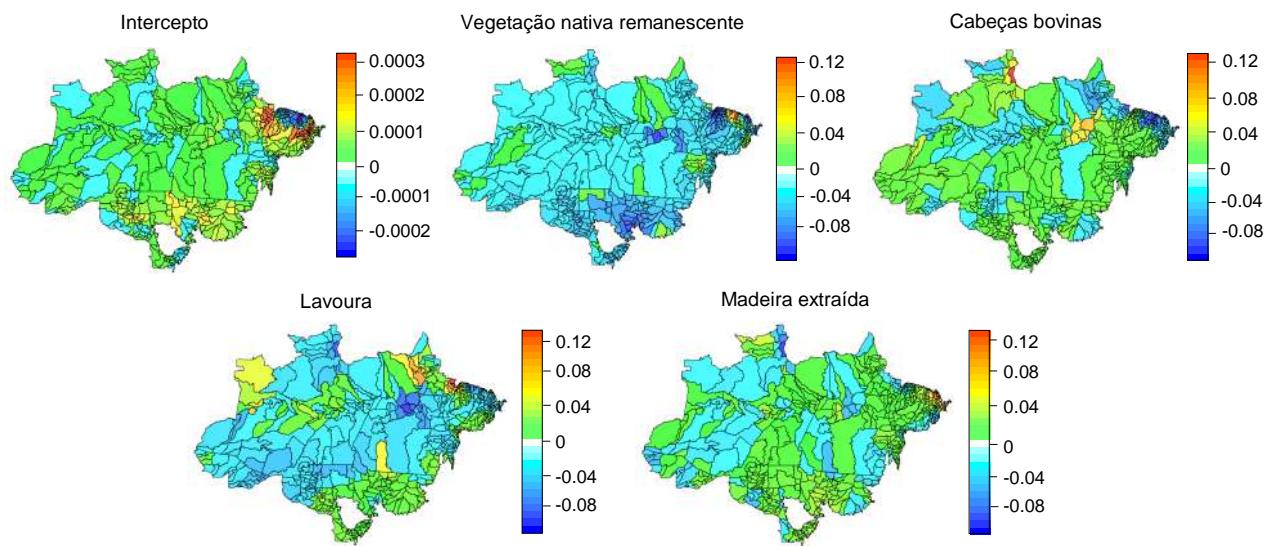


Figura 2.3: Estimativas do intercepto e dos coeficientes das covariáveis, obtidas por meio do modelo GWR, com base nas médias do desmatamento e das variáveis explicativas entre 2002 e 2022.

# Capítulo 3

## Modelo de regressão espaço-temporal com coeficientes variando no espaço e no tempo

Modelos que incorporam variação espacial e temporal nos coeficientes têm ganhado destaque na estatística espacial e espaço-temporal, pois permitem capturar dinâmicas locais e interanuais de fenômenos complexos, ajustando as relações entre variáveis a diferentes contextos geográficos e temporais. Diversos autores têm contribuído para o desenvolvimento dessa abordagem. [Paez, Gamerman e Landim \(2007\)](#) propõem modelos lineares flexíveis para dados em tempo discreto e espaço contínuo, com coeficientes de regressão variando suavemente ao longo do tempo e do espaço, aplicados à modelagem da poluição no Nordeste dos Estados Unidos. Similarmente, [Gelfand, Kim e Sirmans \(2003\)](#) apresentam um modelo de regressão espacial com coeficientes variando localmente, aplicado à modelagem dos preços de imóveis, estendendo a abordagem para modelos espaço-temporais e lineares generalizados.

Avançando nessa linha, [Gelfand, Banerjee e Gamerman \(2005\)](#) propuseram um modelo dinâmico espaço-temporal univariado, no qual os coeficientes variam tanto espacial quanto temporalmente. A variação espacial capta diferenças locais na relação entre a variável resposta e os preditores, enquanto a variação temporal permite detectar tendências e mudanças interanuais.

Portanto, neste capítulo, é adotada a mesma estrutura do modelo proposto por [Gelfand, Banerjee e Gamerman \(2005\)](#) para a modelagem da proporção de área desmatada, com uma variação na componente espaço-temporal. Especificamente, devido à natureza dos dados, que são dados de área, a componente responsável pela variabili-

dade espacial é modelada por um modelo autorregressivo condicional (CAR), conforme proposto por [Besag \(1974\)](#), ao invés de um processo gaussiano, conforme foi assumido por [Gelfand, Banerjee e Gamerman \(2005\)](#).

### 3.1 Modelo proposto

Considere um conjunto de índices  $s = \{s_1, \dots, s_N\}$  para localizações e  $t = \{1, \dots, T\}$  para os instantes de tempo, para os quais observa-se uma proporção de desmatamento  $Y(s, t)$  e um vetor de covariáveis  $\mathbf{x}(s, t)$ ,  $p$ -dimensional, associados a um vetor de coeficientes de regressão  $\tilde{\beta}(s, t)$ .

A resposta  $Y(s, t)$ , é modelada como uma variável que inclui um ruído aleatório,  $\varepsilon(s, t)$ , que representam perturbações gaussianas centradas em zero e não estão correlacionadas no tempo quanto no espaço. O vetor  $\tilde{\beta}(s, t)$ , de dimensão  $(p \times 1)$  decompõe-se em componente puramente temporal  $\beta_t$  e componente espaço-temporal  $\beta(s, t)$ . Podemos expressar a estrutura de modelagem espaço-temporal da seguinte forma:

$$\begin{aligned} Y(s, t) &= \mu(s, t) + \varepsilon(s, t), \quad \varepsilon(s, t) \stackrel{\text{ind}}{\sim} N(0, \sigma_\varepsilon^2), \\ \mu(s, t) &= \mathbf{x}'(s, t) \tilde{\beta}(s, t), \\ \tilde{\beta}(s, t) &= \beta_t + \beta(s, t), \\ \beta_t &= \beta_{t-1} + \boldsymbol{\eta}_t, \quad \boldsymbol{\eta}_t \stackrel{\text{ind}}{\sim} N(\mathbf{0}, \sigma_\eta^2 I_p), \\ \beta(s, t) &= \beta(s, t-1) + \mathbf{w}(s, t), \end{aligned} \tag{3.1}$$

em que  $\mathbf{w}(s, t)$  é a  $s$ -ésima coluna da matriz  $W_t = [\mathbf{w}(s_1, t), \dots, \mathbf{w}(s_N, t)]_{p \times N}$ , cujos elementos são assumidos independentes com variância  $\sigma_w^2$ . Cada linha da matriz  $W_t$  segue um modelo CAR, diferentemente do processo gaussiano adotado por [Gelfand, Banerjee e Gamerman \(2005\)](#).

A partir de estudos realizados com dados artificiais, não foi possível estimar  $\beta(s, t)$  de forma satisfatória. Por esse motivo, optou-se por decompor o vetor  $\tilde{\beta}(s, t)$  em uma componente puramente temporal e uma componente puramente espacial. Assim,  $\beta(s, t) = \beta(s)$ , em que  $\beta(s)$  é a  $s$ -ésima coluna da matriz

$$B = [\beta(s_1), \dots, \beta(s_N)]_{p \times N},$$

considerando os elementos de  $\beta(s)$  independentes, com variância  $\sigma_\beta^2$ , e cada linha da matriz  $B$ , denotada por  $\beta_{(j)} = (\beta_j(s_1), \dots, \beta_j(s_N))$  para cada  $j = 1, \dots, p$ , segue um

modelo CAR com parâmetro  $\alpha$ , isto é,

$$\boldsymbol{\beta}_{(j)} \sim N(0, [D_v - \alpha V]^{-1} \sigma_\beta^2),$$

em que

- $V$  é a matriz de vizinhança ( $N \times N$ ) com valor 1 quando duas localidades  $s_i$  e  $s_k$  são vizinhas, ou seja, compartilham uma fronteira em comum, e 0 quando não são.
- $D_v$  é a matriz diagonal cujos elementos  $(D_v)_{ii} = \sum_{k=1}^N V_{ik}$  correspondem ao número de vizinhos da localização  $i$ .

O modelo CAR é caracterizado por sua abordagem condicional e autoregressiva, na qual a distribuição de cada localidade depende das localidades vizinhas. Isso significa que, ao modelar a variável de interesse em uma região, o modelo considera as informações das áreas ao seu redor, criando uma estrutura de dependência entre as localizações. Essa dependência é definida por uma matriz de vizinhança, que determina como cada localidade influencia as outras. Por essa característica, o modelo é especialmente adequado para analisar dados de área.

A dependência espacial é controlada por um parâmetro  $\alpha$ , que regula a intensidade de dependência entre as localizações. Para que o modelo seja considerado próprio, é necessário que a matriz de precisão  $\Sigma_u = (D_v - \alpha V)^{-1}$  seja positiva definida. Isso ocorre quando  $\alpha \in \left(\frac{1}{\lambda(1)}, \frac{1}{\lambda(n)}\right)$ , em que  $\lambda(1)$  e  $\lambda(n)$  são, respectivamente, o menor e o maior autovalor da matriz  $D_v^{-1/2} V D_v^{-1/2}$ .

Um caso particular ocorre quando  $\alpha = 1$ , conhecido como o modelo CAR intrínseco. Esse modelo oferece uma forma simples e robusta de modelar a dependência espacial, assumindo uma influência uniforme entre todas as localizações vizinhas.

Mais detalhes sobre o modelo CAR podem ser vistos em [Banerjee, Bradley e Gelfand \(2014\)](#) e [Schmidt e Nobre \(2014\)](#).

Assim,  $B$  segue uma distribuição normal matriz variada

$$B \sim N_{p \times N}(0, \sigma_\beta^2 I_p, [D_v - \alpha V]^{-1}),$$

em que  $0$  é a matriz de médias ( $p \times N$ ),  $\sigma_\beta^2 I_p$  é a matriz de covariância entre as linhas ( $p \times p$ ) e  $[D_v - \alpha V]^{-1}$  é a matriz de covariância entre as colunas ( $N \times N$ ).

Então pode-se definir

$$\boldsymbol{\beta} = \text{vec}(B) = (\boldsymbol{\beta}(s_1), \dots, \boldsymbol{\beta}(s_N))' \sim N(\mathbf{0}, [D_v - \alpha V]^{-1} \otimes \sigma_\beta^2 I_p),$$

um vetor de dimensão  $(Np \times 1)$ . Vale ressaltar que esse é um caso particular de um modelo CAR multivariado, cujos detalhes podem ser vistos no [Banerjee, Bradley e Gelfand \(2014\)](#) e [Jin, Carlin e Banerjee \(2005\)](#).

O modelo em (3.1) pode ser escrito em sua forma matricial. Para cada instante de tempo  $t = 1, \dots, T$ , define-se

- $\mathbf{Y}_t = (Y(s_1, t), \dots, Y(s_N, t))'$  como um vetor  $(N \times 1)$ ,
- $F_t = \text{Diag}(\mathbf{x}'(s_1, t), \dots, \mathbf{x}'(s_N, t))$  como uma matriz bloco diagonal de dimensão  $(N \times Np)$ ,
- $\boldsymbol{\varepsilon}_t = (\varepsilon(s_1, t), \dots, \varepsilon(s_N, t))$  como um vetor de erro de dimensão  $(N \times 1)$ .

Dessa forma, o modelo é representado por:

$$\mathbf{Y}_t = F_t \boldsymbol{\theta}_t + \boldsymbol{\varepsilon}_t, \quad \boldsymbol{\varepsilon}_t \sim N(\mathbf{0}, \sigma_\varepsilon^2 I_N), \quad (3.2a)$$

$$\boldsymbol{\theta}_t = \mathbf{1}_N \otimes \boldsymbol{\beta}_t + \boldsymbol{\beta}, \quad (3.2b)$$

$$\boldsymbol{\beta}_t = \boldsymbol{\beta}_{t-1} + \boldsymbol{\eta}_t, \quad \boldsymbol{\eta}_t \sim N(\mathbf{0}, \sigma_\eta^2 I_p), \quad (3.2c)$$

$$\boldsymbol{\beta} \sim N(\mathbf{0}, [D_v - \alpha V]^{-1} \otimes \sigma_\beta^2 I_p), \quad (3.2d)$$

em que  $\otimes$  representa o produto de Kronecker e  $\mathbf{1}_N$  é o vetor coluna de dimensão  $(N \times 1)$  cujos elementos são todos iguais a 1.

Escrever o modelo na forma matricial é importante para o procedimento de inferência, pois facilita a utilização de métodos eficientes para estimação dos parâmetros. Além disso, esse modelo pode ser interpretado como um modelo linear dinâmico (MLD), permitindo a aplicação de técnicas usuais de inferência desenvolvidas para esse tipo de estrutura. Métodos baseados em filtragem e suavização, como os discutidos em [West e Harrison \(1997\)](#), são particularmente úteis para lidar com a dependência temporal presente nos dados.

## 3.2 Procedimento de inferência

O procedimento de inferência será feito sob o enfoque bayesiano. Nesse contexto, assume-se uma distribuição *a priori* para os parâmetros de interesse do modelo, visando obter a distribuição *a posteriori*, que combina as informações das observações com a distribuição *a priori*.

### Distribuição *a priori*

O vetor paramétrico a ser estimado é definido como

$$\Theta = (\beta_0, \beta_1, \dots, \beta_T, \beta(s_1), \dots, \beta(s_N), \sigma_\varepsilon^2, \sigma_\eta^2, \sigma_\beta^2, \alpha),$$

contendo todas as variáveis desconhecidas do modelo em (3.2) a serem estimadas.

Assim, a distribuição *a priori* assumida para o vetor acima será descrita como

$$p(\Theta) = p(\beta_t \mid \beta_{t-1}, \sigma_\eta^2) p(\beta \mid \alpha, \sigma_\beta^2) p(\beta_0) p(\sigma_\varepsilon^2) p(\sigma_\eta^2) p(\sigma_\beta^2) p(\alpha), \quad (3.3)$$

em que

- Para o vetor  $\beta_0$  será assumida uma distribuição *a priori* normal multivariada, isto é,  $\beta_0 \sim N(\mathbf{m}_0, C_0)$ , em que o vetor de médias  $\mathbf{m}_0$  e a matriz de covariância  $C_0$  são conhecidos.
- Os parâmetros  $\sigma_\varepsilon^2, \sigma_\eta^2$  e  $\sigma_\beta^2$  seguem uma distribuição *a priori* gama inversa, isto é,  $\sigma_\varepsilon^2 \sim IG(a_\varepsilon, b_\varepsilon)$ ,  $\sigma_\eta^2 \sim IG(a_\eta, b_\eta)$ ,  $\sigma_\beta^2 \sim IG(a_\beta, b_\beta)$ , em que  $a_\varepsilon, b_\varepsilon, a_\eta, b_\eta, a_\beta, b_\beta$  são valores conhecidos e o parâmetro  $\alpha$  segue uma distribuição *a priori* uniforme.

### Distribuição *a posteriori*

Apos atribuir uma distribuição *a priori* para o vetor paramétrico  $\Theta$ , é possível obter a distribuição *a posteriori*, que é da forma

$$p(\Theta | y) \propto \left[ \prod_{t=1}^T p(\mathbf{Y}_t | \boldsymbol{\beta}_t, \boldsymbol{\beta}, \sigma_\varepsilon^2) \right] \left[ \prod_{t=1}^T p(\boldsymbol{\beta}_t | \boldsymbol{\beta}_{t-1}, \sigma_\eta^2) \right] p(\boldsymbol{\beta} | \alpha, \sigma_\beta^2) p(\boldsymbol{\beta}_0) p(\sigma_\varepsilon^2) p(\sigma_\eta^2) p(\sigma_\beta^2) p(\alpha). \quad (3.4a)$$

$$\begin{aligned} & \propto \left[ \prod_{t=1}^T (\sigma_\varepsilon^{2N})^{-1/2} \exp \left\{ -\frac{1}{2\sigma_\varepsilon^2} (\mathbf{Y}_t - F_t(\mathbf{1}_N \otimes \boldsymbol{\beta}_t + \boldsymbol{\beta}))' \right. \right. \\ & \quad \left. \left. (\mathbf{Y}_t - F_t(\mathbf{1}_N \otimes \boldsymbol{\beta}_t + \boldsymbol{\beta})) \right\} \right] \\ & \quad \times \prod_{t=1}^T (\sigma_\eta^{2p})^{-1/2} \exp \left\{ -\frac{1}{2\sigma_\eta^2} (\boldsymbol{\beta}_t - \boldsymbol{\beta}_{t-1})' (\boldsymbol{\beta}_t - \boldsymbol{\beta}_{t-1}) \right\} \\ & \quad \times (\sigma_\beta^{2Np})^{-1/2} \exp \left\{ -\frac{1}{2} (\boldsymbol{\beta})' \left[ (D_v - \alpha V)^{-1} \otimes \sigma_\beta^2 I_p \right]^{-1} \boldsymbol{\beta} \right\} \\ & \quad \times \exp \left\{ -\frac{1}{2C_0} (\boldsymbol{\beta}_0 - \mathbf{m}_0)' (\boldsymbol{\beta}_0 - \mathbf{m}_0) \right\} \times (\sigma_\varepsilon^2)^{-(a_\varepsilon+1)} \exp \left\{ -\frac{b_\varepsilon}{\sigma_\varepsilon^2} \right\} \\ & \quad \times (\sigma_\eta^2)^{-(a_\eta+1)} \exp \left\{ \frac{-b_\eta}{\sigma_\eta^2} \right\} \times (\sigma_\beta^2)^{-(a_\beta+1)} \exp \left\{ \frac{-b_\beta}{\sigma_\beta^2} \right\}. \end{aligned} \quad (3.4b)$$

A equação acima não é analiticamente conhecida. Para possibilitar a execução do procedimento de inferência e a obtenção de amostras dos parâmetros a partir de sua distribuição *a posteriori*, torna-se essencial a aplicação de métodos de simulação estocástica. Assim, serão empregados os Métodos de Monte Carlo via Cadeias de Markov (MCMC, na sigla em inglês), especificamente, o amostrador de Gibbs. Mais detalhes sobre métodos MCMC podem ser vistos em [Gamerman e Lopes \(2006\)](#).

Para a implementação do amostrador de Gibbs, é necessário, inicialmente, obter as distribuições condicionais completas dos elementos do vetor paramétrico  $\Theta$ . Em particular, para o parâmetro  $\boldsymbol{\beta}_t$ , com  $t = 1, \dots, T$ , utiliza-se o algoritmo *Forward Filtering Backward Sampling* (FFBS) para sortear esses parâmetros a partir de sua distribuição condicional conjunta. A seguir, são descritos os detalhes da implementação do FFBS para  $\boldsymbol{\beta}_t$ , assim como as distribuições condicionais completas dos demais parâmetros do modelo.

### 3.2.1 Sorteando $\beta_t$

O algoritmo FFBS, proposto por [Carter e Kohn \(1994\)](#) e [Frühwirth-Schnatter \(1994\)](#), é uma maneira eficiente de sortear os parâmetros de estado de um MLD, em todos os instantes de tempo, a partir de sua distribuição conjunta, condicionada aos demais parâmetros e a todos os dados observados.

Ao considerar-se a forma matricial do modelo em (3.1), pode-se perceber que ele pertence à classe dos modelos lineares dinâmicos (MLD). Nesse caso, o FFBS será utilizado para amostrar  $\beta_t$  em seus respectivos passos do amostrador de Gibbs.

Considere o vetor  $\beta_t$  e seja  $D_t = \{\mathbf{Y}_1, \dots, \mathbf{Y}_t\}$  a informação disponível até o tempo  $t$ . O método FFBS é composto por duas etapas: inicialmente, realiza-se a filtragem para frente (*Forward Filtering*), utilizando o Filtro de Kalman para calcular as distribuições condicionais dos estados dado  $D_t$  e os demais parâmetros. Posteriormente, aplica-se um processo recursivo pra trás (*Backwards Sampling*), que sorteia os estados desde  $t = T$  até  $t = 1$ . A seguir, se apresentam os detalhes das duas etapas.

#### Filtragem

Considerando o MLD especificado em (3.2a) e (3.2b), e partindo da distribuição

$$\beta_{t-1} \mid D_{t-1} \sim N(\mathbf{m}_{t-1}, C_{t-1}),$$

com vetor de médias  $\mathbf{m}_{t-1}$  e matriz de covariância  $C_{t-1}$ , temos:

(i) *Distribuição a priori em t:*

$$\beta_t \mid D_{t-1} \sim N(\mathbf{a}_t, R_t),$$

em que

$$\begin{aligned} \mathbf{a}_t &= \mathbf{m}_{t-1}, \\ R_t &= C_{t-1} + \sigma_\eta^2 I_p. \end{aligned}$$

(ii) *Distribuição preditiva um passo a frente:*

$$\mathbf{Y}_t \mid D_{t-1} \sim N(\mathbf{f}_t, Q_t),$$

em que

$$\begin{aligned}\mathbf{f}_t &= F_t(1_N \otimes \mathbf{a}_t + \boldsymbol{\beta}), \\ Q_t &= F_t A R_t A' F_t' + \sigma_e^2 I_N, \\ A &= 1_N \otimes I_p.\end{aligned}$$

(iii) *Distribuição a posteriori* em  $t$ :

$$\boldsymbol{\beta}_t \mid D_t \sim N(\mathbf{m}_t, C_t),$$

em que

$$\begin{aligned}\mathbf{m}_t &= \mathbf{a}_t + R_t(F_t A)' Q_t^{-1}(\mathbf{Y}_t - \mathbf{f}_t), \\ C_t &= R_t - R_t(F_t A)' Q_t^{-1}(F_t A) R_t.\end{aligned}$$

### Suavização

O passo *Backwards Sampling* é baseado na decomposição da distribuição *a posteriori* conjunta dos parâmetros de estado da forma

$$p(\boldsymbol{\beta}_1, \dots, \boldsymbol{\beta}_T \mid D_T) = p(\boldsymbol{\beta}_T \mid D_T) \prod_{t=1}^{T-1} p(\boldsymbol{\beta}_t \mid \boldsymbol{\beta}_{t+1}, D_t),$$

pelo teorema de Bayes, para  $t = T-1, \dots, 1$ , pode ser mostrado que

$$p(\boldsymbol{\beta}_t \mid \boldsymbol{\beta}_{t+1}, D_t) \propto p(\boldsymbol{\beta}_t \mid D_t)$$

segue uma distribuição normal com média

$$\mathbf{h}_t = \mathbf{m}_t + C_t R_{t+1}^{-1}(\boldsymbol{\beta}_{t+1} - \mathbf{a}_{t+1})$$

e variância

$$H_t = C_t - C_t R_{t+1}^{-1} C_t.$$

É importante destacar que os  $\boldsymbol{\beta}_t$  desta distribuição são sorteados de forma recursiva.

### 3.2.2 Sorteando os demais parâmetros

Nesta subseção podem ser vistas as distribuições condicionais completas dos demais elementos que compõem o vetor paramétrico  $\boldsymbol{\Theta}$ . Para a maioria deles, as distribuições possuem forma conhecida e podem ser derivadas diretamente a partir das condições

do modelo, com exceção do parâmetro  $\alpha$ . A seguir, são apresentadas as formas das distribuições condicionais completas de cada um dos parâmetros de interesse.

### Vetor $\beta_0$

$$p(\beta_0 | \Theta_{-\beta_0}, y) \propto \exp \left\{ -\frac{1}{2} \left[ \beta_0' ([\sigma_\eta^2 I_p]^{-1} + C_0^{-1}) \beta_0 - 2\beta_0' ([\sigma_\eta^2 I_p]^{-1} \beta_1 + C_0^{-1} \mathbf{m}_0) \right] \right\},$$

o que implica  $\beta_0 | \Theta_{-\beta_0}, y \sim N(\mathbf{m}_{\beta_0}^*, C_{\beta_0}^*)$ , em que

$$\begin{cases} C_{\beta_0}^* = ([\sigma_\eta^2 I_p]^{-1} + C_0^{-1})^{-1}, \\ \mathbf{m}_{\beta_0}^* = C_{\beta_0}^* ([\sigma_\eta^2 I_p]^{-1} \beta_1 + C_0^{-1} \mathbf{m}_0). \end{cases}$$

### Parâmetro $\sigma_\eta^2$

A distribuição condicional completa da componente da variância  $\sigma_\eta^2$  é da forma

$$\sigma_\eta^2 | \Theta_{-\sigma_\eta^2}, y \sim IG \left( p \frac{T}{2} + a_\eta, \frac{1}{2} \sum_{t=1}^T (\beta_t - \beta_{t-1})' (\beta_t - \beta_{t-1}) + b_\eta \right).$$

### Parâmetro $\sigma_\varepsilon^2$

A distribuição condicional completa da componente da variância  $\sigma_\varepsilon^2$  é da forma

$$\sigma_\varepsilon^2 | \Theta_{-\sigma_\varepsilon^2}, y \sim IG \left( \frac{NT}{2} + a_\varepsilon, \frac{1}{2} \sum_{t=1}^T \left( \mathbf{Y}_t - F_t(\mathbf{1}_N \otimes \beta_t + \beta) \right)' \left( \mathbf{Y}_t - F_t(\mathbf{1}_N \otimes \beta_t + \beta) \right) + b_\varepsilon \right).$$

### Parâmetro $\sigma_\beta^2$

A distribuição condicional completa do  $\sigma_\beta^2$  é da forma

$$\sigma_\beta^2 | \Theta_{-\sigma_\beta^2}, y \sim IG \left( \frac{Np}{2} + a_\beta, \frac{1}{2} \beta' [(D_v - \alpha V) \otimes I_p] \beta + b_\beta \right).$$

### Parâmetro $\alpha$

$$p(\alpha | \Theta_{-\alpha}, y) \propto |D_v - \alpha V|^{p/2} \exp \left\{ -\frac{1}{2} \beta' [(D_v - \alpha V) \otimes (\sigma_\beta^2 I_p)^{-1}] \beta \right\}.$$

Observa-se que, neste caso, a distribuição condicional não possui forma conhecida. Assim, será utilizado o algoritmo de Metropolis-Hastings com uma distribuição proposta

normal truncada, permitindo a obtenção de amostras de  $\alpha$  de forma adequada. Essa abordagem é necessária porque o cálculo direto da distribuição condicional não é viável.

### Vetor $\beta$

$$p(\beta | \Theta_{-\beta}, y) \propto \exp \left\{ \frac{1}{2} \left[ \beta' \left( \sum_{t=1}^T F_t' (\sigma_\varepsilon^2)^{-1} F_t + [(D_v - \alpha V) \otimes (\sigma_\beta^2 I_p)^{-1}] \right) \beta - 2\beta' \left( \sum_{t=1}^T F_t' (\sigma_\varepsilon^2)^{-1} Y_t - \sum_{t=1}^T F_t' (\sigma_\varepsilon^2)^{-1} F_t (\mathbb{1}_N \otimes \beta_t) \right) \right] \right\},$$

o que implica  $\beta | \Theta_{-\beta}, y \sim N(\mathbf{m}_\beta^*, C_\beta^*)$ , em que

$$\begin{cases} C_{\beta_s}^* = \left( \sum_{t=1}^T F_t' (\sigma_\varepsilon^2)^{-1} F_t + [(D_v - \alpha V) \otimes (\sigma_\beta^2 I_p)^{-1}] \right)^{-1}, \\ \mathbf{m}_\beta^* = C_\beta^* \left( \sum_{t=1}^T F_t' (\sigma_\varepsilon^2)^{-1} \mathbf{Y}_t - \sum_{t=1}^T F_t' (\sigma_\varepsilon^2)^{-1} F_t (\mathbb{1}_N \otimes \beta_t) \right). \end{cases}$$

Como mencionado anteriormente, em uma primeira abordagem, considerou-se o efeito  $\tilde{\beta}(s, t) = \beta_t + \beta(s, t)$ . Especificamente para a componente espaço temporal, definimos  $\beta_t^* = (\beta(s_1, t), \dots, \beta(s_N, t))'$ , um vetor de dimensão  $(Np \times 1)$ . Assim, essa componente foi formulada da seguinte forma em termos matriciais:

$$\beta_t^* = \beta_{t-1}^* + \mathbf{w}_t, \quad (3.5)$$

sendo

$$\mathbf{w}_t = (\mathbf{w}'(s_1, t), \dots, \mathbf{w}'(s_N, t))' \sim N(\mathbf{0}, [D_v - \alpha V]^{-1} \otimes \sigma_w^2 I_p).$$

Para a estimação dos vetores  $\beta_t$  e  $\beta_t^*$ , empregou-se o algoritmo FFBS. No entanto, essa abordagem não produziu resultados satisfatórios para  $\beta_t^*$ , pois os intervalos de credibilidade obtidos eram excessivamente amplos, refletindo altos níveis de incerteza, o que indicava que essa abordagem não recuperava bem o verdadeiro valor dos parâmetros.

Por esse motivo, optou-se que o efeito  $\tilde{\beta}(s, t)$  esteja composta por uma componente puramente temporal e uma componente puramente espacial, conforme mostrado em (3.2).

### 3.3 Aplicação: dados artificiais

Nesta seção, será apresentado um estudo com dados artificiais gerados a partir do modelo em (3.2). O objetivo dessa simulação é avaliar o desempenho do modelo, verificando sua capacidade de estimar e recuperar corretamente os parâmetros. Para isso, foram simulados dados artificiais considerando  $p = 5$  (intercepto e 4 covariáveis reais), definidas ao longo do período de tempo  $T = 21$  (2002-2022) e das  $N = 144$  localizações correspondendo aos municípios do estado do Pará, conforme destacado na Figura 3.1.

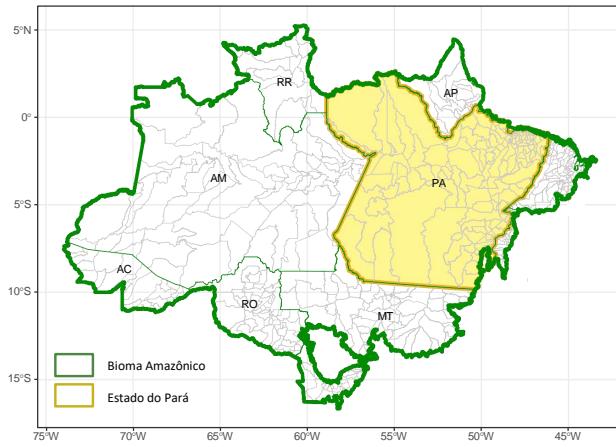


Figura 3.1: Mapa do estado do Pará

Na estimação dos parâmetros do modelo, analisaram-se dois cenários distintos para o parâmetro  $\alpha$ . No primeiro, assumiu-se um modelo CAR com valores de  $\alpha$  restritos ao intervalo  $(0,1)$ , intervalo no qual a distribuição dos efeitos espaciais é própria, conforme determinado pelos autovalores de  $D_v^{-1/2} V D_v^{-1/2}$ . No segundo, considerou-se um modelo CAR intrínseco, com  $\alpha = 1$ .

Nesta aplicação, foram consideradas 12.000 iterações do algoritmo MCMC com aquecimento de 2.000 e espaçamento de 2 iterações entre amostras.

#### 3.3.1 Modelo considerando $0 < \alpha < 1$

Os parâmetros do modelo considerando  $\alpha \in (0, 1)$  foram fixados em  $\sigma_\eta^2 = 0,14$ ,  $\sigma_\varepsilon^2 = 0,20$ ,  $\sigma_\beta^2 = 0,40$  e  $\alpha = 0,80$ . Para o procedimento de estimação, assumiram-se distribuições *a priori* gama inversa com parâmetros de forma 0,1 e de escala 0,3 para  $\sigma_\varepsilon^2$ ,  $\sigma_\beta^2$  e  $\sigma_\eta^2$  e para  $\alpha$  considerou-se uma distribuição *a priori* uniforme  $\alpha \sim U(0, 1)$ . Adicionalmente  $\beta_0 \sim N(0, 100)$ , uma distribuição normal multivariada. Os dados

simulados podem ser vistos na Figura 3.2.

Na Figura 3.3, apresenta-se as estimativas dos coeficientes temporais  $\beta_t$ .  $\beta_{0t}$  refere-se ao intercepto, enquanto os demais representam os coeficientes das variáveis preditoras ao longo do tempo. Os gráficos mostram uma comparação entre as estimativas pontuais, representadas pela média *a posteriori* e os valores verdadeiros simulados. Nota-se que, de forma geral, as estimativas acompanham bem os valores reais, refletindo a capacidade do modelo em capturar as variações temporais.

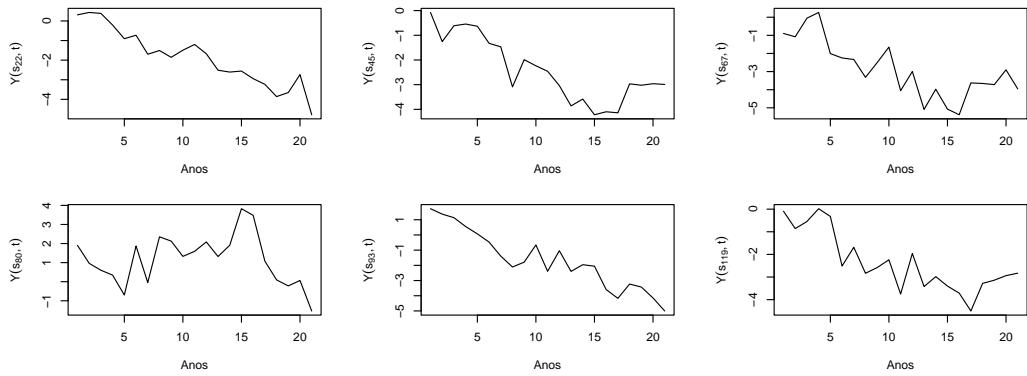


Figura 3.2: Dados artificiais simulado do modelo.

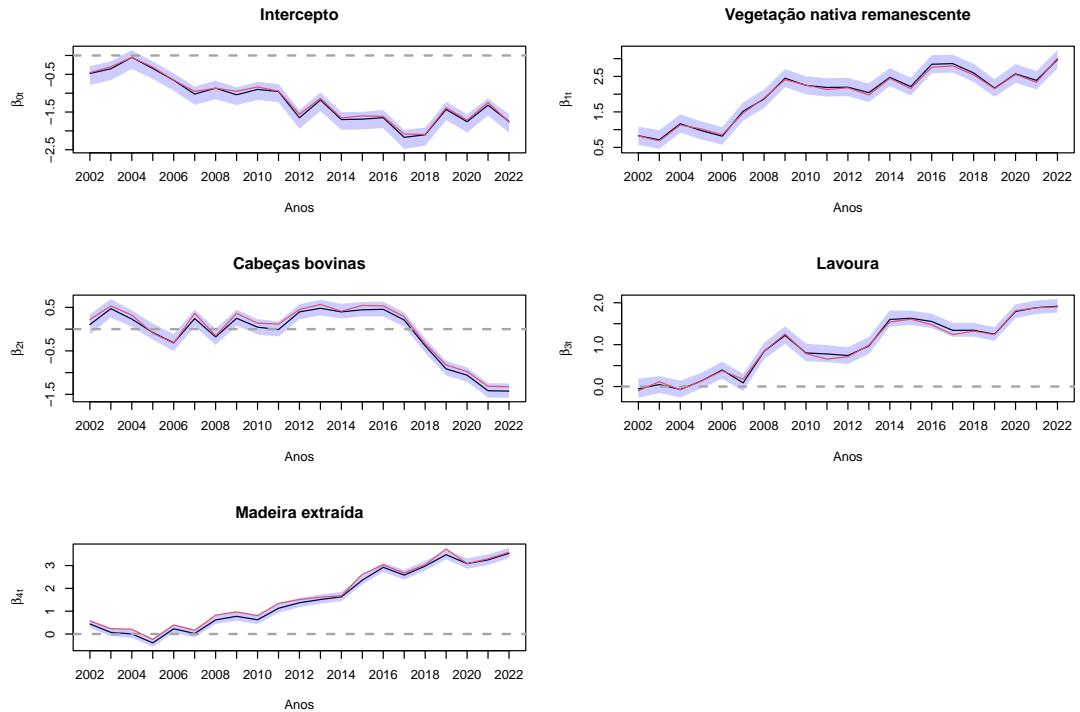


Figura 3.3: Valores estimados (linha preta) e valores verdadeiros (linha vermelha) com respectivos intervalos de 95% de credibilidade *a posteriori* (área hachurada) dos coeficientes temporais ao longo do tempo, para o modelo normal com  $\alpha \in (0, 1)$ .

A Figura 3.4 apresenta os valores verdadeiros, simulados do modelo proposto, e os valores estimados obtidos através da média *a posteriori*, assim como a incerteza associada à estimativa de  $\beta$ , para algumas localizações. Este resultado pode também ser visto ao observar-se a Figura 3.5, que também compara valores verdadeiros e estimados. De maneira geral, o modelo consegue recuperar bem os valores reais, mas observa-se que em alguns municípios as estimativas apresentam discrepâncias maiores. Esse comportamento sugere que certos efeitos são mais difíceis de estimar com precisão. Além disso, esse parâmetro parece ser particularmente desafiador de estimar, o que já havia sido mencionado anteriormente ao se discutir o caso em que foi assumido  $\beta(s, t)$ , no qual não foi possível obter estimativas satisfatórias.

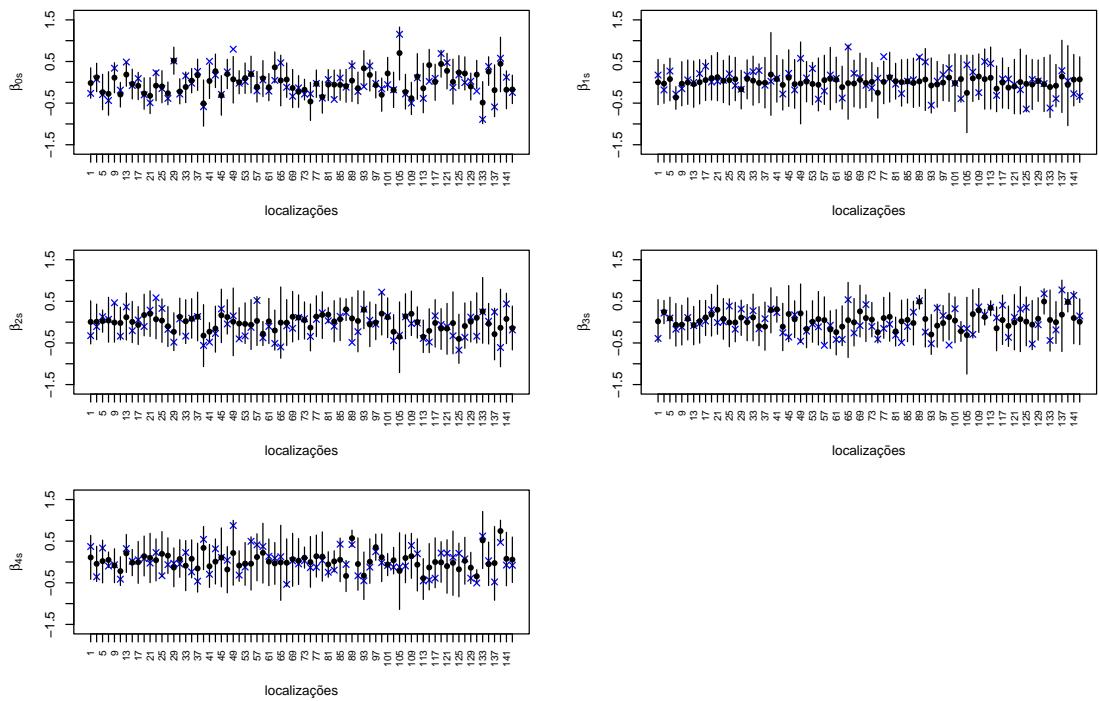


Figura 3.4: Valores estimados (pontos) e valores reais (cruzes) do intercepto e coeficientes espaciais das variáveis preditoras com respectivos intervalos de 95% de credibilidade (linhas verticais) para  $i = 1, \dots, 144$ , no modelo normal com  $\alpha \in (0, 1)$

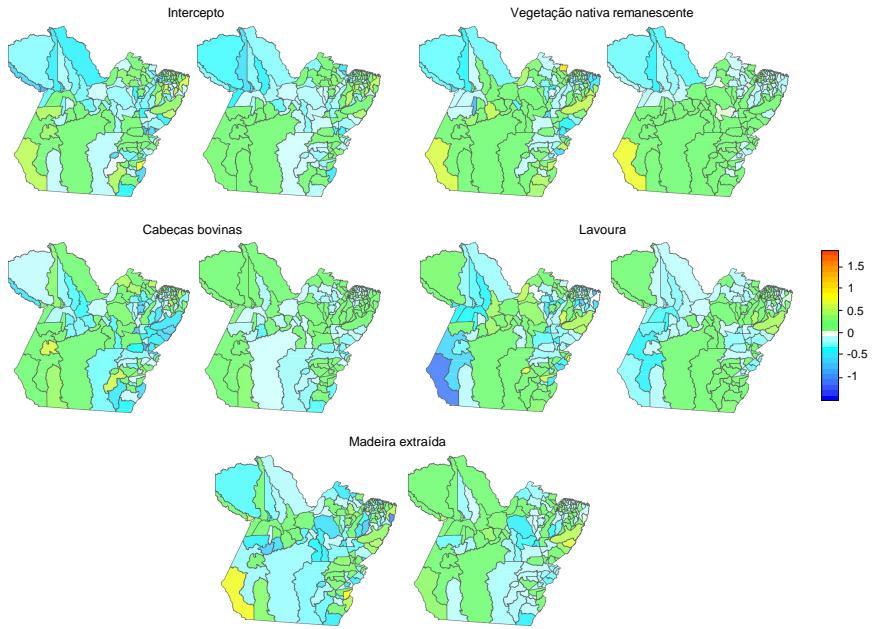


Figura 3.5: Valores verdadeiros e valores estimados do intercepto e dos coeficientes espaciais das variáveis preditoras para todos os municípios, no modelo normal com  $\alpha \in (0, 1)$ .

A Tabela 3.1 apresenta os valores verdadeiros, as médias *a posteriori* e os limites inferior e superior dos intervalos de 95% de credibilidade dos parâmetros do modelo proposto. Observa-se de forma geral que as médias *a posteriori* estão próximos dos valores verdadeiros. Além disso, os intervalos de credibilidade contém os valores verdadeiros. No entanto, percebe-se que o intervalo de credibilidade para  $\alpha$  é relativamente amplo, o que sugere uma maior incerteza na estimação desse parâmetro.

A Figura 3.6 apresenta os traços das cadeias e os histogramas das distribuições *a posteriori* dos parâmetros  $\sigma_\eta^2$ ,  $\sigma_\varepsilon^2$ ,  $\sigma_\beta^2$  e  $\alpha$ . A verificação da convergência das estimativas foi realizada de maneira visual por meio dos traços das cadeias, que permitem avaliar a estabilidade das amostras ao longo das iterações. Já os histogramas fornecem uma visualização da distribuição *a posteriori* desses parâmetros, possibilitando uma análise da incerteza associada às estimativas. Além disso, a linha vermelha indica o valor verdadeiro de cada parâmetro, servindo como referência para avaliar a precisão das estimativas.

Tabela 3.1: Resumo das distribuições *a posteriori* de  $\sigma_\eta^2$ ,  $\sigma_\varepsilon^2$ ,  $\sigma_\beta^2$  e  $\alpha$  no modelo normal.

Parâmetros	Valor verdadeiro	Média <i>a posteriori</i>	Intervalo de credibilidade
$\sigma_\eta^2$	0,14	0,138	(0,078 ; 0,253)
$\sigma_\varepsilon^2$	0,20	0,203	(0,182 ; 0,225)
$\sigma_\beta^2$	0,40	0,383	(0,233 ; 0,585)
$\alpha$	0,80	0,786	(0,303 ; 0,985)

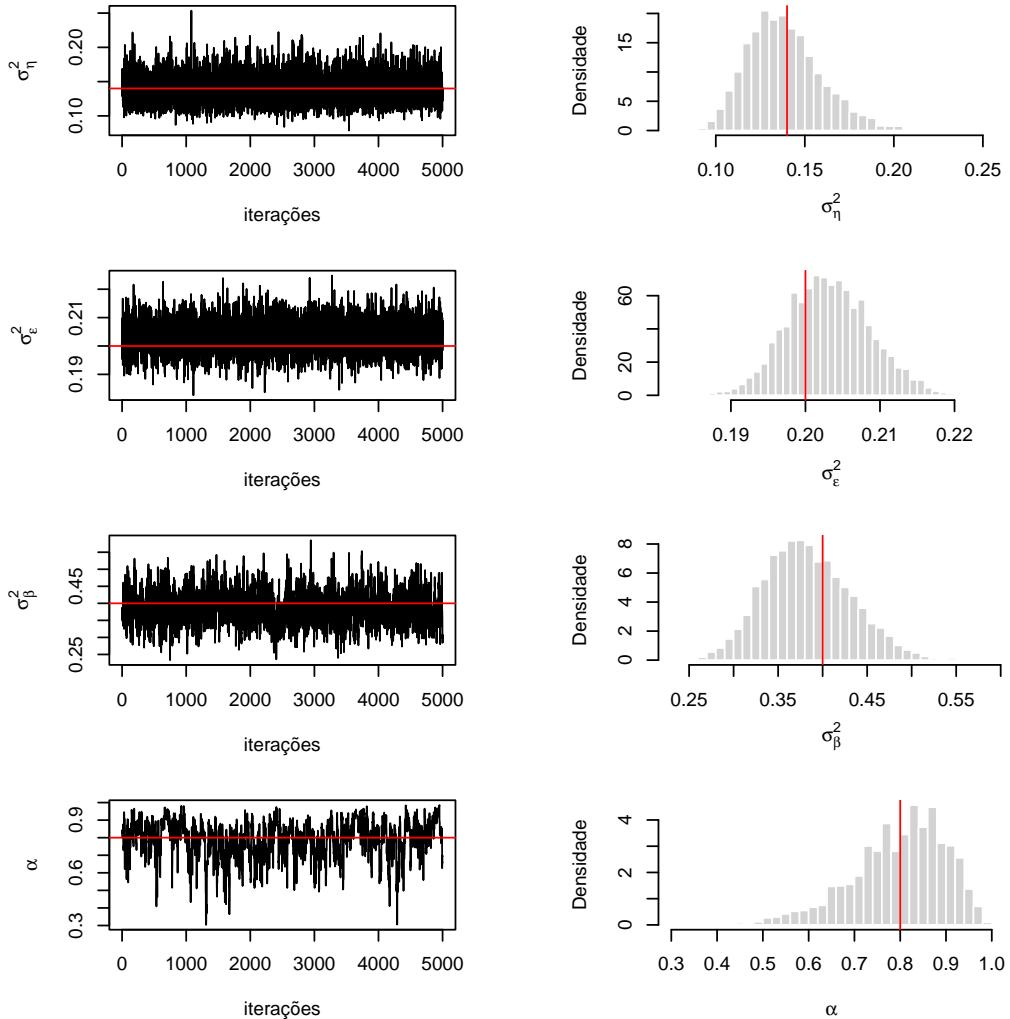


Figura 3.6: Traços das cadeias e histogramas das distribuições *a posteriori* de  $\sigma_\eta^2$ ,  $\sigma_\varepsilon^2$ ,  $\sigma_\beta^2$  e  $\alpha$  do modelo normal. As linhas vermelhas horizontal e vertical indicam o valor verdadeiro.

### 3.3.2 Modelo considerando $\alpha = 1$

Os parâmetros do modelo considerando  $\alpha = 1$  foram fixados em  $\sigma_\eta^2 = 0,16$ ,  $\sigma_\varepsilon^2 = 0,20$  e  $\sigma_\beta^2 = 0,36$ . As distribuições *a priori* são as mesmas do modelo que considera  $\alpha \in (0, 1)$ . Nesse caso, considerou-se a restrição  $\sum_{i=1}^N \beta_j(s_i) = 0$ , em que  $j = 1, \dots, p$ , uma vez que o modelo assume a estrutura CAR intrínseca. Mais detalhes em [Banerjee, Bradley e Gelfand \(2014\)](#) e [Schmidt e Nobre \(2014\)](#).

Os dados simulados podem ser vistos na Figura 3.7. A Figura 3.8 mostra os valores verdadeiros e as estimativas obtidas pela média *a posteriori* do intercepto e dos coeficientes das variáveis preditoras ao longo do tempo. Pode-se ver que, na maioria dos casos, as estimativas estão bem próximas dos valores verdadeiros.

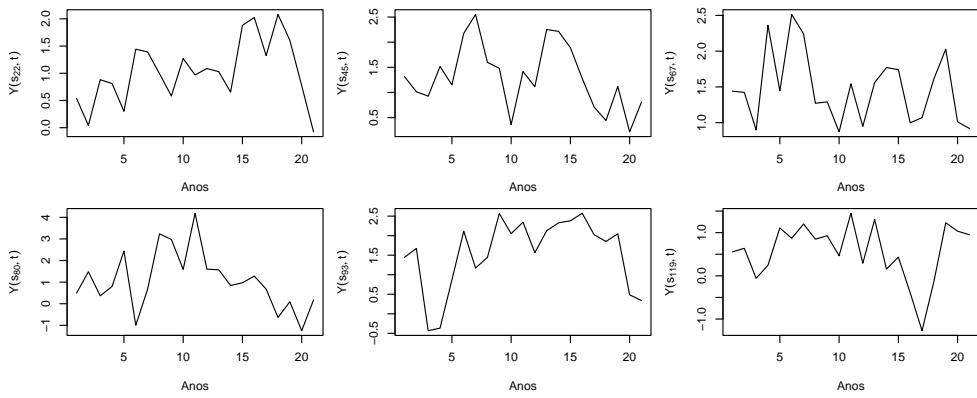


Figura 3.7: Dados artificiais simulado do modelo.

A Figura 3.9 apresenta os valores verdadeiros e estimados do intercepto e dos coeficientes espaciais, juntamente com a incerteza associada à estimativa para alguns municípios na região de interesse. Já a Figura 3.10 apresenta essa mesma informação de forma espacial, por meio de mapas coropléticos que comparam os valores verdadeiros e estimados em cada localidade considerada na simulação. De modo geral, observa-se que os valores estimados se aproximam bem dos valores reais. No entanto, ainda há discrepâncias em alguns municípios, especialmente em áreas ao norte e o sudeste da região analisada, o que indica que o modelo pode ser aprimorado para representar melhor certas variações locais.

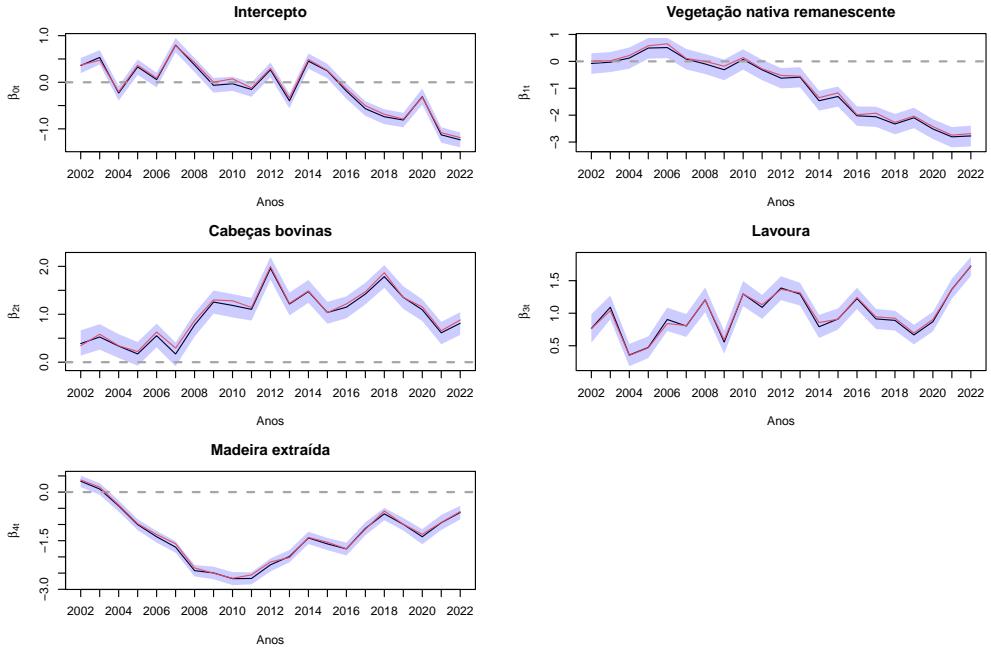


Figura 3.8: Valores estimados através da média *a posteriori* (linha preta) e valores verdadeiros (linha vermelha) com respectivos intervalos de 95% de credibilidade (área hachurada) dos coeficientes temporais ao longo do tempo, para o modelo normal com  $\alpha = 1$ .

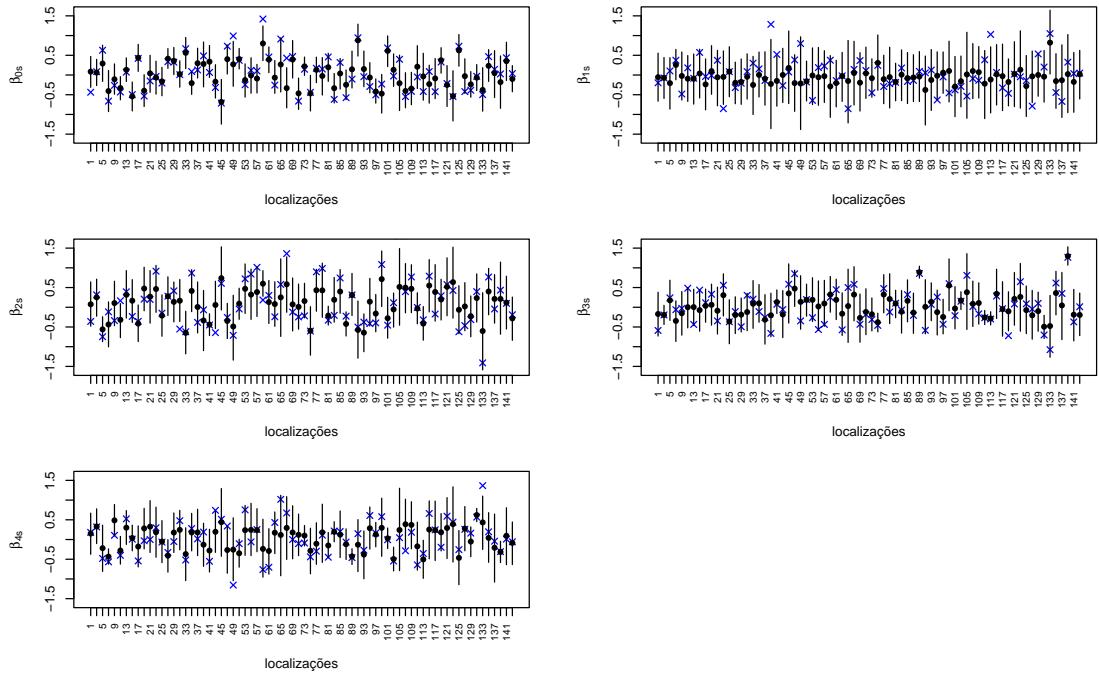


Figura 3.9: Valores estimados (pontos) e valores reais (cruzes) do intercepto e coeficientes espaciais das variáveis preditoras com respectivos intervalos de 95% de credibilidade (linhas verticais) para  $i = 1, \dots, 144$ , no modelo normal com  $\alpha = 1$ .

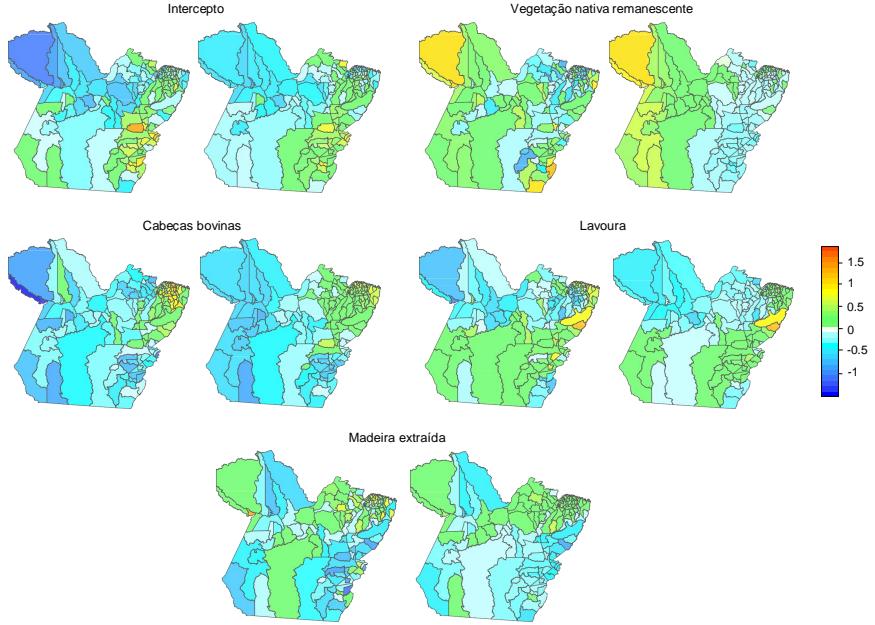


Figura 3.10: Valores verdadeiros simulados e valores estimados através da média *a posteriori* para o intercepto e os coeficientes espaciais das variáveis preditoras para todos os municípios, no modelo normal com  $\alpha = 1$ .

A Tabela 3.2 fornece um resumo das amostras *a posteriori*, incluindo as médias e os intervalos de 95% de credibilidade dos parâmetros do modelo, bem como os respectivos valores verdadeiros.

Tabela 3.2: Resumo das distribuições *a posteriori* de  $\sigma_\eta^2$ ,  $\sigma_\varepsilon^2$ ,  $\sigma_\beta^2$  do modelo normal.

Parâmetros	Valor verdadeiro	Média <i>a posteriori</i>	Intervalo de credibilidade
$\sigma_\eta^2$	0,16	0,170	(0,103 ; 0,312)
$\sigma_\varepsilon^2$	0,20	0,197	(0,177 ; 0,217)
$\sigma_\beta^2$	0,36	0,342	(0,228 ; 0,529)

As cadeias e os histogramas das distribuições *a posteriori* de  $\sigma_\eta^2$ ,  $\sigma_\varepsilon^2$ ,  $\sigma_\beta^2$  podem ser vistos na Figura 3.11. Por meio de uma análise visual desta figura, é possível verificar que as cadeias convergiram e estão próximas do valor verdadeiro do parâmetro utilizado na geração dos dados. Além desses parâmetros, os traços das cadeias de alguns coeficientes temporais  $\beta_t$  e espaciais  $\beta$  podem ser vistas no Apêndice A.1.

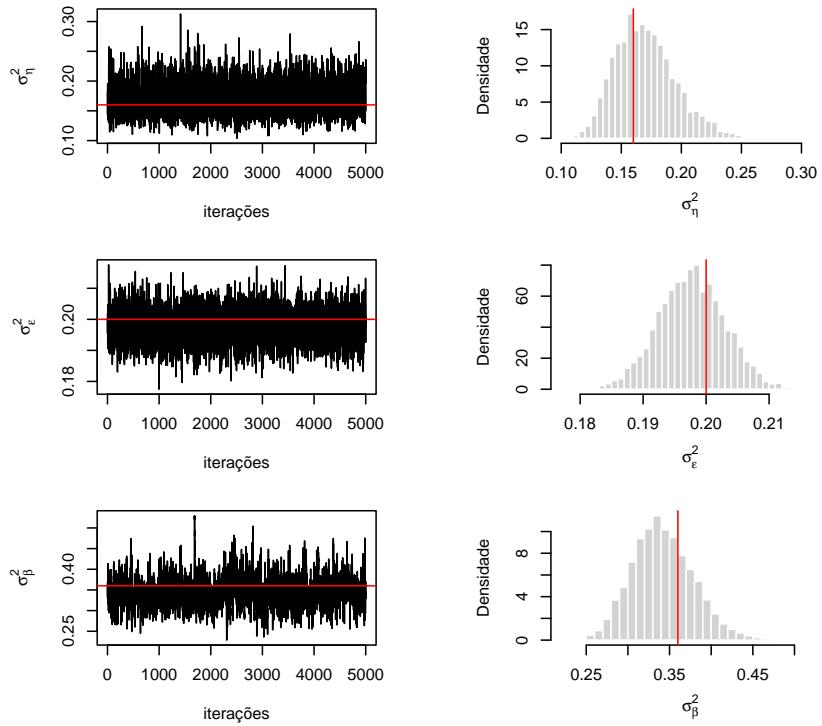


Figura 3.11: Traços das cadeias e histogramas das distribuições *a posteriori* de  $\sigma_\eta^2$ ,  $\sigma_\epsilon^2$ ,  $\sigma_\beta^2$  no modelo normal. As linhas vermelhas horizontal e vertical indicam o valor verdadeiro.

# Capítulo 4

## Modelagem espaço-temporal do desmatamento

Os dados reais foram analisados usando a transformação logit da variável taxa, definida como  $\text{taxa} = (\text{Área desmatada} / \text{Área total do município})$ . A transformação logit foi aplicada para que a variável taxa pudesse assumir qualquer valor na reta real, tornando-a adequada para o modelo normal. Em alguns anos pesquisados, a variável taxa apresentava valores iguais a zero, os quais foram substituídos por 0,00001 para evitar problemas quando aplicamos a função logit.

Para o procedimento de estimação, adotaram-se as mesmas distribuições *a priori* utilizadas nas análises com dados simulados. Para o parâmetro  $\alpha$ , foram considerados dois cenários distintos: o caso em que  $\alpha \in (0, 1)$  e o caso em que  $\alpha = 1$ , a fim de avaliar o impacto das diferentes escolhas do parâmetro.

Para o procedimento de estimação, foram considerados 16.000 iterações com aquecimento de 4.000 e espaçamento de 3 iterações.

A aplicação do modelo foi inicialmente realizada para os municípios do estado do Pará e, em seguida, a análise foi estendida para todos os municípios do Bioma Amazônico.

### 4.1 Análise para o estado do Pará

Nesta seção, realizou-se um estudo com dados reais do estado do Pará, utilizando o modelo proposto em (3.2). Foram consideradas  $p = 5$  (intercepto e 4 covariáveis reais) definidas ao longo de  $T = 21$  anos (2002 a 2022), abrangendo  $N = 144$  localizações,

correspondentes aos municípios do estado do Pará.

#### 4.1.1 Modelo considerando $0 < \alpha < 1$

A Figura 4.1 apresenta a evolução dos coeficientes temporais  $\beta_t$ . Os gráficos exibem as estimativas pontuais, representadas pela média *a posteriori*, acompanhadas de seus respectivos intervalos de 95% de credibilidade. Observa-se que a estimativa temporal do intercepto apresenta variações ao longo do tempo, com uma queda inicial seguida por uma tendência de alta nos anos subsequentes. Os coeficientes das variáveis preditoras, como vegetação nativa remanescente e madeira extraída, apresentam uma tendência de aumento ao longo do tempo. Já no caso do número de cabeças bovinas, não é possível identificar uma tendência consistente de crescimento ou queda. Além disso, é importante destacar que, em diversos anos, os intervalos de credibilidade dos coeficientes incluem o zero. Isso sugere que as variações observadas podem não ser estatisticamente significativas e que esses fatores podem não apresentar mudanças marcantes ao longo do tempo.

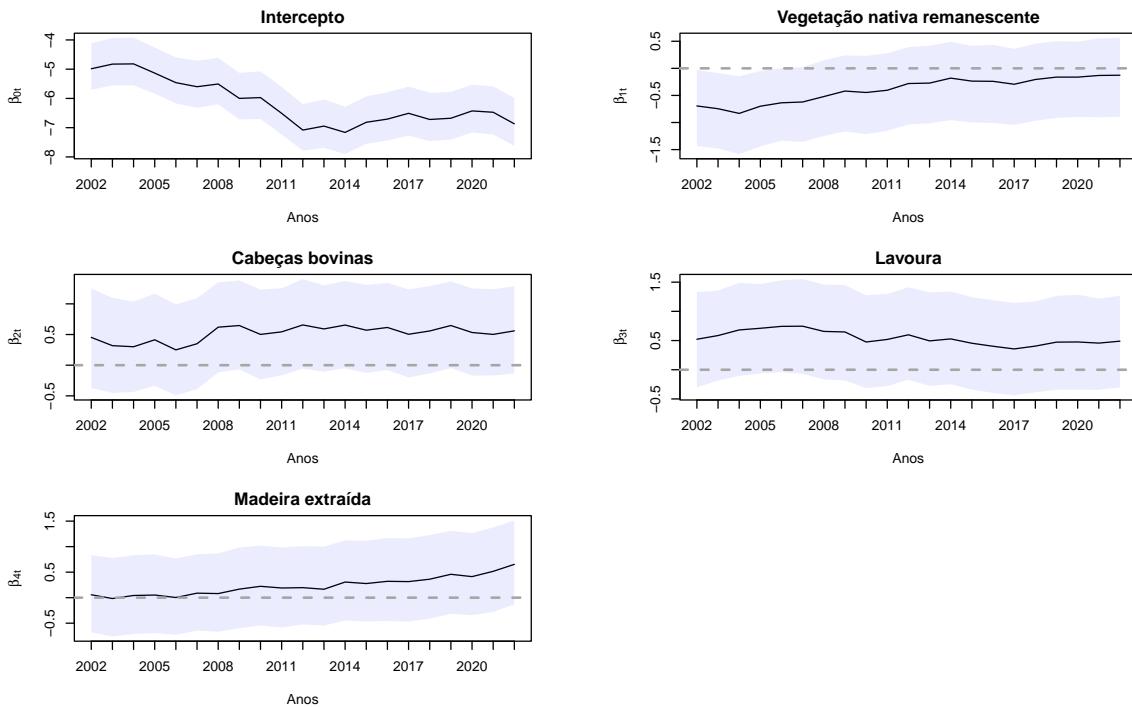


Figura 4.1: Valores estimados através da média *a posteriori* (linha preta) com respectivos intervalos de 95% de credibilidade (área hachurada) dos coeficientes temporais ao longo do tempo, para o modelo normal com  $\alpha \in (0, 1)$ .

Na Figura 4.2 apresenta-se os valores estimados do intercepto e os coeficientes espaciais das variáveis por meio da média *a posteriori* representados em mapas espaciais do estado de Pará. Observa-se por exemplo que no intercepto os valores são maiores na parte sul, o que pode indicar que, nessa região, o desmatamento tende a ser maior independentemente das variáveis preditoras consideradas. A variável cabeças bovinas apresenta coeficientes predominantemente positivos no norte e nordeste do estado. Já a variável lavoura apresenta coeficientes positivos nos municípios localizados no nordeste, sugerindo que a expansão dessa atividade pode estar associada ao desmatamento nessas áreas. Por fim, a variável madeira extraída apresenta coeficientes positivos nos municípios do sul, indicando que essa variável pode estar impulsionando o desmatamento nesses municípios.

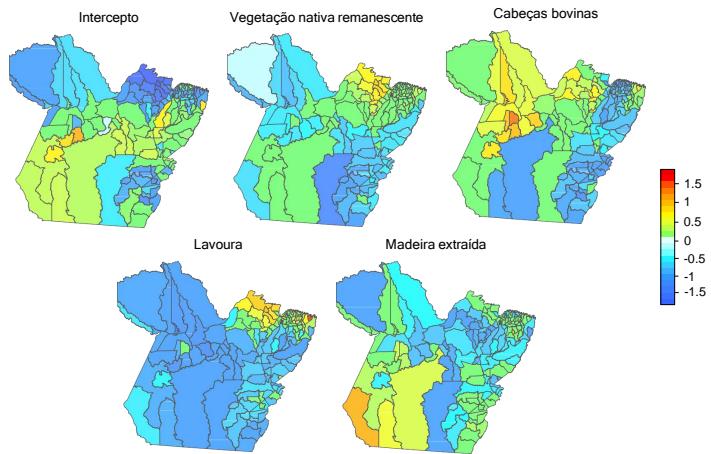


Figura 4.2: Valores estimados através da média *a posteriori* para o intercepto e os coeficientes espaciais das variáveis preditoras para todos os municípios no modelo normal com  $\alpha \in (0, 1)$ .

A Tabela 4.1 apresenta os resultados das estimativas da média *a posteriori* e intervalos de 95% de credibilidade de  $\sigma_\eta^2$ ,  $\sigma_\varepsilon^2$ ,  $\sigma_\beta^2$ , esses hiperparâmetros permitem compreender a variabilidade do processo, destacando diferenças entre a variabilidade temporal, espacial e a variabilidade associada ao erro. Observa-se que a variabilidade temporal  $\sigma_\eta^2$  é pequena, indicando que as mudanças ao longo do tempo são mais estáveis. Por outro lado, a variabilidade espacial, representada por  $\sigma_\beta^2$ , é maior, sugerindo diferenças significativas entre as localidades analisadas.

A Figura 4.3 apresenta-se os traços das cadeias e os histogramas das distribuições *a posteriori* de  $\sigma_\eta^2$ ,  $\sigma_\varepsilon^2$ ,  $\sigma_\beta^2$  e  $\alpha$ . Observa-se que a cadeia do parâmetro  $\alpha$  apresenta bastante autocorrelação e a maioria dos valores está concentrada no intervalo entre 0,96 e 0,99. Isso sugere que, apesar de uma tendência de estabilidade, o parâmetro  $\alpha$  apresenta

uma distribuição mais restrita e com pouca variação ao longo das iterações, o que pode indicar a necessidade de ajustes no modelo ou uma avaliação mais aprofundada da sensibilidade do parâmetro.

Tabela 4.1: Resumo da distribuição *a posteriori* de  $\sigma_\eta^2$ ,  $\sigma_\varepsilon^2$ ,  $\sigma_\beta^2$  e  $\alpha$  do modelo normal aplicado aos dados reais.

Parâmetros	Média <i>a posteriori</i>	Intervalo de credibilidade
$\sigma_\eta^2$	0,042	(0,020 ; 0,085)
$\sigma_\varepsilon^2$	1,510	(1,383 ; 1,670)
$\sigma_\beta^2$	2,436	(1,519 ; 3,775)
$\alpha$	0,974	(0,887 ; 0,989)

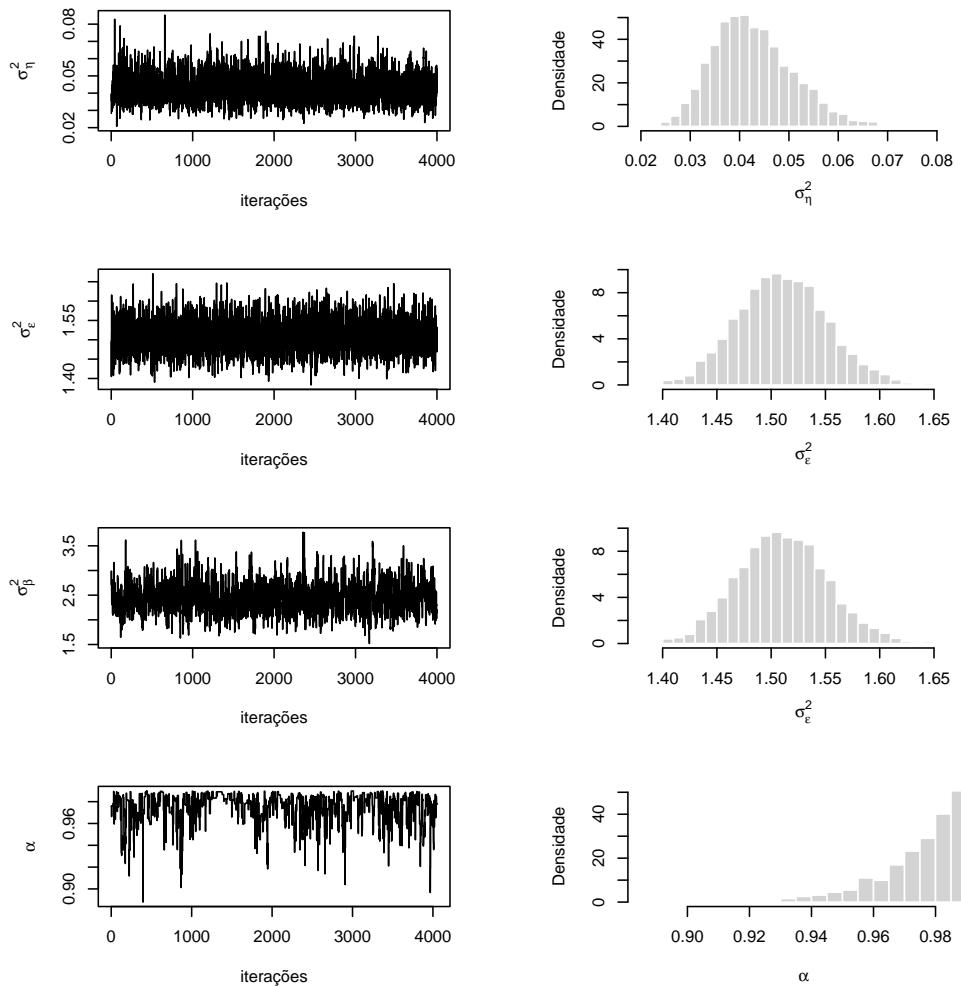


Figura 4.3: Traços das cadeias e histogramas das distribuições *a posteriori* de  $\sigma_\eta^2$ ,  $\sigma_\varepsilon^2$ ,  $\sigma_\beta^2$  e  $\alpha$  no modelo normal para dados reais.

#### 4.1.2 Modelo considerando $\alpha = 1$

A Figura 4.4 apresenta as médias *a posteriori* e os intervalos de 95% de credibilidade *a posteriori* do intercepto e as variáveis explicativas ao longo do tempo. Observa-se que o comportamento dos coeficientes temporais segue um padrão semelhante ao do modelo com  $\alpha \in (0, 1)$ , indicando que a estrutura dos efeitos no tempo se mantém estável. No entanto, para a variável *lavoura*, observa-se uma inclinação mais acentuada, o que sugere uma mudança mais pronunciada na tendência desse coeficiente em comparação com o outro modelo.

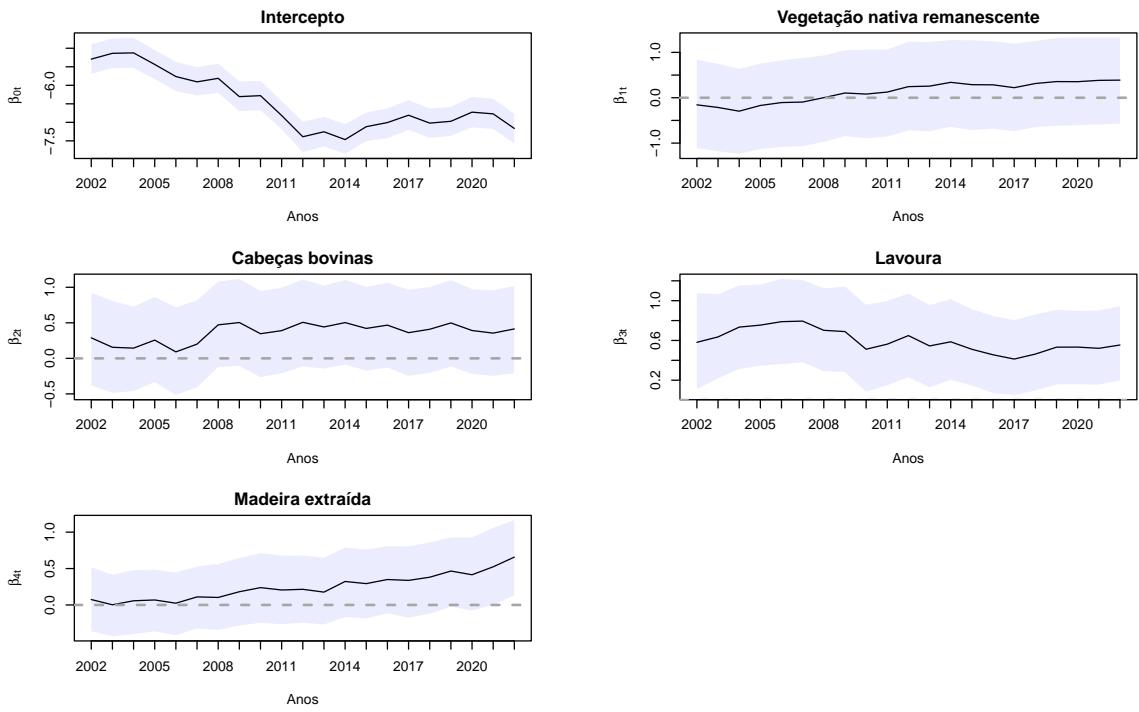


Figura 4.4: Valores estimados através da média *a posteriori* (linha preta) com respectivos intervalos de 95% de credibilidade (área hachurada) dos coeficientes temporais ao longo do tempo, para o modelo normal com  $\alpha = 1$ .

A Figura 4.5 apresenta os coeficientes espaciais estimados para o modelo com  $\alpha = 1$ , representados por mapas coropléticos, que exibem os valores estimados do intercepto e dos coeficientes associados às variáveis explicativas. Observa-se que os padrões espaciais das variáveis *cabeças bovinas*, *lavoura* e *madeira extraída* são bastante semelhantes aos observados no modelo com  $\alpha \in (0, 1)$ . Por outro lado, a variável *vegetação nativa remanescente* apresenta diferenças relevantes, por exemplo os municípios como *Porto de Moz*, *Portel*, *Itaituba* e *Redenção* apresentam coeficiente negativo em relação ao desmatamento.

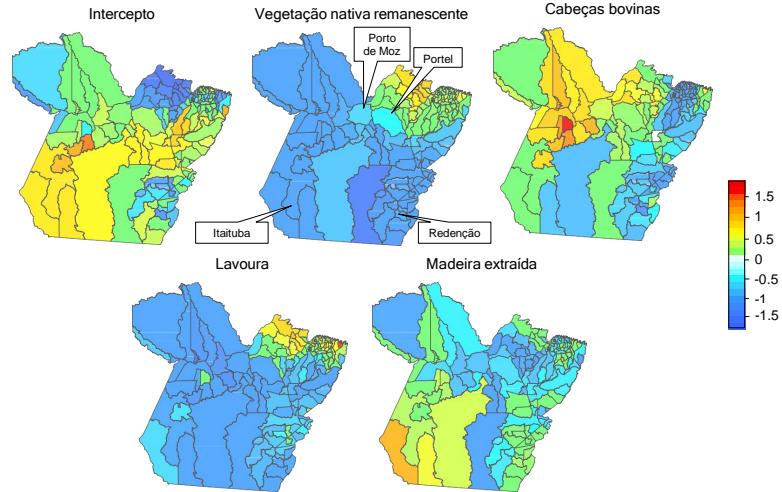


Figura 4.5: Valores estimados através da média *a posteriori* para o intercepto e os coeficientes espaciais das variáveis preditoras para todos os municípios no modelo normal com  $\alpha = 1$

A Tabela 4.2 apresenta as médias *a posteriori* e intervalos de 95% de credibilidade de  $\sigma_\eta^2$ ,  $\sigma_\varepsilon^2$ ,  $\sigma_\beta^2$ . Destaca-se que  $\sigma_\beta^2$  continua sendo o parâmetro com maior incerteza, assim como no caso anterior.

Tabela 4.2: Resumo das distribuições *a posteriori* de  $\sigma_\eta^2$ ,  $\sigma_\varepsilon^2$ ,  $\sigma_\beta^2$  do modelo normal em dados reais.

Parâmetros	Média <i>a posteriori</i>	Intervalo de credibilidade
$\sigma_\eta^2$	0,042	(0,020 ; 0,088)
$\sigma_\varepsilon^2$	1,509	(1,364 ; 1,676)
$\sigma_\beta^2$	2,227	(1,433 ; 3,492)

A Figura 4.6 exibe os traços das cadeias e os histogramas das distribuições *a posteriori* de  $\sigma_\eta^2$ ,  $\sigma_\varepsilon^2$  e  $\sigma_\beta^2$ . Nota-se que todas as cadeias parecem convergir. No Apêndice A.2, incluímos os traços de alguns coeficientes temporais  $\beta_t$  e espaciais  $\beta$  para ilustrar o comportamento dessas amostras.

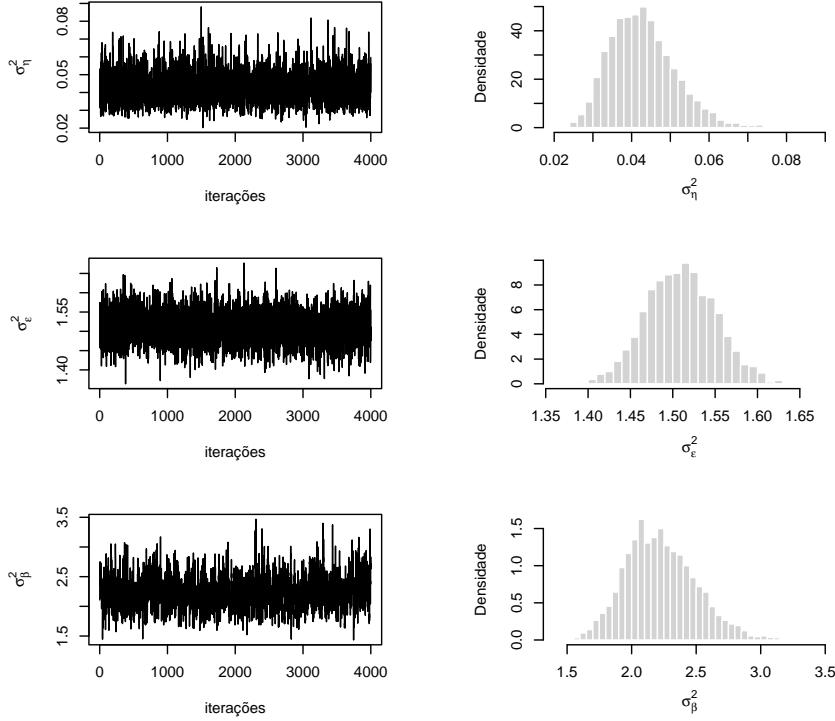


Figura 4.6: Traços das cadeias e histogramas das distribuições *a posteriori* de  $\sigma_\eta^2$ ,  $\sigma_\epsilon^2$ ,  $\sigma_\beta^2$  no modelo normal para dados reais.

A diferença entre os dois casos de  $\alpha$  está na estrutura da matriz de precisão e no efeito da dependência espacial. Quando  $\alpha = 1$ , cada localização é quase totalmente determinada pelos vizinhos, podendo levar à singularidade da matriz e tornar a distribuição imprópria. Quando  $\alpha \in (0, 1)$  a correlação espacial é controlada, garantindo uma matriz de precisão bem condicionada e evitando instabilidade no ajuste do modelo.

Para comparar a qualidade dos modelos que consideram  $\alpha \in (0, 1)$  e  $\alpha = 1$ , utilizamos o Critério de Informação Watanabe-Akaike (WAIC) (Watanabe, 2010). O WAIC é um critério totalmente bayesiano, pois realiza uma média sobre a distribuição *a posteriori*. Para mais detalhes sobre esse critério, ver Apêndice B.

A Tabela 4.3 apresenta os valores do WAIC para cada modelo ajustado, juntamente com suas componentes: lppd, que mede a qualidade do ajuste do modelo, e pWAIC, que representa a penalização associada ao número efetivo de parâmetros do modelo. Observa-se que o modelo com  $\alpha = 1$  apresenta um menor valor de WAIC, sugerindo um melhor ajuste em comparação com o modelo onde  $\alpha \in (0, 1)$ . Isso sugere que a escolha de  $\alpha = 1$  pode ser a mais apropriada dentro do contexto analisado.

Tabela 4.3: Critério de comparação de modelos WAIC.

Modelo	lppd	pWAIC	WAIC
Modelo com $\alpha \in (0, 1)$	-4818,7865	229,9808	10097,5347
Modelo com $\alpha = 1$	-4817,8056	228,3378	10092,2870

A Figura 4.7 apresenta o gráfico de dispersão entre os valores replicados e observados da variável  $Y(s, t)$ , permitindo avaliar a qualidade do ajuste para cada modelo. A linha diagonal representa a referência ideal onde os valores replicados seriam iguais aos observados. Observa-se que, em ambos modelos, a maioria dos pontos estão distribuídos próximos a essa linha, indicando um bom ajuste geral. No entanto, o modelo com  $\alpha = 1$  parece apresentar menor dispersão em torno da diagonal, sugerindo um ajuste ligeiramente superior em comparação com o outro modelo.

Essa diferença pode estar associada à flexibilidade adicional conferida pelo parâmetro  $\alpha$  no primeiro modelo. Embora essa flexibilidade possa melhorar o ajuste aos dados, ela também pode levar a um aumento na complexidade do modelo, o que se reflete na maior penalização pWAIC. Isso sugere que, apesar do modelo com  $\alpha \in (0, 1)$  permitir uma estrutura mais flexível, essa vantagem não se traduz necessariamente em um melhor desempenho, conforme indicado pelo critério de informação.

Em conclusão, os resultados sugerem que a fixação de  $\alpha = 1$  pode levar a um modelo mais parcimonioso, com melhor capacidade preditiva e menor risco de sobreajuste.

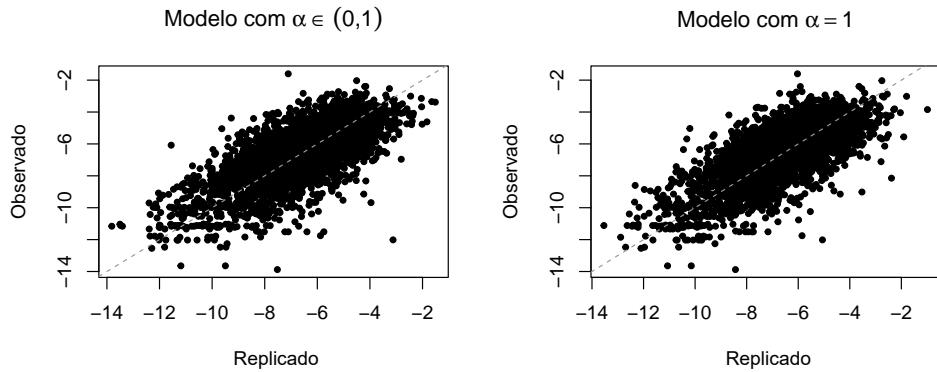


Figura 4.7: Gráfico dos valores replicados versus os valores observados.

A Figura 4.8 apresenta mapas do coeficiente de regressão  $\tilde{\beta}(s, t)$  de algumas co-variáveis para alguns instantes de tempo no modelo considerando  $\alpha = 1$ , mostrando os efeitos significativos e não significativos em diferentes anos. Os municípios estão

coloridos de acordo com os efeitos das covariáveis: vermelho indica efeitos significativamente positivos, azul representa efeitos significativamente negativos, e branco mostra áreas onde não há evidências suficientes para considerar o efeito como significativo.

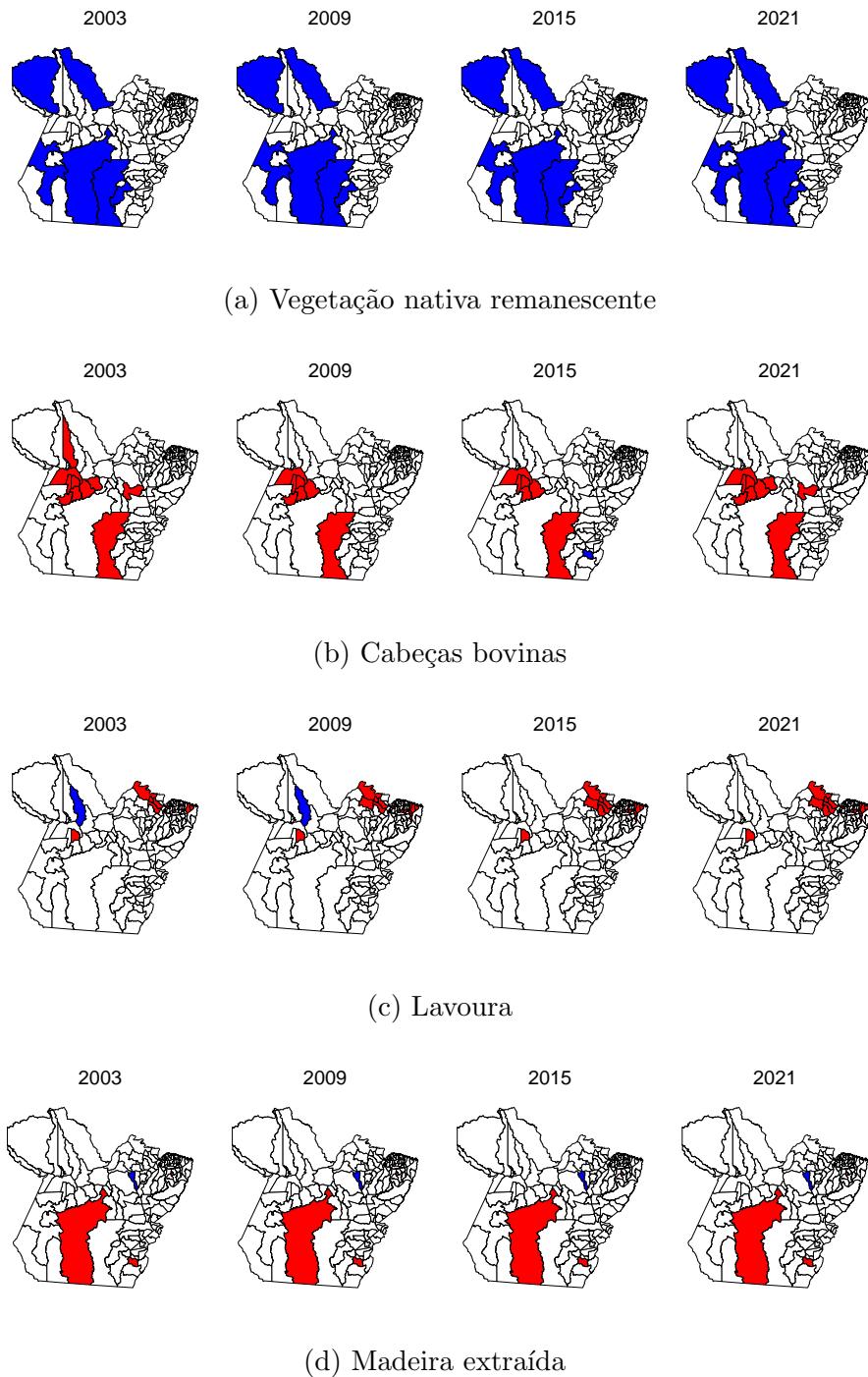


Figura 4.8: Mapas dos  $\tilde{\beta}(s, t)$  para o modelo normal com  $\alpha = 1$  com efeitos significativamente positivos (em vermelho), não significativos (em branco) e negativos (em azul) para alguns anos.

Após a aplicação inicial do modelo aos dados do estado do Pará, ampliamos a análise para abranger todo o Bioma Amazônico. O objetivo é avaliar o desempenho do modelo proposto em uma escala espacial mais ampla, considerando um número maior de municípios.

## 4.2 Análise para o Bioma Amazônico

Nesta seção, apresentamos um estudo baseado em dados reais do Bioma Amazônico considerando o modelo proposto. Foram consideradas  $p = 5$  (intercepto e 4 covariáveis reais) definidas ao longo de  $T = 21$  anos (2002-2022), abrangendo  $N = 559$  localizações, correspondentes aos municípios que compõem o bioma.

Neste caso, não implementamos o modelo para o caso em que  $\alpha \in (0, 1)$  devido ao alto custo computacional e ao tempo que ele demanda para sua execução. Como os resultados anteriores mostram que em uma região geográfica menor já o modelo com  $\alpha = 1$  é um bom modelo, apresentamos a seguir apenas os resultados do modelo considerando  $\alpha = 1$ . Devido ao tempo que ele demanda, foram realizadas apenas 4.000 iterações do MCMC.

A Figura 4.9 apresenta as médias *a posteriori* e intervalos de 95% de credibilidade do intercepto e dos coeficientes temporais ao longo do tempo. A estimativa de  $\beta_{0t}$  apresenta variações ao longo do tempo. A redução inicial pode coincidir com períodos de políticas ambientais mais rigorosas, mas a reversão dessa tendência nos anos seguintes sugere uma mudança no comportamento do desmatamento, que pode estar relacionada a uma série de fatores, incluindo mudanças na fiscalização a partir de 2014. Os coeficientes das variáveis preditoras, como vegetação nativa remanescente e madeira extraída, mostram tendência de aumento. Contudo, em diversos anos, os intervalos de credibilidade de algumas covariáveis incluem o zero, o que sugere que as variações observadas podem não ser estatisticamente significativas, indicando que as mudanças ao longo do tempo podem ser menos expressivas.

A Figura 4.10 apresenta uma visualização espacial do comportamento do intercepto espacial e dos coeficientes das covariáveis selecionadas. A vegetação nativa remanescente apresenta coeficientes negativos na maior parte das regiões, com coeficientes elevados observados em algumas áreas específicas, como no sul do estado de Mato Grosso, no nordeste do Pará e em partes do Tocantins. A variável cabeças bovinas apresenta coeficientes espaciais positivos na maioria dos municípios, com valores mais altos no nordeste do Pará, Acre, Roraima, norte do Amazonas e Amapá.

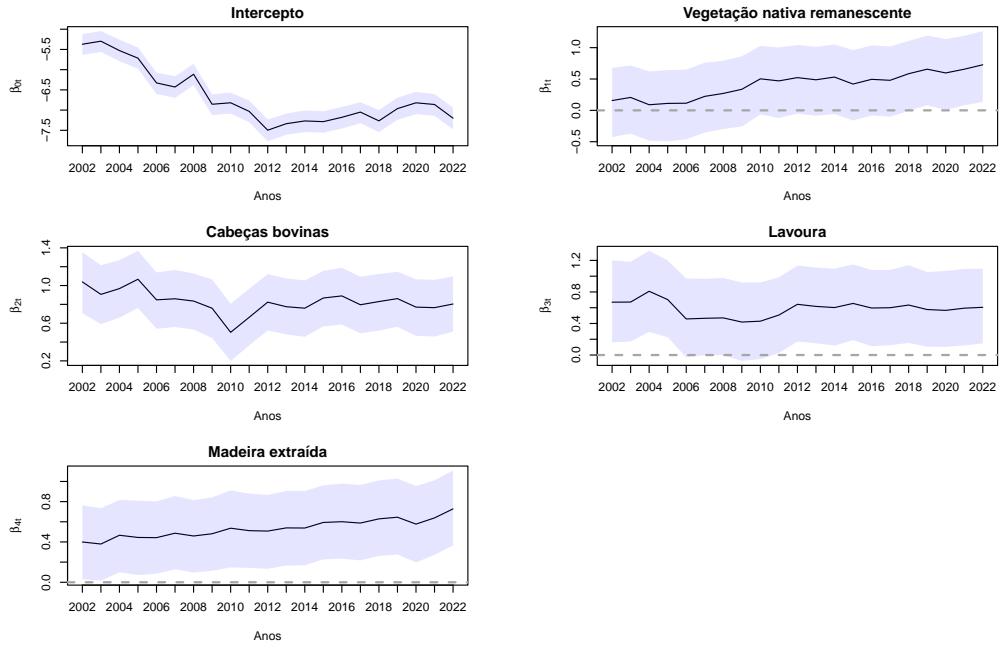


Figura 4.9: Média *a posteriori* (linha preta) com respectivos intervalos 95% credibilidade (área hachurada) dos coeficientes temporais ao longo dos anos, para o Bioma Amazônico.

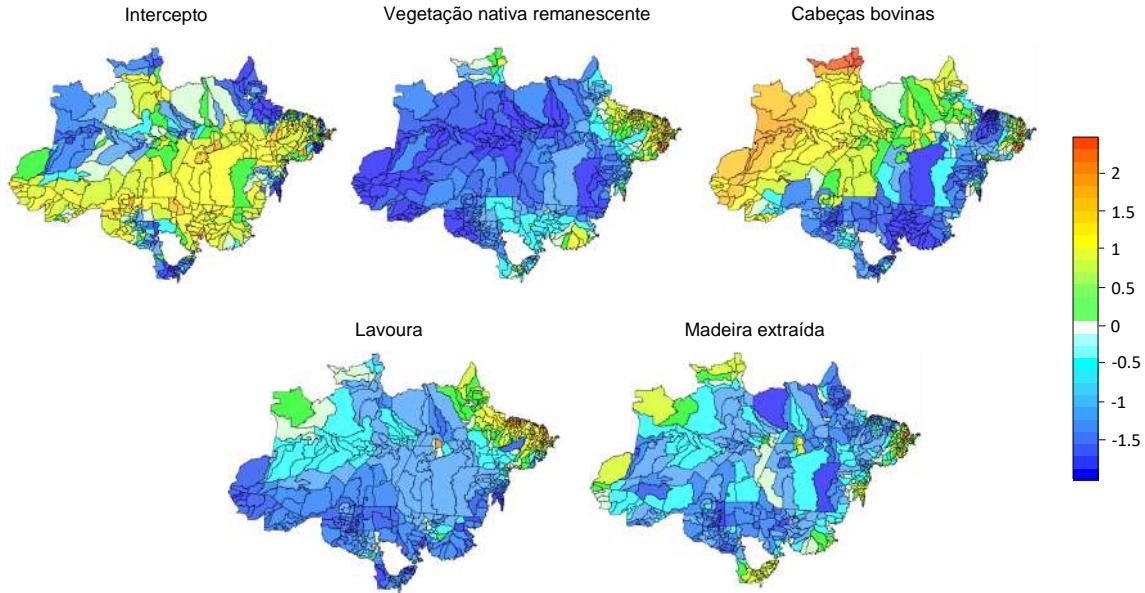


Figura 4.10: Valores estimados do intercepto e dos coeficientes espaciais das covariáveis para o Bioma Amazônico.

A Tabela 4.4 apresenta as médias *a posteriori* e os limites inferior e superior dos intervalos de 95% de credibilidade dos parâmetros do modelo.

Tabela 4.4: Resumo da distribuição *a posteriori* dos parâmetros  $\sigma_\eta^2$ ,  $\sigma_\varepsilon^2$  e  $\sigma_\beta^2$  do modelo normal no Bioma Amazônico.

Parâmetros	Média <i>a posteriori</i>	Intervalo de credibilidade
$\sigma_\eta^2$	0,034	(0,017 ; 0,059)
$\sigma_\varepsilon^2$	1,541	(1,467 ; 1,622)
$\sigma_\beta^2$	2,802	(2,242 ; 3,395)

A Figura 4.11 apresenta os traços das cadeias com seus histogramas. Observa-se que todas convergem.

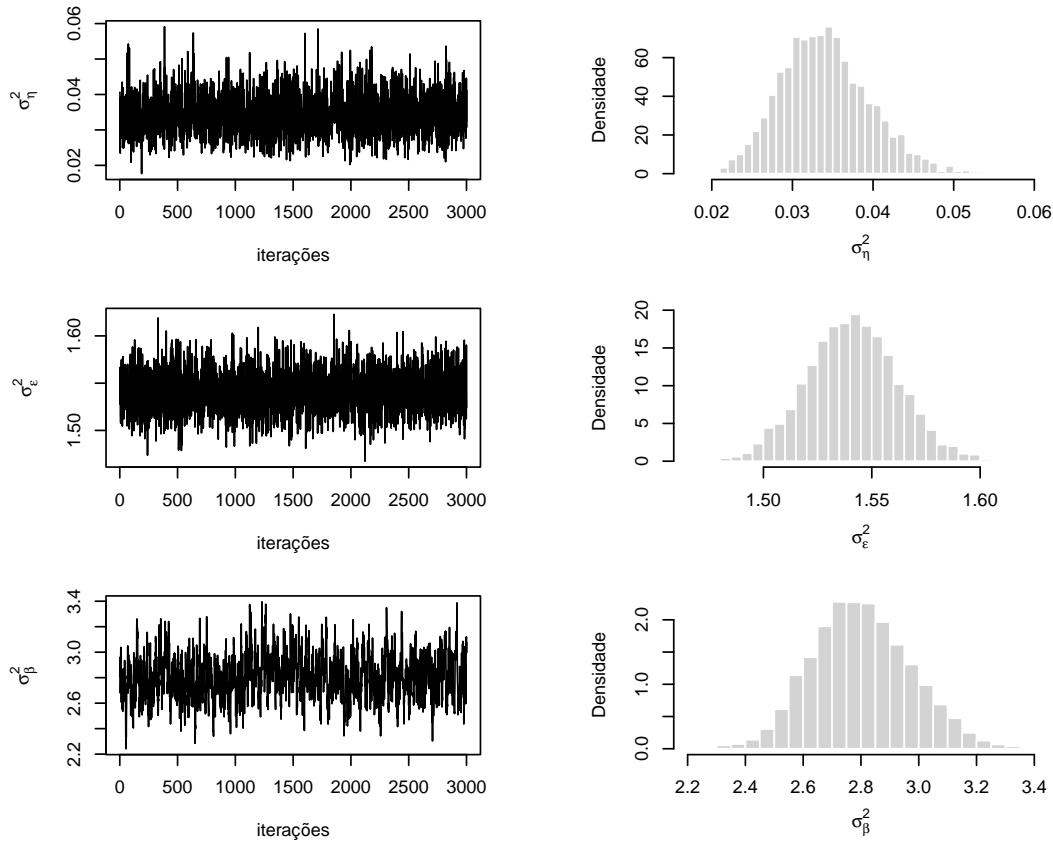


Figura 4.11: Traços das cadeias e histogramas das distribuições *a posteriori* de  $\sigma_\eta^2$ ,  $\sigma_\varepsilon^2$ ,  $\sigma_\beta^2$  no modelo normal para o bioma.

# Capítulo 5

## Modelo de regressão beta espaço-temporal com coeficientes variando no espaço e no tempo

A distribuição beta é bastante versátil para modelar proporções, sendo sua densidade dada por

$$f(y; \alpha, \beta) = \frac{\Gamma(p+q)}{\Gamma(p)\Gamma(q)} y^{p-1} (1-y)^{q-1}, \quad 0 < y < 1$$

em que  $p > 0$  e  $q > 0$  são parâmetros que indexam a distribuição, e  $\Gamma(\cdot)$  denota a função gama.

[Ferrari e Cribari-Neto \(2004\)](#) propuseram uma parametrização para a densidade beta, em que definem  $\mu = p/(p+q)$  e  $\phi = p+q$ , com isso, os parâmetros  $p$  e  $q$  podem ser reescritos da seguinte forma  $p = \mu\phi$  e  $q = (1-\mu)\phi$ . Assim, temos que

$$E(y) = \mu \quad \text{e} \quad \text{var}(y) = \frac{\mu(1-\mu)}{1+\phi}$$

de tal forma que  $\mu$  é a média da variável  $y$  e  $\phi$  pode ser interpretado como um parâmetro de precisão. Com essa nova parametrização, a densidade de  $y$  pode ser escrita como

$$f(y; \mu, \phi) = \frac{\Gamma(\phi)}{\Gamma(\mu\phi)\Gamma((1-\mu)\phi)} y^{\mu\phi-1} (1-y)^{(1-\mu)\phi-1}, \quad 0 < y < 1,$$

em que  $0 < \mu < 1$  e  $\phi > 0$ .

## 5.1 Modelo proposto

Seja  $s = \{s_1, \dots, s_N\}$  o conjunto de localizações e  $t = \{1, \dots, T\}$  os instantes de tempo. Observa-se a proporção de desmatamento  $Y(s, t)$  e um vetor de covariáveis  $\mathbf{x}(s, t)$ , de dimensão  $p$ , com coeficientes de regressão  $\tilde{\beta}(s, t)$ . O modelo inclui um termo de erro  $\varepsilon(s, t)$ , assumido como um ruído gaussiano centrado em zero e não correlacionado no tempo e no espaço. Além disso, o vetor  $\tilde{\beta}(s, t)$ , de dimensão  $(p \times 1)$  se decompõe em componente puramente temporal  $\beta_t$  e componente puramente espacial  $\beta(s)$ . Dessa forma, a estrutura espaço-temporal do modelo é definida por:

$$\begin{aligned} Y(s, t) &\sim \text{Beta}(\mu(s, t), \phi) \\ \zeta(s, t) &= \log \left\{ \frac{\mu(s, t)}{1 - \mu(s, t)} \right\} = \mathbf{x}'(s, t) \tilde{\beta}(s, t) + \varepsilon(s, t), \quad \varepsilon(s, t) \sim N(\mathbf{0}, \sigma_\varepsilon^2), \\ \tilde{\beta}(s, t) &= \beta_t + \beta(s), \\ \beta_t &= \beta_{t-1} + \boldsymbol{\eta}_t, \quad \boldsymbol{\eta}_t \stackrel{\text{ind}}{\sim} N(\mathbf{0}, \sigma_\eta^2 I_p). \end{aligned} \quad (5.1)$$

Considere-se  $\mathbf{y}_t = (y(s_1, t), \dots, y(s_N, t))'$ ,  $\boldsymbol{\mu}_t = (\mu(s_1, t), \dots, \mu(s_N, t))'$ ,  $\boldsymbol{\zeta}_t = (\zeta(s_1, t), \dots, \zeta(s_N, t))'$  e  $\boldsymbol{\varepsilon}_t = (\varepsilon(s_1, t), \dots, \varepsilon(s_N, t))'$  vetores de dimensão  $(N \times 1)$  para cada  $t = 1, \dots, T$ . Assim, o modelo em (5.1) pode ser escrito em sua forma matricial da seguinte forma

$$\begin{aligned} \boldsymbol{\zeta}_t &= F_t \boldsymbol{\theta}_t + \boldsymbol{\varepsilon}_t, \quad \boldsymbol{\varepsilon}_t \sim N(\mathbf{0}, \sigma_\varepsilon^2 I_N), \\ \boldsymbol{\theta}_t &= \mathbf{1}_N \otimes \beta_t + \boldsymbol{\beta}, \\ \beta_t &= \beta_{t-1} + \boldsymbol{\eta}_t, \quad \boldsymbol{\eta}_t \sim N(\mathbf{0}, \sigma_\eta^2 I_p), \\ \boldsymbol{\beta} &\sim N(\mathbf{0}, [D_v - \alpha V]^{-1} \otimes \sigma_\beta^2 I_p). \end{aligned} \quad (5.2)$$

## 5.2 Procedimento de inferência

O procedimento de inferência será feito sob o enfoque bayesiano. Assim, uma distribuição *a priori* será assumida para os parâmetros do modelo a fim da obtenção da distribuição *a posteriori*.

A função de verossimilhança é da forma

$$\begin{aligned}
l(\Theta|y) &\propto \prod_{t=1}^T \prod_{i=1}^N p(y(s_i, t) | \mu(s_i, t), \phi) \\
&\propto \prod_{t=1}^T \prod_{i=1}^N \Gamma(\phi)(1 - y(s_i, t))^{\phi} \exp \left\{ \mu(s_i, t)\phi \log \left( \frac{y(s_i, t)}{1 - y(s_i, t)} \right) - \log \Gamma(\mu(s_i, t)\phi) - \right. \\
&\quad \left. \log \Gamma((1 - \mu(s_i, t))\phi) \right\}.
\end{aligned}$$

### Distribuição *a priori*

O vetor paramétrico a ser estimado é definido como

$$\Theta = (\zeta_t, \beta_0, \beta_1, \dots, \beta_T, \beta(s_1), \dots, \beta(s_N), \sigma_{\varepsilon}^2, \sigma_{\eta}^2, \sigma_{\beta}^2, \alpha, \phi),$$

contendo todas as variáveis desconhecidas do modelo em (5.2) a serem estimadas.

Assim, a distribuição *a priori* assumida para o vetor acima será descrita como

$$p(\Theta) = p(\zeta_t | \beta_t, \beta, \sigma_{\varepsilon}^2) p(\beta_t | \beta_{t-1}, \sigma_{\eta}^2) p(\beta | \alpha, \sigma_{\beta}^2) p(\beta_0) p(\sigma_{\varepsilon}^2) p(\sigma_{\eta}^2) p(\sigma_{\beta}^2) p(\alpha) p(\phi),$$

em que os parâmetros  $\beta_0$ ,  $\sigma_{\varepsilon}^2$ ,  $\sigma_{\eta}^2$  e  $\sigma_{\beta}^2$  tem as mesmas distribuições *a priori* do modelo normal descrito em (3.3). Além disso, para o parâmetro  $\phi$ , considera-se uma distribuição *a priori* independente dos demais parâmetros, dada por  $\phi \sim \text{Gamma}(a_{\phi}, b_{\phi})$ , em que  $a_{\phi}$  e  $b_{\phi}$  são valores conhecidos.

### Distribuição *a posteriori*

Após definir a distribuição *a priori* para o vetor paramétrico  $\Theta$ , pode-se então determinar a distribuição *a posteriori*, que é dada por

$$\begin{aligned}
p(\Theta | y) &\propto l(\Theta|y) \left[ \prod_{t=1}^T p(\zeta_t | \beta_t, \beta, \sigma_{\varepsilon}^2) \right] \left[ \prod_{t=1}^T p(\beta_t | \beta_{t-1}, \sigma_{\eta}^2) \right] \\
&\quad \times p(\beta | \alpha, \sigma_{\beta}^2) p(\beta_0) p(\sigma_{\varepsilon}^2) p(\sigma_{\eta}^2) p(\sigma_{\beta}^2) p(\alpha) p(\phi).
\end{aligned} \tag{5.3}$$

$$\begin{aligned}
&\propto l(\Theta|y) \left[ \prod_{t=1}^T (\sigma_{\varepsilon}^{2N})^{-1/2} \exp \left\{ -\frac{1}{2\sigma_{\varepsilon}^2} (\zeta_t - F_t(\mathbf{1}_N \otimes \beta_t + \beta))^2 \right\} \right] \left[ \prod_{t=1}^T p(\beta_t | \beta_{t-1}, \sigma_{\eta}^2) \right] \\
&\quad \times p(\beta | \alpha, \sigma_{\beta}^2) p(\beta_0) p(\sigma_{\varepsilon}^2) p(\sigma_{\eta}^2) p(\sigma_{\beta}^2) p(\alpha) p(\phi).
\end{aligned} \tag{5.4}$$

A equação acima não é analiticamente tratável. Para que seja possível realizar o procedimento de inferência e obter amostras dos parâmetros a partir da sua distribuição *a posteriori*, faz-se necessária a utilização de Métodos de Monte Carlo via Cadeias de Markov (MCMC) serão utilizados, especificamente, o amostrador de Gibbs.

### 5.2.1 Distribuições condicionais completas

Nesta subseção são mostradas as distribuições condicionais completas dos elementos que compõem o vetor paramétrico  $\Theta$ . Para a maioria deles, essas distribuições possuem forma conhecida e podem ser derivadas diretamente a partir das condições do modelo. No entanto,  $\zeta_t$ ,  $\phi$  e  $\alpha$  não apresentam uma forma conhecida.

A seguir, são apresentadas em detalhe as formas das distribuições condicionais completas de  $\zeta_t$  e  $\phi$ .

#### Vetor $\zeta_t$

A distribuição condicional completa de  $\zeta(s_i, t)$  para cada  $t = 1, \dots, T$  e  $i = 1, \dots, N$  é da forma

$$p(\zeta(s_i, t) | \Theta_{-\zeta(s_i, t)}, y) \propto \exp \left\{ \mu(s_i, t)\phi \log \left( \frac{y(s_i, t)}{1 - y(s_i, t)} \right) - \log \Gamma(\mu(s_i, t)\phi) - \log \Gamma((1 - \mu(s_i, t))\phi) \right\} \times \exp \left\{ -\frac{1}{2} (\zeta(s_i, t) - \mathbf{x}'(s_i, t)\tilde{\beta}(s_i, t))'(\sigma_\varepsilon^2)^{-1} (\zeta(s_i, t) - \mathbf{x}'(s_i, t)\tilde{\beta}(s_i, t)) \right\} \quad (5.5)$$

não é uma distribuição conhecida.

Inicialmente, para obter amostras dessa distribuição utilizamos o algoritmo Metropolis-Hastings com proposta baseada em uma adaptação do algoritmo proposto por [Gamerman \(1997\)](#), em que, no passo  $l$  do algoritmo, os momentos da função de densidade proposta são determinados por:

$$\begin{aligned} \mathbf{m}_t^{(l)} &= ((\sigma_\varepsilon^2)^{-1} + W(\zeta^{(l-1)}(s_i, t)))^{-1} \\ &\quad \times (\sigma_\varepsilon^2)^{-1} \mathbf{x}'(s_i, t)(\beta_t + \beta) + W(\zeta^{(l-1)}(s_i, t))y(\zeta^{(l-1)}(s_i, t)) \\ C_t^{(l)} &= ((\sigma_\varepsilon^2)^{-1} + W(\zeta^{(l-1)}(s_i, t)))^{-1}. \end{aligned}$$

em que  $W(\zeta(s_i, t)) = \mu(s_i, t)(1 - \mu(s_i, t))(\phi + 1)$  é uma matriz diagonal de pesos.

A ideia era utilizar uma distribuição proposta eficiente para sortear  $\zeta(s_i, t)$  para cada  $t$  e  $i$ . No entanto, essa abordagem não gerou bons resultados, razão pela qual optou-se por utilizar o algoritmo Metropolis-Hastings com uma proposta passeio aleatório. Especificamente, uma distribuição normal com média centrada no valor corrente da cadeia.

### Parâmetro $\phi$

$$p(\phi | \Theta_{-\phi}, y) \propto \Gamma(\phi)^{NT} \prod_{t=1}^T \prod_{i=1}^N (1 - y(s_i, t))^{\phi} \phi^{(a_\phi - 1)} \\ \times \exp \left\{ \mu(s_i, t) \phi \log \left( \frac{y(s_i, t)}{1 - y(s_i, t)} \right) - \right. \\ \left. \log \Gamma(\mu(s_i, t) \phi) - \log \Gamma((1 - \mu(s_i, t)) \phi - b_\phi \phi) \right\} \quad (5.6)$$

Nota-se que a forma da distribuição condicional completa de  $\phi$  não é conhecida. Por isso, utilizou-se o algoritmo Metropolis-Hastings com uma proposta normal truncada centrada no valor corrente da cadeia.

As distribuições condicionais completas dos demais parâmetros permanecem idênticas as apresentadas na subseção 3.2.2, dado que o modelo proposto em (3.2) foi adaptado para um contexto de regressão beta. No entanto, no caso do  $\beta$  e  $\sigma_\varepsilon^2$ , há uma modificação específica: a variável  $\mathbf{Y}_t$  é substituída por  $\zeta_t$ .

## 5.3 Aplicação: dados artificiais

Nesta seção, apresenta-se um estudo com dados artificiais gerados a partir do modelo em (5.2). Para isso, foram simulados dados considerando  $p = 5$  (intercepto e 4 covariáveis reais) definidas ao longo de  $T = 21$  anos (2002 a 2022), abrangendo  $N = 144$  localizações, correspondentes aos municípios do estado do Pará.

Nesta aplicação, foram consideradas 12.000 iterações do algoritmo MCMC com aquecimento de 2.000 e espaçamento de 2 iterações, resultando numa amostra de tamanho 5.000.

### 5.3.1 Modelo considerando $0 < \alpha < 1$

Os parâmetros do modelo com  $\alpha \in (0, 1)$  foram fixados em  $\sigma_\eta^2 = 0,1$ ,  $\sigma_\varepsilon^2 = 0,3$ ,  $\sigma_\beta^2 = 0,42$ ,  $\alpha = 0,65$  e  $\phi = 28$ . Para o procedimento de estimação, assumiram-se distribuições *a priori* gama inversa para  $\sigma_\varepsilon^2$ ,  $\sigma_\beta^2$  e  $\sigma_\eta^2$  com parâmetros de forma 0,1 e de escala 0,3 e para  $\alpha$  uma distribuição *a priori* uniforme. Além disso, considerou-se  $\phi \sim \text{Gamma}(5, 3)$ . Essa escolha foi feita de forma a garantir que as estimativas de  $\phi$  não se afastassem excessivamente do valor verdadeiro simulado e, ao mesmo tempo, para minimizar a correlação com  $\sigma_\varepsilon^2$ . Testes com outras configurações de prioris não forneceram estimativas adequadas, o que levou à escolha dessa distribuição específica para  $\phi$ .

A Figura 5.1 apresenta a evolução temporal do intercepto temporal e dos coeficientes temporais das variáveis explicativas. Nela, compara-se os valores verdadeiros simulados com os valores obtidos da média *a posteriori*. Observa-se que as estimativas acompanham bem os valores verdadeiros ao longo do tempo, capturando as tendências gerais das séries.

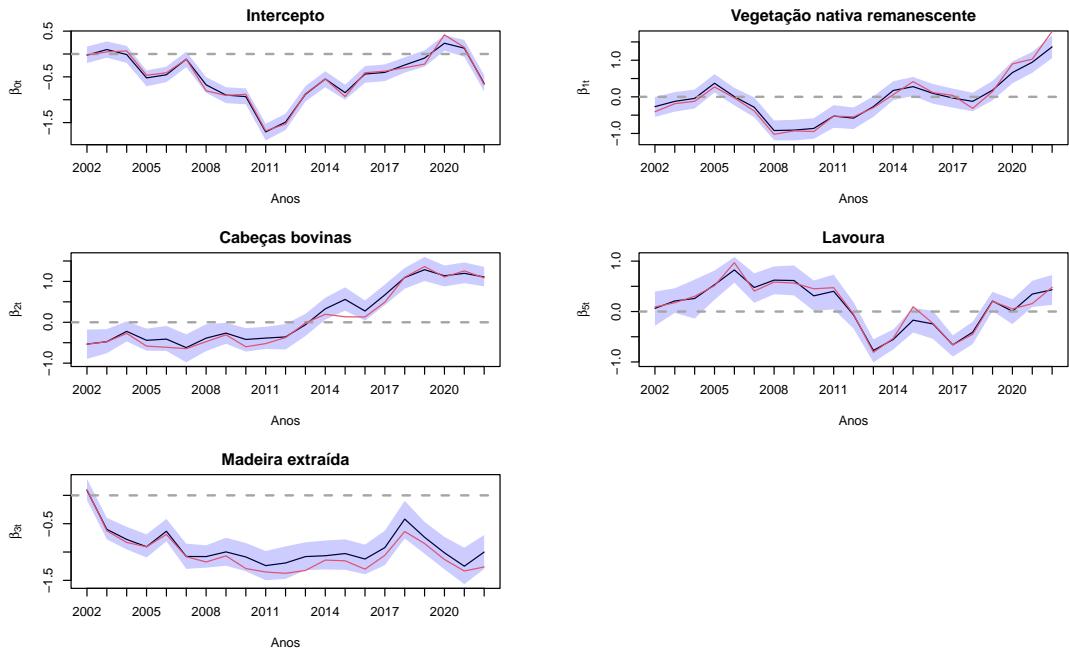


Figura 5.1: Valores estimados através da média *a posteriori* (linha preta) e valores verdadeiros (linha vermelha) com respectivos intervalos de 95% de credibilidade (área hachurada) dos coeficientes temporais ao longo do tempo, para o modelo beta com  $\alpha \in (0, 1)$ .

A Figura 5.2 ilustra a comparação entre os valores verdadeiros e os valores estimados obtidos através da média *a posteriori* do intercepto e dos coeficientes espaciais, des-

tacando também as incertezas associadas a essas estimativas para alguns municípios. Já a Figura 5.3 apresenta essa comparação em mapas coropléticos. Observa-se que, embora os padrões gerais sejam capturados, alguns municípios apresentam discrepâncias nos valores estimados.

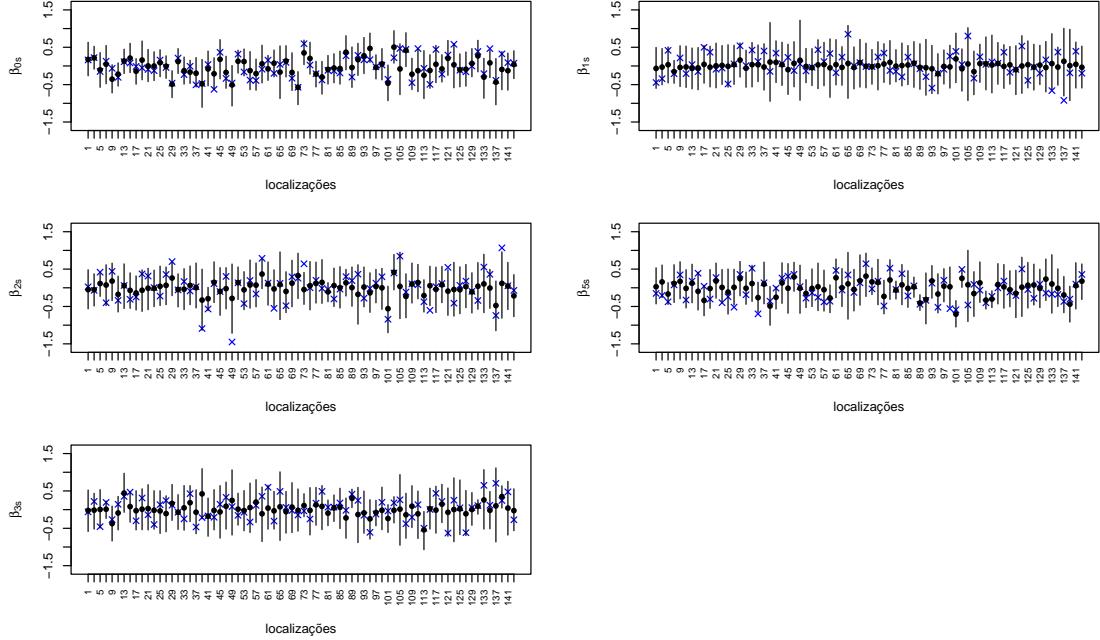


Figura 5.2: Valores estimados (pontos) e valores reais (cruzes) do intercepto e coeficientes espaciais das variáveis preditoras com respectivos intervalos de 95% de credibilidade (linhas verticais) para  $i = 1, \dots, 144$ , no modelo beta com  $\alpha \in (0, 1)$ .

A Tabela 5.1 apresenta os valores verdadeiros, as médias *a posteriori* e os intervalos de 95% de credibilidade para os parâmetros do modelo. Observa-se que as estimativas *a posteriori* são, em sua maioria, próximas dos valores verdadeiros, o que indica um bom desempenho do modelo. No entanto, vale destacar que o intervalo de credibilidade para o parâmetro  $\phi$  sugere uma maior incerteza em torno dessa estimativa, o que pode indicar um ajuste menos preciso em comparação com os demais parâmetros.

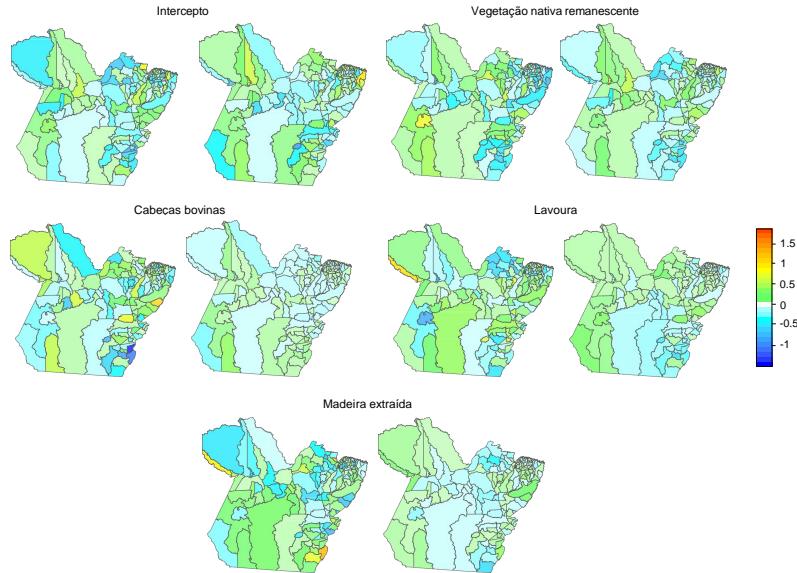


Figura 5.3: Valores verdadeiros do intercepto e dos coeficientes espaciais das variáveis preditoras para o modelo beta com  $\alpha \in (0, 1)$ .

Tabela 5.1: Resumo das distribuições *a posteriori* de  $\sigma_\eta^2$ ,  $\sigma_\varepsilon^2$ ,  $\sigma_\beta^2$ ,  $\alpha$  e  $\phi$  no modelo beta.

Parâmetros	Valor verdadeiro	Média <i>a posteriori</i>	Intervalo de credibilidade
$\sigma_\eta^2$	0,10	0,107	(0,063; 0,189)
$\sigma_\varepsilon^2$	0,30	0,315	(0,258 ; 0,373)
$\sigma_\beta^2$	0,42	0,431	(0,265 ; 0,626)
$\alpha$	0,65	0,705	(0,401; 0,971)
$\phi$	28	27,241	(22,089 ; 33,479)

A Figura 5.4 exibe os traços das cadeias do MCMC dos parâmetros, acompanhados de seus histogramas. A linha vermelha representa o valor verdadeiro do parâmetro, permitindo avaliar a qualidade das estimativas. Complementando essa análise, no Apêndice C.1, apresentam-se os traços de alguns coeficientes temporais  $\beta_t$  e espaciais  $\beta$ , ilustrando seu comportamento amostral.

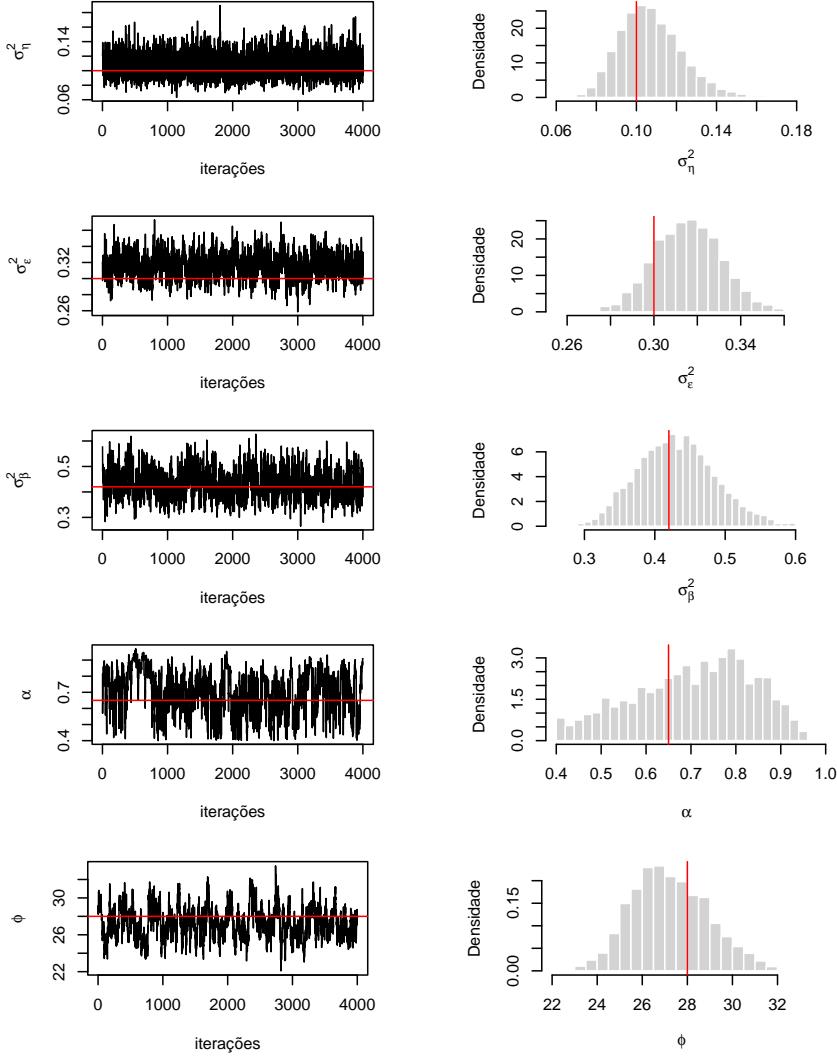


Figura 5.4: Traços das cadeias e histogramas das distribuições *a posteriori* de  $\sigma_\eta^2$ ,  $\sigma_\epsilon^2$ ,  $\sigma_\beta^2$ ,  $\alpha$  e  $\phi$  do modelo beta. As linhas vermelhas horizontal e vertical indicam o valor verdadeiro.

### 5.3.2 Modelo considerando $\alpha = 1$

Os parâmetros do modelo com  $\alpha = 1$  foram fixados em  $\sigma_\eta^2 = 0,05$ ,  $\sigma_\epsilon^2 = 0,44$ ,  $\sigma_\beta^2 = 0,38$  e  $\phi = 24$ . As distribuições *a priori* utilizadas são as mesmas do modelo com  $\alpha \in (0, 1)$ . Além disso, a restrição imposta é que a soma dos valores de  $\beta_j(s_i)$ , para cada  $j = 1, \dots, p$ , seja igual a zero.

A Figura 5.5 apresenta uma comparação entre os valores verdadeiros simulados e a média *a posteriori* do intercepto e dos coeficientes das variáveis preditoras com respectivos intervalos 95% credibilidade ao longo dos anos. Observa-se de forma geral que as médias *a posteriori* estão próximas dos valores verdadeiros, embora em alguns

casos as estimativas estejam um pouco mais afastadas.

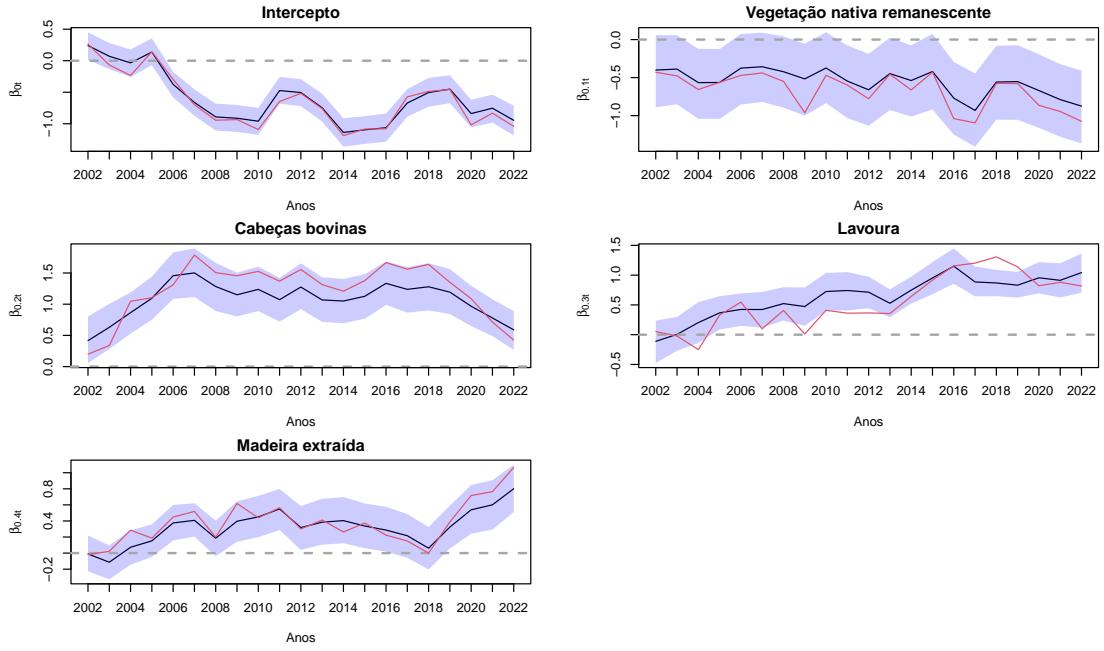


Figura 5.5: Valores estimados através da média *a posteriori* (linha preta) e valores verdadeiros (linha vermelha) com respectivos intervalos de 95% de credibilidade (área hachurada) dos coeficientes temporais ao longo do tempo, para o modelo beta com  $\alpha = 1$ .

A Figura 5.6 compara os valores estimados com os valores verdadeiros, destacando as incertezas associadas às estimativas. Observa-se que, em alguns casos, os valores estimados estão mais distantes dos valores verdadeiros. Já a Figura 5.7 apresenta os mesmos valores, mas agora visualizados em mapas coropléticos, o que possibilita uma análise espacial das discrepâncias entre os valores estimados nos diferentes municípios.

A Tabela 5.2 fornece um resumo das amostras *a posteriori*, incluindo as médias e os intervalos de 95% de credibilidade para os parâmetros do modelo, bem como os respectivos valores verdadeiros.

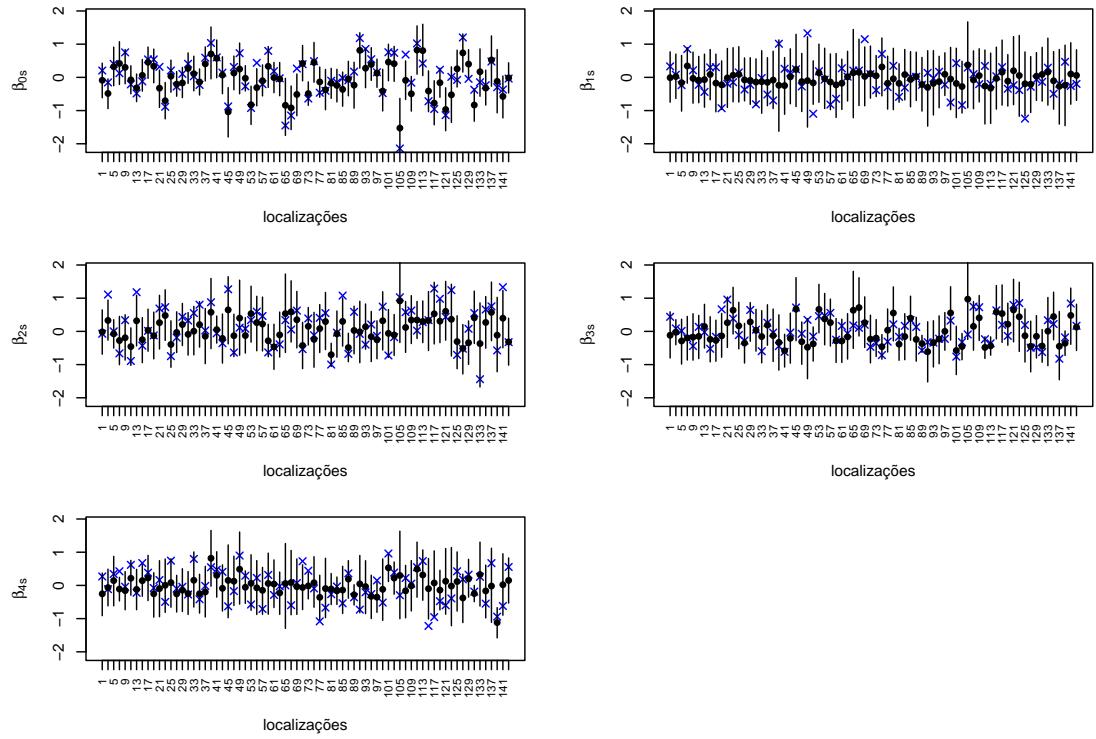


Figura 5.6: Valores estimados (pontos) e valores reais (cruzes) do intercepto e coeficientes espaciais das variáveis preditoras com respectivos intervalos de 95% de credibilidade (linhas verticais) para  $i = 1, \dots, 144$ , para o modelo normal com  $\alpha = 1$ .

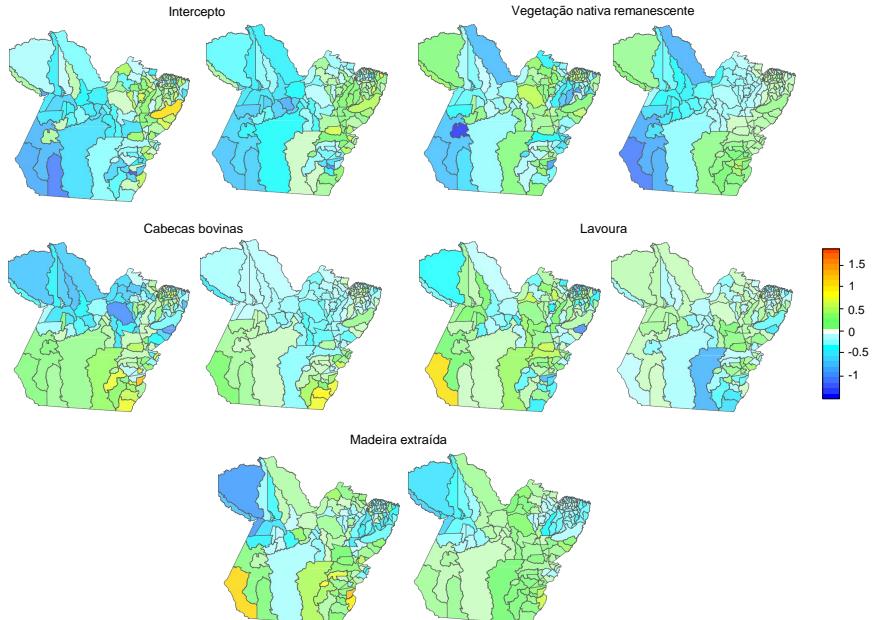


Figura 5.7: Valores verdadeiros e estimados do intercepto e dos coeficientes espaciais das variáveis preditoras para todos os municípios, para o modelo normal com  $\alpha = 1$ .

Tabela 5.2: Resumo das distribuições *a posteriori* de  $\sigma_\eta^2$ ,  $\sigma_\varepsilon^2$ ,  $\sigma_\beta^2$  e  $\phi$  do modelo beta.

Parâmetros	Valor verdadeiro	Média <i>a posteriori</i>	Intervalo de credibilidade
$\sigma_\eta^2$	0,05	0,045	(0,024; 0,088)
$\sigma_\varepsilon^2$	0,46	0,478	(0,399; 0,594)
$\sigma_\beta^2$	0,51	0,508	(0,302 ; 0,857)
$\phi$	23	21,46	(16,677 ; 28,055)

As cadeias e os histogramas das distribuições *a posteriori* dos parâmetros estão apresentados na Figura 5.8. No Apêndice C.2, são mostrados os traços de alguns coeficientes temporais  $\beta_t$  e espaciais  $\beta$ , ilustrando o comportamento dessas amostras.

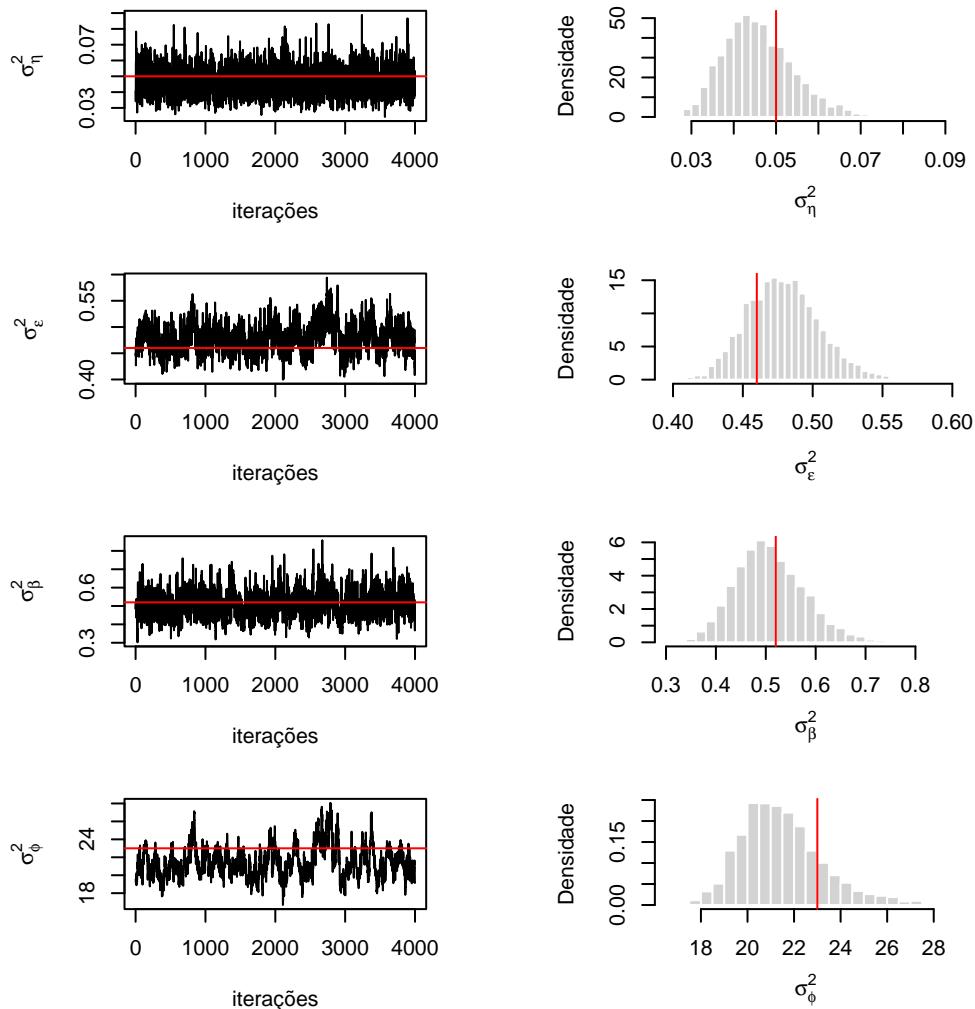


Figura 5.8: Traços das cadeias e histogramas das distribuições *a posteriori* de  $\sigma_\eta^2$ ,  $\sigma_\varepsilon^2$ ,  $\sigma_\beta^2$  e  $\phi$  para o modelo beta.

# Capítulo 6

## Modelagem espaço-temporal do desmatamento com regressão beta

Os dados analisados correspondem à taxa dada por **taxa** = (Área desmatada / Área total do município). Em alguns anos, a variável taxa apresentava valores iguais a zero, os quais foram substituídos por 0,00001 para garantir a adequação ao modelo. Além disso, os valores foram normalizados utilizando a transformação Min-Max, dada por:

$$y = \frac{\text{taxa} - \min(\text{taxa})}{\max(\text{taxa}) - \min(\text{taxa})}$$

A normalização foi realizada devido à dispersão heterogênea, com a maior parte dos valores concentrados próximos de zero. Embora a variável resposta já esteja no intervalo (0,1), a normalização foi aplicada com o objetivo de homogeneizar a variabilidade dos dados, criando um intervalo de amplitude similar onde todos os valores da variável irão residir.

### 6.1 Análise para o estado do Pará

Nesta seção, aplica-se o modelo de regressão beta aos mesmos dados reais do estado do Pará analisados no Capítulo 4.

Nesta aplicação, foram consideradas 12.000 iterações do algoritmo MCMC com aquecimento de 4.000 e espaçamento de 2 iterações, resultando numa amostra de tamanho 4.000.

### 6.1.1 Modelo considerando $\alpha \in (0, 1)$

A Figura 6.1 ilustra a evolução dos coeficientes temporais, mostrando as estimativas pontuais obtidas a partir da média *a posteriori*, juntamente com os respectivos intervalos de 95% de credibilidade.

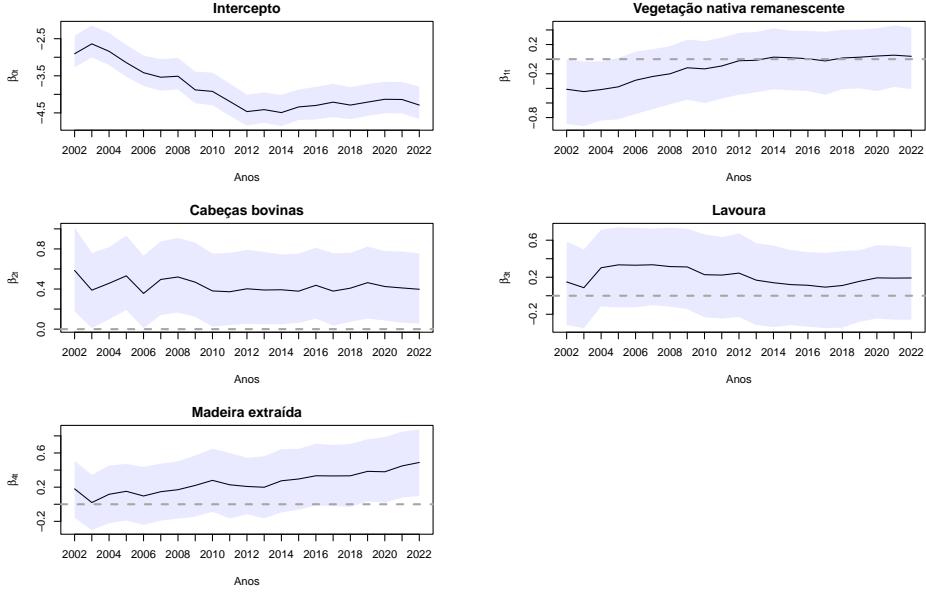


Figura 6.1: Valores estimados através da média *a posteriori* (linha preta) com respectivos intervalos 95% credibilidade (área hachurada) dos coeficientes temporais ao longo dos anos, para o modelo beta com  $\alpha \in (0, 1)$

A Figura 6.2 apresenta mapas coropléticos dos valores estimados pela média *a posteriori* para o intercepto e os coeficientes espaciais das variáveis preditoras no modelo beta. Observa-se que a influência das variáveis não é homogênea entre os municípios. A variável vegetação nativa remanescente apresenta coeficientes predominantemente negativos, enquanto a variável cabeças bovinas apresenta coeficientes positivos, especialmente nos municípios do norte. Já a variável madeira extraída apresenta coeficientes positivos, particularmente nos municípios da região sudoeste.

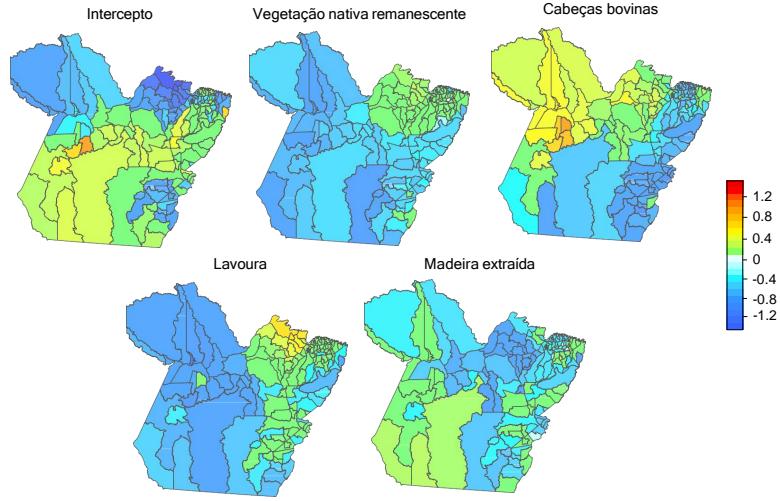


Figura 6.2: Valores estimados através da média *a posteriori* para o intercepto e os coeficientes espaciais das variáveis preditoras para todos os municípios no modelo beta com  $\alpha \in (0, 1)$ .

A Tabela 6.1 apresenta o resumo da distribuição *a posteriori* considerando  $\alpha \in (0, 1)$ . São fornecidas as estimativas dos parâmetros obtidas a partir da média *a posteriori*.

Tabela 6.1: Resumo da distribuição *a posteriori* dos parâmetros  $\sigma_\eta^2$ ,  $\sigma_\varepsilon^2$ ,  $\sigma_\beta^2$ ,  $\alpha$  e  $\phi$  no modelo beta.

Parâmetros	Média <i>a posteriori</i>	Intervalo de credibilidade
$\sigma_\eta^2$	0,024	(0,012 ; 0,045)
$\sigma_\varepsilon^2$	0,329	(0,278 ; 0,384)
$\sigma_\beta^2$	0,554	(0,345 ; 0,905)
$\phi$	61,360	(53,274 ; 69,721)
$\alpha$	0,98	(0,916 ; 0,989)

A Figura 6.3 apresenta os traços das cadeias e os histogramas das distribuições *a posteriori* dos parâmetros. Observa-se que a cadeia do parâmetro  $\alpha$  apresenta muita autocorrelação, com oscilações e quedas abruptas ao longo das iterações. Além disso, a maioria dos valores está concentrada no intervalo entre 0,96 e 0,99.

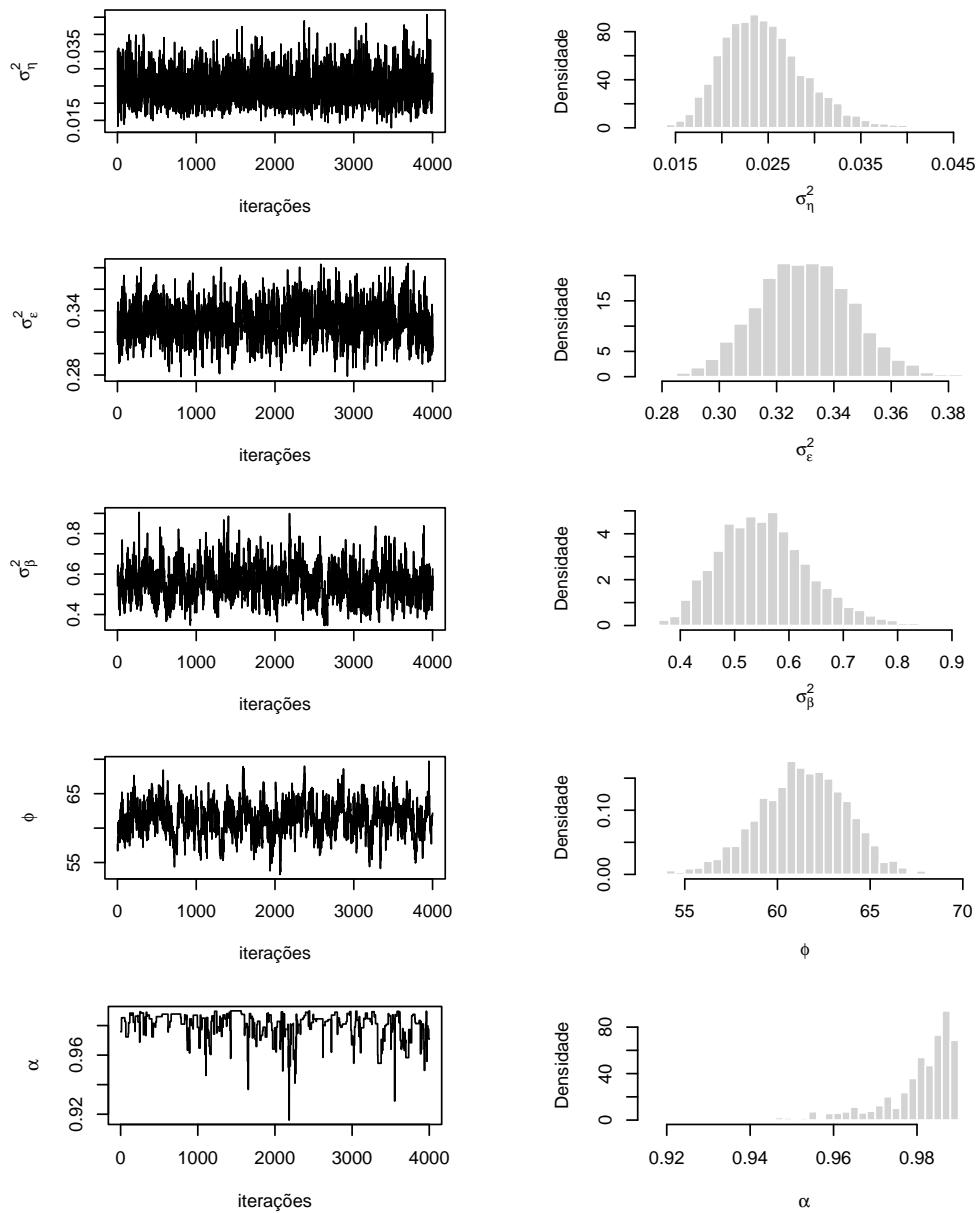


Figura 6.3: Traços das cadeias e histogramas das distribuições *a posteriori* de  $\sigma_\eta^2$ ,  $\sigma_\epsilon^2$ ,  $\sigma_\beta^2$  e  $\phi$  no modelo beta em dados reais.

### 6.1.2 Modelo considerando $\alpha = 1$

A Figura 6.4 apresenta as médias *a posteriori* e os intervalos de 95% de credibilidade do intercepto temporal e dos coeficientes temporais das variáveis preditoras ao longo do tempo.

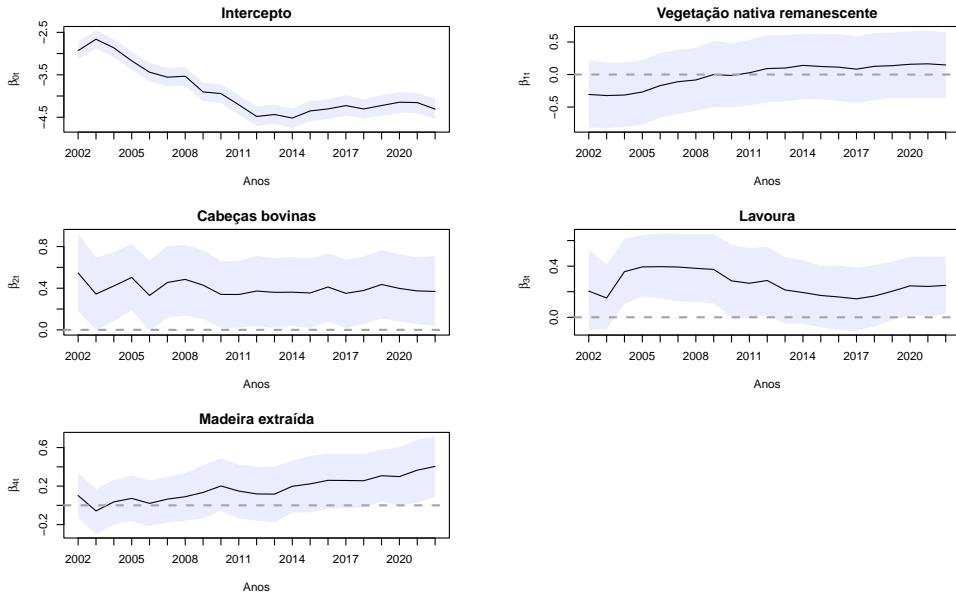


Figura 6.4: Valores estimados através da média *a posteriori* (linha preta) com respectivos intervalos 95% credibilidade (área hachurada) dos coeficientes temporais ao longo dos anos, para o modelo beta com  $\alpha = 1$

A Figura 6.5 apresenta os valores estimados no espaço para o intercepto espacial e os coeficientes espaciais das variáveis preditoras. Observa-se, por exemplo, que para a variável vegetação nativa remanescente, a maioria dos municípios apresenta coeficientes negativos. A variável cabeças bovinas apresenta coeficientes positivos na região norte, enquanto madeira extraída apresenta coeficientes positivos na região oeste dessa mesma área.

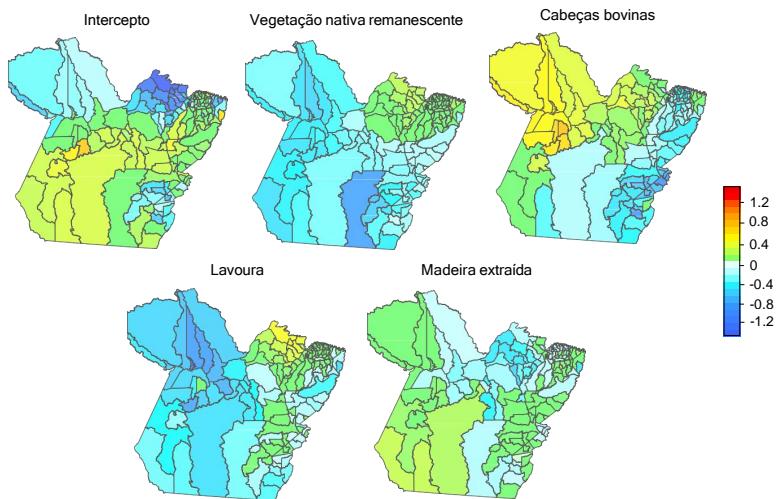


Figura 6.5: Valores estimados através da média *a posteriori* para o intercepto e os coeficientes espaciais das variáveis preditoras para todos os municípios no modelo beta com  $\alpha = 1$ .

A Tabela 6.2 apresenta o resumo da distribuição *a posteriori* para os dados reais do Pará. São fornecidos as estimativas da média *a posteriori* e seus intervalos de 95% de credibilidade.

Tabela 6.2: Resumo das distribuições *a posteriori* de  $\sigma_\eta^2$ ,  $\sigma_\varepsilon^2$ ,  $\sigma_\beta^2$  e  $\phi$  para o modelo beta com  $\alpha = 1$ .

Parâmetros	Média <i>a posteriori</i>	Intervalo de credibilidade
$\sigma_\eta^2$	0,024	(0,013 ; 0,045)
$\sigma_\varepsilon^2$	0,328	(0,276 ; 0,411)
$\sigma_\beta^2$	0,491	(0,301 ; 0,947)
$\phi$	60,730	(53,507 ; 68,180)

A Figura 6.6 apresenta os traços das cadeias e histogramas dos parâmetros. Além disso, os traços das cadeias dos coeficientes temporais  $\beta_t$  e espaciais  $\beta$  podem ser vistos no Apêndice C.3

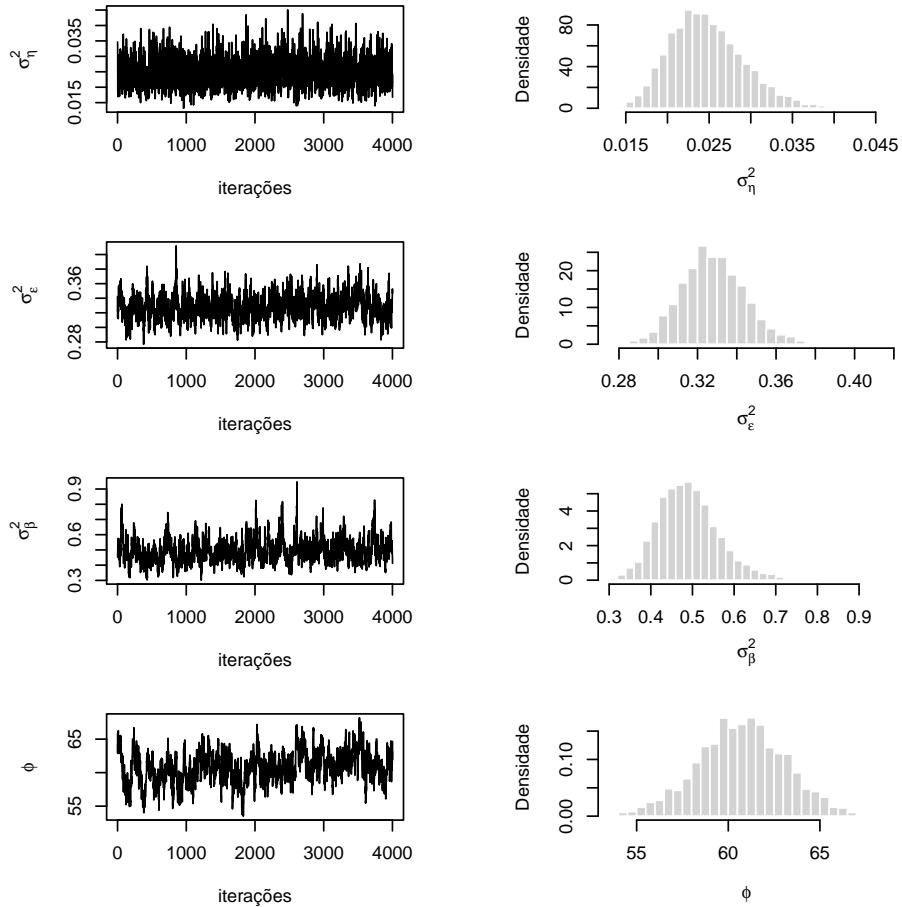


Figura 6.6: Traços das cadeias e histogramas das distribuições *a posteriori* de  $\sigma_\eta^2$ ,  $\sigma_\varepsilon^2$ ,  $\sigma_\beta^2$  e  $\phi$  no modelo beta em dados reais.

A Tabela 6.3 apresenta a comparação entre os modelos beta considerando diferentes especificações para o parâmetro  $\alpha$ . O modelo com  $\alpha = 1$  se destaca por ser mais simples e eficiente, e o critério WAIC indica que ele proporciona o melhor ajuste. Além disso, no modelo com  $\alpha \in (0, 1)$ , as estimativas de  $\alpha$  tendem a se aproximar de 1, reforçando a escolha do modelo com  $\alpha = 1$  como o mais adequado. A Figura 6.3 mostra os gráficos de dispersão entre os valores replicados e observados sugerem que ambos os modelos capturam bem a estrutura dos dados, com uma tendência geral ao longo da diagonal.

Tabela 6.3: Critério de comparação de modelos WAIC para o modelo de regressão beta.

Modelo	lppd	pWAIC	WAIC
Modelo com $\alpha \in (0, 1)$	10994,1846	682,7699	-20622,8294
Modelo com $\alpha = 1$	10996,9361	683,4958	-20625,5411

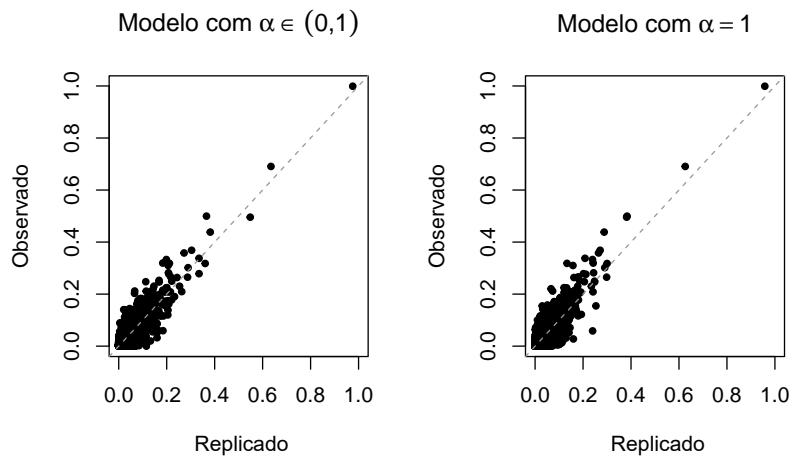


Figura 6.7: Gráfico dos valores replicados versus os valores observados para o modelo beta.

A Figura 6.8 apresenta mapas do coeficiente de regressão  $\tilde{\beta}(s, t)$  de algumas covariáveis para alguns instantes de tempo no modelo beta considerando  $\alpha = 1$ , mostrando os efeitos significativos e não significativos em diferentes anos. Os municípios estão coloridos de acordo com os efeitos das covariáveis: vermelho indica efeitos significativamente positivos, azul representa efeitos significativamente negativos, e branco mostra áreas onde não há evidências suficientes para considerar o efeito como significativo.

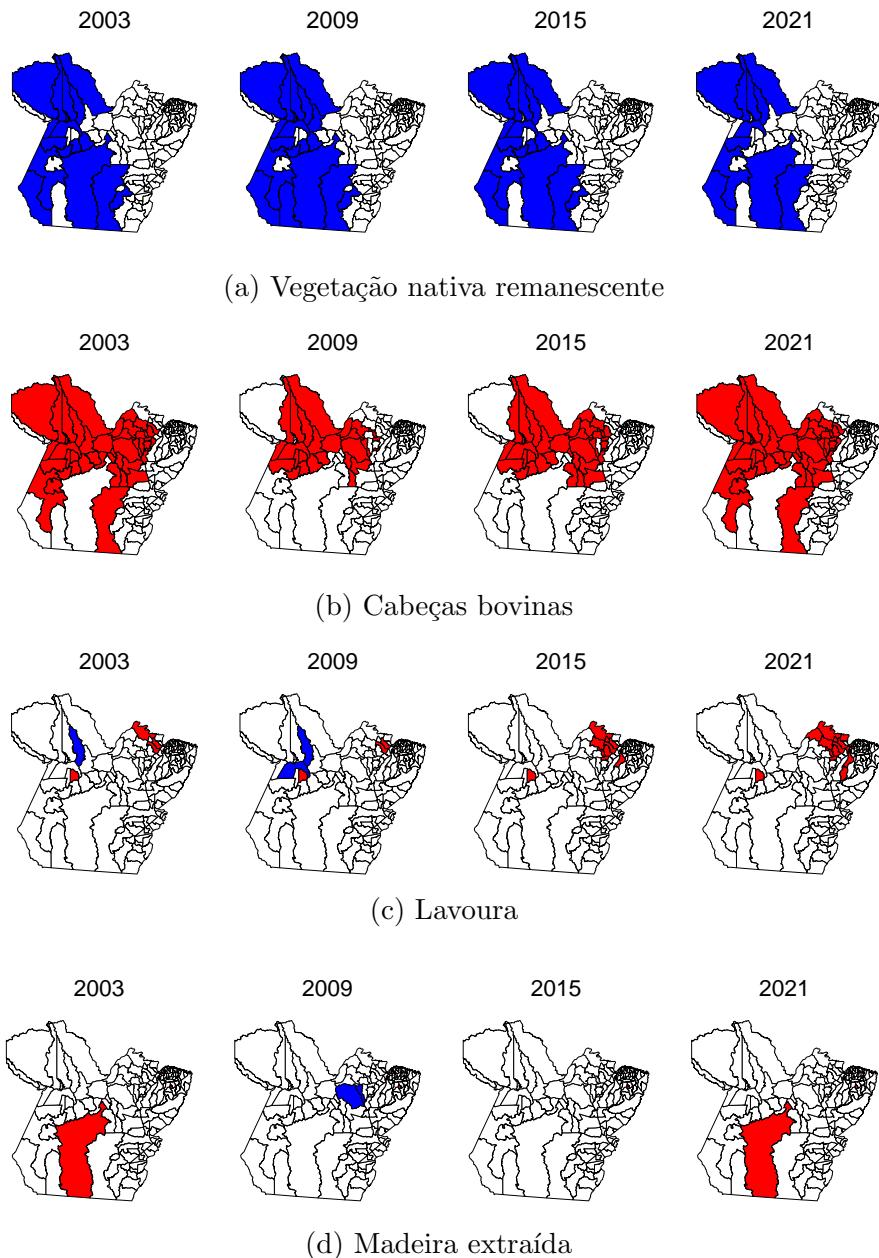


Figura 6.8: Mapas dos  $\tilde{\beta}(s, t)$  para o modelo beta com  $\alpha = 1$  com efeitos significativamente positivos (em vermelho), não significativos (em branco) e negativos (em azul) para alguns anos.

# Capítulo 7

## Conclusões e trabalhos futuros

Neste trabalho, foram desenvolvidos dois modelos espaço-temporais para analisar a dinâmica do desmatamento no Bioma Amazônico. O primeiro modelo corresponde a uma variação do modelo de regressão proposto por [Gelfand, Banerjee e Gamerman \(2005\)](#), mas para dados de área, assumindo que a variabilidade espacial é explicada por um modelo CAR. O segundo modelo é uma extensão do primeiro modelo proposto, assumindo que as observações seguem uma distribuição beta. Essas abordagens permitiram estudar a relação entre as variáveis preditoras e o desmatamento no espaço e ao longo do tempo. Em particular, foi possível entender como os efeitos das variáveis muda espaço-temporalmemente.

Os resultados mostraram que, apesar de os coeficientes de regressão não variarem muito no tempo, eles apresentam variações espaciais significativas, indicando que o impacto das variáveis preditoras não é homogêneo entre os municípios analisados. A seguir, são apresentados os principais achados:

- No estudo com dados artificiais, ambos os modelos CAR, com  $\alpha \in (0, 1)$  e  $\alpha = 1$ , recuperaram bem seus parâmetros. O modelo com  $\alpha \in (0, 1)$  tem a vantagem de ser mais flexível, capturando melhor as dependências espaciais, mas pode ser mais sensível aos dados. O modelo com  $\alpha = 1$  é mais robusto, mas menos flexível para capturar essas dependências.
- Os resultados para dados reais sugerem que o melhor modelo é com  $\alpha = 1$ , sugerindo que a estrutura CAR intrínseca representa melhor a dependência espacial no contexto analisado.
- A escolha entre o modelo normal e o modelo beta envolve um compromisso entre interpretação e complexidade computacional. O modelo normal trabalha com

uma transformação da variável resposta, facilitando a modelagem e a implementação computacional. Por outro lado, o modelo beta opera diretamente na escala original.

## 7.1 Trabalhos futuros

Nesta seção, são mencionados alguns dos trabalhos futuros, que darão continuidade ao trabalho desenvolvido nesta dissertação.

- Pretende-se revisitar a modelagem espaço-temporal inicial que não apresentou bons resultados, agora permitindo que os elementos de  $\mathbf{w}_t$  em 3.5 tenham variâncias distintas para cada  $s$ , permitindo que cada componente do processo espacial tenha sua própria variabilidade em diferentes locais.
- Aplicação a outros biomas: Aplicar a metodologia ao Cerrado, Pantanal e Mata Atlântica, que também sofrem com o desmatamento.
- Pretende-se ajustar o modelo de regressão beta sem considerar o parâmetro  $\sigma_\varepsilon^2$ . Essa simplificação do modelo, entretanto, resulta em maior complexidade no procedimento de inferência, uma vez que não será possível utilizar o algoritmo FFBS para a amostragem dos parâmetros temporais.
- Considerar uma extensão do modelo beta em que o parâmetro  $\phi$  varie no espaço e/ou no tempo, ou ainda como uma função de covariáveis, além de investigar a relação entre  $\phi$  e  $\sigma_\varepsilon^2$  para compreender melhor a variabilidade da variável resposta.
- Pretende-se propor uma adaptação eficiente do algoritmo de [Gamerman \(1997\)](#) para sortear os  $\zeta_t$ .

# Apêndice A

## Traços das cadeias dos coeficientes temporais e espaciais do modelo normal

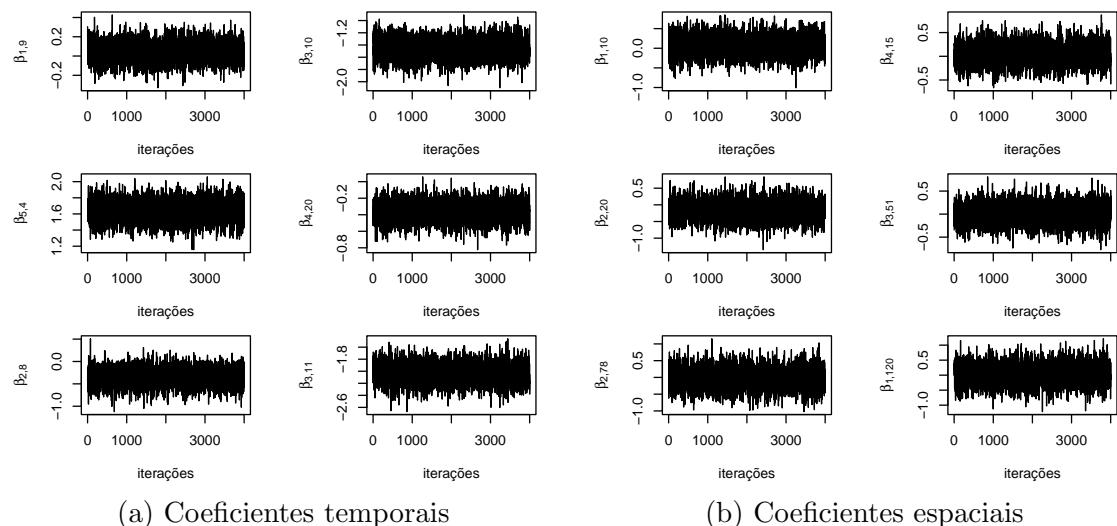
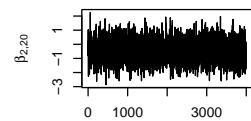
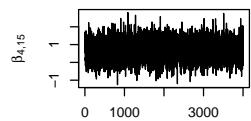
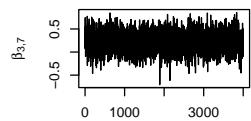
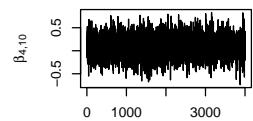
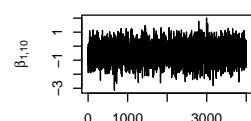
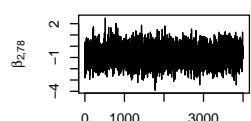
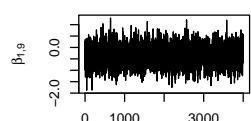
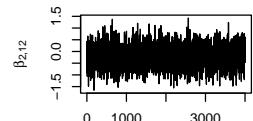
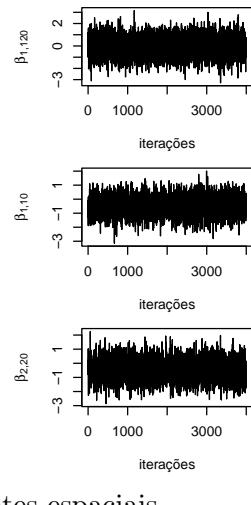
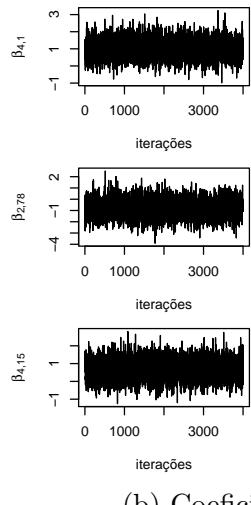
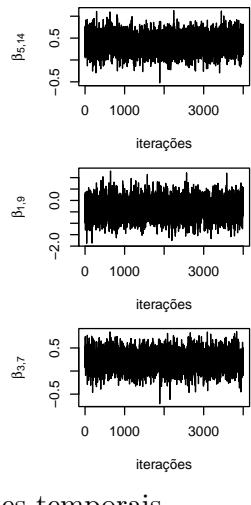
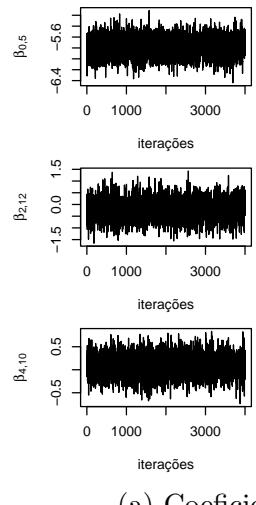


Figura A.1: Traços das cadeias de alguns coeficientes temporais  $\beta_t$  e espaciais  $\beta$  para o modelo normal com  $\alpha = 1$  em dados artificiais.



(a) Coeficientes temporais

(b) Coeficientes espaciais

Figura A.2: Traços das cadeias de alguns coeficientes temporais  $\beta_t$  e espaciais  $\beta$  para o modelo normal com  $\alpha = 1$  para dados reais.

## Apêndice B

### Critério de informação Watanabe Akaike

[Watanabe \(2010\)](#) propôs o critério de informação amplamente disponível (WAIC). WAIC realiza uma média sobre a distribuição a posteriori, em vez de se basear em uma estimativa pontoal. Conforme descrito em [Gelman, Carlin e Stern \(2014\)](#), o WAIC é calculado como:

$$WAIC = -2(lppd - pWAIC)$$

onde lppd representa a densidade preditiva logarítmica ponto a ponto, que avalia a qualidade do ajuste do modelo. O termo pWAIC atua como uma penalização para a complexidade do modelo, e é calculado como

$$\sum_{i=1}^N \sum_{t=1}^T \log \left[ \frac{1}{M} \sum_{m=1}^M p(Y(s_i, t) | \Theta^m) \right]$$

com  $\Theta$  denotando o  $m$ -ésimo valor amostrado da distribuição a posteriori, para  $m = 1, \dots, M$ . O numero efetivo de parâmetros é calculado:

$$pWAIC = \sum_{i=1}^N \sum_{t=1}^T V_{m=1}^M \log [p(Y(s_i, t) | \Theta^m)]$$

# Apêndice C

## Traços das cadeias dos coeficientes temporais e espaciais do modelo de regressão beta

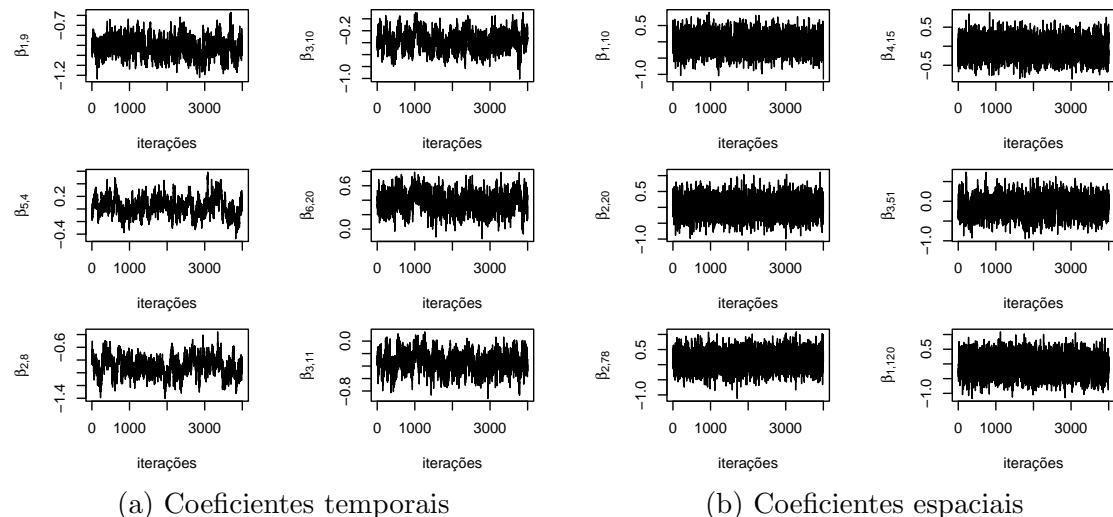


Figura C.1: Traços das cadeias de alguns coeficientes temporais  $\beta_t$  e espaciais  $\beta$  para o modelo beta com  $\alpha \in (0, 1)$  em dados artificiais.

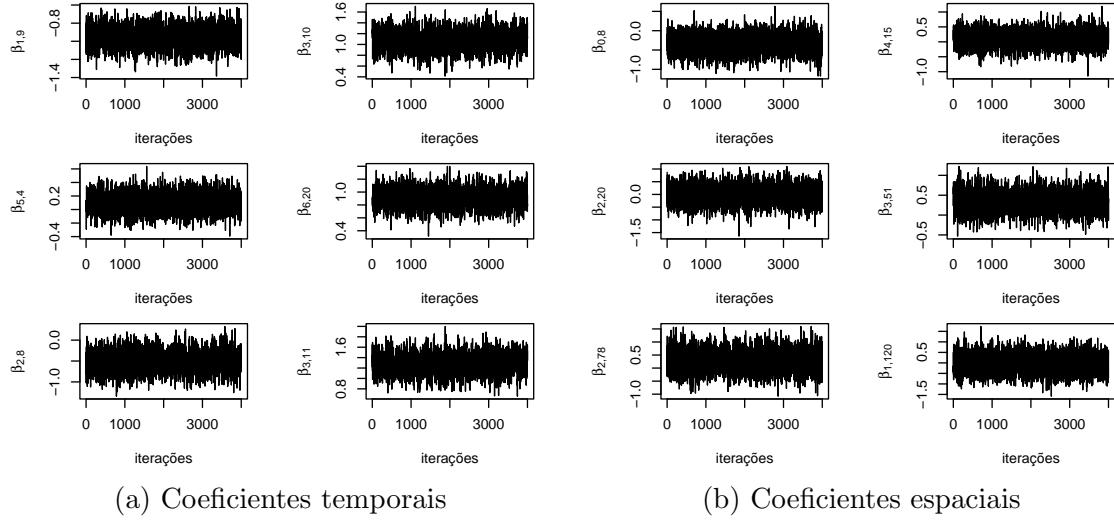


Figura C.2: Traços das cadeias de alguns coeficientes temporais  $\beta_t$  e espaciais  $\beta$  para o modelo beta com  $\alpha = 1$  em dados artificiais.

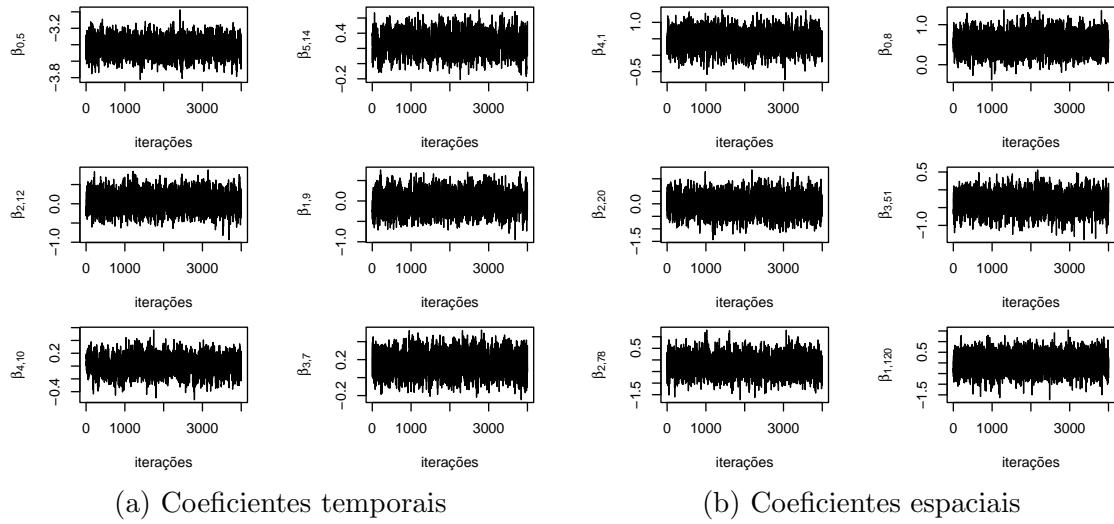


Figura C.3: Traços das cadeias de alguns coeficientes temporais  $\beta_t$  e espaciais  $\beta$  para o modelo beta com  $\alpha = 1$  para dados reais.

# Referências Bibliográficas

- Araujo, C.; Bonjean, C.; Combes, J. Property rights and deforestation in the Brazilian Amazon. *Ecological Economics*, **68**, 2009.
- Assunção, J.; Clarissa, G.; Romero, R. The Effect of Rural Credit on Deforestation: Evidence from the Brazilian Amazon. *Econ. J.*, **130**, p. 290–330, 2020.
- Banerjee, S.; Bradley, P. C.; Gelfand, A. E. *Hierarchical Modeling and Analysis for Spatial Data*. New York: Chapman & Hall, 2014.
- Barona, E.; Navin, R.; Glenn, H. The role of pasture and soybean in deforestation of the Brazilian Amazon. *Environ. Res. Lett.*, **5**, 2010.
- BCB. *Banco Central do Brasil*. 2023. Disponível em: <https://www.bcb.gov.br/estabilidadefinanceira/micrrural>.
- Besag, J. Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society: Series B (Methodological)*, **36**, p. 192–225, 1974.
- Brunsdon, C.; Fotheringham, C.; Charlton, M. Geographically Weighted Regression-Modelling Spatial Non-Stationarity. *The Statistician*, **47**, p. 431–443, 1998.
- Carter, C.; Kohn, R. On gibbs sampling for state space models. *Biometrika*, **81**, p. 541–553, 1994.
- Correia, F.; Manzi, A.; Cândido, L. Balanço de umidade na Amazônia e sua sensibilidade às mudanças na cobertura vegetal. *Mudanças climáticas*, **59**, p. 39–43, 2007.
- Domingues, M.; Bermann, C. O arco de desflorestamento na Amazônia: da pecuária à soja. *Ambient. soc.*, 2012.

Ellwanger, J.; Kulmann-Leal, B.; Kaminski, V. Beyond diversity loss and climate change: Impacts of Amazon deforestation on infectious diseases and public health. *Anais da Academia Brasileira de Ciências*, **92**, 2020.

Fearnside, P. M. Deforestation in Brazilian Amazonia: History, rates, and consequences. *Conservation Biology*, **19**, p. 680–688, 2005.

Ferrari, S.; Cribari-Neto, F. Beta Regression for Modelling Rates and Proportions. *Journal of Applied Statistics*, **53**, p. 799–815, 2004.

Fotheringham, A.; Crespo, R.; Yao, J. Geographical and Temporal Weighted Regression GTWR. *Geographical Analysis*, **47**, p. 431–452, 2015.

Frühwirth-Schnatter, S. Data augmentation and dynamic linear models. *Journal of Time Series Analysis*, **15**, p. 183–202, 1994.

Gamerman, D. Sampling from the posterior distribution in generalized linear mixed models. *Journal of Applied Statistics*, **7**, p. 57–68, 1997.

Gamerman, D.; Lopes, H. *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*. New York: Chapman & Hall, 2006.

Geist, H. J.; Lambin, E. F. What drives tropical deforestation? a meta-analysis of proximate and underlying causes of deforestation based on subnational case study evidence. 2001.

Gelfand, A.; Kim, H.-J.; Sirmans, C. Spatial modeling with spatially varying coefficient processes. *Journal of the American Statistical Association*, **98**, 2003.

Gelfand, A. E.; Banerjee, S.; Gamerman, D. Spatial process modelling for univariate and multivariate dynamic spatial data. *Environmetrics*, **16**, p. 465–479, 2005.

Gelman, A.; Carlin, J. B.; Stern, H. S. *Bayesian Data Analysis*. New York: Chapman and Hall-CRC, 3rd ed, 2014.

Guimarães, P.; Santos, A.; Rudke, A. Análise espacial dos fatores condicionantes do desmatamento do Bioma Mata Atlântica/Brasil. *Revista Geografias*, **19**, 2023.

IBGE. *Biomas e Sistema Costeiro-Marinho do Brasil*. 2019. Disponível em: [⟨https://biblioteca.ibge.gov.br/visualizacao/livros/liv101676.pdf⟩](https://biblioteca.ibge.gov.br/visualizacao/livros/liv101676.pdf).

IBGE. *Sistema IBGE de Recuperação Automática*. 2023. Disponível em: [⟨https://sidra.ibge.gov.br/home/pms/brasil⟩](https://sidra.ibge.gov.br/home/pms/brasil).

INPE. *Inpe. Projeto PRODES*. 2023. Disponível em: [⟨https://terrabrasilis.dpi.inpe.br/app/dashboard/deforestation/biomes/amazon/increments⟩](https://terrabrasilis.dpi.inpe.br/app/dashboard/deforestation/biomes/amazon/increments).

Jin, X.; Carlin, B.; Banerjee, S. Generalized hierarchical multivariate car models for areal data. *Biometrics*, **61**, p. 950–961, 2005.

Malhi, Y.; Roberts, J. T.; Betts, R. Climate change, deforestation, and the fate of the Amazon. *Science*, **319**, 2008.

MapBiomass. *MapBiomass Brasil*. 2023. Disponível em: [⟨https://brasil.mapbiomas.org/⟩](https://brasil.mapbiomas.org/).

Nepstad, D.; Mccgrath, D.; Stickler, C. Slowing Amazon deforestation through public policy and interventions in beef and soy supply chains. *Science*, **344**, p. 1118–1123, 2014.

Paez, M.; Gamerman, D.; Landim, F. Spatially varying dynamic coefficient models. *Journal of Statistical Planning and Inference*, **138**, p. 1038–1058, 2007.

R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria, 2022. Disponível em: [⟨https://www.R-project.org/⟩](https://www.R-project.org/).

Santos, A.; Assunção, C.; Monteiro, P. Deforestation drivers in the Brazilian Amazon: Assessing new spatial predictors. *Journal of Environmental Management*, **294**, 2021.

Schmidt, A. M.; Nobre, W. Conditional Autoregressive (CAR) Model. *Wiley StatsRef: Statistics Reference Online*, 2014.

Vatcheva, K.; Lee, M.; McCormick, J. Multicollinearity in regression analyses conducted in epidemiologic studies. *Epidemiology (Sunnyvale)*, **6(2)**, 2016.

Watanabe, S. Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. *J. Mach. Learn. Res*, **11**, p. 3571–3594, 2010.

West, M.; Harrison, J. *Bayesian Forecasting and Dynamic Models*. New York: Springer, 2nd edn, 1997.