UNICAMP

UNIVERSIDADE ESTADUAL DE CAMPINAS

Instituto de Matemática, Estatística e Computação Científica

DAVID HENRIQUES DA MATTA

Modelo Funcional Espaço-Temporal para Dados com Estrutura de Blocos com Medidas Repetidas: uma abordagem Bayesiana

Campinas

2022

David Henriques da Matta

Modelo Funcional Espaço-Temporal para Dados com Estrutura de Blocos com Medidas Repetidas: uma abordagem Bayesiana

Tese apresentada ao Instituto de Matemática, Estatística e Computação Científica da Universidade Estadual de Campinas como parte dos requisitos exigidos para a obtenção do título de Doutor em Estatística.

Orientadora: Nancy Lopes Garcia

Coorientadora: Mariana Rodrigues Motta

Este trabalho corresponde à versão final da Tese defendida pelo aluno David Henriques da Matta e orientada pela Profa. Dra. Nancy Lopes Garcia.

Campinas

2022

Ficha catalográfica
Universidade Estadual de Campinas
Biblioteca do Instituto de Matemática, Estatística e Computação Científica
Ana Regina Machado - CRB 8/5467

M429m
Matta, David Henriques da, 1979-
Modelo funcional espaço-temporal para dados com estrutura de blocos com medidas repetidas : uma abordagem bayesiana / David Henriques da Matta. – Campinas, SP : [s.n.], 2022.

Orientador: Nancy Lopes Garcia.
Coorientador: Mariana Rodrigues Motta.
Tese (doutorado) – Universidade Estadual de Campinas, Instituto de Matemática, Estatística e Computação Científica.

1. Teoria do spline. 2. Análise multivariada. 3. Análise espacial (Estatística). 4. Inferência bayesiana. 5. Métodos MCMC (Estatística). I. Garcia, Nancy Lopes, 1964-. II. Motta, Mariana Rodrigues, 1975-. III. Universidade Estadual de Campinas. Instituto de Matemática, Estatística e Computação Científica. IV. Título.

Informações Complementares

Título em outro idioma: Spatio-temporal regression model with block structure with repeated measures : a Bayesian approach
Palavras-chave em inglês:
Spline theory
Multivariate analysis
Spatial analysis (Statistics)
Bayesian inference
MCMC methods (Statistics)
Área de concentração: Estatística
Titulação: Doutor em Estatística
Banca examinadora:
Mariana Rodrigues Motta [Coorientador]
Guilherme Vieira Nunes Ludwig
Dani Gamerman
Márcio Poletti Laurini
Clarice Garcia Borges Demétrio
Data de defesa: 28-11-2022
Programa de Pós-Graduação: Estatística

Identificação e informações acadêmicas do(a) aluno(a)
- ORCID do autor: https://orcid.org/0000-0003-0199-2075
- Currículo Lattes do autor: http://lattes.cnpq.br/0274497958438413

Tese de Doutorado defendida em 28 de novembro de 2022 e aprovada pela banca examinadora composta pelos Profs. Drs.

Prof(a). Dr(a). MARIANA RODRIGUES MOTTA

Prof(a). Dr(a). GUILHERME VIEIRA NUNES LUDWIG

Prof(a). Dr(a). DANI GAMERMAN

Prof(a). Dr(a). MÁRCIO POLETTI LAURINI

Prof(a). Dr(a). CLARICE GARCIA BORGES DEMÉTRIO

A Ata da Defesa, assinada pelos membros da Comissão Examinadora, consta no SIGA/Sistema de Fluxo de Dissertação/Tese e na Secretaria de Pós-Graduação do Instituto de Matemática, Estatística e Computação Científica.

Este trabalho é dedicado a Deus por encaminhar minha vida.
A minha família pelo exemplo, incentivo e amor.

Agradecimentos

Gostaria de agradecer primeiramente a Deus e ao meu mentor espiritual que tanto me ajudaram em mais esta etapa.

Agradeço à Profa. Dra. Nancy Lopes Garcia, minha orientadora, por me receber como aluno de doutorado. Obrigado pela confiança no meu trabalho, pelo respeito e por me ensinar. Sua dedicação, paciência e ensinamentos serviram como pilares de sustentação não somente para a conclusão deste trabalho, como também para a consolidação da minha carreira acadêmica. Sinto-me privilegiado por ter sido seu orientado tanto de mestrado quanto de doutorado.

Agradeço à minha coorientadora, Profa. Dra. Mariana Rodrigues Motta, pela confiança, pela paciência e por prontamente me ajudar sempre que a procurei. Obrigado por me manter motivado durante todo o processo. Sua dedicação e conhecimento foram fundamentais para a conclusão deste projeto.

Agradeço aos meus pais, Teresinha de Almeida da Matta e Ruy Henriques da Matta Filho (*in memoriam*), que juntos enfrentaram tantas dificuldades para que eu pudesse estudar. Obrigado à minha família, por compreender as minhas ausências por causa do desenvolvimento deste trabalho. Finalmente, gostaria de agradecer aos meus amigos pelo apoio nos momentos que mais precisei.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

"A persistência é o menor caminho do êxito".
(Charles Chaplin)

Resumo

A análise de dados espaço-temporais têm sido objeto de pesquisas em diversas áreas do conhecimento. Um dos principais objetivos de tais pesquisas é a necessidade de avaliar o comportamento dos efeitos climáticos em determinadas regiões ao longo de um período de tempo. Quando determinados padrões climáticos atuam por vários dias ou até mesmo semanas, fazendo com que as áreas por eles afetadas tenham o mesmo tipo de clima por um longo período de tempo, o uso de blocos para esses fenômenos pode ser uma boa estratégia. Além disso, ter medidas repetidas para observações dentro de blocos ajuda a controlar as diferenças entre as observações, ganhando assim uma maior sensibilidade. Diante de tais perspectivas, este trabalho apresenta um modelo de regressão espaço-temporal com estrutura de blocos com medidas repetidas incorporando como preditores variáveis funcionais de natureza fixa e aleatória. Para acomodar estruturas complexas espaciais, temporais e de blocos, foram considerados componentes funcionais baseados em efeitos aleatórios, além da estrutura de covariância de classe *Matérn*, responsável por contabilizar a covariância espacial. Esta tese é motivada por um conjunto de dados de precipitação coletados ao longo do tempo (mensalmente) em diversas estações meteorológicas localizadas no Estado de Goiás no Brasil, entre os anos de 1980 e 2001 (21 anos). Neste contexto, os efeitos espaciais são representados por diferentes estações meteorológicas, o efeito temporal é representado por meses, o efeito de bloco por padrões climáticos e as medidas repetidas por anos dentro dos padrões climáticos. O modelo proposto apresentou resultados satisfatórios nos estudos de simulação, bem como quando aplicado ao problema de estimação da precipitação nos dados disponíveis.

Palavras-chave: Análise de dados funcionais, Modelos multivariados espaço-temporais, Inferência Bayesiana, Cadeias de Markov e Monte Carlo.

Abstract

The analysis of spatio-temporal data has been the object of research in several areas of knowledge. One of the main objectives of such research is the need to evaluate the behavior of climate effects in certain regions across a period of time. When certain climate patterns appear for several days or even weeks, causing the areas affected by them to have the same kind of weather for an extended period of time, the use of blocks for these phenomena may be a good strategy. Additionally, having repeated measures for observations within blocks helps to control for differences between observations, thus gaining more statistical power. In view of these perspectives, this thesis presents a spatio-temporal regression model with block structure with repeated measures incorporating as predictors functional variables of fixed and random nature. To accommodate complex spatial, temporal and block structures, functional components based on random effects were considered in addition to the class *Matérn* covariance structure, which was resposible to account for spatial covariance. This thesis is motivated by a precipitation dataset collected over time (monthly) from several meteorological stations located in Goiás State, Brazil, between the years 1980 and 2001 (21 years). In this context, spatial effects are represented by different meteorological stations, time effect are represented by months, block effect by climate patterns and repested measures by years inside climate patterns. The proposed model presented satisfactory results in the simulation studies as well as when applied for estimating precipitation in the available data.

Keywords: Functional data analysis, Multivariated spatio-temporal models, Bayesian inference, Markov Chain Monte Carlo.

Lista de ilustrações

Lista de tabelas

Lista de abreviaturas e siglas

UNICAMP	Universidade Estadual de Campinas
IMECC	Instituto de Matemática, Estatística e Computação Científica
ADF	Análise de Dados Funcionais
ENOS	El Niño Oscilação do Sul
LPML	Logaritmo da Pseudo Verossimilhança Marginal
DIC	Critério de Informação Baseado na Função Desvio
DIC7	Critério de Informação Baseado na Função Desvio Adaptado para a Presença de Efeitos Aleatórios
CPO	Ordenada Preditiva Condicional
MCMC	Cadeia de Markov e Monte Carlo
INMET	Instituto Brasileiro de Meteorologia
ANA	Agência Nacional de Águas
SIMEHGO	Sistema Meteorológico e Hídrico do Estado de Goiás
EMBRAPA	Empresa Brasileira de Pesquisa Agropecuária
SST	Temperatura da Superfície do Mar
NOAA	Administração Nacional Oceânica e Atmosférica
ONI	Índice Oceânico Niño
ASO	Meses de Agosto, Setembro e Outubro
SON	Meses de Setembro, Outubro e Novembro
OND	Meses de Outubro, Novembro e Dezembro
NDJ	Meses de Novembro, Dezembro e Janeiro
DJF	Meses de Dezembro, Janeiro e Fevereiro
JFM	Meses de Janeiro, Fevereiro e Março
FMA	Meses de Fevereiro, Março e Abril

MAM	Meses de Março, Abril e Maio
Me	Média
Med	Mediana
Mo	Moda
DP	Desvio Padrão
AR1	Estrutura de Correlação Autoregressiva de Ordem Um

Lista de símbolos

$\otimes$	Produto matricial de Kronecker.
$\mathbf{1}_i$	Matriz coluna de tamanho $i \times 1$ com todos elementos iguais a 1.
$\mathbf{I}_i$	Matriz identidade de tamanho i.
I	Inteiro positivo que representa o número de blocos observados (página 32).
J_i	Inteiro positivo que representa o número de medidas repetidas do bloco i. (página 32).
τ	Inteiro positivo que representa o número de tempos observados (página 33).
n	Inteiro positivo que representa o número de localizações observadas (página 33).
K_{μ_i}	Inteiro positivo que representa o número de bases *B-splines* cúbicos utilizadas na i-ésima dimensão da função média geral (página 32).
K_{ζ_i}	Inteiro positivo que representa o número de bases *B-splines* cúbicos utilizadas na i-ésima dimensão do efeito espacial aleatório (página 33).
K_{Γ}	Inteiro positivo que representa o número de bases *B-splines* cúbicos utilizadas no efeito temporal aleatório (página 33).
Υ_{μ_i}	Conjunto de nós interiores utilizados na i-ésima dimensão da função média geral (página 32).
Υ_{ζ_i}	Conjunto de nós interiores utilizados na i-ésima dimensão do efeito espacial aleatório (página 33).
Υ_{Γ}	Conjunto de nós interiores utilizados no efeito temporal aleatório (página 33).
$\boldsymbol{\Sigma}_{\boldsymbol{\Theta}}{}^{(i)}$	Matriz de covariância do vetor de efeitos aleatórios da i-ésima componente espacial (página 33).
$\boldsymbol{\Sigma}_{\boldsymbol{\vartheta}}{}^{(i)}$	Matriz de covariância do vetor de efeitos aleatórios da i-ésima componente temporal (página 33).

$\boldsymbol{\Sigma}_{\epsilon_i}$	Matriz de covariância dos erros aleatórios da i-ésima componente espacial (página 34).
$\boldsymbol{\xi}$	Vetor de parâmetros do modelo proposto (página 37).
LPMP	Critério de seleção de modelos baseado no logaritmo da pseudo verossimilhança marginal (página 40)
DIC	Critério de informação baseado na função desvio (página 42)
DIC7	Critério de informação baseado na função desvio adaptado para a presença de efeitos aleatórios (página 42)
$\boldsymbol{\Delta}$	Vetor de hiperparâmetros da distribuição a priori (página 65)

Sumário

1 Introdução

O desenvolvimento de modelos estatísticos capazes de prever variáveis climatológicas são fundamentais para o avanço de estudos que visam um melhor entendimento das interações entre a cultura e o ambiente. Os resultados desses modelos permitem preencher falhas em séries climatológicas e também criar séries climatológicas em sítios aonde as mesmas não são mensuradas. No avanço da agricultura, os dados climatológicos são aplicados na predição do efeito do fenômeno El Niño Oscilação Sul (ENOS) no calendário agrícola, na interação cultura e clima e ainda na interação manejo e cultura (veja, por exemplo, Santos et al. (2021) e Heinemann et al. (2022)).

De modo geral, as variáveis climatológicas são consideradas processos físicos complexos que podem variar amplamente no espaço e no tempo. Como exemplo, podemos citar a precipitação e a temperatura. No caso da precipitação, o conhecimento da sua climatologia permite avaliar riscos de plantio, bem como planejar uma correta gestão dos recursos hídricos disponíveis em uma determinada região. Neste contexto, Stauffer et al. (2017) propôs um modelo espaço-temporal, cujo objetivo é estimar a distribuição da precipitação diária a partir de uma distribuição normal censurada à esquerda.

Nas mais diversas áreas do conhecimento, diferentes métodos vem sendo utilizados para se estimar dados espaço-temporais. Por exemplo, Blangiardo et al. (2013) propôs uma abordagem por Aproximação de Laplace Aninhada Integrada (*Integrated Nested Laplace Approximation* - INLA), como uma alternativa computacional para estimar dados epidemiológicos caracterizados em uma estrutura espacial e/ou temporal. Por outro lado, Gamerman et al. (2022) propôs uma abordagem de redução da dimensionalidade espaço-temporal, a partir de equações estruturais dinâmicas, como forma de avaliar os efeitos da poluição do ar em dados de hospitalização.

Diante da extensa literatura existente para dados espaço-temporais, e também da motivação deste trabalho (dados de precipitação), daremos ênfase, daqui por diante, nas contribuições apresentadas para o estudo de dados climatológicos. Neste contexto, dentre as diversas abordagens apresentadas na literatura, Aryaputera et al. (2015) prôpos a inclusão de covaríaveis adicionais em algumas formas de *Kriging*, denotadas por *Kriging* espaço-temporais, como forma de explicar a dependência entre os diversos tipos de topografia, permitindo que informações de outras variáveis possam ajudar no processo de previsão. Veja, também, Snepvangers, Heuvelink e Huisman (2003).

Daly et al. (2008) e Thornton, Running e White (1997) avaliam padrões climáticos a partir de modelos de regressão regional, no qual para uma determinada localidade um modelo de regressão usual é ajustado, a partir de um subconjunto de estações vizinhas. Neste caso, a aplicação do método torna-se inviável, quando as covariáveis que influenciam uma determinada variável climática não estiverem disponíveis em uma localização geográfica de interesse.

Usando modelos aditivos generalizados, Lewis et al. (2006) considera funções (possivelmente não lineares) como possíveis preditores em um modelo de regressão espaço-temporal. O interessante de tal técnica é a possibilidade de representar efeitos cíclicos temporais a partir de componentes funcionais (Apêndice A) unidimensionais, além da possibilidade de descrever a distribuição espacial por meio de componentes funcionais bidimensionais (veja, por exemplo, Guan et al. (2009) e Vicente-Serrano, Saz-Sánchez e Cuadrat (2003)).

Do ponto de vista Bayesiano, Laurini (2019) propôs um modelo espaço-temporal para investigar a existência de mudanças climáticas em séries de temperatura. A estrutura do modelo proposto por Laurini (2019) consiste de uma componente de efeito fixo, descrita por covariáveis observadas no espaço e no tempo, por componentes de tendência, sazonalidade e ciclo, que evoluem dinamicamente segundo passeios aleatórios independentes e, finalmente, por uma componente aleatória espaço-temporal com função de covariância espacial pertencente à classe *Matérn*.

Existem vários motivos para se utilizarem componentes funcionais baseadas no espaço e no tempo como possíveis preditores lineares, o primeiro é que pode haver indisponibilidade de informação (covariáveis) em uma localização geográfica de interesse, o que poderia inviabilizar, por exemplo, técnicas do tipo regressão regional (DALY et al., 2008). O segundo motivo, também destacado em Guan et al. (2009), é que pode existir uma relação não linear entre a variável de interesse e as condições de espaço e tempo. O terceiro e último motivo está associado à possibilidade da inclusão de efeitos aleatórios funcionais que nos permitem incluir correlação espacial e temporal, além daquela admitida no erro aleatório do modelo (STAUFFER et al., 2017).

O objetivo deste trabalho é desenvolver um modelo de regressão espaço-temporal com estrutura de blocos com medidas repetidas, incorporando variáveis funcionais de natureza fixa e aleatória como preditoras da variável resposta. A nossa proposta é utilizar uma abordagem de Análise de Dados Funcionais (ADF), a qual é adequada para dados cuja estrutura natural é intrinsecamente funcional, o que é condizente à motivação desta tese (dados de precipitação). Para isso, considere $Y_{ij}(\mathbf{x}, t)$, $i = 1, \ldots, I$, $j = 1, \ldots, J_i$, a j-ésima medida repetida do bloco i observada no tempo $t \in D_T \subset \mathbb{N}$ e na localização $\mathbf{x} = (x_1, x_2) \in D_{lat} \times D_{long} \subset \mathbb{R}^2$. Sob esta definição, cada ponto Y_{ij} é uma observação em algum espaço de funções (RAMSAY; SILVERMAN, 1993).

Em nossa abordagem, cada observação $Y_{ij}(\mathbf{x}, t)$, $i = 1, ..., I$, $j = 1, ..., J_i$, $\mathbf{x} \in D_{lat} \times D_{long} \subset \Re^2$, $t \in D_T$, será descrita por um modelo de efeitos espaço-temporais fixos e aleatórios, que podem ser muito bem aproximados por combinações lineares de produtos tensoriais de bases *B-splines* cúbicos avaliados no tempo e no espaço (DE BOOR, 1978).

As estruturas de covariância assumidas nesse trabalho admitem que unidades observacionais tomadas em diferentes localizações e/ou tempos, serão correlacionadas segundo uma estrutura espaço-temporal somente se tais medidas pertencerem a um mesmo bloco e medida repetida. Por outro lado, devido à forma em que propomos as componentes espacias e temporais do modelo proposto, expansão em *B-splines* cúbicos (Apêndice A), acomodamos estruturas de correlação espaciais e temporais intra bloco, provenientes das formas funcionais assumidas.

Em dados climatológicos, a sazonalidade ilustra os efeitos das mudanças climáticas no ambiente de acordo com o período do ano em que se encontra, podendo variar intensamente dependendo do local em que se está situado. Neste trabalho, propomos a inclusão de uma estrutura de bloco como forma de resgatar a sazonalidade presente nos dados de precipitação, o que em Laurini (2019) foi resgatada somente a partir de efeitos aleatórios. No mais, também adotaremos o ponto de vista Bayesiano para a estimação dos parâmetros, assim como Laurini (2019).

O modelo proposto é motivado por um conjunto de dados de precipitação que foram registrados ao longo do tempo (mensalmente) em diversas estações meteorológicas localizadas no Estado de Goiás no Brasil, durante o período compreendido entre os anos de 1980 a 2001 (21 anos). As distâncias observadas entre as estações meteorológicas avaliadas neste trabalho são inferiores a 790.9 km, no qual 50% das distâncias concentram-se entre 153.1 km e 348.5 km.

Consideramos, nesse estudo, que as observações são medidas entre os meses de outubro a março, estação das chuvas no Estado de Goiás, e estão distribuídas entre os blocos de efeitos climáticos El Niño, La Niña e Neutro, conforme descrito em NOAA (2019). Logo, os anos observados sob um mesmo efeito climático serão considerados medidas repetidas dentro daquele bloco. Em outras palavras, as observações são coletadas em diversos intervalos de tempo (mensalmente), em uma certa região (superfície), sob possível efeito de bloco.

O interessante em tal estrutura de dados é que, apesar de as observações serem medidas somente em algumas coordenadas geográficas ao longo do tempo e do espaço, a variável resposta a ser modelada é de natureza estritamente contínua no tempo e no espaço, tal que os modelos multivariados não são adequados à análise de tais conjuntos de dados. É importante observar que a terminologia funcional é referência à estrutura pressuposta para os dados, que na prática são coletados de forma discreta.

A seguir, descrevemos a estrutura desta tese. No Capítulo 2, exibimos o modelo proposto neste trabalho considerando um modelo com resposta e covariáveis funcionais no espaço-tempo. No Capítulo 3, desenvolvemos inferência Bayesiana para o modelo proposto. No Capítulo 4, descrevemos um critério baseado no logaritmo da pseudo verossimilhança marginal (LPML), proposto por Geisser (1993) e por Geisser e Eddy (1979), e um outro critério derivado a partir da função desvio (DIC), adaptado para o caso de presença de efeitos aleatórios (CELEUX et al., 2006), que foram utilizados como uma ferramenta para resgatar o verdadeiro valor de uma das componentes de variância espacial do modelo proposto, a dizer o parâmetro κ que compõe a estrutura de correlação *Matérn* (CRESSIE; HUANG, 1999).

No Capítulo 5, trazemos os resultados de um intenso conjunto de simulações com o intuito de avaliar o desempenho do modelo proposto quanto às seguintes perspectivas: sensibilidade da distribuição a priori das componentes de variância, efeito do tamanho amostral e do número de medidas repetidas no processo de estimação, determinação do parâmetro κ, a capacidade preditiva do modelo proposto e, finalmente, a adequabilidade da estimação empírica dos parâmetros em espaços ortogonais. Destacamos que, dentre as dificuldades encontradas na construção do algoritmo de estimação, escrito nas primeiras versões exclusivamente em linguagem *R* (TEAM, 2013), o tempo de processamento foi o grande entrave identificado. Logo, reestruturamos nossa linguagem de programação utilizando o pacote Rcpp (EDDELBUETTEL et al., 2017), que integra a linguagem C++ e R, o que possibilitou significativo ganho no tempo de processamento, viabilizando todo o estudo de simulação apresentado neste trabalho.

No Capítulo 6, exibimos os resultados de estimação e de predição para o modelo proposto para os dados de precipitação no Estado de Goiás, coletados entre os anos 1980 a 2001. Os Capítulos 7 e 8 trazem, respectivamente, as considerações finais e as sugestões de novos estudos envolvendo possíveis reestruturações do modelo proposto neste trabalho.

2 Modelo Funcional Espaço-Temporal para Dados com Estrutura de Blocos com Medidas Repetidas

Este trabalho propõe a utilização de estruturas funcionais (Apêndice A) de efeitos fixos e aleatórios como preditores em modelos de regressão com estrutura de blocos com medidas repetidas. O nosso conjunto de dados consiste de medidas repetidas no espaço e no tempo, agrupadas em I blocos. Denote por $Y_{ij}(\mathbf{x}, t)$, $i = 1, \ldots, I$, $j = 1, \ldots, J_i$, a j-ésima medida repetida no bloco i observada no tempo $t \in D_T \subset \mathbb{N}$ e na localização $\mathbf{x} = (x_1, x_2) \in D_{lat} \times D_{long} \subset \mathbb{R}^2$, em que x_1 é a latitude e x_2 a longitude, ambas avaliadas em graus decimais. O modelo proposto é dado por

$$Y_{ij}(\mathbf{x}, t) = \mu(\mathbf{x}, t) + \zeta_i(\mathbf{x}) + \Gamma_i(t) + \epsilon_{ij}(\mathbf{x}, t), \tag{2.1}$$

os processos $\epsilon_{ij}(\mathbf{x}, t)$ representam os erros aleatórios do modelo, $\mu(\mathbf{x}, t)$ representa a função média geral na localização $\mathbf{x}$ e no tempo t, $\zeta_i(\mathbf{x})$ representa o efeito espacial aleatório associado à localização $\mathbf{x}$ no bloco i e $\Gamma_i(t)$ acomoda o efeito temporal aleatório associado ao tempo t e bloco i. Contextualizando o modelo proposto ao conjunto de dados reais, devidamente apresentado no Capítulo 6, temos que a variável aleatória $Y_{ij}(\mathbf{x}, t)$ representará a precipitação média avaliada no j-ésimo ano do i-ésimo bloco de efeito climático, avaliada na localização $\mathbf{x}$ e no mês t.

Neste trabalho, vamos considerar que o efeito fixo $\mu(\mathbf{x}, t)$ depende somente da localização $\mathbf{x}$ e do tempo t, porém é possível agregar covariáveis explicativas por meio de um modelo linear ou não linear, desde que as covariáveis de interesse sejam conhecidas em todo $\mathbf{x}$ e t. Assim, consideramos que $\mu(\mathbf{x}, t)$ pertence ao espaço de funções suaves que podem ser escritas como

$$\mu(\mathbf{x}, t) = \sum_{a=1}^{K_{\mu_1}} \sum_{s=1}^{K_{\mu_2}} \sum_{d=1}^{K_{\mu_3}} \beta_{asd} M_a^{(\Upsilon_{\mu_1})}(x_1) M_s^{(\Upsilon_{\mu_2})}(x_2) M_d^{(\Upsilon_{\mu_3})}(t), \tag{2.2}$$

em que K_{μ_1}, K_{μ_2} e K_{μ_3} são inteiros positivos que representam o número de bases relacionadas aos espaços da latitude, longitude e tempo, respectivamente, cujos $K_{\mu_j} - 4$ nós interiores são tomados segundo os conjuntos Υ_{μ_j}, $j = 1, 2, 3$, com $\Upsilon_{\mu_1} \subset D_{lat}$, $\Upsilon_{\mu_2} \subset D_{long}$ e $\Upsilon_{\mu_3} \subset D_T$. Temos ainda que $M_a^{(\Upsilon_{\mu_1})}(.)$, $M_s^{(\Upsilon_{\mu_2})}(.)$ e $M_d^{(\Upsilon_{\mu_3})}(.)$ são bases *B-splines* cúbicos avaliadas em D_{lat}, D_{long} e D_T, respectivamente, e $\boldsymbol{\beta} = (\beta_{111}, \beta_{112}, \ldots, \beta_{11K_{\mu_3}}, \ldots, \beta_{1K_{\mu_2}1}, \beta_{1K_{\mu_2}2}, \ldots, \beta_{1K_{\mu_2}K_{\mu_3}}, \ldots, \beta_{K_{\mu_1}11}, \ldots, \beta_{K_{\mu_1}K_{\mu_2}K_{\mu_3}})$ é o vetor de efeitos fixos considerado na expansão por bases *B-splines*. A seguir, representamos a componente $\zeta_i(\mathbf{x})$ pelo produto tensorial de bases *B-splines* cúbicos avaliadas em D_{lat} e D_{long}, tal que

$$\zeta_i(\mathbf{x}) = \sum_{f=1}^{K_{\zeta_1}} \sum_{g=1}^{K_{\zeta_2}} \theta_{fg}^{(i)} M_f^{(\Upsilon_{\zeta_1})}(x_1) M_g^{(\Upsilon_{\zeta_2})}(x_2), \tag{2.3}$$

em que $M_f^{(\Upsilon_{\zeta_1})}(.)$ e $M_g^{(\Upsilon_{\zeta_2})}(.)$ são as bases *B-splines* cúbicos avaliadas em D_{lat} e D_{long}, nesta ordem, K_{ζ_1} e K_{ζ_2} representam o número de bases associadas às dimensões da latitude e longitude, respectivamente, nas quais os $K_{\zeta_j} - 4$ nós interiores são descritos nos conjuntos Υ_{ζ_j}, $j \in \{4, 5\}$, com $\Upsilon_{\zeta_4} \subset D_{lat}$ e $\Upsilon_{\zeta_5} \subset D_{long}$, e $\boldsymbol{\Theta}^{(i)} = (\theta_{11}^{(i)}, \ldots, \theta_{1K_{\zeta_2}}^{(i)}, \ldots, \theta_{K_{\zeta_1}1}^{(i)}, \ldots, \theta_{K_{\zeta_1}K_{\zeta_2}}^{(i)})$ $\sim \mathbf{N}(\mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{\Theta}}{}^{(i)})$ é o vetor de coeficientes aleatórios considerado na expansão por *B-splines* da componente espacial. Similarmente, $\Gamma_i(t)$ será representado por

$$\Gamma_i(t) = \sum_{l=1}^{K_\Gamma} \vartheta_l^{(i)} M_l^{(\Upsilon_\Gamma)}(t), \tag{2.4}$$

em que $M_l^{(\Upsilon_\Gamma)}(.)$ são as K_Γ bases *B-splines* cúbicos avaliadas em D_T, com $\Upsilon_\Gamma \subset D_T$ representando o conjunto dos $K_\Gamma - 4$ nós interiores, e $\boldsymbol{\vartheta}^{(i)} = (\vartheta_1^{(i)}, \ldots, \vartheta_{K_\Gamma}^{(i)}) \sim \mathbf{N}(\mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{\vartheta}}{}^{(i)})$ é o vetor de efeitos aleatórios considerado na expansão por *B-splines* da componente temporal. Vamos supor ainda que $\boldsymbol{\Theta}^{(i)}$ e $\boldsymbol{\vartheta}^{(i)}$ são variáveis aleatórias independentes para $i = 1, \ldots, I$, cujas respectivas estruturas de covariância acomodam as correlações entre as medidas repetidas no bloco i, vistas por meio das bases *B-splines*.

Sejam $\mathbf{M}^{(\Upsilon_{\mu_3})}$ (de ordem $\tau \times K_{\mu_3}$) e $\mathbf{M}^{(\Upsilon_\Gamma)}$ (de ordem $\tau \times K_\Gamma$) as matrizes contendo as bases *B-splines* cúbicos avaliadas em $\{t_1, t_2, \ldots, t_\tau\} \subset D_T$ considerando K_{μ_3} e K_Γ bases, respectivamente, com nós interiores dados por $\Upsilon_{\mu_3} \subset D_T$ e $\Upsilon_\Gamma \subset D_T$, nessa ordem. Considere também $\mathbf{M}^{(\Upsilon_{\mu_1})}$ (de ordem $n \times K_{\mu_1}$) e $\mathbf{M}^{(\Upsilon_{\zeta_1})}$ (de ordem $n \times K_{\zeta_1}$) as matrizes de bases *B-splines* cúbicos calculadas nos pontos $\{x_{11}, x_{12}, \ldots, x_{1n}\} \subset \mathbf{D}_{lat}$ considerando K_{μ_1} e K_{ζ_1} bases, respectivamente, com nós interiores dados por $\Upsilon_1 \subset D_{lat}$ e $\Upsilon_4 \subset D_{lat}$, por essa ordem, e finalmente, sejam $\mathbf{M}^{(\Upsilon_{\mu_2})}$ (de ordem $n \times K_{\mu_2}$) e $\mathbf{M}^{(\Upsilon_{\zeta_2})}$ (de ordem $n \times K_{\zeta_2}$) as matrizes de bases *B-splines* cúbicos considerando $\{x_{21}, x_{22}, \ldots, x_{2n}\} \subset \mathbf{D}_{long}$ levando em conta K_{μ_2} e K_{ζ_2} bases, nessa ordem, com nós interiores dados respectivamente por $\Upsilon_{\mu_2} \subset D_{long}$ e $\Upsilon_{\zeta_2} \subset D_{long}$. Assim, tomando o produto de *Kronecker* temos que o vetor de observações $\mathbf{Y}_{ij}$ pode ser escrito na forma

$$\mathbf{Y}_{ij} = \mathbf{M}\boldsymbol{\beta} + \mathbf{P}\boldsymbol{\Theta}^{(i)} + \mathbf{N}\boldsymbol{\vartheta}^{(i)} + \mathbf{E}_{ij}, \tag{2.5}$$

em que

$$\mathbf{M} = \begin{pmatrix} \mathbf{M}^{(\Upsilon_{\mu_1})}[1,] \otimes \mathbf{M}^{(\Upsilon_{\mu_2})}[1,] \otimes \mathbf{M}^{(\Upsilon_\Gamma)} \\ \mathbf{M}^{(\Upsilon_{\mu_1})}[2,] \otimes \mathbf{M}^{(\Upsilon_{\mu_2})}[2,] \otimes \mathbf{M}^{(\Upsilon_\Gamma)} \\ \vdots \\ \mathbf{M}^{(\Upsilon_{\mu_1})}[n,] \otimes \mathbf{M}^{(\Upsilon_{\mu_2})}[n,] \otimes \mathbf{M}^{(\Upsilon_\Gamma)} \end{pmatrix}, \tag{2.6}$$

$$\mathbf{P} = \begin{pmatrix} \mathbf{1}_{\tau\times 1} \otimes \mathbf{M}^{(\Upsilon_{\varsigma_1})}[1,] \otimes \mathbf{M}^{(\Upsilon_{\varsigma_2})}[1,] \\ \mathbf{1}_{\tau\times 1} \otimes \mathbf{M}^{(\Upsilon_{\varsigma_1})}[1,] \otimes \mathbf{M}^{(\Upsilon_{\varsigma_2})}[1,] \\ \vdots \\ \mathbf{1}_{\tau\times 1} \otimes \mathbf{M}^{(\Upsilon_{\varsigma_1})}[n,] \otimes \mathbf{M}^{(\Upsilon_{\varsigma_2})}[n,] \end{pmatrix}, \tag{2.7}$$

$\mathbf{N} = (\mathbf{1}_{n\times 1} \otimes \mathbf{M}_{\Gamma(T)})$ e $\mathbf{E}_{ij}$ é o vetor de erros aleatórios que segue uma distribuição normal com matriz de covariância espaço temporal que denotaremos momentaneamente por $\mathbf{\Sigma}_{\epsilon_{ij}}$. Portanto, segue que $\mathbf{Y}_{ij} \sim \mathbf{N}(\mathbf{M}\boldsymbol{\beta}, \mathbf{P}\mathbf{\Sigma_\Theta}^{(i)}\mathbf{P}^T + \mathbf{N}\mathbf{\Sigma}_{\boldsymbol{\vartheta}}{}^{(i)}\mathbf{N}^T + \mathbf{\Sigma}_{\epsilon_{ij}})$.

Podemos observar que o modelo proposto em (2.1) permite que para cada bloco i, $i = 1, 2, \ldots, I$, sejam consideradas J_i medidas repetidas, não sendo portanto, necessariamente, um modelo para dados balanceados. Logo, para o vetor de variáveis resposta do i-ésimo bloco podemos definir

$$\mathbf{Y}_i = \mathbf{M}_i\boldsymbol{\beta} + \mathbf{P}_i\mathbf{\Theta}^{(i)} + \mathbf{N}_i\boldsymbol{\vartheta}^{(i)} + \mathbf{E}_i, \tag{2.8}$$

em que $\mathbf{M}_i = (\mathbf{1}_{J_i} \otimes \mathbf{M})$, $\mathbf{P}_i = (\mathbf{1}_{J_i} \otimes \mathbf{P})$, $\mathbf{N}_i = (\mathbf{1}_{J_i} \otimes \mathbf{N})$, $\mathbf{E}_i$ segue uma distribuição normal com matriz de covariância bloco diagonal $\mathbf{\Sigma}_{\epsilon_i} := \mathbf{BD}(\{\mathbf{\Sigma}_{\epsilon_{ij}}\}_{j=1}^{J_i})$ e, consequentemente, $\mathbf{Y}_i \sim \mathbf{N}(\mathbf{M}_i\boldsymbol{\beta}, \mathbf{P}_i\mathbf{\Sigma_\Theta}^{(i)}\mathbf{P}_i^T + \mathbf{N}_i\mathbf{\Sigma}_{\boldsymbol{\vartheta}}{}^{(i)}\mathbf{N}_i^T + \mathbf{\Sigma}_{\epsilon_i})$. Finalmente, tomando $J_1, J_2, \ldots J_I$ medidas repetidas para os blocos $1, 2, \ldots, I$, respectivamente, temos que $\mathbf{Y} := (\mathbf{Y}_1^T, \mathbf{Y}_2^T, \ldots, \mathbf{Y}_I^T)^T$ pode ser definido por meio da regressão

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Q}\mathbf{\Theta} + \mathbf{R}\boldsymbol{\vartheta} + \mathbf{E}, \tag{2.9}$$

em que $\mathbf{X} = (\mathbf{M}_1^T, \mathbf{M}_2^T, \ldots, \mathbf{M}_I^T)^T$, $\mathbf{Q} = \mathbf{BD}(\{\mathbf{P}_i\}_{i=1}^I)$, $\mathbf{R} = \mathbf{BD}(\{\mathbf{N}_i\}_{i=1}^I)$, $\mathbf{\Theta} = (\mathbf{\Theta}^{(1)}, \mathbf{\Theta}^{(2)}, \ldots, \mathbf{\Theta}^{(I)})$, $\boldsymbol{\vartheta} = (\boldsymbol{\vartheta}^{(1)}, \boldsymbol{\vartheta}^{(2)}, \ldots, \boldsymbol{\vartheta}^{(I)})$ e $\mathbf{E}$ segue uma distribuição normal com matriz de covariância bloco diagonal $\mathbf{\Sigma}_\epsilon := \mathbf{BD}(\{\mathbf{\Sigma}_{\epsilon_i}\}_{i=1}^I)$.

Note que é possível construir extensões do modelo proposto, já que podemos propor diferentes estruturas de covariância para os efeitos aleatórios, espaciais e temporais, bem como para a estrutura de covariância imposta para o erro aleatório do modelo. Dentre as possíveis estruturas de covariância que podem ser admitidas tomamos $\mathbf{\Sigma_\Theta}^{(i)} = \sigma_{\theta_i}^2\mathbf{I}_{K_\varsigma}$,

$\boldsymbol{\Sigma}_{\boldsymbol{\vartheta}}^{(i)} = \sigma^2_{\vartheta_i}\mathbf{I}_{K_\Gamma}$ e $\boldsymbol{\Sigma}_{\epsilon_{ij}} := \boldsymbol{\Sigma}_{\epsilon} = \omega^2\boldsymbol{\Sigma}_S(\kappa,\phi)\otimes\boldsymbol{\Sigma}_T(\varphi)$, em que $\boldsymbol{\Sigma}_T(\varphi)$ representa uma estrutura de correlação com decaimento exponencial de parâmetro φ e $\boldsymbol{\Sigma}_S(\kappa,\phi)$ acomoda uma estrutura de correlação *Matérn* com parâmetros κ e ϕ definida por

$$\boldsymbol{\Sigma}_S(d \mid \kappa,\phi) = \{2^{\kappa-1}\Gamma(\kappa)\}^{-1}(d/\phi)^{\kappa}B_{\kappa}(d/\phi), \tag{2.10}$$

em que $B_\kappa(.)$ representa a função modificada de *Bessel* (ABRAMOWITZ; STEGUN, 1970).

As estruturas de covariância adotadas para o modelo proposto admitem que observações em diferentes localizações e/ou tempos serão correlacionadas segundo uma estrutura espaço-temporal ($\boldsymbol{\Sigma}_\epsilon$) somente se tais medidas pertencerem a um mesmo bloco i, $i = 1,..,I$, e a uma mesma medida repetida j, $j = 1,..,J_i$. Ou seja, entre observações avaliadas em diferentes medidas repetidas, intra ou entre blocos, não admitimos correlação espaço-temporal.

Por outro lado, devido à forma que propomos as componentes espacias e temporais do modelo proposto, expansão em *B-splines* cúbicos, acomodamos ainda estruturas de correlação espaciais e temporais intra bloco provenientes das formas funcionais assumidas. De modo geral, podemos descrever a estrutura de covariância do modelo proposto por

$$\begin{aligned} cov(Y_{ij}(\mathbf{x}_r,t_s),Y_{i'j'}(\mathbf{x}_p,t_q)) &= \sigma_{\theta_i}\{\mathbf{M}^{(\Upsilon_{\varsigma_1})}[r,]\otimes\mathbf{M}^{(\Upsilon_{\varsigma_2})}[r,]\} \\ &\times \{\mathbf{M}^{(\Upsilon_{\varsigma_1})}[p,]\otimes\mathbf{M}^{(\Upsilon_{\varsigma_2})}[p,]\}'\mathbf{I}_{\{i=i'\}} \\ &+ \sigma_{\vartheta_i}\mathbf{M}^{(\Upsilon_\Gamma)}[s,]\mathbf{M}^{(\Upsilon_\Gamma)'}[q,]\mathbf{I}_{\{i=i'\}} \\ &+ \omega^2\boldsymbol{\Sigma}_S(\mathbf{x}_r,\mathbf{x}_s \mid \kappa,\phi)\boldsymbol{\Sigma}_T(t_s.t_q|\varphi)\mathbf{I}_{\{i=i'\}}\mathbf{I}_{\{j=j'\}}. \end{aligned} \tag{2.11}$$

Em relação aos dados reais, apresentados no Capítulo 6, a estrutura de covariância descrita em (2.11) pressupõe que estações metrológicas avaliadas em diferentes anos não serão correlacionadas segundo uma estrutura espaço-temporal. No entanto, estações avaliadas em diferentes anos de um mesmo bloco de efeito climático terão covariância espacial e temporal positiva. É importante observar que o fato de assumirmos sempre covariância positiva não se contrapõe aos dados reais, pois nesse estudo, estamos avaliando precipitação média mensal.

Para melhor entendermos o impacto da escolha do número de bases temos, por exemplo, que ao tomarmos uma quantidade de K bases *B-splines* cúbicos para uma determinada dimensão, estaremos admitindo $K-4$ nós interiores neste domínio de variação, sendo que, quanto maior for o número de nós interiores, menor será o grau de suavidade associado à componente funcional, chegando até mesmo a um processo de interpolação quando este número for grande. No mais, vamos supor que os conjuntos de nós interiores, para todas as dimensões do modelo proposto em (2.1), serão tomados igualmente espaçados em suas respectivas dimensões, sendo que adotaremos ainda $K_\mu := K_{\mu_1} = K_{\mu_2} = K_{\mu_3}$ e $K_\zeta := K_{\zeta_1} = K_{\zeta_2}$.

3 Estimação Bayesiana

Neste capítulo, descrevemos o procedimento de estimação Bayesiana do modelo proposto em (2.1). Para isso, definiremos a função de verossimilhança do modelo, as distribuições a prioris dos parâmetros e a distribuição a posteriori conjunta, bem como um esquema de amostragem da distribuição a posteriori a partir do amostrador de Gibbs (GAMERMAN; LOPES, 2006).

De acordo com Stein (1999), não é recomendado estimar o parâmetro κ disposto na estrutura de correlação *Matérn* (2.10). Logo, o mesmo será escolhido por meio de técnicas de seleção de modelos e portanto, será considerado como uma quantidade fixa, de modo que o vetor de parâmetros para o modelo proposto em (2.1) pode ser representado por $\boldsymbol{\xi} = (\boldsymbol{\beta}, \sigma^2_{\theta_1}, ..., \sigma^2_{\theta_I}, \sigma^2_{\vartheta_1}, ..., \sigma^2_{\vartheta_I}, \omega^2, \phi, \varphi)$ sendo a função de verossimilhança dada por

$$L(\boldsymbol{\xi}|\mathbf{Y}) = \prod_{i=1}^{I} p(\mathbf{Y}_i|\boldsymbol{\xi}), \tag{3.1}$$

em que $\mathbf{Y}_i \sim \mathbf{N}(\mathbf{M}_i\boldsymbol{\beta}, \sigma_{\theta_i}\mathbf{P}_i\mathbf{P}_i^T + \sigma_{\vartheta_i}\mathbf{N}_i\mathbf{N}_i^T + \boldsymbol{\Sigma}_{\epsilon_i})$, independentes em i, $i = 1, 2, ..., I$. Denote por $\Pi(\boldsymbol{\xi}|\boldsymbol{\Delta})$ a distribuição conjunta a priori de $\boldsymbol{\xi}$ em que $\boldsymbol{\Delta}$ é um vetor de hiperparâmetros conhecido. Desta forma, a distribuição conjunta a posteriori de $\boldsymbol{\xi}$, aumentada em $\boldsymbol{\Theta}$ e $\boldsymbol{\vartheta}$, é descrita por

$$\begin{aligned} p(\boldsymbol{\xi}, \boldsymbol{\Theta}, \boldsymbol{\vartheta}|\mathbf{y}, \boldsymbol{\Delta}) &\propto p(\mathbf{y}|\boldsymbol{\xi}, \boldsymbol{\Theta}, \boldsymbol{\vartheta})p(\boldsymbol{\Theta}|\sigma^2_{\theta_i})p(\boldsymbol{\vartheta}|\sigma^2_{\vartheta_i})\Pi(\boldsymbol{\xi}|\boldsymbol{\Delta}) \\ &= \prod_{i=1}^{I}\prod_{j=1}^{J_i} p(\mathbf{Y}_{ij}|\boldsymbol{\xi}, \boldsymbol{\Theta}^{(i)}, \boldsymbol{\vartheta}^{(i)})p(\boldsymbol{\Theta}^{(i)}|\sigma^2_{\theta_i})p(\boldsymbol{\vartheta}^{(i)}|\sigma^2_{\vartheta_i})\Pi(\boldsymbol{\xi}|\boldsymbol{\Delta}), \end{aligned} \tag{3.2}$$

em que $p(\mathbf{Y}_{ij}|\boldsymbol{\xi}, \boldsymbol{\Theta}^{(i)}, \boldsymbol{\vartheta}^{(i)})$ representa a densidade de uma distribuição normal com média $\mathbf{M}\boldsymbol{\beta} + \mathbf{P}\boldsymbol{\Theta}^{(i)} + \mathbf{N}\boldsymbol{\vartheta}^{(i)}$ e matriz de covariância $\omega^2\boldsymbol{\Sigma}_S(\kappa, \phi) \otimes \boldsymbol{\Sigma}_T(\varphi)$, $p(\boldsymbol{\Theta}^{(i)}|\sigma^2_{\theta_i})$ é a densidade de uma distribuição normal com média zero e matriz de covariância $\sigma^2_{\theta_i}\mathbf{I}_{K^2_\varsigma}$ e, finalmente, $p(\boldsymbol{\vartheta}^{(i)}|\sigma^2_{\vartheta_i})$ é a densidade de uma distribuição normal com média zero e matriz covariância $\sigma^2_{\vartheta_i}\mathbf{I}_{K^2_\Gamma}$.

Consideramos que as distribuições a priori dos parâmetros que compõem $\boldsymbol{\xi}$ são independentes e podem variar de acordo com as famílias de distribuições apresentadas na Tabela 1. Como dito anteriormente, o parâmetro κ será escolhido a partir de critérios de seleção de modelos, devidamente apresentados no Capítulo 4.

Devido à estrutura do modelo proposto a distribuição a posteriori conjunta de $(\boldsymbol{\xi}, \boldsymbol{\Theta}, \boldsymbol{\vartheta})$, apresentada em (3.2), não possui forma fechada e, nesse caso, recorremos ao procedimento de amostrador de Gibbs (ROBERT; CASELLA; CASELLA, 1999) para obter amostras da posteriori conjunta. O amostrador de Gibbs descreve um procedimento para se gerar amostras da distribuição de interesse a partir das distribuições condicionais completas que, usualmente, possuem forma fechada conhecida (GAMERMAN; LOPES, 2006). Nos casos em que a distribuição condicional completa não possui a forma fechada conhecida, foi utilizado um passo do algoritmo de Metropolis-Hastings (GAMERMAN; LOPES, 2006).

Na prática, a fim de garantir a convergência do algoritmo são descartados um número suficientemente grande de iterações e, desta forma, estaremos amostrando de uma distribuição que convergiu para a distribuição a posteriori conjunta de $(\boldsymbol{\xi}, \boldsymbol{\Theta}, \boldsymbol{\vartheta})$. Esse procedimento é conhecido por aquecimento, do inglês *burn-in*. Finalmente, para diminuir a dependência dos valores simulados é determinado um valor k de espaçamento, do inglês *thinning*, e assim a amostra considerada é representada pelo valor da cadeia a cada k-ésima iteração.

As distribuições condicionais completas também são apresentadas na Tabela 1. Como pode ser observado, as condicionais completas para os parâmetros φ e ϕ não possuem a forma fechada conhecida, necessitando portanto do algoritmo de *Metropolis-Hastings* para obter amostras de tais distribuições. As distribuições condicionais completas, descritas na Tabela 1, consideram

$$\tilde{\boldsymbol{\beta}} = [\mathbf{X}^T\boldsymbol{\Sigma}_\epsilon^{-1}\mathbf{X} + \omega^2/\sigma^2_\beta\mathbf{I}]^{-1}\mathbf{X}^T\boldsymbol{\Sigma}_\epsilon^{-1}(\mathbf{Y} - \mathbf{Q}\boldsymbol{\Theta} - \mathbf{R}\boldsymbol{\vartheta}), \tag{3.3}$$

$$\tilde{\boldsymbol{\Theta}}^{(i)} = [\mathbf{P}_i^T\boldsymbol{\Sigma}_{\epsilon_i}^{-1}\mathbf{P}_i + \omega^2/\sigma^2_{\theta_i}\mathbf{I}]^{-1}\mathbf{P}_i^T\boldsymbol{\Sigma}_{\epsilon_i}^{-1}(\mathbf{Y}_i - \mathbf{M}_i\boldsymbol{\beta} - \mathbf{N}_i\boldsymbol{\vartheta}^{(i)}), \tag{3.4}$$

$$\tilde{\boldsymbol{\vartheta}}^{(i)} = [\mathbf{N}_i^T\boldsymbol{\Sigma}_{\epsilon_i}^{-1}\mathbf{N}_i + \omega^2/\sigma^2_{\vartheta_i}\mathbf{I}]^{-1}\mathbf{N}_i^T\boldsymbol{\Sigma}_{\epsilon_i}^{-1}(\mathbf{Y}_i - \mathbf{M}_i\boldsymbol{\beta} - \mathbf{P}_i\tilde{\boldsymbol{\Theta}}^{(i)}) \tag{3.5}$$

e

$$\tilde{s}^2_{\omega^2} = (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Q}\boldsymbol{\Theta} - \mathbf{R}\boldsymbol{\vartheta})^T\boldsymbol{\Sigma}_\epsilon^{-1}(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Q}\boldsymbol{\Theta} - \mathbf{R}\boldsymbol{\vartheta}). \tag{3.6}$$

Tabela 1 – Proposta de distribuições a priori e respectivas distribuições condicionais completas para os elementos em $\boldsymbol{\xi}$.

Parâmetro	Distribuição a priori	Distribuição condicional completa
$\boldsymbol{\beta}$	$\mathbf{N}(\mathbf{m}_\beta, \sigma^2_\beta \mathbf{I})$	$\mathbf{N}[\tilde{\boldsymbol{\beta}}, \omega^2(\mathbf{X}^T \boldsymbol{\Sigma}_E^{-1} \mathbf{X} + \omega^2/\sigma^2_\beta \mathbf{I})]$
$\boldsymbol{\Theta}^{(i)}$	$\mathbf{N}(\mathbf{0}, \sigma^2_{\theta_i} \mathbf{I})$	$\mathbf{N}[\tilde{\boldsymbol{\Theta}}^{(i)}, \omega^2(\mathbf{P}_i^T \boldsymbol{\Sigma}_{E_i}^{-1} \mathbf{P}_i + \omega^2/\sigma^2_{\theta_i} \mathbf{I})]$
$\boldsymbol{\vartheta}^{(i)}$	$\mathbf{N}(\mathbf{0}, \sigma^2_{\vartheta_i} \mathbf{I})$	$\mathbf{N}[\tilde{\boldsymbol{\vartheta}}^{(i)}, \omega^2(\mathbf{N}_i^T \boldsymbol{\Sigma}_{E_i}^{-1} \mathbf{N}_i + \omega^2/\sigma^2_{\vartheta_i} \mathbf{I})]$
	Qui-Quadrado Escalonada Inversa$(v_{\theta_i}, s^2_{\theta_i})$	Qui-Quadrado Escalonada Inversa $\left(K_\zeta^2 + v_{\theta_i}, \dfrac{\boldsymbol{\Theta}^{(i)^T}\boldsymbol{\Theta}^{(i)} + v_{\theta_i} s^2_{\theta_i}}{K_\zeta^2 + v_{\theta_i}}\right)$
$\sigma^2_{\theta_i}$	Gama Inversa $(a_{\theta_i}, b_{\theta_i})$	Gama Inversa $\left(\dfrac{K_\zeta^2}{2} + a_{\theta_i}, \dfrac{\boldsymbol{\Theta}^{(i)^T}\boldsymbol{\Theta}^{(i)}}{2} + b_{\theta_i}\right)$
	Meia Cauchy(h_{θ_i})	$p(\mathbf{y}\|\boldsymbol{\xi}, \boldsymbol{\Theta}, \boldsymbol{\vartheta}) p(\sigma^2_{\theta_i}\|h_{\theta_i})$
	Qui-Quadrado Escalonada Inversa$(v_{\vartheta_i}, s^2_{\vartheta_i})$	Qui-Quadrado Escalonada Inversa $\left(K_\Gamma + v_{\vartheta_i}, \dfrac{\boldsymbol{\vartheta}^{(i)^T}\boldsymbol{\vartheta}^{(i)} + v_{\vartheta_i} s^2_{\vartheta_i}}{K_\Gamma + v_{\vartheta_i}}\right)$
$\sigma^2_{\vartheta_i}$	Gama Inversa$(a_{\vartheta_i}, b_{\vartheta_i})$	Gama Inversa $\left(\dfrac{K_\Gamma}{2} + a_{\theta_i}, \dfrac{\boldsymbol{\vartheta}^{(i)^T}\boldsymbol{\vartheta}^{(i)}}{2} + b_{\vartheta_i}\right)$
	Meia Cauchy(h_{ϑ_i})	$p(\mathbf{y}\|\boldsymbol{\xi}, \boldsymbol{\Theta}, \boldsymbol{\vartheta}) p(\sigma^2_{\vartheta_i}\|h_{\vartheta_i})$
ω^2	Qui-Quadrado Escalonada Inversa$(v_{\omega^2}, s^2_{\omega^2})$	Qui-Quadrado Escalonada Inversa $\left[n\tau(J_1 + J_2 + J_3) + v_{\omega^2}, \dfrac{\tilde{s}^2_{\omega^2} + v_{\omega^2} s^2_{\omega^2}}{n\tau(J_1 + J_2 + J_3) + v_{\omega^2}}\right]$
	Gama Inversa$(a_{\omega^2}, b_{\omega^2})$	Gama Inversa $\left(\dfrac{n\tau(J_1 + J_2 + J_3)}{2} + a_{\omega^2}, \dfrac{\tilde{s}^2_{\omega^2}}{2} + b_{\omega^2}\right)$
ϕ	Gama(c_ϕ, d_ϕ)	$p(\mathbf{y}\|\boldsymbol{\xi}, \boldsymbol{\Theta}, \boldsymbol{\vartheta}) p(\phi\|c_\phi, d_\phi)$
φ	Gama(c_φ, d_φ)	$p(\mathbf{y}\|\boldsymbol{\xi}, \boldsymbol{\Theta}, \boldsymbol{\vartheta}) p(\varphi\|c_\varphi, d_\varphi)$

4 Critérios de Seleção de Modelos

Neste trabalho, a utilização dos critérios de seleção de modelos será a ferramenta utilizada para selecionar os possíveis valores do parâmetro κ que compõe a estrutura de correlação *Matérn* dada em (2.10). Iremos considerar dois critérios de comparação de modelos amplamente utilizados na literatura: um critério baseado no logaritmo da pseudo verossimilhança marginal (LPML), derivado a partir da ordenada preditiva condicional (CPO), proposta por Geisser (1993) e por Geisser e Eddy (1979), e um critério derivado a partir da função desvio (DIC) adaptado para o caso de presença de efeitos aleatórios (CELEUX et al., 2006).

4.1 Critério LPML

Seja $\mathbf{y} = (\mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_I)$ uma amostra observada com componentes vetoriais $\mathbf{y}_i$ independentes para todo $i = 1, 2, ..., I$. Denote por $\mathbf{y}_{(i)}$ a amostra sem a componente vetorial $\mathbf{y}_i$. Então, a ordenada preditiva condicional (CPO) associado à observação i é definida como a densidade marginal preditiva a posteriori condicional às observações restantes em $\mathbf{y}_{(i)}$ dada por

$$CPO_i = f(\mathbf{y}_i|\mathbf{y}_{(i)}) = \int f(\mathbf{y}_i|\boldsymbol{\varsigma})p(\boldsymbol{\varsigma}|\mathbf{y}_{(i)})d\boldsymbol{\varsigma}, \tag{4.1}$$

em que $\boldsymbol{\varsigma}$ é o vetor de parâmetros, $f(\mathbf{y}_i|\boldsymbol{\varsigma})$ representa a distribuição condicional de $\mathbf{y}_i$ dado $\boldsymbol{\varsigma}$ e $p(\boldsymbol{\varsigma}|\mathbf{y}_{(i)})$ é a distribuição a posteriori associada ao vetor $\mathbf{y}_{(i)}$ (GELFAND, 1996). Segundo Carlin e Louis (2008) o produto dos CPOs individuais podem ser vistos como uma pseudo verossimilhança marginal e, desta forma, o critério baseado no logaritmo da pseudo verossimilhança marginal ($LPML$) é dado por

$$LPML = \sum_i \log(CPO_i). \tag{4.2}$$

A proposta do CPO é comparar os modelos por meio das densidades preditivas $f(\mathbf{y}_i|\mathbf{y}_{(i)})$ por meio de validação cruzada (GELFAND, 1996). Note ainda que as distribuições preditivas das observações que foram removidas são obtidas para examinar se os pontos removidos pertencem a regiões com densidade significativa. Portanto, com base nesse critério, o modelo com maior poder de predição em termos da densidade preditiva a posteriori é aquele em que, em média, os pontos removidos tem maior chance de serem explicados, ou seja, valores mais altos da estatística LPML indicam um melhor ajuste do modelo.

O CPO pode ser visto como uma medida do efeito da observação $\mathbf{y}_i$ na densidade preditiva a priori de $\mathbf{y}$, $f(\mathbf{y}) = \displaystyle\int f(\mathbf{y}_i|\boldsymbol{\varsigma})p(\boldsymbol{\varsigma}|\mathbf{y}_{(i)})d\boldsymbol{\varsigma}$, dado que

$$CPO_i = f(\mathbf{y}_i|\mathbf{y}_{(i)}) = \frac{f(\mathbf{y})}{f(\mathbf{y}_{(i)})}, \tag{4.3}$$

tal que o modelo necessita ser reajustado a cada observação retirada. Entretanto, como $\mathbf{y}_i$ e $\mathbf{y}_{(i)}$ são independentes, podemos observar que

$$\begin{aligned}(CPO_i)^{-1} &= \frac{1}{f(\mathbf{y}_i|\mathbf{y}_{(i)})} = \frac{f(\mathbf{y}_{(i)})}{f(\mathbf{y})} = \int \frac{f(\mathbf{y}_{(i)}|\boldsymbol{\varsigma})\pi(\boldsymbol{\varsigma})}{f(\mathbf{y})}d\boldsymbol{\varsigma} \\ &= \int \frac{f(\mathbf{y}_i|\boldsymbol{\varsigma})}{f(\mathbf{y}_i|\boldsymbol{\varsigma})}\frac{f(\mathbf{y}_{(i)}|\boldsymbol{\varsigma})\pi(\boldsymbol{\varsigma})}{f(\mathbf{y})}d\boldsymbol{\varsigma} = \int \frac{1}{f(\mathbf{y}_i|\boldsymbol{\varsigma})}\frac{f(\mathbf{y}|\boldsymbol{\varsigma})\pi(\boldsymbol{\varsigma})}{f(\mathbf{y})}d\boldsymbol{\varsigma} \\ &= \int \frac{1}{f(\mathbf{y}_i|\boldsymbol{\varsigma})}p(\boldsymbol{\varsigma}|\mathbf{y})d\boldsymbol{\varsigma} = E_{\boldsymbol{\varsigma}|\mathbf{y}}\left[\frac{1}{f(\mathbf{y}_i|\boldsymbol{\varsigma})}\right]. \end{aligned} \tag{4.4}$$

Dey, Gelfand e Peng (1997) mostram que uma estimativa por Monte Carlo de (4.1), obtida por meio da aproximação por média harmônica de (4.4), é dada por

$$\widehat{CPO_i} = \left(\frac{1}{M}\sum_{m=1}^{M}\frac{1}{f(\mathbf{y}_i|\boldsymbol{\varsigma}^{(m)})}\right)^{-1}, \tag{4.5}$$

em que $\boldsymbol{\varsigma}^{(1)}, \boldsymbol{\varsigma}^{(2)}, ..., \boldsymbol{\varsigma}^{(M)}$ são amostras da posteriori $p(\boldsymbol{\varsigma}|\mathbf{y})$. Logo, temos uma estimativa que não necessita que o modelo seja reajustado a cada observação removida.

No caso de modelos com efeitos aleatórios o cálculo de $f(\mathbf{y}_i|\boldsymbol{\varsigma}^{(m)})$ depende de uma integral com respeito à distribuição dos efeitos aleatórios. Essas integrais podem trazer uma maior estabilidade ao estimador (4.6) uma vez que o espaço paramétrico é reduzido (RAFTERY et al., 2006). Neste trabalho, conforme apresentado em (3.2), o vetor de parâmetros é $\boldsymbol{\xi}$ com a função de verossimilhança aumentada em $\boldsymbol{\Theta}$ e $\boldsymbol{\vartheta}$ sendo que, nesse caso, os efeitos aleatórios são de alta dimensão, tornando as integrais em (4.4) computacionalmente dispendiosas. Assim, para modelos com efeitos aleatórios as estimativas dos $CPO_{i´\mathrm{s}}$ são dadas por

$$\widehat{CPO_i} = \left(\frac{1}{M}\sum_{m=1}^{M}\frac{1}{f(\mathbf{y}_i|\boldsymbol{\Theta}^{(i)^{(m)}}, \boldsymbol{\vartheta}^{(i)^{(m)}}, \boldsymbol{\beta}^{(m)}, \omega^{(m)}, \phi^{(m)}, \varphi^{(m)})}\right)^{-1}, \tag{4.6}$$

em que $\boldsymbol{\Theta}^{(i)^{(m)}}$, $\boldsymbol{\vartheta}^{(i)^{(m)}}$, $\boldsymbol{\beta}^{(m)}$, $\omega^{(m)}$, $\phi^{(m)}$ e $\varphi^{(m)}$ são amostras obtidas a partir da posteriori.

4.2 Critério DIC7

Outro critério que também iremos utilizar para selecionar possíveis valores para o parâmetro κ, que compõe a estrutura de correlação *Matérn* dada em (2.10), foi proposto inicialmente por Spiegelhalter et al. (2002) como um critério de comparação de modelos que considera o ajuste e a complexidade do modelo em seu cálculo, denominado DIC. No caso dos modelos estruturados somente com efeitos fixos, essa complexidade é dada pelo número efetivo de parâmetros p_D. Por outro lado, para os modelos com efeitos aleatórios essa complexidade demanda uma alternativa ao DIC tradicional (CELEUX et al., 2006).

O DIC é definido por:

$$DIC = \overline{D(\boldsymbol{\varsigma})} + p_D,$$

em que $\overline{D(\boldsymbol{\varsigma})} = E[D(\boldsymbol{\varsigma})|\mathbf{y}]$ é a esperança a posteriori da função desvio do modelo, com $D(\boldsymbol{\varsigma}) = -2\log(f(\mathbf{y})|\boldsymbol{\varsigma})$, e $p_D = \overline{D(\boldsymbol{\varsigma})} - D(\tilde{\boldsymbol{\varsigma}}) = E[D(\boldsymbol{\varsigma})|\mathbf{y}] - D(\tilde{\boldsymbol{\varsigma}})$ é a diferença entre a esperança a posteriori da função desvio e a função desvio avaliada em $\tilde{\boldsymbol{\varsigma}}$ (uma estimativa de $\boldsymbol{\varsigma}$). Assim, o DIC pode ser escrito como:

$$DIC = -4E[\log(f(\mathbf{y}|\boldsymbol{\varsigma}))] + 2\log(f(\mathbf{y}|\tilde{\boldsymbol{\varsigma}})). \tag{4.7}$$

Para o modelo proposto, apresentado na Equação (2.9), vamos considerar o critério DIC7 apresentado em Celeux et al. (2006), o qual é baseado na verossimilhança condicional $f(\mathbf{y}|\boldsymbol{\Theta}, \boldsymbol{\vartheta}, \boldsymbol{\beta}, \omega, \phi, \varphi)$ e dado por:

$$DIC7 = -4E[\log(f(\mathbf{y}|\boldsymbol{\Theta}, \boldsymbol{\vartheta}, \boldsymbol{\beta}, \omega, \phi, \varphi))] + 2\log(f(\mathbf{y}|\hat{\boldsymbol{\Theta}}, \hat{\boldsymbol{\vartheta}}, \hat{\boldsymbol{\beta}}, \hat{\omega}, \hat{\phi}, \hat{\varphi})), \tag{4.8}$$

em que $\hat{\boldsymbol{\Theta}}$ e $\hat{\boldsymbol{\vartheta}}$ são as médias a posteriori dos efeitos aleatórios $\boldsymbol{\Theta}$ e $\boldsymbol{\vartheta}$, nessa ordem, $\hat{\boldsymbol{\beta}}$, $\hat{\omega}$, $\hat{\phi}$, $\hat{\varphi}$ são as médias a posteriori de $\boldsymbol{\beta}$, ω, ϕ e φ, respectivamente, estimativas estas obtidas por meio das suas respectivas distribuições a posteriori. Finalmente, aproximamos a esperança a posteriori da função desvio $D(\boldsymbol{\Theta}, \boldsymbol{\vartheta}, \boldsymbol{\beta}, \omega, \phi, \varphi) = E[\log(f(\mathbf{y}|\boldsymbol{\Theta}, \boldsymbol{\vartheta}, \boldsymbol{\beta}, \omega, \phi, \varphi))]$ por

$$1/M \sum_{m=1}^{M} \sum_{i=1}^{I} \log(f(\mathbf{y}_i|\boldsymbol{\Theta}^{(i)^{(m)}}, \boldsymbol{\vartheta}^{(i)^{(m)}}, \boldsymbol{\beta}^{(m)}, \omega^{(m)}, \phi^{(m)}, \varphi^{(m)})), \tag{4.9}$$

em que $\boldsymbol{\Theta}^{(i)^{(m)}}$, $\boldsymbol{\vartheta}^{(i)^{(m)}}$, $\boldsymbol{\beta}^{(m)}$, $\omega^{(m)}$, $\phi^{(m)}$ e $\varphi^{(m)}$ são amostras obtidas a partir da posteriori. Neste caso, o melhor modelo é aquele com menor valor para o DIC7.

5 Estudos de Simulação

Neste capítulo, apresentamos um estudo de simulação com intuito de avaliar o desempenho do modelo proposto em (2.1) sob os seguintes aspectos: na Seção 5.1 avaliamos a sensibilidade da distribuição a priori das componentes de variância, na Seção 5.2 consideramos o efeito do tamanho amostral e do número de medidas repetidas no processo de estimação, na Seção 5.3 estudamos a determinação do parâmetro κ, na Seção 5.4 avaliamos a capacidade preditiva do modelo proposto e na Seção 5.5 estudamos a adequabilidade da estimação empírica dos parâmetros em espaços ortogonais.

5.1 Estudo da Sensibilidade da Distribuição a Priori das Componentes de Variância

O uso de certas distribuições a priori para as componentes de variância, ainda que não informativas, podem levar a um achatamento da posteriori em determinadas regiões caracterizando uma restrição do espaço amostral (GELMAN, 2006). Assim, a fim de avaliarmos tal impacto sobre nosso processo de estimação foram gerados 50 conjuntos de dados independentes, a partir dos valores apresentados na Tabela 2, mantendo $\kappa = 0.2$ e o número de bases $K_\mu = 9$, $K_\zeta = 5$ e $K_\Gamma = 7$. Foram considerados três blocos, com medidas repetidas $J_1 = 9$ e $J_2 = J_3 = 6$, sendo o número de tempos observados $\tau = 6$, imitando a estrutura do conjunto de dados reais.

Tabela 2 – Valores propostos para os parâmetros do modelo proposto em (2.1), utilizados para gerar 50 conjuntos de dados independentes mantendo $\kappa = 0.2$, $K_\mu = 9$, $K_\zeta = 5$ e $K_\Gamma = 7$.

Parâmetros	Valores Propostos
ω^2	11.02
$\sigma^2_{\theta_1}$	0.09
$\sigma^2_{\theta_2}$	0.06
$\sigma^2_{\theta_3}$	0.10
$\sigma^2_{\vartheta_1}$	0.62
$\sigma^2_{\vartheta_2}$	0.18
$\sigma^2_{\vartheta_3}$	1.05
φ	2.04
ϕ	705

De acordo com o Capítulo 3, as distribuições condicionais completas para as componentes de variância φ e ϕ do modelo proposto em (2.1) não apresentam forma fechada e portanto, em cada passo do amostrador de Gibbs, é introduzido um passo do algoritmo de *Metropolis-Hastings*. Estudos preliminares de simulação realizados para os parâmetros φ e ϕ, em que os demais parâmetros do modelo proposto foram mantidos fixos, mostraram resultados promissores quando a distribuição Gama é considerada como priori de φ e ϕ, havendo menor período de convergência das cadeias para estes parâmetros. Portanto, nos estudos que serão apresentados fixamos as prioris Gama$(1, 0.01)$ e Gama$(1, 0.001)$ para os parâmetros φ e ϕ, respectivamente.

Na Figura 1, apresentamos as medianas a posteriori para as componentes de variância ω^2, φ e ϕ, do modelo proposto em (2.1), obtidas ao longo dos 50 conjuntos de dados simulados considerando as famílias de prioris Gama Inversa e Qui-Quadrado Escalonada Inversa para os parâmetros $\sigma^2_{\theta_i}$ e $\sigma^2_{\vartheta_i}$, $i = 1, 2, 3$, e fixando Gama$(1, 0.01)$ e Gama$(1, 0.001)$ para os parâmetros φ e ϕ, respectivamente. Ainda na Figura 1, observamos que as medianas a posteriori para os parâmetros φ e ϕ não são impactadas por mudanças nas distribuições a priori dos demais parâmetros em estudo, dado que o comportamento das medianas se mantiveram praticamente estáveis ao longo de todos os cenários avaliados para os demais parâmetros. Com relação ao parâmetro ω^2, notamos uma maior robustez à escolha das prioris, pois o comportamento das medianas a posteriori apresentou certa similaridade dentre todos cenários avaliados. Vale ressaltar, que essas mesmas características também foram observadas para a média e moda a posteriori, em que foi possível observar ainda indícios acerca da simetria das distribuições a posteriori, uma vez que nestes casos as estimativas da mediana e da média variam em regiões semelhantes (veja Apêndice B).

Com relação às medianas a posteriori das componentes de variância dos efeitos aleatórios espaciais e temporais, denotados, respectivamente, por $\sigma^2_{\theta_i}$ e $\sigma^2_{\vartheta_i}$, $i = 1, 2, 3$, ilustradas nas Figuras 2 e 3, respectivamente, observamos que a família Qui-Quadrado Escalonada Inversa apresenta um maior achatamento das estimativas nas proximidades do zero, sendo que o melhor resultado foi obtido na distribuição a priori Gama Inversa$(0.01, 0.01)$, pois neste caso as estimativas estão mais próximas dos verdadeiros valores dos respectivos parâmetros. Destacamos ainda que a média e a moda a posteriori das componentes de variância dos efeitos aleatórios espaciais e temporais também apresentaram achatamento das estimativas nas proximidades do zero, além disso, observamos evidências acerca da assimetria das distribuições a posteriori, apontadas pela distribuição das estimativas mediana e moda a posteriori nos respectivos gráficos. Para maiores detalhes veja Apêndice B.

Quanto às probabilidades de cobertura, apresentadas nas Tabelas 3 e 4, observamos que os parâmetros em estudo são levemente impactados por mudanças nas distribuições a priori. Os piores resultados foram alcançados quando a família de prioris Qui-Quadrado Escalonada Inversa é utilizada e, de maneira geral, os melhores resultados foram atingidos quando a priori Gama Inversa $(0.001, 0.001)$ é adotada.

Segundo Gelman (2006) o achatamento da distribuição a posteriori para de-

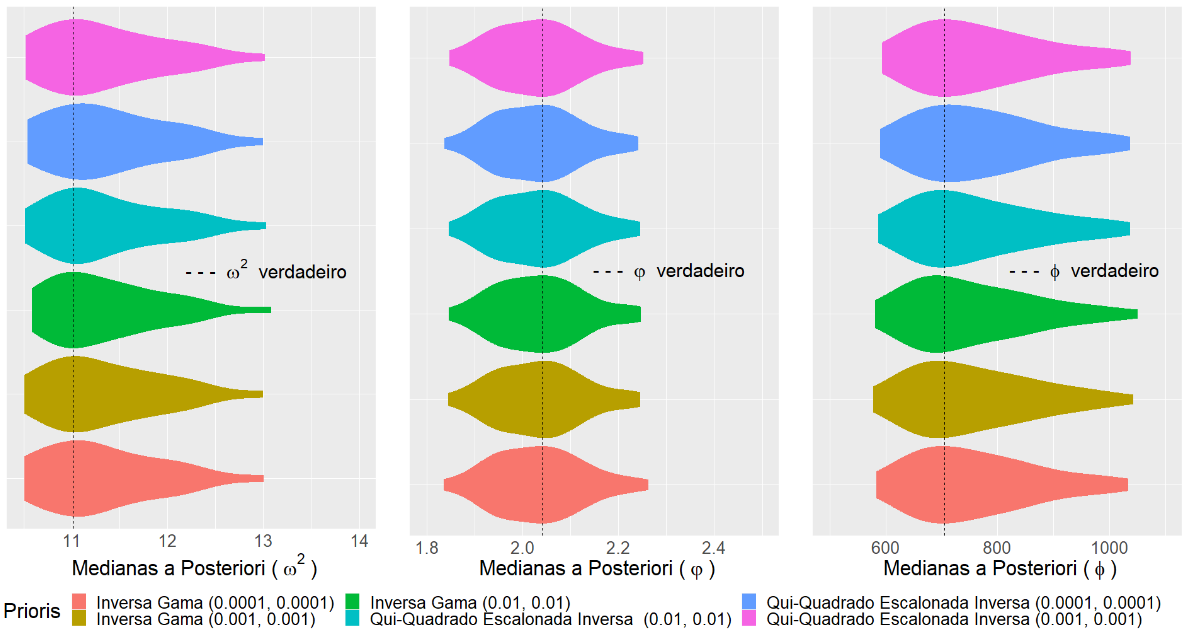

Figura 1 – Medianas a posteriori das componentes de variância ω^2, φ e ϕ, do modelo proposto em (2.1), tomadas ao longo dos 50 conjuntos de dados simulados considerando as famílias de prioris Gama Inversa e Qui-Quadrado Escalonada Inversa para os parâmetros ω^2, $\sigma^2_{\theta_i}$ e $\sigma^2_{\vartheta_i}$, $i = 1, 2, 3$, e fixando Gama$(1, 0.01)$ e Gama$(1, 0.001)$ para os parâmetros φ e ϕ, respectivamente.

terminadas componentes de variância pode ser amenizado com o uso de uma priori pertencente à família de distribuição Meia Cauchy. Assim, consideramos o seu uso no estudo de simulação tal que para os parâmetros ω^2, φ e ϕ atribuímos distribuições a priori Gama Inversa$(0.01, 0.01)$ e para as distribuições a priori das componentes de variância dos efeitos aleatórios espaciais e temporais, $\sigma^2_{\theta_i}$ e $\sigma^2_{\vartheta_i}$, i=1,2,3, foram tomadas Meia Cauchy(25), caracterizando assim prioris próprias não informativas. A Figura 4 apresenta os resultados obtidos para esse cenário. Como podemos observar, as medianas a posteriori não apresentam melhores comportamentos quando comparadas aos demais casos de simulação apresentados nessa seção, além disso, os ganhos observados nos percentuais das probabilidades de cobertura (Tabela 5) podem ser explicados pelo aumento nas variâncias a posteriori.

Portanto, a partir dos resultados apresentados nesta seção, concluímos que a priori Gama Inversa $(0.01, 0.01)$ apresenta os melhores resultados e será considerada para os parâmetros ω^2, $\sigma^2_{\theta_i}$ e $\sigma^2_{\vartheta_i}$, $i = 1, 2, 3$. Verificamos por meio do pacote Coda (PLUMMER et al., 2019), desenvolvido no programa R (TEAM, 2013), que um período de aquecimento da cadeia de 120.000 passos e espaçamento de 100 passos são suficientes para garantir a convergência das cadeias e que as amostras geradas sejam consideradas independentes.

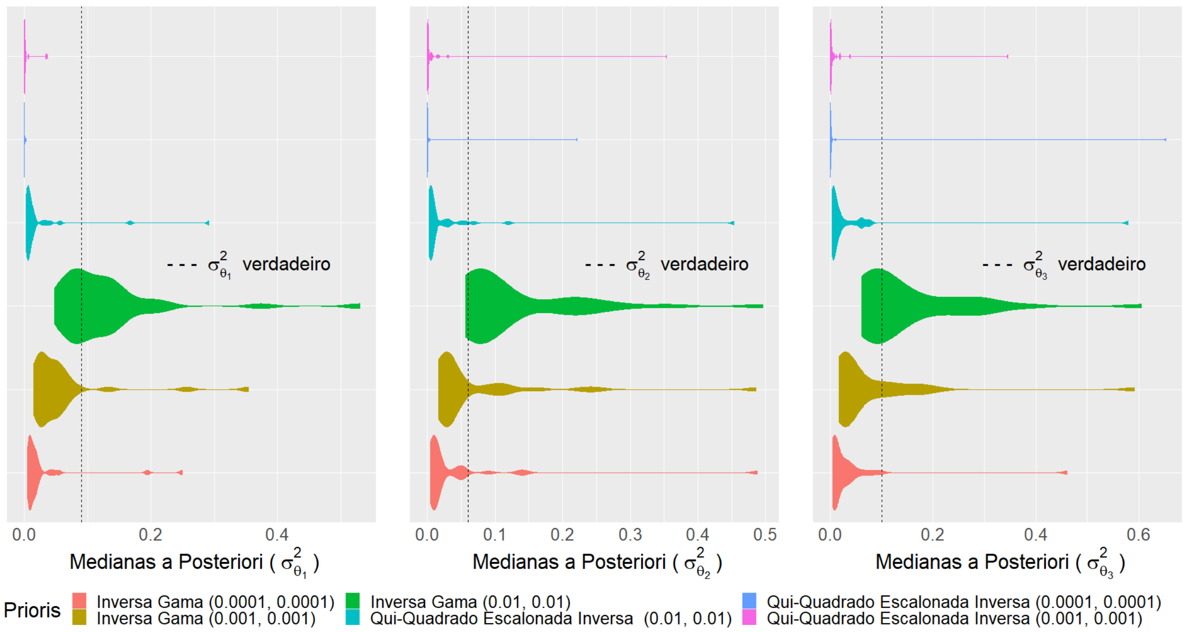

Figura 2 – Medianas a posteriori das componentes de variância do efeito aleatório espacial $\sigma^2_{\theta_i}$, $i = 1, 2, 3$, do modelo proposto em (2.1), tomadas ao longo dos 50 conjuntos de dados simulados quando consideradas as famílias de prioris Gama Inversa e Qui-Quadrado Escalonada Inversa para os parâmetros ω^2, $\sigma^2_{\theta_i}$ e $\sigma^2_{\vartheta_i}$, $i = 1, 2, 3$, e fixando Gama$(1, 0.001)$ e Gama$(1, 0.01)$ para os parâmetros φ e ϕ, respectivamente.

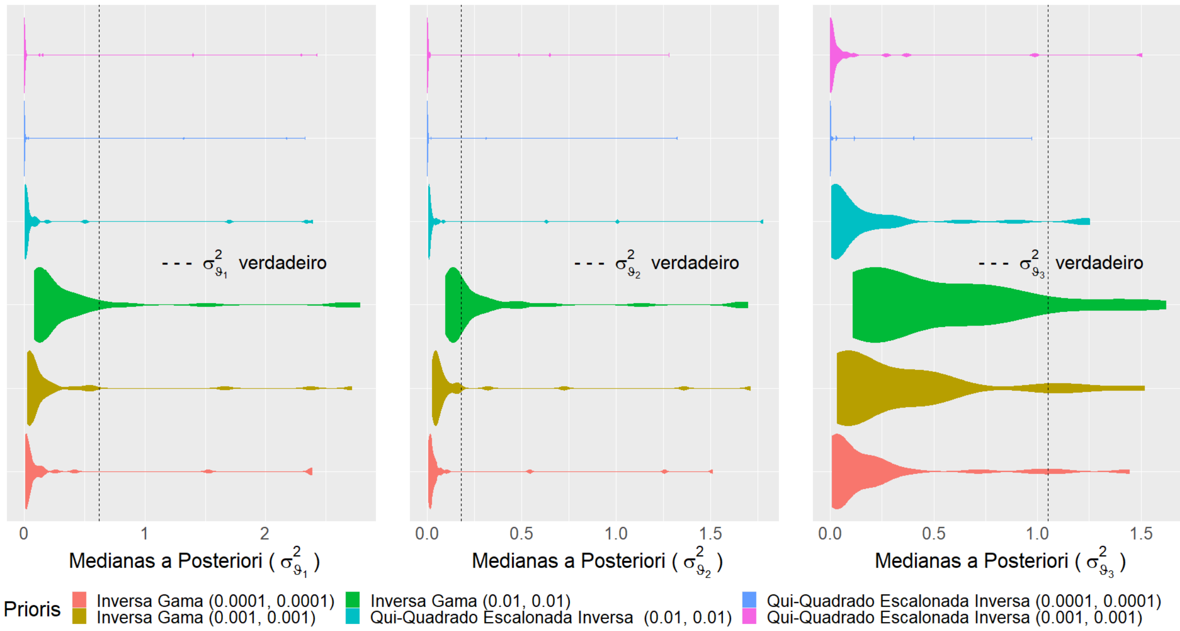

Figura 3 – Medianas a posteriori das componentes de variância do efeito aleatório temporal $\sigma^2_{\vartheta_i}$, $i = 1, 2, 3$, do modelo proposto em (2.1), tomadas ao longo dos 50 conjuntos de dados simulados quando consideradas as famílias de prioris Gama Inversa e Qui-Quadrado Escalonada Inversa para os parâmetros ω^2, $\sigma^2_{\theta_i}$ e $\sigma^2_{\vartheta_i}$, $i = 1, 2, 3$, e fixando Gama$(1, 0.001)$ e Gama$(1, 0.01)$ para os parâmetros φ e ϕ, respectivamente.

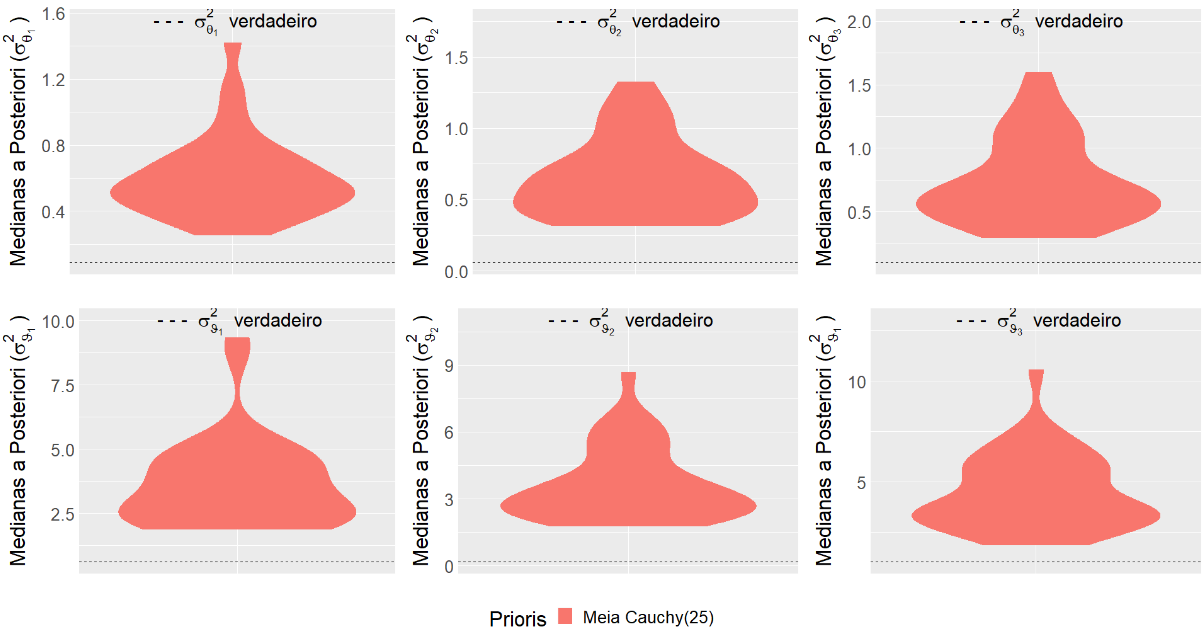

Figura 4 – Medianas a posteriori das componentes de variância do efeito aleatório temporal $\sigma^2_{\vartheta_i}$, $i = 1, 2, 3$, do modelo proposto em (2.1), tomadas ao longo dos 50 conjuntos de dados simulados considerando a priori Gama Inversa$(0.01, 0.01)$ para os parâmetros ω^2, φ e ϕ, e Meia Cauchy(25) para as componentes de variância $\sigma^2_{\theta_i}$ e $\sigma^2_{\vartheta_i}$, $i = 1, 2, 3$.

Tabela 3 – Probabilidade de cobertura (PC) para os parâmetros do modelo proposto em (2.1) tomadas ao longo dos 50 conjuntos de dados gerados independentemente aplicando as famílias de prioris Gama Inversa para os parâmetros ω^2, $\sigma^2_{\theta_i}$ e $\sigma^2_{\vartheta_i}$, $i = 1, 2, 3$, e fixando Gama$(1, 0.01)$ e Gama$(1, 0.001)$ para os parâmetros φ e ϕ, respectivamente.

Distribuição a Priori	Componete de Variância	PC
Gama Inversa $(0.0001, 0.0001)$	ω^2	92%
	$\sigma^2_{\theta_1}$	100%
	$\sigma^2_{\theta_2}$	100%
	$\sigma^2_{\theta_3}$	100%
	$\sigma^2_{\vartheta_1}$	100%
	$\sigma^2_{\vartheta_2}$	100%
	$\sigma^2_{\vartheta_3}$	96%
Gama$(1, 0.01)$	φ	90%
Gama$(1, 0.001)$	ϕ	92%
Gama Inversa $(0.001, 0.001)$	ω^2	94%
	$\sigma^2_{\theta_1}$	100%
	$\sigma^2_{\theta_2}$	100%
	$\sigma^2_{\theta_3}$	100%
	$\sigma^2_{\vartheta_1}$	100%
	$\sigma^2_{\vartheta_2}$	100%
	$\sigma^2_{\vartheta_3}$	100%
Gama$(1, 0.01)$	φ	90%
Gama$(1, 0.001)$	ϕ	92%
Gama Inversa $(0.01, 0.01)$	ω^2	92%
	$\sigma^2_{\theta_1}$	100%
	$\sigma^2_{\theta_2}$	100%
	$\sigma^2_{\theta_3}$	100%
	$\sigma^2_{\vartheta_1}$	100%
	$\sigma^2_{\vartheta_2}$	100%
	$\sigma^2_{\vartheta_3}$	100%
Gama$(1, 0.01)$	φ	88%
Gama$(1, 0.001)$	ϕ	90%

Tabela 4 – Probabilidade de cobertura (PC) para os parâmetros do modelo proposto em (2.1) tomadas ao longo dos 50 conjuntos de dados gerados independentemente aplicando as famílias de prioris Qui-Quadrado Escalonada Inversa para os parâmetros ω^2, $\sigma^2_{\theta_i}$ e $\sigma^2_{\vartheta_i}$, $i = 1, 2, 3$, e fixando Gama$(1, 0.01)$ e Gama$(1, 0.001)$ para os parâmetros φ e ϕ, respectivamente.

Distribuição a Priori	Componete de Variância	PC
Qui-Quadrado Escalonada Inversa $(0.0001, 0.0001)$	ω^2	90%
	$\sigma^2_{\theta_1}$	98%
	$\sigma^2_{\theta_2}$	100%
	$\sigma^2_{\theta_3}$	96%
	$\sigma^2_{\vartheta_1}$	56%
	$\sigma^2_{\vartheta_2}$	100%
	$\sigma^2_{\vartheta_3}$	80%
Gama$(1, 0.01)$	φ	90%
Gama$(1, 0.001)$	ϕ	88%
Qui-Quadrado Escalonada Inversa $(0.001, 0.001)$	ω^2	90%
	$\sigma^2_{\theta_1}$	100%
	$\sigma^2_{\theta_2}$	100%
	$\sigma^2_{\theta_3}$	100%
	$\sigma^2_{\vartheta_1}$	78%
	$\sigma^2_{\vartheta_2}$	100%
	$\sigma^2_{\vartheta_3}$	80%
Gama$(1, 0.01)$	φ	90%
Gama$(1, 0.001)$	ϕ	92%
Qui-Quadrado Escalonada Inversa $(0.01, 0.01)$	ω^2	90%
	$\sigma^2_{\theta_1}$	100%
	$\sigma^2_{\theta_2}$	100%
	$\sigma^2_{\theta_3}$	100%
	$\sigma^2_{\vartheta_1}$	96%
	$\sigma^2_{\vartheta_2}$	100%
	$\sigma^2_{\vartheta_3}$	96%
Gama$(1, 0.01)$	φ	88%
Gama$(1, 0.001)$	ϕ	90%

Tabela 5 – Probabilidades de cobertura (PC) para os parâmetros do modelo proposto em (2.1) tomadas ao longo dos 50 conjuntos de dados gerados independentemente aplicando a priori Gama Inversa$(0.01, 0.01)$ para os parâmetros ω^2, φ e ϕ, e Meia Cauchy(25) para as componentes de variância $\sigma^2_{\theta_i}$ e $\sigma^2_{\vartheta_i}$, $i = 1, 2, 3$.

Distribuição a Priori	Componete de Variância	PC
Gama Inversa $(0.01, 0.01)$	ω^2	92%
Meia Cauchy (25)	$\sigma^2_{\theta_1}$	100%
	$\sigma^2_{\theta_2}$	100%
	$\sigma^2_{\theta_3}$	100%
	$\sigma^2_{\vartheta_1}$	100%
	$\sigma^2_{\vartheta_2}$	98%
	$\sigma^2_{\vartheta_3}$	100%
Gama$(1, 0.01)$	φ	88%
Gama$(1, 0.001)$	ϕ	92%

5.2 Estudo da Sensibilidade do Número de Localizações (n) e do Número de Medidas Repetidas (J_i, $i = 1, 2, 3$)

Nesta seção, abordaremos um estudo de simulação para avaliar o impacto do número de localizações (n) e do número de medidas repetidas por bloco (J_i, $i = 1, 2, 3$) nas estimativas dos parâmetros do modelo proposto em (2.1). Para efeito de contextualização, no conjunto de dados reais que será apresentado no Capítulo 6, n e J_i, $i = 1, 2, 3$, representam respectivamente o número de estações meteorológicas avaliadas e o número de anos avaliados em cada bloco climático.

O conjunto de distribuições a priori adotado nos estudos de simulação apresentados nesta seção está disposto na Tabela 6. Em todos os cenários propostos, mantivemos um período de aquecimento da cadeia de 120.000 passos e espaçamento de 100 passos. Utilizamos o pacote Coda (PLUMMER et al., 2019) do R (TEAM, 2013) para verificar a convergência das cadeias e assim amostrar da posteriori de forma apropriada. Foram verificados ainda os critérios de convergência das cadeias apresentados em Gelman e Rubin (1992).

Tabela 6 – Distribuições a priori consideradas no estudo de simulação realizado para avaliar o impacto sobre as estimativas dos parâmetros do modelo proposto em (2.1) quando o número de localizações (n) e o número de medidas repetidas por bloco (J_i, $i = 1, 2, 3$) são alteradas.

Parâmetros	Distribuição a Priori
$\boldsymbol{\beta}$	$\mathbf{N}(\mathbf{0}, 10^3\mathbf{I})$
ω^2	Gama Inversa$(0.01, 0.01)$
$\sigma^2_{\theta_i}$	Gama Inversa$(0.01, 0.01)$
$\sigma^2_{\vartheta_i}$	Gama Inversa$(0.01, 0.01)$
φ	Gama$(1, 0.01)$
ϕ	Gama$(1, 0.001)$

O primeiro estudo de simulação busca avaliar o impacto do número de localizações (n) e o número de períodos no tempo (τ) sobre o processo de estimação. Para tanto, foram gerados 50 conjuntos de dados independentes considerando 112 ($n = 112$) localizações avaliadas ao longo de 9 períodos no tempo ($\tau = 9$). As configurações geográficas, longitude e latitude, foram herdadas dos dados reais e complementadas a partir de valores sorteados aleatoriamente dentro dos respectivos domínios de variação. A complementação dos referenciais geográficos, longitude e latitude, se faz necessária uma vez que os dados reais estão dispostos em apenas 87 estações. Ainda dentro das configurações dos dados gerados, foram utilizadas as medidas repetidas $J_1 = 12$ e $J_2 = J_3 = 9$, com os parâmetros do modelo tomados a partir dos valores apresentados na Tabela 2, utilizando para isso $\kappa = 0.2$, $K_\mu = 9$, $K_\zeta = 5$ e $K_\Gamma = 7$. Lembramos que a condição $J_2 = J_3$ será adotada em todas as simulações que iremos apresentar nesta seção, uma vez que este comportamento também é observado no conjunto de dados reais.

Na Figura 5, apresentamos os comportamentos da mediana a posteriori para os parâmetros ω^2, φ e ϕ, considerando os 50 conjuntos de dados simulados quando variamos o número de localizações (n) em 62, 87 e 112, e a quantidade de meses observados (τ) entre 3, 6 e 9. A configuração $n = 62$ e $\tau = 3$ não aparece nos gráficos dos parâmetros em estudo, pois neste caso as cadeias não convergiam, mesmo após tentativas de mudanças nos valores iniciais no processo de estimação. É possível observar ainda que ω^2, φ e ϕ apresentam, em todos os cenários, estimativas a posteriori concentradas em torno dos verdadeiros valores dos respectivos parâmetros e, neste caso, o acréscimo em qualquer uma das dimensões avaliadas, n e/ou τ, acarreta em uma concentração ainda maior das estimativas em torno dos respectivos valores verdadeiros com redução da variabilidade nas amostras a posteriori. Em outras palavras, o aumento de n e/ou τ produz estimativas mais próximas dos respectivos verdadeiros valores dos parâmetros, assim como aumenta a precisão do processo de estimação. Vale ressaltar ainda que essas mesmas características foram observadas para a média e moda a posteriori (veja Apêndice C).

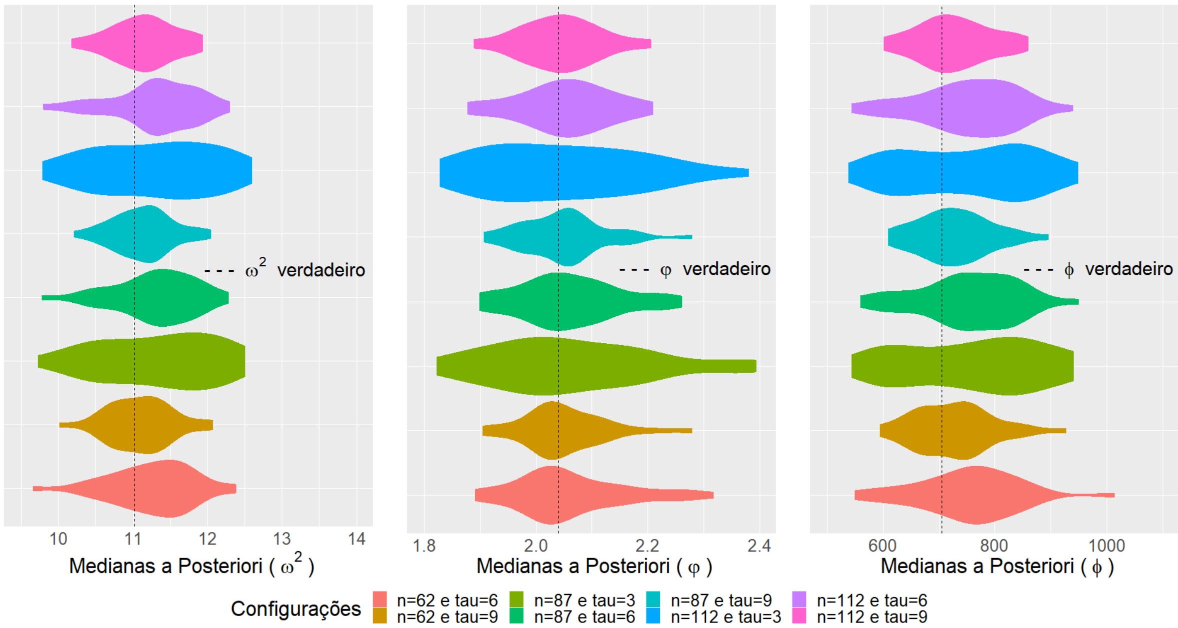

Figura 5 – Medianas a posteriori para os parâmetros ω^2, φ e ϕ, tomadas ao longo dos 50 conjuntos de dados simulados independentemente variando o número de localizações (n) entre 62, 87 e 112, e o número de períodos observados (τ) entre 3, 6 e 9.

Para as medianas a posteriori dos parâmetros $\sigma^2_{\theta_i}$ e $\sigma^2_{\vartheta_i}$, $i = 1, 2, 3$, que representam respectivamente as componentes de variância dos efeitos espaciais e temporais, ilustrados nas Figuras 6 e 7, nessa ordem, observamos um comportamento assimétrico das estimativas obtidas nas amostras a posteriori que, em geral, estão subestimadas quando comparadas aos respectivos verdadeiros valores. Neste caso, notamos ainda que aparentemente o aumento em qualquer uma das dimensões (n e/ou J_i, $i = 1, 2, 3$) não impacta diretamente em um aumento de precisão para estes parâmetros. É importante ressaltar que este comportamento é similar para a média e moda a posteriori (veja Apêndice C).

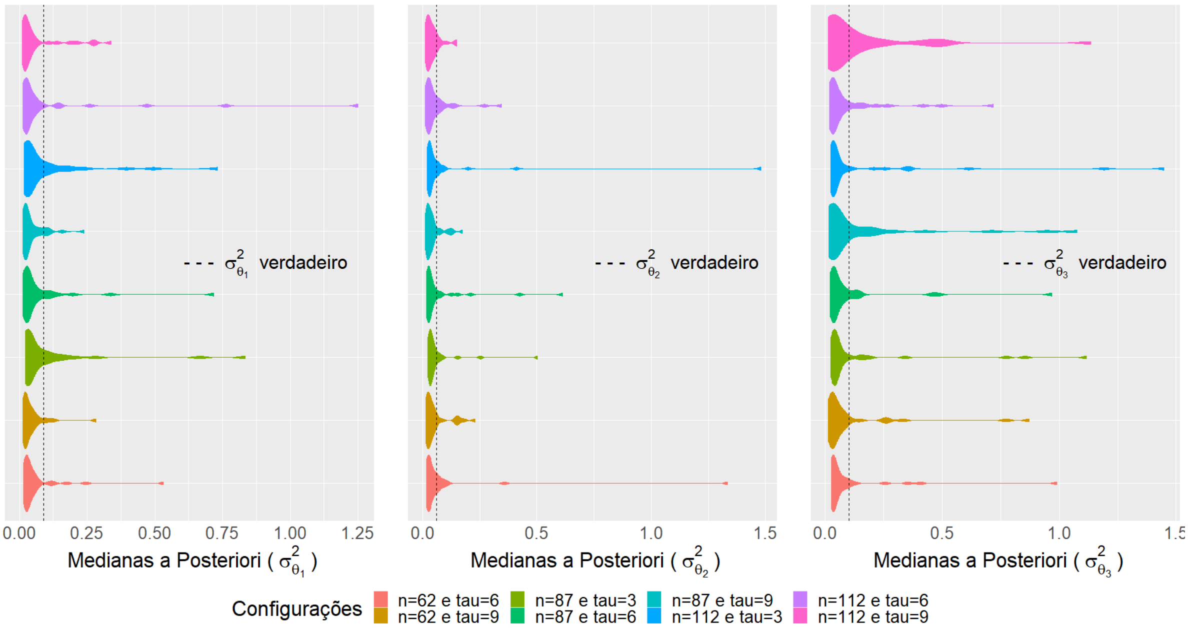

Figura 6 – Medianas a posteriori para os parâmetros $\sigma^2_{\theta_i}$, $i = 1, 2, 3$, tomadas ao longo dos 50 conjuntos de dados simulados independentemente variando o número de localizações (n) entre 62, 87 e 112, e o número de períodos observados (τ) entre 3, 6 e 9.

Na Tabela 7, apresentamos as probabilidades de cobertura para os parâmetros do modelo proposto em (2.1), obtidas a partir desse estudo de simulação. Observamos que os percentuais de cobertura obtidos estão próximos de 100% sendo que, em todos os casos, o limitante inferior para os percentuais avaliados foi de 90%.

O segundo estudo de simulação abordado nesta seção tem como objetivo encontrar evidências do impacto do incremento no número de medidas repetidas J_1, J_2 e J_3, dentro de cada um dos três blocos, nas estimativas dos parâmetros do modelo proposto em (2.1). Para tanto, foram gerados 50 conjuntos de dados independentes considerando $J_1 = 12$, $J_2 = J_3 = 9$, $n = 87$ e $\tau = 6$, sendo que as configurações geográficas foram herdadas dos dados reais com valores dos parâmetros tomados a partir do conjunto de dados apresentado na Tabela 2. Adotados ainda $\kappa = 0.2$, $K_\mu = 9$, $K_\zeta = 5$ e $K_\Gamma = 7$.

A Figura 8 apresenta o comportamento da mediana a posteriori para os parâmetros ω^2, φ e ϕ, do modelo proposto em (2.1), tomadas ao longo dos 50 conjuntos de dados simulados. Para tanto, variamos o número de medidas repetidas em cada um dos blocos presentes no modelo respeitando a variação de + ou − 3 unidades em torno da

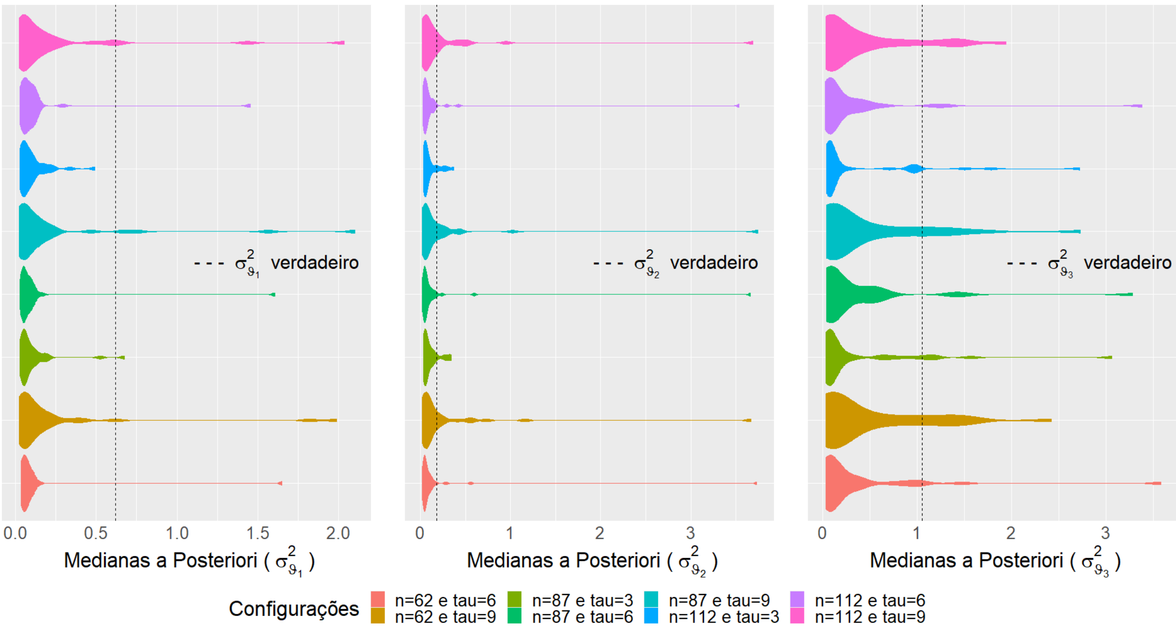

Figura 7 – Medianas a posteriori para os parâmetros $\sigma^2_{\vartheta_i}$, $i = 1, 2, 3$, tomadas ao longo dos 50 conjuntos de dados simulados independentemente variando o número de localizações (n) entre 62, 87 e 112, e o número de períodos observados (τ) entre 3, 6 e 9.

Tabela 7 – Probabilidades de cobertura para os parâmetros do modelo proposto em (2.1), tomadas ao longo dos 50 conjuntos de dados gerados independentemente variando o número de localizações (n) entre 62, 87 e 112, e o número de meses observados (τ) entre 3, 6 e 9.

	Probabilidades de Cobertura								
	$\tau = 3$			$\tau = 6$			$\tau = 9$		
Parâmetros	$n = 62$	$n = 87$	$n = 112$	$n = 62$	$n = 87$	$n = 112$	$n = 62$	$n = 87$	$n = 112$
ω^2	$--$	100%	98%	96%	98%	96%	96%	100%	100%
$\sigma^2_{\theta_1}$	$--$	100%	100%	100%	100%	100%	100%	100%	100%
$\sigma^2_{\theta_2}$	$--$	100%	100%	100%	100%	100%	100%	100%	100%
$\sigma^2_{\theta_3}$	$--$	100%	100%	100%	100%	100%	100%	100%	100%
$\sigma^2_{\vartheta_1}$	$--$	100%	100%	100%	100%	100%	100%	100%	100%
$\sigma^2_{\vartheta_2}$	$--$	100%	100%	100%	100%	100%	100%	100%	100%
$\sigma^2_{\vartheta_3}$	$--$	100%	100%	100%	96%	96%	100%	98%	98%
φ	$--$	92%	92%	96%	94%	90%	94%	92%	90%
ϕ	$--$	100%	100%	96%	96%	96%	98%	98%	100%

dimensão dos dados reais ($J_1 = 9$, $J_2 = J_3 = 6$). Como podemos observar, de modo geral, fixado o número de medidas em qualquer uma das dimensões do modelo proposto, ou seja, em J_1 ou $J_2 = J_3$, o aumento na dimensão restante acarretará em uma concentração das medianas a posteriori em torno dos respectivos valores verdadeiros. No Apêndice C, é possível verificar ainda que tanto a média quanto a moda a posteriori possuem o mesmo comportamento descrito pela mediana a posteriori, explicado a partir do impacto do número de medidas repetidas sobre o processo de estimação dos parâmetros. Para maiores detalhes veja Apêndice C.

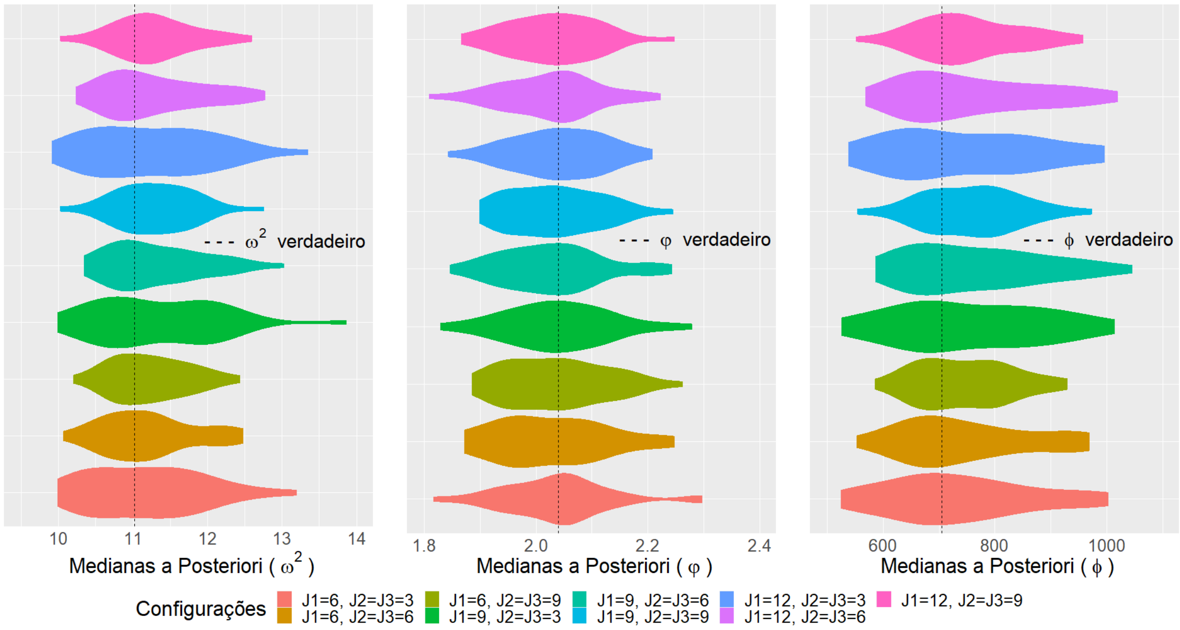

Figura 8 – Medianas a posteriori para os parâmetros ω^2, φ e ϕ, tomadas ao longo dos 50 conjuntos de dados simulados independentemente variando o número de repetidas por bloco, J_i, $i = 1, 2, 3$.

Quanto às medianas a posteriori para os parâmetros $\sigma^2_{\theta_i}$ e $\sigma^2_{\vartheta_i}$, $i = 1, 2, 3$, que representam as componentes de variância dos efeitos espaciais e temporais, respectivamente, ilustrados reciprocamente nas Figuras 9 e 10, podemos observar que, de modo geral, o incremento no número de medidas repetidas não acarreta em melhoria da precisão para o processo de estimação destas componentes de variância. Essa mesma característica pode ser observada tanto para média quanto para a moda a posteriori (veja Apêndice C).

Finalmente, na Tabela 8, apresentamos as probabilidades de cobertura para o conjunto das quantidades desconhecidas do modelo proposto em (2.1), no qual não foi possível identificar o impacto devido às mudanças atribuídas aos números de medidas repetidas. Ademais, notamos ainda que para as componentes com maior assimetria ($\sigma^2_{\theta_i}$ e $\sigma^2_{\vartheta_i}$, $i = 1, 2, 3$) observamos percentuais de probabilidades de cobertura próximos a 100%, sendo que os demais parâmetros apresentam percentuais de probabilidades de cobertura maiores ou iguais a 88%.

O intenso estudo de simulação apresentado nesta seção mostrou que, dentre os cenários avaliados, as estimativas a posteriori dos parâmetros ω^2, φ e ϕ são afetadas por n, τ e também pelo número de medidas repetidas J_i, $i = 1, 2, 3$. Por outro lado, as componentes de variância dos efeitos espaciais e temporais, denotadas respectivamente por $\sigma^2_{\theta_i}$ e $\sigma^2_{\vartheta_i}$, $i = 1, 2, 3$, não são afetadas por n, τ e J_i, $i = 1, 2, 3$. Em outras palavras, temos indícios estatísticos de que quanto maior for a quantidade de informação apresentada no conjunto

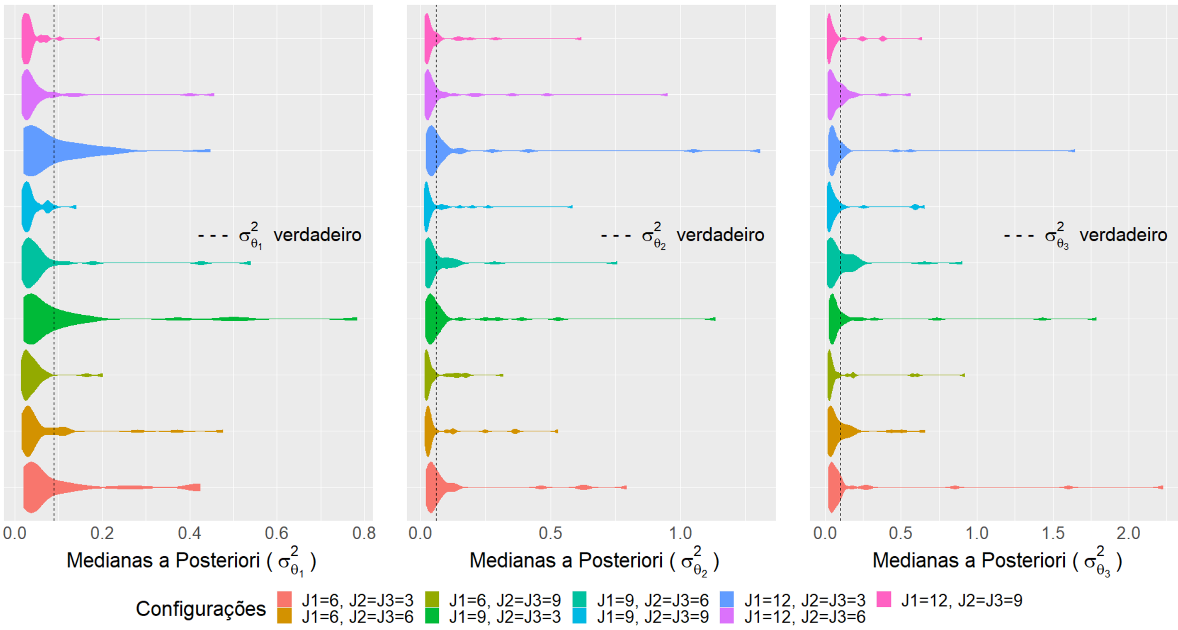

Figura 9 – Medianas a posteriori para os parâmetros $\sigma^2_{\theta_i}$, $i = 1, 2, 3$, tomadas ao longo dos 50 conjuntos de dados simulados independentemente variando o número de repetidas por bloco, J_i, $i = 1, 2, 3$.

Tabela 8 – Probabilidades de cobertura para os parâmetros do modelo proposto em (2.1), tomadas ao longo dos 50 conjuntos de dados gerados independentemente variando o número de repetidas por bloco, J_i, $i = 1, 2, 3$.

	Probabilidades de Cobertura								
	$j_2 = j_3 = 3$			$j_2 = j_3 = 6$			$j_2 = j_3 = 9$		
Parâmetros	$j_1 = 6$	$j_1 = 9$	$j_1 = 12$	$j_1 = 6$	$j_1 = 9$	$j_1 = 12$	$j_1 = 6$	$j_1 = 9$	$j_1 = 12$
ω^2	96%	96%	90%	94%	92%	88%	96%	94%	88%
$\sigma^2_{\theta_1}$	100%	100%	100%	100%	100%	100%	100%	100%	100%
$\sigma^2_{\theta_2}$	100%	100%	100%	100%	100%	100%	100%	100%	100%
$\sigma^2_{\theta_3}$	100%	100%	100%	100%	100%	100%	100%	100%	100%
$\sigma^2_{\vartheta_1}$	100%	100%	100%	100%	100%	100%	100%	98%	100%
$\sigma^2_{\vartheta_2}$	100%	100%	100%	100%	100%	98%	100%	100%	100%
$\sigma^2_{\vartheta_3}$	100%	98%	98%	86%	100%	90%	90%	100%	98%
φ	96%	92%	96%	90%	90%	92%	90%	92%	90%
ϕ	96%	92%	92%	96%	94%	96%	100%	92%	92%

de dados a ser estimado, melhor será o comportamento a posteriori das estimativas dos parâmetros ω^2, φ e ϕ. É importante salientar que as estruturas algébricas das componentes espacial e temporal são compostas por produtos tensoriais que podem conduzir a uma maior sensibilidade no processo de estimação dos parâmetros. Por outro lado, tal estrutura algébrica não apresenta um reflexo direto no processo de previsão/predição. Neste sentido, apresentaremos, na Seção 5.4, um estudo dedicado ao comportamento preditivo do modelo proposto.

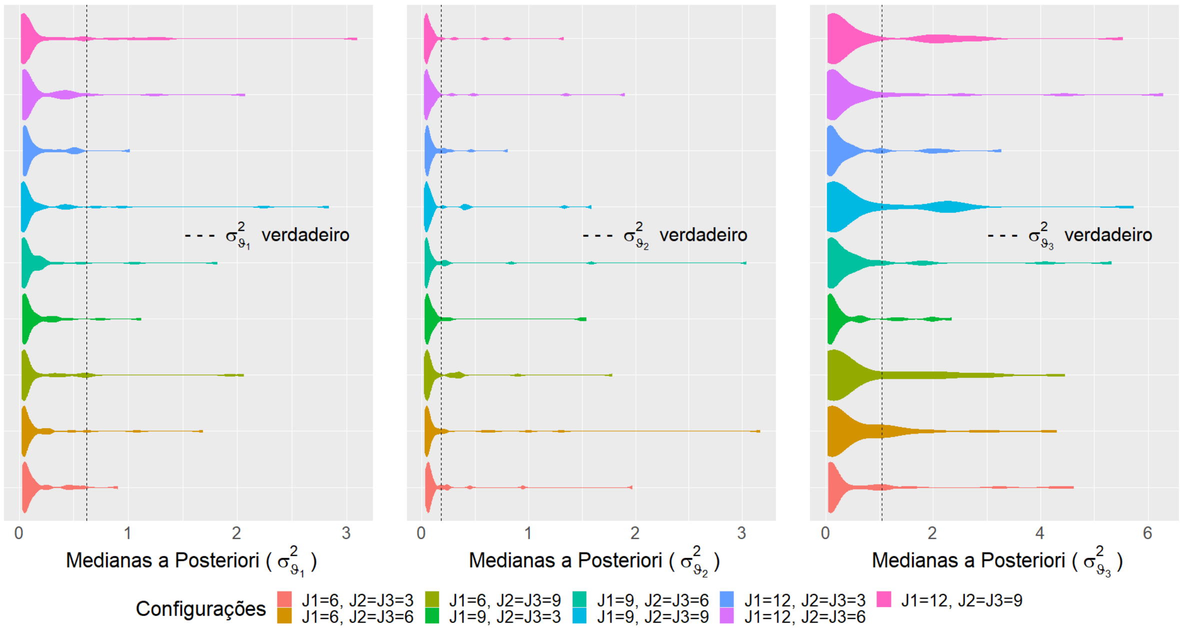

Figura 10 – Medianas a posteriori para os parâmetros $\sigma^2_{\vartheta_i}$, $i = 1, 2, 3$, tomadas ao longo dos 50 conjuntos de dados simulados independentemente variando o número de repetidas por bloco, J_i, $i = 1, 2, 3$.

5.3 Escolha do Parâmetro κ

Nesta seção, realizamos um estudo de simulação para avaliar a capacidade dos critérios de seleção de modelos LPML e DIC7, apresentados no Capítulo 4, em resgatar o verdadeiro valor do parâmetro κ, apresentado em (2.10), a partir da adequabilidade do modelo ajustado. Com essa finalidade, para cada valor de $\kappa \in U = \{0.2, 0.5, 1, 1.5, 2\}$ foram gerados 100 conjuntos de dados independentes utilizando os valores dos parâmetros apresentados na Tabela 2, mantendo $K_\mu = 9$, $K_\zeta = 5$, $K_\Gamma = 7$, $\tau = 6$, $J_1 = 9$ e $J_2 = J_3 = 6$. O conjunto de distribuições a priori utilizadas nesta seção está apresentado na Tabela 6.

Para cada conjunto de dados gerado, o modelo proposto em (2.1) foi ajustado fixando-se $\kappa \in U$. Em seguida, os critérios LPML e DIC7 foram considerados para apontar o melhor ajuste. O ideal é que o valor de κ indicado pelos critérios de seleção de modelos seja igual ao utilizado na geração do respectivo conjunto de dados. A Tabela 9 apresenta os resultados encontrados. Podemos observar que ambos os critérios apresentaram taxas de acerto superiores a 98%, o que nos traz indícios favoráveis à utilização dos critérios LPML e DIC7 para o processo de estimação do parâmetro κ.

Tabela 9 – Porcentagem de acertos dos critérios de seleção de modelos LPML e DIC7 associados a escolha do verdadeiro valor de κ descrito na estrutura de correlação *Matérn* (2.10).

Verdadeiro valor de κ	Acerto (%)	
	LPML	DIC7
0.2	100 %	100 %
0.5	99 %	99 %
1.0	98 %	98 %
1.5	98 %	98 %
2.0	99 %	99 %

5.4 Distribuição Preditiva a Posteriori

Nesta seção, abordamos um estudo de simulação para avaliar a capacidade preditiva do modelo proposto em (2.1). A distribuição preditiva acomoda a incerteza quando queremos prever o comportamento de uma unidade experimental não considerada no ajuste do modelo. De modo geral, utilizamos a distribuição a posteriori para avaliar a incerteza de uma observação futura, marginalizando a verossimilhança a partir da distribuição a posteriori.

Considere $\mathbf{y}_{new}$ um novo conjunto de respostas ainda não observadas. Assim, dada a resposta não observada $\mathbf{y}_{new}$, a distribuição preditiva a posteriori segundo o modelo proposto em (2.1) é dada por

$$
\begin{aligned}
p(\mathbf{y}_{new}|\mathbf{y}) &= \int_{\boldsymbol{\xi}}\int_{\boldsymbol{\vartheta}}\int_{\boldsymbol{\Theta}} p(\mathbf{y}_{new}|\mathbf{y},\boldsymbol{\xi},\boldsymbol{\vartheta},\boldsymbol{\Theta})p(\boldsymbol{\xi},\boldsymbol{\vartheta},\boldsymbol{\Theta}|\mathbf{y})d\boldsymbol{\xi}d\boldsymbol{\vartheta}d\boldsymbol{\Theta} \\
&\int_{\boldsymbol{\xi}}\int_{\boldsymbol{\vartheta}}\int_{\boldsymbol{\Theta}} p(\mathbf{y}_{new}|\boldsymbol{\xi},\boldsymbol{\vartheta},\boldsymbol{\Theta})p(\boldsymbol{\xi},\boldsymbol{\vartheta},\boldsymbol{\Theta}|\mathbf{y})d\boldsymbol{\xi}d\boldsymbol{\vartheta}d\boldsymbol{\Theta},
\end{aligned} \tag{5.1}
$$

no qual a segunda igualdade considera a independência condicional entre $\mathbf{y}_{new}$ e $\mathbf{y}$ e, portanto, $p(\mathbf{y}_{new}|\mathbf{y}) = E_{\boldsymbol{\xi},\boldsymbol{\vartheta},\boldsymbol{\Theta}}(p(\mathbf{y}_{new}|\boldsymbol{\xi},\boldsymbol{\vartheta},\boldsymbol{\Theta}))$. Utilizando MCMC é possível obter amostras da distribuição conjunta a posteriori (GAMERMAN; LOPES, 2006) a partir da estimativa da distribuição preditiva dada por

$$
\widehat{p(\mathbf{y}_{new}|\mathbf{y})} = \frac{1}{\eta}\sum_{h=1}^{\eta} p(\mathbf{y}_{new}|\boldsymbol{\xi}^h,\boldsymbol{\vartheta}^h,\boldsymbol{\Theta}^h), \tag{5.2}
$$

em que $\boldsymbol{\xi}^h$, $\boldsymbol{\vartheta}^h$ e $\boldsymbol{\Theta}^h$, $h = 1,...,\eta$, são amostras da distribuição a posteriori $p(\boldsymbol{\xi},\boldsymbol{\vartheta},\boldsymbol{\Theta}|\mathbf{y})$ obtidas por meio do amostrador de Gibbs descrito no Capítulo 3.

5.4.1 Capacidade Preditiva do Modelo Proposto em (2.1)

Para avaliarmos os resultados do estudo sobre o comportamento preditivo do modelo proposto em (2.1), tomamos $\kappa = 0.2$, $K_\mu = 9$, $K_\zeta = 5$ e $K_\Gamma = 7$. As distribuições a priori são apresentadas na Tabela 6. Para um primeiro estudo, será gerado um novo conjunto de dados utilizando 113 localizações ($n = 113$), $\tau = 6$, 13 medidas repetidas para o bloco $i = 1$ ($J_1 = 13$) e 10 medidas repetidas para cada um dos blocos $i = 2, 3$ ($J_2 = J_3 = 10$), imitando a estrutura avaliada na Seção 5.2.

Para obtermos dados de teste em uma nova localização retiramos do banco de dados a última medida repetida gerada ao longo de cada um dos blocos e, em seguida, removemos de forma aleatória uma das 113 localizações inicialmente consideradas. Desta forma, o conjunto de dados de treino terá uma dimensão máxima de $n = 112$, $\tau = 6$, $J_1 = 12$ e $J_2 = J_3 = 9$. As Figuras 11 a 13 apresentam os intervalos de predição de 95%, avaliados na localização teste para cada um dos blocos, quando variamos a partir de sorteios aleatórios a quantidade de informação a partir de J_i, $i = 1, 2, 3$, dentro do processo de treino do modelo. Como podemos notar, em todos os casos as medidas de teste estão inteiramente contidas nos seus respectivos intervalos de predição. No entanto, o acréscimo de informação provinda a partir da variação no número de medidas repetidas (J_i, $i = 1, 2, 3$) dentro de cada um dos blocos não gera uma redução da banda de credibilidade nos intervalos de predição.

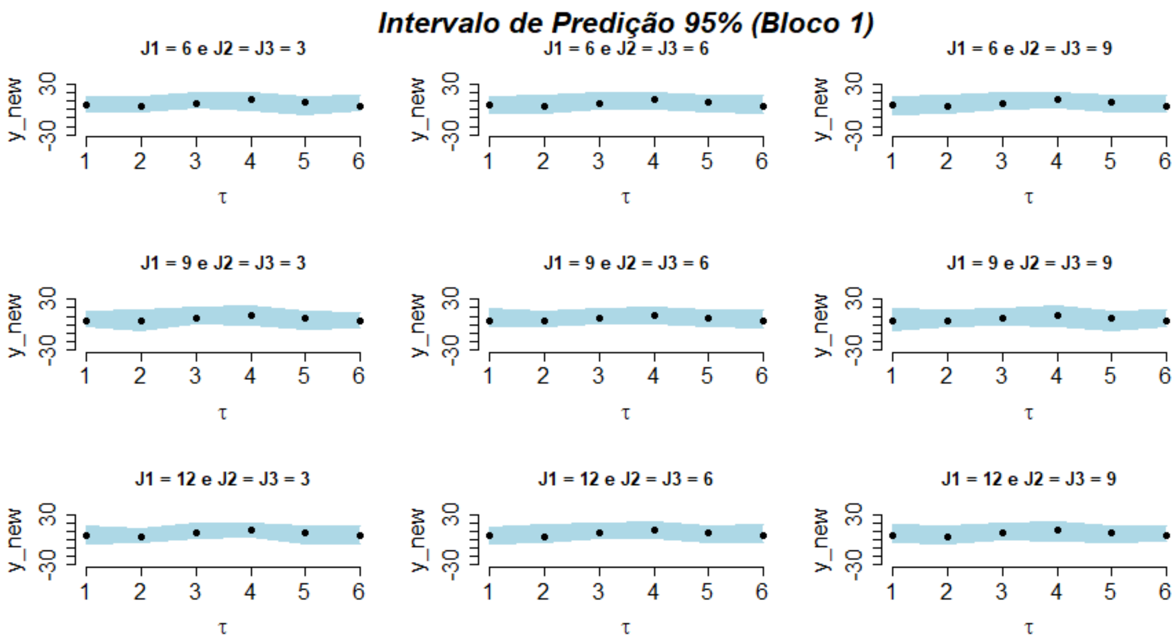

Figura 11 – Intervalos de predição de 95% para uma localização teste avaliada no bloco $i = 1$, tomados a partir das estimativas do modelo proposto em (2.1) variando J_i, $i = 1, 2, 3$, utilizando $\tau = 6$, $\kappa = 0.2$, $K_\mu = 9$, $K_\zeta = 5$ e $K_\Gamma = 7$.

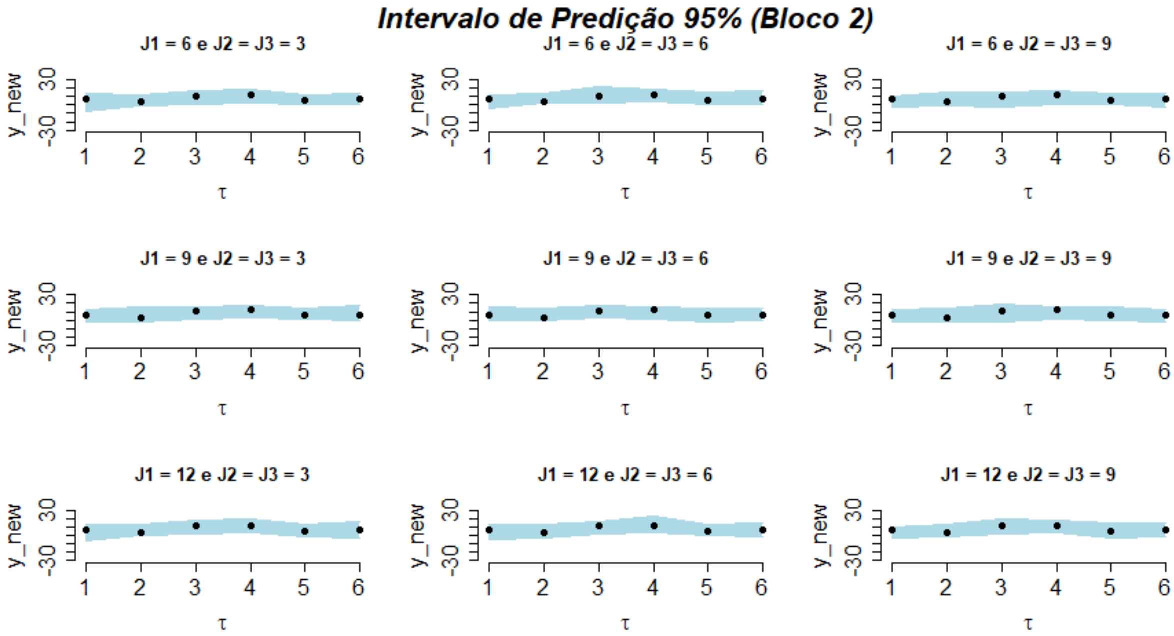

Figura 12 – Intervalos de predição de 95% para uma localização teste avaliada no bloco $i = 2$, obtidos a partir das estimativas do modelo proposto em (2.1) variando J_i, $i = 1, 2, 3$, utilizando $\tau = 6$, $\kappa = 0.2$, $K_\mu = 9$, $K_\zeta = 5$ e $K_\Gamma = 7$.

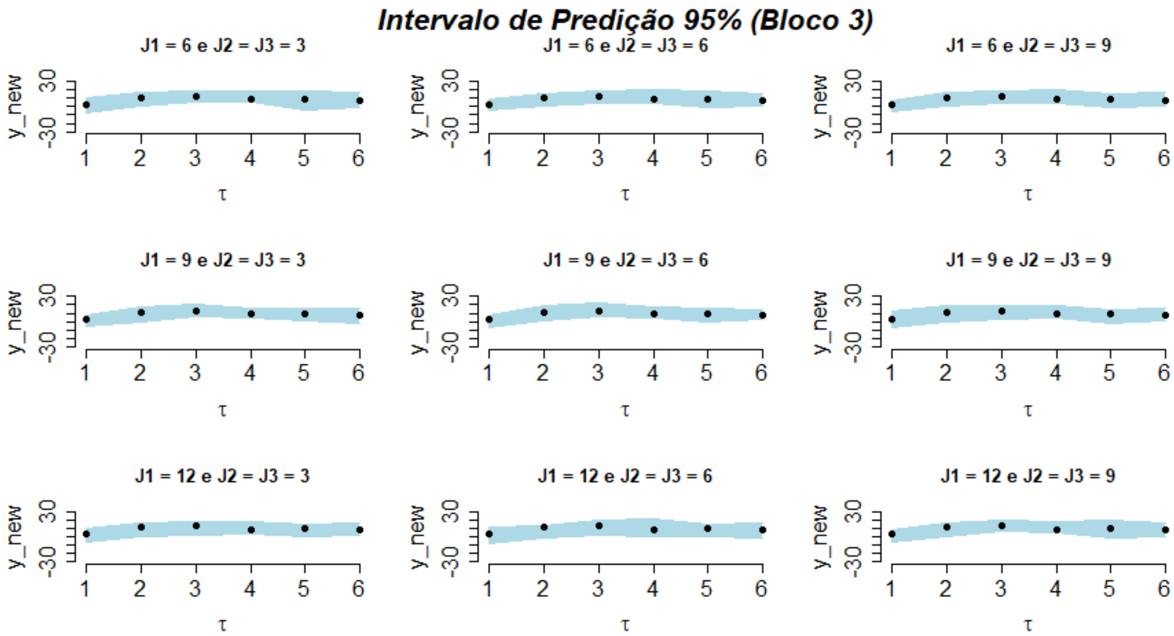

Figura 13 – Intervalos de predição de 95% para uma localização teste avaliada no bloco $i = 3$, obtidos a partir das estimativas do modelo proposto em (2.1) variando J_i, $i = 1, 2, 3$, utilizando $\tau = 6$, $\kappa = 0.2$, $K_\mu = 9$, $K_\zeta = 5$ e $K_\Gamma = 7$.

Para o nosso próximo estudo, foi gerado um conjunto de dados utilizando 113 localizações ($n = 113$), $\tau = 9$, 10 medidas repetidas para o bloco $i = 1$ ($J_1 = 10$) e 7 medidas repetidas para cada um dos blocos $i = 2, 3$ ($J_2 = J_3 = 7$), imitando a estrutura avaliada nos dados reais. Conforme descrito anteriormente, foi selecionada aleatoriamente uma localização teste e a última medida repetida gerada ao longo de cada um dos blocos foi retirada, reduzindo portanto o conjunto de dados treino à dimensão máxima de $n = 112$, $\tau = 9$, $J_1 = 9$ e $J_2 = J_3 = 6$. O objetivo agora é avaliar o comportamento preditivo do modelo, em uma localização teste, quando variamos a partir de sorteios aleatórios os valores de n e τ dentro do processo de treino do modelo.

As Figuras 14 a 16 apresentam os intervalos de predição de 95% para os cenários supracitados. Conforme já mencionado anteriormente, o caso $n = 62$ e $\tau = 3$ não foi ilustrado uma vez que, neste caso, as cadeias não convergiram, não sendo admissível extrair amostras das distribuições a posteriori dos parâmetros do modelo proposto em (2.1). É possível observar então que, em todos os casos, as novas medidas estão inteiramente contidas nos seus respectivos intervalos de predição sendo que, o acréscimo de informação provinda a partir da variação de n ou τ contribui claramente para uma menor magnitude dos intervalos, auxiliando assim a capacidade preditiva do modelo. Vale ressaltar ainda que outras simulações similares às apresentadas aqui foram realizadas, retornando resultados equivalentes aos apresentados nesta seção.

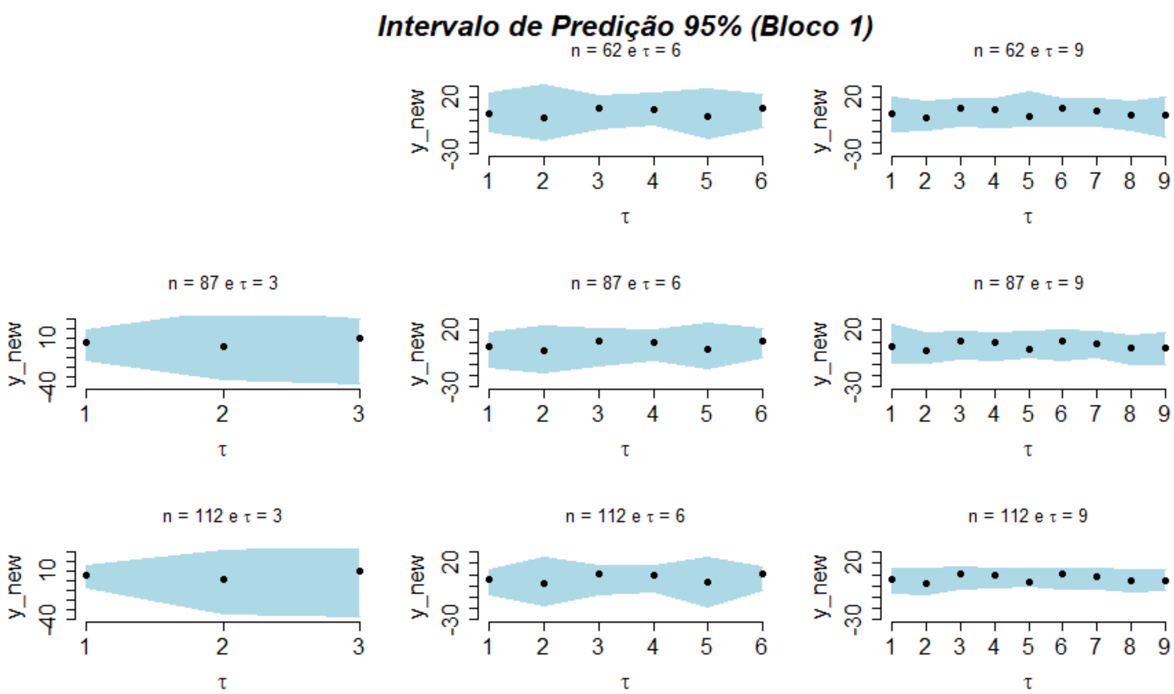

Figura 14 – Intervalos de predição de 95% para uma localização teste avaliada no bloco $i = 1$, obtidos a partir das estimativas do modelo proposto em (2.1) variando n e τ, utilizando $J_1 = 9$, $J_2 = J_3 = 6$, $\kappa = 0.2$, $K_\mu = 9$, $K_\zeta = 5$ e $K_\Gamma = 7$.

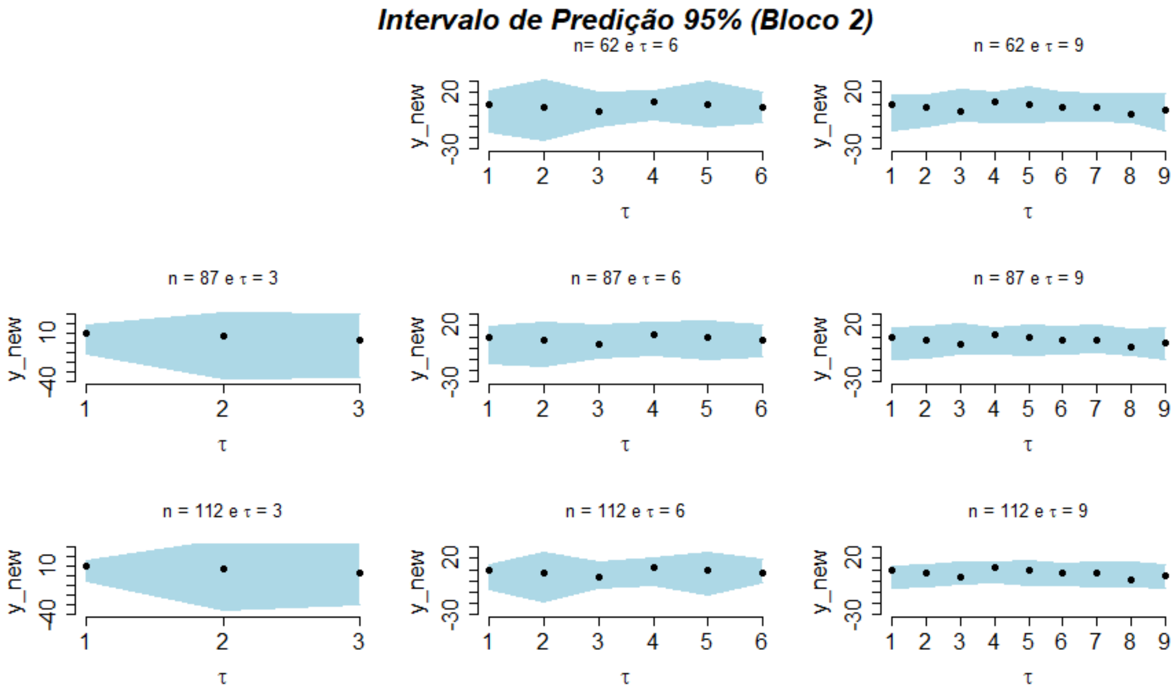

Figura 15 – Intervalos de predição de 95% para uma localização teste avaliada no bloco $i = 2$, obtidos a partir das estimativas do modelo proposto em (2.1) variando n e τ, utilizando $J_1 = 9$, $J_2 = J_3 = 6$, $\kappa = 0.2$, $K_\mu = 9$, $K_\zeta = 5$ e $K_\Gamma = 7$.

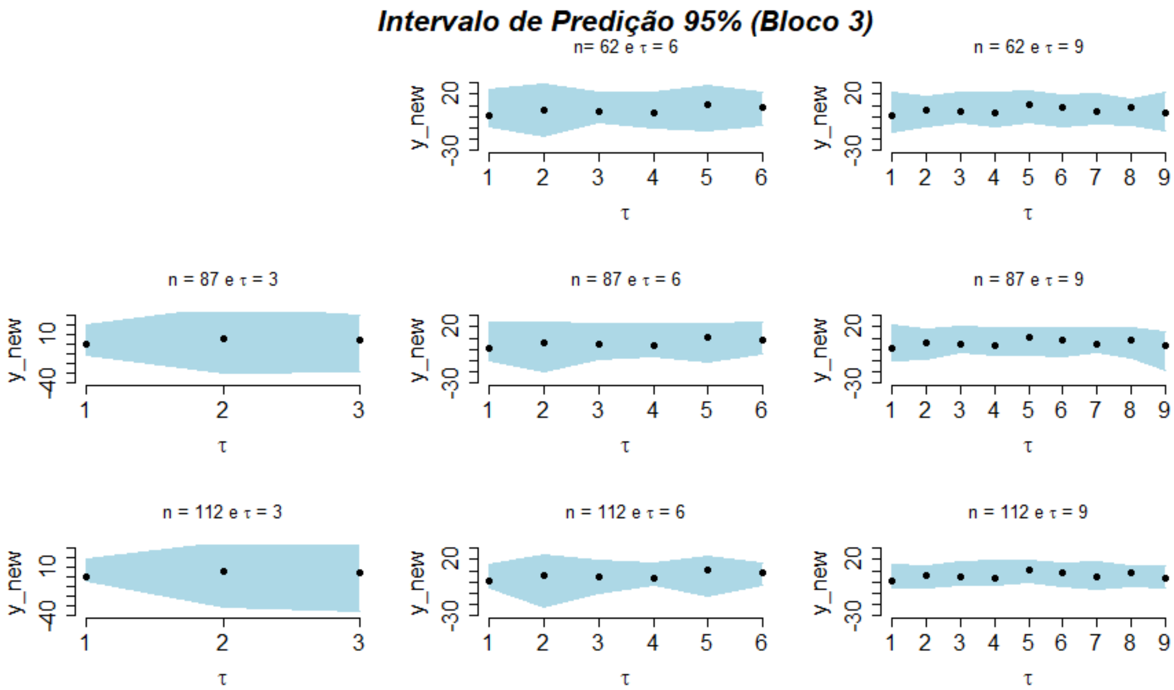

Figura 16 – Intervalos de predição de 95% para uma localização teste avaliada no bloco $i = 3$, obtidos a partir das estimativas do modelo proposto em (2.1) variando n e τ, utilizando $J_1 = 9$, $J_2 = J_3 = 6$, $\kappa = 0.2$, $K_\mu = 9$, $K_\zeta = 5$ e $K_\Gamma = 7$.

5.5 Adequabilidade da estimação empírica dos parâmetros em espaços ortogonais: Resultados Empíricos

A abordagem proposta nesta seção surgiu como uma alternativa ao alto custo computacional requerido durante o processo de estimação dos parâmetros do modelo proposto em (2.1), e tem como objetivo obter estimativas em espaços ortogonais. Essa nova perspectiva gera uma redução no tempo computacional necessário para a execução do algoritmo MCMC, pois neste caso, os efeitos fixos são obtidos diretamente por mínimos quadrados enquanto que os outros parâmetros por meio de MCMC.

Na literatura, outros autores já propuseram a utilização de transformações como uma estratégia para a obtenção de estimadores em modelos de regressão. Como exemplo, temos o método da máxima verossimilhança restrita proposto por Patterson e Thompson (1971).

É importante ressaltar que na tentativa de encontrarmos a comprovação matemática para os resultados que iremos apresentar, reduzimos a parametrização do modelo proposto em (2.1) tomando $\sigma^2_{\theta_i} = \sigma^2_{\theta}$, $\sigma^2_{\vartheta_i} = \sigma^2_{\vartheta}$ e $\mathbf{\Sigma}_{\epsilon_{ij}} = \omega^2\mathbf{I}$. Vale ressaltar que apesar de os estudos de simulação apresentarem promissores resultados, não foi possível encontrar um limitante teórico para a diferença entre as estimativas.

5.5.1 Projeção em Espaços Ortogonais

Para este estudo vamos denotar por $\mathbf{H} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$ e assumir que as matrizes $\mathbf{Q}$ e $\mathbf{R}$, apresentadas em (2.9), não são ortogonais ao espaço coluna de $\mathbf{X}$, ou seja, $(\mathbf{I}-\mathbf{H})\mathbf{Q} \neq \mathbf{0}$ e $(\mathbf{I}-\mathbf{H})\mathbf{R} \neq \mathbf{0}$. Projetando o modelo proposto no espaço coluna de $\mathbf{X}$ temos que:

$$(\mathbf{I}-\mathbf{H})\mathbf{Y} = (\mathbf{I}-\mathbf{H})\mathbf{Q}\boldsymbol{\Theta} + (\mathbf{I}-\mathbf{H})\mathbf{R}\boldsymbol{\vartheta} + (\mathbf{I}-\mathbf{H})\mathbf{E}, \tag{5.3}$$

uma vez que $(\mathbf{I}-\mathbf{H})\mathbf{X} = \mathbf{0}$. Note que essa projeção gera uma restrição sobre o vetor de parâmetro $\boldsymbol{\beta}$, que agora deixa de representar um parâmetro direto do modelo e passa a ser definido pela seguinte relação:

$$\tilde{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T(\mathbf{y} - \mathbf{Q}\tilde{\boldsymbol{\Theta}} - \mathbf{N}\tilde{\boldsymbol{\vartheta}}), \tag{5.4}$$

em que $\tilde{\boldsymbol{\Theta}}$ e $\tilde{\boldsymbol{\vartheta}}$ são as estimativas a posteriori de $\boldsymbol{\Theta}$ e $\boldsymbol{\vartheta}$, respectivamente.

Como podemos observar, a projeção gera uma estrutura de estimação muito parecida como a que foi descrita para o modelo proposto em (2.1), sendo que agora

$(\mathbf{I}-\mathbf{H})\mathbf{E} \sim \mathbf{N}[\mathbf{0}, (\mathbf{I}-\mathbf{H})\boldsymbol{\Sigma}_\epsilon(\mathbf{I}-\mathbf{H})^T]$. Como $(\mathbf{I}-\mathbf{H})$ tem posto incompleto segue, portanto, que o vetor de erros possui matriz de covariância semi-positiva definida, o que caracteriza uma distribuição normal degenerada. Essa condição implica que as distribuições condicionais completas de $\boldsymbol{\Theta}$ e $\boldsymbol{\vartheta}$ também serão degeneradas (veja Tabela 1). Logo, ao utilizarmos o algoritmo amostrador de Gibbs para se obterem amostras da distribuição a posteriori, a partir de condicionais completas degeneradas, adotaremos as técnicas propostas em Srivastava e Rosen (2002). Vale ressaltar que tais técnicas supõem que as auto-funções das distribuições degeneradas sejam não colineares (veja Dony et al. (2001)).

Um resultado importante que pode ser observado ao compararmos o núcleo da distribuição a posteriori das componentes de variância, a saber ω^2, σ_θ^2 e σ_ϑ^2, quando avaliado tanto para o modelo proposto em (2.1) quanto para o modelo apresentado na forma projetada, proposta em (5.3), é que sobre determinada condição o quociente obtido entre esses núcleos é proporcional a uma constante. Ou seja, estabelecida determinada condição, a projecão $(\mathbf{I}-\mathbf{H})$ não afeta o processo de obtenção de amostras a posteriori para as componentes de variância.

Para melhor entendermos a condição necessária para que o resultado supracitado seja válido, considere inicialmente o modelo apresentado na forma usual proposta em (2.1). De acordo com (3.2), a distribuição conjunta a posteriori de ω^2, σ_θ^2 e σ_ϑ^2 aumentada em $\boldsymbol{\Theta}$ e $\boldsymbol{\vartheta}$ é descrita por:

$$\begin{aligned} p(\boldsymbol{\beta}, \boldsymbol{\Theta}, \boldsymbol{\vartheta}, \omega^2, \sigma_\theta^2, \sigma_\vartheta^2|\mathbf{y}, \boldsymbol{\Delta}) &\propto p(\mathbf{y}|\boldsymbol{\beta}, \boldsymbol{\Theta}, \boldsymbol{\vartheta}, \omega^2, \sigma_\theta^2, \sigma_\vartheta^2)p(\boldsymbol{\beta}|\sigma_\beta^2)p(\boldsymbol{\Theta}|\sigma_\theta^2)p(\boldsymbol{\vartheta}|\sigma_\vartheta^2) \\ &\times p(\omega^2, \sigma_\theta^2, \sigma_\vartheta^2|\boldsymbol{\Delta}), \end{aligned} \tag{5.5}$$

em que $p(\mathbf{y}|\boldsymbol{\beta}, \boldsymbol{\Theta}, \boldsymbol{\vartheta}, \omega^2, \sigma_\theta^2, \sigma_\vartheta^2)$ representa a densidade de uma distribuição normal com média $\mathbf{X}\boldsymbol{\beta} + \mathbf{Q}\boldsymbol{\Theta} + \mathbf{R}\boldsymbol{\vartheta}$ e matriz de covariância $\omega^2\mathbf{I}_{n(J_1+J_2+J_3)\tau}$, $p(\boldsymbol{\Theta}|\sigma_\theta^2)$ é a densidade de uma distribuição normal com média zero e matriz de covariância $\sigma_\theta^2\mathbf{I}_{K_\zeta^2}$, $p(\boldsymbol{\vartheta}|\sigma_\vartheta^2)$ é a densidade de uma distribuição normal com média zero e matriz covariância $\sigma_\vartheta^2\mathbf{I}_{K_\Gamma^2}$ e, finalmente, $\boldsymbol{\Delta}$ é o vetor de hiperparâmetros. Integrando (5.5) em relação a $\boldsymbol{\beta}$, temos que:

$$\begin{aligned} p(\boldsymbol{\Theta}, \boldsymbol{\vartheta}, \omega^2, \sigma_\theta^2, \sigma_\vartheta^2|\mathbf{y}, \boldsymbol{\Delta}) &\propto p(\boldsymbol{\Theta}|\sigma_\theta^2)p(\boldsymbol{\vartheta}|\sigma_\vartheta^2)p(\omega^2, \sigma_\theta^2, \sigma_\vartheta^2|\boldsymbol{\Delta}) \\ &\times \int_{\boldsymbol{\beta}} p(\mathbf{y}|\boldsymbol{\beta}, \boldsymbol{\Theta}, \boldsymbol{\vartheta}, \omega^2, \sigma_\theta^2, \sigma_\vartheta^2)p(\boldsymbol{\beta}|\sigma_\beta^2)d\boldsymbol{\beta}. \\ &= p(\boldsymbol{\Theta}|\sigma_\theta^2)p(\boldsymbol{\vartheta}|\sigma_\vartheta^2)p(\omega^2, \sigma_\theta^2, \sigma_\vartheta^2|\boldsymbol{\Delta}) \end{aligned}$$

$$
\begin{aligned}
&\times\ (2\pi)^{-n(J_1+J_2+J_3)\tau/2}(\omega^2)^{-n(J_1+J_2+J_3)\tau/2}(2\pi)^{-K_\mu^3/2}(\sigma_\beta^2)^{-K_\mu^3/2} \\
&\times\ \int_{\boldsymbol{\beta}} \exp\{\frac{-1}{2\omega^2}[(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}-\mathbf{Q}\boldsymbol{\Theta}-\mathbf{R}\boldsymbol{\vartheta})'(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}-\mathbf{Q}\boldsymbol{\Theta}-\mathbf{R}\boldsymbol{\vartheta}) \\
&+\ \boldsymbol{\beta}'(\omega^2/\sigma_\beta^2 I)\boldsymbol{\beta}]\}d\boldsymbol{\beta} \\
&=\ p(\boldsymbol{\Theta}|\sigma_\theta^2)p(\boldsymbol{\vartheta}|\sigma_\vartheta^2)p(\omega^2,\sigma_\theta^2,\sigma_\vartheta^2|\boldsymbol{\Delta}) \\
&\times\ (2\pi)^{-n(J_1+J_2+J_3)\tau/2}(\omega^2)^{-n(J_1+J_2+J_3)\tau/2}(2\pi)^{-K_\mu^3/2}(\sigma_\beta^2)^{-K_\mu^3/2} \\
&\times\ \exp\{\frac{-1}{2\omega^2}\mathbf{W}_1'\mathbf{W}_1\} \\
&\times\ \int_{\boldsymbol{\beta}} \exp\{\frac{-1}{2\omega^2}[-\mathbf{W}_1'\mathbf{X}\boldsymbol{\beta}-\boldsymbol{\beta}'\mathbf{X}'\mathbf{W}_1\boldsymbol{\beta}'\mathbf{X}(\mathbf{X}'\mathbf{X}+\omega^2/\sigma_\beta^2\mathbf{I})\boldsymbol{\beta}]\}d\boldsymbol{\beta},
\end{aligned} \tag{5.6}
$$

em que $\mathbf{W}_1 = \mathbf{y}-\mathbf{Q}\boldsymbol{\Theta}-\mathbf{R}\boldsymbol{\vartheta}$. Completando o quadrado no termo exponencial contido na integração obtemos que:

$$
\begin{aligned}
p(\boldsymbol{\Theta},\boldsymbol{\vartheta},\omega^2,\sigma_\theta^2,\sigma_\vartheta^2|\mathbf{y},\boldsymbol{\Delta}) &\propto\ p(\boldsymbol{\Theta}|\sigma_\theta^2)p(\boldsymbol{\vartheta}|\sigma_\vartheta^2)p(\omega^2,\sigma_\theta^2,\sigma_\vartheta^2|\boldsymbol{\Delta}) \\
&\times\ (2\pi)^{-n(J_1+J_2+J_3)\tau/2}(\omega^2)^{-n(J_1+J_2+J_3)\tau/2}(2\pi)^{-K_\mu^3/2}(\sigma_\beta^2)^{-K_\mu^3/2} \\
&\times\ \exp\{\frac{-1}{2\omega^2}[\mathbf{W}_1'\mathbf{W}_1-\mathbf{W}_1'\mathbf{X}(\mathbf{X}'\mathbf{X}+\omega^2/\sigma_\beta^2\mathbf{I})^{-1}\mathbf{X}'\mathbf{W}_1]\} \\
&\times\ \int_{\boldsymbol{\beta}} \exp\{\frac{-1}{2\omega^2}[\boldsymbol{\beta}-(\mathbf{X}'\mathbf{X}+\omega^2/\sigma_\beta^2\mathbf{I})^{-1}\mathbf{X}'\mathbf{W}_1]'(\mathbf{X}'\mathbf{X}+\omega^2/\sigma_\beta^2\mathbf{I}) \\
&\times\ [\boldsymbol{\beta}-(\mathbf{X}'\mathbf{X}+\omega^2/\sigma_\beta^2\mathbf{I})^{-1}\mathbf{X}'\mathbf{W}_1]\}d\boldsymbol{\beta} \\
&=\ p(\boldsymbol{\Theta}|\sigma_\theta^2)p(\boldsymbol{\vartheta}|\sigma_\vartheta^2)p(\omega^2,\sigma_\theta^2,\sigma_\vartheta^2|\boldsymbol{\Delta}) \\
&\times\ (2\pi)^{-n(J_1+J_2+J_3)\tau/2}(\omega^2)^{-(n(J_1+J_2+J_3)\tau-K_\mu^3)/2}(\sigma_\beta^2)^{-K_\mu^3/2} \\
&\times\ |\mathbf{X}'\mathbf{X}+\omega^2/\sigma_\beta^2\mathbf{I}|^{-1/2} \\
&\times\ \exp\{\frac{-1}{2\omega^2}(\mathbf{y}-\mathbf{Q}\boldsymbol{\Theta}-\mathbf{R}\boldsymbol{\vartheta})'[\mathbf{I}-\mathbf{X}(\mathbf{X}'\mathbf{X}+\omega^2/\sigma_\beta^2\mathbf{I})^{-1}\mathbf{X}'] \\
&\times\ (\mathbf{y}-\mathbf{Q}\boldsymbol{\Theta}-\mathbf{R}\boldsymbol{\vartheta})\} \\
&=\ p(\boldsymbol{\Theta}|\sigma_\theta^2)p(\boldsymbol{\vartheta}|\sigma_\vartheta^2)p(\omega^2,\sigma_\theta^2,\sigma_\vartheta^2|\boldsymbol{\Delta}) \\
&\times\ (2\pi)^{-n(J_1+J_2+J_3)\tau/2}(\omega^2)^{-(n(J_1+J_2+J_3)\tau-K_\mu^3)/2}(\sigma_\beta^2)^{-K_\mu^3/2} \\
&\times\ |\mathbf{X}'\mathbf{X}+\omega^2/\sigma_\beta^2\mathbf{I}|^{-1/2} \\
&\times\ \exp\{\frac{-1}{2\omega^2}(\mathbf{y}-\mathbf{Q}\boldsymbol{\Theta}-\mathbf{R}\boldsymbol{\vartheta})'\mathbf{G}_1(\mathbf{y}-\mathbf{Q}\boldsymbol{\Theta}-\mathbf{R}\boldsymbol{\vartheta})\},
\end{aligned} \tag{5.7}
$$

em que $\mathbf{G_1} := \mathbf{I}-\mathbf{X}(\mathbf{X}'\mathbf{X}+\omega^2/\sigma_\beta^2\mathbf{I})^{-1}\mathbf{X}'$. Tomando $\mathbf{W_2} = \mathbf{y}-\mathbf{R}\boldsymbol{\vartheta}$ e integrando (5.7) em $\boldsymbol{\Theta}$ temos que:

$$\begin{aligned}
p(\boldsymbol{\vartheta}, \omega^2, \sigma^2_\theta, \sigma^2_\vartheta | \mathbf{y}, \boldsymbol{\Delta}) &\propto p(\boldsymbol{\vartheta}|\sigma^2_\vartheta)p(\omega^2, \sigma^2_\theta, \sigma^2_\vartheta|\boldsymbol{\Delta}) \\
&\times (2\pi)^{-n(J_1+J_2+J_3)\tau/2}(\omega^2)^{-(n(J_1+J_2+J_3)\tau-K^3_\mu)/2}(\sigma^2_\beta)^{-K^3_\mu/2}(2\pi)^{-K^2_\zeta/2} \\
&\times (\sigma^2_\theta)^{-K^2_\zeta/2}|\mathbf{X}'\mathbf{X} + \omega^2/\sigma^2_\beta\mathbf{I}|^{-1/2} \\
&\times \int_{\boldsymbol{\Theta}} \exp\{\frac{-1}{2\omega^2}[(\mathbf{W_2} - \mathbf{Q}\boldsymbol{\Theta})'\mathbf{G_1}(\mathbf{W_2} - \mathbf{Q}\boldsymbol{\Theta}) + \boldsymbol{\Theta}'(\omega^2/\sigma^2_\theta\mathbf{I})\boldsymbol{\Theta}]\}d\boldsymbol{\Theta} \\
&= \frac{1}{p(\mathbf{y})}p(\boldsymbol{\vartheta}|\sigma^2_\vartheta)p(\omega^2, \sigma^2_\theta, \sigma^2_\vartheta|\boldsymbol{\Delta}) \\
&\times (2\pi)^{-n(J_1+J_2+J_3)\tau/2}(\omega^2)^{-(n(J_1+J_2+J_3)\tau-K^3_\mu-K^2_\zeta)/2}(\sigma^2_\beta)^{-K^3_\mu/2}(\sigma^2_\theta)^{-K^2_\zeta/2} \\
&\times |\mathbf{X}'\mathbf{X} + \omega^2/\sigma^2_\beta\mathbf{I}|^{-1/2} \\
&\times \exp\{\frac{-1}{2\omega^2}\mathbf{W_2}'[\mathbf{G_1} - \mathbf{G_1}\mathbf{Q}(\mathbf{Q}'\mathbf{G_1}\mathbf{Q} + \omega^2/\sigma^2_\theta\mathbf{I})^{-1}\mathbf{Q}'\mathbf{G_1}]\mathbf{W_2}\} \\
&\times |\mathbf{Q}'\mathbf{G_1}\mathbf{Q} + \omega^2/\sigma^2_\theta\mathbf{I}|^{-1/2}. \qquad (5.8)
\end{aligned}$$

Para efeito de notação considere $\mathbf{G_2} := \mathbf{G_1} - \mathbf{G_1}\mathbf{Q}(\mathbf{Q}'\mathbf{G_1}\mathbf{Q} + \omega^2/\sigma^2_\beta\mathbf{I})^{-1}\mathbf{Q}'\mathbf{G_1}$. Integrando (5.8) em $\boldsymbol{\vartheta}$, temos que:

$$\begin{aligned}
p(\omega^2, \sigma^2_\theta, \sigma^2_\vartheta | \mathbf{y}, \boldsymbol{\Delta}) &\propto p(\omega^2, \sigma^2_\theta, \sigma^2_\vartheta|\boldsymbol{\Delta}) \\
&\times (2\pi)^{-n(J_1+J_2+J_3)\tau/2}(\omega^2)^{-(n(J_1+J_2+J_3)\tau-K^3_\mu-K^2_\zeta)/2}(\sigma^2_\beta)^{-K^3_\mu/2}(\sigma^2_\theta)^{-K^2_\zeta/2} \\
&\times (2\pi)^{-K_\Gamma/2}(\sigma^2_\vartheta)^{-K_\Gamma/2}|\mathbf{X}'\mathbf{X} + \omega^2/\sigma^2_\beta\mathbf{I}|^{-1/2} \\
&\times |\mathbf{Q}'\mathbf{G_1}\mathbf{Q} + \omega^2/\sigma^2_\theta\mathbf{I}|^{-1/2} \\
&\times \int_{\boldsymbol{\vartheta}} \exp\{\frac{-1}{2\omega^2}[(\mathbf{y} - \mathbf{R}\boldsymbol{\vartheta})'\mathbf{G_2}(\mathbf{y} - \mathbf{R}\boldsymbol{\vartheta}) + \boldsymbol{\vartheta}'(\omega^2/\sigma^2_\theta\mathbf{I})\boldsymbol{\vartheta}]\}d\boldsymbol{\vartheta} \\
&= p(\omega^2, \sigma^2_\theta, \sigma^2_\vartheta|\boldsymbol{\Omega}) \\
&\times (2\pi)^{-n(J_1+J_2+J_3)\tau/2}(\omega^2)^{-(n(J_1+J_2+J_3)\tau-K^3_\mu-K^2_\zeta-K_\Gamma)/2}(\sigma^2_\beta)^{-K^3_\mu/2}(\sigma^2_\theta)^{-K^2_\zeta/2} \\
&\times (\sigma^2_\vartheta)^{-K_\Gamma/2}|\mathbf{X}'\mathbf{X} + \omega^2/\sigma^2_\beta\mathbf{I}|^{-1/2} \\
&\times |\mathbf{Q}'\mathbf{G_1}\mathbf{Q} + \omega^2/\sigma^2_\theta\mathbf{I}|^{-1/2} \\
&\times |\mathbf{R}'\mathbf{G_2}\mathbf{R} + \omega^2/\sigma^2_\vartheta\mathbf{I}|^{-1/2} \\
&\times \exp\{\frac{-1}{2\omega^2}\mathbf{y}'[\mathbf{G_2} - \mathbf{G_2}\mathbf{R}(\mathbf{R}'\mathbf{G_2}\mathbf{R} + \omega^2/\sigma^2_\vartheta\mathbf{I})^{-1}\mathbf{R}'\mathbf{G_2}]\mathbf{y}\}. \qquad (5.9)
\end{aligned}$$

Considere agora o modelo apresentado na forma projetada proposta em (5.3). Neste caso, a distribuição conjunta a posteriori de ω^2, σ^2_θ e σ^2_ϑ aumentada em $\boldsymbol{\Theta}$ e $\boldsymbol{\vartheta}$ é descrita por:

$$\begin{aligned}
p^\star(\boldsymbol{\Theta}, \boldsymbol{\vartheta}, \omega^2, \sigma^2_\theta, \sigma^2_\vartheta | (\mathbf{I} - \mathbf{H})\mathbf{y}, \boldsymbol{\Delta}) &\propto p^\star((\mathbf{I} - \mathbf{H})\mathbf{y}|\boldsymbol{\Theta}, \boldsymbol{\vartheta}, \omega^2, \sigma^2_\theta, \sigma^2_\vartheta)p(\boldsymbol{\Theta}|\sigma^2_\theta)p(\boldsymbol{\vartheta}|\sigma^2_\vartheta) \\
&\times p(\omega^2, \sigma^2_\theta, \sigma^2_\vartheta|\boldsymbol{\Delta}), \qquad (5.10)
\end{aligned}$$

em que $p^{\star}((\mathbf{I}-\mathbf{H})\mathbf{y}|\boldsymbol{\Theta},\boldsymbol{\vartheta},\omega^2,\sigma^2_{\theta},\sigma^2_{\vartheta})$ representa a densidade de uma distribuição normal degenerada com média $\mathbf{X}\boldsymbol{\beta}+\mathbf{Q}\boldsymbol{\Theta}+\mathbf{R}\boldsymbol{\vartheta}$ e matriz de covariância semi-positiva definida $\omega^2(\mathbf{I}-\mathbf{H})$ (SRIVASTAVA; ROSEN, 2002). De forma análoga ao caso usual, conseguimos demonstrar que:

$$\begin{aligned} p^{\star}(\omega^2,\sigma^2_{\theta},\sigma^2_{\vartheta}|\mathbf{y}^{\star},\boldsymbol{\Delta}) &\propto p(\omega^2,\sigma^2_{\theta},\sigma^2_{\vartheta}|\boldsymbol{\Delta}) \\ &\times (2\pi)^{-(n(J_1+J_2+J_3)\tau-K^3_{\mu})/2}(\omega^2)^{-(n(J_1+J_2+J_3)\tau-K^3_{\mu}-K^2_{\zeta}-K_{\Gamma}/2)/2}|(\sigma^2_{\theta})^{-K^2_{\zeta}/2} \\ &\times (\sigma^2_{\vartheta})^{-K_{\Gamma}/2}|\boldsymbol{\Lambda}|^{-1/2}|\mathbf{Q}^{'}(\mathbf{I}-\mathbf{H})\mathbf{Q}+\omega^2/\sigma^2_{\theta}\mathbf{I}|^{-1/2} \\ &\times |\mathbf{R}^{'}\mathbf{G_2}^{\star}\mathbf{R}+\omega^2/\sigma^2_{\vartheta}\mathbf{I}|^{-1/2} \\ &\times \exp\{\frac{-1}{2\omega^2}\mathbf{y}^{'}[\mathbf{G_2}^{\star}-\mathbf{G_2}^{\star}\mathbf{R}(\mathbf{R}^{'}\mathbf{G_2}^{\star}\mathbf{R}+\omega^2/\sigma^2_{\vartheta}\mathbf{I})^{-1}\mathbf{R}^{'}\mathbf{G_2}^{\star}]\mathbf{y}\}, \quad (5.11) \end{aligned}$$

em que $\mathbf{G_2}^{\star} := (\mathbf{I}-\mathbf{H})-(\mathbf{I}-\mathbf{H})\mathbf{Q}(\mathbf{Q}^{'}(\mathbf{I}-\mathbf{H})\mathbf{Q}+\omega^2/\sigma^2_{\theta}\mathbf{I})^{-1}\mathbf{Q}^{'}(\mathbf{I}-\mathbf{H})$ e $|\boldsymbol{\Lambda}|$ é o produto dos autovalores não nulos de $(\mathbf{I}-\mathbf{H})$. Logo, a partir dos resultados obtidos, temos que:

$$\begin{aligned} \frac{p(\omega^2,\sigma^2_{\theta},\sigma^2_{\vartheta}|\mathbf{y},\boldsymbol{\Delta})}{p^{\star}(\omega^2,\sigma^2_{\theta},\sigma^2_{\vartheta}|\mathbf{y}^{\star},\boldsymbol{\Delta})} &\propto \frac{(\sigma^2_{\beta})^{-K^3_{\mu}/2}}{(2\pi)^{K^3_{\mu}/2}|\boldsymbol{\Lambda}|^{-1/2}} \\ &\times |\mathbf{X}^{'}\mathbf{X}+\omega^2/\sigma^2_{\beta}\mathbf{I}|^{-1/2} \\ &\times \frac{|\mathbf{Q}^{'}\mathbf{G_1}\mathbf{Q}+\omega^2/\sigma^2_{\theta}\mathbf{I}|^{-1/2}}{|\mathbf{Q}^{'}(\mathbf{I}-\mathbf{H})\mathbf{Q}+\omega^2/\sigma^2_{\theta}\mathbf{I}|^{-1/2}} \\ &\times \frac{|\mathbf{R}^{'}\mathbf{G_2}\mathbf{R}+\omega^2/\sigma^2_{\vartheta}\mathbf{I}|^{-1/2}}{|\mathbf{R}^{'}\mathbf{G_2}^{\star}\mathbf{R}+\omega^2/\sigma^2_{\vartheta}\mathbf{I}|^{-1/2}} \\ &\times \frac{\exp\{\frac{-1}{2\omega^2}\mathbf{y}^{'}[\mathbf{G_2}-\mathbf{G_2}\mathbf{R}(\mathbf{R}^{'}\mathbf{G_2}\mathbf{R}+\omega^2/\sigma^2_{\vartheta}\mathbf{I})^{-1}\mathbf{R}^{'}\mathbf{G_2}]\mathbf{y}\}}{\exp\{\frac{-1}{2\omega^2}\mathbf{y}^{'}[\mathbf{G_2}^{\star}-\mathbf{G_2}^{\star}\mathbf{R}(\mathbf{R}^{'}\mathbf{G_2}^{\star}\mathbf{R}+\omega^2/\sigma^2_{\vartheta}\mathbf{I})^{-1}\mathbf{R}^{'}\mathbf{G_2}^{\star}]\mathbf{y}\}}, (5.12) \end{aligned}$$

tal que se $\omega^2/\sigma^2_{\beta}$ for suficientemente próximo de zero, em outras palavras, se adotarmos uma priori suficientemente vaga para o parâmetro $\boldsymbol{\beta}$, teremos que $\mathbf{G_1}\approx\mathbf{I}-\mathbf{H}$ o que por sua vez acarretará em $\mathbf{G_2}\approx\mathbf{G_2}^{\star}$ e, consequentemente, (5.12) será proporcional a uma constante.

Para melhor entendermos como a medida $\omega^2/\sigma^2_{\beta}$ afeta a condição $\mathbf{G_1}\approx\mathbf{I}-\mathbf{H}$, calculamos a norma de Frobenius (BÖTTCHER; WENZEL, 2008) da matriz diferença $\mathbf{D}:=\mathbf{G_1}-(\mathbf{I}-\mathbf{H})$, tomando os valores de $\omega^2/\sigma^2_{\beta}$ no conjunto $\{0.0001, 0.001, 0.01, 0.1\}$. Para tanto, adotamos na construção da matriz $\mathbf{X}$ as configurações $K_{\mu}=5$, $K_{\zeta}=7$, $K_{\Gamma}=6$, $D_{lat}\times D_{long}=[0,1]\times[0,1]$, $I=3$, $J_i=J\in\{2,4,6\}$, $\tau=12$ e $n\in\{100,150\}$. Podemos observar na Tabela 10 que quanto menor for $\omega^2/\sigma^2_{\beta}$, mais próxima a matriz $\mathbf{D}$ estará da matriz nula, segundo a norma de Frobenius. Além disso, foi possível verificar ainda que o aumento no número de localizações (n) e/ou do número de medidas repetidas (J), quando adotadas bases *B-splines* cúbicos, também auxiliam nessa aproximação.

Tabela 10 – Norma de Frobenius da matriz diferença $\mathbf{D} := \mathbf{G_1} - (\mathbf{I} - \mathbf{H})$, tomando os valores de ω^2/σ^2_β no conjunto $\{0.0001, 0.001, 0.01, 0.1\}$, adotando $K_\mu = 5$, $K_\zeta = 7$, $K_\Gamma = 6$, $D_{lat} \times D_{long} = [0,1] \times [0,1]$, $I = 3$, $J_i = J \in \{2,4,6\}$, $\tau = 12$ e $n \in \{100, 150\}$.

		Norma de Frobenius	
	ω^2/σ^2_β	$n = 100$	$n = 150$
J=2	0.0001	0.000947	0.000297
	0.001	0.008599	0.002876
	0.01	0.049196	0.022041
	0.1	0.140388	0.088507
J=4	0.0001	0.000236	0.00007
	0.001	0.002252	0.000733
	0.01	0.015880	0.006302
	0.1	0.053390	0.031021
J=6	0.0001	0.000105	0.00003
	0.001	0.001018	0.000328
	0.01	0.007887	0.002950
	0.1	0.029901	0.016450

5.5.2 Estudo de Simulação Aplicado ao Modelo Projetado Ortogonalmente

A seguir, apresentaremos um estudo de simulação para comparar o comportamento a posteriori dos parâmetros que compõem o modelo proposto em (2.1), estimados tanto para forma projetada (dada em (5.3)) quanto para a forma usual (dada em (2.9)). Para tanto, foram gerados 50 conjuntos de dados independentes e, sem perda de generalidade, utilizamos nesta simulação $K_\mu = 5$, $K_\zeta = 7$, $K_\Gamma = 6$, $D_{lat} \times D_{long} = [0,1] \times [0,1]$, $I = 3$, $J_i = J = 6$, $\tau = 12$ e $n = 150$. Ademais, tomamos ainda $\sigma^2_{\theta_i} = \sigma^2_\theta$, $\sigma^2_{\vartheta_i} = \sigma^2_\vartheta$ e $\mathbf{\Sigma}_{\epsilon_{ij}} = \omega^2\mathbf{I}$. A Tabela 11 apresenta os valores tomados para os parâmetros do modelo.

Tabela 11 – Valores propostos para os parâmetros do modelo proposto em (2.1), utilizados para gerar 50 conjuntos de dados independentes mantendo $K_\mu = 5$, $K_\zeta = 7$, $K_\Gamma = 6$, $D_{lat} \times D_{long} = [0,1] \times [0,1]$, $I = 3$, $J_i = 6$, $\tau = 12$ e $n = 150$, com $\sigma^2_{\theta_i} = \sigma^2_\theta$, $\sigma^2_{\vartheta_i} = \sigma^2_\vartheta$ e $\mathbf{\Sigma}_{\epsilon_{ij}} = \omega^2\mathbf{I}$.

Parâmetros	Valores Propostos
ω^2	5.3
σ^2_θ	3.2
σ^2_ϑ	2.1

Neste cenário de estudo o vetor de parâmetros para o modelo apresentado na forma usual é $\boldsymbol{\xi} = (\boldsymbol{\beta}, \sigma^2_\theta, \sigma^2_\vartheta, \omega^2)$ sendo que, para a forma projetada, a componente $\boldsymbol{\beta}$ passa a ser representada por (5.4). A Tabela 12 apresenta as distribuições a priori utilizadas nas simulações tanto para o modelo na forma usual quanto para o modelo na forma projetada, neste último caso, omitindo-se a componente $\boldsymbol{\beta}$. Ressaltamos ainda que por meio do pacote Coda (PLUMMER et al., 2019) foi possível verificar que um período de aquecimento da cadeia de 100.000 passos e espaçamento de 100 passos são suficientes para garantir a convergência das cadeias e que as amostras geradas sejam independentes. Foram verificados, também, os critérios de convergência descritos em Gelman e Rubin (1992).

Tabela 12 – Distribuições a prioris utilizadas nas simulações realizadas para comparar as estimativas dos parâmetros obtidas tanto para o modelo na forma usual (proposta em (2.9)) quanto para o modelo na forma projetada (proposta em (5.3)), neste último caso, omitindo-se a componente $\boldsymbol{\beta}$.

Parâmetro	Distribuição a Priori
$\boldsymbol{\beta}$	$\mathbf{N(0, 10^3 I)}$
ω^2	Gama Inversa $(0.05, 0.05)$
σ^2_ϑ	Gama Inversa$(0.05, 0.05)$
σ^2_θ	Gama Inversa $(0.05, 0.05)$

Para cada um dos diferentes cenários obtidos por meio da variação do número de localizações (n) e do número de medidas repetidas (J), tomamos a mediana a posteriori ao longo dos 50 conjuntos de dados simulados. A Figura 18 apresenta os resultados obtidos para o modelo na forma projetada, proposta em (5.3). Como podemos observar, para o parâmetro ω^2 o aumento de n e/ou J resulta em melhores estimativas da mediana a posteriori, uma vez que esse aumento favorece a concentração das estimativas nas proximidades do verdadeiro valor do parâmetro. Já para os parâmetros σ^2_θ e σ^2_ϑ notamos que o aumento das componentes n e/ou J, de modo geral, não acarretam melhoras nas estimativas da mediana a posteriori. Esse mesmo comportamento das estimativas da mediana a posteriori pode ser observado para a média e moda a posteriori (veja Apêndice D). Ressaltamos ainda que ao avaliarmos os valores de ω^2/σ^2_β, obtidos a partir do conjunto de amostras a posteriori geradas para ω^2, observamos uma ordem de grandeza máxima de 10^{-3}.

Com relação ao modelo na forma usual, proposto em (2.9), podemos observar, na Figura 17, que o comportamento das estimativas da mediana a posteriori é similar ao descrito pelas estimativas da mediana a posteriori obtidas no modelo apresentado na forma projetada (veja na Figura 18). Essa similaridade pode ser estendida para a média e moda a posteriori (veja Apêndice D).

Nas Tabelas 13 e 14, apresentamos as probabilidades de cobertura para as estimativas dos parâmetros obtidas para o modelo na forma usual e para o modelo na forma projetada, respectivamente. Os resultados indicam que aumentando n e/ou J não implica necessariamente em um aumento nas probabilidades de cobertura. Ressaltamos ainda que as probabilidades de cobertura obtidas para o modelo na forma usual e para o modelo na forma projetada foram próximas e superiores a 88%.

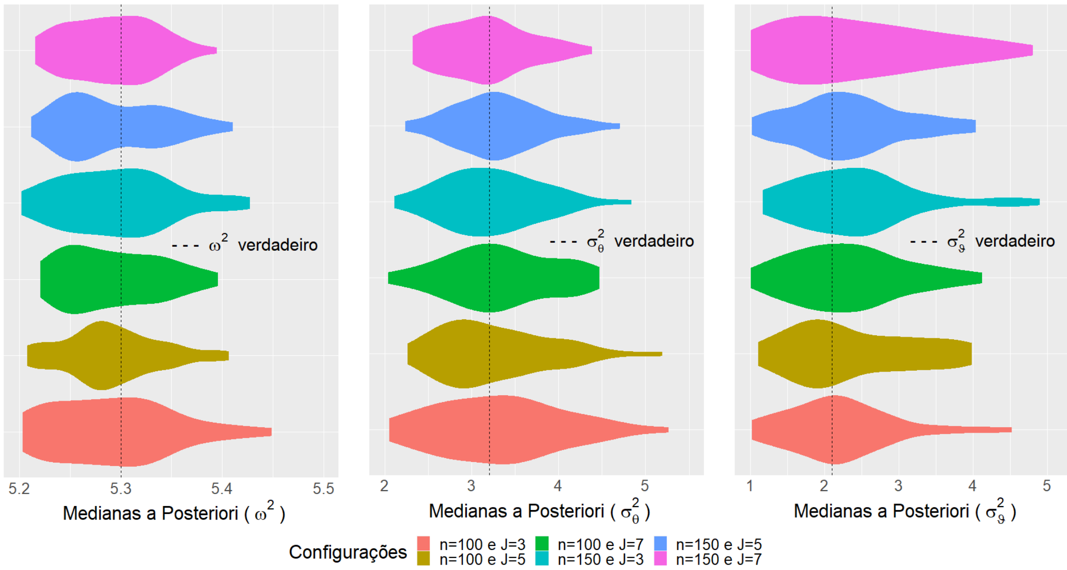

Figura 17 – Medianas a posteriori dos parâmetros ω^2, σ^2_θ e σ^2_ϑ do modelo apresentado na forma usual, obtidas ao longo dos 50 conjuntos de dados simulados independentemente variando o número de localizações (n) e do número de medidas repetidas (J).

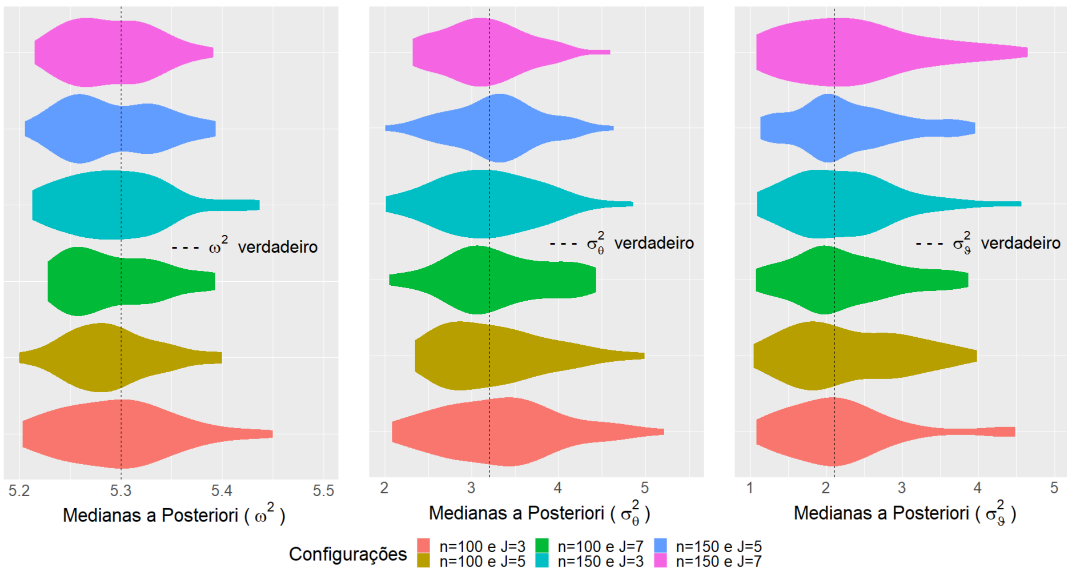

Figura 18 – Medianas a posteriori dos parâmetros ω^2, σ^2_θ e σ^2_ϑ do modelo apresentado na forma projetada, obtidas ao longo dos 50 conjuntos de dados simulados independentemente variando o número de localizações (n) e do número de medidas repetidas (J).

Tabela 13 – Probabilidades de cobertura para os parâmetros do modelo apresentado na forma projetada, obtidas ao longo dos 50 conjuntos de dados gerados independentemente avaliados em cada um dos cenários obtidos por meio da variação dos valores de n e J.

	Parâmetros	Probabilidade de Cobertura $n = 100$	$n = 150$
J=2	ω^2	94%	98%
	σ_θ^2	96%	90%
	σ_ϑ^2	92%	96%
J=4	ω^2	94%	96%
	σ_θ^2	96%	94%
	σ_ϑ^2	98%	94%
J=6	ω^2	96%	98%
	σ_θ^2	92%	94%
	σ_ϑ^2	94%	88%

Tabela 14 – Probabilidades de cobertura para os parâmetros do modelo apresentado na forma usual, obtidas ao longo dos 50 conjuntos de dados gerados independentemente avaliados em cada um dos cenários obtidos por meio da variação dos valores de n e J.

	Parâmetros	Probabilidade de Cobertura $n = 100$	$n = 150$
J=2	ω^2	96%	96%
	σ_θ^2	96%	90%
	σ_ϑ^2	90%	96%
J=4	ω^2	96%	96%
	σ_θ^2	94%	96%
	σ_ϑ^2	98%	92%
J=6	ω^2	96%	96%
	σ_θ^2	90%	92%
	σ_ϑ^2	92%	88%

5.5.3 Capacidade Preditiva do Modelo Projetado Ortogonalmente

Conforme ilustrado na Seção 5.5.2, a técnica de estimação aplicada a modelos projetados em espaços ortogonais (caracterizada em (5.3)) apresentou um comportamento a posteriori das estimativas similar àquele obtido a partir do modelo apresentado na forma usual, descrito em (2.1), quando utilizada a seguinte redução na estrutura paramétrica do modelo: $\sigma_{\theta_i}^2 = \sigma_\theta^2$, $\sigma_{\vartheta_i}^2 = \sigma_\vartheta^2$ e $\boldsymbol{\Sigma}_{\epsilon_{ij}} = \omega^2\mathbf{I}$. Além disso, foram tomados $K_\mu = 5$, $k_\zeta = 7$, $K_\Gamma = 6$, $D_{lat} \times D_{long} = [0,1] \times [0,1]$, $\tau = 12$ e $I = 3$. Assim, apresentaremos nesta seção um estudo preditivo que permitirá uma abordagem comparativa dos resultados encontrados em ambas as perspectivas.

Para esta simulação, foi gerado um conjunto de dados utilizando 151 localizações ($n = 151$) e 8 medidas repetidas ($J = 8$). Em seguida, foi retirada aleatoriamente uma localização teste bem como a última medida repetida gerada ao longo de cada um dos I blocos, reduzindo assim o conjunto de dados treino à dimensão máxima de $n = 150$ e $J = 6$. O objetivo será mensurar o comportamento preditivo do modelo tanto na forma usual quanto na forma projetada, avaliados na localização teste, representada pela localização retirada de forma aleatória, para dados futuros fornecidos pelas últimas medidas repetidas que também foram suprimidas em cada um dos blocos. Finalmente, avaliaremos ainda o desempenho preditivo de tais perspectivas quando variamos a partir de sorteios aleatórios a quantidade de informação, valores de n e J, dentro do processo de treino do modelo.

As Figuras 19 a 21 apresentam os intervalos de predição de 95% para os dados de treino, avaliados ao longo dos diferentes blocos e também a partir da variação dos valores de n e J, quando utilizamos a técnica de estimação aplicada ao modelo proposto na forma usual. Por outro lado, as Figuras 22 a 24 apresentam essa mesma análise porém, utilizando as estimativas obtidas pelo modelo apresentado na forma projetada. Como podemos observar, ambas as propostas geraram intervalos de predição pouco impactados pelo acréscimo de informação dado pela variação dos valores de n e J sendo que, além disso, os intervalos se mostraram praticamente invariantes ao tipo de técnica de estimação utilizada. Portanto, esses resultados sugerem que a técnica de estimação aplicada a modelos projetados em espaços ortogonais é uma boa alternativa ao tratarmos dados de alta dimensão, uma vez que, reduz o número de parâmetros a serem estimados.

Como podemos observar, ainda que empiricamente, os resultados obtidos por meio de soluções em espaços ortogonais são satisfatórios. Infelizmente, não foi possível mensurar teoricamente as diferenças entre as distribuições a posteriori do modelo na forma usual e do modelo na forma projetada, tal que as análises apresentadas neste trabalho foram baseadas no modelo na forma usual.

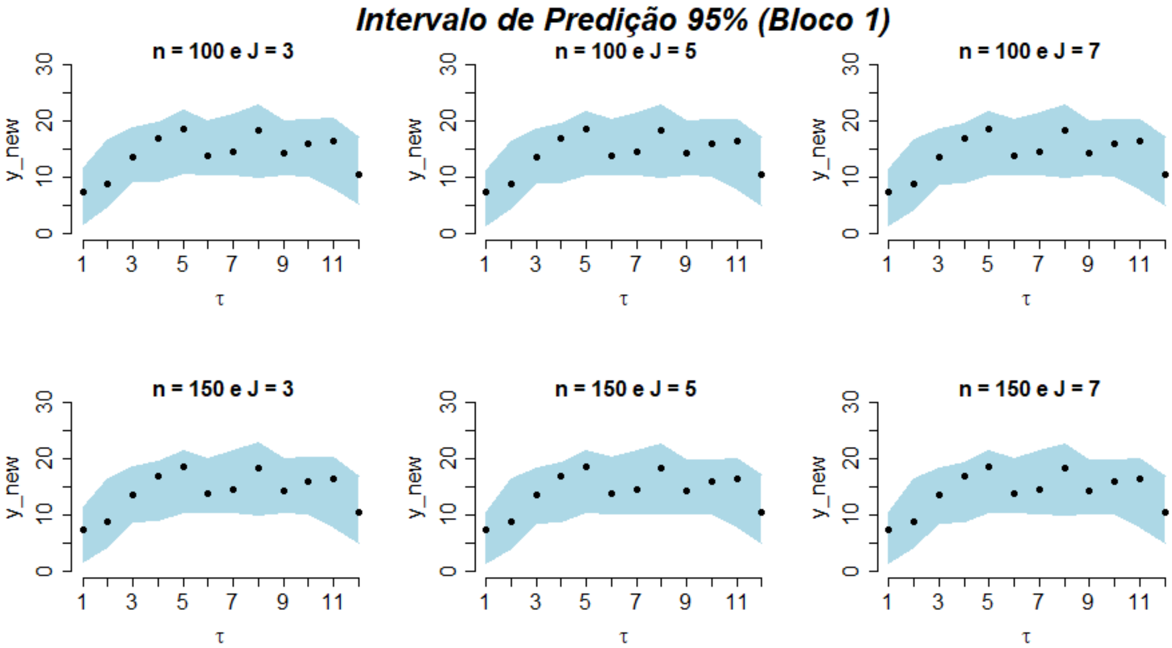

Figura 19 – Intervalos de predição medidos em uma localização teste ao longo do bloco $i = 1$, obtidos a partir das estimativas do modelo na forma usual, caracterizado em (2.1), variando os valores de n e J diante da seguinte redução na estrutura paramétrica: $J_i = J$, $\sigma^2_{\theta_i} = \sigma^2_\theta$, $\sigma^2_{\vartheta_i} = \sigma^2_\vartheta$ e $\mathbf{\Sigma}_{\epsilon_{ij}} = \omega^2\mathbf{I}$.

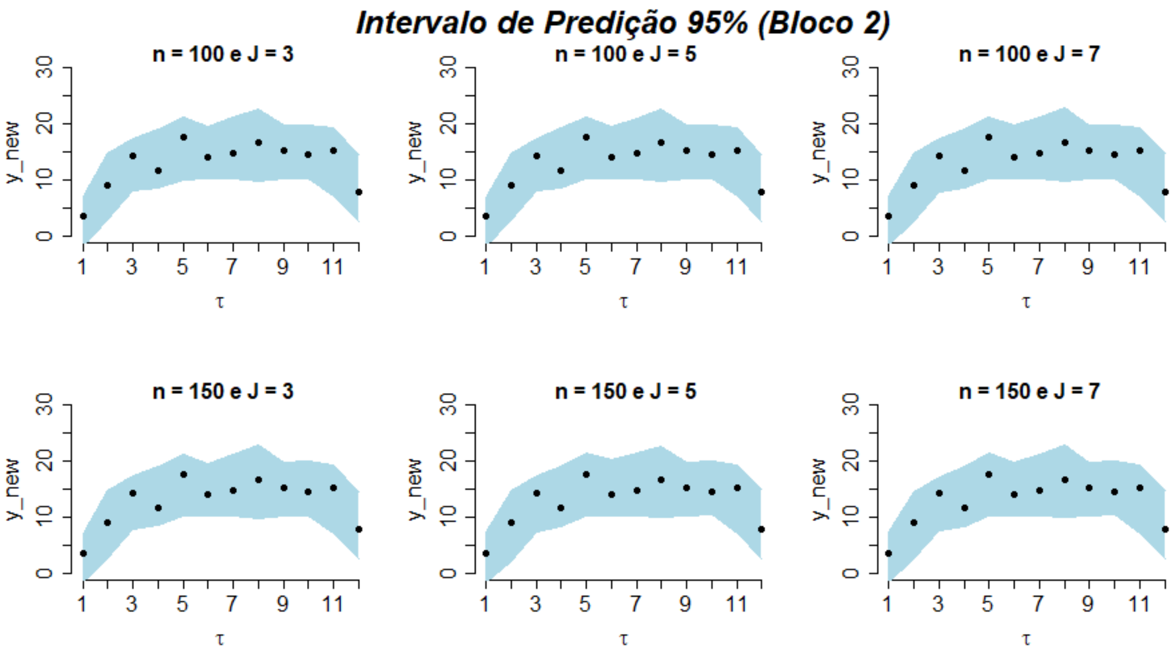

Figura 20 – Intervalos de predição medidos em uma localização teste ao longo do bloco $i = 2$, obtidos a partir das estimativas do modelo na forma usual, caracterizado em (2.1), variando os valores de n e J diante da seguinte redução na estrutura paramétrica: $J_i = J$, $\sigma^2_{\theta_i} = \sigma^2_\theta$, $\sigma^2_{\vartheta_i} = \sigma^2_\vartheta$ e $\mathbf{\Sigma}_{\epsilon_{ij}} = \omega^2\mathbf{I}$.

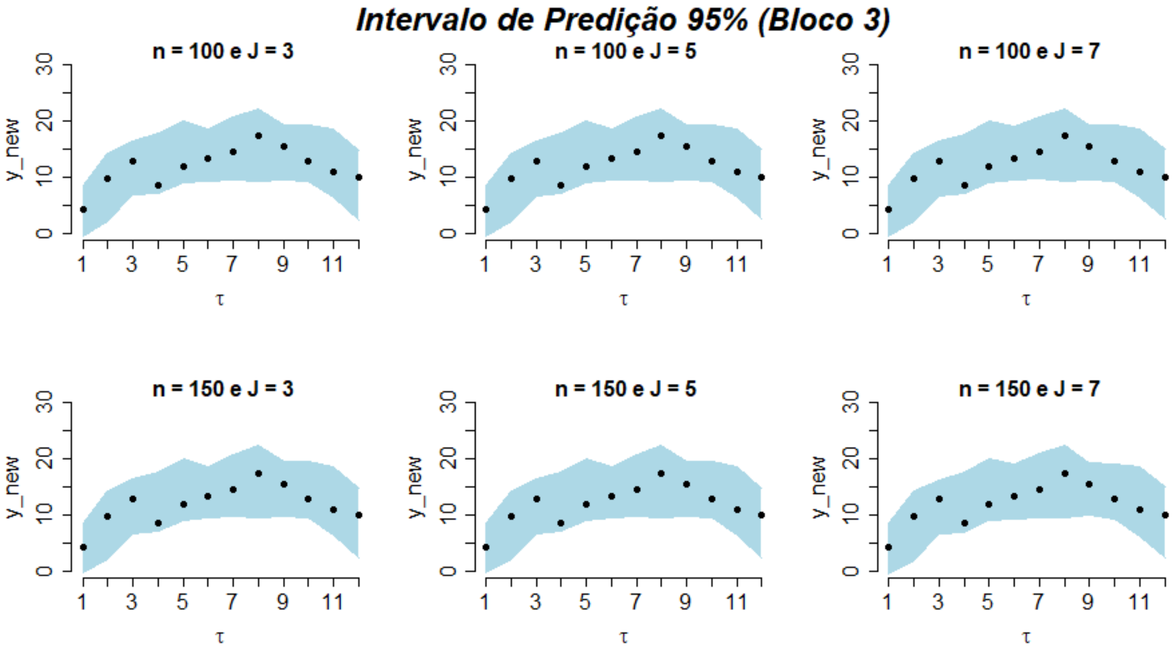

Figura 21 – Intervalos de predição medidos em uma localização teste ao longo do bloco $i = 3$, obtidos a partir das estimativas do modelo na forma usual, caracterizado em (2.1), variando os valores de n e J diante da seguinte redução na estrutura paramétrica: $J_i = J$, $\sigma^2_{\theta_i} = \sigma^2_{\theta}$, $\sigma^2_{\vartheta_i} = \sigma^2_{\vartheta}$ e $\mathbf{\Sigma}_{\epsilon_{ij}} = \omega^2\mathbf{I}$.

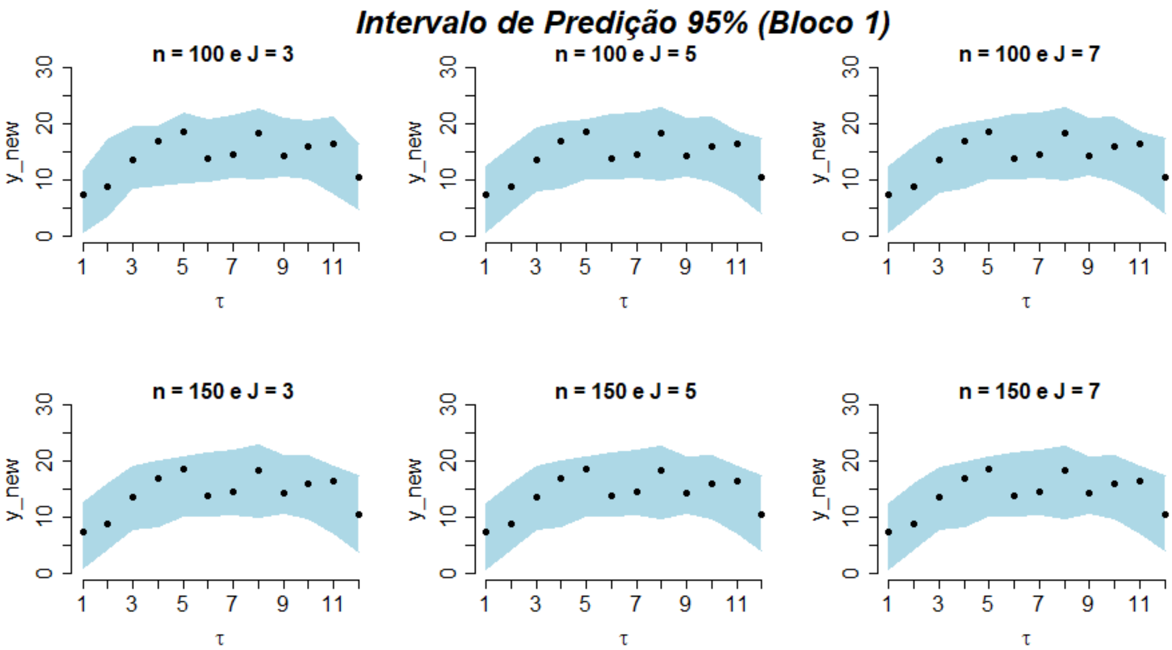

Figura 22 – Intervalos de predição medidos em uma localização teste ao longo do bloco $i = 1$, obtidos a partir das estimativas do modelo na forma projetada, caracterizado em (5.3), variando os valores de n e J diante da seguinte redução na estrutura paramétrica:$J_i = J$, $\sigma^2_{\theta_i} = \sigma^2_{\theta}$, $\sigma^2_{\vartheta_i} = \sigma^2_{\vartheta}$ e $\mathbf{\Sigma}_{\epsilon_{ij}} = \omega^2\mathbf{I}$.

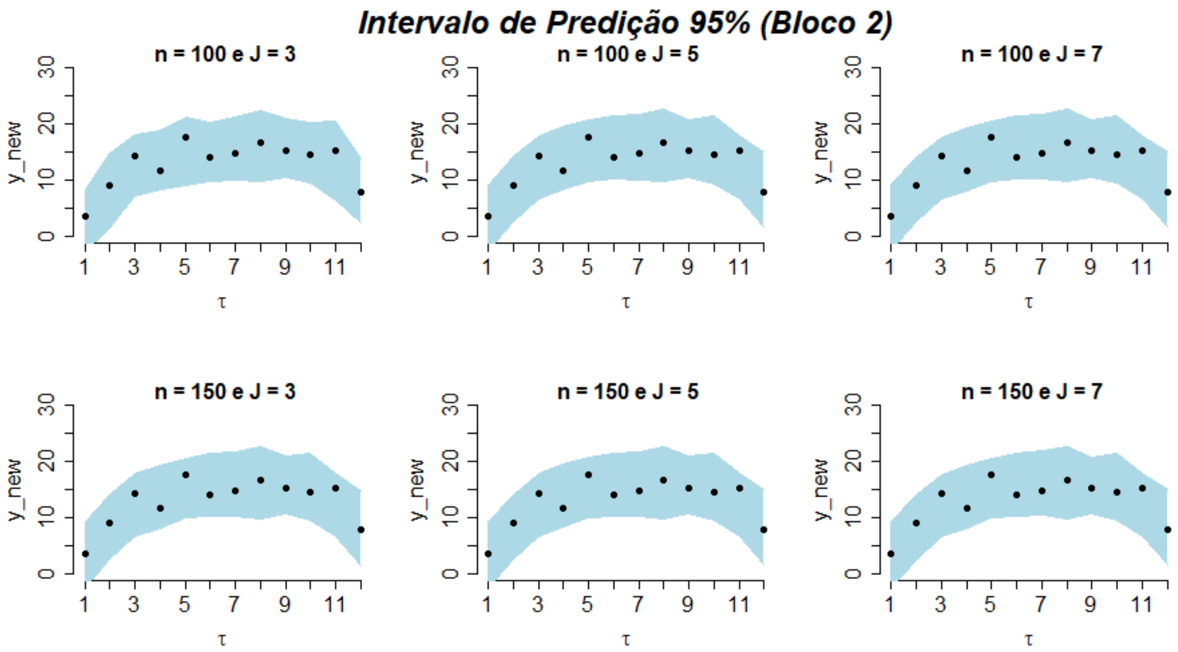

Figura 23 – Intervalos de predição medidos em uma localização teste ao longo do bloco $i = 2$, obtidos a partir das estimativas do modelo na forma projetada, caracterizado em (5.3), variando os valores de n e J diante da seguinte redução na estrutura paramétrica: $J_i = J$, $\sigma^2_{\theta_i} = \sigma^2_{\theta}$, $\sigma^2_{\vartheta_i} = \sigma^2_{\vartheta}$ e $\mathbf{\Sigma}_{\epsilon_{ij}} = \omega^2\mathbf{I}$.

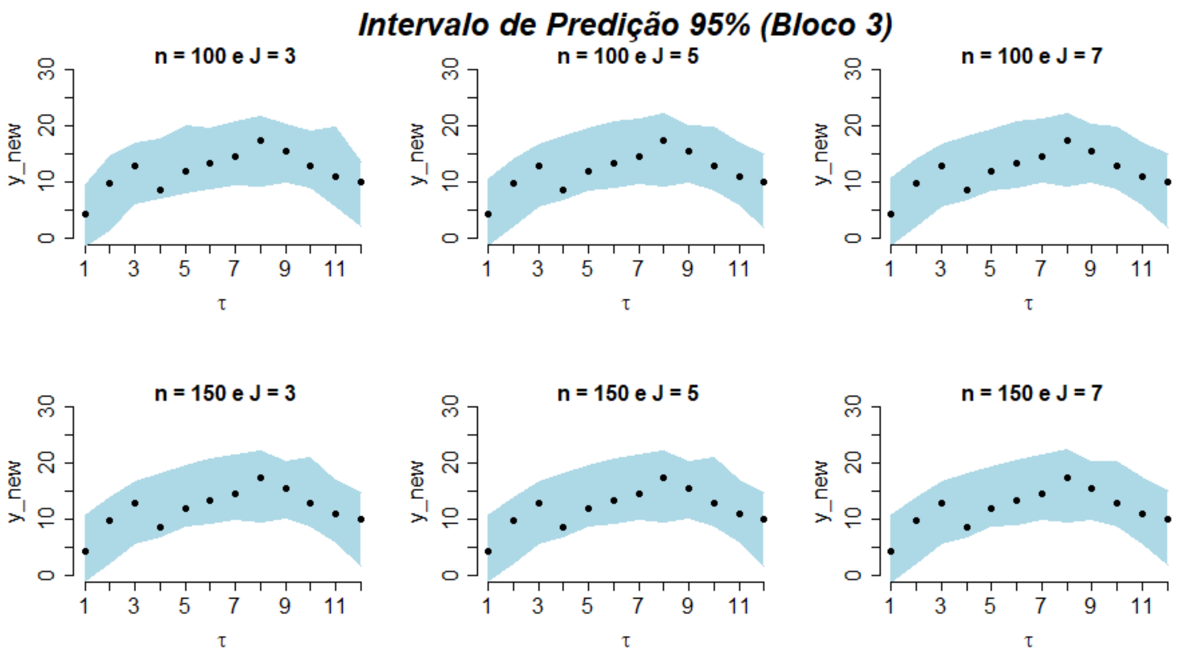

Figura 24 – Intervalos de predição medidos em uma localização teste ao longo do bloco $i = 3$, obtidos a partir das estimativas do modelo na forma projetada, caracterizado em (5.3), variando os valores de n e J diante da seguinte redução na estrutura paramétrica: $J_i = J$, $\sigma^2_{\theta_i} = \sigma^2_{\theta}$, $\sigma^2_{\vartheta_i} = \sigma^2_{\vartheta}$ e $\mathbf{\Sigma}_{\epsilon_{ij}} = \omega^2\mathbf{I}$.

6 Aplicação do Modelo Proposto a Dados de Precipitação no Estado de Goiás

Neste capítulo, apresentamos os resultados do modelo proposto em (2.1) obtidos para os dados de precipitação no Estado de Goiás coletados entre 1980 e 2001.

6.1 Sobre os Dados

As estações meteorológicas que iremos avaliar neste estudo estão localizadas no Estado de Goiás. A região tem uma área superficial de 340.086 km² com latitude e longitude variando na faixa de 19° a 15° e 46° a 51°, respectivamente. O clima predominante na região é a savana tropical, com regime de chuvas apresentando um padrão com apenas duas estações no ano, úmida e seca (ALVARES et al., 2013), sendo que 80% da precipitação anual concentra-se na época úmida, entre os meses de outubro a março.

Nesse estudo, utilizamos dados de séries temporais de precipitação de diferentes fontes: Instituto Brasileiro de Meteorologia - INMET; Agência Nacional de Águas - ANA; Sistema Meteorológico e Hídrico do Estado de Goiás - SIMEHGO e Empresa Brasileira de Pesquisa Agropecuária - EMBRAPA. Foram consideradas, para representar a região de estudo, 87 estações meteorológicas, avaliadas entre os meses de outubro a março, divididas politicamente em 5 mesoregiões. A distribuição geográfica das observações está representada na Figura 25.

Os registros meteorológicos são de 1980 a 2001 (21 anos) e foram classificados de acordo com as condições climáticas observadas nos respectivos períodos de avaliação. As condições dos fenômenos El Niño Oscilação Sul (ENOS) são definidas pelas variações da temperatura da superfície do mar (SST) e sua persistência ao longo do Oceano Pacífico Equatorial (NOAA, 2019). A Administração Nacional Oceânica e Atmosférica (NOAA) define os eventos El Niño e La Niña com base em uma anomalia de temperatura limite de ± 0,5 ºC no Índice Oceânico Niño (ONI), que por sua vez é calculada como a média corrente de três meses de anomalias SST no Pacífico Equatorial Oriental.

No Estado de Goiás, a estação das chuvas ocorre entre outubro e março. Logo, as médias dos valores do ONI foram calculadas com base nas estimativas que fazem referência aos meses de agosto, setembro e outubro (ASO) a março, abril e maio (MAM), conforme descrito em NOAA (2019). Assim, para os fins de nossa análise, consideramos valores inferiores a -0,5 ºC anos predominantemente La Niña, consideramos valores superiores a 0,5 ºC anos, anos predominantemente El Niño e, finalmente, consideramos valores entre

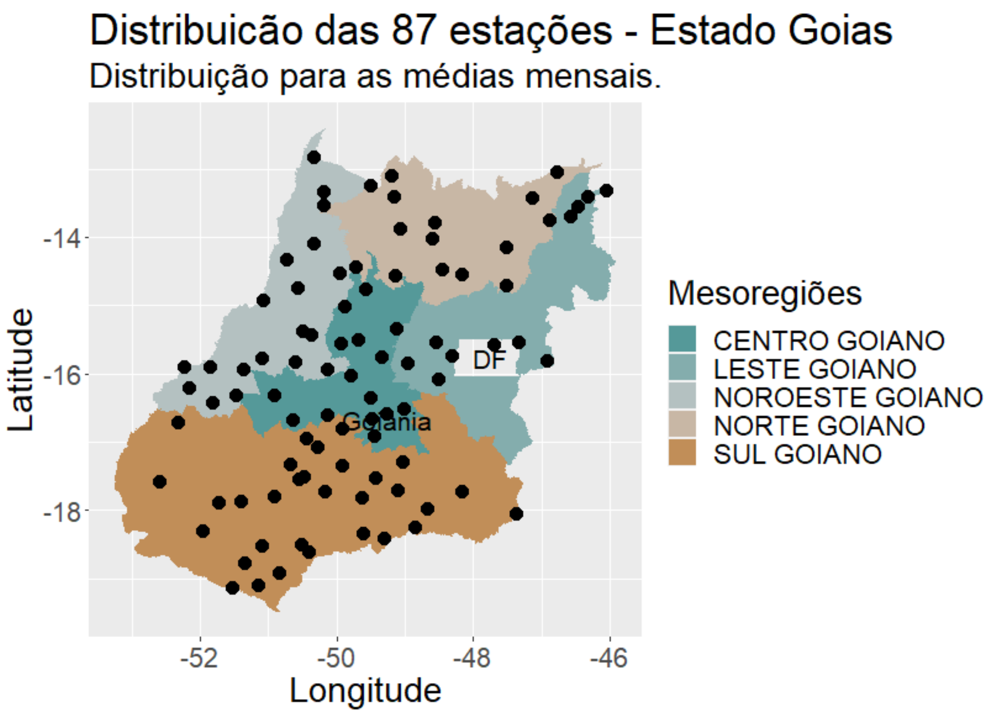

Figura 25 – Distribuição espacial das 87 estações meteorológicas sob estudo no Estado de Goiás.

-0,5 ºC e 0,5 ºC anos Neutros (NOAA, 2019). Na Tabela 15 apresentamos o processo de construção da classificação para o período compreendido entre 1980 a 1983. A classificação completa do conjunto de dados em estudo está apresentada na Tabela 16.

Tabela 15 – Ilustração da classificação dos efeitos climáticos preponderantes para o período compreendido entre 1980 a 1983.

Período	Índice Oceânico Niño (ONI)									
	ASO	SON	OND	NDJ	DJF	JFM	FMA	MAM	Média	Efeito
1980-1981	-0.1	0.0	0.1	0.0	-0.3	-0.5	-0.5	-0.4	-0.21	Neutro
1981-1982	-0.2	-0.1	-0.2	-0.1	0.0	0.1	0.2	0.5	0.02	Neutro
1982-1983	1.6	2.0	2.2	2.2	2.2	1.9	1.5	1.3	1.86	El Niño

Na Figura 26, avaliamos por meio de um boxplot o comportamento da distribuição das distâncias observadas entre as estações meteorológicas avaliadas nos dados reais. Neste caso, é possível observar que 50% das distâncias concentram-se entre 153.1 km e 348.5 km, com valor máximo em 790.9 km. A matriz das distâncias é uma das entradas necessárias à estrutura de correlação *Matérn* apresentada em (2.10).

Na Figura 27, apresentamos, para os diferentes efeitos climáticos, os boxplots funcionais do índice pluviométrico (mm^3) médio mensal avaliados no período de 1980 a 2001, com bandas de profundidade em 95%. Note que a curva mediana para o efeito neutro está disposta acima dos $5mm^3$, com pico no mês de dezembro. Já nos efeitos La Niña e El

Tabela 16 – Classificação dos efeitos climáticos para o período compreendido entre 1980 a 2001.

	Efeito Climático Preponderante		
	La Niña	Neutro	El Niño
Períodos	1984-1985	1980-1991	1982-1983
	1988-1989	1981-1982	1986-1987
	1995-1996	1983-1984	1987-1988
	1998-1999	1985-1986	1991-1992
	1999-2000	1989-1990	1994-1995
	2000-2001	1990-1991	1997-1998
	—	1992-1993	—
	—	1993-1994	—
	—	1996-1997	—

Ninõ as curvas medianas também estão localizadas nas proximidades dos $5mm^3$, com os piores índices presenciados no mês de outubro, porém com picos em janeiro e fevereiro, respectivamente. Assim, como o modelo proposto em (2.1) adota uma média comum a todos os efeitos climáticos, sendo perturbada por meio de efeitos aleatórios espaciais e temporais, temos indícios visuais de que o modelo proposto é compatível com o formato apresentado pelos dados reais conforme apresentado na Figura 27.

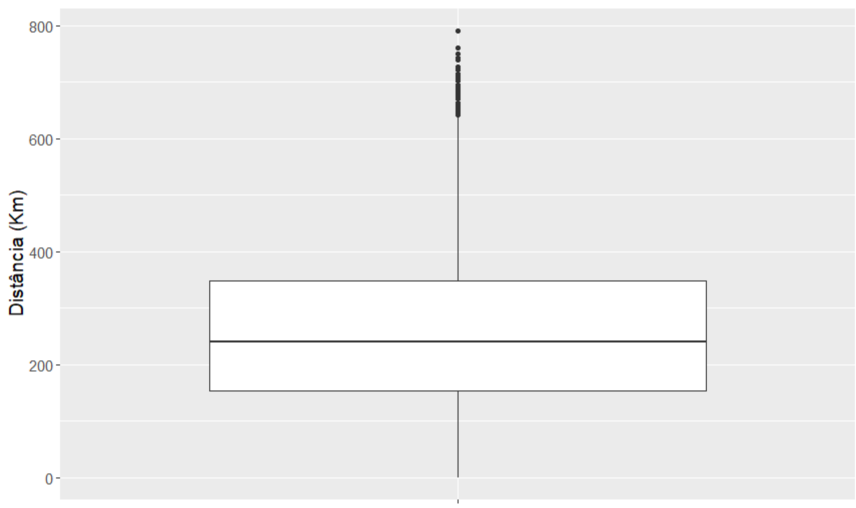

Figura 26 – Boxplot das distâncias observadas entre as estações meteorológicas avaliadas nos dados de precipitação no Estado de Goiás.

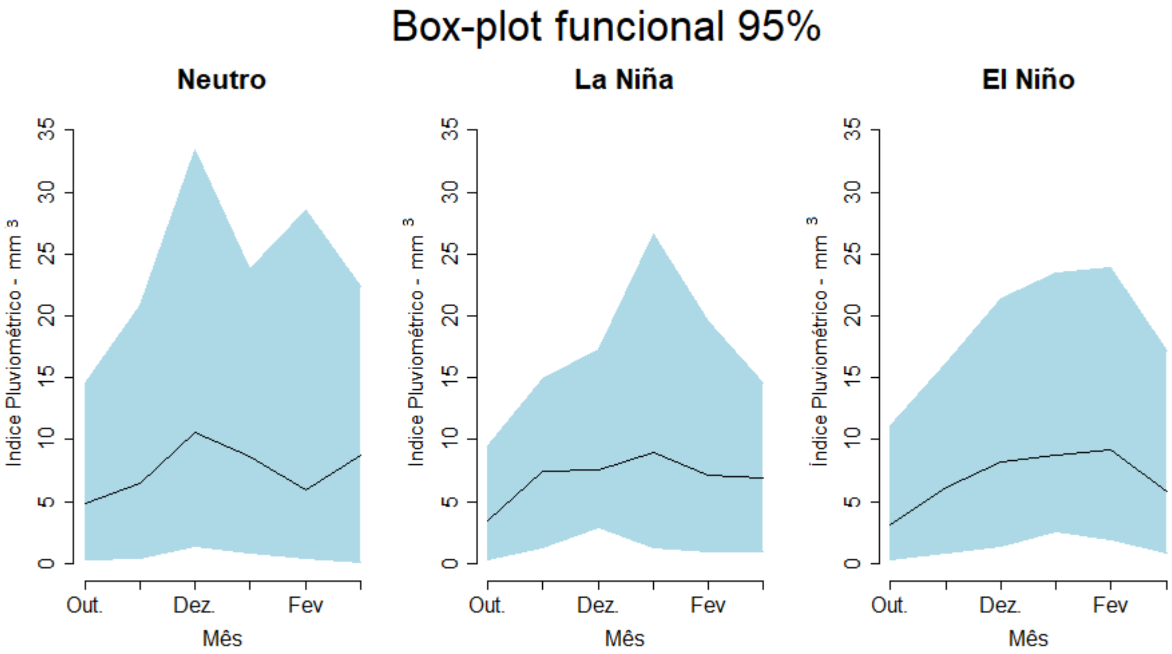

Figura 27 – Boxplot funcional para o índice pluviométrico (mm^3) médio mensal, com bandas de profundidade em 95%, avaliadas segundo efeito climático para o período de 1980 a 2001.

6.2 Estimação dos Parâmetros

Para o ajuste do modelo proposto em (2.1) é necessário fornecer o número de bases utilizadas nas estruturas funcionais (2.2), (2.3) e (2.4), responsáveis pelos efeitos fixo, temporal e espacial, respectivamente. Além disso, é preciso fornecer o valor do parâmetro κ da estrutura de correlação *Matérn* apresentada em (2.10). Assim, iniciamos com um estudo na Seção 6.2.1 para avaliar qual o valor apropriado de κ e do número de bases utilizadas nas expansões dos efeitos fixo , temporal e espacial. Posteriormente, usamos estes valores para ajustar o modelo proposto aos dados de precipitação no Estado de Goiás e as estimativas encontram-se na Seção 6.2.2.

6.2.1 Escolha do Número de Bases e do Parâmetro κ

Para avaliarmos o parâmetro κ da estrutura de correlação *Matérn* em (2.10), tomamos o conjunto $\{0.1, 0.2, ..., 0.5, 1, 1.5, ..., 3.5\}$ contendo os possíveis valores deste parâmetro. O limite superior é 3.5, pois a partir desse ponto, estudos prévios de simulação mostraram que o processo de estimação tende a não identificar bem o valor de κ além de afetar as estimativas do parâmetro ϕ que também compõe esta estrutura de correlação.

De acordo com os resultados do Capítulo 5, tomamos os números de bases consideradas nas expansões dos efeitos funcionais fixo (K_μ), espacial (K_ζ) e temporal

(K_Γ) pertencentes ao conjunto $\{5, 6, 7, 8, 9\}$. As distribuições prioris utilizadas nos estudos realizados nesta seção estão apresentadas na Tabela 6.

Fixadas as condições para este estudo de simulação, utilizaremos o conjunto de dados reais, que conta com 87 localizações ($n = 87$) avaliadas nos meses de outubro a março ($\tau = 6$) entre o período de 1980 a 2001, que resulta em $J_1 = 9$ e $J_2 = J_3 = 6$. O ajuste aos dados reais foi feito para todas as combinações de κ, K_μ, K_ζ e K_Γ, respeitando o conjunto de valores possíveis pré determinados. A Tabela 17 apresenta os cinco ajustes que mais se adequaram aos dados reais, segundo os critérios de seleção de modelos LPML e DIC7.

Tabela 17 – Resultados LPML e DIC7 obtidos nos ajustes do modelo proposto em (2.1) considerando diversos valores de κ e números de bases usadas nas expansões dos efeitos funcionais fixo, espacial e temporal, quando aplicado ao conjunto de dados reais.

Posição	Critério	Estimativas	Número de Bases	κ
1ª	LPML	-23940.0	$K_\mu = 9$, $K_\zeta = 8$ e $K_\Gamma = 9$	0.2
	DIC7	47925.4	$K_\mu = 9$, $K_\zeta = 5$ e $K_\Gamma = 9$	0.2
2ª	LPML	-23940.8	$K_\mu = 9$, $K_\zeta = 5$ e $K_\Gamma = 7$	0.2
	DIC7	47925.5	$K_\mu = 9$, $K_\zeta = 8$ e $K_\Gamma = 7$	0.2
3ª	LPML	-23941.3	$K_\mu = 9$, $K_\zeta = 6$ e $K_l = 9$	0.2
	DIC7	47925.7	$K_\mu = 9$, $K_\zeta = 5$ e $K_\Gamma = 7$	0.2
4ª	LPML	-23942.6	$K_\mu = 9$, $K_\zeta = 5$ e $K_\Gamma = 8$	0.2
	DIC7	47925.9	$K_\mu = 9$, $K_\zeta = 8$ e $K_\Gamma = 8$	0.2
5ª	LPML	-23943.4	$K_\mu = 9$, $K_\zeta = 5$ e $K_\Gamma = 9$	0.2
	DIC7	47926.4	$K_\mu = 9$, $K_\zeta = 9$ e $K_\Gamma = 8$	0.2

Observamos em todos os casos da Tabela 17 que o valor de κ escolhido é de 0.2. Por outro lado, em termos da escolha das possíveis combinações dos números de bases utilizadas nas componentes funcionais do modelo proposto, temos que os critérios LPML e DIC7 se divergem em todos os níveis do ranqueamento sendo, por exemplo, a composição $K_\mu = 9$, $K_\zeta = 8$ e $K_\Gamma = 9$ a mais indicada pelo critério LPML enquanto que para o DIC7 obtivemos $K_\mu = 9$, $K_\zeta = 5$ e $K_\Gamma = 9$. No geral, podemos destacar ainda que a composição $K_\mu = 9$, $K_\zeta = 5$ e $K_\Gamma = 7$, configurou na 2ª e 3ª posições para os critérios LPML e DIC7, respectivamente.

6.2.2 Ajuste do Modelo aos Dados de Precipitação no Estado de Goiás Considerando o Estudo da Seção 6.2.1

Nesta seção, apresentamos os ajustes encontrados durante o processo de estimação dos parâmetros do modelo proposto em (2.1) quando aplicado aos dados de precipitação no Estado de Goiás. O objetivo desse estudo é verificar se após tomadas as configurações iniciais, apresentadas na Seção 6.2.1, as estimativas dos parâmetros de covariância do modelo proposto contemplam as condições de convergência e independência, admitidas durante o procedimento inferencial Bayesiano. Tais averiguações se fazem necessárias, uma vez que, no próximo capítulo utilizaremos tais estimativas a posteriori para gerar intervalos de confiança e bandas de predição para os dados observados.

Para esse estudo foram selecionados três cenários, apresentados na Tabela 18, que se diferenciam pelos números de bases que descrevem os efeitos fixos e aleatórios, pois o valor do parâmetro $\kappa = 0.2$ foi escolhido para todos os cenários, conforme resultados expostos na Tabela 17. Os números de bases utilizados foram selecionados a partir das seguintes condições: a primeira posição apresentada pelo critério LPML, a primeira posição tomada pelo critério DIC7 e, finalmente, a composição que configurou na 2ª e 3ª posições para os critérios LPML e DIC7, respectivamente.

Tabela 18 – Valor do parâmetro κ e do número de bases usadas nas expansões dos efeitos funcionais do modelo em (2.1), aplicados aos dados de precipitação no Estado de Goiás, selecionados para o estudo de convergência e independência admitidas durante o procedimento inferencial Bayesiano.

Cenários	Posição	Critério	Número de Base	Parâmetro κ
1	1ª	LPML	$K_\mu = 9$, $K_\zeta = 8$ e $K_\Gamma = 9$	0.2
2	1ª	DIC7	$K_\mu = 9$, $K_\zeta = 5$ e $K_\Gamma = 9$	0.2
3	2ª	LPML	$K_\mu = 9$, $K_\zeta = 5$ e $K_\Gamma = 7$	0.2
	3ª	DIC7	$K_\mu = 9$, $K_\zeta = 5$ e $K_\Gamma = 7$	0.2

O conjunto das distribuições a priori para o vetor de parâmetros $\xi = (\boldsymbol{\beta}, \sigma^2_{\theta_1}, \sigma^2_{\theta_2}, \sigma^2_{\theta_3}, \sigma^2_{\vartheta_1}, \sigma^2_{\vartheta_2}, \sigma^2_{\vartheta_3}, \omega^2, \phi, \varphi)$ do modelo proposto em (2.1) está disposto na Tabela 6 e vai ao encontro dos resultados observados na Seção 5.1. Foram utilizadas cadeias de 1.5×10^5 passos, com aquecimento de 10^5 passos e amostradas a cada 100 passos. Para o processo de verificação da convergência das cadeias e da independência das amostras geradas, realizado a partir do pacote Coda (PLUMMER et al., 2019), foram construídas duas cadeias de Markov para cada um dos parâmetros do modelo, que foram obtidas a partir dos diferentes valores iniciais descritos na Tabela 19.

A Tabela 20 apresenta os valores da estatística de Gelman e Rubin (veja Gelman e Rubin (1992)) para os parâmetros estimados segundo os cenários descritos na Tabela 18. Valores próximos a 1 representam a convergência das cadeias de Markov. Logo, a convergência acontece para todos os parâmetros estimados. Esta mesma evidência pode ser encontrada nos limites superiores dos intervalos de confiança das estimativas da estatística de Gelman e Rubin dado que os valores se encontram na vizinhança de 1. Para maiores detalhes veja Apêndice E.

Nas Figuras 28 a 30 mostramos a evolução das amostras das cadeias consideradas. Como podemos notar, as trajetórias das duas cadeias sempre se sobrepõem indicando

Tabela 19 – Valores iniciais utilizados para gerar dois grupos de cadeias de Markov para os parâmetros do modelo em (2.1), quando aplicados aos dados de precipitação no Estado de Goiás, usadas no esquema de amostrador de Gibbs.

		Valores Iniciais	
		Cadeia 1	Cadeia 2
Parâmetros	ω^2	0.1	100
	$\sigma^2_{\theta_1}$	0.01	10
	$\sigma^2_{\theta_2}$	0.01	10
	$\sigma^2_{\theta_3}$	0.01	10
	$\sigma^2_{\vartheta_1}$	0.01	10
	$\sigma^2_{\vartheta_2}$	0.01	10
	$\sigma^2_{\vartheta_3}$	0.01	10
	φ	0.1	100
	ϕ	10	1000

Tabela 20 – Estatística de Gelman e Rubin para os parâmetros do modelo proposto em (2.1), quando aplicado aos dados de precipitação no Estado de Goiás, segundo diferentes cenários apresentados na Tabela 18.

		Estatística de Gelman e Rubin		
		Cenário 1	Cenário 2	Cenário 3
Parâmetros	ω^2	1.007	1.018	1.000
	$\sigma^2_{\theta_1}$	1.000	1.008	1.000
	$\sigma^2_{\theta_2}$	1.002	1.002	1.000
	$\sigma^2_{\theta_3}$	1.029	0.999	1.010
	$\sigma^2_{\vartheta_1}$	1.004	0.999	1.000
	$\sigma^2_{\vartheta_2}$	1.012	1.001	1.000
	$\sigma^2_{\vartheta_3}$	0.999	1.001	1.000
	φ	1.003	1.000	1.000
	ϕ	1.005	1.000	1.000
	$\boldsymbol{\xi}$	1.030	1.020	1.010

convergência. Foi verificado ainda que o espaçamento adotado, 150 passos, foi suficiente para garantir correlação zero entre os termos de uma mesma cadeia de Markov (veja Apêndice E).

Finalmente, na Tabela 21 apresentamos a mediana (Med) e o desvio padrão (DP) a posteriori obtidos a partir da mistura das duas cadeias de Markov geradas para cada um dos parâmetros do modelo proposto. Como podemos observar, as componentes de variância dos efeitos aleatórios espaciais e temporais, designados por $\sigma^2_{\theta_i}$ e $\sigma^2_{\vartheta_i}$, $i = 1, 2, 3$, respectivamente, são as mais impactadas por alterações no número de bases utilizadas. Tal característica também é encontrada na média (Me) e moda (Mo) a posteriori, sendo que foi possível observar ainda um grave achatamento das estimativas moda (Mo) a posteriori na vizinhança do zero. Para maiores informações veja o Apêndice E.

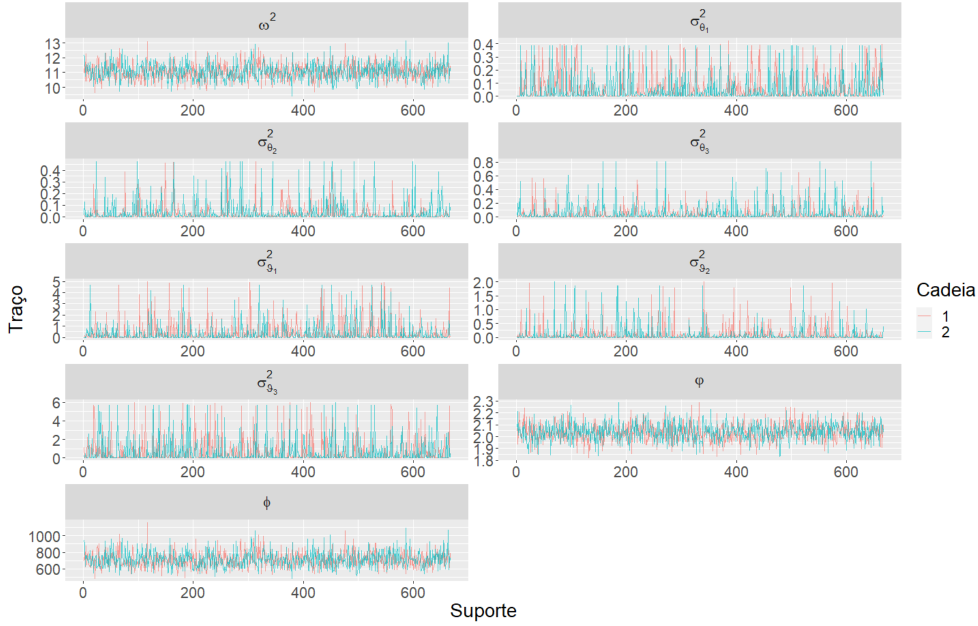

Figura 28 – Trajetória de duas cadeias de Markov, desconsiderando o aquecimento de 10^5 passos, para cada um dos parâmetros que compõem o modelo proposto em (2.1), quando ajustado aos dados de precipitação no Estado de Goiás, segundo Cenário 1.

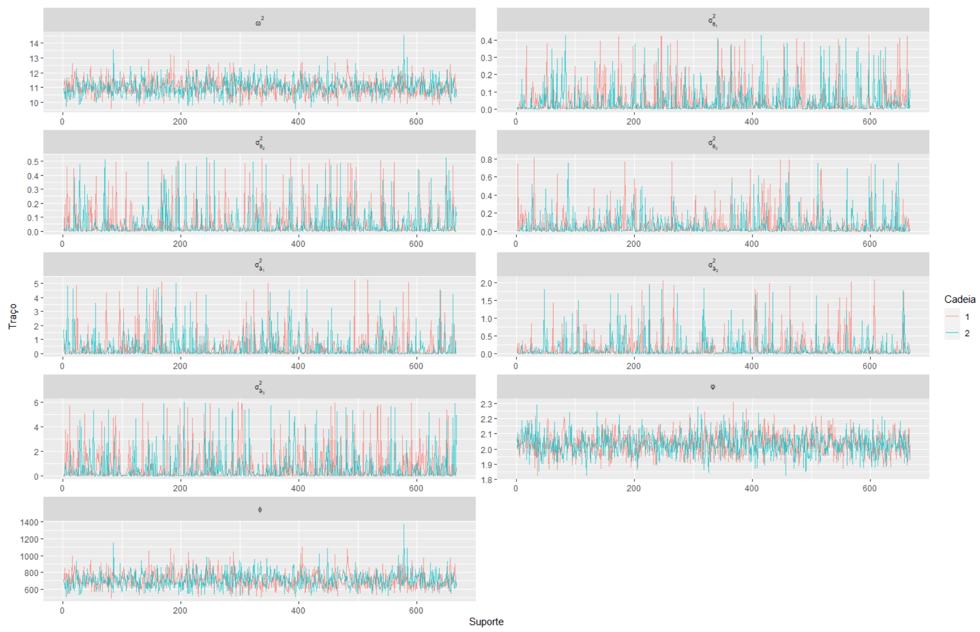

Figura 29 – Trajetória de duas cadeias de Markov, desconsiderando o aquecimento de 10^5 passos, para cada um dos parâmetros que compõem o modelo proposto em (2.1), quando ajustado aos dados de precipitação no Estado de Goiás, segundo Cenário 2.

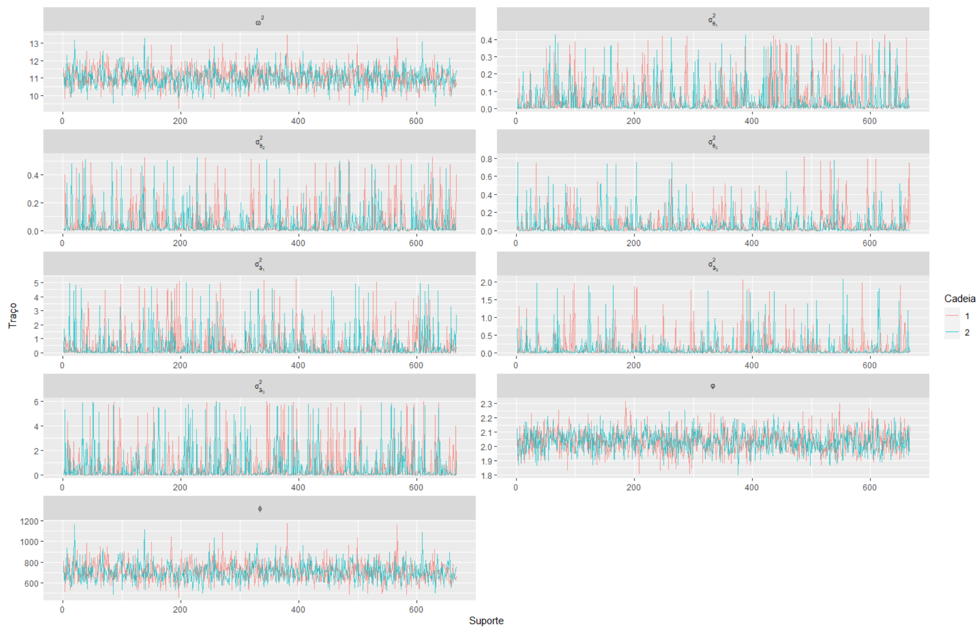

Figura 30 – Trajetória de duas cadeias de Markov, desconsiderando o aquecimento de 10^5 passos, para cada um dos parâmetros que compõem o modelo proposto em (2.1), quando ajustado aos dados de precipitação no Estado de Goiás, segundo Cenário 3.

Tabela 21 – Estimativas a posteriori dos parâmetros que compõem o modelo proposto em (2.1): mediana (Med) e desvio padrão (DP), obtidas a partir de amostras de duas cadeias de Markov geradas para cada um dos parâmetros do modelo proposto, segundo os cenários apresentados na Tabela 18.

		Estatísticas a Posteriori					
		Cenário 1		Cenário 2		Cenário 3	
		Med	DP	Med	DP	Med	DP
Parâmetros	ω^2	11.041	0.577	10.983	0.599	10.989	0.582
	$\sigma^2_{\theta_1}$	0.008	0.100	0.007	0.084	0.006	0.091
	$\sigma^2_{\theta_2}$	0.005	0.081	0.006	0.103	0.007	0.105
	$\sigma^2_{\theta_3}$	0.006	0.107	0.007	0.124	0.006	0.123
	$\sigma^2_{\vartheta_1}$	0.032	0.848	0.025	0.852	0.035	0.992
	$\sigma^2_{\vartheta_2}$	0.007	0.284	0.008	0.313	0.008	0.313
	$\sigma^2_{\vartheta_3}$	0.045	1.302	0.049	1.296	0.037	1.358
	φ	2.033	0.076	2.030	0.077	2.030	0.079
	ϕ	702.879	96.085	696.873	99.219	697.908	95.379

6.3 Resultados

Nesta seção, avaliaremos os resultados fornecidos pelo modelo proposto em (2.1), quando aplicado aos dados de precipitação no Estado de Goiás, a partir da construção de intervalos de credibilidade e de bandas de predição para determinadas estações meteorológicas. Para tanto, selecionamos aleatoriamente e sem reposição cinco estações pertencentes ao conjunto de dados, classificadas aqui como "Treino com predição", que estão distribuídas dentro de cada uma das mesoregiões do Estado de Goiás. Ademais, para cada uma das mesoregiões foram coletados ainda dados adicionais de cinco estações meteorológicas, que serão classificadas aqui como "Teste". A distribuição geográfica das estações, assim como as demais características temporais e espaciais são representadas na Figura 31 e na Tabela 22, respectivamente.

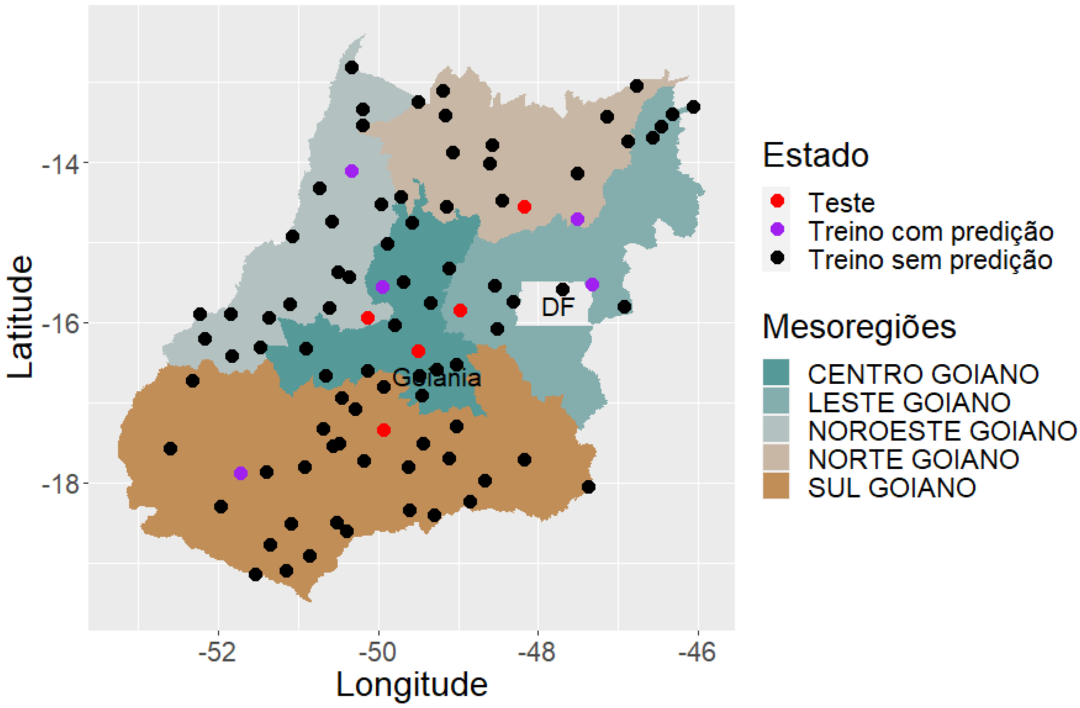

Figura 31 – Distribuição espacial das estações meteorológicas classificadas como "Treino", "Teste com predição" e "Teste sem predição", utilizadas para verificar os resultados fornecidos pelo modelo proposto em (2.1), quando aplicado aos dados de precipitação no Estado de Goiás.

Considerando as estações amostradas para o treino com predição, apresentadas na Tabela 22, construímos intervalos de credibilidade para os períodos 1996-1997, 1997-1998 e 2000-2001, que representam as últimas observações para os efeitos climáticos Neutro, La Niña e EL Niño, respectivamente (veja Tabela 16). É possível observar nas Figuras 32 a 36 que as tendências intrínsecas às diferentes regiões e efeitos climáticos, avaliadas ao longo dos modelos descritos na Tabela 18, são capturadas pelas respectivas bandas de credibilidade. No entanto, em alguns casos podemos destacar até mesmo três pontos externos, porém próximos, aos respectivos intervalos de credibilidade.

Ainda nas Figuras 32 a 36 podemos observar que o comportamento das bandas de credibilidade não é muito afetado entre os cenários ajustados. No entanto, é possível notar uma pequena vantagem para determinados casos específicos. Por exemplo, podemos

Tabela 22 – Município, latitude, longitude e classificação como "Teste" ou "Treino com predição", das estações meteorológicas utilizadas para verificar os resultados fornecidos pelo modelo proposto em (2.1), quando aplicado aos dados de precipitação no Estado de Goiás.

Mesoregião	Estações Avaliadas			
	Munícipio	Latitude	Longitude	Estado
Centro Goiano	Itapuranga	-15.56	-49.94	Treino com predição
	Inhumas	-16.35	-49.5	Teste
Leste Goiano	Formosa	-15.53	-47.33	Treino com predição
	Pirenópolis	-15.85	-48.97	Teste
Noroeste Goiano	Nova Crixas	-14.1	-50.33	Treino com predição
	Goiás	-15.94	-50.14	Teste
Norte Goiano	São João D'Aliança	-14.71	-47.52	Treino com predição
	Niquelândia	-14.55	-48.17	Teste
Sul Goiano	Jataí	-17.88	-51.72	Treino com predição
	Edéia	-17.34	-49.93	Teste

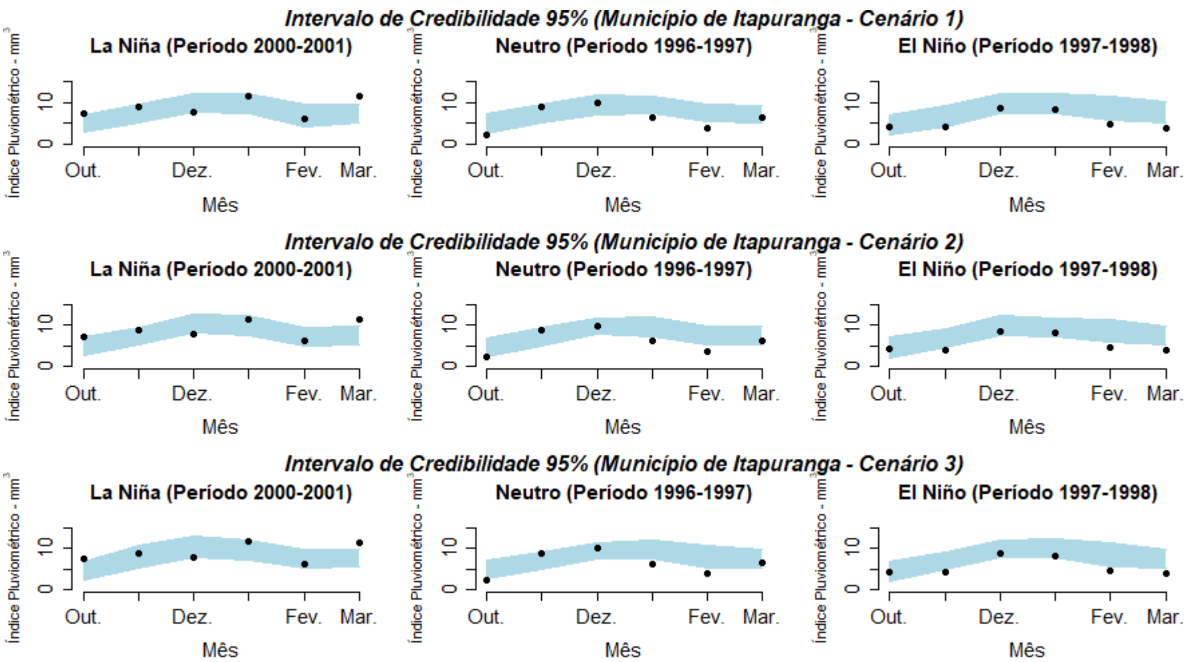

Figura 32 – Intervalos de credibilidade para os índices pluviométricos médios mensais estimados para a estação meteorológica "Treino com predição", localizada no município de Itapuranga-GO, considerando os cenários descritos na Tabela 18 ao longo dos últimos períodos observados em cada um dos efeitos climáticos.

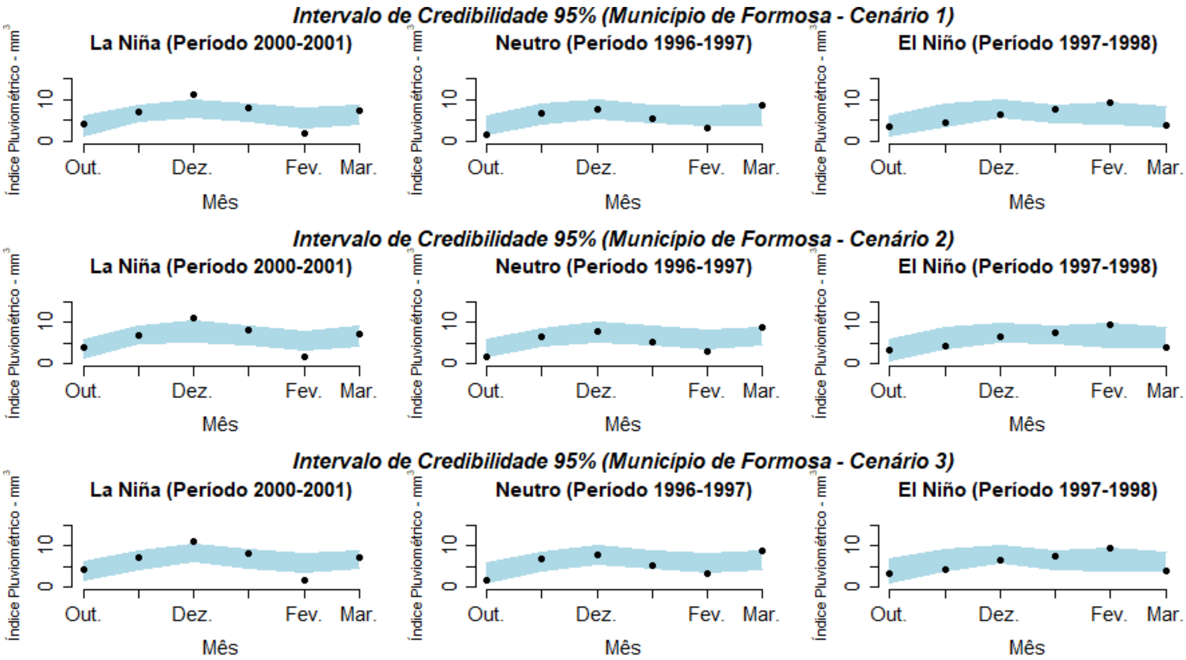

Figura 33 – Intervalos de credibilidade para os índices pluviométricos médios mensais estimados para a estação meteorológica “Treino com predição”, localizada no município de Formosa-GO, considerando os cenários descritos na Tabela 18 ao longo dos últimos períodos observados em cada um dos efeitos climáticos.

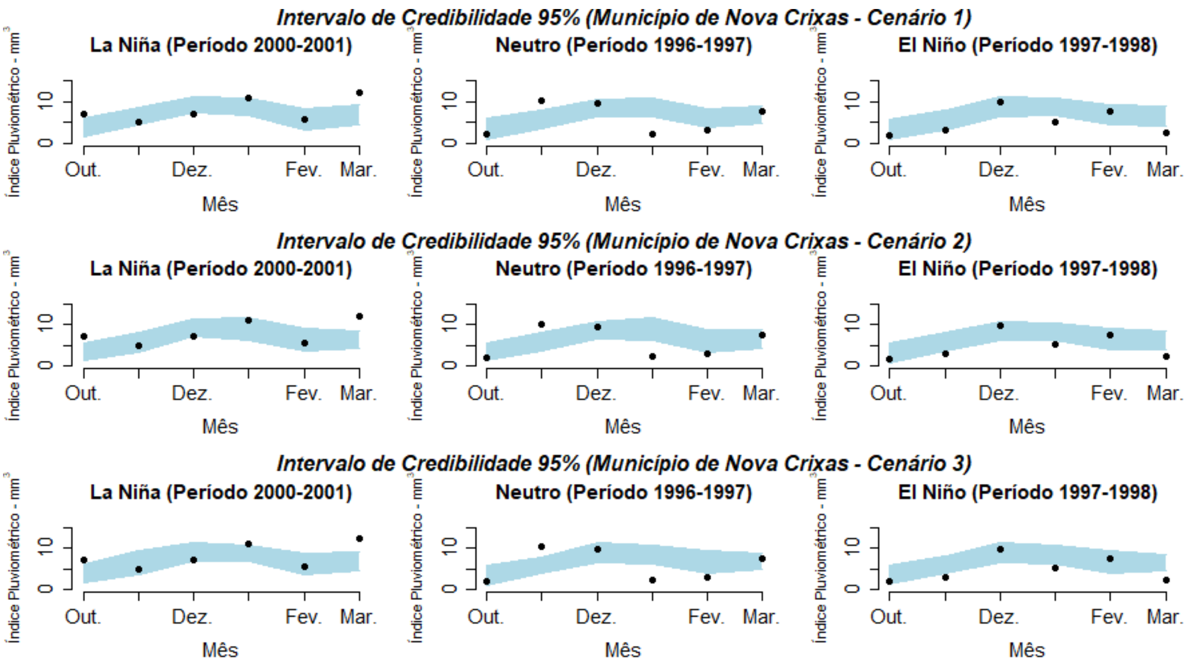

Figura 34 – Intervalos de credibilidade para os índices pluviométricos médios mensais estimados para a estação meteorológica “Treino com predição”, localizada no município de Nova Crixas-GO, considerando os cenários descritos na Tabela 18 ao longo dos últimos períodos observados em cada um dos efeitos climáticos.

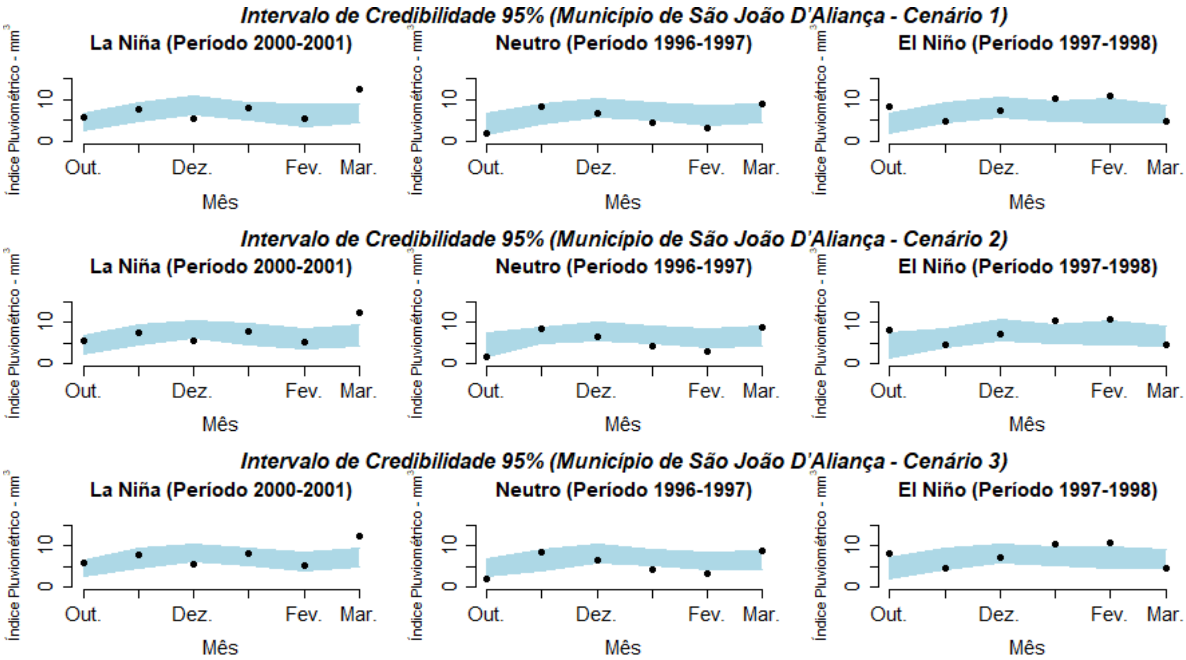

Figura 35 – Intervalos de credibilidade para os índices pluviométricos médios mensais estimados para a estação meteorológica “Treino com predição”, localizada no município de São João D´Aliança-GO, considerando os cenários descritos na Tabela 18 ao longo dos últimos períodos observados em cada um dos efeitos climáticos.

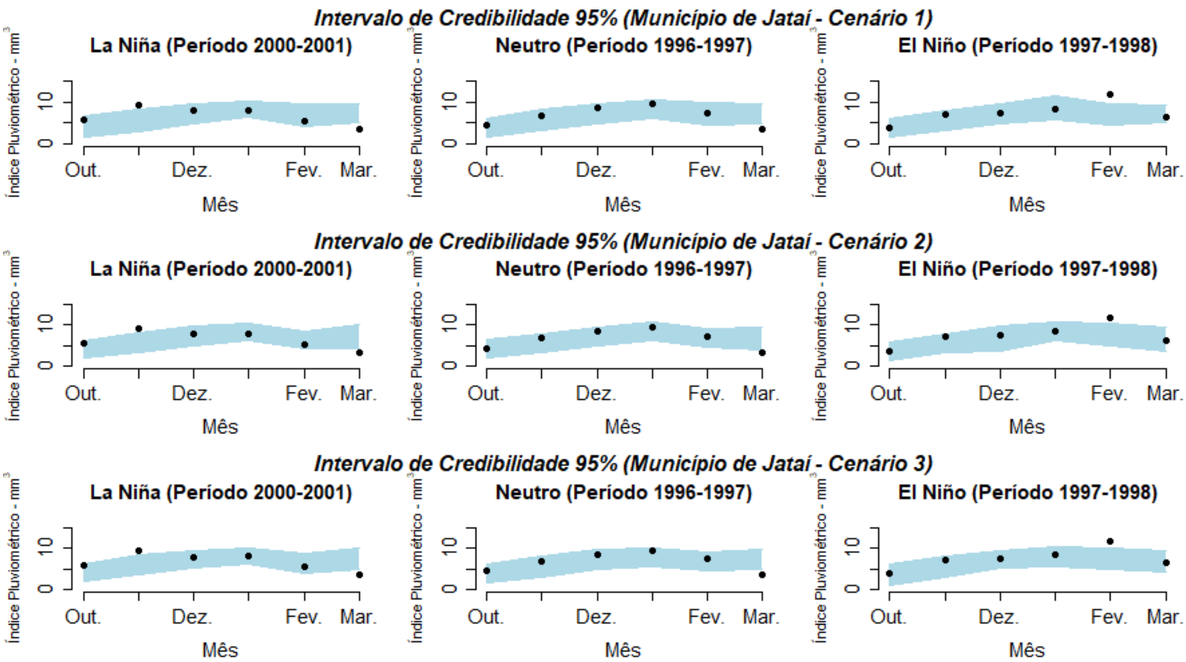

Figura 36 – Intervalos de credibilidade para os índices pluviométricos médios mensais estimados para a estação meteorológica localizada “Treino com predição”, no município de Jataí-GO, considerando os cenários descritos na Tabela 18 ao longo dos últimos períodos observados em cada um dos efeitos climáticos.

observar no efeito El Niño da Figura 34 uma pequena vantagem para o Cenário 1, uma vez que apenas dois pontos estão fora das bandas de credibilidade, meses de janeiro e março, enquanto que os demais cenários apresentaram um terceiro ponto que representa o índice pluviométrico médio observado no mês de novembro. Num outro exemplo, tomamos a estação localizada no município de Jataí, ilustrada na Figura 36, onde o Cenário 2 apresentou melhor desempenho para o efeito Neutro uma vez que todos os índices pluviométricos médios são cobertos pelos respectivos intervalos de credibilidade, o que não aconteceu nos demais cenários.

Na classificação apresentada em NOAA (2019), observamos que para períodos posteriores aos que foram avaliados neste estudo, ou seja, após 2000-2001, os efeitos climáticos Neutro, El Niño e La Niña, ocorreram inicialmente em 2001-2002, 2002-2003 e 2007-2008, respectivamente. Logo, estes serão os períodos utilizados para avaliar a capacidade preditiva para todos os cenários avaliados neste estudo. As Figuras 37 a 41 apresentam as bandas preditivas para as estações classificadas com estado de "Treino com predição", conforme descrito na Tabela 22, avaliados ao longo dos cenários descritos na Tabela 18.

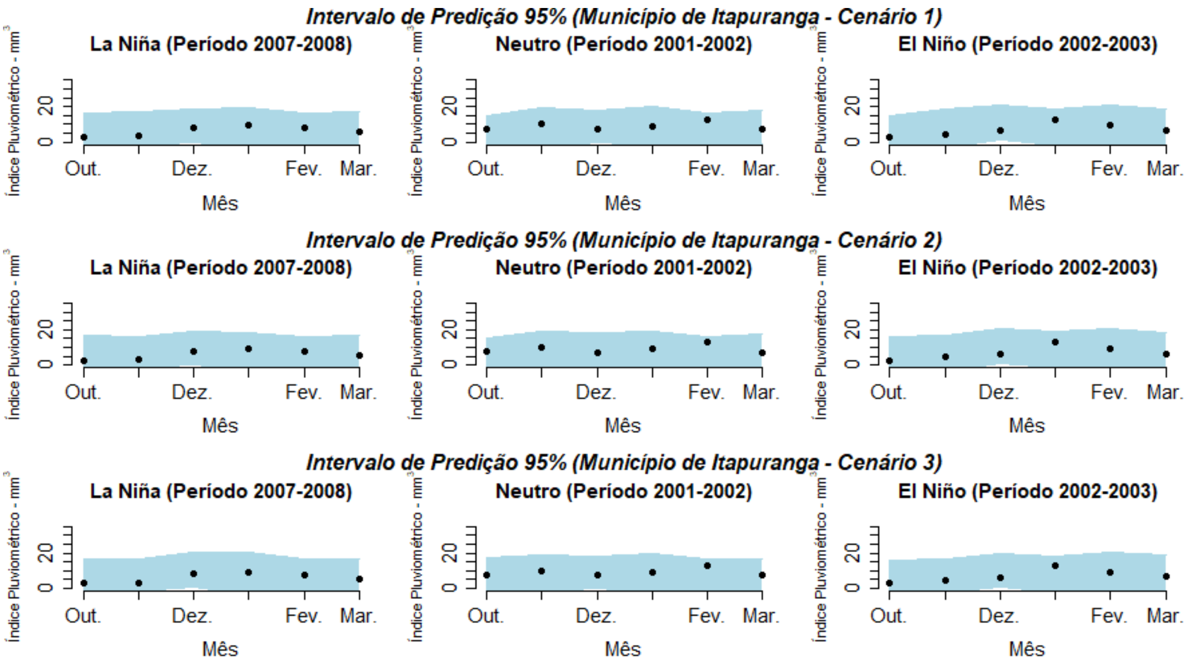

Figura 37 – Intervalos de predição para os índices pluviométricos médios mensais observados para a estação meteorológica "Treino com predição", localizada no município de Itapuranga-GO, considerando os cenários descritos na Tabela 18 ao longo dos efeitos climáticos observados após o período de 2000-2001.

Para todos os cenários ilustrados nas Figuras 37 a 41, é possível notar que os intervalos de predição possuem alta amplitude e que também contemplam todos os índices pluviométricos médios observados. Ao avaliarmos o impacto dado pela escolha do

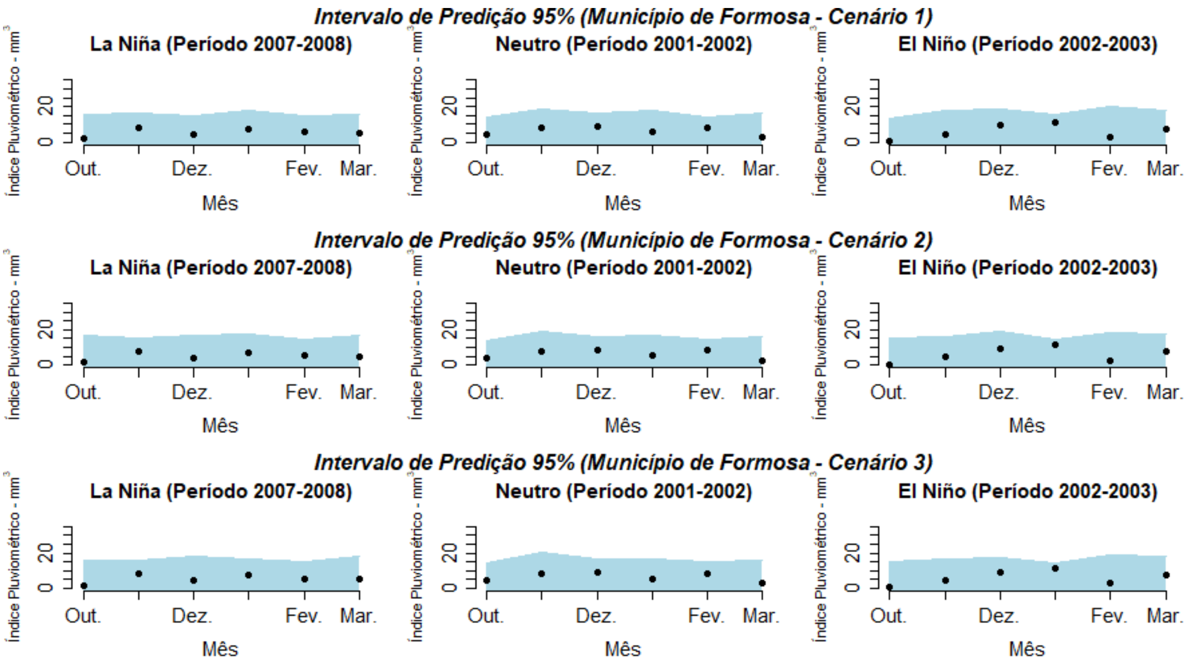

Figura 38 – Intervalos de predição para os índices pluviométricos médios mensais observados para a estação meteorológica "Treino com predição", localizada no município de Formosa-GO, considerando os cenários descritos na Tabela 18 ao longo dos efeitos climáticos observados após o período de 2000-2001.

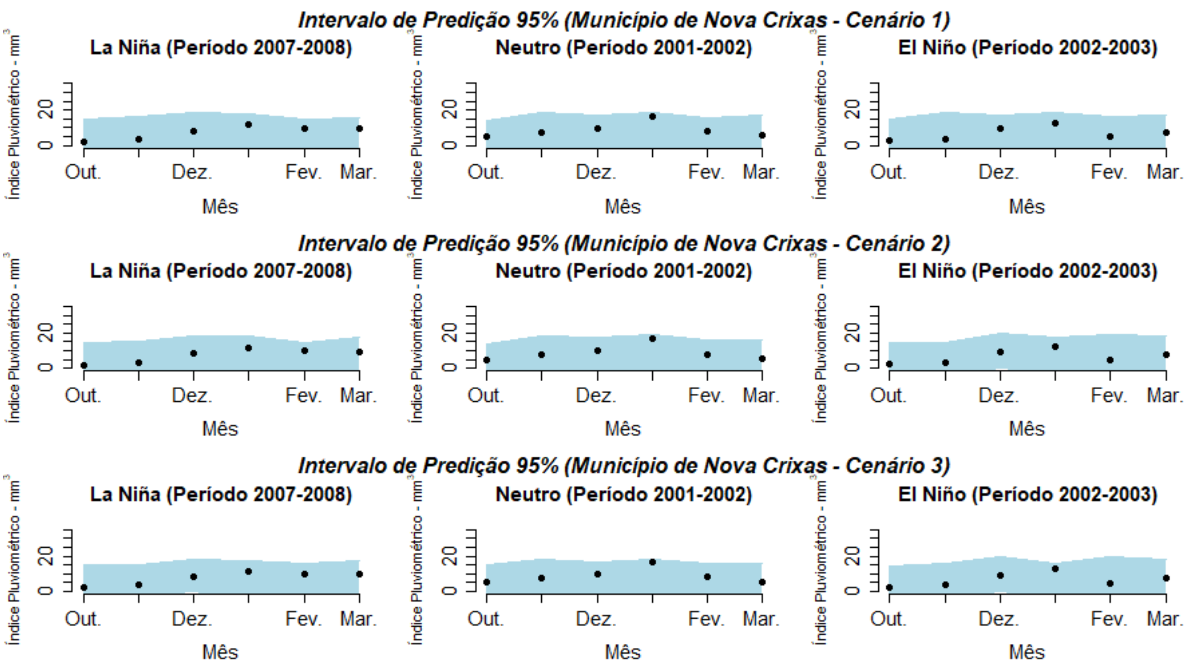

Figura 39 – Intervalos de predição para os índices pluviométricos médios mensais observados para a estação meteorológica "Treino com predição", localizada no município de Nova Crixas-GO, considerando os cenários descritos na Tabela 18 ao longo dos efeitos climáticos observados após o período de 2000-2001.

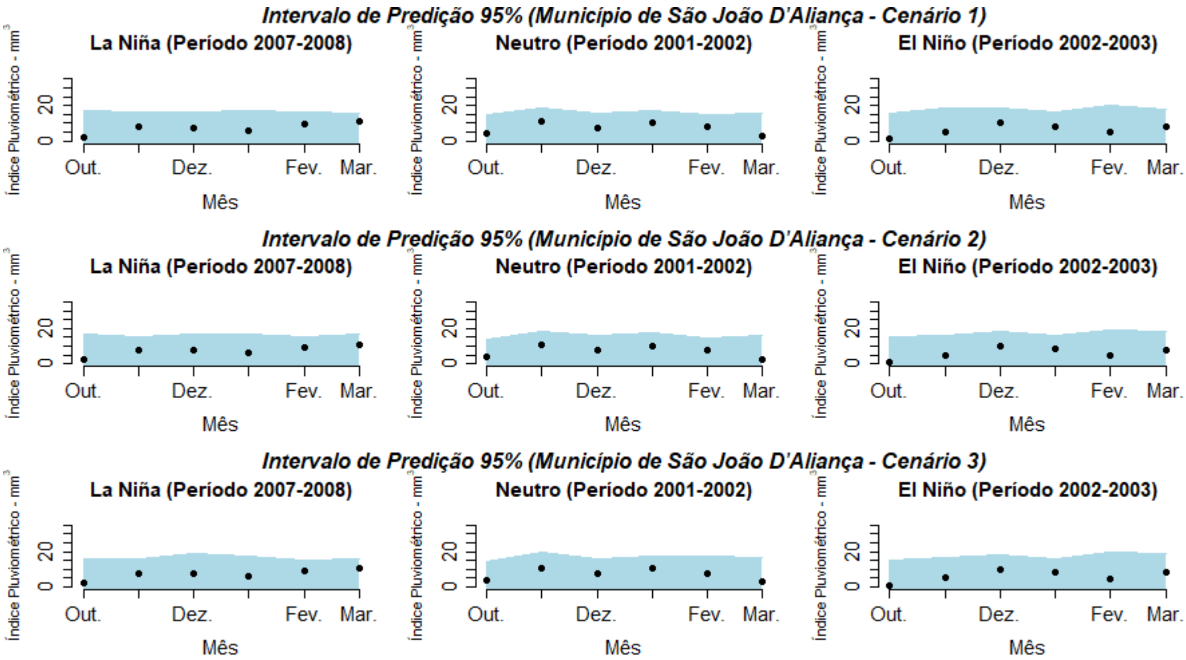

Figura 40 – Intervalos de predição para os índices pluviométricos médios mensais observados para a estação meteorológica "Treino com predição", localizada no município de São João D'Aliança-GO, considerando os cenários descritos na Tabela 18 ao longo dos efeitos climáticos observados após o período de 2000-2001.

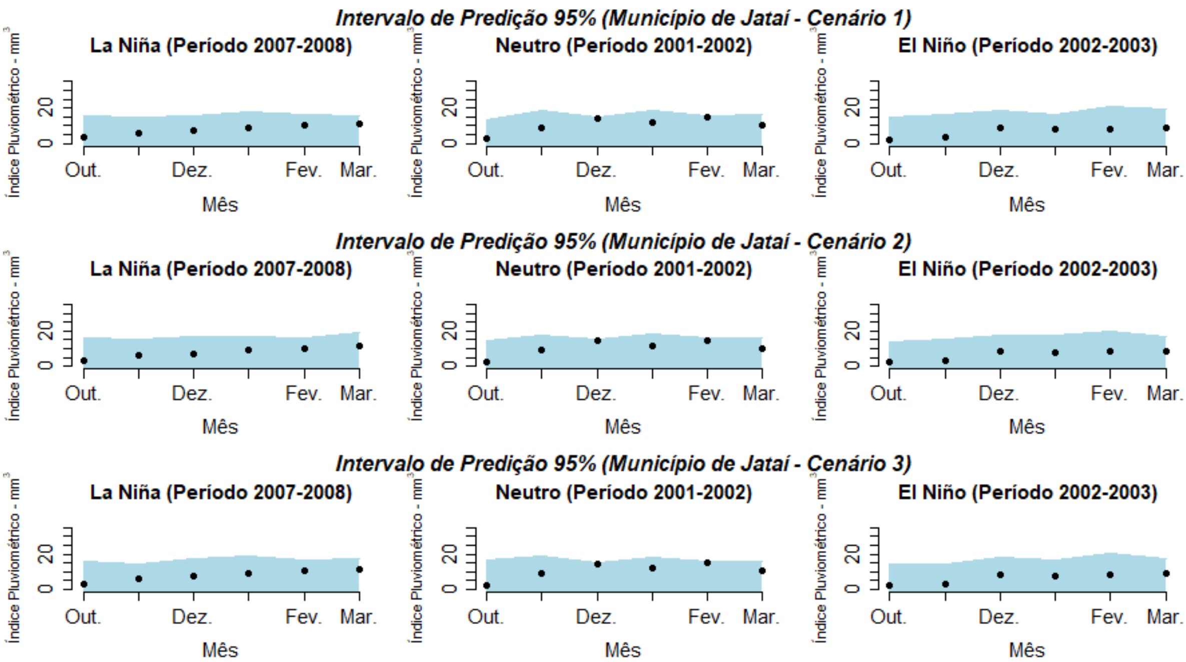

Figura 41 – Intervalos de predição para os índices pluviométricos médios mensais observados para a estação meteorológica "Treino com predição", localizada no município de Jatai-GO, considerando os cenários descritos na Tabela 18 ao longo dos efeitos climáticos observados após o período de 2000-2001.

cenário sobre o comportamento das bandas de predição, notamos que existe uma pequena vantagem para o Cenário 3, pois em alguns casos específicos, ele apresenta uma menor amplitude da banda de predição ajustada. Por exemplo, no município de Nova Crixas-GO (Figura 39) temos que o Cenário 3 apresentou uma menor amplitude da banda de predição ajustada para o efeito climático La Niña, período 2007-2008, quando comparado aos demais cenários neste caso. Este mesmo fato pode ser observado para os municípios de Formosa-GO e de Jataí-GO, Figura 38 e Figura 41 respectivamente, porém agora para a predição do efeito El Niño (período 2002-2003).

Ilustramos, nas Figuras 42 a 46, os intervalos de predição ajustados para as estações classificadas como "Teste", conforme descrito na Tabela 22, observadas nos períodos 2001-2002, 2002-2003 e 2007-2008, referente aos efeitos climáticos Neutro, El Niño e La Niña, respectivamente. Note que as bandas de predição possuem um comportamento similar quanto à alta amplitude e, também, por cobrirem todos os valores observados. Destacamos novamente que o Cenário 3 apresenta uma pequena vantagem em algumas situações específicas. Por exemplo, no município de Pirenópolis-GO (Figura 43) temos que o Cenário 3 apresentou uma leve redução na amplitude da banda de predição ajustada para o efeito climático Neutro, período 2001-2002, quando comparado aos demais cenários neste casos. Gostaríamos de destacar, que as amplitudes das bandas de predição para estações meteorológicas "Teste", ou seja, estações que não se encontram em regiões geográficas contempladas nos dados observados, foram similares às apresentadas pelas estações "Treino com predição".

Finalmente, nas Figuras 47 e 48, apresentamos respectivamente os efeitos espaciais e temporais preditos para os cenários destacados na Tabela 18, quando avaliados nos dados de precipitação do Estado de Goiás. Como podemos observar, em todos os cenários avaliados, os blocos de efeitos climáticos se diferenciam na amplitude dos intervalos de predição gerados. Por exemplo, para o efeito espacial predito, obtido ao longo das 87 estações avaliadas nos diferentes blocos de efeitos climáticos (Figura 47), a maior discrepância entre as amplitudes dos intervalos de predição aconteceu quando o Cenário 1 é acionado. Por outro lado, para o efeito temporal predito ao longo dos meses de Outubro a Março, avaliados nos diversos blocos de efeitos climáticos (Figura 48), notamos que os cenários estabelecidos apresentam comportamentos similares, com uma maior amplitude para o intervalo de predição observada nos blocos de efeitos climáticos La Ninã e El Niño.

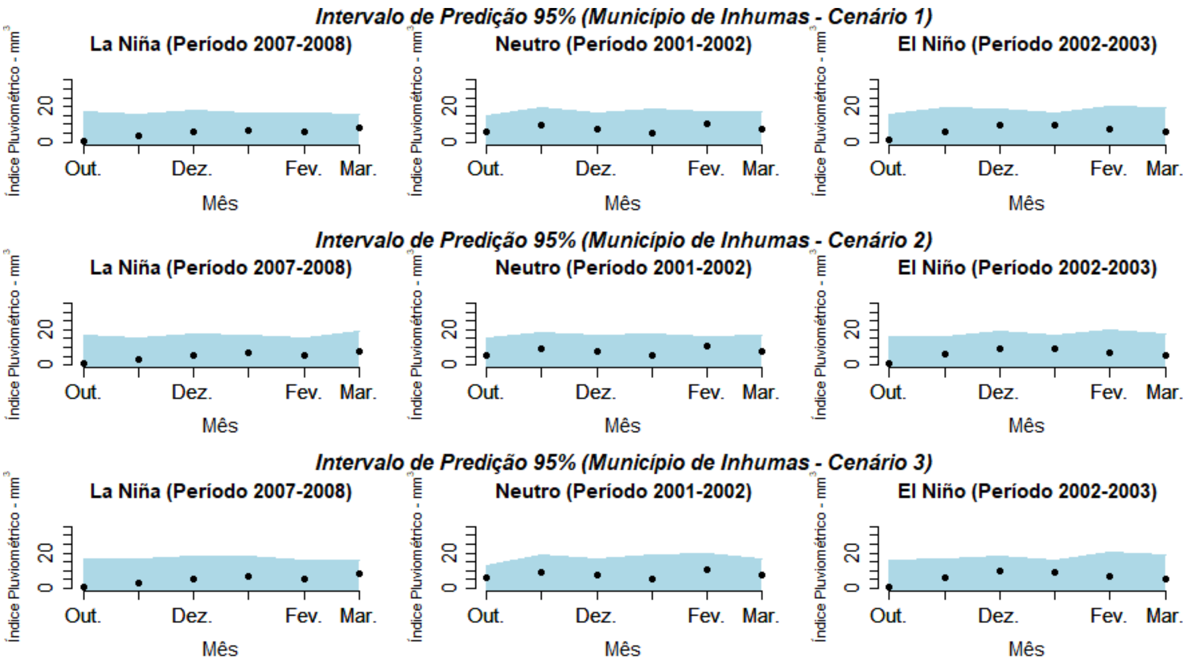

Figura 42 – Intervalos de predição para os índices pluviométricos médios mensais observados para a estação meteorológica "Teste", localizada no município de Inhumas-GO, considerando os cenários descritos na Tabela 18 ao longo de cada um dos efeitos climáticos observados após o período de 2000-2001.

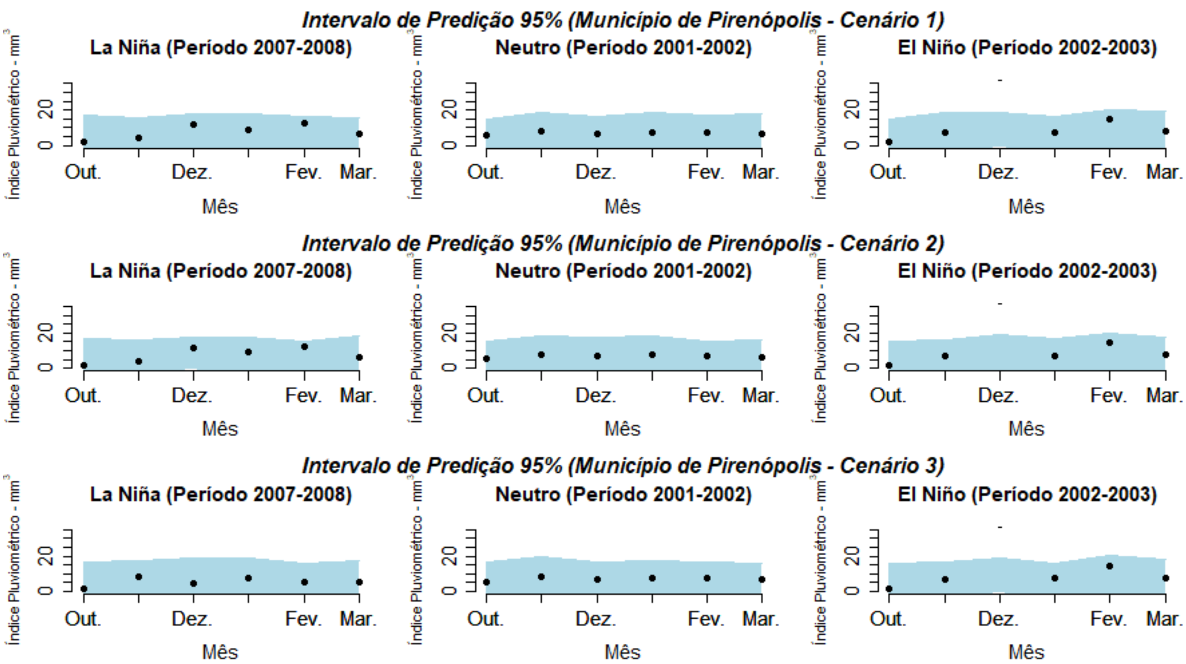

Figura 43 – Intervalos de predição para os índices pluviométricos médios mensais observados para a estação meteorológica "Teste", localizada no município de Pirenópolis-GO, considerando os cenários descritos na Tabela 18 ao longo de cada um dos efeitos climáticos observados após o período de 2000-2001.

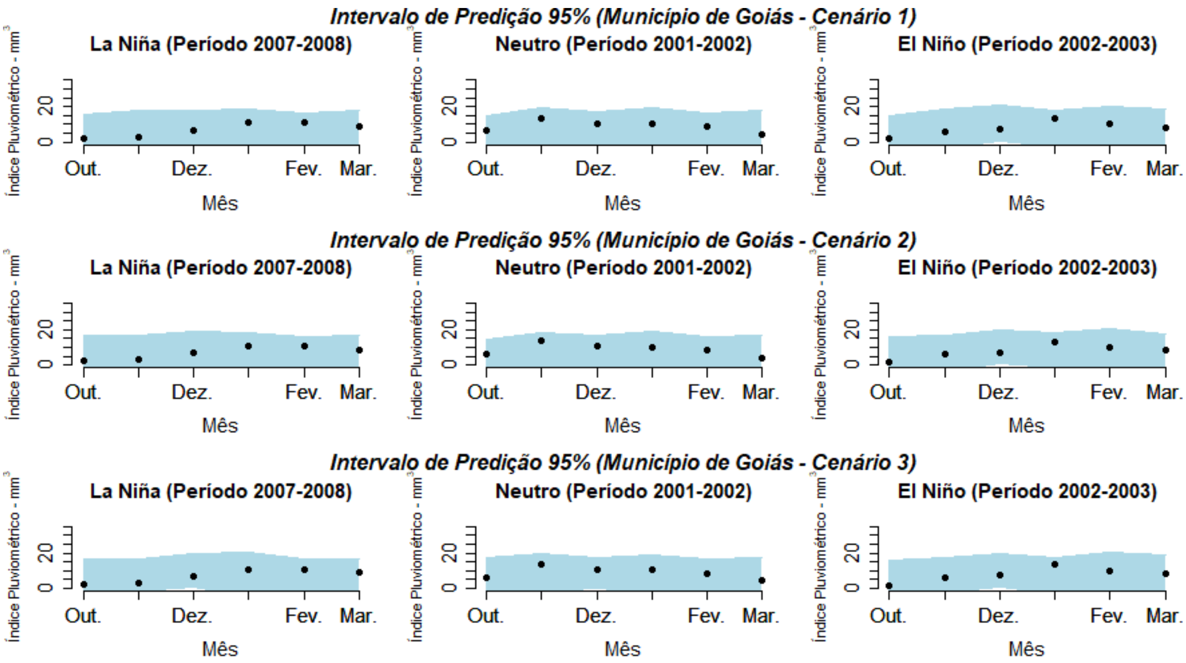

Figura 44 – Intervalos de predição para os índices pluviométricos médios mensais observados para a estação meteorológica “Teste” localizada no município de Goiás-GO, considerando os cenários descritos na Tabela 18, ao longo de cada um dos efeitos climáticos observados após o período de 2000-2001.

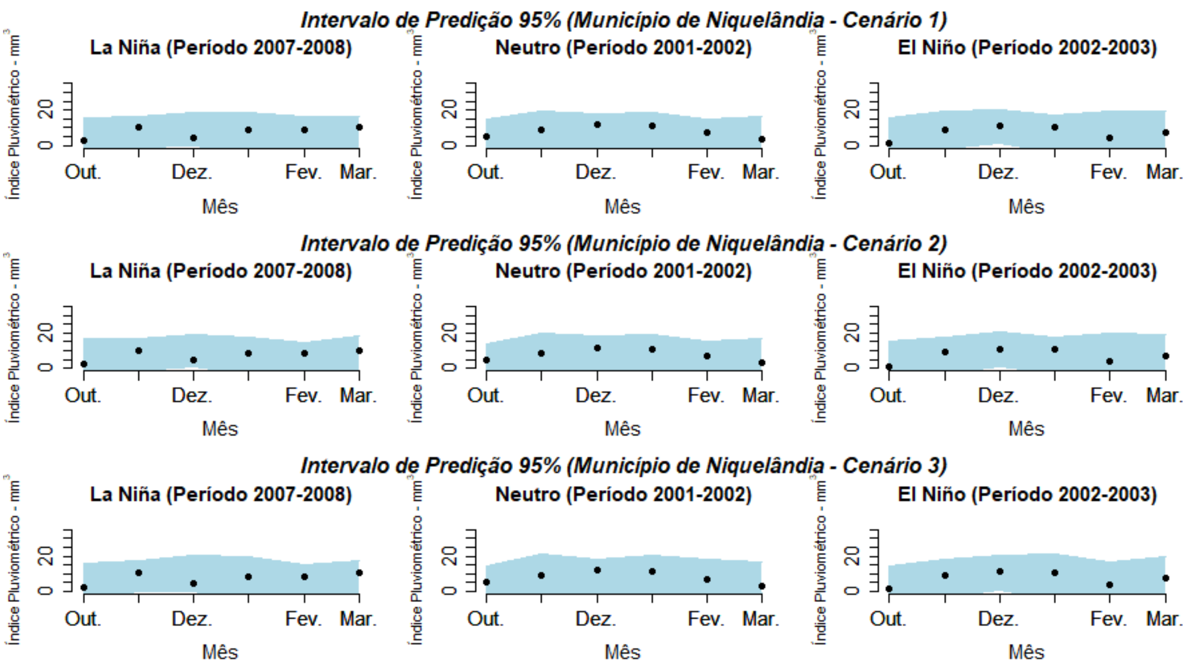

Figura 45 – Intervalos de predição para os índices pluviométricos médios mensais observados para a estação meteorológica “Teste”, localizada no município de Niquelândia-GO, considerando os cenários descritos na Tabela 18 ao longo de cada um dos efeitos climáticos observados após o período de 2000-2001.

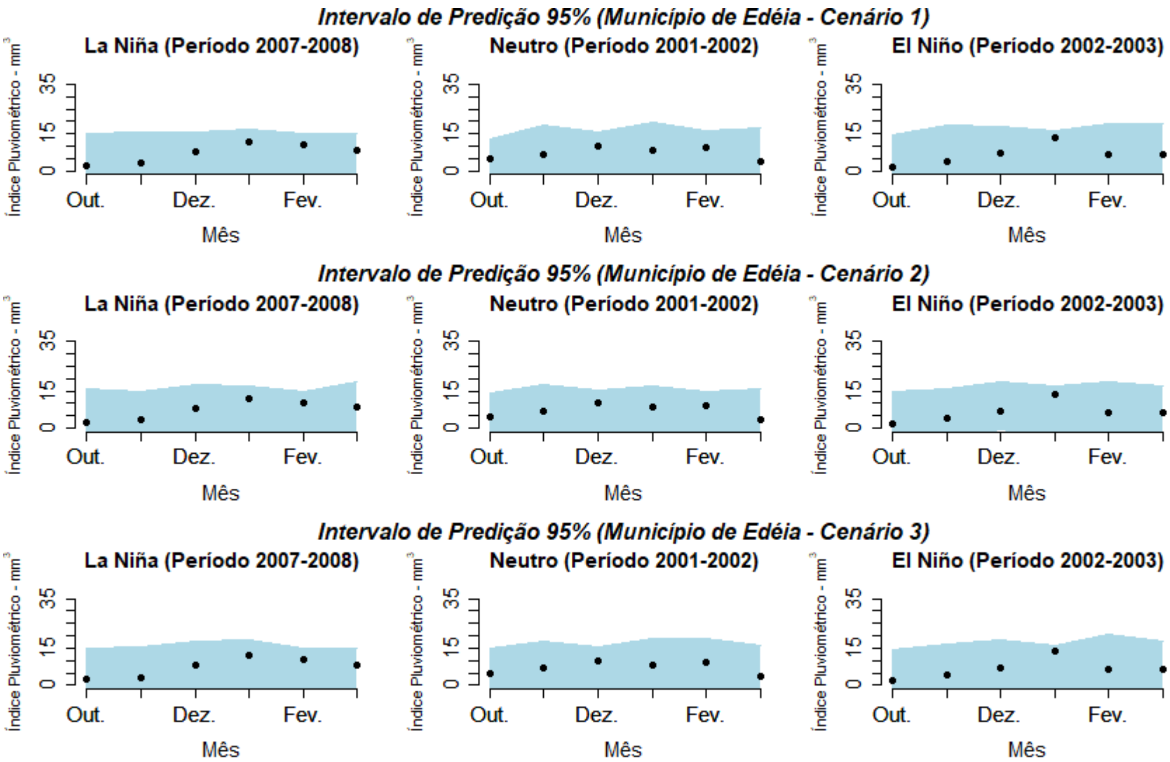

Figura 46 – Intervalos de predição para os índices pluviométricos médios mensais observados para a estação meteorológica "Teste", localizada no município de Edéia-GO, considerando os cenários descritos na Tabela 18 ao longo de cada um dos efeitos climáticos observados após o período de 2000-2001.

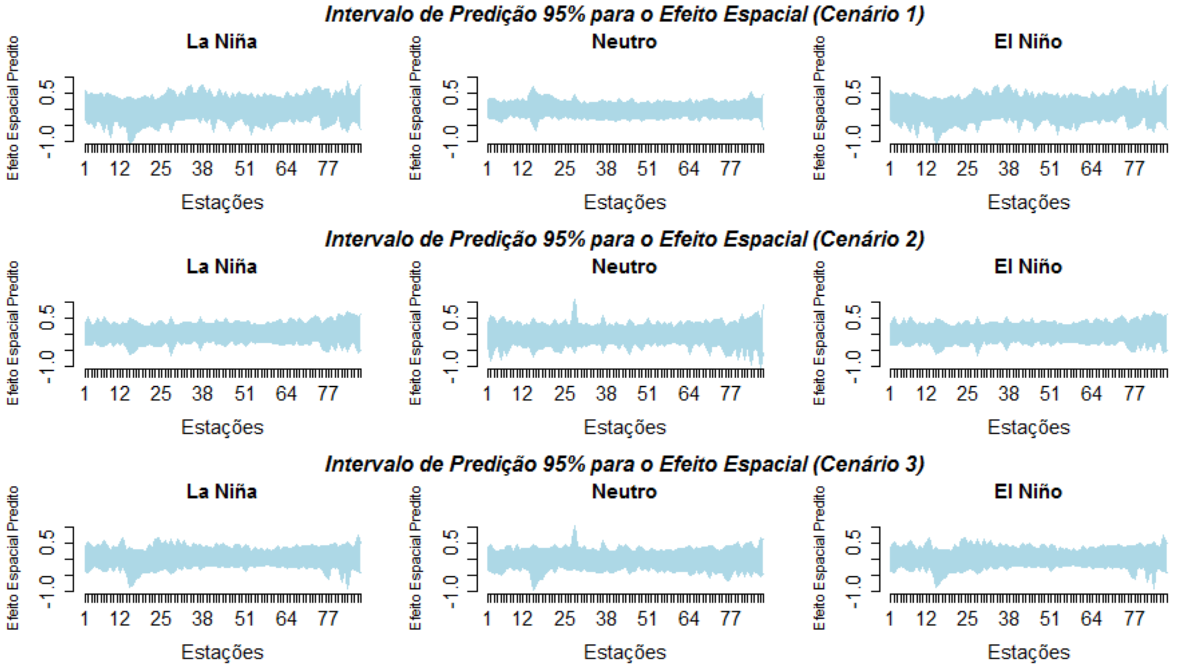

Figura 47 – Intervalo de predição 95% para os efeitos espaciais preditos ao longo das 87 estações avaliadas nos diferentes blocos de efeitos climático, quando ajustados aos dados de precipitação no Estado de Goiás.

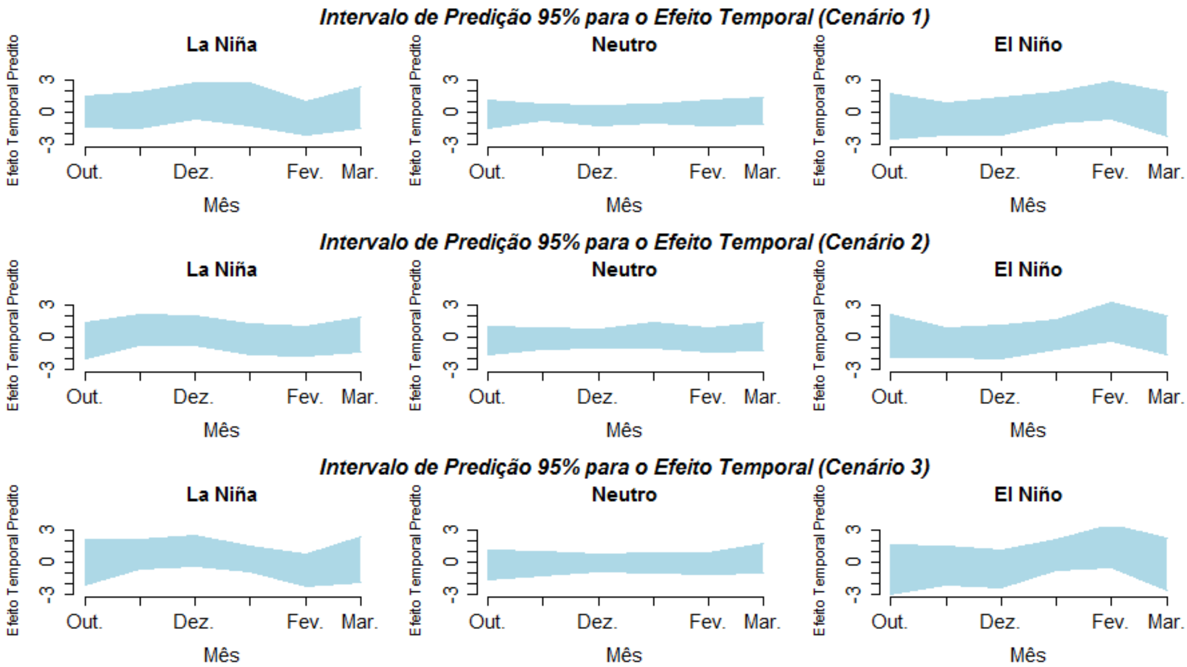

Figura 48 – Intervalo de predição 95% para os efeitos temporais preditos ao longo dos meses de Outubro a março, avaliadas nos diferentes blocos de efeitos climático, quando ajustados aos dados de precipitação no Estado de Goiás.

7 Conclusões

Nesse trabalho apresentamos modelos de regressão espaço-temporais para dados funcionais, com estrutura de agrupamento e de medida repetida, os quais foram incorporados por meio de componentes aleatórias espaciais e temporais, construídas a partir de expansões por funções bases *B-splines* cúbicos com coeficientes normais.

Uma das dificuldades encontradas no decorrer desse estudo foi o alto custo computacional envolvido no processo de estimação dos parâmetros, o que trouxe à tona uma proposta de um estudo empírico baseado em projeções ortogonais, ilustrado na Seção 5.5. Os resultados de simulação mostraram-se promissores. Entretanto, não foi possível mensurar analiticamente a diferença entre as distribuições a posteriori do modelo usual (2.1) e do modelo ajustado a partir de uma projeção ortogonal (5.3).

A complexidade do modelo proposto, dada pelo grande número de parâmetros a serem estimados, tornaram a técnica de estimação por máxima verossimilhança ineficaz. Por outro lado, a utilização da abordagem Bayesiana para a estimação dos parâmetros do modelo mostrou resultados satisfatórios, conforme estudos de simulação. É importante destacar que a utilização do pacote Rcpp (EDDELBUETTEL et al., 2017), que integra a biblioteca Armadillo C++ (SANDERSON; CURTIN, 2016) ao R (TEAM, 2013), foi o responsável por um significativo ganho no tempo de processamento requerido pelo amostrador de Gibbs, minimizando o custo computacional envolvido no processo de estimação dos parâmetros do modelo.

No decorrer do estudo de simulação, foi possível constatar que o processo de estimação do parâmetro κ, presente na estrutura de correlação *Matérn* (2.10), apresentou inconsistências durante a execução do amostrador de Gibbs, não sendo possível obter a convergência das cadeias geradas, mesmo após inúmeras tentativas de mudanças nos valores iniciais no processo de estimação, reforçando, portanto, as recomendações contrárias a estimação do parâmetro κ, conforme descrito em Stein (1999).

As estimativas do parâmetro κ, conjuntamente com o número de bases utilizadas nas estruturas funcionais, apresentadas em (2.2), (2.3) e (2.4), responsáveis pelos efeitos fixo, temporal e espacial, respectivamente, foram tratadas por intermédio dos critérios de comparação de modelos LPML e DIC7 que, por sua vez, foram eficientes na recuperação do verdadeiro valor do parâmetro κ. Cabe ressaltar ainda, que estudos preliminares de simulação mostraram que a aplicação de um mesmo número de bases *B-splines*, para todas as componentes funcionais do modelo proposto, geram graves problemas de confundimento no processo de estimação dos parâmetros.

No estudo sobre a sensibilidade das distribuições prioris foi possível verificar que as componentes de variância do ruído branco, ω^2, φ e ϕ, são invariantes à escolha das prioris, não apresentando grandes variações em suas respectivas estimativas a posteriori e probabilidades de cobertura. Com relação as componentes de variância dos efeitos aleatórios espaciais e temporais, denotados respectivamente por $\sigma^2_{\theta_i}$ e $\sigma^2_{\vartheta_i}$, $i = 1, 2, 3$, observamos que a distribuição a priori Gama Inversa$(0.01, 0.01)$ produz uma menor sensibilidade a posteriori. Outro fato relevante, em relação as componentes de variância dos efeitos funcionais, foi que a família de prioris Qui-Quadrado Escalonada Inversa apresentou um maior achatamento das estimativas a posteriori próximas ao zero, impactando fortemente o processo de estimação.

Após a escolha das distribuições a priori foi conduzido um estudo de simulação que permitiu evidenciar que o aumento no dimensionamento dos dados, descrito na prática pelo número de estações observadas (n), quantidade de meses apurados (τ) e pelo número de medidas repetidas (J_i, $i = 1, 2, 3$), contribui para uma melhoria da precisão para o processo de estimação para os parâmetros ω^2, φ e ϕ do modelo proposto em (2.1). Por outro lado, não foi possível identificar essa propriedade para as componentes de variância dos efeitos espaciais e temporais, denotadas respectivamente por $\sigma^2_{\theta_i}$ e $\sigma^2_{\vartheta_i}$, $i = 1, 2, 3$. É importante salientar que as estruturas algébricas das componentes espacial e temporal são compostas por produtos tensoriais que conduziram a uma maior sensibilidade no processo de estimação dos parâmetros, no entanto, esse fato não produziu reflexo direto no processo de previsão/predição.

O modelo proposto em (2.1) foi ajustado aos dados de precipitação de 87 estações meteorológicas localizadas nas diversas mesoregiões que constituem o Estado de Goiás, avaliadas nos meses de outubro a março para os períodos compreendidos entre 1980 a 2001 (21 anos). Ademais, cada período foi classificado de acordo com as condições climáticas definidas pelas variações da temperatura da superfície do mar e sua persistência ao longo do Oceano Pacífico Equatorial (NOAA, 2019). A partir das técnicas de seleção de modelos LPML e DIC7 foram selecionadas três configurações, que contemplam tanto a escolha do parâmetro κ quanto o número de bases utilizadas nas estruturas funcionais. Na análise dos intervalos de credibilidade dos dados de precipitação foi possível constatar que, em geral, os modelos destacados possuem comportamentos similares quanto à cobertura dos pontos observados sendo que, em sua maioria, encontramos pontos próximos ou internos às respectivas bandas de credibilidade. Já para os intervalos de predição dos dados de precipitação foram observadas altas variabilidades nos modelos destacados, no entanto, todos os pontos observados são encontrados internos às respectivas bandas de predição. É importante ressaltar que estudos de simulação, avaliados em dados simulados, indicaram que o aumento nas dimensões dos dados observados, seja pelo número de estações observadas (n) ou quantidade de meses apurados (τ), contribuem diretamente tanto para a redução da variabilidade da distribuição a posteriori quanto para uma menor amplitude dos intervalos de credibilidade e predição.

Portanto, o modelo proposto neste trabalho mostrou resultados promissores nos estudos de simulação e também quando aplicado ao problema de estimação do índice pluviométrico no Estado de Goiás. A flexibilidade da estrutura funcional permite uma boa adaptação à natureza dos dados, acomodando relações além da linear. No mais, descrever componentes espacias e temporais a partir de expansões em *B-splines* cúbicos, permite acomodar estruturas de correlação espaciais e temporais entre observações avaliadas em um mesmo bloco, atribuindo ao erro aleatório do modelo a tarefa de explicar a correlação espaço-temporal, entre observações em uma mesma medida repetida de um determinado bloco. É importante ressaltar que a disposição do modelo proposto em (2.1) permite ainda elaborar variações nas estruturas de correlações das componentes funcionais, o que será abordado nos estudos futuros.

8 Estudos Futuros

Estudos realizados em Matta et al. (2021) mostram que a precipitação acumulada no Estado de Goiás, ao longo das últimas três décadas, tem sido cada vez menos identificada pelas condições El Niño Oscilação Sul (ENOS). Diante deste cenário, apresentamos uma reformulação do modelo proposto em (2.1) que tem como objetivo retirar a estrutura de blocos, responsável por representar as diferentes classificações climáticas (La Niña, Neutro e El Niño), e supor que as medidas observadas ao longo dos diferentes anos, em uma determinada estação, sejam correlacionadas a partir de uma estrutura de correlação autoregressiva de ordem um (AR1).

Para isso, considere $Y_j(\mathbf{x}, t)$, $j = 1, ..., J$, a j-ésima medida observada no tempo $t \in D_T \subset \mathbb{N}$ e na localização $\mathbf{x} = (x_1, x_2) \in D_{lat} \times D_{long} \subset \mathbb{R}^2$. O modelo é dado por:

$$Y_j(\mathbf{x}, t) = \mu(\mathbf{x}, t) + \zeta_j(\mathbf{x}) + \Gamma_j(t) + \epsilon_j(\mathbf{x}, t). \tag{8.1}$$

Os erros aleatórios do modelo são descritos por $\epsilon_j(\mathbf{x}, t)$, $\mu(\mathbf{x}, t)$ representa a função média geral na localização e no tempo, apresentada em (2.2), $\zeta_j(\mathbf{x})$ representa o efeito espacial aleatório associado à localização $\mathbf{x}$ na observação j e $\Gamma_j(t)$ acomoda o efeito temporal aleatório associado ao tempo t na observação j. A componente $\zeta_j(\mathbf{x})$ será descrita por um efeito aleatório funcional com estrutura de correlação autoregressiva de ordem um (AR(1)). Logo, se considerarmos

$$\zeta_0(\mathbf{x}) = \sum_{f=1}^{K_{\zeta_1}} \sum_{g=1}^{K_{\zeta_2}} \theta_{fg} M_f^{(\Upsilon_{\zeta_1})}(x_1) M_g^{(\Upsilon_{\zeta_2})}(x_2), \tag{8.2}$$

teremos que:

$$\begin{aligned} \zeta_j(\mathbf{x}) &= \psi \zeta_{j-1}(\mathbf{x}) + \nu_j(\mathbf{x}) \\ &= \psi^j \zeta_0(\mathbf{x}) + \sum_{l_1=0}^{j-1} \psi^{l_1} \nu_{j-l_1}(\mathbf{x}), \end{aligned} \tag{8.3}$$

em que $|\psi| < 1$, $\boldsymbol{\Theta} = (\theta_{11}, \ldots, \theta_{1K_{\zeta_2}}, \ldots, \theta_{K_{\zeta_1}1}, \ldots, \theta_{K_{\zeta_1}K_{\zeta_2}}) \sim \mathbf{N}(\mathbf{0}, \boldsymbol{\Sigma_\Theta})$ e $\nu_j(\mathbf{x})$ são processos gaussianos independentes de média zero, com variância $\sigma_\nu^2 > 0$ e $Cov(\nu_j(\mathbf{x}_r), \nu_j(\mathbf{x}_s)) = 0$ para todo $\mathbf{x}_r$, $\mathbf{x}_s \in D_{lat} \times D_{long}$, $j \in \{1, 2, ..., J\}$. Por outro lado, considerando $\Gamma_0(t) = \sum_{l=1}^{K_\Gamma} \vartheta_l^{(i)} M_l^{(\Upsilon_\Gamma)}(t)$ teremos que $\Gamma_j(t)$ será representado por:

$$\begin{aligned} \Gamma_j(t) &= \eta\Gamma_{j-1}(t) + \alpha_j(t) \\ &= \eta^j\Gamma_0(t) + \sum_{l_2=0}^{j-1} \eta^{l_2}\alpha_{j-l_2}(t), \end{aligned} \quad (8.4)$$

em que $|\eta| < 1$, $\boldsymbol{\vartheta} = (\vartheta_1, \ldots, \vartheta_{K_\Gamma}) \sim \mathbf{N}(\mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{\vartheta}})$ e $\alpha_j(t)$ são processos gaussianos independentes de média zero, com variância $\sigma_\alpha^2 > 0$ e $Cov(\alpha_j(t_r), \alpha_j(t_s)) = 0$ para todo t_r, $t_s \in \mathbf{D}_T$, $j \in \{1, 2, ..., J\}$. Finalmente, consideramos ainda que os coeficientes $\boldsymbol{\Theta}$ e $\boldsymbol{\vartheta}$ são independentes.

É importante ressaltar que o modelo proposto em (2.1) admite dados em que não existe estrutura de grupos porém, nesse caso, iriamos assumir que todas as medidas são igualmente correlacionadas o que, por sua vez, pode não ser uma boa suposição principalmente quando trabalhamos com dados observados em diferentes décadas.

Referências

ABRAMOWITZ, M.; STEGUN, I. Handbook of mathematical functions pover. *New York*, 1970. Citado na página 35.

ALVARES, C. A.; STAPE, J. L.; SENTELHAS, P. C.; GONCALVES, J. L. D. M.; SPAROVEK, G. Koppen's climate classification map for Brazil. *Meteorologische Zeitschrift*, v. 22, n. 6, p. 711 – 728, 2013. Citado na página 76.

ARYAPUTERA, A. W.; YANG, D.; ZHAO, L.; WALSH, W. M. Very short-term irradiance forecasting at unobserved locations using spatio-temporal kriging. *Solar Energy*, Elsevier, v. 122, p. 1266–1278, 2015. Citado na página 28.

BLANGIARDO, M.; CAMELETTI, M.; BAIO, G.; RUE, H. Spatial and spatio-temporal models with R-inla. *Spatial and spatio-temporal epidemiology*, Elsevier, v. 4, p. 33–49, 2013. Citado na página 28.

BÖTTCHER, A.; WENZEL, D. The Frobenius norm and the commutator. *Linear algebra and its applications*, Elsevier, v. 429, n. 8-9, p. 1864–1885, 2008. Citado na página 67.

CARLIN, B. P.; LOUIS, T. A. *Bayesian Methods for Data Analysis*. New York: CRC Press, 2008. Citado na página 40.

CELEUX, G.; FORBES, F.; ROBERT, C.; TITTERINGTON, D. Deviance Information Criteria for Missing Data Models. *Bayesian Analysis*, v. 1, n. 4, p. 309–321, 2006. Citado 3 vezes nas páginas 31, 40 e 42.

CRESSIE, N.; HUANG, C. Classes of Nonseparable, Spatio-Temporal Stationary Covariance Functions. *Journal of the American Statistical Association*, v. 94, n. 448, p. 631–647, 1999. Citado na página 31.

DALY, C.; HALBLEIB, M.; SMITH, J. I.; GIBSON, W. P.; DOGGETT, M. K.; TAYLOR, G. H.; CURTIS, J.; PASTERIS, P. P. Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United States. *International Journal of Climatology: a Journal of the Royal Meteorological Society*, Wiley Online Library, v. 28, n. 15, p. 2031–2064, 2008. Citado 2 vezes nas páginas 28 e 29.

DE BOOR, C. *A Practical Guide to Splines*. New York: Springer-Verlag, 1978. Citado 2 vezes nas páginas 30 e 109.

DEY, D. K.; GELFAND, A. E.; PENG, F. Over Dispersed Generalized Linear Models. *Journal of Statistical Planning and Inference*, v. 1, n. 64, p. 93–107, 1997. Citado na página 41.

DONY, R. et al. Karhunen-loeve transform. *The transform and data compression handbook*, CRC Press Boca Raton, London, New York, Washington, DC, v. 1, n. 1-34, p. 29, 2001. Citado na página 64.

EDDELBUETTEL, D.; FRANCOIS, R.; ALLAIRE, J.; USHEY, K.; KOU, Q.; RUSSELL, N.; BATES, D.; CHAMBERS, J. *Rcpp: Seamless R and C++ Integration*. [S.l.], 2017. R package version 0.12.14. Citado 2 vezes nas páginas 31 e 97.

GAMERMAN, D.; IPPOLITI, L.; VALENTINI, P. et al. A dynamic structural equation approach to estimate the short-term effects of air pollution on human health. *Journal of the Royal Statistical Society Series C*, Royal Statistical Society, v. 71, n. 3, p. 739–769, 2022. Citado na página 28.

GAMERMAN, D.; LOPES, H. F. *Markov Chain Monte Carlo : Stochastic Simulation for Bayesian Inference*. Boca Raton: Chapman and Hall, 2006. Citado 3 vezes nas páginas 37, 38 e 58.

GEISSER, S. *Predictive Inference*. New York: CRC Press, 1993. Citado 2 vezes nas páginas 31 e 40.

GEISSER, S.; EDDY, W. F. A Predictive Approach to Model Selection. *Journal of the American Statistical Association*, v. 74, n. 365, p. 153–160, 1979. Citado 2 vezes nas páginas 31 e 40.

GELFAND, A. E. Model Determination Using Sampling-based Methods. *Markov Chain Monte Carlo in Practice*, p. 145–161, 1996. Citado na página 40.

GELMAN, A. Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper). *Bayesian Analysis*, v. 1, n. 3, p. 515–534, 2006. Citado 2 vezes nas páginas 43 e 44.

GELMAN, A.; RUBIN, D. B. Inference from iterative simulation using multiple sequences. *Statistical Science*, JSTOR, p. 457–472, 1992. Citado 3 vezes nas páginas 51, 69 e 81.

GUAN, B. T.; HSU, H.-W.; WEY, T.-H.; TSAO, L.-S. Modeling monthly mean temperatures for the mountain regions of taiwan by generalized additive models. *Agricultural and Forest Meteorology*, Elsevier, v. 149, n. 2, p. 281–290, 2009. Citado na página 29.

HE, X.; SHI, P. Bivariate tensor-product B-splines in a partly linear model. *Journal of Multivariate Analysis*, Elsevier, v. 58, n. 2, p. 162–181, 1996. Citado na página 110.

HEINEMANN, A. B.; MATTA, D. H. da; FERNANDES, I. K.; FRITSCHE-NETO, R.; COSTA-NETO, G. M. Enviromic prediction is useful to define the limits of climate adaptation: A case study of common beans in Brazil. *bioRxiv*, Cold Spring Harbor Laboratory, 2022. Citado na página 28.

LAURINI, M. P. A spatio-temporal approach to estimate patterns of climate change. *Environmetrics*, Wiley Online Library, v. 30, n. 1, p. e2542, 2019. Citado 2 vezes nas páginas 29 e 30.

LEWIS, K.; ALLEN, J.; RICHARDSON, A.; HOLT, J. Error quantification of a high resolution coupled hydrodynamic-ecosystem coastal-ocean model: Part3, validation with Continuous Plankton Recorder data. *Journal of Marine Systems*, Elsevier, v. 63, n. 3-4, p. 209–224, 2006. Citado na página 29.

MATTA, D. H.; COELHO, C. A. S.; SANTOS, L. L.; STONE, L. F.; HEINEMANN, A. B. The impact of different phases of the El Niño-Southern Oscillation phenomenon on Goiás State rainfalls and temperature characteristics across three decades. Preprint. *Theoretical and Applied Climatology*, 2021. Disponível em: https://www.researchsquare.com/article/rs-762347/v1. Acesso em: Dezembro 2021. Citado na página 100.

NOAA. Historical enso episodes (1950–present): Cold and warm episodes by season. national weather service, climate prediction center. In: . [s.n.], 2019. Disponível em: http://www.cpc.ncep.noaa.gov/products/analysis_monitoring/ensostuff/ensoyears_ERSSTv3b.shtml. Acesso em: Junho 2019. Citado 5 vezes nas páginas 30, 76, 77, 89 e 98.

PATTERSON, H. D.; THOMPSON, R. Recovery of inter-block information when block sizes are unequal. *Biometrika*, Oxford University Press, v. 58, n. 3, p. 545–554, 1971. Citado na página 63.

PLUMMER, M.; BEST, N.; COWLES, K.; VINES, K.; SARKAR, D.; BATES, D.; ALMOND, R.; MAGNUSSON, A. *coda: Output Analysis and Diagnostics for MCMC*. [S.l.], 2019. R package version 0.19-3. Citado 4 vezes nas páginas 45, 51, 69 e 81.

RAFTERY, A. E.; NEWTON, M. A.; SATAGOPAN, J. M.; KRIVITSKY, P. N. Estimating the Integrated Likelihood via Posterior Simulation Using the Harmonic Mean Identity. *Bayesian Statistics (eds J. M. Bernardo, M. J. Bayarri, J. O. Berger, A. P. Dawid, D. Heckerman, A. F. M. Smith and M. West)*, v. 8, p. 1–45, 2006. Citado na página 41.

RAMSAY, J. O. When the data are functions. *Psychometrika*, n. 47, p. 379–396, 1982. Citado 2 vezes nas páginas 107 e 108.

RAMSAY, J. O.; SILVERMAN, B. W. *Applied Functional Data Analysis: Methods and Case Studies*. New York: Springer, 1993. Citado 2 vezes nas páginas 29 e 108.

RICE, J. A. Functional and longitudinal data analysis: perspectives on smoothing. *Statistica Sinica*, v. 1, n. 14, p. 1330–1339, 2004. Citado na página 108.

ROBERT, C. P.; CASELLA, G.; CASELLA, G. *Monte Carlo statistical methods*. [S.l.]: Springer, 1999. v. 2. Citado na página 38.

SANDERSON, C.; CURTIN, R. Armadillo: a template-based c++ library for linear algebra. *Journal of Open Source Software*, v. 1, n. 2, p. 26, 2016. Citado na página 97.

SANTOS, M. P. dos; HEINEMANN, A. B.; STONE, L. F.; MATTA, D. H. da; CASTRO, J. R. de; SANTOS, A. B. dos. Nitrogen determination in irrigated rice using spectral reflectance. *Agronomy Journal*, Wiley Online Library, v. 113, n. 6, p. 5087–5101, 2021. Citado na página 28.

SCHUMAKER, L. *Spline Functions: Basic Theory*. New York: Wiley, 1981. Citado na página 109.

SNEPVANGERS, J.; HEUVELINK, G.; HUISMAN, J. Soil water content interpolation using spatio-temporal kriging with external drift. *Geoderma*, Elsevier, v. 112, n. 3-4, p. 253–271, 2003. Citado na página 28.

SPIEGELHALTER, D. J.; BEST, N. G.; CARLIN, B. P.; LINDE, A. v. d. Bayesian Measures of Model Complexity and Fit. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, v. 4, n. 64, p. 583–639, 2002. Citado na página 42.

SRIVASTAVA, M.; ROSEN, D. V. Models with Unknown Singular Covariance Matrix. *Linear Algebra Appl.*, n. 354, p. 255–273, 2002. Citado 2 vezes nas páginas 64 e 67.

STAUFFER, R.; MAYR, G. J.; MESSNER, J. W.; UMLAUF, N.; ZEILEIS, A. Spatio-temporal precipitation climatology over complex terrain using a censored additive regression model. *International Journal of Climatology*, Wiley Online Library, v. 37, n. 7, p. 3264–3275, 2017. Citado 2 vezes nas páginas 28 e 29.

STEIN, M. L. *Interpolation of spatial data: some theory for kriging*. [S.l.]: Springer Science & Business Media, 1999. Citado 2 vezes nas páginas 37 e 97.

TEAM, R. C. *R: A Language and Environment for Statistical Computing*. Vienna, Austria, 2013. Disponível em: http://www.R-project.org/. Citado 4 vezes nas páginas 31, 45, 51 e 97.

THORNTON, P. E.; RUNNING, S. W.; WHITE, M. A. Generating surfaces of daily meteorological variables over large regions of complex terrain. *Journal of hydrology*, Elsevier, v. 190, n. 3-4, p. 214–251, 1997. Citado na página 28.

VICENTE-SERRANO, S. M.; SAZ-SÁNCHEZ, M. A.; CUADRAT, J. M. Comparative analysis of interpolation methods in the middle Ebro Valley (Spain): application to annual precipitation and temperature. *Climate research*, v. 24, n. 2, p. 161–180, 2003. Citado na página 29.

Apêndices

APÊNDICE A – Dados Funcionais

A ADF é um ramo da estatística não paramétrica que permite estudar conjunto de dados em que uma ou mais variáveis são dadas por uma curva, superfície ou outro tipo de função (RAMSAY, 1982). Por exemplo, suponha que estamos interessados em avaliar o índice pluviométrico médio mensal no município de Paraúna-GO, durante o período de 1980 a 1985, entre os meses de outubro a março. Como ilustração desta situação, na Figura 49 temos cinco curvas, observadas em seis pontos (meses de outubro a março).

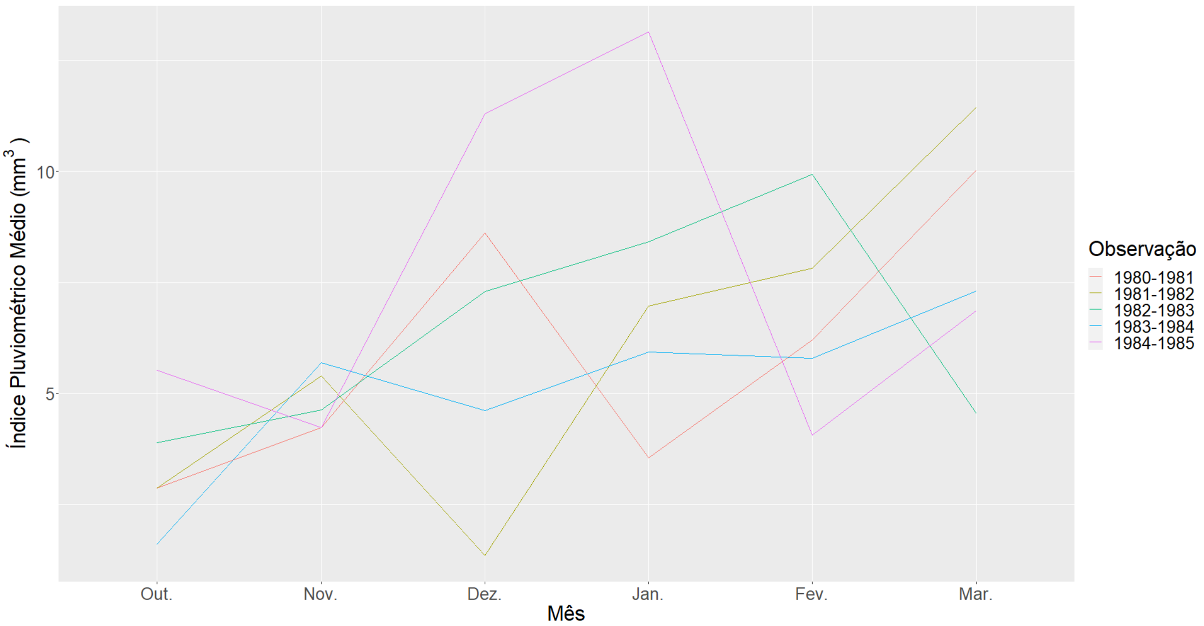

Figura 49 – Precipitação média mensal no município de Paraúna-GO observada durante o período de 1980 a 1985, entre os meses de outubro a março.

Considere (y_{ij}, t_j), $i = 1, 2, \ldots, n$, $j = 1, 2, \ldots, \tau$, modelados a partir da regressão não paramétrica

$$y_{ij} = f(t_j) + \epsilon_{ij}, \tag{A.1}$$

em que f pertence a um espaço funcional infinito dimensional, tal que $f(t_j) = \sum_{i=1}^{\infty} \alpha_i w_i(t_j)$ para algum conjunto de funções conhecidas $\{w_i\}$, e ϵ_{ij} são variáveis aleatórias independentes e identicamente distribuídas de média zero e variância σ^2. Assim, um dos objetivos da regressão não paramétrica é estimar a função f a partir de $f = \sum_{i=1}^{K} \alpha_i w_i$, em que K é um natural positivo que representa o número de funções w_j necessárias para estimar f. Logo, na regressão não paramétrica não é preciso estipular uma estrutura pré-determinada para f, sendo imprescindível apenas a escolha de um espaço de funções apropriado que, de modo geral, estará associado ao grau de suavidade que se deseja para a curva de regressão f.

Na determinação da componente funcional de um conjunto de dados, o grau de suavidade aplicado é determinado pelo número de funções base a serem utilizadas, sendo que, quanto maior for o valor de K, melhor será o ajuste dos dados. Porém, quanto maior for K, maior será o risco de ajustarmos também o ruído (que deveria ser ignorado), ocorrendo, por exemplo, uma interpolação no ajuste. Por outro lado, se K for muito pequeno podemos perder alguns aspectos importantes da suavidade da curva no processo de estimação. É importante ressaltar que a escolha das funções base deverá se pautar também pelas características que se relacionem com aquelas encontradas nas funções que serão estimadas, isto é, tal escolha deverá proporcionar uma adequada aproximação utilizando um menor número de bases. Para maiores detalhes a respeito dessa ótica veja Ramsay e Silverman (1993) e Rice (2004).

Dentre as possíveis escolhas de espaços funcionais, destacamos o espaço das funções definidas em um intervalo $[a, b]$ quadrado integrável, ou seja, $f \in L^2(\mathbb{R}) = \{f : \int f^2 < \infty\}$, no qual uma função f pode ser expressa por meio de uma componente funcional descrita pela combinação linear de estruturas de construção funcional elementares intituladas funções base, que permite a passagem dos dados da forma discreta (maneira como são coletados) para a forma de curvas (RAMSAY, 1982). Segundo Ramsay e Silverman (1993), é importante ressaltar que técnicas de suavização por bases, quando mal ajustadas, podem não representar bem as curvas e suas características de interesse.

Considerando as diferentes funções f que podem ser utilizadas, temos que as funções *Spline* possuem estruturas simplificadas e com boas propriedades para aproximar funções. De modo geral, um intervalo $R = [a, b]$ é dividido em subintervalos menores da forma $[a, r_1], \ldots, [r_K, b]$ e assim, um polinômio p_i de grau pequeno, no nosso caso grau 3 (grau = ordem do polinômio - 1), é empregado para a aproximação da função em cada subintervalo $[r_i, r_{i+1}]$, $i = 0, \ldots, K$. Essa técnica permite estimar f por meio de funções polinomiais por partes $S(.)$, em que $S(r) = p_i(r)$ para $r \in [r_i, r_{i+1}]$, $i = 0, \ldots, K$. Os valores extremos dos subintervalos são denominados de nós (*Knots*), sendo a e b os nós exteriores e os demais $r_1, \ldots, r_K$ chamados de nós interiores.

Geralmente, se os pedaços de polinômio $p_i(r)$ são construídos independentemente dos demais não se obtém uma função contínua $S(r)$, o que não é apropriado quando desejamos aproximar uma função suave. Logo, faz-se necessário que os polinômios sejam unidos suavemente nos nós interiores $r_1, \ldots, r_K$ e, também, que sejam deriváveis um certo número de vezes. Como resultado, obtém-se uma função polinomial por partes, suave, chamada de função *Spline*.

Uma função *Spline* de ordem m com K nós interiores em $r_1, \dots, r_K$ é qualquer função da forma

$$S(r) = \sum_{i=0}^{m-1} \alpha_i r^i + \sum_{i=1}^{K} \chi_i (r - r_i)_+^{m-1}, \tag{A.2}$$

em que os coeficientes $\alpha_0, \dots, \alpha_{m-1}, \chi_0, \dots, \chi_K$ são números reais e, a função de potência truncada para um dado grau q é definida como:

$$u_+^q = \begin{cases} u^q, \text{ se } u \geqslant 0 \\ 0, \text{ se } u < 0. \end{cases} \tag{A.3}$$

Assim, uma função *Spline* é portanto uma combinação linear de $m + K$ funções base. Considerando o conjunto de nós internos $r_1, \dots, r_K$, de acordo com (A.2), teremos então as seguintes funções base $\{1, r, \dots r^{m-1}, (r - r_1)^{m-1}, \dots, (r - r_K)^{m-1}\}$.

O conjunto de funções *Spline* de ordem m e com nós interiores em $r_1, \dots, r_K$ é chamado de espaço *Spline*, sendo denotado por $S_m(r_1, \dots, r_K)$. Neste caso, teremos então um espaço linear de dimensão $m + K$ em que os chamados *B-splines* formam uma base do espaço *Spline* (SCHUMAKER, 1981).

Um *B-spline* de ordem m consiste de m pedaços de polinômio, cada um de ordem m. Tais pedaços de polinômio se juntam de forma especial em $m - 1$ nós internos. Nos pontos de união, as derivadas de ordem até $m - 2$ são contínuas. O *B-spline* também é positivo no domínio dos $m + 1$ nós e zero fora dele. Exceto nas fronteiras, o *B-spline* sobrepõe $2(m - 1)$ pedaços de polinômio de seus vizinhos, sendo que para um certo r, m os *B-splines* são não nulos.

As funções *B-spline* são blocos de construção (bases funcionais) extremamente flexíveis para estimação de curvas. Logo, uma função $f \in L^2(\mathbb{R})$ pode ser aproximada pela seguinte combinação linear

$$f(t) = \sum_{i=1}^{n} c_i B_{i,K}^{(\Upsilon)}(t), \tag{A.4}$$

em que $B_{i,K}^{(\Upsilon_j)}(.)$ são bases *B-splines* de ordem K, com nós interiores contidos no conjunto Υ (SCHUMAKER, 1981). De Boor (DE BOOR, 1978) desenvolveu um algoritmo para calcular um *B-spline* de qualquer ordem por meio de um *B-spline* de ordem menor, ou seja, é possível calcular os *B-splines* por meio de uma relação de recorrência. Devido ao fato de um *B-spline* de ordem 1 ser uma constante em um intervalo entre dois nós, o cálculo de um *B-spline* de qualquer ordem é facilitado. Esse algoritmo funciona quando os nós são igualmente espaçados ou também quando não são. A Figura 50 mostra um *B-spline* de ordem 4 com 10 funções base e nós igualmente espaçados.

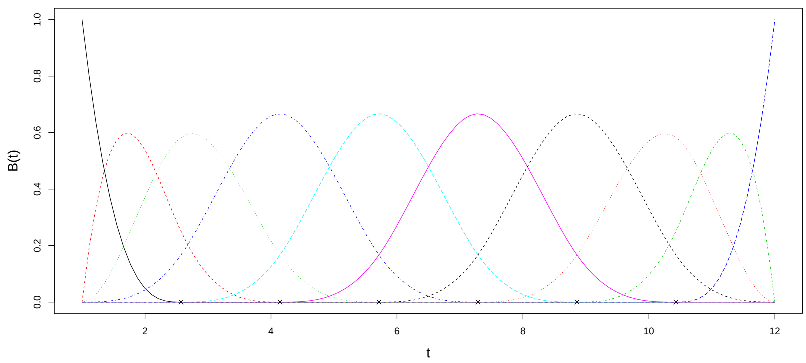

Figura 50 – Ilustração de um B-spline de ordem 4 com 10 funções base.

Os resultados obtidos para o caso unidimensional podem ser estendidos para várias dimensões por meio do uso de produto tensorial *Spline*. Neste caso, cada dimensão é tratada como um subespaço *Spline* (HE; SHI, 1996). Em outras palavras, uma função de produto tensorial ($f : \mathbb{R}^s \longrightarrow \mathbb{R}$) pode ser aproximada por

$$f(t_1, ..., t_s) = \sum_{i_1=1}^{n_1} ... \sum_{i_s=1}^{n_s} c_{i_1,...,i_s} B_{i_1,K_1}^{(\Upsilon_1)}(t_1)....B_{i_s,K_s}^{(\Upsilon_s)}(t_s), \tag{A.5}$$

em que $B_{i_j,K_j}^{(\Upsilon_j)}(t_j)$ são bases *B-splines* de ordem K_j, com nós interiores contidos no conjunto Υ_j. Para maiores detalhes dessa técnica veja He e Shi (1996).

APÊNDICE B – Estudo da Sensibilidade da Distribuição a Priori das Componentes de Variância

Nas Figuras 51 e 52 apresentamos respectivamente as médias e modas a posteriori das componentes de variância ω^2, φ e ϕ, do modelo proposto em (2.1), tomadas ao longo dos 50 conjuntos de dados simulados considerando as famílias de prioris Gama Inversa e Qui-Quadrado Escalonada Inversa para os parâmetros ω^2, $\sigma^2_{\theta_i}$ e $\sigma^2_{\vartheta_i}$, $i = 1, 2, 3$, e fixando Gama$(1, 0.01)$ e Gama$(1, 0.001)$ para os parâmetros φ e ϕ, respectivamente. Como podemos observar, as médias e modas a posteriori para os parâmetros φ e ϕ, ilustradas reciprocamente nas Figuras 51 e 52, não são muito impactadas por mudanças nas distribuições a priori dos demais parâmetros em estudo. Com relação ao parâmetro ω^2, notamos uma maior robustez à escolha das prioris, não apresentando grandes variações para as médias e modas a posteriori. Além disso, observamos indícios acerca da simetria das distribuições a posteriori, uma vez que nestes casos as estimativas da mediana e da média, Figuras 1 e 51, respectivamente, variam em suportes semelhantes.

Nas Figuras 53 e 54 apresentamos as médias e as modas a posteriori das componentes de variância dos efeitos aleatórios espaciais $\sigma^2_{\theta_i}$, $i = 1, 2, 3$, tomadas durante o estudo supracitado. Podemos observar indícios de achatamento das estimativas nas proximidades do zero, além de presenciarmos evidências acerca da assimetria das distribuições a posteriori devido ao posicionamento das estimativas mediana e média a posteriori, Figuras 2 e 53, respectivamente. Estas mesmas características também são encontradas nas componentes de variância dos efeitos aleatórios temporais, $\sigma^2_{\vartheta_i}$, $i = 1, 2, 3$, evidenciadas nas Figuras 3, 55 e 56.

Nas Figuras 55 e 56 apresentamos respectivamente as médias e modas a posteriori das componentes de variância dos efeitos aleatórios temporais, denotados respectivamente por $\sigma^2_{\vartheta_i}$, $i = 1, 2, 3$, do modelo proposto em (2.1), tomadas ao longo dos 50 conjuntos de dados simulados quando consideradas as famílias de prioris Gama Inversa e Qui-Quadrado Escalonada Inversa para os parâmetros ω^2, $\sigma^2_{\theta_i}$ e $\sigma^2_{\vartheta_i}$, $i = 1, 2, 3$, e fixando Gama$(1, 0.001)$ e Gama$(1, 0.01)$ para os parâmetros φ e ϕ, respectivamente.

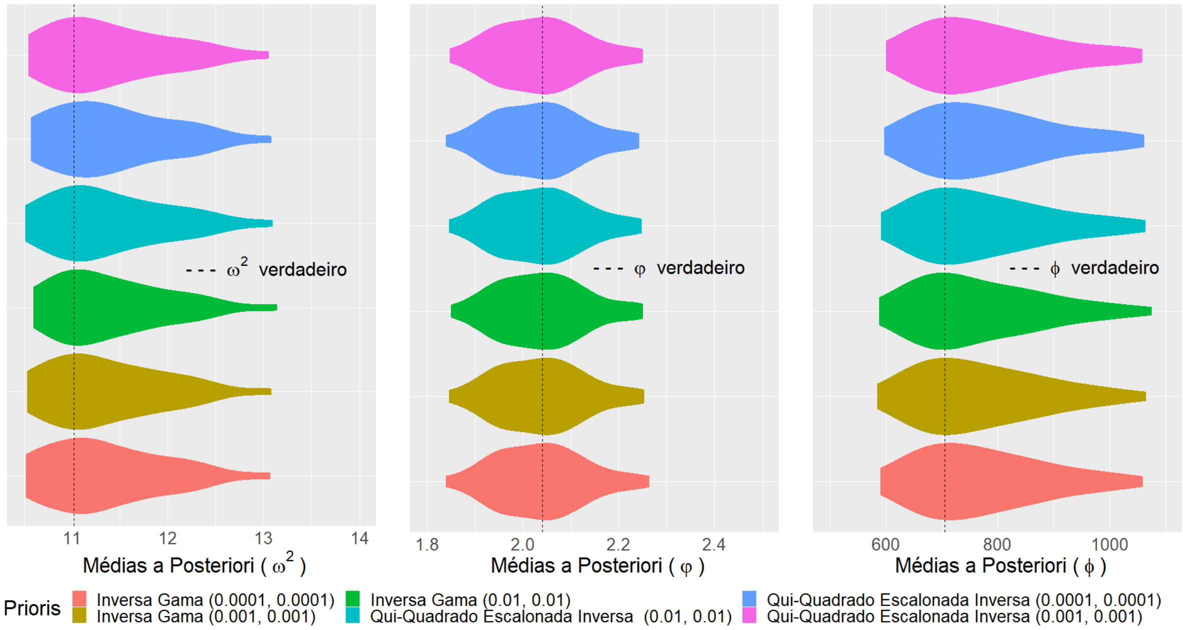

Figura 51 – Médias a posteriori das componentes de variância ω^2, φ e ϕ, do modelo proposto em (2.1), tomadas ao longo dos 50 conjuntos de dados simulados considerando as famílias de prioris Gama Inversa e Qui-Quadrado Escalonada Inversa para os parâmetros ω^2, $\sigma^2_{\theta_i}$ e $\sigma^2_{\vartheta_i}$, $i = 1, 2, 3$, e fixando Gama$(1, 0.01)$ e Gama$(1, 0.001)$ para os parâmetros φ e ϕ, respectivamente.

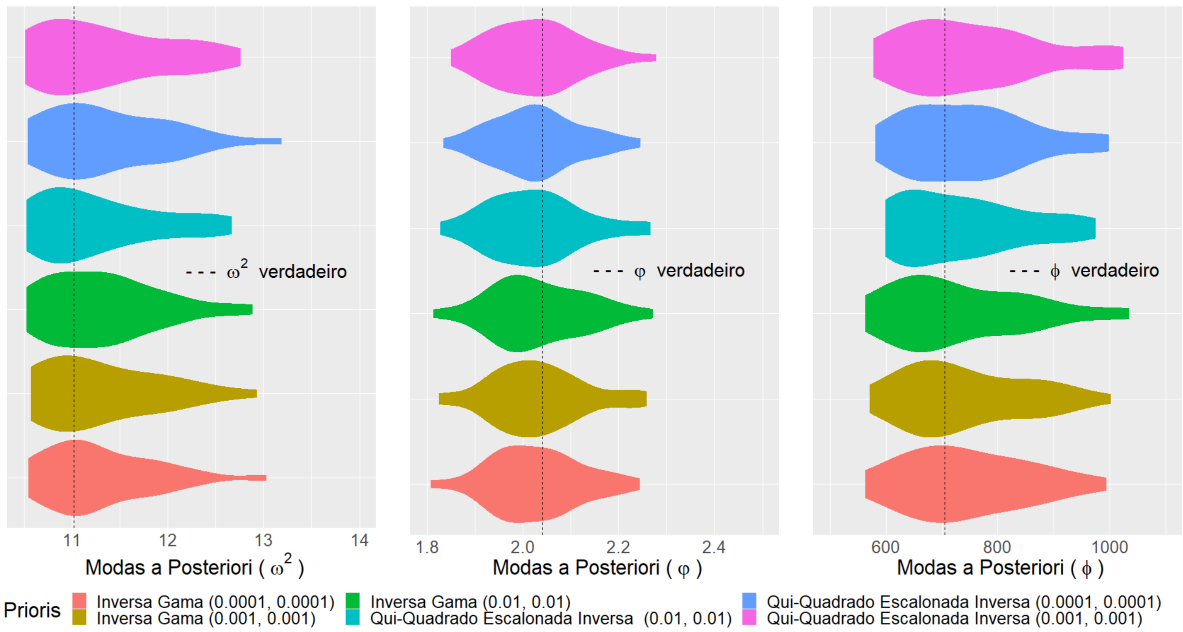

Figura 52 – Modas a posteriori das componentes de variância ω^2, φ e ϕ, do modelo proposto em (2.1), tomadas ao longo dos 50 conjuntos de dados simulados considerando as famílias de prioris Gama Inversa e Qui-Quadrado Escalonada Inversa para os parâmetros ω^2, $\sigma^2_{\theta_i}$ e $\sigma^2_{\vartheta_i}$, $i = 1, 2, 3$, e fixando Gama$(1, 0.01)$ e Gama$(1, 0.001)$ para os parâmetros φ e ϕ, respectivamente.

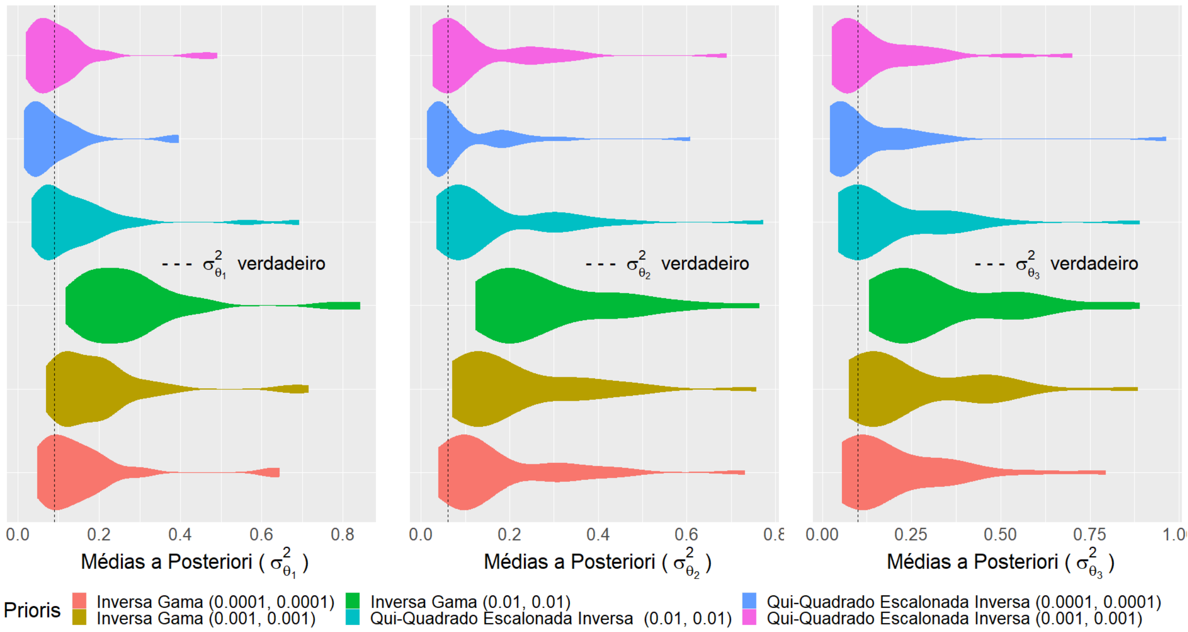

Figura 53 – Médias a posteriori das componentes de variância dos efeitos aleatórios espaciais, denotados respectivamente por $\sigma^2_{\theta_i}$, $i = 1, 2, 3$, do modelo proposto em (2.1), tomadas ao longo dos 50 conjuntos de dados simulados quando consideradas as famílias de prioris Gama Inversa e Qui-Quadrado Escalonada Inversa para os parâmetros ω^2, $\sigma^2_{\theta_i}$ e $\sigma^2_{\vartheta_i}$, $i = 1, 2, 3$, e fixando Gama$(1, 0.001)$ e Gama$(1, 0.01)$ para os parâmetros φ e ϕ, respectivamente.

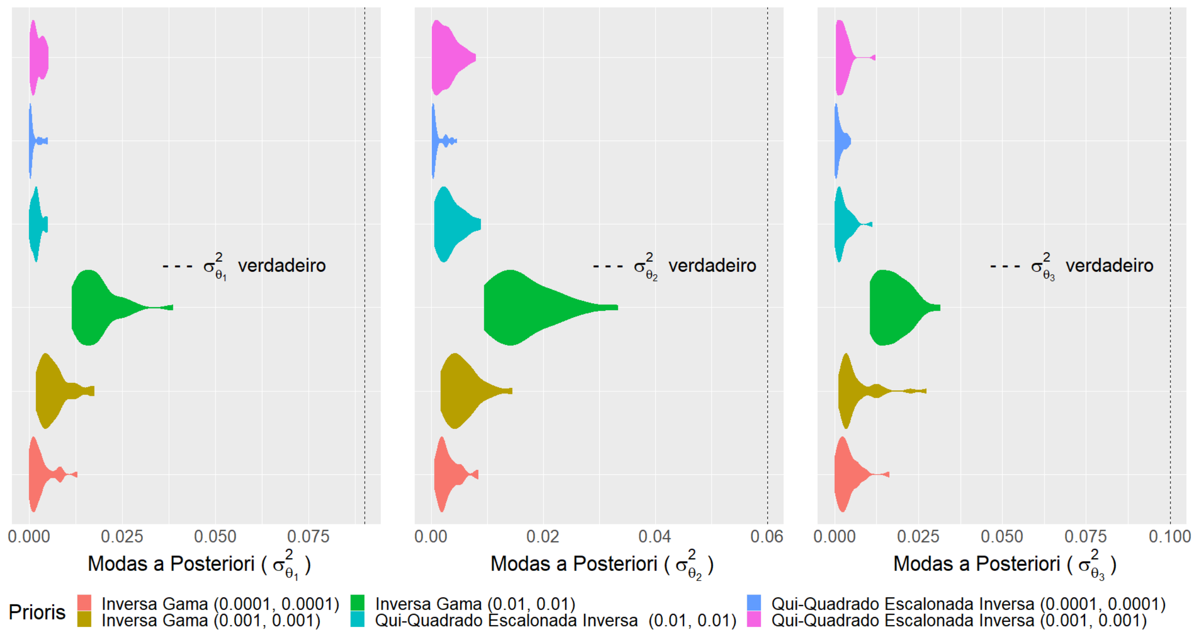

Figura 54 – Modas a posteriori das componentes de variância dos efeitos aleatórios espaciais, denotados respectivamente por $\sigma^2_{\theta_i}$, $i = 1, 2, 3$, do modelo proposto em (2.1), tomadas ao longo dos 50 conjuntos de dados simulados quando consideradas as famílias de prioris Gama Inversa e Qui-Quadrado Escalonada Inversa para os parâmetros ω^2, $\sigma^2_{\theta_i}$ e $\sigma^2_{\vartheta_i}$, $i = 1, 2, 3$, e fixando Gama$(1, 0.001)$ e Gama$(1, 0.01)$ para os parâmetros φ e ϕ, respectivamente.

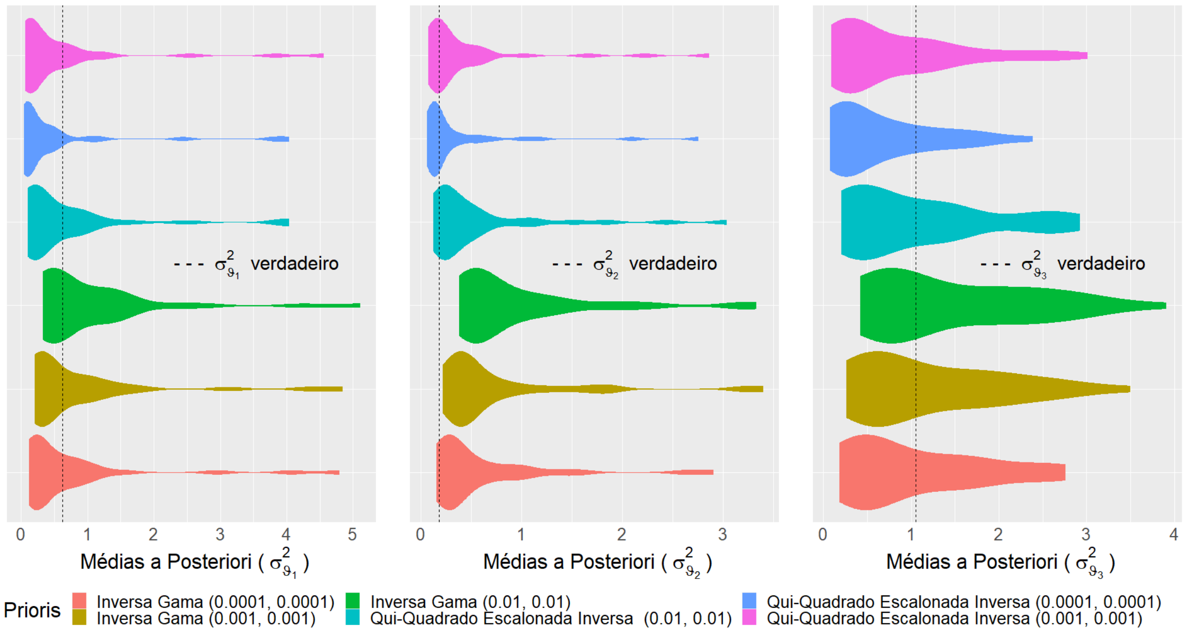

Figura 55 – Médias a posteriori das componentes de variância dos efeitos aleatórios temporais, denotados respectivamente por $\sigma^2_{\vartheta_i}$, $i = 1, 2, 3$, do modelo proposto em (2.1), tomadas ao longo dos 50 conjuntos de dados simulados quando consideradas as famílias de prioris Gama Inversa e Qui-Quadrado Escalonada Inversa para os parâmetros ω^2, $\sigma^2_{\theta_i}$ e $\sigma^2_{\vartheta_i}$, $i = 1, 2, 3$, e fixando Gama$(1, 0.001)$ e Gama$(1, 0.01)$ para os parâmetros φ e ϕ, respectivamente.

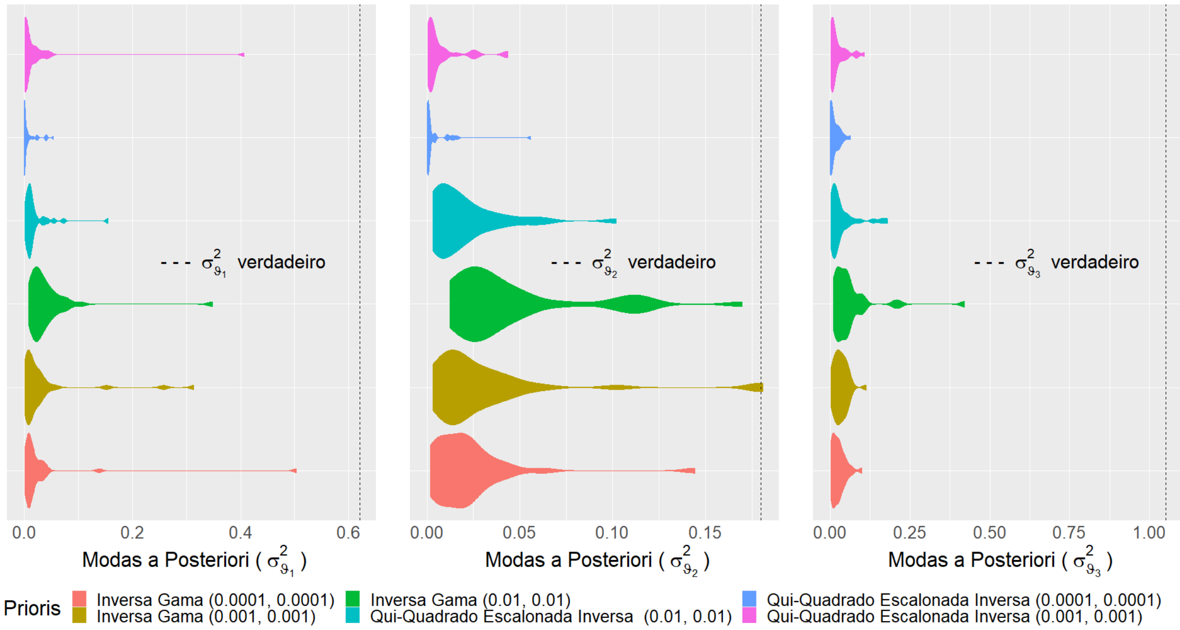

Figura 56 – Modas a posteriori das componentes de variância dos efeitos aleatórios temporais, denotados respectivamente por $\sigma^2_{\vartheta_i}$, $i = 1, 2, 3$, do modelo proposto em (2.1), tomadas ao longo dos 50 conjuntos de dados simulados quando consideradas as famílias de prioris Gama Inversa e Qui-Quadrado Escalonada Inversa para os parâmetros ω^2, $\sigma^2_{\theta_i}$ e $\sigma^2_{\vartheta_i}$, $i = 1, 2, 3$, e fixando Gama$(1, 0.001)$ e Gama$(1, 0.01)$ para os parâmetros φ e ϕ, respectivamente.

APÊNDICE C – Estudo da Sensibilidade do Número de Localizações (n) e do Número de Medidas Repetidas (J_i, $i = 1, 2, 3$)

As Figuras 57 e 58 apresentam respectivamente os comportamentos da média e da moda a posteriori para os parâmetros ω^2, φ e ϕ, obtidas ao longo dos 50 conjuntos de dados simulados quando variamos o número de localizações (n) em 62, 87 e 112, e a quantidade de meses observados (τ) em 3, 6 e 9. A configuração $n = 62$ e $\tau = 3$ não aparece nos gráficos dos parâmetros em estudo, pois neste caso as cadeias não convergiam, mesmo após tentativas de mudanças nos valores iniciais no processo de estimação, não sendo possível portanto extrair amostras independentes das distribuições a posteriori, o que indica um certo limite mínimo de informação necessária ao processo de estimação. É possível observar ainda que ω^2, φ e ϕ apresentam, em todos os cenários, as médias e modas a posteriori mais concentradas em torno dos respectivos verdadeiros valores e, neste caso, o acréscimo em qualquer uma das dimensões avaliadas, n e/ou τ, acarreta em uma concentração ainda maior das estimativas em torno dos respectivos valores verdadeiros com redução da variabilidade nas amostras a posteriori.

Para as médias a posteriori dos parâmetros $\sigma^2_{\theta_i}$ e $\sigma^2_{\vartheta_i}$, $i = 1, 2, 3$, que representam respectivamente as componentes de variância dos efeitos espaciais e temporais, ilustrados nas Figuras 59 e 60, nessa ordem, observamos um comportamento assimétrico das estimativas que, em geral, estão inflacionadas quando comparadas aos respectivos verdadeiros valores. Neste caso, notamos ainda que aparentemente o aumento em qualquer uma das dimensões (n e/ou J_i, $i = 1, 2, 3$) não impacta diretamente em um aumento de precisão para o processo de estimação. No mais, esse mesmo comportamento observado para a média a posteriori, descrito pelo impacto das estimativas sobre o processo de estimação, é presenciado para a moda a posteriori, ilustrado nas Figuras 61 e 62, em que é possível notar ainda um achatamento das estimativas nas proximidades do zero.

As Figuras 63 e 64 apresentam respectivamente o comportamento da média e da moda a posteriori para os parâmetros ω^2, φ e ϕ, do modelo proposto em (2.1), tomadas ao longo dos 50 conjuntos de dados gerados de forma independente considerando $J_1 = 12$, $J_2 = J_3 = 9$, $n = 87$, $\tau = 6$ e os demais parâmetros tomados a partir da Tabela 2. Para tanto, variamos o número de medidas repetidas em cada um dos blocos presentes no modelo respeitando a variação de $+$ ou $-$ 3 unidades em torno da dimensão dos dados reais ($J_1 = 9$, $J_2 = J_3 = 6$). Como podemos observar, de modo geral, fixando o número de

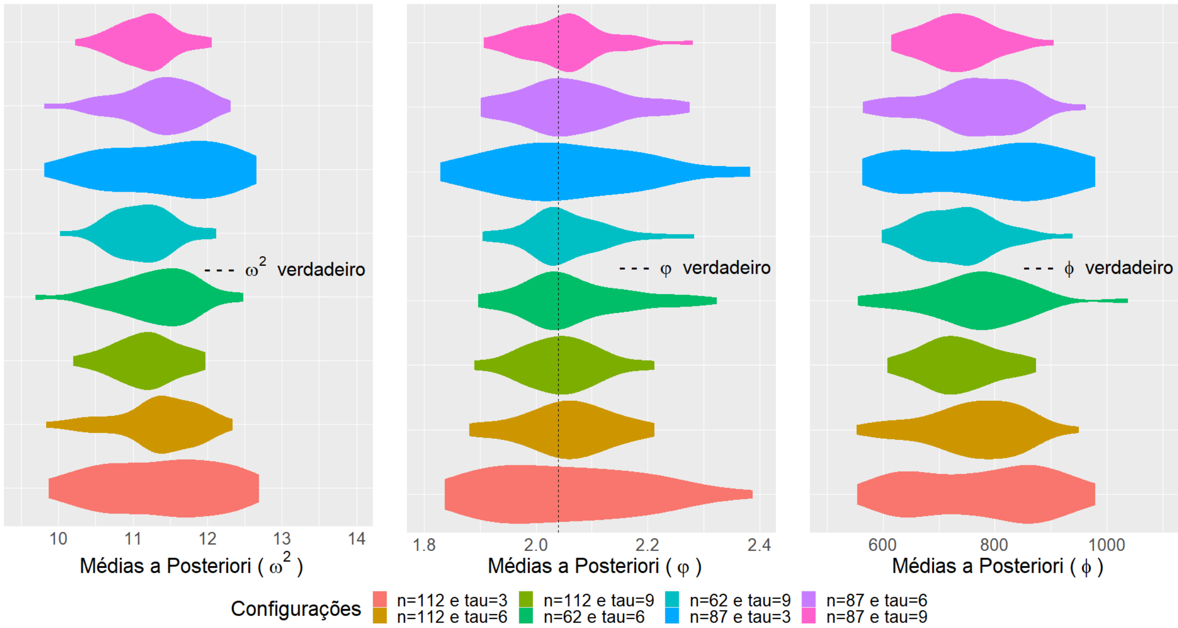

Figura 57 – Médias a posteriori para os parâmetros ω^2, φ e ϕ, tomadas ao longo dos 50 conjuntos de dados simulados independentemente variando o número de localizações (n) entre 62, 87 e 112, e o número de períodos observados (τ) entre 3, 6 e 9.

medidas em qualquer uma das dimensões do modelo proposto, ou seja J_1 ou $J_2 = J_3$, o incremento na dimensão restante acarreta em uma concentração das médias e/ou modas a posteriori em torno dos respectivos valores verdadeiros.

Quanto as médias a posteriori para os parâmetros $\sigma^2_{\theta_i}$ e $\sigma^2_{\vartheta_i}$, $i = 1, 2, 3$, que representam as componentes de variância dos efeitos espaciais e temporais, respectivamente, ilustrados reciprocamente nas Figuras 65 e 66, podemos observar que, de modo geral, o incremento no número de medidas repetidas não acarreta em melhoria da precisão para o processo de estimação. Essa mesma característica pode ser observada para moda a posteriori apresentada nas Figuras 67 e 68.

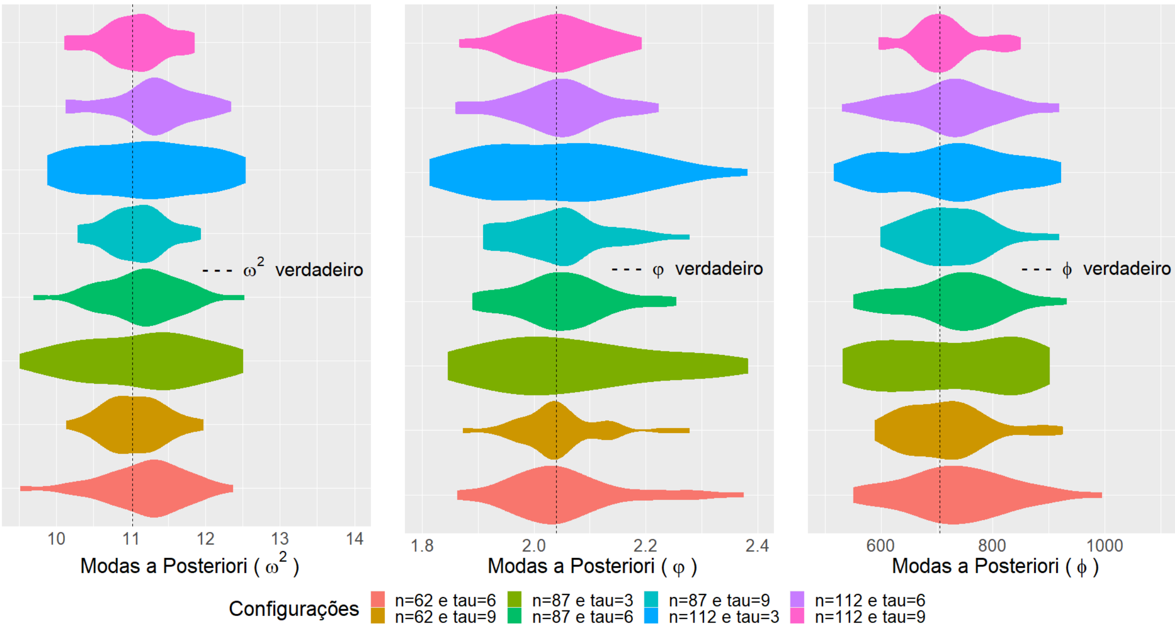

Figura 58 – Modas a posteriori para os parâmetros ω^2, φ e ϕ, tomadas ao longo dos 50 conjuntos de dados simulados independentemente variando o número de localizações (n) entre 62, 87 e 112, e o número de períodos observados (τ) entre 3, 6 e 9.

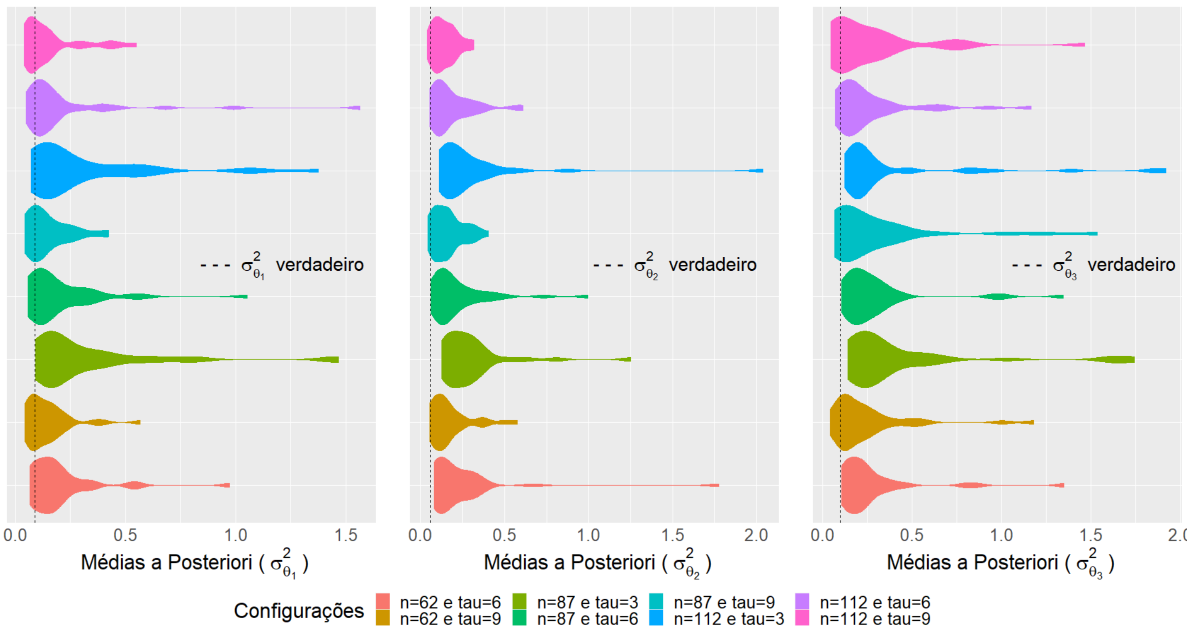

Figura 59 – Médias a posteriori para os parâmetros $\sigma^2_{\theta_i}$, $i = 1, 2, 3$, tomadas ao longo dos 50 conjuntos de dados simulados independentemente variando o número de localizações (n) entre 62, 87 e 112, e o número de períodos observados (τ) entre 3, 6 e 9.

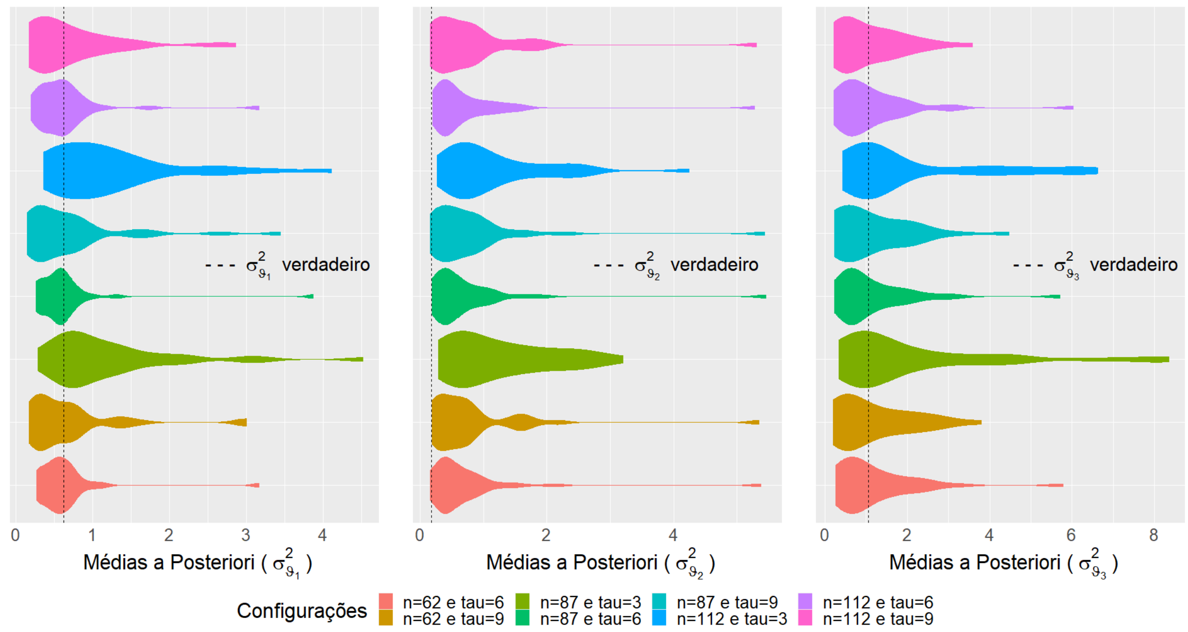

Figura 60 – Médias a posteriori para os parâmetros $\sigma^2_{\vartheta_i}$, $i = 1, 2, 3$, tomadas ao longo dos 50 conjuntos de dados simulados independentemente variando o número de localizações (n) entre 62, 87 e 112, e o número de períodos observados (τ) entre 3, 6 e 9.

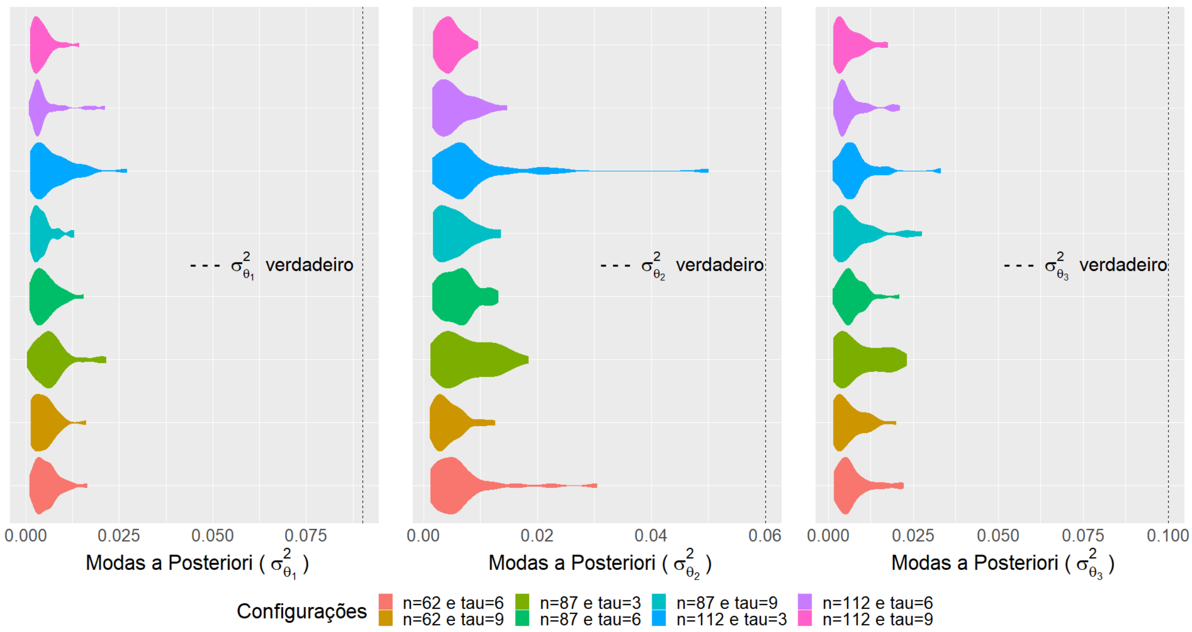

Figura 61 – Modas a posteriori para os parâmetros $\sigma^2_{\theta_i}$, $i = 1, 2, 3$, tomadas ao longo dos 50 conjuntos de dados simulados independentemente variando o número de localizações (n) entre 62, 87 e 112, e o número de períodos observados (τ) entre 3, 6 e 9.

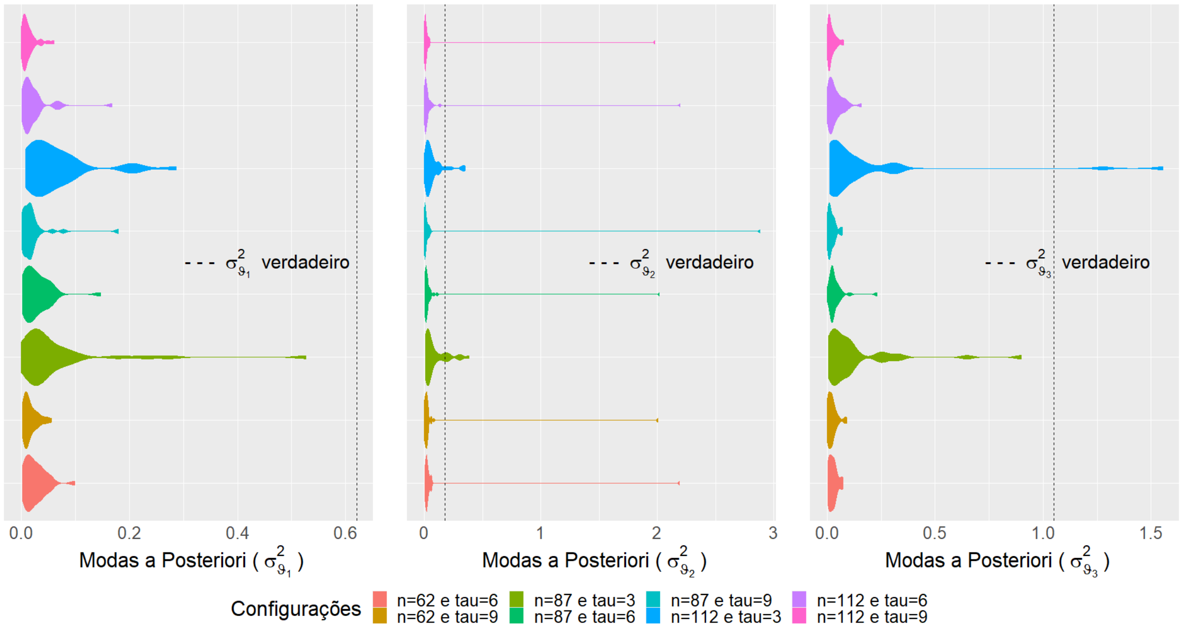

Figura 62 – Modas a posteriori para os parâmetros $\sigma^2_{\vartheta_i}$, $i = 1, 2, 3$, tomadas ao longo dos 50 conjuntos de dados simulados independentemente variando o número de localizações (n) entre 62, 87 e 112, e o número de períodos observados (τ) entre 3, 6 e 9.

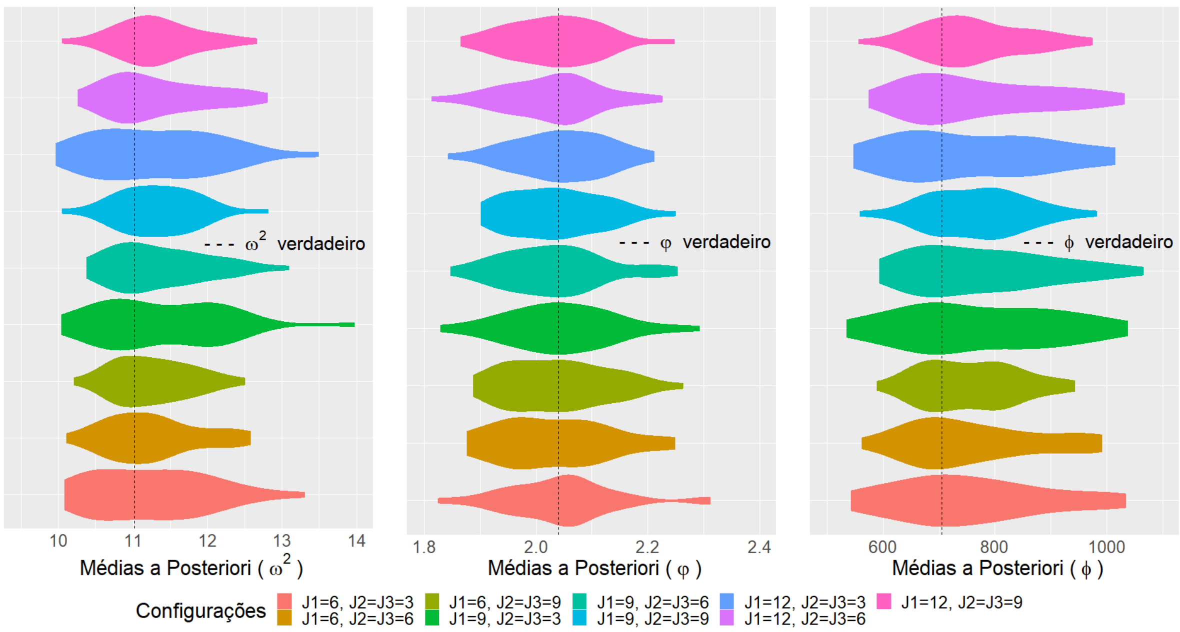

Figura 63 – Médias a posteriori para os parâmetros ω^2, φ e ϕ, tomadas ao longo dos 50 conjuntos de dados simulados independentemente variando o número de repetidas por agrupamento, J_i, $i = 1, 2, 3$.

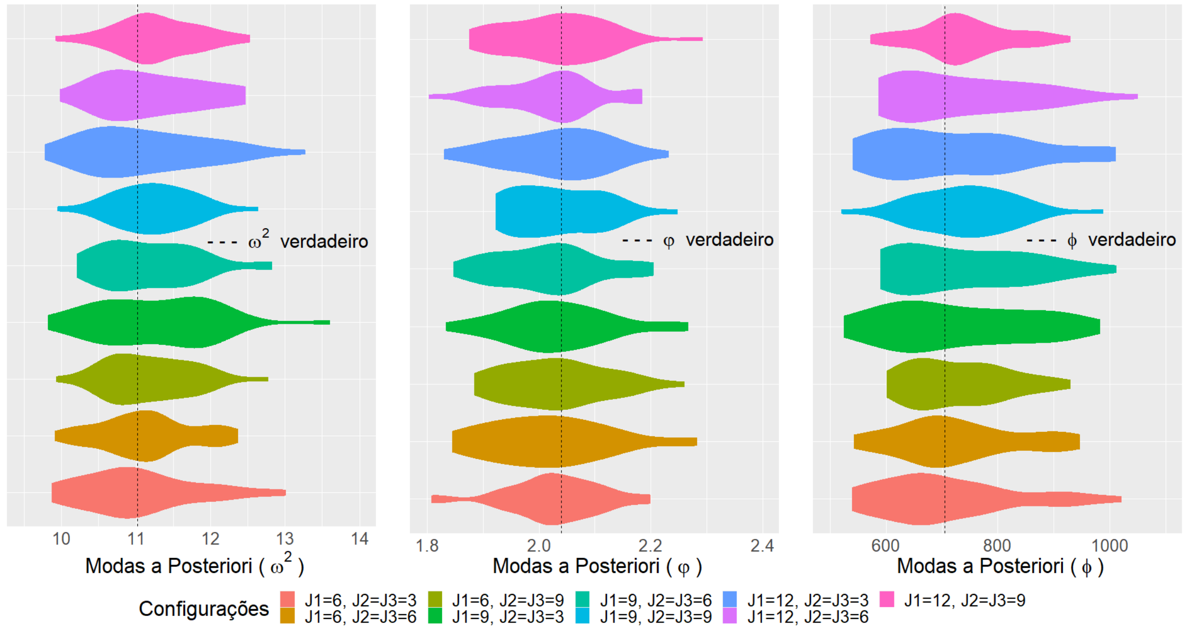

Figura 64 – Modas a posteriori para os parâmetros ω^2, φ e ϕ, tomadas ao longo dos 50 conjuntos de dados simulados independentemente variando o número de repetidas por agrupamento, J_i, $i = 1, 2, 3$.

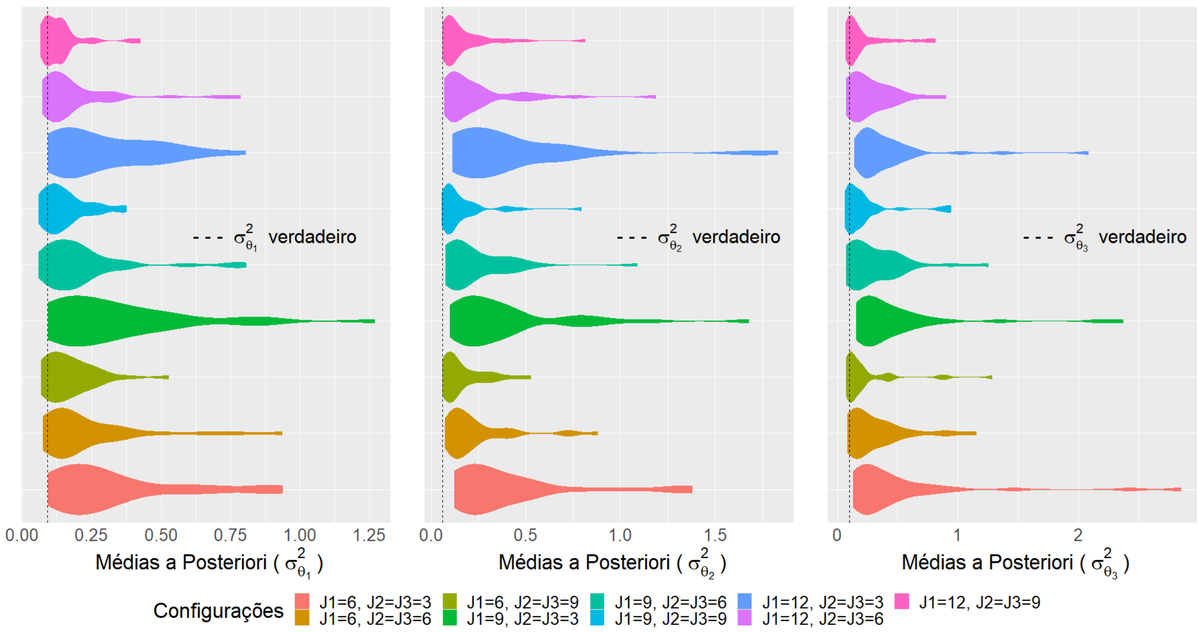

Figura 65 – Médias a posteriori para os parâmetros $\sigma^2_{\vartheta_i}$, $i = 1, 2, 3$, tomadas ao longo dos 50 conjuntos de dados simulados independentemente variando o número de repetidas por agrupamento, J_i, $i = 1, 2, 3$.

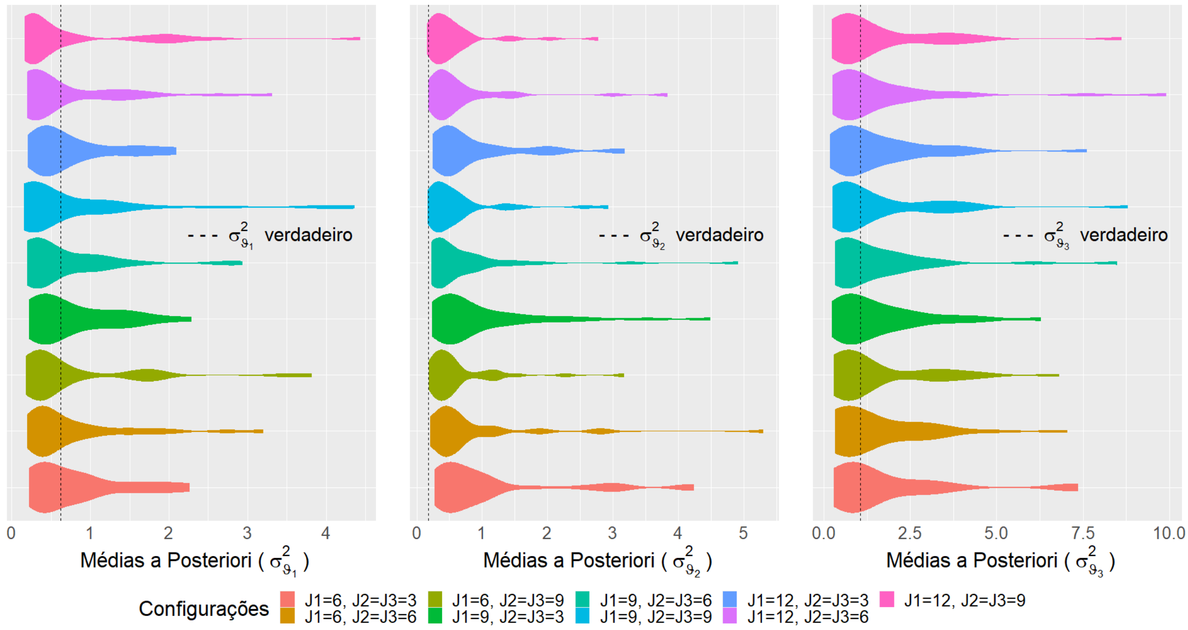

Figura 66 – Médias a posteriori para os parâmetros $\sigma^2_{\vartheta_i}$, $i = 1, 2, 3$, tomadas ao longo dos 50 conjuntos de dados simulados independentemente variando o número de repetidas por agrupamento, J_i, $i = 1, 2, 3$.

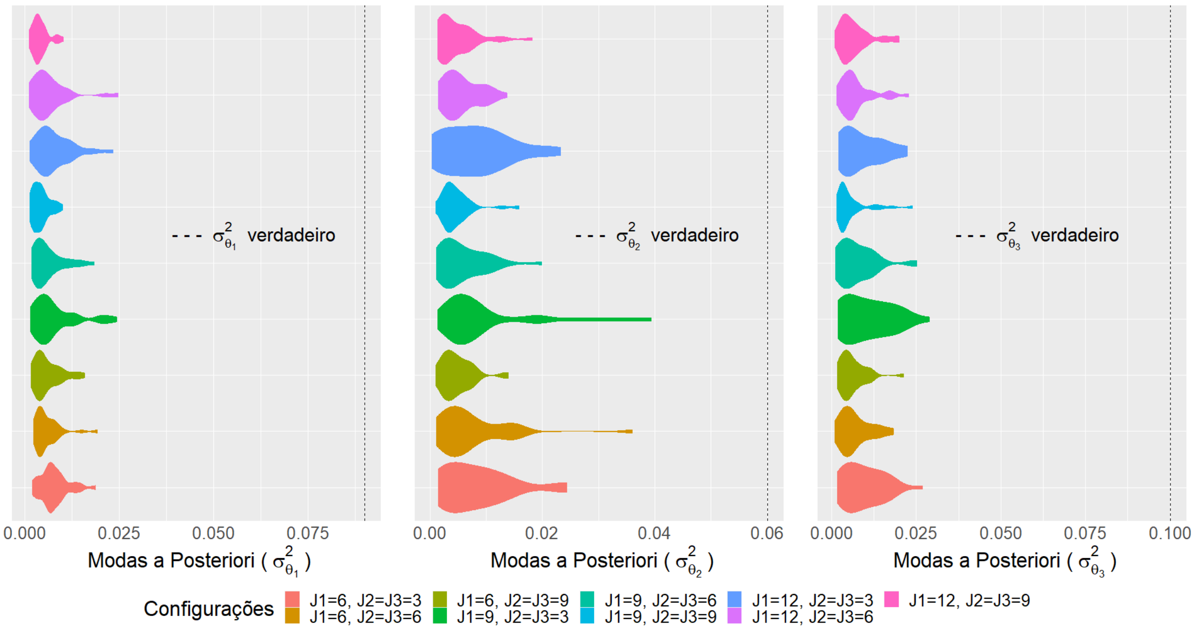

Figura 67 – Modas a posteriori para os parâmetros $\sigma^2_{\vartheta_i}$, $i = 1, 2, 3$, tomadas ao longo dos 50 conjuntos de dados simulados independentemente variando o número de repetidas por agrupamento, J_i, $i = 1, 2, 3$.

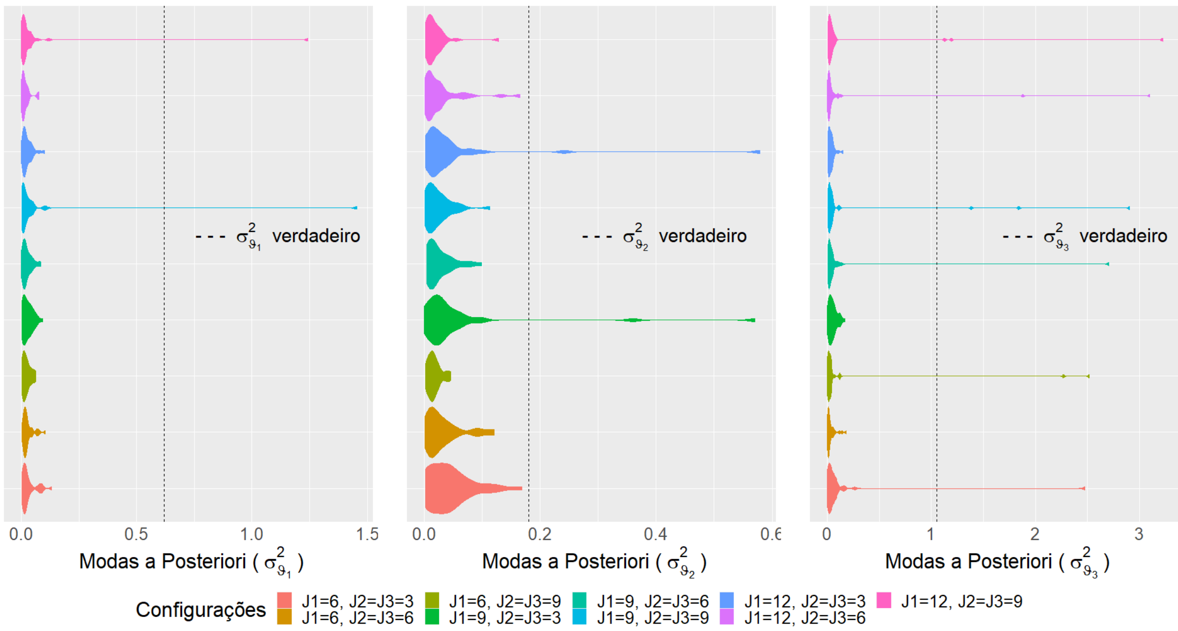

Figura 68 – Modas a posteriori para os parâmetros $\sigma^2_{\vartheta_i}$, $i = 1, 2, 3$, tomadas ao longo dos 50 conjuntos de dados simulados independentemente variando o número de repetidas por agrupamento, J_i, $i = 1, 2, 3$.

APÊNDICE D – Projeção em Espaços Ortogonais

As Figuras 69 e 70 apresentam respectivamente as estimativas da média e da moda a posteriori para o modelo na forma projetada, proposta em (5.3), obtidas ao longo dos diferentes cenários gerados por meio da variação do número de localizações (n) e do número de medidas repetidas (J). Como podemos observar, para o parâmetro ω^2 o aumento de n e/ou J resulta em melhores estimativas tanto da média quanto da moda a posteriori, uma vez que esse aumento favorece a concentração das estimativas nas proximidades dos verdadeiros valores dos respectivos parâmetros. Já para os parâmetros σ^2_θ e σ^2_ϑ notamos que o aumento das componentes n e/ou J, de modo geral, não acarretam melhoras nas estimativas da média e da moda a posteriori. Resultados análogos são encontrados para a média a posteriori (Figura 71) e para a moda a posteriori (Figura 72) quando acionamos o modelo na forma usual, proposta em (2.9).

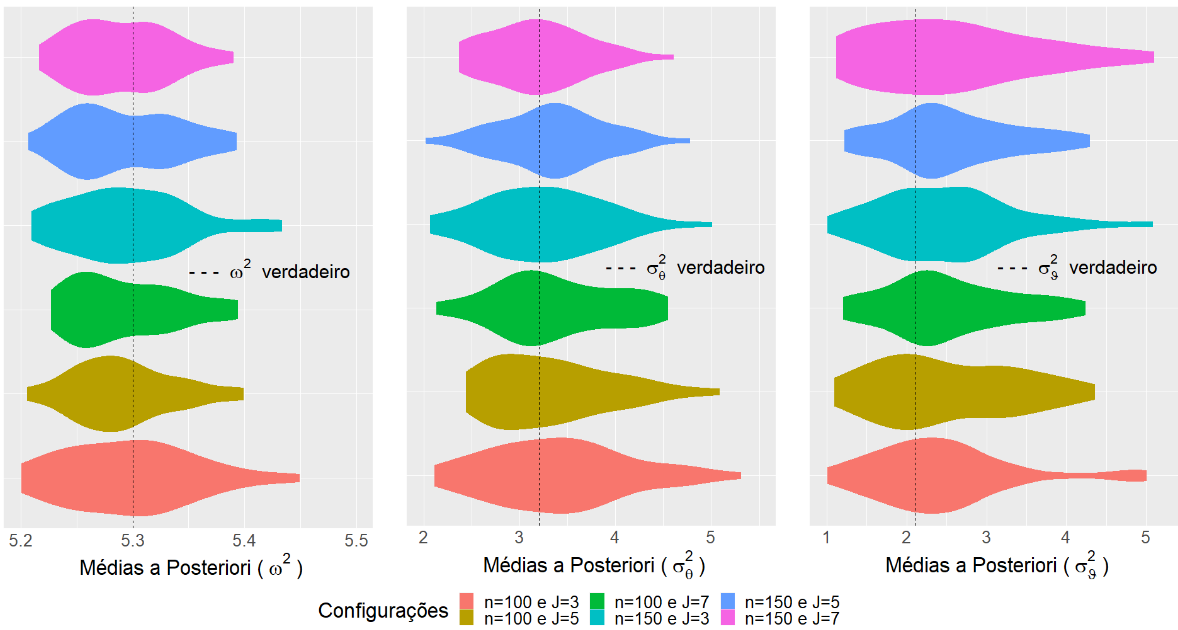

Figura 69 – Médias a posteriori dos parâmetro ω^2, σ^2_θ e σ^2_ϑ do modelo apresentado na forma projetada (proposta (5.3)), obtidas ao longo dos 50 conjuntos de dados simulados independentemente variando o número de localizações (n) e do número de medidas repetidas (J).

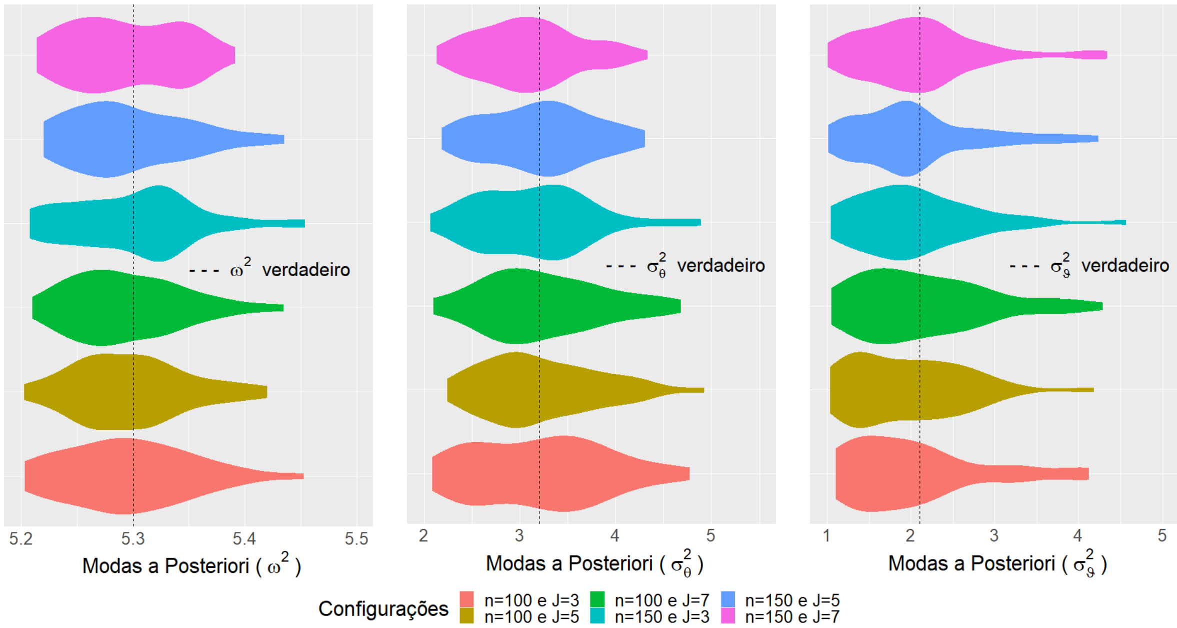

Figura 70 – Modas a posteriori dos parâmetros ω^2, σ_θ^2 e σ_ϑ^2 do modelo apresentado na forma projetada (proposta (5.3)), obtidas ao longo dos 50 conjuntos de dados simulados independentemente variando o número de localizações (n) e do número de medidas repetidas (J).

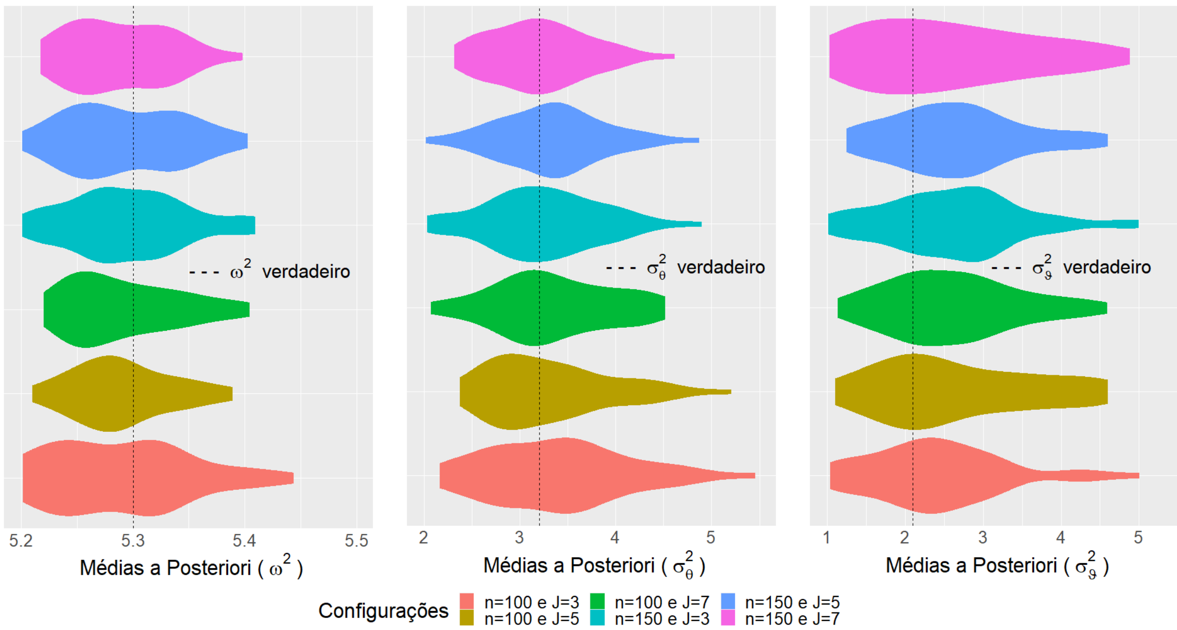

Figura 71 – Médias a posteriori dos parâmetros ω^2, σ_θ^2 e σ_ϑ^2 do modelo apresentado na forma usual (proposta em (2.9)), obtidas ao longo dos 50 conjuntos de dados simulados independentemente variando o número de localizações (n) e do número de medidas repetidas (J).

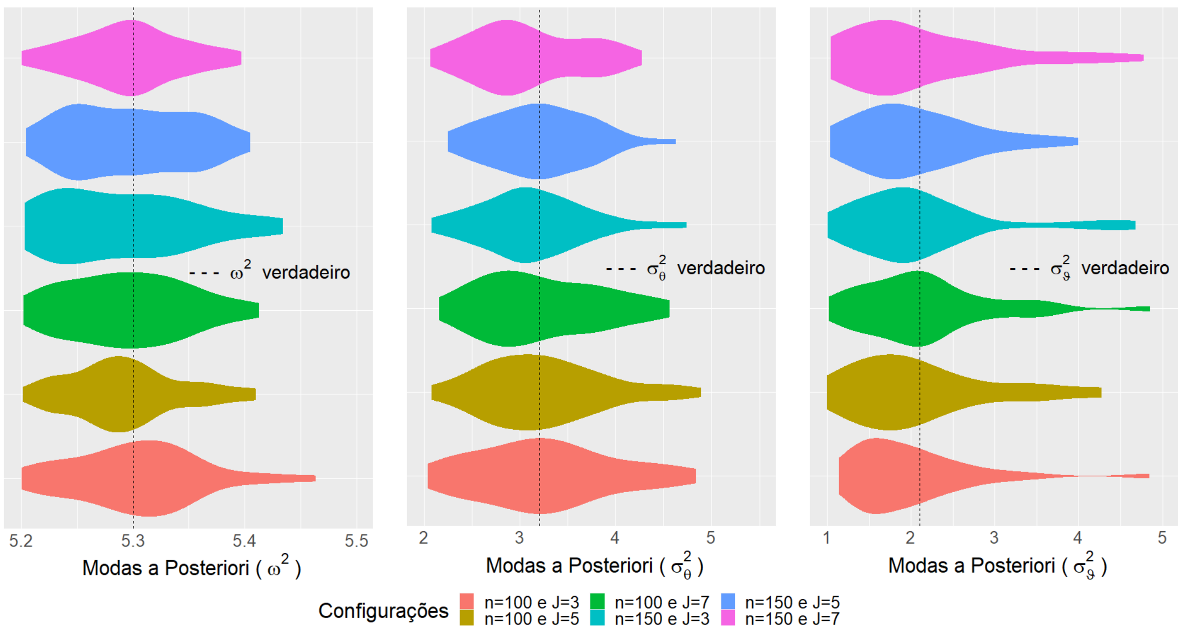

Figura 72 – Modas a posteriori dos parâmetros ω^2, σ^2_θ e σ^2_ϑ do modelo apresentado na forma usual (proposta em (2.9)), obtidas ao longo dos 50 conjuntos de dados simulados independentemente variando o número de localizações (n) e do número de medidas repetidas (J).

APÊNDICE E – Ajuste do Modelo aos Dados de Precipitação no Estado de Goiás Considerando o Estudo da Seção 6.2.1

Na Tabela 23 apresentamos os limites superiores dos intervalos de confiança da estatística de Gelman e Rubin para os parâmetros do modelo proposto em (2.1), quando aplicado aos dados de precipitação no Estado de Goiás, segundo os diferentes cenários apresentados na Tabela 18. Como podemos observar, temos indícios da convergência das cadeias de Markov uma vez que os limites superiores das estimativas se encontram próximas a 1.

Tabela 23 – Limites Superiores (L.S.) dos intervalos de confiança da estatística de Gelman e Rubin para os parâmetros do modelo proposto em (2.1), quando aplicado aos dados de precipitação no Estado de Goiás, segundo os diferentes cenários apresentados na Tabela 18.

		Estatística de Gelman e Rubin		
		Cenário 1	Cenário 2	Cenário 3
Parâmetros	ω^2	1.01	1.09	1.00
	$\sigma^2_{\theta_1}$	1.00	1.04	1.02
	$\sigma^2_{\theta_2}$	1.00	1.01	1.01
	$\sigma^2_{\theta_3}$	1.12	1.00	1.03
	$\sigma^2_{\vartheta_1}$	1.03	1.00	1.00
	$\sigma^2_{\vartheta_2}$	1.06	1.01	1.02
	$\sigma^2_{\vartheta_3}$	1.00	1.00	1.01
	φ	1.02	1.01	1.01
	ϕ	1.01	1.07	1.00

Nas Figuras 73 a 74 apresentamos os gráficos de auto-correlação para as duas cadeias de Markov geradas, para cada um dos parâmetros que compõem o modelo proposto em (2.1), quando ajustado aos dados de precipitação no Estado de Goiás, para os cenários descritos na Tabela 18. Como podemos observar, temos indícios de que o espaçamento adotado, 150 passos, foi suficiente para garantir correlação zero entre os termos de uma mesma cadeia de Markov.

Na Tabela 24 apresentamos as estimativas a posteriori dos parâmetros que compõem o modelo proposto em (2.1): média (Me) e moda (Mo), obtidas a partir da mistura das duas cadeias de Markov geradas para cada um dos parâmetros do modelo proposto, segundo os cenários apresentados na Tabela 18. Como podemos observar, as componentes de variância dos efeitos aleatórios espaciais e temporais, designados por $\sigma^2_{\theta_i}$ e

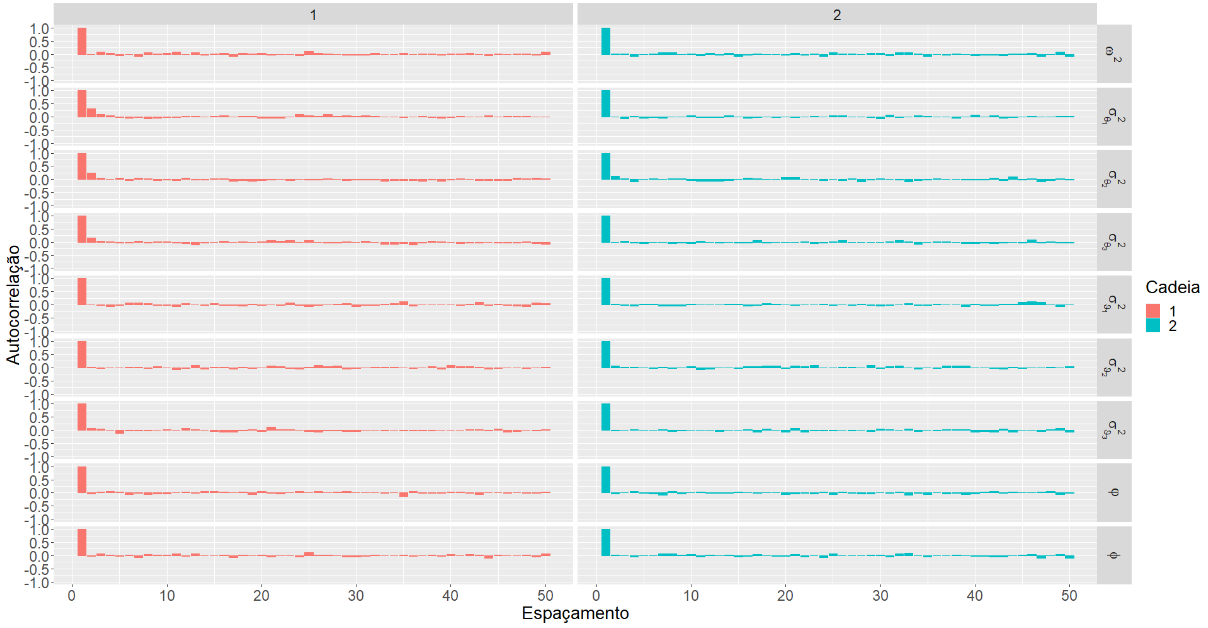

Figura 73 – Autocorrelação nas duas cadeias de Markov geradas, para cada um dos parâmetros que compõem o modelo proposto em (2.1), quando ajustado aos dados de precipitação no Estado de Goiás, segundo Cenário 1.

$\sigma^2_{\vartheta_i}$, $i = 1, 2, 3$, respectivamente, são as mais impactadas por alterações no número de bases utilizadas. Notamos ainda um grave achatamento das estimativas moda (Mo) a posteriori na vizinhança do zero.

Tabela 24 – Estimativas a posteriori dos parâmetros que compõem o modelo proposto em (2.1): média (Me) e moda (Mo), obtidas a partir da mistura das duas cadeias de Markov geradas para cada um dos parâmetros do modelo proposto, segundo os cenários apresentados na Tabela 18.

		Estatísticas a Posteriori					
		Cenário 1		Cenário 2		Cenário 3	
		Me	Mo	Me	Mo	Me	Mo
Parâmetros	ω^2	11.062	11.109	11.022	10.986	11.019	10.893
	$\sigma^2_{\theta_1}$	0.056	0.003	0.043	0.003	0.046	0.002
	$\sigma^2_{\theta_2}$	0.036	0.002	0.048	0.002	0.050	0.002
	$\sigma^2_{\theta_3}$	0.046	0.003	0.059	0.003	0.056	0.003
	$\sigma^2_{\vartheta_1}$	0.383	0.017	0.378	0.011	0.461	0.018
	$\sigma^2_{\vartheta_2}$	0.106	0.003	0.123	0.005	0.119	0.003
	$\sigma^2_{\vartheta_3}$	0.649	0.019	0.650	0.029	0.640	0.013
	φ	2.038	2.027	2.033	2.023	2.035	2.016
	ϕ	712.486	683.595	707.277	680.586	706.063	691.738

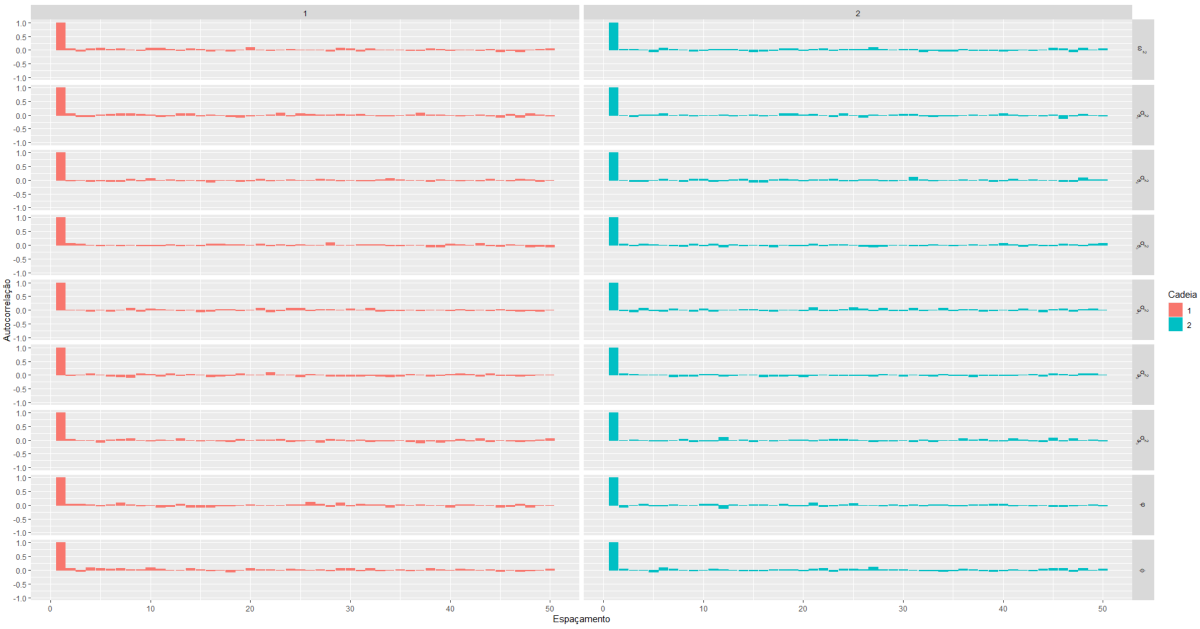

Figura 74 – Autocorrelação nas duas cadeias de Markov geradas, para cada um dos parâmetros que compõem o modelo proposto em (2.1), quando ajustado aos dados de precipitação no Estado de Goiás, segundo Cenário 2.