# Modelos multivariados espaço-temporais com covariâncias variando no tempo

Lucas Moura Faria e Silva

Universidade Federal do Rio de Janeiro
Instituto de Matemática

Rio de Janeiro
2023

LUCAS MOURA FARIA E SILVA

# Modelos multivariados espaço-temporais com covariâncias variando no tempo

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

Orientadora: Profa. Dra. Thaís Cristina Oliveira da Fonseca

Rio de Janeiro

2023

LUCAS MOURA FARIA E SILVA

# Modelos multivariados espaço-temporais com covariâncias variando no tempo

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

Aprovado em ___ de _____ de _____

BANCA EXAMINADORA:

---

Thaís Cristina Oliveira da Fonseca
Doutorado (DME/IM - UFRJ)

---

Kelly Cristina Mota Gonçalves
Doutorado (DME/IM - UFRJ)

---

Marcos Oliveira Prates
Doutorado (Dpto. Estatística - UFMG)

CIP - Catalogação na Publicação

S586m

Silva, Lucas Moura Faria e
Modelos multivariados espaço-temporais com covariâncias variando no tempo / Lucas Moura Faria e Silva. -- Rio de Janeiro, 2023.
80 f.

Orientadora: Thaís Cristina Oliveira da Fonseca.
Dissertação (mestrado) - Universidade Federal do Rio de Janeiro, Instituto de Matemática, Programa de Pós-Graduação em Estatística, 2023.

1. modelos espaço-temporais multivariados. 2. modelos de mistura. 3. abordagem condicional. 4. processo Dirichlet. 5. esparsidade. I. Fonseca, Thaís Cristina Oliveira da, orient. II. Título.

Elaborado pelo Sistema de Geração Automática da UFRJ com os dados fornecidos pelo(a) autor(a), sob a responsabilidade de Miguel Romeu Amorim Neto - CRB-7/6283.

## RESUMO

Recentemente, modelos espaço-temporais multivariados têm atraído o interesse de muitos pesquisadores para modelagem de dados por acomodarem estruturas de dependência complexas, sendo aplicados em diversas áreas como ciências ambientais, ecologia, epidemiologia, entre outras. Estes modelos buscam descrever a relação existente num processo considerando a dependência das variáveis entre si, em diferentes locais e ao longo do tempo. Em muitos casos, a aplicação de tais modelos é dificultada pelo alto custo computacional envolvido na sua implementação, além da necessidade da especificação das covariâncias cruzadas, que pode ser uma tarefa desafiadora. Para lidar com isso, uma abordagem condicional foi considerada para a especificação de um processo Gaussiano multivariado. Esta abordagem especifica a distribuição conjunta de um processo espacial através de um produtório de condicionais. Esparsidade pode ser obtida na matriz de correlação limitando o número de condicionantes em cada fator. Neste trabalho, foi proposto um modelo de mistura que inclui os modelos especificados sob diferentes organizações das variáveis condicionantes. Também foi considerado um método de seleção com base nas distribuições preditivas para obter um subconjunto de modelos e limitar o número de componentes no modelo de mistura. Neste modelo os pesos da mistura evoluem no tempo seguindo um processo Dirichlet, resultando em flexibilidade na estrutura de dependência entre as variáveis do vetor, permitindo que esta possa variar ao longo do tempo. Estudos simulados com diversos cenários de interesse são apresentados e ilustra a flexibilidade do modelo proposto. Além disso, uma aplicação em dados de poluentes atmosféricos foi realizada.



**Palavras-chave**: modelos espaço-temporais multivariados, modelos de mistura, abordagem condicional, processo Dirichlet, esparsidade.

## ABSTRACT

Recently, multivariate space-time models have attracted the interest of many researchers for data modeling because they accommodate complex dependency structures, being applied in several areas such as environmental sciences, ecology, epidemiology, among others. These models attempt to describe the existent relationship in a process considering the dependency between the variables, in different locations and through time. In many cases, the application of such models is difficult due to the high computational cost involved in their implementation and the need to specify cross-covariances, which can be challenging. To deal with it, a conditional approach was considered to define a multivariate Gaussian process. A conditional approach was considered to specify a multivariate Gaussian process. This approach specifies a joint distribution of a spatial process through a product of conditionals. Sparsity can be obtained in the correlation matrix by limiting the number of conditioning in each factor. In this work, a mixture model was proposed, it includes models specified under different organizations of conditioning variables. Also, it was considered a model selection method based on predictive distribution to obtain a smaller subset and limitate the number of components in the mixture. In this model, the mixture weights evolve through time following a Dirichlet process, resulting in flexibility in the dependency structure between the variables, allowing this to vary through time. Simulated studies with different scenarios of interest are presented and illustrate the flexibility of the proposed model. Furthermore, an application on air pollution data was performed.



**Keywords**: multivariate space-time models, mixture models, conditional approach, Dirichlet process, sparsity.

**SUMÁRIO**

# LISTA DE ILUSTRAÇÕES

## LISTA DE TABELAS

# 1 INTRODUÇÃO

O avanço na capacidade de coleta e no armazenamento de dados, tem possibilitado cada vez mais estudos de dados georreferenciados. Modelos espaço-temporais multivariados têm atraído o interesse de muitos pesquisadores para modelagem de dados desse tipo devido a sua flexibilidade e sua capacidade de acomodar estruturas de dependência complexas. Idealmente, os modelos devem considerar a variação no espaço e no tempo, além da correlação entre as variáveis, sendo aplicados em diversas áreas como ciências ambientais, ecologia, epidemiologia, entre outras. Estes modelos buscam descrever o padrão espacial e a dinâmica temporal existente num processo acomodando dependência entre as variáveis, em diferentes locais e ao longo do tempo, simultaneamente.

Mesmo sem considerar a variação temporal, determinar uma estrutura de covariância para dados espaciais multivariados pode ser uma tarefa desafiadora. Além de definir uma estrutura espacial para cada variável, ainda é necessário especificar as covariâncias cruzadas para descrever a relação espacial de cada par de componentes. Essa tarefa se torna ainda mais desafiadora quando a estrutura de dependência espacial entre as variáveis pode mudar ao longo do tempo. Por exemplo, poluentes do ar são variáveis que podem sofrer influência da direção do vento, que é um fenômeno que costuma apresentar grande variabilidade no tempo. Desta forma, em uma modelagem para dados de poluentes atmosféricos, pode ser necessário especificar diferentes covariâncias no espaço de modo que o modelo tenha flexibilidade suficiente para acomodar diferentes padrões espaciais para diferentes períodos de tempo. Para ilustrar um cenário deste tipo, a Figura 1 mostra gráficos de rosa dos ventos para os seis primeiros meses de 2023, onde podem ser visualizados medições sobre a velocidade, direção e frequência do vento, coletados em estações de monitoramento localizadas no Reino Unido, que capturam níveis de concentração de partículas no ar que auxiliam na medição da qualidade do ar. Além disso, nos mesmos locais foram capturados dados de dois poluentes: $NO_2$ e $PM_{10}$ (partículas inaláveis de diâmetro inferior a 10 micrómetros). Estes são dois dos principais poluentes atmosféricos. Na Tabela 1 podem ser vistas as correlações empíricas da média mensal desses poluentes.

Pela Figura 1, nota-se que o comportamento do vento ao longos dos meses apresentou bastante variação e o mesmo foi observado na correlação entre os poluentes. Nos três primeiros meses, os ventos foram mais intensos direcionados com mais frequência a oeste e sudoeste. Além disso, como pode ser visto na Tabela 1, nesses meses foi foram observadas a correlação mais forte e a mais fraca, de modo que foram o período onde há

| Janeiro | Fevereiro | Março | Abril | Maio | Junho |
|---|---|---|---|---|---|
| 0.5197 | 0.7188 | 0.3511 | 0.4040 | 0.4408 | 0.3984 |

Tabela 1 – Correlação empírica da média mensal da concentração de $NO_2$ e $PM_{10}$ no ar.

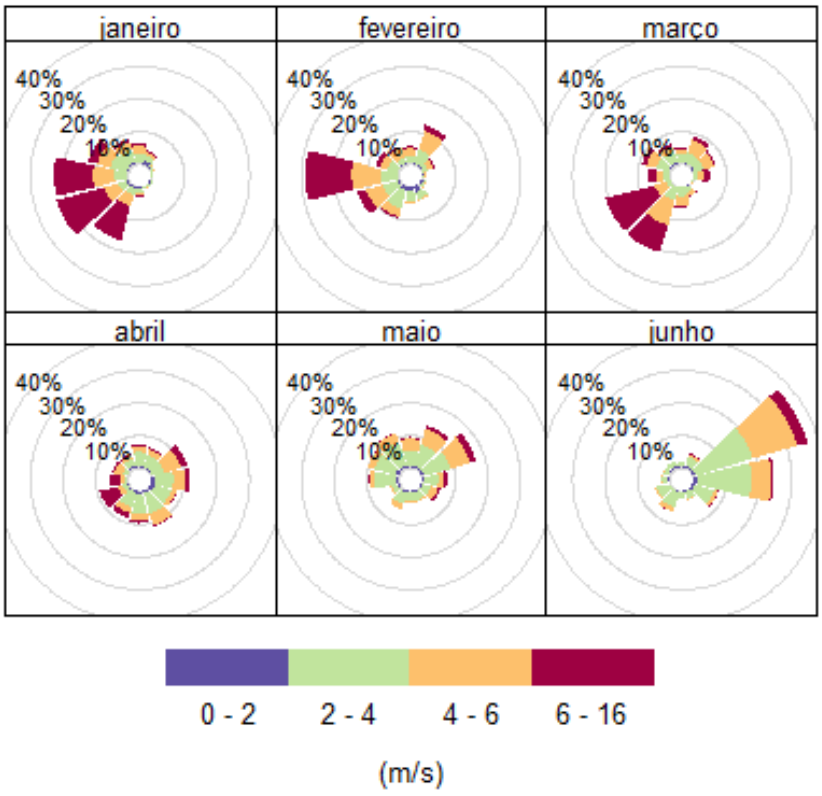



Figura 1 – Gráficos das rosas dos ventos coletados em estações de monitoramento da qualidade do ar no Reino Unido durante os seis primeiros meses de 2023.

mais variabilidade na relação entre os poluentes. Nos meses seguintes, a intensidade do vento diminuiu e passou a tomar direções mais variadas, exceto em junho. Com isso, as correlações foram mais constantes ao longo destes períodos.

Na literatura, existem algumas propostas para especificação das funções de covariância para dados multivariados. Uma proposta muito utilizada assume separabilidade das funções de covariância (MARDIA; GOODALL, 1993). Essa especificação é muito restritiva, pois impõe simetria na covariância cruzada e assume que todos os componentes modelados possuem o mesmo alcance espacial. Dependendo dos dados analisados, esta hipótese é totalmente inviável, como é o caso para dados de poluentes que são influenciados pelo vento, que foi mostrado no exemplo anterior. Outra abordagem amplamente utilizada é o Modelo Linear de Corregionalização (GOULARD; VOLTZ, 1992; WACKERNAGEL, 2003). Essa abordagem consiste em decompor o processo espacial em várias componentes espaciais não correlacionas. Isso garante diferentes alcances espaciais para cada componente de interesse, mas ainda mantém a simetria da covariância cruzada. Existem ainda proposta que são capazes de que promover assimetria na covariância cruzada (SAIN; CRESSIE, 2007; SAIN; FURRER; CRESSIE, 2011; MARTINEZ-BENEITO, 2013). Mais recentemente, Ribeiro, Junior e Bonat (2022) propuseram uma nova abordagem para especificação de covariâncias para modelagem multivariada dentro de um espaço contínuo. Esta proposta

determina uma estrutura a partir das especificações marginais de cada componente, o que fornece flexibilidade para a especificação da estrutura de covariância espaço-componentes. Essa abordagem foi desenvolvida sob a abordagem clássica e não considera a variação no tempo.

No contexto espaço-temporal, é possível observar assimetria no tempo (STEIN, 2005). Fonseca e Steel (2011) propuseram uma nova classe de funções de covariância não separáveis baseadas na convolução de funções de covariâncias válidas puramente espacial e puramente temporal. Erbisti (2019) propôs uma extensão dessa classe, para modelagem de componentes e dependência espacial, se baseando na convolução de funções de covariância separáveis e permite diferentes alcances no espaço.

Cressie e Zammit-Mangion (2016) desenvolveram uma abordagem condicional para a construção de estruturas de covariância para dados espaciais multivariados. Esta abordagem especifica a distribuição conjunta de um processo espacial através de um produtório de condicionais. Uma das vantagens dessa abordagem, é que ela permite determinar toda estrutura de covariância apenas especificando funções de covariâncias univariadas, que são funções conhecidas e discutidas na literatura (CRESSIE, 1993; STEIN, 1999; DIGGLE; JR, 2007; CHILES; DELFINER, 2012), e funções de interação que descrevem a relação entre as componentes. O artigo demonstra a generalidade dessa abordagem, mostrando sua conexão com a regressão e com modelos multivariados definidos por redes espaciais, além de toda a flexibilidade obtida ao definir diferentes funções de interação. Uma questão que existe ao utilizar a abordagem condicional é sobre como organizar as condicionais de forma a obter a melhor estrutura de relações entre as componentes para especificar o modelo, uma vez que essa especificação não é única.

Neste trabalho, foi considerado a abordagem condicional proposta por Cressie e Zammit-Mangion (2016) para especificação da função de covariância conjunta. Em geral, o custo computacional envolvido na implementação de modelos espaço-temporais multivariados é bastante alta. Desta forma, a abordagem condicional pode ajudar a aliviar o custo computacional envolvido na implementação do modelo diminuindo a ordem das matrizes de covariância. Além disso, esparsidade pode ser obtida na matriz de covariância limitando o número de condicionantes em cada fator, de forma similar ao que foi realizado em Stein, Chi e Welty (2004) e Vecchia (1988) para aproximação das funções de verossimilhança, porém ao invés de limitar as condicionais do vetor espacial, a ideia é limitar as condicionais das componentes do processo espacial. Isso reduz o número de funções e parâmetros necessários para especificar a estrutura de covariância do modelo pela abordagem condicional. Neste caso, o produtório de condicionais pode ser organizado de diferentes formas, resultando em modelos com diferentes conjuntos de dependências entre as componentes.

O objetivo do trabalho é propor um modelo multivariado espaço-temporal que possua flexibilidade para acomodar diferentes estruturas de dependência no espaço em diferentes períodos do tempo, sob uma abordagem Bayesiana. Para isso, desenvolvemos um modelo

de mistura que considera modelos especificados sob a abordagem condicional com diferentes organizações das condicionais. Nós assumimos que os pesos da mistura podem variar ao longo do tempo, de acordo com um processo Dirichlet (FONSECA; FERREIRA, 2017; SARTÓRIO; FONSECA, 2020). Essa especificação garante mais flexibilidade, permitindo que diferentes modelos e, portanto, diferentes estruturas de correlação possam ser selecionadas ao longo do tempo e, pelo processo Dirichlet, permitindo que essas mudanças ocorram de forma suave no tempo. Além disso, para limitar o número de modelos a serem considerados no modelo de mistura, consideramos um método de seleção proposto por Madigan e Raftery (1994) que se baseia nas distribuições preditivas para selecionar um subconjunto menor de modelos antes da modelagem.

O trabalho está organizado da seguinte forma. No Capítulo 2, a especificação e os métodos utilizado para a construção do modelo de mistura são descritos. No Capítulo 3, é apresentado um estudo de simulação realizada para comparar o modelo completo que considera todas as relações entre os componentes e um modelo simplificado que induz esparsidade. No Capítulo 4, outro estudo de simulação é apresentado com o intuito de avaliar o modelo de mistura proposto no trabalho. No Capítulo 5, uma aplicação em dados de poluentes atmosféricos são apresentados. No Capítulo 6, se encontra a conclusão e discussão dos resultados apresentados no trabalho.

# 2 MODELO PROPOSTO

De acordo com Cressie (1993), a estatística espacial pode ser dividida em três grandes áreas: geoestatística, dados de área e processos pontuais. Neste trabalho, foi considerado a geoestatística, que é a área utilizada para o tratamento de dados georreferenciados. Desta forma, os dados representam a mensuração de uma ou mais variáveis observadas em localizações fixa no espaço. Dentro desse contexto, a especificação da estrutura de covariância é realizada definindo funções de covariância para um processo espacial $W$, geralmente um processo Gaussiano, definido dentro de uma região de interesse.

Seja $D$ o conjunto de todas as localizações possíveis dentro de uma certa região de interesse. Sejam $s_1, s_2, \ldots, s_n$ as localizações observadas em $D \subset \mathbb{R}^d$, isto é, $s_i \in D, i = 1, 2, \ldots, n$. Assumimos $d = 2$. Vamos assumir que temos dados observados em diferentes instantes de tempo, $t = 1, 2, \ldots T$. Suponha que os dados observados são multivariados: $\boldsymbol{Y}^{(t)}(s_i) = \left(Y_1^{(t)}(s_i), Y_2^{(t)}(s_i), \ldots, Y_m^{(t)}(s_i)\right)'$, $s_i \in D$, $i = 1, 2, \ldots, n$ e $t = 1, 2, \ldots T$. Seja $\{W_j(s) : s \in D, j = 1, \ldots, m\}$ o processo espacial latente que descreve a estrutura espacial dos dados. Assim, para quaisquer número de pontos observado em $D$, a distribuição de $W_j$ é uma normal multivariada com vetor de médias definido por uma função de média $\mu(\cdot)$ e uma matriz de covariância determinada por uma função de covariância $C(\cdot, \cdot)$ que descreve a relação espacial do processo. Essa função de covariância precisa ser uma função definida positiva para ser capaz de construir uma matriz de covariância válida para o processo. Além disso, geralmente, a função de covariância adota uma relação em que a dependência espacial entre localizações mais próximas é mais fortes do que a dependência entre localizações mais distantes. Por fim, $\boldsymbol{y}$ denota todas as componentes do vetor das variáveis respostas que foram observadas.

Para a especificação geral do modelo de regressão para dados multivariados espaço-temporais, foi considerado uma estrutura hierárquica seguindo a descrição dada em Banerjee, Carlin e Gelfand (2014). Essa especificação traz flexibilidade ao modelo e torna possível incorporar diferentes fontes de informação e variabilidade para descrever os dados de interesse. O modelo hierárquico segue a construção:

$$[Dados|Processo, Parâmetros][Processo|Parâmetros][Parâmetros]$$

Deste modo, a equação geral do modelo é dada por:

$$\boldsymbol{Y}^{(t)}(s) = \boldsymbol{\mu}^{(t)}(s) + \boldsymbol{W}(s) + \boldsymbol{\varepsilon}^{(t)}(s), \quad \boldsymbol{\varepsilon}^{(t)}(s) \sim N_m(\boldsymbol{0}, V), \tag{2.1}$$

em que $\boldsymbol{\mu}^{(t)}(\cdot)$ é um preditor linear que pode ter índice de espaço e tempo, $\mu_j^{(t)}(\cdot) = X_j^{(t)}(\cdot)\beta_j^{(t)}, j = 1, \ldots, m$, onde $\boldsymbol{\beta}_j^{(t)}$ é um vetor de coeficientes de regressão associados a $j$-ésima variável resposta e $X_j^{(t)}(\cdot)$ são um conjunto de covariáveis que podem variar no tempo e/ou no espaço. A princípio, a dinâmica temporal dos dados é incorporada

através da média $\boldsymbol{\mu}$. $\boldsymbol{\varepsilon}^{(t)}(\cdot)$ é um vetor de ruídos branco independentes e com distribuição normal associados aos erros de medição. $V$ é uma matriz diagonal, com elementos $\sigma_j^2, j = 1, \ldots, m$.

Como consequência da estrutura hierárquica, temos que:

$$\boldsymbol{Y}^{(t)}(s_i)|\boldsymbol{\mu}^{(t)}(s_i), \boldsymbol{W}(s_i), V \sim N_m\left(\boldsymbol{\mu}^{(t)}(s_i) + \boldsymbol{W}(s_i), V\right), i = 1, \ldots, n; \tag{2.2}$$

$$\boldsymbol{W}|\boldsymbol{\theta} \sim N_{m\times n}\left(\mathbf{0}, \Sigma_W(\boldsymbol{\theta})\right) \tag{2.3}$$

Em que $\Sigma_W$ é uma matriz de covariância conjunta definida por alguma função de covariância multivariada $C_w(\boldsymbol{\theta})$ que determina tanto a relação espacial do vetor $\boldsymbol{W}_j$, quanto a interação espacial entre cada par de componentes $\boldsymbol{W}_j$ e $\boldsymbol{W}_q$. $\boldsymbol{\theta}$ é um vetor com todos os parâmetros que determinam a função de covariância $C_w$. Além disso, o vetor de médias de $\boldsymbol{W}$ é um vetor de 0's. Como $V$ é uma matriz diagonal, ao seguir a especificação hierárquica, as variáveis respostas $\boldsymbol{Y}$ são independentes quando condicionadas no processo espacial $\boldsymbol{W}$ e no preditor linear $\boldsymbol{\mu}$. Assim, a função de verossimilhança é dada por um conjunto de produtórios:

$$f\left(\boldsymbol{y}|\boldsymbol{\mu}, \boldsymbol{W}, V\right) = \prod_{j=1}^{m}\prod_{t=1}^{T}\prod_{i=1}^{n} N\left(y_j^{(t)}(s_i);\ \mu_j^{(t)}(s_i) + W_j(s_i),\ \sigma_j^2\right) \tag{2.4}$$

Em que, $N\left(y;\ \mu,\ \sigma^2\right)$ denota a função de densidade de probabilidade de uma normal com média $\mu$ e variância $\sigma^2$.

## 2.1 REGRESSÃO DINÂMICA

Na regressão especificada em (2.1), toda a dinâmica temporal é incorporada na modelagem através do preditor linear $\boldsymbol{\mu}^{(t)}(\cdot) = \left(\boldsymbol{\mu}_1^{(t)}(\cdot), \ldots, \boldsymbol{\mu}_m^{(t)}(\cdot)\right)'$. Além disso, as covariáveis também são incluídas no modelo através dessa componente, onde $\boldsymbol{\mu}_j^{(t)}(\cdot) = X_j^{(t)}(\cdot)\boldsymbol{\beta}_j^{(t)}, j = 1, \ldots, m$ . Assim, para estimar o preditor linear é necessário apenas estimar os coeficientes de regressão $\boldsymbol{\beta}_j$. Para fazer isso, vamos utilizar os métodos de Modelos Lineares Dinâmicos (WEST; HARRISON, 1997). Esta é uma classe de modelos usualmente utilizadas na modelagem de séries temporais que possuem bastante flexibilidade, sendo capazes de modelar todas as componentes de uma série temporal, como tendência, sazonalidade e de modelar séries não estacionárias. Em particular, neste trabalho estamos considerando a especificação geral de uma regressão dinâmica:

$$\boldsymbol{Y}_{j,*}^{(t)} = X_j^{(t)}\boldsymbol{\beta}_j^{(t)} + v_t, \quad v_t \sim N(0, V_t) \tag{2.5}$$

$$\boldsymbol{\beta}_j^{(t)} = G_t\boldsymbol{\beta}_j^{(t-1)} + w_t, \quad w_t \sim N(0, W_t) \tag{2.6}$$

Para $j = 1, \ldots, m$.

A Equação (2.5) é chamada de equação observacional e a Equação (2.6) é chamada de equação de sistema. A primeira equação como dados evoluem ao longo do tempo em

função do conjunto de covariáveis e estados latentes que são os coeficientes de regressão, além de um choque aleatório $v_t$ com distribuição normal independente. A segunda equação descreve a dinâmica temporal dos coeficientes de regressão, através da matriz $G_t$, que é matriz de evolução do sistema e um erro aleatório $w_t$ também com distribuição normal. Em particular, para o nosso modelo, temos que $\boldsymbol{Y}_{j,*}^{(t)} = \boldsymbol{Y}_j^{(t)} - \boldsymbol{W}_j$.

Para a estimação dos coeficientes de regressão, nós utilizamos o algoritmo FFBS (*Forward Filter Backwards Sampler*), que é uma combinação do filtro de Kalman com amostragem restrospectiva. O filtro de Kalman é um algoritmo recursivo usado para obter as distribuições de filtragem, que é a distribuição de $\boldsymbol{\beta}_j^{(t)}$ condicionada a informação disponível até o período $t$, que vamos denotar por $D_t$. A amostragem retrospectiva é outro algoritmo recursivo que obtem e simula da distribuição $\boldsymbol{\beta}_j^{(t)}$ condicionada a toda informação disponível, $D_T$. Além disso, especificamos a matriz $W_t$ por meio de um fator de desconto $\delta$, que determina a taxa de informação perdida ao passar do período $t-1$ ao período $t$. Desta forma, o fator de desconto consegue controlar o nível de flexibilidade da curva ajustada. Se $\delta = 1$, temos um modelo constante e quando $\delta = 0$, temos um modelo sem perdas de informação. Para mais detalhes sobre o filtro de Kalman, algoritmo FFBS e fator de desconto, veja (WEST; HARRISON, 1997; PETRIS; PETRONE; CAMPAGNOLI, 2009).

## 2.2 ESPECIFICAÇÃO DO PROCESSO ESPACIAL PELA ABORDAGEM CONDICIONAL

Para completar a especificação do processo espacial, é preciso determinar uma função de covariância multivariada para definir $\Sigma_W(\boldsymbol{\theta})$. Existem algumas propostas na disponíveis literatura, desde modelos mais simples e restritos que supõe separabilidade de $\Sigma_W$ (MARDIA; GOODALL, 1993) a modelos mais complexos e flexíveis, que são capazes de promover assimetria na covariância cruzada (SAIN; CRESSIE, 2007; SAIN; FURRER; CRESSIE, 2011; MARTINEZ-BENEITO, 2013). Neste trabalho, foi utilizado a metodologia proposta por Cressie e Zammit-Mangion (2016), que usa a abordagem condicional para representar a distribuição conjunta. Nesta abordagem, a distribuição conjunta é representada por um produtório de condicionais:

$$[\boldsymbol{W}_1(\cdot), \ldots, \boldsymbol{W}_m(\cdot)] = [\boldsymbol{W}_1(\cdot)] \prod_{q=2}^{m} [\boldsymbol{W}_q(\cdot) | \boldsymbol{W}_1(\cdot), \ldots, \boldsymbol{W}_{q-1}(\cdot)] \tag{2.7}$$

Neste caso, a matriz de covariância conjunta $\Sigma_W(\boldsymbol{\theta})$ é construída de forma recursiva, através da especificiação de um conjunto de funções de covariância univariadas e um conjunto de funções $b(\cdot,\cdot)$ que descrevem a relação existente entre cada par de componentes do processo espacial. $b(\cdot,\cdot)$ pode ser qualquer função integrável tal que $b : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ e aqui é chamada de função de interação. Deste modo, ao invés de considerar uma única matriz de covariância de ordem $m \times n$, se considera $m$ matrizes de ordem $n$, o que é

menos custoso computacionalmente. Além disso, uma outra vantagem dessa abordagem é que podemos aproveitar a literatura existente sobre as funções de covariância univariadas (STEIN, 1999; BANERJEE; CARLIN; GELFAND, 2014; CHILES; DELFINER, 2012). Essa abordagem fornece bastante flexibilidade para construir toda a estrutura de correlação do processo espacial e (CRESSIE; ZAMMIT-MANGION, 2016) mostram que essa abordagem constói uma estrutura válida para processos multivariados.

Suponha um processo Gaussiano multivariado $\boldsymbol{W}$ com $m$ componentes, como definido anteriormente. Defina $C_{11}(s,u) = Cov(W_1(s), W_1(u))$ a função de covariância marginal de $W_1$. Assuma que $C_{11}(s,u)$ é uma função de covariância válida. Para as definições a seguir nesta seção, assuma $q = 2, \ldots, m$. Agora, defina:

$$E\left[W_1(s)\right] = 0; \tag{2.8}$$

$$E\left[W_q(s)|W_1(s), \ldots, W_{q-1}(s)\right] = \sum_{r=1}^{q-1} \int_D b_{qr}(s,v) W_r(v) dv, \tag{2.9}$$

onde $s \in D$ e $b_{qr}(\cdot,\cdot)$ é a função de interação que descreve a relação entre as componentes $W_q(s)$ e $W_r(s)$ $(r < q)$. Portanto, a média condicional da componente $W_q(s)$ depende das componentes condicionantes através das funções de interação. Além disso, se $b_{qr} = 0$, como resultado $\boldsymbol{W}_q$ e $\boldsymbol{W}_r$ são independentes. Também são definidas funções de covariância para as covariâncias condicionais:

$$Cov\left(W_q(s), W_q(u)|W_1(\cdot), \ldots, W_{q-1}(\cdot)\right) = C_{q|r<q}(s,u), \quad s,u \in D. \tag{2.10}$$

A partir das especificações das funções de interação e das funções de covariância marginal $C_{11}$ e condicionais $C_{q|r<q}$, podemos obter as funções de covariância marginais e funções de covariância cruzada recursivamente da seguinte forma:

$$C_{qq}(s,u) = C_{q|r<q}(s,u) + \sum_{r=1}^{q-1} \sum_{r'=1}^{q-1} \int_D \int_D b_{qr}(s,v) C_{rr'}(v,w) b_{qr'}(u,w) dv dw; \tag{2.11}$$

$$C_{rq}(s,u) = Cov(W_r(s), W_q(u)) = \sum_{r'=1}^{q-1} \int_D b_{qr'}(u,w) C_{rr'}(s,w) dw, \quad r < q, s, u \in D. \tag{2.12}$$

Onde $C_{qq}$ denota a função de covariância marginal para $W_q$ e $C_{rq}$ denota a função de covariância cruzada, para $r < q$. Então, para construir a matriz de covariância conjunta do processo espacial multivariado $\{W_j(s) : s \in D, j = 1, \ldots, m\}$, é preciso especificar as funções de interação $(b_{qr}, r < q)$, a função de covariância marginal de $W_1$ $(C_{11})$ e as funções de covariância condicional $(C_{q|r<q})$. Em todos os casos, são funções de covariância univariadas.

Nas Equações (2.9), (2.11) e (2.12), dependendo da especificação das funções de interação e das funções de covariância, não há uma solução analítica para a integral. Porém, podemos assumir que $D$ é um conjunto finito, compondo um grid regular em uma certa

região de interesse contínua no espaço. Neste caso, as integrais das Equações (2.9), (2.11) e (2.12) podem ser aproximadas por somatórios em cima do grid construído. Um exemplo dessa aproximação para a construção da matriz de covariância conjunta para um cenário com duas variáveis pode ser visto em (CRESSIE; ZAMMIT-MANGION, 2016).

### 2.2.1 Funções de interação e Covariância

Cressie e Zammit-Mangion (2016) utilizam três funções de interação diferentes em seu artigo. Uma função que induz independência, fixando $b_{qr} = 0, \forall, q, r$, outra que induz uma dependência pontual, onde a relação entre $\boldsymbol{W}_q$ e $\boldsymbol{W}_r$ existe apenas em dados nas mesmas localizações, e uma função com dependência difusa, onde a relação das componentes é difundida no espaço. Neste trabalho, foi considerado apenas essa última, que é definida como:

$$b(h) = \begin{cases} A\left[1 - \left(\frac{||h-\Delta||}{r}\right)^2\right]^2 & , \text{ se} ||h-\Delta|| \leq r, \\ 0 & , \text{ caso contrário} \end{cases}$$

Em que $h = s - u$, $s, u \in D$. O parâmetro $\Delta$ incorpora assimetria na relação das componentes. No caso de um espaço bidimensional, $d = 2$, $\boldsymbol{\Delta} = (\Delta_1, \Delta_2)$, com um elemento associado a cada coordenada geográfica, determinando uma direção onde a correlação entre as componentes é mais forte. O parâmetro $A$ é um parâmetro de escala que controla o quão forte é a relação entre as componentes e pode, inclusive, assumir valores negativos. $r$ é chamado de parâmetro de abertura. Ele assume apenas valores positivos e determina o quanto a relação se difunde em relação a direção determinada por $\boldsymbol{\Delta}$. Valores baixos de $r$ geram relações muito restritas a direção $\boldsymbol{\Delta}$, enquanto valores mais altos fornecem uma relação mais suave em torno de $\boldsymbol{\Delta}$. Se fixarmos $r$ em um valor menor do que a distância mínima entre pontos, a função de interação com dependência difusa é simplificada para a função de interação com dependência pontual.

Para especificar as funções de covariância univariadas existem diversas possibilidades, como a função exponencial, Gaussiana, Matern, entre outros (BANERJEE; CARLIN; GELFAND, 2014). Neste trabalho, foi utilizado apenas a função Matern, que é uma função fortemente recomendada por Stein (1999), por ser uma função flexível, incluindo um parâmetro que controla o nível de suavidade do processo espacial. A função Matern é definida da seguinte forma:

$$C(h) = \frac{\phi \pi^{0.5}}{2^{\nu-1}\Gamma(\nu + 0.5)\alpha^{2\nu}}(\alpha|h|)^{\nu}\kappa_{\nu}(\alpha|h|) \tag{2.13}$$

em que, $\kappa_\nu$ é a função Bessel modificada de segundo tipo. $\phi$ é um parâmetro de escala e $\alpha$ é um parâmetro de alcance que determina a distância máxima onde ainda há correlação espacial significativa; $\nu$ é o parâmetro que controla o nível de suavidade do processo espacial. Temos um processo mais suave para valores maiores de $\nu$. Se fixarmos $\nu = 0.5$,

a função Matern é simplificada para função exponencial, e se $\nu \to \infty$, é simplificada para a função Gaussiana.

### 2.2.2 Esparsidade

Para um vetor de processos espaciais de tamanho $m$, é necessário especificar $m(m+1)/2$ funções de interação e de covariância. Então, o número de funções necessárias para determinar toda a estrutura de covariância dos processos espaciais cresce rapidamente em função de $m$, assim como o número de parâmetros que precisaria ser estimado pelo modelo. Assim, para valores altos de $m$, pode ser inviável definir tantas funções para a especificação do modelo, onde , inclusive, podemos ter questões de identificabilidade. Para lidar com este cenário, neste trabalho propomos induzir esparsidade dentro da estrutura de covariância cruzada entre as componentes do processo espacial, seguindo um caminho similar ao que foi proposto em Stein, Chi e Welty (2004) e Vecchia (1988).

Diferente deles que utilizam essa ideia para aproximar a função de verossimilhança de dados espaciais, como estamos seguindo uma abordagem hierárquica para especificação do modelo de regressão, essa esparsidade deve ser induzida diretamente no processo espacial como ocorre em Datta et al. (2016) para limitar o número máximo de condicionantes de cada condicional. No caso de Datta et al. (2016) a esparsidade é induzida na estrutura espacial, enquanto aqui a ideia é induzir a esparsidade entre as componentes do processo espacial. Seja $\Pi_q$ um subconjunto de $\{W_1(\cdot), \ldots, W_{q-1}(\cdot)\}$ com no máximo $L$ elementos. Assim, a Equação 2.7 é reescrita como:

$$[W_1(\cdot), \ldots, W_m(\cdot)] = [W_1(\cdot)] \prod_{q=2}^{m} [W_q(\cdot)|\Pi_q] \tag{2.14}$$

Esta configuração nos permite aproximar a distribuição conjunta através de um produtório de condicionais sem perder muita informação da estrutura geral da correlação entre as componentes do processo espacial. Essencialmente, ao induzir esparsidade desta forma, estamos estabelecendo que algumas componentes são condicionalmente independentes. Além disso, o número de funções e de parâmetros do modelo é reduzido consideravelmente quando o valor de $m$ é alto, como pode ser visto na Tabela 2.

| | Número de componentes | | | | | |
|---|---|---|---|---|---|---|
| Limite de condicionantes | 3 | 4 | 5 | 6 | 8 | 10 |
| Sem limite | 21 | 36 | 55 | 78 | 136 | 210 |
| 2 | 21 | 32 | 43 | 54 | 76 | 98 |
| 1 | 17 | 24 | 31 | 38 | 52 | 66 |

Tabela 2 – Número de parâmetros necessários para determinar a estrutura de correlação seguindo a abordagem condicional, considerando as funções Matern e de interação com dependência difusa, com $d = 2$.

Stein, Chi e Welty (2004) e Vecchia (1988) mostraram que a aproximação era melhor quando se utiliza os dados mais próximos (ou menos distantes) geograficamente nas condicionais. No caso deste trabalho, como essa esparsidade não é induzida no vetor espacial de cada componente, mas na interação entre elas, essa noção natural de distância se perde. Uma alternativa, é considerar alguma outra medida de distância para tentar identificar quais componentes estão mais próximas entre si, como por exemplo, a própria correlação poderia ser utilizada, Mahalanobis, Camberra, entre outras (COX; COX, 2000).

## 2.3 MODELO DE MISTURA

Uma questão que existe com respeito a abordagem condicional é sobre como organizar as condicionais. De fato, a Equação (2.7) não é única e para valores altos de $m$ pode ser organizada de muitas maneiras distintas, principalmente se for considerado a possibilidade da existência de independências condicionais. Cressie e Zammit-Mangion (2016) argumenta que uma forma de tentar lidar com isso é considerar a relação causal e que na maior parte das aplicações essa organização pode ser realizada com base na natureza das variáveis. O problema é que nem sempre pode estar totalmente claro na aplicação e também exige que o pesquisador possua algum grau de conhecimento sobre os dados, o que nem sempre é o caso.

Quando a esparsidade é introduzida limitando o número máximo de condicionantes nas condicionais da abordagem condicional, diferentes estruturas de correlação são automaticamente determinadas reorganizando as condicionais. Por exemplo, para um modelo com um processo espacial de três componentes: $\boldsymbol{W}_1$, $\boldsymbol{W}_2$ e $\boldsymbol{W}_3$ e limitando em uma condicionante no máximo, há três formas distintas de estabelecer relação entre os processos: $[\boldsymbol{W}_1][\boldsymbol{W}_2|\boldsymbol{W}_1][\boldsymbol{W}_3|\boldsymbol{W}_1]$, $[\boldsymbol{W}_1][\boldsymbol{W}_2|\boldsymbol{W}_1][\boldsymbol{W}_3|\boldsymbol{W}_1]$ ou $[\boldsymbol{W}_1][\boldsymbol{W}_3|\boldsymbol{W}_1][\boldsymbol{W}_2|\boldsymbol{W}_3]$. Note que cada configuração corta a relação entre um par de componentes, isto é, estabelece um par de componentes que são condicionalmente independentes. No primeiro, ocorre com $\boldsymbol{W}_2$ e $\boldsymbol{W}_3$, no segundo entre $\boldsymbol{W}_1$ e $\boldsymbol{W}_3$ e no último entre $\boldsymbol{W}_1$ e $\boldsymbol{W}_2$. Qualquer outra organização das condicionais leva a uma estrutura equivalente a um dessas três configurações. Essas relações podem ser representadas através de grafos acíclicos direcionados, como na Figura 2, em que, se não há uma flecha conectando dois nós associados às componentes, então essas componentes são condicionalmente independentes.

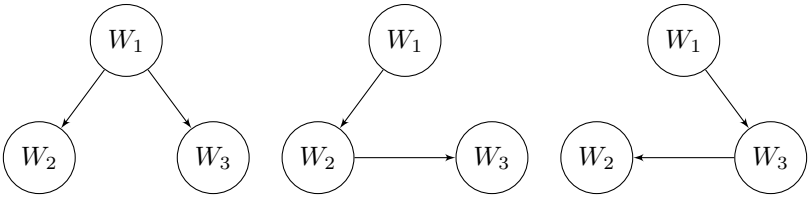



Figura 2 – Grafos direcionados representando a relação condicional entre as componentes do processo espacial.

Neste trabalho, propomos considerar as diferentes estruturas de correlações geradas pela indução de esparsidade e pelas reorganizações das condicionais em um modelo de mistura, onde cada modelo é especificado com uma das diferentes estrutura de correlações. Assumimos que os pesos da mistura podem variar ao longo do tempo. Desta forma, o modelo de mistura proporciona flexibilidade para acomodar diferentes estruturas de dependência no espaço ao longo do tempo. Seja $K$ o número de modelos considerados nessa mistura e $\boldsymbol{\pi} = \{\boldsymbol{\pi}_1, \boldsymbol{\pi}_2, \ldots, \boldsymbol{\pi}_K\}$ uma lista indexando as diferentes estruturas correlações da mistura determinadas pelas condicionais. Nesse contexto, o valor de $K$ é conhecido. Defina $\boldsymbol{\eta}^{(t)} = \left(\eta_1^{(t)}, \ldots, \eta_K^{(t)}\right)'$ os pesos da mistura associados a cada modelo. Assim, para um período $t$, temos que:

$$f\left(\boldsymbol{y}_t|\boldsymbol{\mu}_t, \boldsymbol{W}, V, \boldsymbol{\eta}^{(t)}\right) = \sum_{k=1}^{K} \eta_k^{(t)} f_k\left(\boldsymbol{y}_t|\boldsymbol{\mu}_t, \boldsymbol{W}^{(k)}, V\right), \tag{2.15}$$

$$f\left(\boldsymbol{W}^{(k)}|\boldsymbol{\theta}_k\right) = f\left(\boldsymbol{W}_{\pi_k(1)}|\boldsymbol{\theta}_k\right) \times \prod_{j=2}^{m} f\left(\boldsymbol{W}_{\pi_k(j)}|\boldsymbol{\theta}_k, \Pi_{\pi_k(j-1)}\right) \tag{2.16}$$

em que $f$ denota uma função de densidade de probabilidade, $\boldsymbol{\theta}_k$ denota o vetor de hiperparâmetros associado a estrutura de correlação dos processos espaciais do $k$-ésimo modelo e $\Theta = (\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_K)'$. Podemos reescrever a expressão do modelo de mistura definido em (2.15) usando o princípio do aumento de dados para incluir artificialmente variáveis latentes $\boldsymbol{Z} = (Z_1, \ldots, Z_T)'$ tais que $P\left(Z_t = k\right) = \eta_k^{(t)}$ que indicam qual é o modelo selecionado em cada período, resultando numa nova expressão para o modelo mistura:

$$f\left(\boldsymbol{Y}_t, Z_t|\boldsymbol{\mu}_t, \boldsymbol{W}, V, \boldsymbol{\eta}^{(t)}\right) = \prod_{k=1}^{K} \left[f_k\left(\boldsymbol{y}_t|\boldsymbol{\mu}_t, \boldsymbol{W}^{(k)}, V\right) \eta_k^{(t)}\right]^{1_{\{Z_t=k\}}} \tag{2.17}$$

em que $1_A$ é a função indicadora do evento $A$. Marginalizando $Z_t$ na Equação (2.17), a relação da Equação (2.15) é recuperada.

É importante permitir que os pesos da mistura $\eta_k^{(t)}$ possam variar ao longo do tempo, pois produz uma grande flexibilidade para a estrutura de dependência entre as variáveis, permitindo que a organização das condicionais possa ser alterada ao longo do tempo. Desse modo, o modelo pode ser capaz de capturar mudanças de comportamento das variáveis que podem modificar a estrutura de correlação entre as variáveis e no espaço. Em particular, assumimos que os pesos da mistura evoluem com o tempo de acordo com um processo de evolução Dirichlet. Esse processo permite que os pesos da mistura $\eta_k^{(t)}$ possam mudar ao longo do tempo de forma suave, evitando que mudanças bruscas possam ocorrer de um período a outro.

### 2.3.1 Processo de Evolução Dirichlet

Seja $\boldsymbol{e}_k$ é um vetor com $K$ elementos tal que o $k$-ésimo elemento é igual a 1 e o resto são todos iguais a 0. Além disso, seja $1_K$ um vetor de 1's de tamanho $K$. Então, seguindo

essa notação, $\boldsymbol{e}_{Z_t}|\boldsymbol{\eta}^{(t)} \sim Multinomial\left(1_K, \boldsymbol{\eta}^{(t)}\right)$ é um vetor latente que vai auxiliar a evolução Dirichlet. Fonseca e Ferreira (2017) definem um processo de evolução Dirchlet como se segue. Seja $\boldsymbol{\psi}_t = (\psi_{t1}, \ldots, \psi_{tK})'$, vetor com elementos independentes tal que $\psi_{tk} \sim Beta\left(\delta c_{t-1,k}, (1-\delta)c_{t-1,k}\right)$, $k = 1, \ldots, K$, $\delta \in (0, 1]$ e $c_{t-1,k} > 0, \forall k$. Defina $S_t = \boldsymbol{\psi}_t' \boldsymbol{\eta}^{(t-1)}$. Então, o Processo de evolução Dirichlet é definido por:

$$\boldsymbol{\eta}^{(t)} = \frac{1}{S_t} \boldsymbol{\psi}_t \odot \boldsymbol{\eta}^{(t-1)}, \tag{2.18}$$

onde $\odot$ denota o produto elemento a elemento de um vetor.

O hiperparâmetro $\boldsymbol{c}_t = (c_{t,1}, c_{t,2} \ldots, c_{t,K})$ resume a informação que se tem sobre o peso $\boldsymbol{\eta}^{(t)}$ até o período $t$. O parâmetro $\delta$ é o fator de desconto, que determina a taxa de informação que se perde ao passar do período $t-1$ para o período $t$. Deste modo, $\delta$ indica a velocidade da mudança dos pesos $\boldsymbol{\eta}^{(t)}$ ao longo do tempo. Valores de $\delta$ menores indicam que a mudança nos pesos ocorrem mais rapidamente, enquanto valores mais altos indicam mudanças mais suaves nos valores de $\boldsymbol{\eta}^{(t)}$. Quando $\delta = 0$, os pesos passam a evoluir de forma independente e quando $\delta = 1$ os pesos se tornam constantes no tempo. Suponha que $D_0$ denota a informação *a priori* sobre $\boldsymbol{\eta}^{(0)}$ e que $D_t = \{D_{t-1}, e_{Z_t}\}$ denota a informação que se tem até o período $t$.

**Teorema 1 (Forward Filter)** *Assuma a distribuição a priori $\boldsymbol{\eta}^{(0)}|D_0 \sim Dirichlet(\boldsymbol{c}_0)$. Além disso, considere que $\boldsymbol{e}_{Z_t}|\boldsymbol{\eta}^{(t)} \sim Multinomial\left(\mathbf{1}_l, \boldsymbol{\eta}^{(t)}\right)$ e suponha o Processo de Evolução Dirichlet definido anteriormente. Então, para $t = 1, \ldots, T$, temos:*

$$\boldsymbol{\eta}^{(t-1)}|\delta, D_{t-1} \sim Dirichlet\left(\boldsymbol{c}_{t-1}\right); \tag{2.19}$$

$$\boldsymbol{\eta}^{(t)}|\delta, D_{t-1} \sim Dirichlet\left(\delta \boldsymbol{c}_{t-1}\right); \tag{2.20}$$

$$\boldsymbol{\eta}^{(t)}|\delta, D_t \sim Dirichlet\left(\boldsymbol{c}_t\right), \tag{2.21}$$

*em que $\boldsymbol{c}_t = \delta \boldsymbol{c}_{t-1} + e_{Z_t}$.*

Por esse resultado, observe que: $E\left[\boldsymbol{\eta}^{(t)}|D_{t-1}, \delta\right] = E\left[\boldsymbol{\eta}^{(t-1)}|D_{t-1}, \delta\right]$. Portanto, condicionado aos dados que se tem até o período $t-1$, o valor esperado de $\boldsymbol{\eta}^{(t)}$ e $\boldsymbol{\eta}^{(t-1)}$ são iguais.

**Teorema 2 (Backwards Sampler)** *Após a aplicação do Teorema 1, é possível amostrar diretamente de $\boldsymbol{\eta}^{(T)}|\delta, D_T \sim Dirichlet\left(\boldsymbol{c}_T\right)$, pela distribuição de filtragem. Em seguida, para $t = T, \ldots, 2$, é possível amostrar, recursivamente, de $\boldsymbol{\eta}^{(t-1)}$ condicionado em toda informação disponível $\left\{\delta, D_T, \boldsymbol{\eta}^{(t)}, S_t\right\}$ através dos seguintes passos:*

1. *Amostre: $S_t|D_T, \boldsymbol{\eta}^{(t)}, \delta \sim Beta\left(\delta \mathbf{1}_K \boldsymbol{c}_{t-1}, (1-\delta)\mathbf{1}_K \boldsymbol{c}_{t-1}\right)$;*
2. *Amostre: $u \sim Dirichlet\left((1-\delta)\boldsymbol{c}_{t-1},\right)$;*
3. *Calcule: $\boldsymbol{\eta}^{(t-1)} = S_t \boldsymbol{\eta}^{(t)} + (1 - S_t)u$.*

A aplicação conjunta dos Teoremas 1 e 2 resulta em um algoritmo FFBS que permite simular os pesos $\boldsymbol{\eta}_1, \ldots, \boldsymbol{\eta}_T$, recursivamente, de acordo com o Processo de Evolução Dirichlet. Para mais detalhes do processo Dirichlet, veja (FONSECA; FERREIRA, 2017). O fator de desconto $\delta$ pode ser estimado pelo modelo, de modo que $\delta$ indique o nível volatilidade na estrutura de correlações espaciais apresentados nos dados ou pode ser fixado. Neste trabalho, vamos trabalhar com o fator de desconto fixado.

### 2.3.2 Estimação do Modelo

Como estamos adotando uma abordagem Bayesiana neste trabalho, a estimação do modelo é realizado através da distribuição *a posteriori* conjunta de todos os parâmetros, hiperparâmetros e variáveis latentes: $P(\boldsymbol{W}, \boldsymbol{\beta}, V, \Theta, \boldsymbol{Z}, \boldsymbol{\eta}|\boldsymbol{y})$, que é calculada pelo Teorema de Bayes. Entretanto, não é possível obter essa distribuição analiticamente, devido a impossibilidade de se obter a distribuição preditiva $p(\boldsymbol{y})$. Desta forma, é necessário recorrer a métodos computacionais para obter aproximações da distribuição *a posteriori* do modelo. Para isso, foram considerados os métodos MCMC (*Markov Chain Monte Carlo*).

Os métodos MCMC são algoritmos iterativos que geram cadeias de Markov cuja distribuição estacionária é a própria distribuição *a posteriori*. As propriedades do MCMC garantem que o algoritmo, após um determinado número de iterações, converge para a distribuição estacionária, que nesse caso será a distribuição *a posteriori*, de forma que a cadeia gerada represente uma amostra gerada diretamente da distribuição *a posteriori*. Assim, combinando as propriedades de uma cadeia de Markov com os métodos de Monte Carlo para realizar aproximações de quantidades de interesse, os métodos MCMC permitem a realização de inferência acerca da distribuição *a posteriori*. Dentre os diferentes algoritmos MCMC, dois que se destacam são o algoritmo de Metropolis-Hastings e o amostrador de Gibbs (CHIB; GREENBERG, 1995; GEMAN; GEMAN, 1984).

No algoritmo MCMC desenvolvido para a implementação do modelo, foram utilizados uma combinação de passos de Gibbs com o dois algoritmos FFBS, um para os coeficientes de regressão $\boldsymbol{\beta}$ e outro para simular os pesos da mistura $\boldsymbol{\eta}$ de acordo com o processo Dirichlet. Os parâmetros $\boldsymbol{W}^{(k)}$, $\phi_j^k$, $\sigma_j^2$, $A_{jq}^{(k)}$, $Z_t$ possuem distribuições condicionais completas conhecidas, então seguimos amostrando diretamente por elas. Para os parâmetros cuja distribuições condicionais completas não eram conhecidas: $\alpha_{jq}^{(k)}$, $\nu_{jq}^{(k)}$, $r_{jq}^{(k)}$ e $\boldsymbol{\Delta}_{jq}^{(k)}$, seguindo o mesmo caminho usado em Cressie e Zammit-Mangion (2016) foi utilizado um Slice Sampling (NEAL, 2003), que tende a ser mais custoso computacionalmente do que um passo de Metropolis, mas é mais eficiente, já que nos permite gerar da distribuição condicional sem a rejeição de valores. Ainda assim, para altas dimensões, os parâmetros $\alpha_{jq}^{(k)}$, $\nu_{jq}^{(k)}$ podem ser atualizados conjuntamente por meio de um passo de Metropolis para reduzir o custo computacional envolvido no cálculo da função Matern. A seguir um pseudo-código do algoritmo MCMC explicitando o procedimento usado na implementação do modelo de mistura.

Algoritmo MCMC para implementação do modelo de mistura

1. Inicialização de todos os parâmetros;
2. Para $i = 1, \ldots, N$:
   - 2.1 Para $j = 1, \ldots, m$:
     - 2.1.1 Gera $\boldsymbol{\mu}_{j,i}$ através do algoritmo FFBS com $\boldsymbol{y}_j = \boldsymbol{Y}_j - \boldsymbol{W}_j$;
     - 2.1.2 Gera $\sigma_j^2 \sim P(\sigma_j^2|\cdot)$;
   - 2.2 Para $k = 1, \ldots, K$:
     - 2.2.1 Para $j = 1, \ldots, m$:
       - 2.2.1.1 Gera $\boldsymbol{W}_{j,i}^{(k)} \sim P\left(\boldsymbol{W}_j^{(k)}|\cdot\right)$ ;
       - 2.2.1.2 Gera $\phi_{j,i}^{(k)} \sim P\left(\phi_j^{(k)}|\cdot\right)$ ;
       - 2.2.1.3 Gera $\nu_{j,i}^{(k)}, \alpha_{j,i}^{(k)} \sim P\left(\nu_{j,i}^{(k)}, \alpha_j^{(k)}|\cdot\right)$, pelo slice sampling;
     - 2.2.2 Para $q = 2, \ldots, m$ e $j < q$:
       - 2.2.2.1 Gera $A_{jq,i}^{(k)} \sim P\left(A_{jq}^{(k)}|\cdot\right)$ ;
       - 2.2.2.2 Gera $\boldsymbol{\Delta}_{jq,i}^{(k)}, r_{jq,i}^{(k)} \sim P\left(\boldsymbol{\Delta}_{jq}^{(k)}, r_{jq}^{(k)}|\cdot\right)$, pelo slice sampling;
   - 2.3 Simula $\boldsymbol{\eta}_i^{(1)}, \ldots, \boldsymbol{\eta}_i^{(T)}$ pelo algoritmo FFBS do processo Dirichlet;
   - 2.4 Para $t = 1, \ldots, T$, gera $Z_{t,i} \sim P\left(Z_t|\cdot\right)$.

### 2.3.3 Especificações finais dos Modelos

Nesta seção, vamos completar as especificações dos modelos, atribuindo as distribuições *a priori* dos parâmetros e explicitando os diferentes modelos considerados no trabalho. Ao todo, são considerados três modelos. O primeiro é modelo sem mistura, onde o processo espacial é especificado sob abordagem condicional usando a relação da Equação (2.7), onde todas as relações cruzadas são consideradas, vamos chamar esse modelo de Modelo Completo ($MC$). O segundo é modelo de mistura com os pesos fixos no tempo, onde $\boldsymbol{\eta}^{(t)} = \boldsymbol{\eta}, \forall t$, que chamamos de Modelo de Mistura Estático ($MME$). Por fim, é modelo de mistura com os pesos evoluindo no tempo seguindo um processo Dirichlet, que vamos chamar de Modelo de Mistura Dinâmico ($MMD$).

Seja $B_{qr}$ uma matriz $n \times n$ construída pela função de interação com dependência difusa $b_{qr}$ e $C_q$ a função de covariância da $q$-ésima componente do processo espacial. Assuma que $\Pi_q \subset \{\boldsymbol{W}_1, \ldots, \boldsymbol{W}_{q-1}\}$ representa o subconjunto de condicionais em que $\boldsymbol{W}_q$ está condicionado. Então a distribuição *a priori* dos processos espaciais é dado por:

$$\boldsymbol{W}_1|\boldsymbol{\theta}_1 \sim N\left(\boldsymbol{0}, C_1\left(\boldsymbol{\theta}_1\right)\right), \quad \boldsymbol{\theta}_1 = \left(\phi_1, \nu_1, \alpha_1\right);$$

$$\boldsymbol{W}_q|\boldsymbol{\theta}_q, \Pi_q \sim N\left(\sum_{j \in \Pi_q} B_{jq}\boldsymbol{W}_j, C_q\left(\boldsymbol{\theta}_q\right)\right), \quad \boldsymbol{\theta}_q = \left(\phi_q, \nu_q, \alpha_q\right), q = 2, \ldots, m.$$

Para os Hiperparâmetro que compõem a estrutura de dependência no espaço dos

processos espaciais, as distribuições *a priori* consideradas foram:

$$\phi_q^{-1} \sim Gama(a_q, \gamma_q);\ \nu_q \sim Gama(a_{\nu_q}, \gamma_{\nu_q});\ \alpha_q \sim Gama(a_{\alpha_q}, \gamma_{\alpha_q});$$
$$A_{jq} \sim N(m_{A_{jq}}, V_{A_{jq}});\ \log(r_{jq}) \sim N(m_{r_{jq}}, V_{r_{jq}});\ \boldsymbol{\Delta}_{jq} \sim N(\boldsymbol{m}_{\Delta_{jq}}, V_{\Delta_{jq}})$$

Para os demais parâmetros, temos $\sigma_j^{-2} \sim Gama(a, b)$, como inicialização do algoritmo FFBS da regressão dinâmica, temos $\boldsymbol{\beta}_j^{(0)} \sim N(m_0, C_0)$ e para os modelos de mistura, temos $P(Z_t = k) = \eta_k^{(t)}$, com $\boldsymbol{\eta}^{(0)} \sim Dirichlet(c_0)$, como inicialização do algoritmo FFBS para o processo Dirichlet. Nos modelos de mistura, as mesmas distribuições *a priori* são consideradas para cada modelo componente da mistura. As distribuições condicionais completas são apresentadas no Apêndice A.

## 2.4 SELEÇÃO DE MODELOS

O número de modelos considerados na mistura cresce rapidamente a medida que $m$ aumenta. Então, para valores altos de $m$, o custo computacional para a implementação do modelo pode ficar muito elevado. Questões similares podem ser encontrados no contexto de *Bayesian Model Averaging* (HOETING et al., 1999). Dentro desse contexto, Madigan e Raftery (1994) desenvolveram um algoritmo de busca para selecionar um subconjuntos de modelos com base nas distribuições preditivas. Esta metodologia utiliza o princípio da navalha de Occam e propõe selecionar os modelos que produzirem as maiores distribuições preditivas ou mais próximas da maior preditiva. O método é realizado em duas etapas, onde a segunda é opcional e não foi utilizada neste trabalho. Na primeira etapa, determina-se um conjunto $M$ tal que, se a razão da maior distribuição preditiva dentre todos modelos com a preditiva do $k$-ésimo modelo $M_k$ for maior que uma determinada constante $C$, o modelo $M_k$ é descartado, pois é um modelo que fornece uma capacidade preditiva muito pior do que o "melhor" modelo. A constante $C$ é determinada pelo usuário do método.

Assim, para o procedimento de inferência do modelo de mistura, iremos considerar um procedimento em 2 passos. No passo 1, selecionamos um subconjunto de modelos com base em uma distribuição preditiva aproximada, uma vez que não conseguimos obter a distribuição preditiva analiticamente. Apenas os modelos selecionados dentro desse subconjunto são considerados no modelo de mistura. No segundo passo, condicionado a esse subconjunto de modelos, estimamos os parâmetros do modelo de mistura são estimados.

### 2.4.1 Distribuição preditiva

De forma geral, a distribuição preditiva de $\boldsymbol{y}$ é dada por:

$$p(\boldsymbol{y}) = \int_{\Theta} p(\boldsymbol{y}|\boldsymbol{\theta}) p(\boldsymbol{\theta}) d\theta \tag{2.22}$$

Em que $\boldsymbol{\theta}$, aqui, representa o conjunto dos parâmetros do modelo. No caso do modelo de mistura, como já foi comentado, não é possível obter essa distribuição de forma analítica.

Deste modo, para calcular a distribuição preditiva de forma analítica, que é necessário para realizar a seleção de modelos, consideramos uma aproximação que é uma versão simplificada do modelo, onde assumimos independência no espaço e no tempo, simultaneamente. Nesse caso, para calcular a preditiva de forma analítica, basta especificar as variâncias de $\boldsymbol{\mu}_t$, $\boldsymbol{W}$ de forma conveniente. A função de interação, é definida como uma função pontual, assim só há correlação entre os pontos de mesma localização. Desta forma, especificamos $Var(\boldsymbol{W}) = A = \sigma^2 A'$, onde $A'$ é uma matriz de covariância das componentes do vetor $Y_(s_i)^{(t)}$. Para $\boldsymbol{\mu}_t$, consideramos uma especificação semelhante: $Var(\boldsymbol{\mu}_t) = \lambda I_m = \sigma^2 \lambda I'_m$ Assim, definimos que:

$$\begin{aligned}
\boldsymbol{Y}^{(t)}(s_i)|\boldsymbol{\mu}^{(t)}(s_i), \sigma^2, \boldsymbol{W}(s_i) &\sim N_m\left(\boldsymbol{\mu}^{(t)}(s_i) + \boldsymbol{W}(s_i), \sigma^2 I_m\right); \\
\boldsymbol{W}(s_i)|\sigma^2 &\sim N_m\left(\mathbf{0}, A\right); \\
\boldsymbol{\mu}^{(t)}(s_i)|\sigma^2 &\sim N_m(\mathbf{0}, \lambda I_m), i = 1, \ldots, n; \\
\sigma^2 &\sim Ga(a_0, b_0).
\end{aligned}$$

Assim, integrando o processo espacial $\boldsymbol{W}$, a média $\boldsymbol{\mu}^{(t)}(s_i)$ e a variância $\sigma^2$, temos como resultado a distribuição preditiva de $\boldsymbol{Y}^{(t)}(s_i)$:

$$\boldsymbol{Y}^{(t)}(s_i) \sim t_{m,2a_0}\left(\mathbf{0}, \frac{\beta_0}{a_0}(\lambda' I_m + A' + I_m)\right), i = 1, \ldots, n, \tag{2.23}$$

Os hiperparâmetros $a_0$, $b_0$ e $\lambda'$ são conhecidos, portanto para conhecer a distribuição preditiva basta estimar $A'$. Para isso, utilizamos simplesmente o método Bayes Empírico e tomamos um valor de $A'$ maximizando a função de verossimilhança.

## 2.5 INTERPOLAÇÃO E PREDIÇÃO

Quando estamos trabalhando com modelos espaciais em uma superfície contínua, em muitos casos, estamos interessados em fazer uma interpolação na região de interesse, para compreender o comportamento da variável modelada ao longo de toda a superfície. Neste caso, o interesse é na predição de $\boldsymbol{y}(s_0)|\boldsymbol{y}$, para $s_0 \in D$. De forma similar a Equação (2.22), tem-se, de forma geral, que:

$$p(\boldsymbol{y}(s_0)|\boldsymbol{y}) = \int_\Theta p(\boldsymbol{y}(s_0)|\boldsymbol{\theta}) p(\boldsymbol{\theta}|\boldsymbol{y}) d\theta = E_{\boldsymbol{\theta}|\boldsymbol{y}}\left[p(\boldsymbol{y}(s_0)|\boldsymbol{\theta})\right] \tag{2.24}$$

Assim, podemos utilizar *composition sampling* para gerar valores da distribuição de $\boldsymbol{y}(s_0)|\boldsymbol{y}$, usando o resultado do MCMC utilizado para ajustar o modelo. Deste modo, se $s_0 \in \{s_1, \ldots, s_n\}$, o conjunto de localizações onde os dados foram observados, podemos gerar uma amostra de $\boldsymbol{y}(s_0)|\boldsymbol{y}$ diretamente de $N_m\left(\boldsymbol{\mu}(s_0) + \boldsymbol{W}(s_0), V\right)$, para cada usando os valores dos parâmetros da cadeia final do MCMC. Se $s_0$ for uma nova localização,

então é preciso adotar um procedimento de duas etapas. Primeiro, é preciso amostrar $\boldsymbol{W}(s_0)|\boldsymbol{y}$, o que pode ser feito pelo *composition sampling*, gerando $W_1(s_0), \ldots, W_m(s_0)$ das suas distribuições condicionais completas, e em seguida, amostrar $\boldsymbol{y}(s_0)$, seguindo o mesmo procedimento anterior. Note que para a média $\boldsymbol{\mu}(s_0)$, é a covariável que determina a variação espacial. Então, nesse caso ou a covariável é fixa no espaço ou precisa ser interpolada. Em alguns casos, essa interpolação é simples, por exemplo, se as covariáveis forem a latitude e a longitude.

Dado que somos capazes de realizar predições das variáveis respostas em novas localizações $s_0$, a interpolação em toda a região de interesse, ocorre por meio da construção de um grid regular com pontos gerados ao longo de toda a região. Em seguida, aplicamos o procedimento de predição para obter os valores das variáveis respostas em cada ponto do grid. Dessa forma, quanto mais fino, for o grid construído, mais preciso é a interpolação. Além disso, podemos aproveitar as amostras geradas pelo *composition sampling*, para calcular a correlação *a posteriori* dos processos espaciais, através da aplicação do método de Monte Carlo. Nesse caso, $Cor(W_j(s), W_q(u)) \approx Cor\left(W_j^{(1:N)}(s), W_q^{(1:N)}(u)\right)$, onde representa a amostra de tamanho $N$ de $W_j$ gerada na localização $s$.

## 2.6 MEDIDAS DE COMPARAÇÃO

Para avaliar a capacidade preditiva dos modelos nos estudos de simulação e na aplicação em dados reais, foram consideradas três medidas: RMSE (Raiz quadrada do Erro Quadrático Médio), MAE (Erro Absoluto Médio) e IS (*Interval Score*). As duas primeiras avaliam o quão distante as predições se distanciam dos valores verdadeiros. Ambas são medidas que avaliam a estimação pontual dos modelos e são calculadas da seguinte forma:

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\hat{y}_i - y_i)^2} \tag{2.25}$$

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|\hat{y}_i - y_i| \tag{2.26}$$

Em que, $\hat{y}_i$ é o valor estimado de $y_i$, é o valor observado. No contexto bayesiano em que o modelo foi ajustado via MCMC, cada $\hat{y}_i$ pode ser uma amostra da cadeia de Markov gerada pelo algoritmo e pela fórmula, podemos gerar uma nova cadeia para o $RMSE$ e $MAE$. Neste caso, tomamos a média dessas medidas como referência para a comparação dos modelos.

A terceira medida considerada é o IS, que é um critério que avalia o intervalo preditivo do modelo, penalizando os modelos cujo intervalo preditivo é muito estreito e não contém o verdadeiro valor do parâmetro. Mais detalhes podem ser vistos em (GNEITING; RAFTERY, 2007). Seja $y_0$ o valor observado, $q_1$ e $q_2$ o limite inferior e o limite superior

do intervalo preditivo e $1 - \alpha$ a probabilidade do intervalo, então o IS é definido como:

$$IS = (q_2 - q_1) + \frac{2}{\alpha} I(y_0 < q_1) + \frac{2}{\alpha} I(y_0 > q_2) \tag{2.27}$$

Neste trabalho, para todas as aplicações dessa medida, foi considerado $\alpha = 0.05$. Outras medidas que servem como critérios multivariados para previsões probabilísticas que consideram amostras finitas de uma distribuição de interesse poderiam ser considerados, como *Energy Score* (GNEITING; RAFTERY, 2007) e *Variogram Score* (SCHEUERER; HAMILL, 2015).

# 3 ESTUDO DE SIMULAÇÃO 1

O número de funções necessárias para determinar a estrutura de covariância dos processos espaciais e, principalmente, o número de parâmetros cresce rapidamente em função de $m$. Para valores altos de $m$, a estimação do modelo pode se tornar difícil, pela necessidade de definir tantas funções com tantos parâmetros para sua especificação. Para lidar com este cenário, propomos induzir esparsidade dentro da estrutura de covariância cruzada entre as componentes do processo espacial limitando o número máximo de condicionantes de cada fator do produtório. Para avaliar o quanto essa simplificação da estrutura de correlação pode impactar a qualidade do ajuste e da predição do modelo, um estudo de simulação foi realizado sob quatro cenários diferentes de relação entre as componentes, ajustando o modelo completo, onde a distribuição conjunta é representada como na Equação (2.7) e ajustando o modelo esparso onde limitamos a Equação (2.14) em no máximo uma condicionante em cada fator. Além disso, também queremos avaliar o método de seleção de modelos. Para isso, propositalmente, os dados foram gerados estabelecendo relações mais fortes entre algumas variáveis e mais fraco entre outras. Neste estudo, não consideramos o tempo, apenas o espaço.

Um vetor de processos espaciais com três componentes: $\boldsymbol{W}_1$, $\boldsymbol{W}_2$ e $\boldsymbol{W}_3$ e o vetor da variável resposta foram gerados em 150 pontos dispostos aleatoriamente no quadrado $D = [0, 1] \times [0, 1]$, onde 100 localizações foram utilizadas para ajustar os modelos e 50 localizações foram usadas para avaliar as suas capacidades preditivas. Ao todo, foram considerados quatro cenários com diferentes estruturas de dependência espacial entre as componentes. No primeiro cenário, assumimos uma relação espacialmente simétrica entre todas as componentes, com $\boldsymbol{W}_2$ tendo mais impacto na geração de $\boldsymbol{W}_3$. No segundo cenário, $\boldsymbol{W}_2$ continua sendo mais influente em $\boldsymbol{W}_3$, mas adicionamos assimetria na relação entre $\boldsymbol{W}_1$ e $\boldsymbol{W}_3$. No terceiro cenário, a assimetria foi adicionada na relação entre $\boldsymbol{W}_2$ e $\boldsymbol{W}_3$. No quarto cenário, temos assimetria em todas as relações de $\boldsymbol{W}_3$, com ambos $\boldsymbol{W}_1$ e $\boldsymbol{W}_2$ tendo a mesma influência. Em todos os cenários a estrutura geral considerada gerava os dados pela estrutura: $[\boldsymbol{W}_1][\boldsymbol{W}_2|\boldsymbol{W}_1][\boldsymbol{W}_3|\boldsymbol{W}_1, \boldsymbol{W}_2]$, com as mesmas funções de covariância utilizadas. A diferença entre os cenários ocorre pelas diferentes especificações das funções de interação entre as componentes. No Apêndice B, são mostrados os hiperparâmetros utilizados na geração dos dados em cada cenário.

O procedimento para o ajuste do modelo esparso leva em consideração todos os possíveis modelos com diferentes estruturas de correlação entre as componentes do processo espacial. Neste caso, as possíveis especificações das condicionais são: $[\boldsymbol{W}_1][\boldsymbol{W}_2|\boldsymbol{W}_1][\boldsymbol{W}_3|\boldsymbol{W}_1]$, $[\boldsymbol{W}_1][\boldsymbol{W}_2|\boldsymbol{W}_1][\boldsymbol{W}_3|\boldsymbol{W}_2]$ e $[\boldsymbol{W}_1][\boldsymbol{W}_3|\boldsymbol{W}_1][\boldsymbol{W}_2|\boldsymbol{W}_3]$, que são as estruturas representadas na Figura 2. Qualquer outra organização dos fatores e reordenação dos processos levam a uma covariância cruzada equivalente a um dos modelos. No fim, ajustamos apenas o

modelo com dependência espacial que apresentou o maior valor na distribuição preditiva aproximada dada por (2.23). Para ambos os modelos, os hiperparâmetros das distribuições *a priori* consideradas foram $a_q = \gamma_q = a_{\alpha_q} = \gamma_{\alpha_q} = 0.1$, $a_{\nu_q} = \gamma_{\nu_q} = 1$, para os hiperparâmetros das funções de covariânica e $m_{A_{jq}} = m_{r_{jq}} = m_{\Delta_{jq}} = 0$, $V_{A_{jq}} = V_{\Delta_{jq}} = 0.1$, $V_{r_{jq}} = 1, \forall j, q$ para os hiperparâmetros das funções de interação. Na Tabela 3 as distribuições preditivas avaliadas dos modelos em cada cenário são apresentados. Nas Figuras 3, 4, 5 e 6 são apresentados os processos espaciais interpolados dentro do quadrado $D$, comparando o processo verdadeiro com os processos estimados pelo modelo completo e pelo modelo esparso. Por fim, na Tabela 4 são mostrados algumas medidas para avaliar os erros de predição obtidos nos ajustes com base nos 50 pontos separados inicialmente.

| Modelo | Cenário 1 | Cenário 2 | Cenário 3 | Cenário 4 |
|---|---|---|---|---|
| $[\boldsymbol{W}_1][\boldsymbol{W}_2\|\boldsymbol{W}_1][\boldsymbol{W}_3\|\boldsymbol{W}_2]$ | -716.28 | -700.11 | -728.97 | -731.61 |
| $[\boldsymbol{W}_1][\boldsymbol{W}_2\|\boldsymbol{W}_1][\boldsymbol{W}_3\|\boldsymbol{W}_1]$ | -738.76 | -737.95 | -736.56 | -735.10 |
| $[\boldsymbol{W}_1][\boldsymbol{W}_3\|\boldsymbol{W}_1][\boldsymbol{W}_2\|\boldsymbol{W}_3]$ | -720.88 | -705.49 | -737.03 | -739.48 |

Tabela 3 – Avaliação das distribuições preditivas aproximadas dos modelos calculadas em cada cenário.

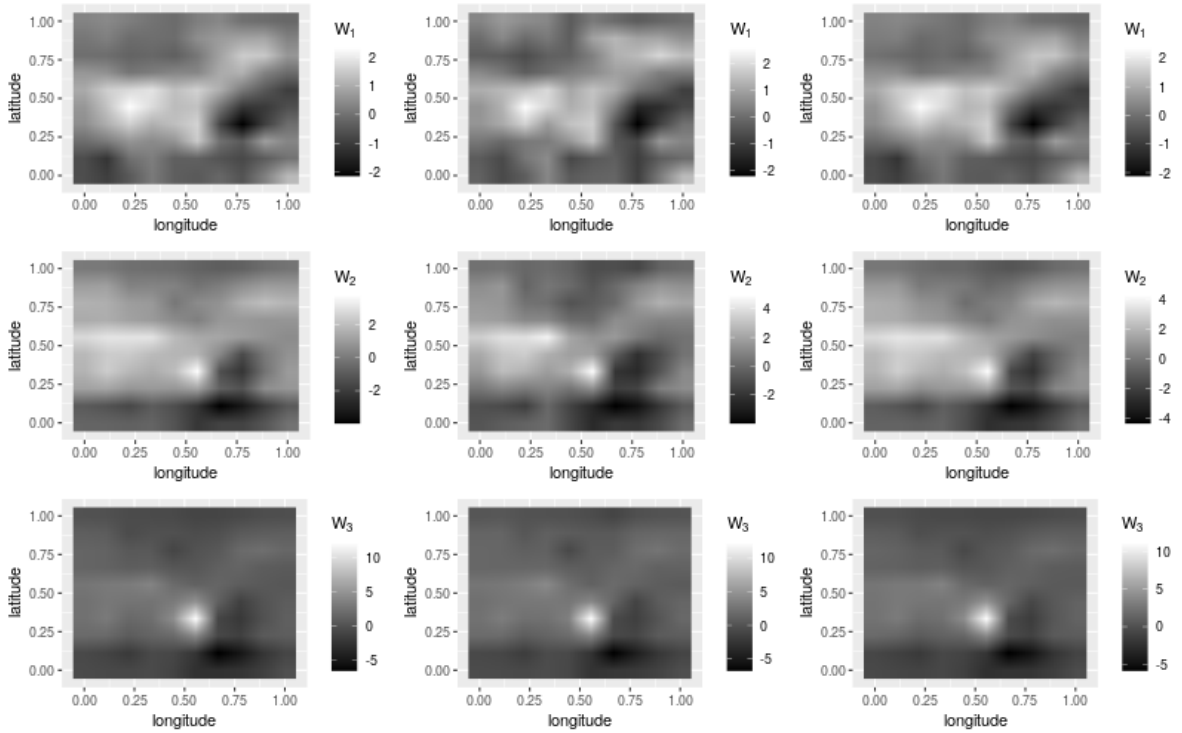



Figura 3 – Interpolação e predição dos processos espaciais no primeiro cenário. No meio temos o processo verdadeiro, na esquerda o processo resultante do modelo esparso e na direita do modelo completo.

Em todos os cenários o modelo selecionado foi aquele com a estrutura $[\boldsymbol{W}_1][\boldsymbol{W}_2|\boldsymbol{W}_1][\boldsymbol{W}_3|\boldsymbol{W}_2]$, o modelo que assume independência condicional entre $\boldsymbol{W}_1$ e $\boldsymbol{W}_3$. De fato, nos três primeiros cenários esse seria o melhor modelo, uma vez que as relações entre $\boldsymbol{W}_3$ e $\boldsymbol{W}_2$ e

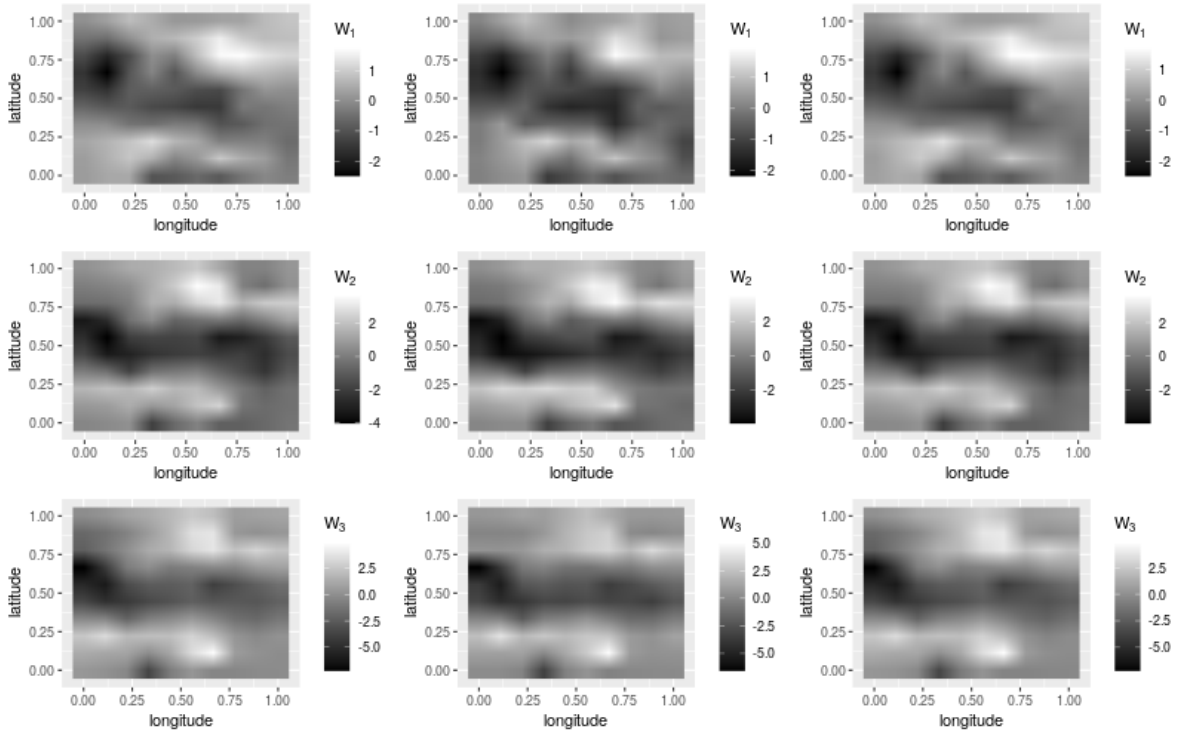



Figura 4 – Interpolação e predição dos processos espaciais no segundo cenário. No meio temos o processo verdadeiro, na esquerda o processo resultante do modelo esparso e na direita do modelo completo.

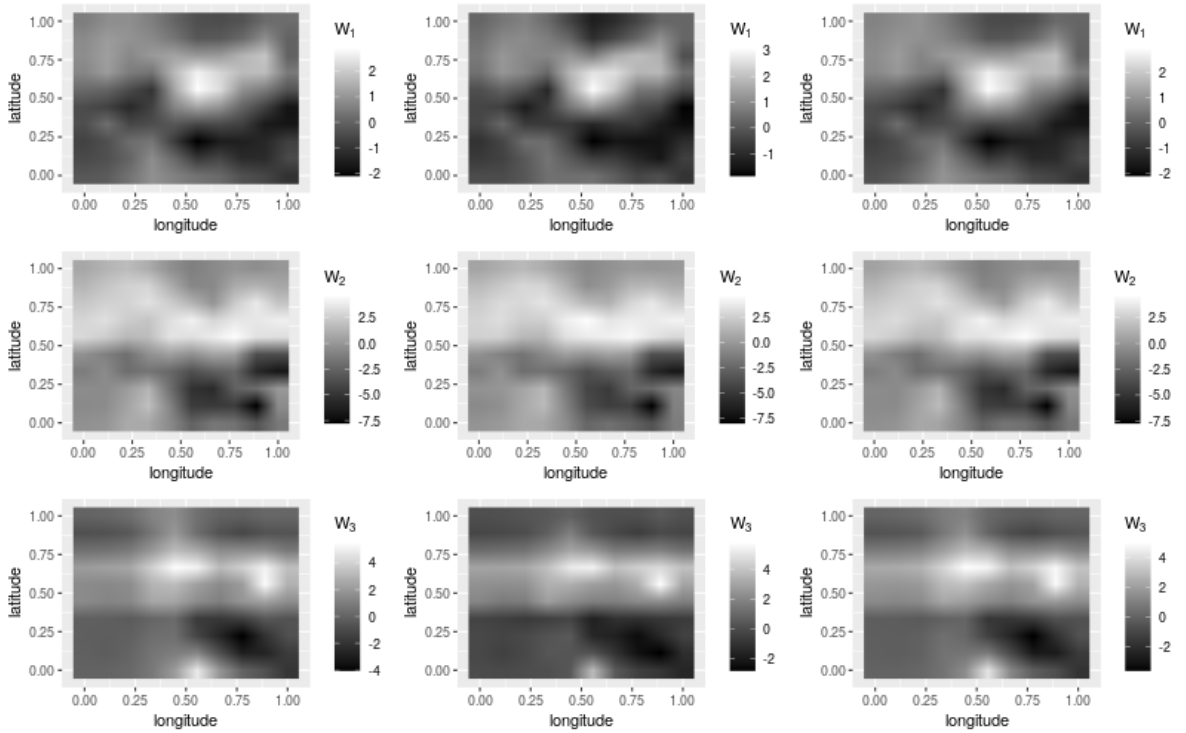



Figura 5 – Interpolação e predição dos processos espaciais no terceiro cenário. No meio temos o processo verdadeiro, na esquerda o processo resultante do modelo esparso e na direita do modelo completo.

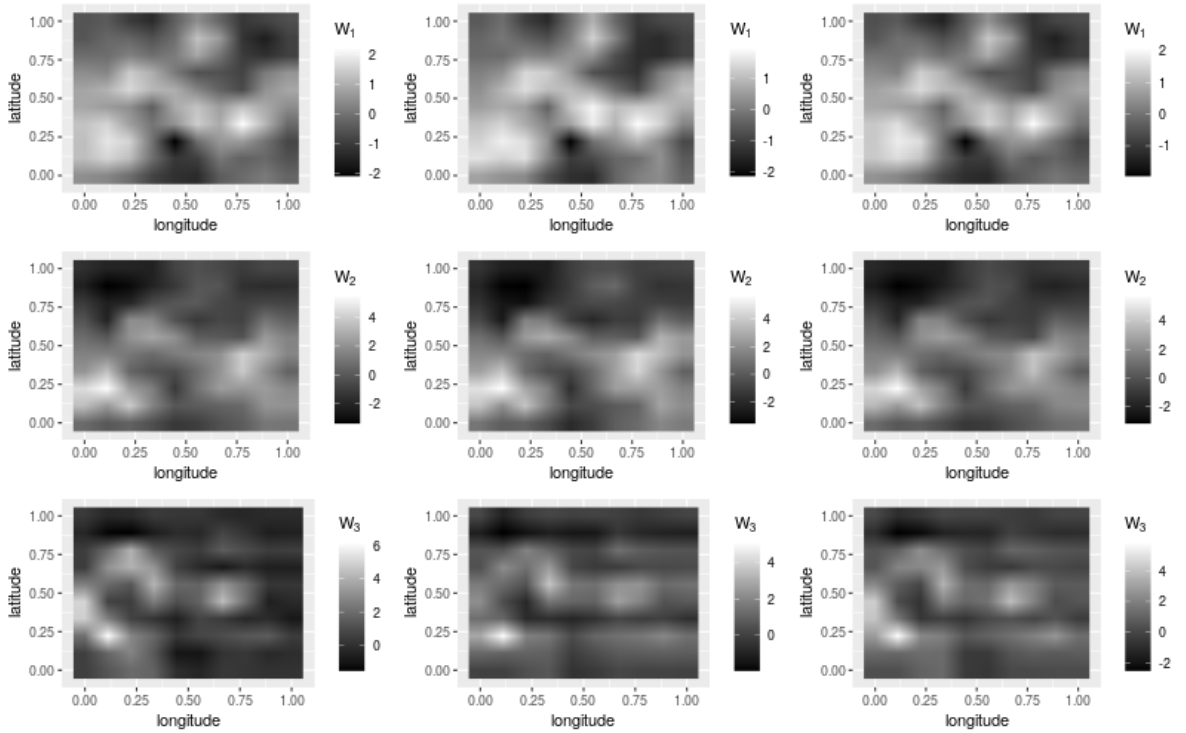



Figura 6 – Interpolação e predição dos processos espaciais no quarto cenário. No meio temos o processo verdadeiro, na esquerda o processo resultante do modelo esparso e na direita do modelo completo.

entre $\boldsymbol{W}_1$ e $\boldsymbol{W}_2$ são mais fortes do que a relação entre $\boldsymbol{W}_3$ e $\boldsymbol{W}_1$. No último cenário as relações entre $\boldsymbol{W}_3$ e $\boldsymbol{W}_2$ e $\boldsymbol{W}_3$ e $\boldsymbol{W}_1$ são equivalentes, então a primeira ou a segunda estrutura deveriam ser selecionados. Portanto, o método de seleção baseado na preditiva aproximada selecionou o modelo mais adequado em todos os cenários.

| Modelo | Medida | Cenário 1 | Cenário 2 | Cenário 3 | Cenário 4 |
|---|---|---|---|---|---|
| Completo | RMSE | 0.8387 | 0.5833 | 0.6012 | **0.8081** |
| | MAE | 0.5616 | 0.4586 | **0.4508** | **0.6094** |
| | IS | 2.5862 | 1.9224 | **1.9289** | **2.5964** |
| Esparso | RMSE | **0.7839** | **0.5592** | **0.5952** | 0.9217 |
| | MAE | **0.5612** | **0.4380** | 0.4554 | 0.6993 |
| | IS | **2.3267** | **1.7808** | 1.9936 | 3.4629 |

Tabela 4 – Erros de predição obtidos nos ajustes em cada cenário, usando as medidas RMSE, MAE e IS.

Podemos observar nos gráficos de interpolação nas Figuras 3, 5, 4 e 6, que em todos os cenários o padrão espacial resultante do modelo esparso e do modelo completo estão muito próximos do padrão espacial verdadeiro, isso é um indicativo de que ambos os modelos se ajustaram bem aos dados, nos quatro cenários. Além disso, pela Tabela 4, observamos que, de forma geral, a diferença nos erros de predição entre o modelo cheio e o modelo esparso estão muito próximos em todos os cenários, inclusive com o modelo esparso apresentando erros menores no primeiro e no segundo cenário. No último cenário,

onde a influência de $\boldsymbol{W}_1$ e $\boldsymbol{W}_2$ em $\boldsymbol{W}_3$ são equivalentes, a perda observada na simplificação foi maior, de acordo com as medidas de erro consideradas. O terceiro cenário foi o mais equilibrado, onde os erros mais ficaram próximos, com o modelo esparso apresentando RMSE menor, mas MAE e IS maiores. Isso demonstra que a nossa simplificação não teve um impacto negativo na estimação do processo espacial, mesmo em diferentes cenários.

No Apêndice B, podem ser vistos as estimativas obtidas no ajuste do modelo esparso e do modelo completo em cada cenário. Pode-se destacar como não há diferença nas estimativas dos hiperparâmetros das funções de covariância, principalmente nos três primeiros cenários. No último, o parâmetro $\phi_3$, que se refere a escala da função de covariância da distribuição condicional $W_3$, foi um pouco maior, mas não significativamente, no modelo esparso, o que indica que a variância explicada pelos condicionantes é menor no modelo esparso do que no modelo completo, o que era esperado, já que, nesse caso, $W_1$ que tem uma relação forte com $W_3$ não é considerado na condicional. Em geral, as estimativas estiveram próximas dos valores usados na geração dos dados, com os intervalos de credibilidade de 95% sempre incluindo o valor verdadeiro.

Entre os hiperparâmetros das funções de interação, de modo geral, também foram observados estimativas muito próximas entre o modelo esparso e o modelo completo, entre os hiperparâmetros que são estimados pelos dois modelos. Note que, em todos os cenários a distribuição condicional $W_3|W_2$ é considerada, portanto os hiperparâmetros associados a função de interação de $W_1$ e $W_3$ não são estimados. Nesses casos $A_{13}$ é fixado em 0. Em relação ao modelo completo, nota-se que no primeiro cenário o modelo estimou uma relação mais forte entre $W_1$ e $W_3$ e uma relação quase nula entre $W_2$ e $W_3$, com $A_{23}$ sendo estimado muito próximo de 0. Como consequência disso, os parâmetros de abertura e de assimetria apresentaram muita incerteza em suas estimações. A estimação de uma relação mais forte entre $W_1$ e $W_3$ também ocorre no último cenário, mas nesse caso os demais parâmetros foram bem estimados. Essas questões podem ter ocorrido pelo fato de $W_1$ e $W_2$ serem altamente correlacionadas, gerando um possível confundimento na estimação do nível das relações dessas variáveis com $W_3$. No segundo e no terceiro cenário o modelo completo estimou uma relação mais forte entre $W_2$ e $W_3$ e uma relação mais fraca entre $W_1$ e $W_3$, sendo quase nula no segundo cenário, gerando, para este cenário, um grau de incerteza muito elevado na estimação dos demais hiperparâmetros associados a relação de $W_1$ e $W_3$. Pela Tabela 4, nota-se que nos cenários 1 e 2, onde os hiperparâmetros apresentaram muita incerteza, foi onde o modelo completo perdeu em nível preditivo para o modelo esparso.

# 4 ESTUDO DE SIMULAÇÃO 2

Um segundo estudo de simulação foi realizado para comparar o modelo sem mistura, com uma única estrutura de covariância, e o modelo de mistura, que é capaz de acomodar diferentes estruturas de covariâncias ao longo do tempo. Para este estudo, foram considerados três diferentes cenários, especificados com duas estruturas de covariâncias distintas ao longo do tempo. Desta forma, queremos observar se o modelo de mistura é capaz de acomodar dois modelos distintos especificados sob diferentes covariâncias, de forma que cada modelo esteja associado a uma das estruturas simuladas. Além disso, queremos também avaliar os possíveis ganhos que temos na capacidade preditiva ao utilizar o modelo de mistura quando a estrutura de covariâncias dos dados se altera ao longo do tempo.

Assim, como no primeiro estudo simulado, foram gerados dados trivariados, em 150 localizações dispostas aleatoriamente em $D = [0, 1] \times [0, 1]$, onde 100 foram usadas para o ajuste dos modelos e 50 para avaliar a capacidade preditiva. Além disso, os dados forma simulados em 50 períodos de tempo. Nos dois primeiros cenários deste estudo, a primeira estrutura de dependência, foi a mesma utilizada no terceiro cenário do primeiro estudo de simulação. A diferença entre eles ocorre na especificação da segunda estrutura de dependência. No primeiro cenário consideramos uma mudança mais drástica de uma estrutura para a outra. Neste caso, a segunda estrutura fornece uma leve mudança na assimetria existente na relação entre $\boldsymbol{W}_2$ e $\boldsymbol{W}_3$ e, além disso, essa estrutura considera uma relação mais forte entre $\boldsymbol{W}_1$ e $\boldsymbol{W}_3$. Já no segundo cenário, a ideia é provocar uma mudança mais suave de uma estrutura de covarância para outra. Neste cenário, a segunda estrutura mantém a hierárquia da força das relações e assimetria entre as variáveis, com apenas algumas mudanças sutis, na força dessas relações. Em ambos os casos, os dados forma simulados com a primeira estrutura de covariância nos 25 primeiros períodos, e a segunda estrutura nos últimos 25 períodos. Nestes dois cenários, o objetivo principal é comparar os modelos de mistura com o modelo sem mistura.

No último cenário, tentamos criar uma situação mais realista, onde a mudança da primeira estrutura para a segunda ocorre de forma mais suave, passando por um período de transição. Nesse caso, nós consideramos a mudança nas estruturas ocorrendo apenas em função da assimetria das relações entre $\boldsymbol{W}_1$ e $\boldsymbol{W}_3$ e entre $\boldsymbol{W}_2$ e $\boldsymbol{W}_3$, por exemplo, como se o vento estivesse mudando de direção de forma gradual. Desta forma, no período de transição, essa assimetria segue gradativamente da direção especificada na primeira estrutura de covariância para a direção da segunda estrutura. Neste cenário, o intuito é fazer uma comparação entre os modelos de mistura, com pesos estáticos e pesos evoluindo no tempo e verificar como cada um desses modelos acomoda as estruturas de covariâncias consideradas. No Apêndice B, são mostrados os hiperparâmetros usados na geração dos dados em cada um dos três cenários.

Como são três variáveis, se considerarmos a especificação simplificada que condiciona cada componente a no máximo uma condicionante, como foi no primeiro estudo simulado, nós temos três opções de modelos a serem considerados nos modelos de mistura, que são as mesmas opções do estudo anterior. Ao todos foram ajustados cinco modelos em cada cenário. Os fatores de desconto da regressão dinâmica e do processo Dirichlet foram fixados em 0.9. As mesmas configurações de *prioris* do primeiro estudo de simulação foram usadas para os hiperparâmetros das funções de covariância e interação, para todos os modelos. Para facilitar a notação da análise, na Tabela 5 são apresentados uma descrição e uma notação usada como referência para cada um dos cinco modelos. Essa notação foi utilizada nas análises subsequentes.

| Notação | Descrição |
|---|---|
| $MC$ | Modelo sem mistura que considera as relações entre todas as componentes. |
| $MME1$ | Modelo de mistura com pesos estáticos no tempo que considera apenas os dois melhores modelos de acordo com o método de seleção de modelos. |
| $MME2$ | Modelo de mistura com pesos estáticos no tempo que considera os três possíveis modelos. |
| $MMD1$ | Modelo de mistura com pesos evoluindo no tempo de acordo com um processo Dirichlet que considera apenas os dois melhores modelos de acordo com o método de seleção de modelos. |
| $MMD2$ | Modelo de mistura com pesos evoluindo no tempo de acordo com um processo Dirichlet que considera os três possíveis modelos. |

Tabela 5 – A notação utilizada para referenciar cada um dos modelos considerados no estudo de simulação junto com a descrição do modelo a qual essa notação se refere.

Na Tabela 6 são mostradas as distribuições preditivas de cada modelo em cada cenário, que foram utilizadas para fazer a seleção de modelos para as misturas de $MMD1$ e $MME1$. Nas Figuras 7, 8 e 9 são apresentados os gráficos com a interpolação dos processos espaciais referentes a primeira estrutura de covariância, no primeiro cenário. Já nas Figuras 10, 11 e 12, temos as interpolações para a segunda estrutura de covariância. Na Tabela 7 são mostrados os erros de predição de cada modelo ajustado em cada cenário.

| Modelo | Cenário 1 | Cenário 2 | Cenário 3 |
|---|---|---|---|
| $[\boldsymbol{W}_1][\boldsymbol{W}_2\|\boldsymbol{W}_1][\boldsymbol{W}_3\|\boldsymbol{W}_2]$ | -33215.68 | -33371.50 | -29188.97 |
| $[\boldsymbol{W}_1][\boldsymbol{W}_2\|\boldsymbol{W}_1][\boldsymbol{W}_3\|\boldsymbol{W}_1]$ | -33216.25 | -33371.71 | -29224.65 |
| $[\boldsymbol{W}_1][\boldsymbol{W}_3\|\boldsymbol{W}_1][\boldsymbol{W}_2\|\boldsymbol{W}_3]$ | -33216.32 | -33372.15 | -29264.03 |

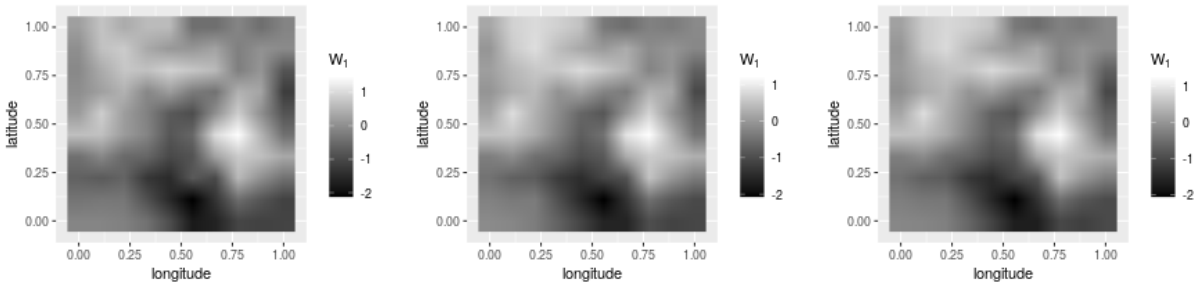
Tabela 6 – Distribuições preditivas dos modelos calculadas em cada cenário, na escala logarítmica.

Pela Tabela 6, podemos ver que em todos cenários os modelos selecionados foram aqueles com as estruturas $[\boldsymbol{W}_1][\boldsymbol{W}_2|\boldsymbol{W}_1][\boldsymbol{W}_3|\boldsymbol{W}_2]$ e $[\boldsymbol{W}_1][\boldsymbol{W}_2|\boldsymbol{W}_1][\boldsymbol{W}_3|\boldsymbol{W}_1]$. Nos dois primeiros cenários, a diferença nos valores da distribuição preditiva foi relativamente baixa,

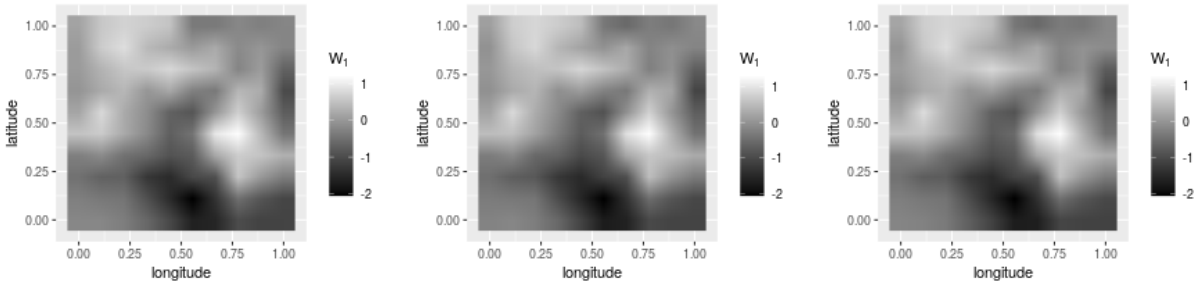
enquanto no último cenário essa diferença foi maior. Neste caso, esses foram os dois modelos considerados nos modelos $MMD1$ e $MME1$. Em relação aos ajustes, na maior parte dos casos os modelos de mistura foram capazes de acomodar uma estrutura de covariância para a primeira estrutura simulada e outra estrutura de covariância para a segunda estrutura simulada. Em outras palavras, ao longo dos ajustes, todos os modelos de mistura, após o período de *burn-in*, selecionou um modelo para os períodos correspondentes a cada estrutura simulada. Nesse sentido, os modelos $MMD1$ e $MME1$, que consideraram a seleção de modelos previamente, apresentaram resultados idênticos, selecionando os mesmos modelos em cada estrutura, nos três cenários. Este resultado é positivo, pois demonstra que os modelos de mistura são capazes de identificar mudanças na estrutura de covariância, mesmo quando essas mudanças são sutis, como no segundo cenário.

A diferença ocorreu quando as misturas consideravam os três modelos possíveis para três variáveis. Neste caso, os modelos precisam identificar dois modelos, um para cada estrutura de covariância, entre os três disponíveis. Em particular, no último cenário foi onde pudemos observar uma diferença mais significativa entre o modelo de mistura com pesos fixos e com pesos evoluindo no tempo. Enquanto o modelo com os pesos evoluindo com um processo Dirichlet, $MMD2$, seguiu os resultados anteriores e selecionou um modelo a cada estrutura de covariância, o modelo de mistura com pesos fixos $MME2$ entendeu o período de transição de uma estrutura para outra como uma nova estrutura de correlação, o que não é verdade. Deste modo, o modelo de mistura considerou um terceiro modelo para esse período. Assim, o modelo de mistura com processo Dirichlet demonstrou mais robustez para lidar com este tipo de cenário.

Pelos gráficos de interpolação das Figuras 7, 8, 9, 10, 11 e 12 pode-se observar que todos os modelos de mistura conseguiram capturar o padrão verdadeiro com bastante precisão, em ambas as estruturas consideradas no primeiro cenário, indicando que os modelos de mistura se ajustaram bem aos dados. Já o modelo sem mistura teve mais dificuldade devido a sua limitação de considerar uma única estrutura de covariância. Como resultado do ajuste, pode-se ver que a interpolação resultante desse modelo é uma ponderação da interpolação espacial dos processos verdadeiros em cada estrutura. Para os outros dois cenários, os gráficos de interpolações podem ser visualizados no Apêndice B. Os resultados foram similares ao do primeiro cenário. No segundo cenário, as duas estruturas de covariância levaram a um padrão espacial muito parecido, consequentemente, foi neste cenário que o modelo $MC$ apresentou os melhores resultados.

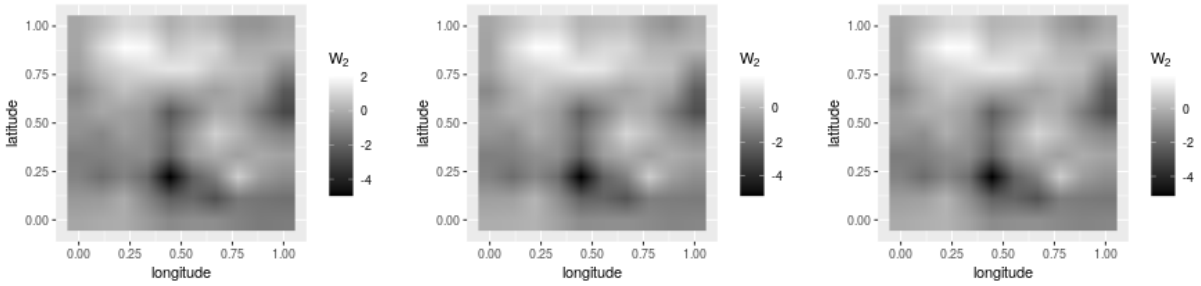



(a) Processo verdadeiro. (b) MMD1. (c) MME1.

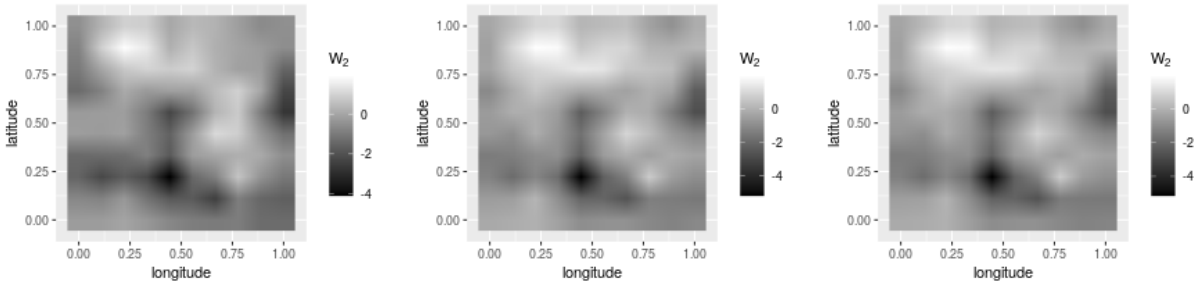



(d) MC. (e) MMD2. (f) MME2.

Figura 7 – Interpolação de $\boldsymbol{W}_1$ na primeira estrutura de dependência espacial.

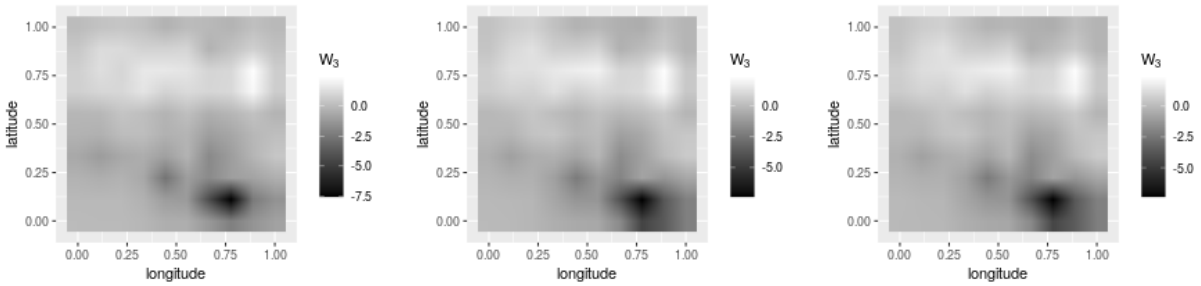



(a) Processo verdadeiro. (b) MMD1. (c) MME1.




(d) MC. (e) MMD2. (f) MME2.

Figura 8 – Interpolação de $\boldsymbol{W}_2$ na primeira estrutura de dependência espacial.



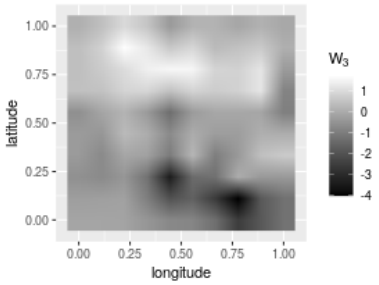

(a) Processo verdadeiro.



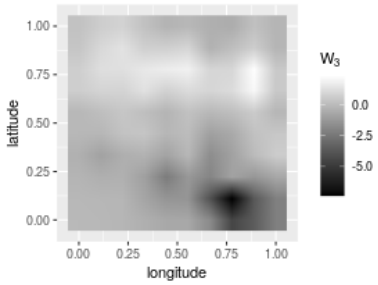

(b) MMD1.



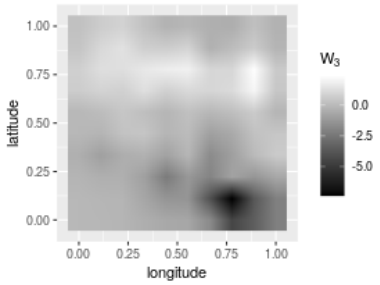

(c) MME1.




(d) MC.

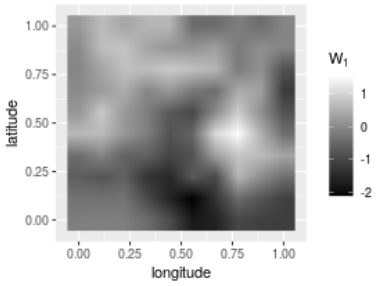



(e) MMD2.

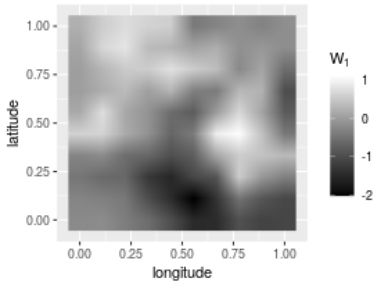



(f) MME2.

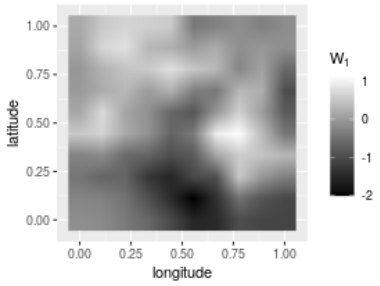
Figura 9 – Interpolação de $\boldsymbol{W}_3$ na primeira estrutura de dependência espacial.



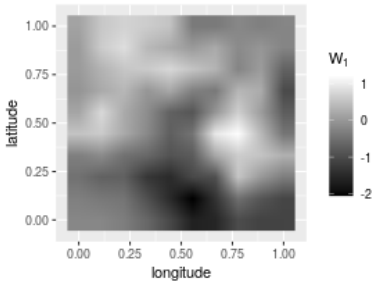

(a) Processo verdadeiro.



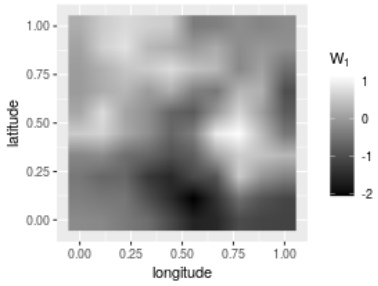

(b) MMD1.



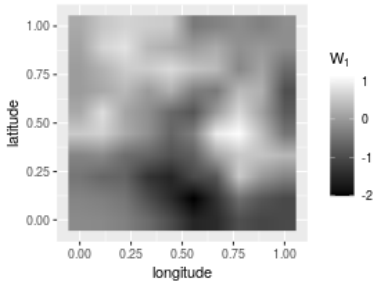

(c) MME1.




(d) MC.

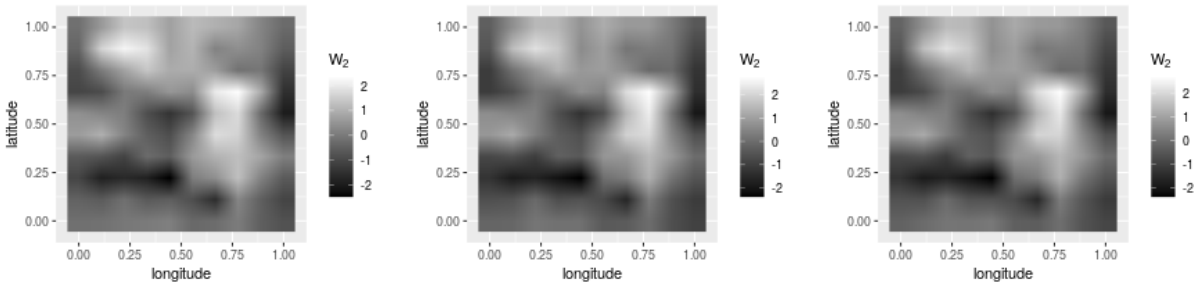



(e) MMD2.

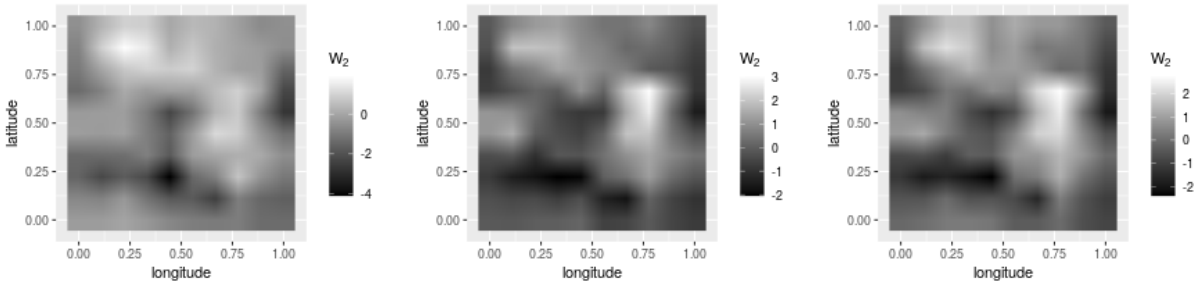



(f) MME2.

Figura 10 – Interpolação de $\boldsymbol{W}_1$ na segunda estrutura de dependência espacial.

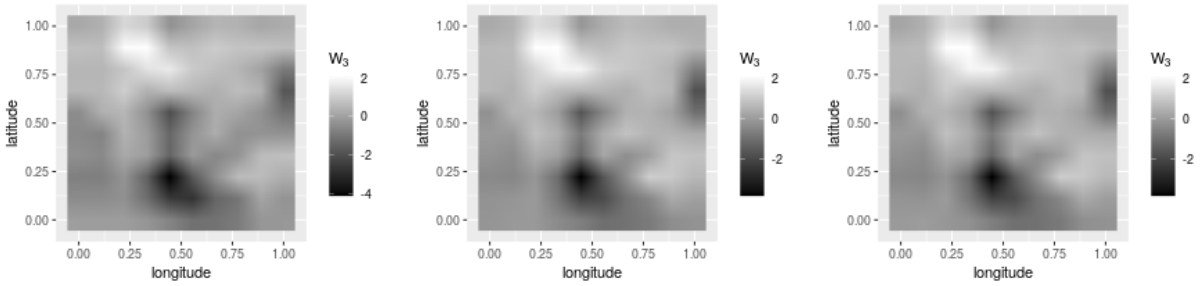



(a) Processo verdadeiro. (b) MMD1. (c) MME1.

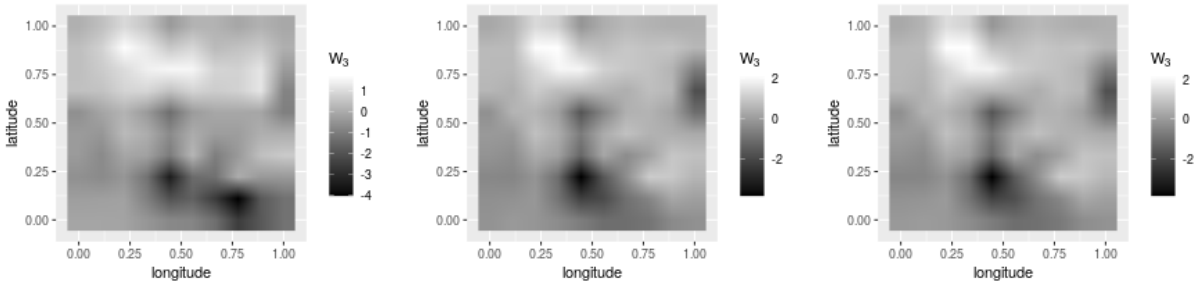



(d) MC. (e) MMD2. (f) MME2.

Figura 11 – Interpolação de $\boldsymbol{W}_2$ na segunda estrutura de dependência espacial.




(a) Processo verdadeiro. (b) MMD1. (c) MME1.




(d) MC. (e) MMD2. (f) MME2.

Figura 12 – Interpolação de $\boldsymbol{W}_3$ na segunda estrutura de dependência espacial.

| | Medidas | MC | MME1 | MMD1 | MME2 | MMD2 |
|---|---|---|---|---|---|---|
| Cenário 1 | RMSE | 0.5560 | 0.4575 | 0.4561 | **0.4546** | 0.5049 |
| | MAE | 0.4211 | 0.3468 | 0.3462 | **0.3451** | 0.3838 |
| | IS | 3.5371 | 1.5227 | 1.5117 | **1.5092** | 1.9306 |
| Cenário 2 | RMSE | 0.4187 | 0.4137 | **0.4111** | 0.6004 | 0.5831 |
| | MAE | 0.3161 | 0.3086 | **0.3071** | 0.4425 | 0.4316 |
| | IS | 1.8629 | 1.4165 | **1.3947** | 2.8125 | 2.7762 |
| Cenário 3 | RMSE | 0.7765 | **0.5883** | **0.5883** | 0.7070 | 0.6056 |
| | MAE | 0.5521 | 0.4364 | **0.4357** | 0.5208 | 0.4546 |
| | IS | 5.9824 | 2.4981 | **2.4870** | 2.8487 | 2.6230 |

Tabela 7 – Erros de predição obtidos nos ajustes em cada cenário. As medidas consideradas foram as mesmas do primeiro estudo de simulação: RMSE, MAE e IS.

Pela Tabela 7, pode-se ver que no primeiro e no último cenário, os modelos de mistura apresentaram uma capacidade preditiva superior ao modelo sem mistura ($MC$). No segundo cenário, que é onde as estruturas de covariância são muito parecidas, o modelo $MC$ se saiu muito bem, apresentando erros muito próximos dos modelos de mistura $MME1$ e $MMD1$. Neste cenário os modelos $MME2$ e $MMD2$, os modelos de mistura que não considera a seleção de modelos previamente, tiveram mais dificuldade de predição, com $MMD2$ apresentando menores. Ao comparar os modelos de mistura entre si, percebe-se, claramente, um desempenho superior dos modelos de mistura que consideraram a seleção de modelos previamente ($MMD1$ e $MME1$). Os modelos $MME2$ e $MMD2$ mostraram muita variabilidade. No primeiro cenário $MME2$ obteve o melhor desempenho, enquanto que nos outros dois foi o pior entre os modelos de mistura. Já o modelo $MMD2$ teve o pior desempenho no primeiro cenário e um desempenho mais aceitável no último. Ainda assim, no geral, os modelos de mistura apresentaram um desempenho superior em relação ao modelo sem mistura, no que diz respeito a capacidade preditiva.

# 5 APLICAÇÃO EM DADOS DE POLUENTES ATMOSFÉRICOS

O modelo proposto foi considerado na análise de dados de poluentes atmosféricos, coletados em estações de monitoramento da qualidade do ar no Reino Unido. Dados de poluentes comumente sofrem a influência de correntes de ar e mudanças de direção do vento. Portando acredita-se que mudanças na estrutura de correlação entre variáveis possam ocorrer e seriam capturadas pela proposta dessa dissertação. Esses dados estão disponíveis para o público e podem ser coletados através do software R (R Core Team, 2023) com o auxílio do pacote *openair* (CARSLAW; ROPKINS, 2012). Este pacote fornece ferramentas para a coleta e visualização desses dados em diferentes períodos de tempo (horas, dias, meses, anos), assim como informações sobre o comportamento do vento ao longo desses períodos.

A base é composta por dados sobre a concentração no ar de quatro dos principais poluentes atmosféricos: $NO$, $NO_2$, $PM_{10}$ e $PM_{2.5}$, em 72 estações de monitoramento, no período que começa em primeiro de julho de 2022 até 22 de junho de 2023. Algumas estações de monitoramento apresentaram dados faltantes em determinados períodos de tempo, o que é um problema comum para dados de poluentes, por diversas razões como falta de energia, falhas do sistema de coleta, entre outros (LI et al., 1999). Para a análise, foram calculadas as médias semanais da concentração de cada poluente ao longo desse período, totalizando 52 semanas. A disposição espacial das estações de monitoramento pode ser visualizada na Figura 13.

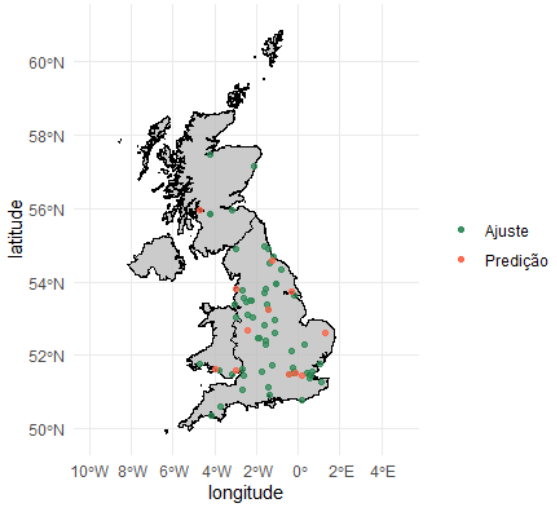



Figura 13 – Mapas do Reino unido. No mapa da esquerda são mostrados as localizações das 72 estações de monitoramento consideradas na análise. Os pontos em verde indicam as estações de monitoramento usadas para ajustar os modelos e os pontos em vermelho indicam as estações de monitoramento que foram retiradas para predição.

## 5.1 ANÁLISE EXPLORATÓRIA

Os poluentes atmosféricos são variáveis que são afetadas pela intensidade e direção do vento. A Figura 14 apresenta gráficos com os padrões de vento observados ao longo dos meses do período analisado. Nas Figuras 15, 16, 17, 18, 19 e 20 são apresentados gráficos em coordenadas polares, onde pode ser visualizado a concentração de um poluente em função da direção e da velocidade do vento. Além disso, para observar a relação entre os poluentes, esses gráficos foram separados de acordo com uma categorização de um segundo poluente. Isso permite realizar uma análise direcional na relação entre os poluentes.

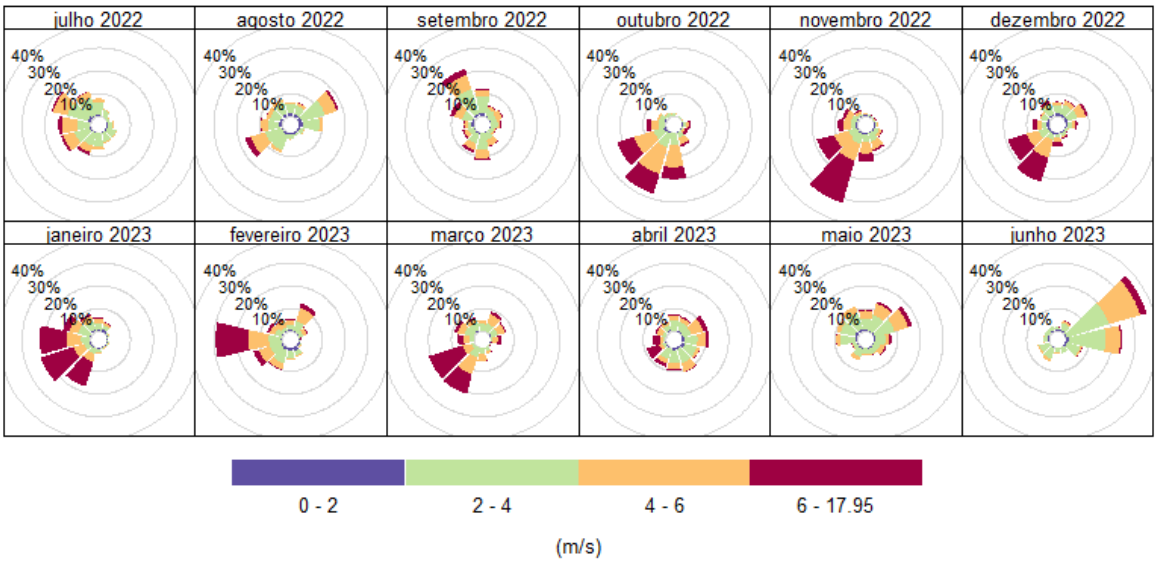



Figura 14 – Gráficos de rosas do vento ao longo dos 12 meses considerados na aplicação do modelo.

Pela Figura 14 observa-se um certo padrão nos primeiros meses observados, com as direções do vento se mostrando mais variadas e as intensidades baixas. Em seguida, nos meses de outubro de 2022 a março de 2023 os ventos estavam muito concentrados nas direções oeste e sudoeste, apresentando intensidades fortes com alta frequência. A partir de abril de 2023, as direções dos se tornam mais variadas, com exceção do mês de junho de 2023, quando os ventos estiveram mais concentrados na direção leste. A intensidade dos ventos nesses últimos meses foi baixa, similar ao que foi observado nos primeiros meses.

Pelos gráficos da Figura 15, nota-se claramente que, quanto maior a concentração de $NO_2$, maior é a concentração de $NO$. Essa relação parece se estabelecer de forma bem uniforme nos gráficos. Apenas na categoria de maior concentração de $NO_2$ é possível destacar alguma mudança de comportamento da concentração de $NO$ em função do vento, onde $NO$ parece ficar mais concentrado quando as velocidades dos ventos são mais baixas. Já na Figura 16, também é observado uma tendência crescente na concentração de $NO$

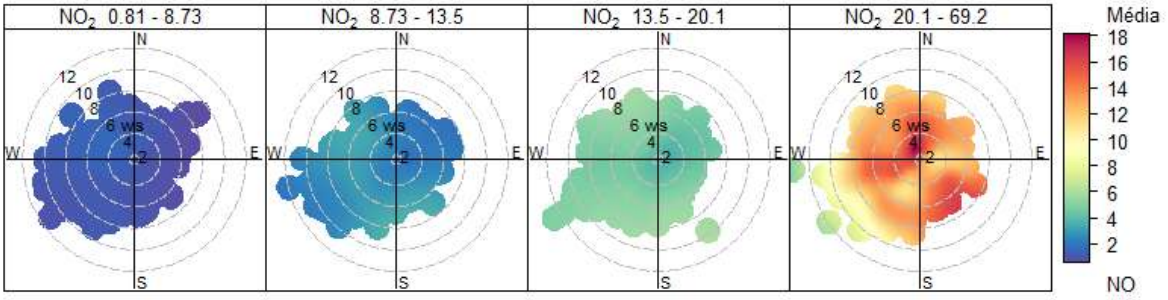



Figura 15 – Gráfico que mostra a concentração de $NO$ em função de uma categorização da concentração de $NO_2$ e em função da direção e velocidade do vento.

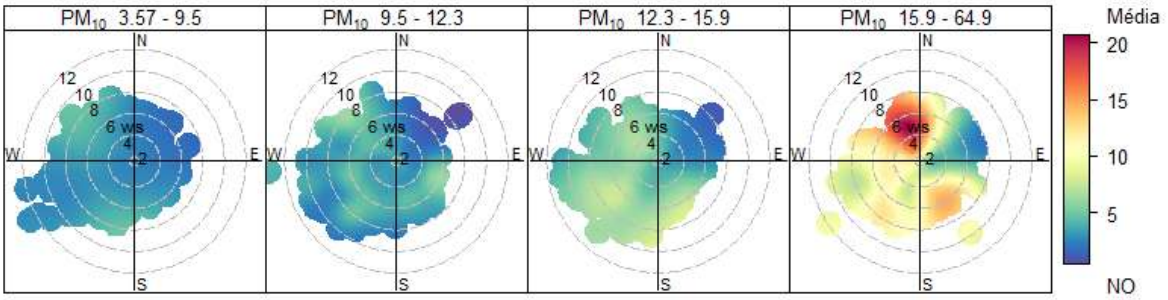



Figura 16 – Gráfico que mostra a concentração de $NO$ em função de uma categorização da concentração de $PM_{10}$ e em função da direção e velocidade do vento.

conforme a concentração de $PM_{10}$ é maior. Mas diferente de antes, aqui o padrão de concentração de $NO$ é mais irregular em função do vento. Quando a concentração de $PM_{10}$ é alta, o nível de concentração de $NO$ sobe muito quando os ventos estão direcionados para noroeste (quase norte) e diminui bastante quando os ventos estão direcionados para leste. Na Figura 17 é possível observar o mesmo padrão da Figura 18, indicando o mesmo comportamento $NO$ em relação aos poluentes $PM_{10}$ e $PM_{2.5}$.

Na Figura 18 pode ser visto que a concentração de $NO_2$ aumenta conforme a concentração de $PM_{10}$ é maior de forma ainda mais evidente do que a concentração de $NO$. Do mesmo modo que antes, os níveis de concentração de $NO_2$ foram maiores quando os ventos se direcionaram a noroeste e menores quando se direcionaram a leste. Esse padrão também é observado no segundo e no terceiro gráfico, onde os níveis de concentração

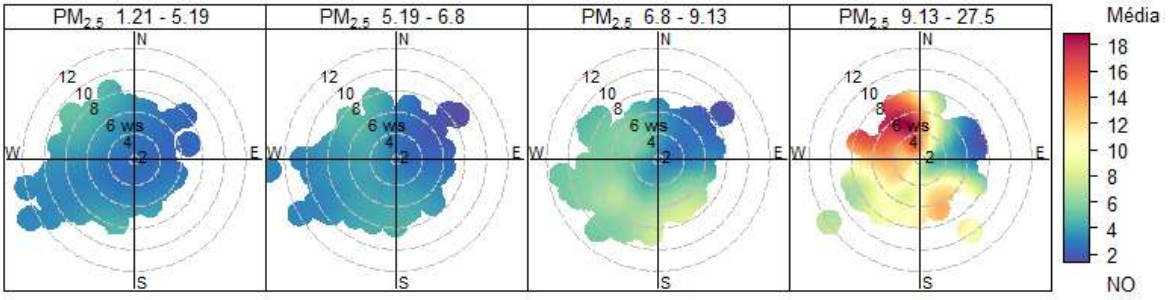



Figura 17 – Gráfico que mostra a concentração de $NO$ em função de uma categorização da concentração de $PM_{2.5}$ e em função da direção e velocidade do vento.

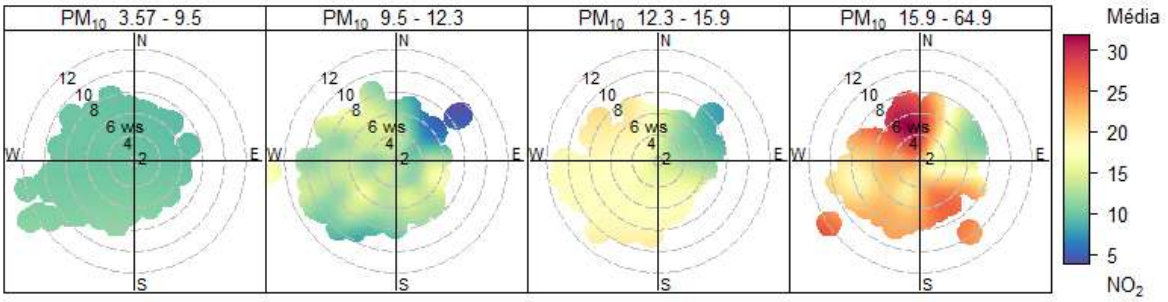



Figura 18 – Gráfico que mostra a concentração de $NO_2$ em função de uma categorização da concentração de $PM_{10}$ e em função da direção e velocidade do vento.

de $PM_{10}$ são intermediários e não apenas na última categoria onde as concentrações de $PM_{10}$ são mais altas, como foi o caso do $NO$. O mesmo padrão é observado na Figura 19, indicando uma relação bem parecida de $NO_2$ com $PM_{10}$ e $PM_{2.5}$, assim como ocorre com o $NO$. Por fim, pela Figura 20, nota-se que a medida que a concentração de PM2.5 aumenta, também aumenta a concentração de PM10. Isso parece ocorrer de forma uniforme em relação a velocidade e a direção do vento.

Para ilustrar as correlações cruzadas entre os poluentes, um modelo de regressão linear multivariado $\boldsymbol{y} = \boldsymbol{\mu} + \boldsymbol{\varepsilon}$ foi ajustado para cada semana considerada na análise, de forma independente. Nesta regressão foram considerados os logaritmos das concentrações dos poluentes e foram usados a latitude e a longitude como covariáveis para remover possíveis padrões espaciais presente nos dados. Na Figura 21 são mostrados as correlações cruzadas

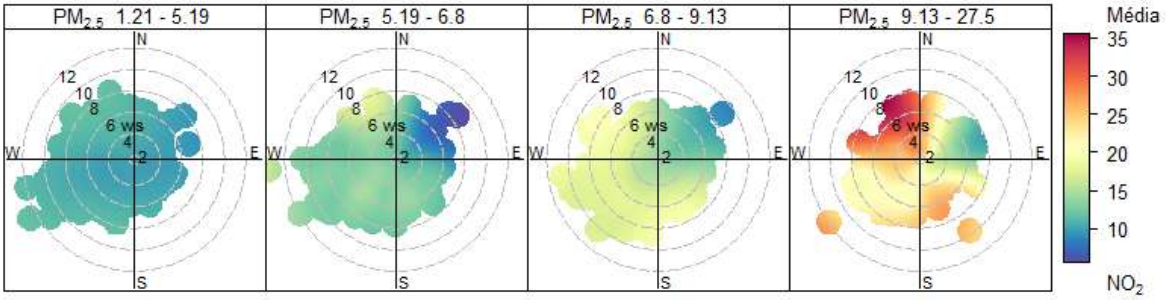



Figura 19 – Gráfico que mostra a concentração de $NO_2$ em função de uma categorização da concentração de $PM_{2.5}$ e em função da direção e velocidade do vento.

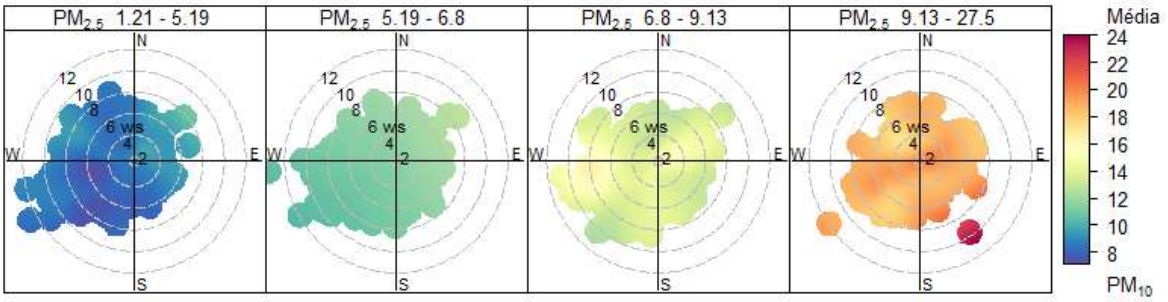
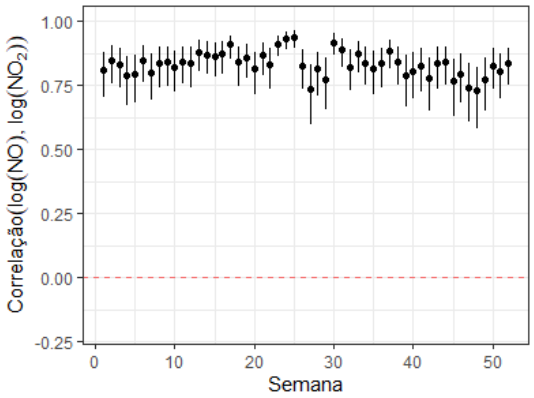



Figura 20 – Gráfico que mostra a concentração de $PM_{10}$ em função de uma categorização da concentração de $PM_{2.5}$ e em função da direção e velocidade do vento.

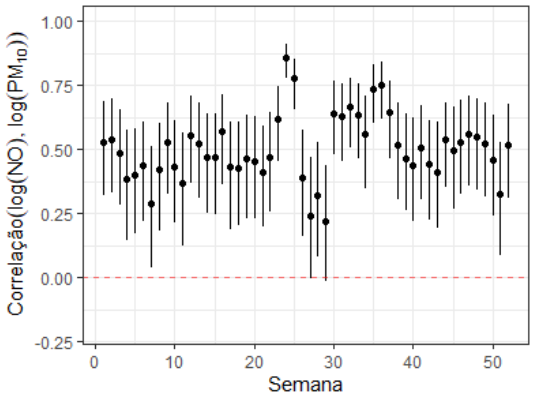
para cada par de poluentes resultantes da matriz de correlação estimada para os erros aleatórios dos modelos ajustados, ao longo das 52 semanas.

Pelas correlações cruzadas observadas na Figura 21, nota-se que os índices de correlações mais altas ocorreram entre o $NO$ e $NO_2$ e entre $PM_{10}$ e $PM_{2.5}$. Em ambos os casos, as correlações parecem subir levemente ao longo das primeiras 20 semanas, apresentando picos por nas semanas 24 e 25. Nas semanas 26 a 29, há uma queda significativa nas correlações e um aumento significativo nas semanas 30 e 31. Em seguida, a partir da semana 31, as correlações voltam a um nível similar ao das primeiras 20 semanas, com uma tendência leve de decrescimento.

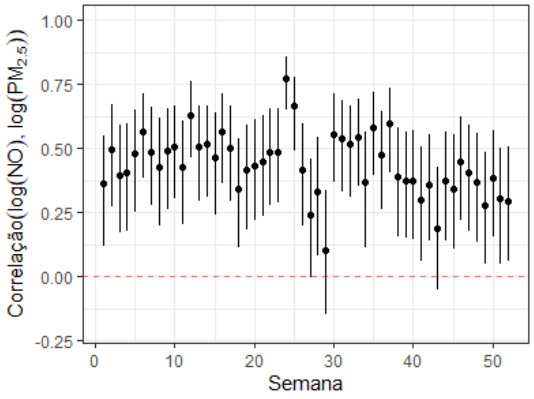
As correlações cruzadas entre $NO$ e $NO_2$ com $PM_{10}$ e $PM_{2.5}$ foram menores e apresentaram mais variabilidades, com intervalos de credibilidade mais amplos. Pode-se notar



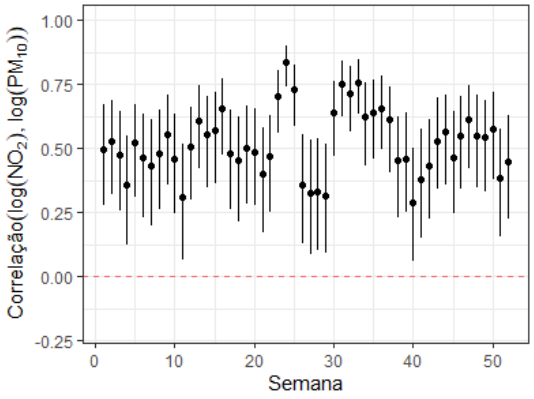

(a) $NO \times NO_2$.



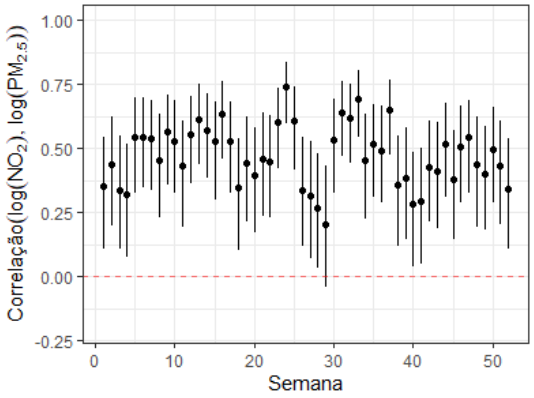

(b) $NO \times PM_{10}$.



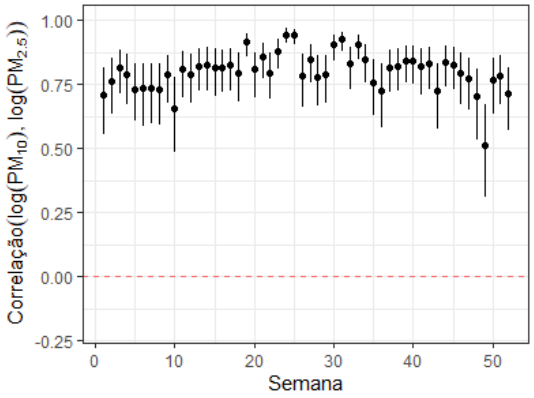

(c) $NO \times PM_{2.5}$.




(d) $NO_2 \times PM_{10}$.




(e) $NO_2 \times PM_{2.5}$.




(f) $PM_{10} \times PM_{2.5}$.

Figura 21 – Mediana *a posteriori* e intervalos de credibilidade de 95% das correlações cruzadas do logaritmo da concentração de cada poluente obtidas pelos modelos de regressão, em cada semana.

comportamentos comuns em todos os gráficos, com picos de correlação por volta das semanas 24 e 25 e uma queda nas semanas 26 a 29. Também são observadas correlações mais altas nas semanas 30 e 31. Em alguns casos essa correlação mais se estende para mais algumas semanas, como é caso das correlações de $NO$ e $PM_{10}$, $NO$ e $PM_{2.5}$ e $NO_2$ $PM_{10}$.

## 5.2 MODELOS AJUSTADOS

As concentrações de poluentes foram transformados utilizando a função logarítmica, assim os dados apresentam um comportamento mais próximo de uma distribuição normal. Para a avaliação da capacidade preditiva de cada modelo considerado nessa aplicação, 12 localizações foram selecionadas aleatoriamente, enquanto as demais foram usadas para o ajuste. A disposição espacial das estações de monitoramento usadas no ajuste e das estações utilizadas para a predição pode ser observada na Figura 13. Como ordenamento, foi considerado $\boldsymbol{Y}_1 : \log(NO)$, $\boldsymbol{Y}_2 : \log(NO_2)$, $\boldsymbol{Y}_3 : \log(PM_{10})$, $\boldsymbol{Y}_4 : \log(PM_{2.5})$, escolhidos de forma arbitrária. Seguindo esse ordenamento, na Tabela 8 podem ser visualizadas as aproximações das distribuições preditivas das especificações que apresentaram os maiores valores.

| Limite de condicionantes | Especificação das condicionais | Preditiva |
|---|---|---|
| $L = 2$ | E1: $[\boldsymbol{W}_1][\boldsymbol{W}_2\|\boldsymbol{W}_1][\boldsymbol{W}_3\|\boldsymbol{W}_2, \boldsymbol{W}_1][\boldsymbol{W}_4\|\boldsymbol{W}_3, \boldsymbol{W}_1]$ | -31106.54 |
| $L = 1$ | E2: $[\boldsymbol{W}_1][\boldsymbol{W}_4\|\boldsymbol{W}_1][\boldsymbol{W}_3\|\boldsymbol{W}_4][\boldsymbol{W}_2\|\boldsymbol{W}_3]$ | -31623.11 |
| $L = 1$ | E3: $[\boldsymbol{W}_1][\boldsymbol{W}_2\|\boldsymbol{W}_1][\boldsymbol{W}_3\|\boldsymbol{W}_2][\boldsymbol{W}_4\|\boldsymbol{W}_3]$ | -31689.29 |
| $L = 2$ | E4: $[\boldsymbol{W}_1][\boldsymbol{W}_2\|\boldsymbol{W}_1][\boldsymbol{W}_4\|\boldsymbol{W}_2, \boldsymbol{W}_1][\boldsymbol{W}_3\|\boldsymbol{W}_4, \boldsymbol{W}_2]$ | -32641.49 |
| $L = 2$ | E5: $[\boldsymbol{W}_1][\boldsymbol{W}_2\|\boldsymbol{W}_1][\boldsymbol{W}_4\|\boldsymbol{W}_2, \boldsymbol{W}_1][\boldsymbol{W}_3\|\boldsymbol{W}_4, \boldsymbol{W}_1]$ | -32980.77 |

Tabela 8 – As 5 melhores especificações de acordo com a aproximação das distribuições preditivas, na escala logarítmica.

De acordo com (MADIGAN; RAFTERY, 1994), apenas os modelos cuja capacidade preditiva esteja próximo do melhor modelo deveria ser considerado. Entretanto, como pode ser visto na Tabela 8, a diferença do melhor modelo para o segundo já é elevada. Apesar da diferença, as três primeiras especificações foram consideradas, pois se destacam ainda mais das demais. Deste modo, foram ajustados quatro modelos, dois modelos sem mistura e dois modelos de mistura. Para as análises a seguir, considere as notações definidas na Tabela 9.

As configurações do MCMC utilizadas para todos os ajustes considerou ao todo 50000 iterações, com um período *burn-in* de 20000 iterações e um espaçamento de 30 iterações, resultando em cadeias finais de tamanho 1000 para cada modelo. Para esta aplicação, a suavidade foi fixada, similar ao que foi feito em Cressie e Zammit-Mangion (2016). Porém, aqui, a suavidade foi fixada em 0.5 para todos os poluentes, simplificando a função Matern em uma função exponencial. As configurações de *priori* usadas foram as mesmas dos estudos de simulação. O fator de desconto do processo Dirichlet foi fixado em 0.8. Essa

| Notação | Descrição |
|---|---|
| $M1$ | Modelo sem mistura que considera as relações entre todas as componentes. |
| $M2$ | Modelo sem mistura com a especificação do melhor modelo, segundo a distribuição preditiva (E1). |
| $M3$ | Modelo de mistura com pesos evoluindo no tempo de acordo com um processo Dirichlet considerando as três possíveis especificações (E1, E2, E3). |
| $M4$ | Modelo de mistura com pesos estáticos no tempo que considera as três possíveis especificações (E1, E2, E3). |

Tabela 9 – Notação utilizada para referenciar cada modelo considerado na análise.

escolha representa um nível flexível para a variação das estruturas mantendo um certo nível de suavidade no processo Dirichlet. Na regressão dinâmica, a matriz de covariância da equação de sistema foi estimada, assumindo que esta é estática e diagonal. Para mostrar a convergência do algoritmo MCMC, um gráfico com as cadeias da log-verossimilhança de cada modelo se encontra no Apêndice C.

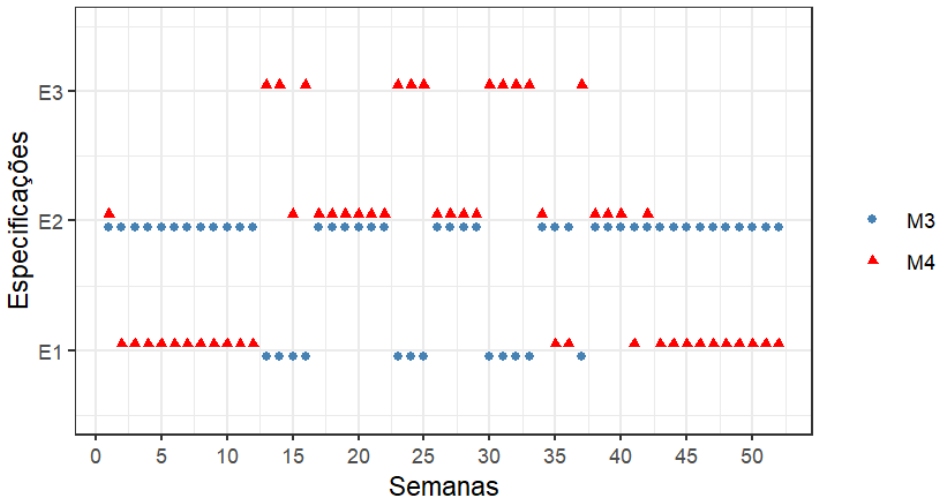



Figura 22 – Seleção das especificações consideradas nos modelos de mistura ao longo das 52 semanas analisadas.

O algoritmo de estimação utilizado para cada modelo de mistura seleciona para cada semana uma dentre as três especificações $E1$, $E2$ e $E3$. Nos ajustes, o modelo de mistura $M3$, que considera os pesos evoluindo no tempo, seguindo o processo Dirichlet, selecionou apenas duas das três especificações: $E1$ e $E2$, como componentes para a mistura, enquanto o modelo de mistura $M4$, que possui pesos estáticos, selecionou as três especificações, como pode ser visto na Figura 22. Observe que, o modelo $M3$, selecionou a especificação $E1$ nas mesmas semanas em que o modelo $M4$ selecionou a especificação $E3$, exceto pela semana 15. Essas semanas representam alguns períodos onde foram observados picos no nível de concentração dos poluentes, como podem ser visualizados nas séries temporais apresentadas nas Figuras 25, 26, 27 e 28. No caso modelo do $M3$, a especificação $E2$ foi selecionada para as semanas restantes, onde o comportamento dos dados aparentam ter

menos variabilidade, enquanto que no modelo $M4$, nessas semanas o algoritmo selecionou as especificações $E1$ e $E2$.

| | | | M3 | | M4 | | |
|---|---|---|---|---|---|---|---|
| Par. | M1 | M2 | E1 | E2 | E1 | E2 | E3 |
| $\phi_1$ | 0.716(0.502, 1.059) | 0.721(0.515, 1.069) | 1.020(0.717, 1.582) | 0.720(0.507, 1.065) | 0.66(0.461, 0.976) | 0.821(0.588, 1.211) | 1.068(0.745, 1.574) |
| $\phi_2$ | 0.078(0.054, 0.121) | 0.079(0.055, 0.115) | 0.047(0.032, 0.078) | 0.073(0.026, 0.17) | 0.101(0.07, 0.162) | 0.091(0.047, 0.208) | 0.044(0.028, 0.07) |
| $\phi_3$ | 0.070(0.020, 0.194) | 0.076(0.022, 0.207) | 0.096(0.051, 0.191) | 0.019(0.012, 0.034) | 0.03(0.016, 0.118) | 0.025(0.015, 0.05) | 0.09(0.05, 0.188) |
| $\phi_4$ | 0.024(0.011, 0.069) | 0.022(0.011, 0.061) | 0.029(0.013, 0.067) | 0.308(0.207, 0.487) | 0.024(0.011, 0.067) | 0.395(0.266, 0.68) | 0.03(0.013, 0.075) |
| $\alpha_1$ | $< 10^{-3}(0, 0.018)$ | $< 10^{-3}(0, 0.017)$ | $< 10^{-3}(0, 0.134)$ | $< 10^{-3}(0, 0.014)$ | $< 10^{-3}(0, 0.015)$ | $< 10^{-3}(0, 0.017)$ | 0.001(0, 0.149) |
| $\alpha_2$ | $< 10^{-3}(0, 0.024)$ | $< 10^{-3}(0, 0.027)$ | $< 10^{-3}(0, 0.210)$ | 1.443(0, 3.615) | $< 10^{-3}(0, 0.02)$ | 2.049(0.679, 4.806) | $< 10^{-3}(0, 0.198)$ |
| $\alpha_3$ | 1.358(0, 4.145) | 1.554192(0, 4.299) | 2.931(1.452, 6.173) | 0.001(0, 0.81) | 0.070(0, 2.759) | 0.001(0, 1.225) | 2.658(1.314, 5.46) |
| $\alpha_4$ | 1.769(0, 6.571) | 0.981703(0, 5.001) | 2.213(0.358, 6.082) | 0.001(0, 0.156) | 1.458(0, 5.118) | 0.065(0, 0.48) | 2.334(0.567, 6.086) |
| $A_{12}$ | 0.585(0.477, 0.724) | 0.578(0.475, 0.727) | 0.553(0.466, 0.717) | 0.153(−0.172, 0.292) | 0.588(0.467, 0.766) | 0.175(−0.018, 0.357) | 0.547(0.465, 0.716) |
| $A_{13}$ | −0.135(−0.404, 0.369) | −0.103(−0.383, 0.197) | 0.281(−0.052, 0.542) | – | −0.128(−0.233, 0.161) | – | – |
| $A_{23}$ | 0.506(−0.077, 0.940) | 0.532(0.163, 0.918) | 0.031(−0.315, 0.694) | 0.944(0.704, 1.213) | 0.331(−0.008, 0.782) | 1.021(0.765, 1.257) | 0.459(0.321, 0.702) |
| $A_{14}$ | 0.052(−0.418, 0.460) | 0.040(−0.447, 0.446) | 0.013(−0.286, 0.271) | – | 0.001(−0.395, 0.39) | – | – |
| $A_{24}$ | 0.034(−0.493, 0.561) | – | – | – | – | – | – |
| $A_{34}$ | 0.766(0.469, 1.031) | 0.743(0.536, 0.993) | 0.833(0.638, 1.09) | 0.149(−0.464, 0.699) | 0.751(0.515, 1.038) | −0.043(−0.517, 0.608) | 0.852(0.654, 1.077) |
| $r_{12}$ | 0.239(0.074, 0.488) | 0.279(0.081, 0.528) | 0.075(0.045, 0.128) | 0.233(0.071, 8.914) | 0.271(0.079, 0.57) | 0.154(0.04, 0.545) | 0.072(0.043, 0.118) |
| $r_{13}$ | 3.722(0.076, 9.674) | 1.791(0.101, 10.350) | 0.074(0.036, 0.989) | – | 3.59(0.131, 10.732) | – | – |
| $r_{23}$ | 0.102(0.045, 1.807) | 0.097(0.041, 0.258) | 0.312(0.054, 5.573) | 0.166(0.05, 0.401) | 0.141(0.06, 2.195) | 0.206(0.047, 0.447) | 0.06(0.038, 0.094) |
| $r_{14}$ | 0.322(0.085, 22.761) | 0.252(0.077, 6.923) | 0.485(0.073, 7.832) | – | 0.386(0.078, 6.249) | – | – |
| $r_{24}$ | 0.316(0.081, 14.107) | – | – | – | – | – | – |
| $r_{34}$ | 0.051(0.022, 0.333) | 0.059(0.025, 0.397) | 0.05(0.022, 0.094) | 0.609(0.146, 3.447) | 0.044(0.019, 0.506) | 0.735(0.127, 7.287) | 0.048(0.021, 0.083) |
| $\Delta_{12}^1$ | 0.008(−0.059, 0.070) | 0.012(−0.054, 0.087) | 0.008(−0.01, 0.022) | 0.011(−0.366, 0.351) | −0.018(−0.115, 0.036) | 0.018(−0.051, 0.124) | 0.006(−0.01, 0.02) |
| $\Delta_{13}^1$ | 0.007(−0.56, 0.644) | −0.032(−0.6, 0.606) | 0.013(−0.41, 0.233) | – | 0.031(−0.568, 0.607) | – | – |
| $\Delta_{23}^1$ | 0.018(−0.204, 0.283) | 0.018(−0.008, 0.056) | 0.017(−0.576, 0.559) | 0.007(−0.054, 0.054) | 0.015(−0.14, 0.327) | 0.003(−0.062, 0.07) | 0.014(−0.001, 0.026) |
| $\Delta_{14}^1$ | −0.249(−0.646, 0.893) | −0.273(−0.619, 0.738) | −0.007(−0.668, 0.755) | – | −0.079(−0.642, 0.588) | – | – |
| $\Delta_{24}^1$ | −0.124(−0.65, 0.748) | – | – | – | – | – | – |
| $\Delta_{34}^1$ | 0.001(−0.025, 0.052) | 0.002(−0.051, 0.058) | 0.003(−0.009, 0.015) | −0.031(−0.594, 0.522) | 0.003(−0.058, 0.04) | −0.037(−0.646, 0.589) | 0.002(−0.009, 0.012) |
| $\Delta_{12}^2$ | −0.007(−0.072, 0.069) | $< 10^{-3}(-0.068, 0.086)$ | −0.003(−0.022, 0.026) | 0.069(−0.372, 0.346) | −0.007(−0.08, 0.083) | 0.037(−0.042, 0.128) | −0.003(−0.022, 0.024) |
| $\Delta_{13}^2$ | 0.043(−0.472, 0.704) | −0.004(−0.604, 0.691) | 0.025(−0.287, 0.166) | – | 0.088(−0.534, 0.665) | – | – |
| $\Delta_{23}^2$ | 0.049(−0.25, 0.201) | 0.0488(−0.019, 0.083) | −0.071(−0.653, 0.561) | −0.001(−0.075, 0.052) | 0.055(−0.25, 0.191) | 0(−0.065, 0.057) | 0.015(−0.008, 0.04) |
| $\Delta_{14}^2$ | −0.097(−0.527, 0.619) | −0.101(−0.510, 0.569) | −0.039(−0.623, 0.653) | – | 0.029(−0.598, 0.751) | – | – |
| $\Delta_{24}^2$ | 0.004(−0.593, 0.972) | – | – | – | – | – | – |
| $\Delta_{34}^2$ | $< 10^{-3}(-0.024, 0.053)$ | $< 10^{-3}(-0.039, 0.067)$ | −0.001(−0.014, 0.017) | −0.003(−0.631, 0.593) | −0.003(−0.033, 0.054) | 0.003(−0.63, 0.614) | −0.001(−0.014, 0.017) |
| $\sigma^2$ | 0.081(0.079, 0.083) | 0.081(0.079, 0.083) | 0.067(0.065, 0.069) | | 0.060(0.059, 0.062) | | |

Tabela 10 – Estimativas pontuais dos hiperparâmetros dos processos espaciais em cada um dos modelos e da variância $\sigma^2$. Entre parênteses temos os limites dos intervalos de credibilidade de 95%.

Na Tabela 10 são apresentadas as estimativas dos hiperparâmetros dos processos espaciais de cada modelo. Na Figura 23 essas estimativas podem ser visualizadas com seus respectivos intervalos de credibilidade. É possível observar que as estimativas dos dois modelos sem mistura foram muito parecidas e repare que, nos modelos de mistura, cada especificação apresentou estimativas para pelo menos alguns hiperparâmetros, indicando a construção de diferentes estruturas no espaço. Observe também que as especificações $E2$ apresentaram estimações parecidas, assim como a especificação $E1$ do modelo $M3$ e $E3$ de $M4$, que foram selecionadas nas mesmas semanas. Os modelos estimaram um alcance espacial pequeno para $NO$ e $NO_2$ e maiores para $PM_{10}$ e $PM_{2.5}$, com os casos mais extremos ocorrendo na estimação das especificações $E1$ e $E3$ dos modelos $M3$ e $M4$, respectivamente. Isso significa que nas semanas onde as concentrações foram mais fortes, que são justamente as semanas onde essas especificações foram selecionadas, a autocorrelação de $NO$ e $NO_2$ é restrita a apenas localizações muito próximas, enquanto que para $PM_{10}$ e $PM_{2.5}$ a autocorrelação é significativa mesmo para localizações mais distantes. Em particular nos modelos de mistura, a especificação $E2$, em ambos os casos, estimou um alcance espacial mais alto para $NO_2$ e mais baixo para $PM_{10}$ e $PM_{2.5}$, diferente das outras especificações e dos outros modelos.

Em relação aos hiperparâmetros das funções de interação, nota-se que, pelas estima-

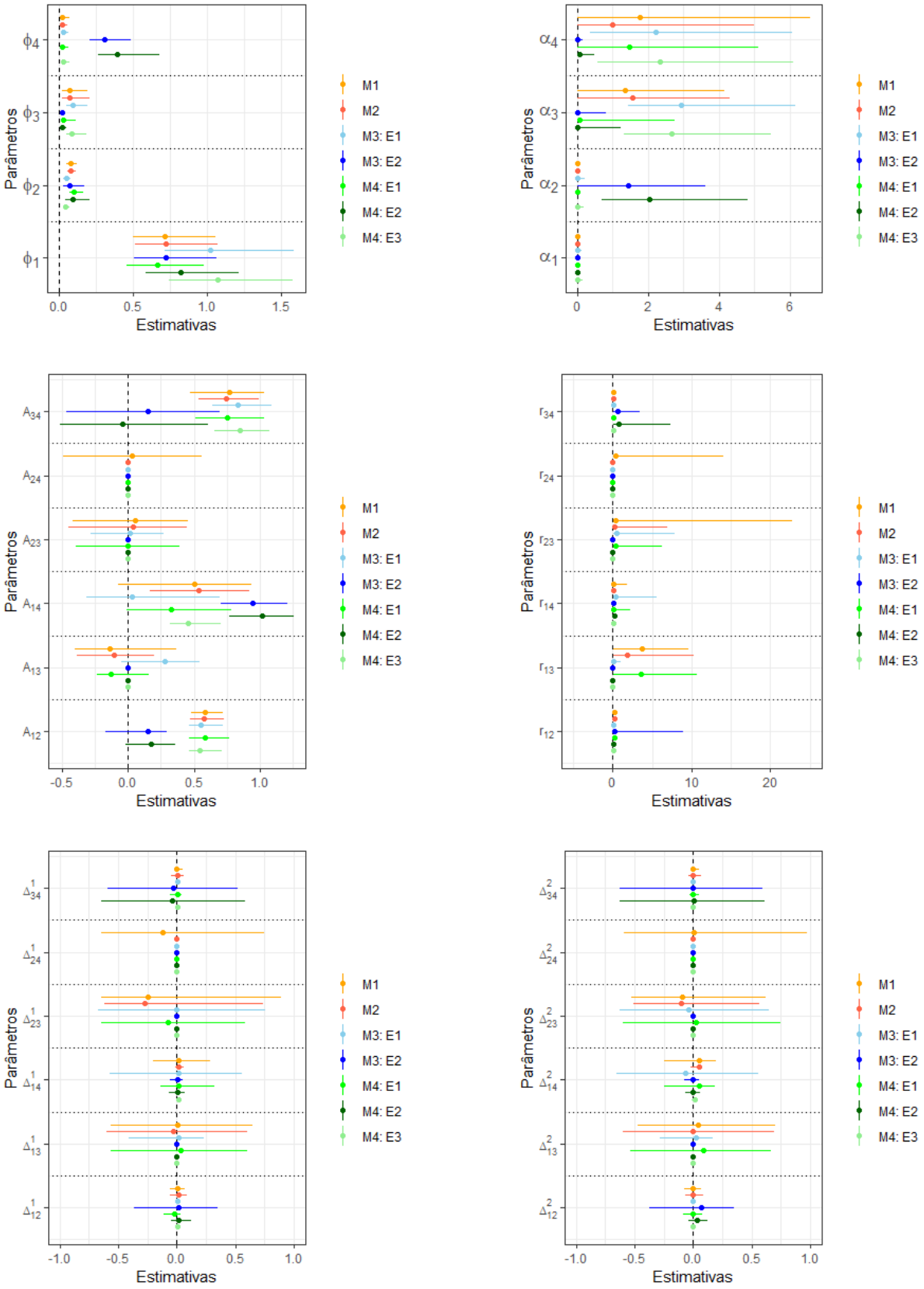


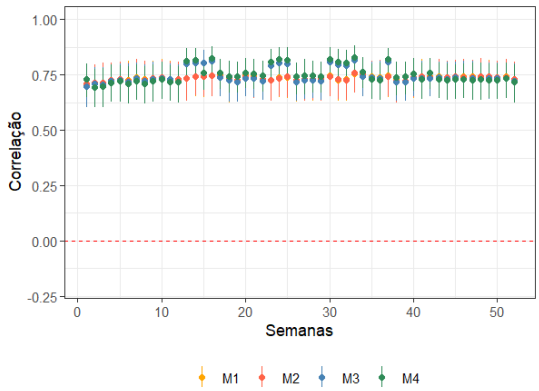

Figura 23 – Estimativas dos hiperparâmetros dos processos espaciais em cada um dos modelos, com os respectivos intervalos de credibilidade de 95%.

tivas do parâmetro de escala $A$, os modelos indicaram uma relação mais forte entre $NO$ e $NO_2$ e entre $PM_{10}$ e $PM_{2.5}$, porém nas semanas onde os modelos de mistura selecionaram a especificação $E2$, essas relações não foram significativas. Por outro lado, nesses mesmos períodos a relação entre $NO_2$ e $PM_{10}$ foi bem forte, de acordo com as estimativas de $A_{23}$ nos modelos $M3$ e $M4$. Para os parâmetros de assimetria: $\Delta_1, \Delta_2$, nenhum dos modelos estimou que algum deles foi significativamente diferente de 0. Os modelos de mistura estimaram uma variância ($\sigma^2$) menor do que os modelos sem mistura. Isso significa que os componentes dos modelos de mistura conseguiram capturar mais incerteza no espaço e no tempo das variáveis respostas do que os componentes dos modelos sem mistura.

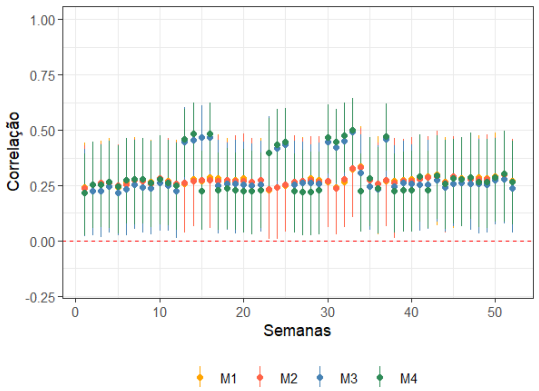
Na Figura 24 podem ser visualizadas as correlações cruzadas estimadas ao longo de todas as estações de monitoramento consideradas no ajuste dos modelos, para cada semana analisada. É evidente que os modelos sem mistura não apresentaram a mesma flexibilidade dos modelos de mistura, uma vez que as todas correlações cruzadas estimadas pelos modelos $M1$ e $M2$ seguiram quase constantes ao longo das semanas, enquanto que os modelos $M3$ e $M4$, por outro lado, foram capazes de estimar correlações diferentes para cada semana, de acordo com a especificação selecionada, embora, sendo rigoroso, a diferença em si não seja estatisticamente significativa, por conta dos intervalos de credibilidade. Entre os modelos de mistura, observa-se um padrão muito parecido, com ambos os modelos apresentando aumentos de correlações nas mesmas semanas, em que as especificações $E1$ no modelo $M3$ e $E3$ no modelo $M4$ foram selecionadas, exceto na semana 15, quando a correlação estimada por $M4$ foi bem menor. Nas demais semanas, há uma pequena diferença nas correlações estimadas para $M3$ e $M4$, que ocorre pelo fato do modelo $M4$ ter selecionado uma terceira especificação para modelagem. Essas diferenças são mais evidentes nas correlações cruzadas entre $NO_2$ e $PM_{10}$ e entre $PM_{10}$ e $PM_{2.5}$. Entretanto, num geral as diferenças nas estimativas do modelo $M4$ são pequenas, e não parecem justificar a seleção de uma terceira estrutura de correlação para a modelagem dessas semanas.

As estimativas dos coeficientes da regressão dinâmica dos efeitos da longitude e da latitude sobre os poluentes podem ser visualizados no Apêndice C, nas Figuras 50 e 51, respectivamente. De modo geral, a variação desses resultados para cada modelo foi baixa. A latitude teve um efeito positivo significativo sobre todos os poluentes, e apresentou bastante variabilidade ao longo das semanas, apresentando estimativas mais altas em semanas em que ocorreram picos de concentração dos poluentes. Por outro lado, a longitude parece ter efeitos poucos significativos sobre os poluentes, e, além disso, apresenta pouca variabilidade ao longo das semanas.

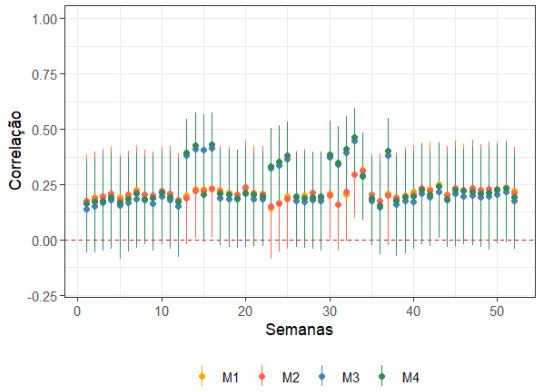
As Figuras 25, 26, 27 e 28 apresentam a série temporal da concentração de cada poluente no bairro de Horninglow (Inglaterra), que foi selecionada aleatoriamente. As curvas ajustadas e os intervalos preditivos obtidos pelos modelos sem mistura são idênticos, e o mesmo ocorre entre os modelos com mistura. De um modo geral, é possível perceber que



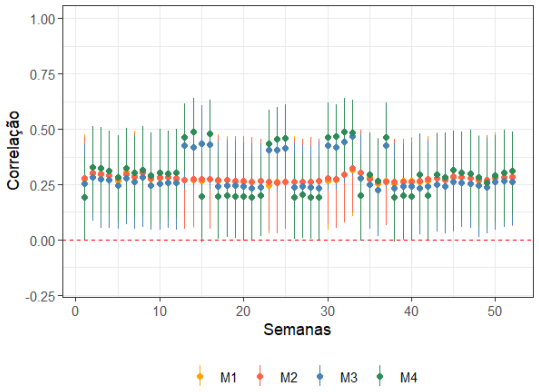

(a) $NO \times NO_2$.



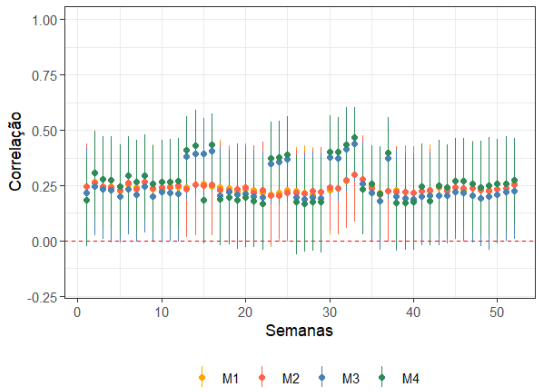

(b) $NO \times PM_{10}$.



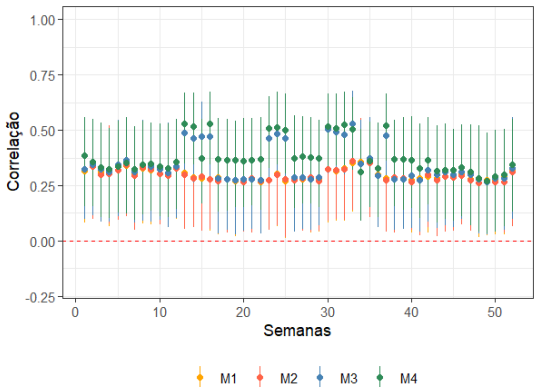

(c) $NO \times PM_{2.5}$.




(d) $NO_2 \times PM_{10}$.




(e) $NO_2 \times PM_{2.5}$.




(f) $PM_{10} \times PM_{2.5}$.

Figura 24 – Correlações cruzadas do logaritmo da concentração de cada poluente ao longo de todas as localizações no decorrer das 52 semanas analisadas, de acordo com cada modelo.

os modelos parecem ter se ajustado bem aos dados, gerando curvas que seguem as tendências dos dados e procuram capturar os picos de concentração em algumas semanas. Nesse sentido, os modelos de mistura foram capazes de se ajustar melhor na presença de picos de concentração, com suas curvas ajustadas se aproximando mais dos dados. Na Figura 25 é onde fica mais evidente essa diferença. Note que, nas semanas 24 e 25, os intervalos preditivos dos modelos sem mistura não incluem as concentrações de $NO$ observadas, enquanto que nos modelos de mistura, esses dados estão contidos nos intervalos.

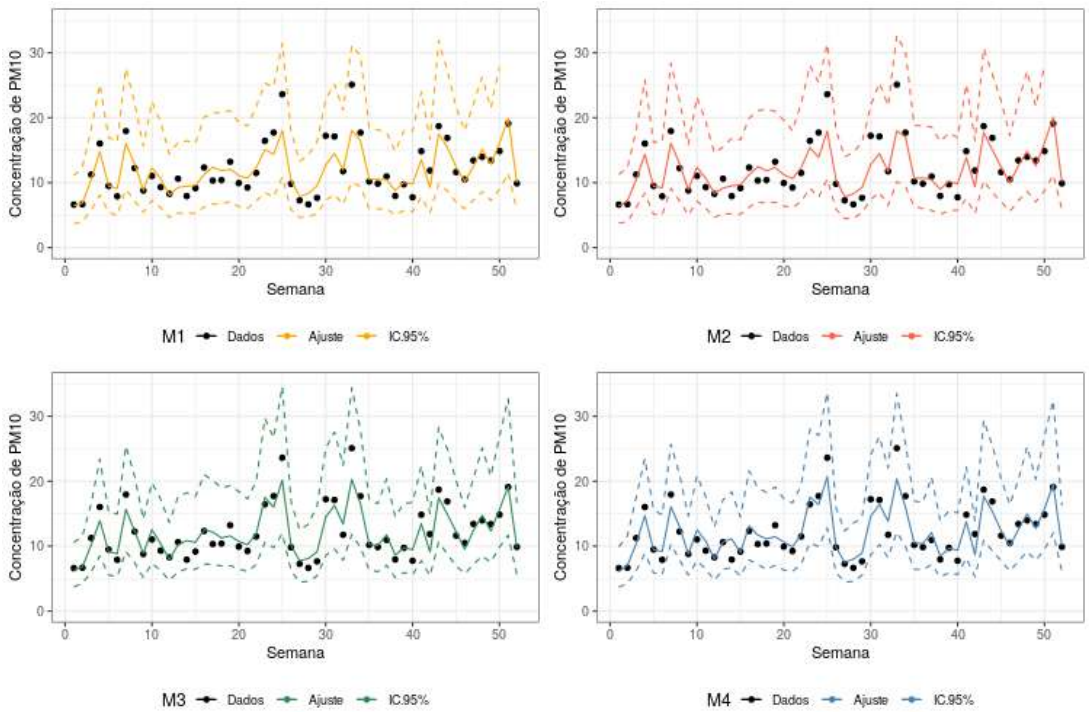



Figura 27 – Série temporal da concentração de $PM_{10}$ na estação localizada em Horninglow, na Inglaterra. Os pontos pretos indicam os dados observados. A linha contínua representa a curva ajustada pelo o modelo e as linhas tracejadas o intervalo preditivo de 95%.

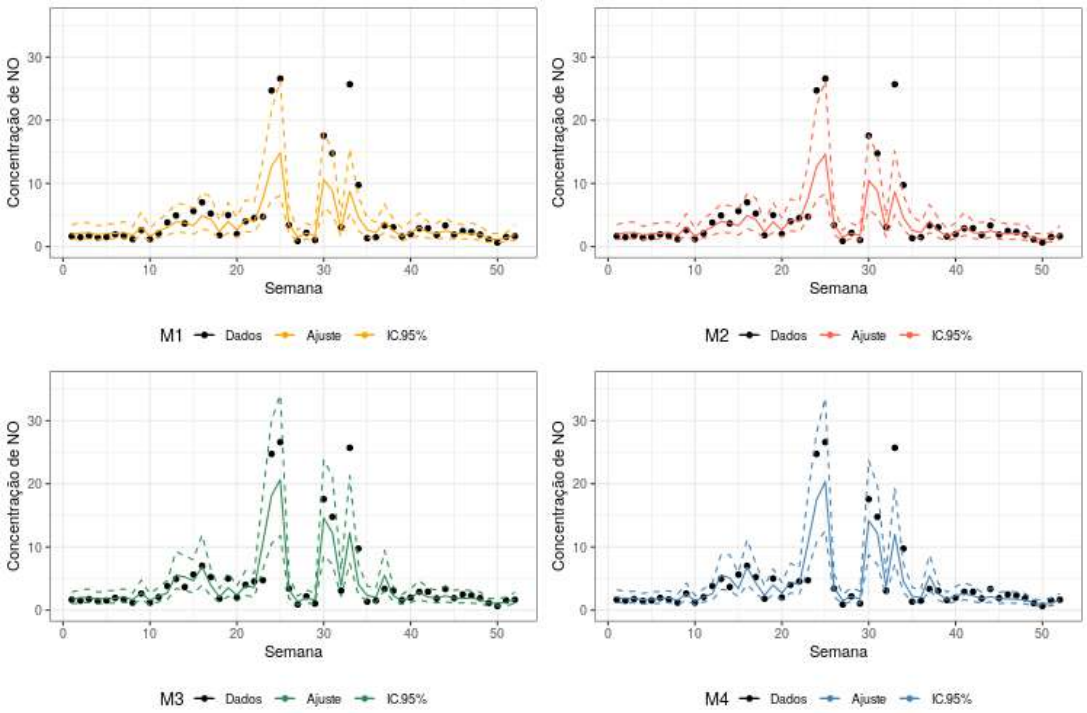



Figura 25 – Série temporal da concentração de $NO$ na estação localizada em Horninglow, na Inglaterra. Os pontos pretos indicam os dados observados. A linha contínua representa a curva ajustada pelo o modelo e as linhas tracejadas o intervalo preditivo de 95%.

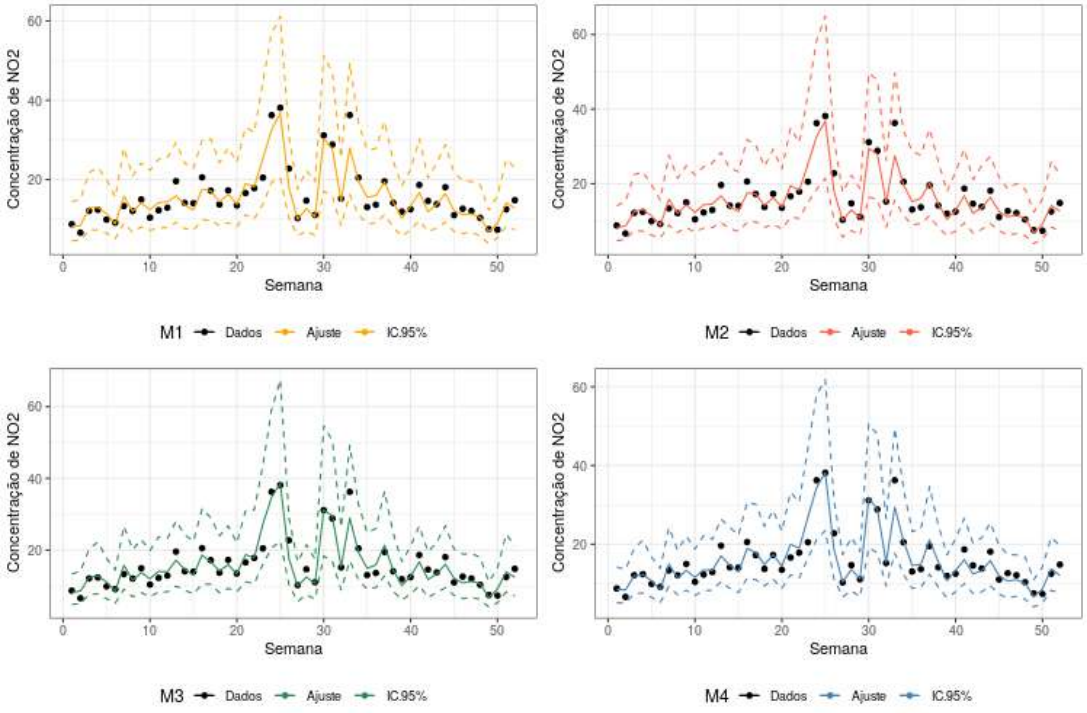



Figura 26 – Série temporal da concentração de $NO_2$ na estação localizada em Horninglow, na Inglaterra. Os pontos pretos indicam os dados observados. A linha contínua representa a curva ajustada pelo o modelo e as linhas tracejadas o intervalo preditivo de 95%.

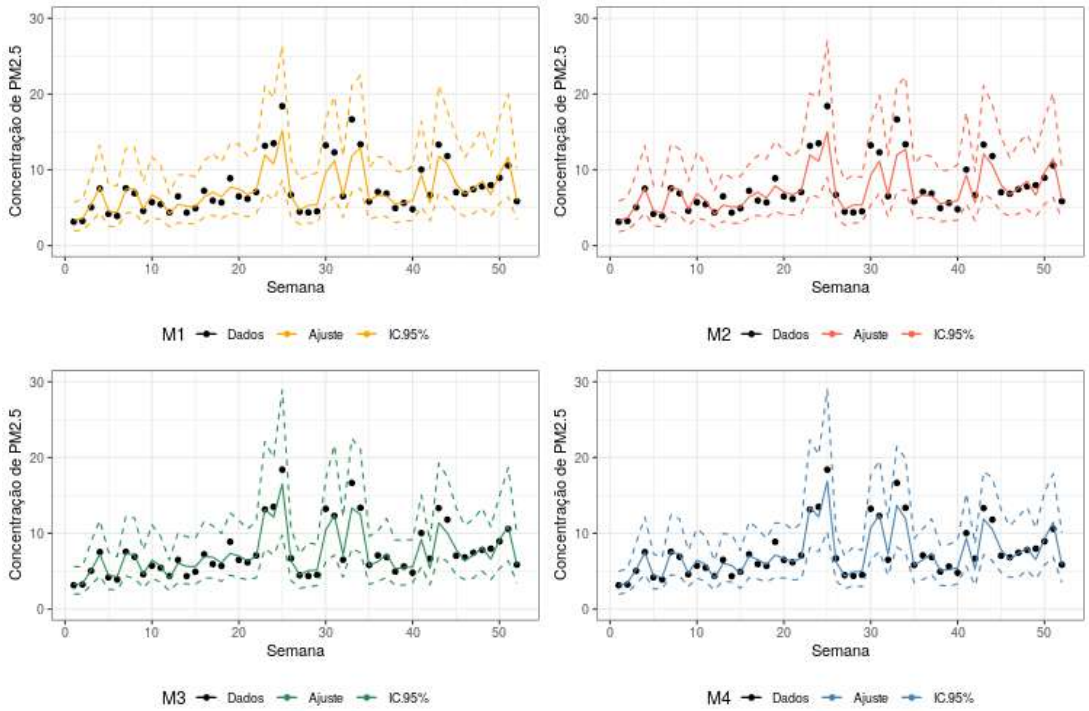



Figura 28 – Série temporal da concentração de $PM_{2.5}$ na estação localizada em Horninglow, na Inglaterra. Os pontos pretos indicam os dados observados. A linha contínua representa a curva ajustada pelo o modelo e as linhas tracejadas o intervalo preditivo de 95%.

A Tabela 11 mostra os erros de predição obtidos por cada modelo, de acordo com as medidas RMSE, MAE e IS (Interval Score). O modelo de mistura com os pesos estáticos ($M4$) foi quem apresentou o menor RMSE e MAE, mas a diferença entre os modelos foi pequena, o que é um indicativo que mesmo os modelos sem mistura apresentaram um bom nível de predição. Para o Interval Score a diferença entre os modelos foi maior, com o modelo de mistura com pesos evoluindo no tempo de acordo com um processo Dirichlet ($M3$) se destacando com o menor Interval Score.

| Medidas | $M1$ | $M2$ | $M3$ | $M4$ |
|---|---|---|---|---|
| RMSE | 0.8419 | 0.8439 | 0.8492 | **0.8381** |
| MAE | 0.6082 | 0.6019 | 0.6053 | **0.5942** |
| IS | 2.9205 | 2.9692 | **2.6101** | 2.9830 |

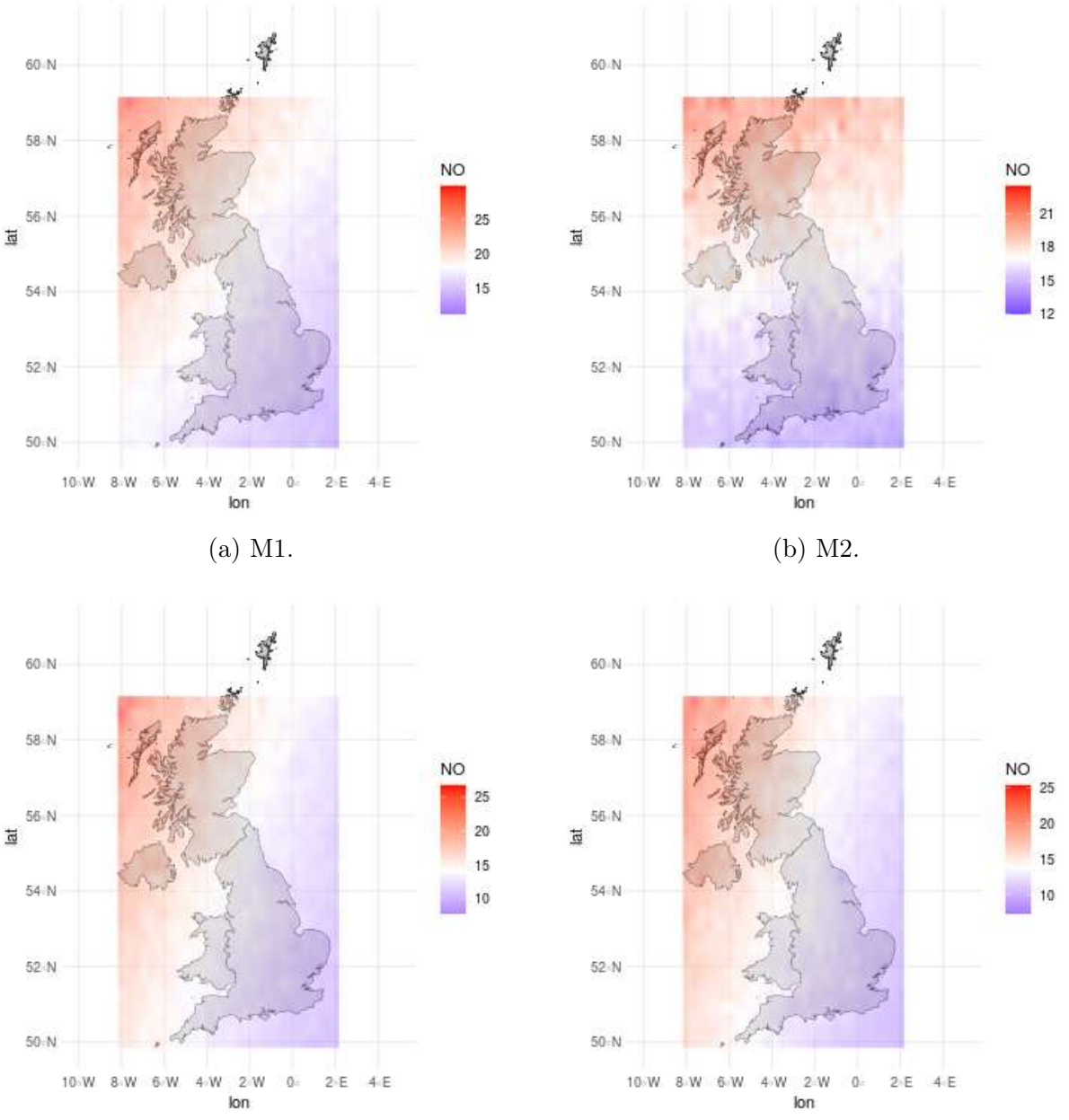
Tabela 11 – Medidas de erros obtidas por cada modelo ajustado ao longo das 12 localizações que não foram utilizadas nos ajustes.

As Figuras 29, 30, 31 e 32 apresentam a interpolação das concentrações dos poluentes dentro de um retângulo ao redor do Reino Unido, na 25ª semana, que é uma semana em que foram observados picos de concentrações dos poluentes e os modelos apresentam algumas diferenças. No apêndice C são apresentados as interpolações para a semana 1 e para a semana 52, que é a última semana analisada.

Os modelos de mistura apresentaram padrões espaciais muito parecidos entre si em praticamente todos os cenários (poluente e semana). Os modelos sem mistura destoam um pouco mais nas interpolações para $NO$, mas estão bem parecidos nas interpolações dos demais poluentes. Ao longo das semanas apresentadas, é perceptível que para os poluentes $NO_2$, $PM_{10}$ e $PM_{2.5}$ as direções se mantém mais estáveis, apresentando menos oscilações nas regiões onde são estimadas as concentrações mais fortes. Além disso, em todas as três semanas, pode-ser perceber que as interpolações para o $NO$ apresentam um comportamento menos suave em relação aos outros poluentes.

Nota-se que, na primeira semana, as concentrações mais fortes de $NO$ ocorrem na região oeste do retângulo, e as concentrações mais baixas na direção oposta. Já para os outros três poluentes, esse comportamento é invertido, com as concentrações mais altas ocorrendo nas regiões a leste do retângulo. Na semana 25, esse comportamento muda para $NO$. As concentrações mais altas passam a ocorrer mais ao norte ($M2$) ou noroeste ($M1$, $M3$ e $M4$). No mesmo período para $NO_2$, observa-se um comportamento similar, as concentrações mais altas estão na parte nordeste do retângulo. Já na semana 52, $NO_2$ volta a apresentar concentrações mais fortes nas regiões leste. Por outro lado, para $NO$ o comportamento para cada modelo, com $M1$ e $M4$ estimando concentrações mais fortes nas regiões norte, $M2$ nas leste e $M3$ nas regiões nordeste. Os outros dois poluentes, $PM_{10}$ e $PM_{2.5}$ continuaram apresentando concentrações mais fortes na parte leste do retângulo, tanto na semana 25 quanto na semana 52.

Pode-se destacar também, o comportamento das interpolações de cada modelo para $PM_{10}$ e $PM_{2.5}$. Os dois poluentes apresentam padrões parecidos. Na primeira semana, os modelo sem mistura indicam um nível de concentração mais alto na Inglaterra e no sul do País de Gales e mais baixo na Escócia de ambos os poluentes, padrão esse que segue para as outras duas semanas apresentadas. Os modelos de mistura apresentam um padrão similar para $PM_{2.5}$ e indicam índices de concentração mais próximos da mediana nos países para $PM_{10}$, na primeira semana. O mesmo comportamento pode ser observado na semana 52. Já na semana 25, que é a semana onde os picos de concentração foram observados, nota-se claramente que os modelos de mistura apresentam um pico de concentração desses poluentes muito mais forte na Inglaterra e no sul do País de Gales. Ao mesmo tempo foi estimado concentrações mais baixas na Escócia. Neste sentido, os modelos de mistura apresentam mais flexibilidade, mostrando que são capazes de alterar o padrão espacial para se adequar melhor aos dados, principalmente na presença de uma mudança mais acentuada nas observações de um período para outro.



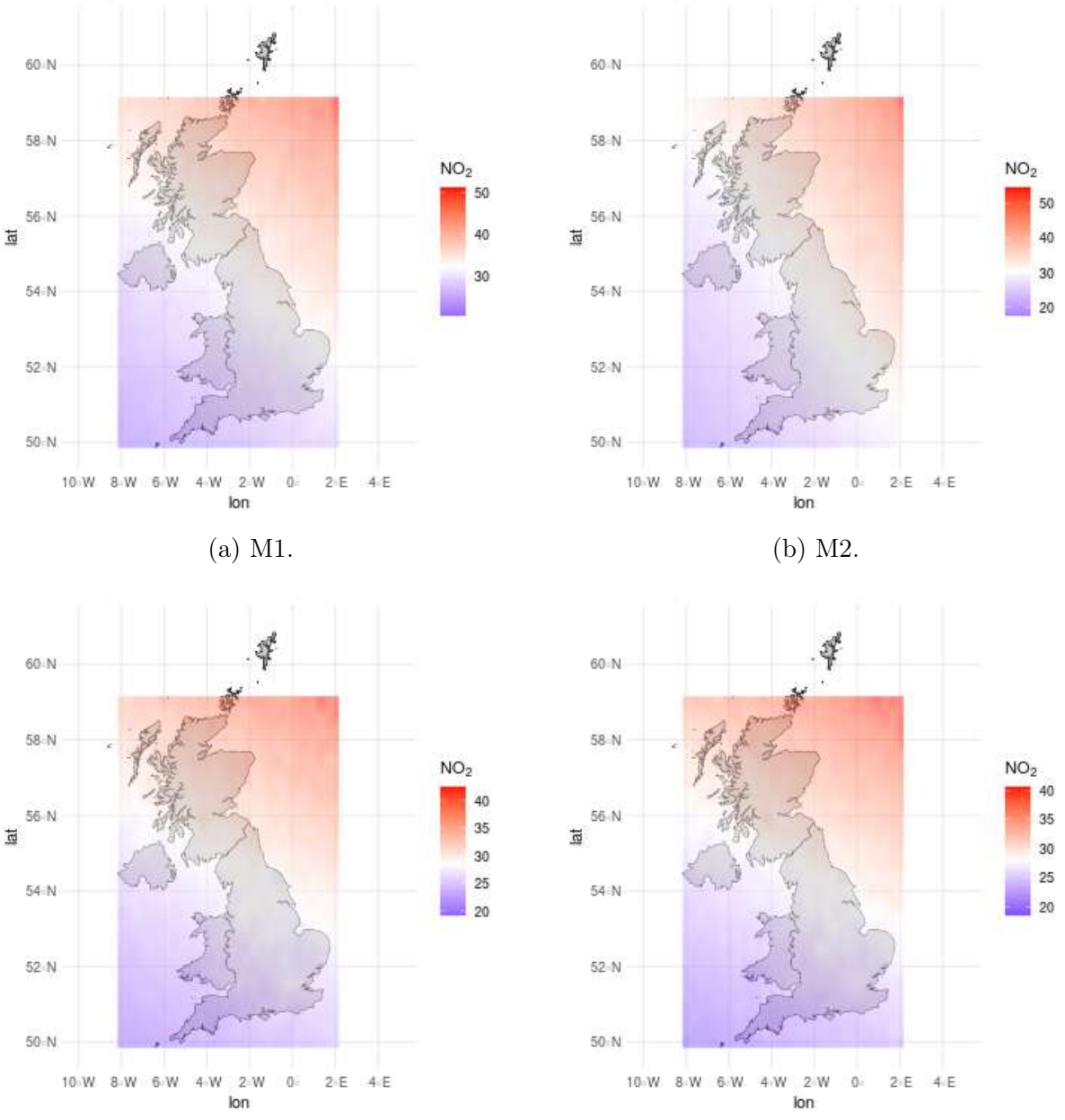

(a) M1.

(b) M2.

(c) M3.

(d) M4.

Figura 29 – Interpolação da concentração de $NO$ ao longo de um grid regular construído ao redor do Reino Unido, na 25ª semana do período analisado. A cor azul indica níveis mais baixos, a cor branca representa níveis intermediários e a cor vermelha, níveis elevados de concentração do poluente.



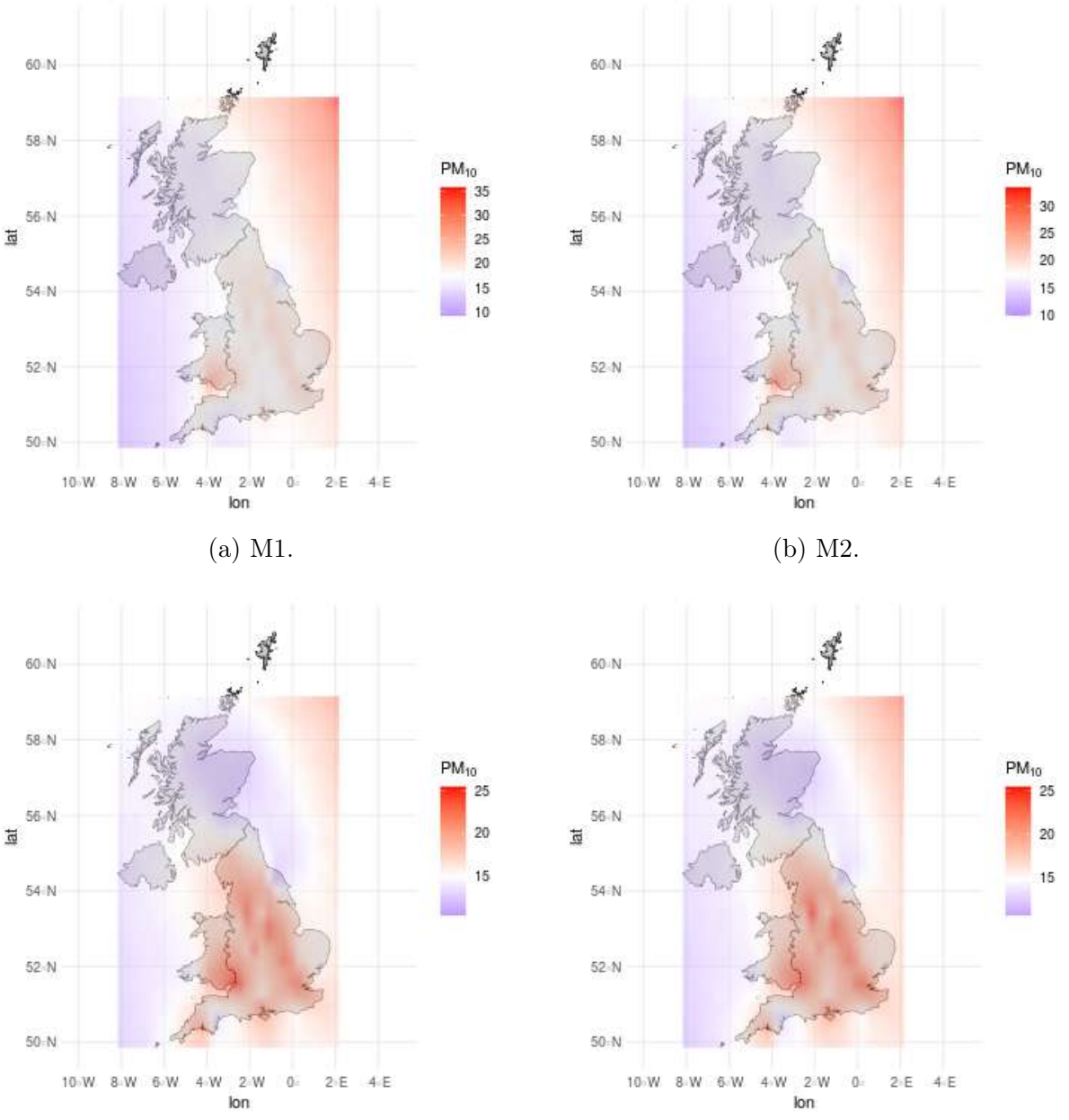

(a) M1.

(b) M2.

(c) M3.

(d) M4.

Figura 30 – Interpolação da concentração de $NO_2$ ao longo de um grid regular construído ao redor do Reino Unido, na 25ª semana do período analisado. A cor azul indica níveis mais baixos, a cor branca representa níveis intermediários e a cor vermelha, níveis elevados de concentração do poluente.




(a) M1.




(b) M2.

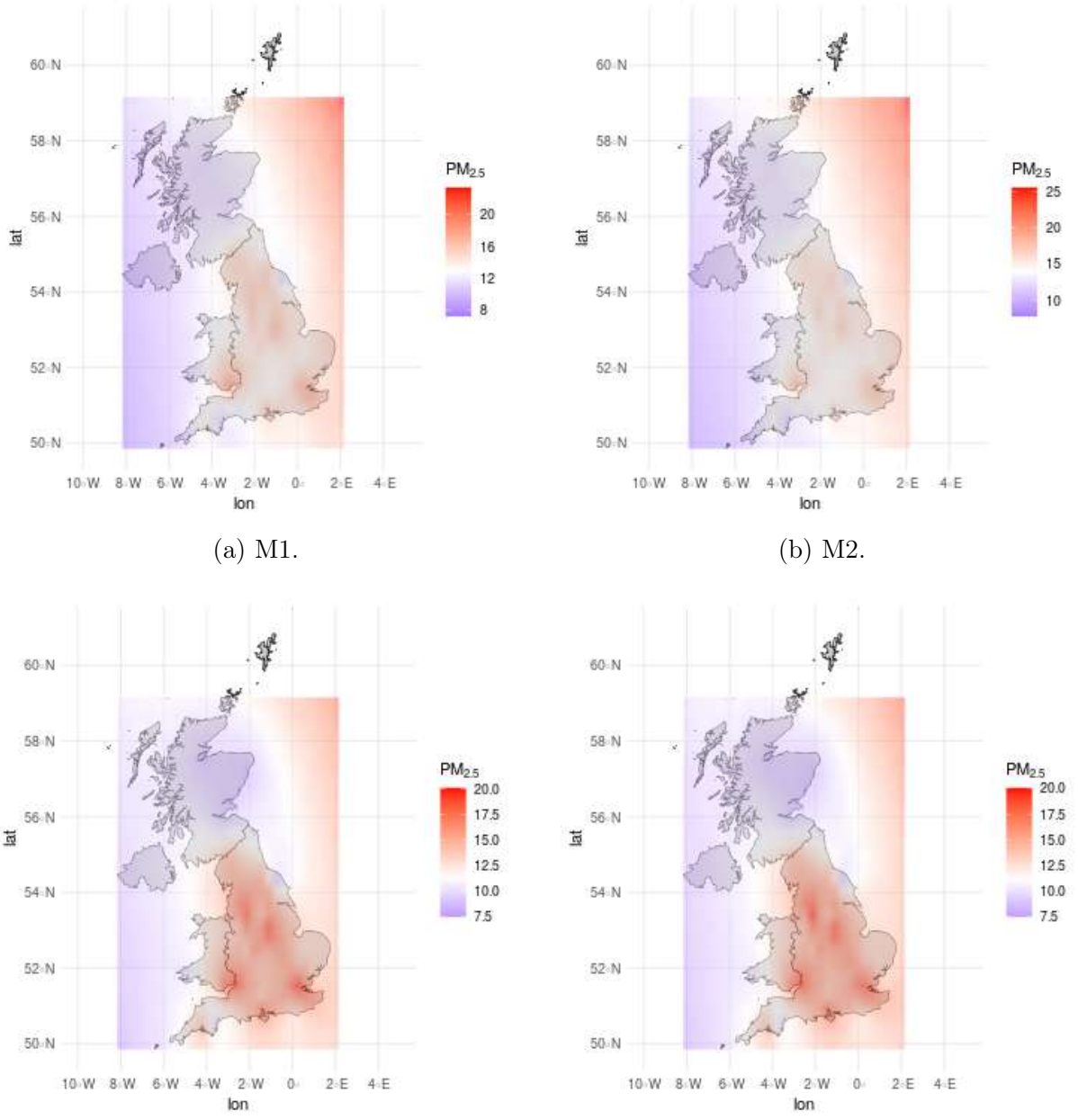



(c) M3.




(d) M4.

Figura 31 – Interpolação da concentração de $PM_{10}$ ao longo de um grid regular construído ao redor do Reino Unido, na 25ª semana do período analisado. A cor azul indica níveis mais baixos, a cor branca representa níveis intermediários e a cor vermelha, níveis elevados de concentração do poluente.




(a) M1.

(b) M2.

(c) M3.

(d) M4.

Figura 32 – Interpolação da concentração de $PM_{2.5}$ ao longo de um grid regular construído ao redor do Reino Unido, na 25ª semana do período analisado. A cor azul indica níveis mais baixos, a cor branca representa níveis intermediários e a cor vermelha, níveis elevados de concentração do poluente.

# 6 CONCLUSÃO

Os modelos multivariado espaço-temporais usualmente utilizados, que levam em consideração uma única estrutura de correlação, podem não ser opções viáveis em cenários onde a relação entre as variáveis de interesse pode ser alterada ao longo do tempo. Como pode ser o caso dos poluentes atmosféricos. Sob uma abordagem Bayesiana, este trabalho teve como objetivo propor um modelo com flexibilidade para acomodar diferentes estruturas de dependência no espaço ao longo do tempo. Para isso, foi desenvolvido um modelo de mistura, em que cada modelo considerado nessa mistura foi especificado sob uma estrutura de covariância diferente, por meio de uma abordagem condicional. Além disso, foi permitido que os pesos dessa mistura variem ao longo do tempo, seguindo um processo de evolução Dirichlet, para garantir que diferentes estruturas de covariância possam ser selecionadas em diferentes períodos de tempo.

Um estudo de simulação foi realizado para comparar um modelo completo, que considera a relação entre todas as variáveis, e um modelo simplificado que induz esparsidade ao limitar as relações atribuídas na especificação das condicionais. Este estudo foi organizado em quatro cenários, e em todos o modelo esparso se mostrou bastante competitivo em relação ao modelo completo, apresentando erros de predição muito próximos. Além disso, os gráficos de interpolação mostraram que tanto o modelo completo quanto o modelo esparso foram capazes de recuperar o padrão espacial verdadeiro, usado na geração dos dados, indicando que ambos os modelos obtiveram bons ajustes.

Em seguida, outro estudo de simulação foi realizado. Desta vez para avaliar o modelo de mistura proposto no trabalho e comparar com outro modelo de mistura, onde os pesos eram estáticos no tempo e um modelo sem mistura. Assim como o primeiro estudo de simulação, este também foi organizado em diferentes cenários. No geral, os modelos de mistura apresentaram resultados superiores ao modelo sem mistura, com erros de predições menores em todos os cenários e gráficos de interpolações mais precisos com relação ao padrão espacial verdadeiro. Também é preciso destacar que o modelo de mistura com pesos evoluindo de acordo com um processo Dirichlet se mostrou mais robusto à variação da estrutura de covariância do que o modelo de mistura com pesos estáticos.

Na aplicação dos modelos em dados reais de poluentes atmosféricos, foi possível observar alguns resultados positivos na comparação entre os modelos de mistura com os modelos sem mistura. Primeiramente, os modelos de mistura apresentaram medidas erros de predição menores do que do que os modelos sem mistura, indicando uma capacidade preditiva melhor para esses dados. Além disso, foi observado que os modelos de mistura conseguiram lidar melhor com os picos de concentração que ocorreram em algumas semanas do que os modelos sem mistura, demonstrando mais flexibilidade e produzindo intervalos preditivos melhores. Pode-se destacar também que os modelos de mistura

apresentam mais flexibilidade nas interpolações, demonstrando serem capaz de alterar o padrão espacial para se adequar melhor aos dados. Por fim, o modelo de mistura com pesos evoluindo no tempo de acordo com um processo Dirichlet selecionou as estruturas espaciais de forma mais suave do que o modelo de mistura com pesos estáticos no tempo, o que é um resultado positivo, uma vez que uma estrutura alternativa foi selecionada apenas quando os dados realmente apresentaram um comportamento mais atípico.

Um possível projeto futuro seria uma extensão do modelo proposto para lidar com dados faltantes, que é um problema muito comum para dados de poluentes atmosféricos. Uma das principais vantagens dos modelos multivariados é que se pode obter mais informações de uma variável a partir das outras variáveis modeladas, principalmente quando existe correlação entre elas. Além disso, a inferência Bayesiana tem uma forma bem natural de lidar com esse tipo de situação, através das distribuições preditivas. Além disso, o fator de desconto do Processo de Evolução Dirichlet foi tratado como uma constante fixa neste trabalho, mas poderia ser estimada e indicaria o nível de volatilidade da estrutura espacial presente nos dados ao longo do tempo.

# REFERÊNCIAS


BANERJEE, S.; CARLIN, B. P.; GELFAND, A. E. **Hierarchical modeling and analysis for spatial data**. [S.l.]: CRC press, 2014.

CARSLAW, D. C.; ROPKINS, K. openair — an r package for air quality data analysis. **Environmental Modelling  Software**, v. 27–28, n. 0, p. 52–61, 2012. ISSN 1364-8152.

CHIB, S.; GREENBERG, E. Understanding the metropolis-hastings algorithm. **The american statistician**, Taylor & Francis, v. 49, n. 4, p. 327–335, 1995.

CHILES, J.-P.; DELFINER, P. **Geostatistics: modeling spatial uncertainty**. [S.l.]: John Wiley & Sons, 2012. v. 713.

COX, T.; COX, M. **Multidimensional Scaling, Second Edition**. [S.l.]: Taylor & Francis, 2000. (Chapman & Hall/CRC Monographs on Statistics & Applied Probability). ISBN 9781584880943.

CRESSIE, N. **Statistics for spatial data**. [S.l.]: John Wiley & Sons, 1993.

CRESSIE, N.; ZAMMIT-MANGION, A. Multivariate spatial covariance models: a conditional approach. **Biometrika**, Oxford University Press, v. 103, n. 4, p. 915–935, 2016.

DATTA, A. et al. Hierarchical nearest-neighbor gaussian process models for large geostatistical datasets. **Journal of the American Statistical Association**, Taylor & Francis, v. 111, n. 514, p. 800–812, 2016.

DIGGLE, P.; JR, P. J. R. **Model-Based Geostatistics**. [S.l.]: Springer, 2007.

ERBISTI, R. S. **Flexible covariance modeling of multivariate spatio-temporal processes**. [S.l.: s.n.], 2019.

FONSECA, T. C.; FERREIRA, M. A. Dynamic multiscale spatiotemporal models for poisson data. **Journal of the American Statistical Association**, Taylor & Francis, v. 112, n. 517, p. 215–234, 2017.

FONSECA, T. C.; STEEL, M. F. A general class of nonseparable space–time covariance models. **Environmetrics**, Wiley Online Library, v. 22, n. 2, p. 224–242, 2011.

GEMAN, S.; GEMAN, D. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. **IEEE Transactions on pattern analysis and machine intelligence**, IEEE, n. 6, p. 721–741, 1984.

GNEITING, T.; RAFTERY, A. E. Strictly proper scoring rules, prediction, and estimation. **Journal of the American statistical Association**, Taylor & Francis, v. 102, n. 477, p. 359–378, 2007.

GOULARD, M.; VOLTZ, M. Linear coregionalization model: tools for estimation and choice of cross-variogram matrix. **Mathematical Geology**, Springer, v. 24, p. 269–286, 1992.

HOETING, J. A. et al. Bayesian model averaging: a tutorial (with comments by m. clyde, david draper and ei george, and a rejoinder by the authors. **Statistical science**, Institute of Mathematical Statistics, v. 14, n. 4, p. 382–417, 1999.

LI, K. H. et al. Spatial–temporal models for ambient hourly pm10 in vancouver. **Environmetrics: The official journal of the International Environmetrics Society**, Wiley Online Library, v. 10, n. 3, p. 321–338, 1999.

MADIGAN, D.; RAFTERY, A. E. Model selection and accounting for model uncertainty in graphical models using occam's window. **Journal of the American Statistical Association**, Taylor & Francis, v. 89, n. 428, p. 1535–1546, 1994.

MARDIA, K. V.; GOODALL, C. R. Spatial-temporal analysis of multivariate environmental monitoring data. **Multivariate environmental statistics**, Elsevier Amsterdam, v. 6, n. 76, p. 347–385, 1993.

MARTINEZ-BENEITO, M. A. A general modelling framework for multivariate disease mapping. **Biometrika**, Oxford University Press, v. 100, n. 3, p. 539–553, 2013.

NEAL, R. M. Slice sampling. **The annals of statistics**, Institute of Mathematical Statistics, v. 31, n. 3, p. 705–767, 2003.

PETRIS, G.; PETRONE, S.; CAMPAGNOLI, P. Dynamic linear models. In: **Dynamic linear models with R**. [S.l.]: Springer, 2009. p. 31–84.

R Core Team. **R: A Language and Environment for Statistical Computing**. Vienna, Austria, 2023. Disponível em: https://www.R-project.org/.

RIBEIRO, A. M. T.; JUNIOR, P. J. R.; BONAT, W. H. A kronecker-based covariance specification for spatially continuous multivariate data. **Stochastic Environmental Research and Risk Assessment**, Springer, v. 36, n. 12, p. 4087–4102, 2022.

SAIN, S. R.; CRESSIE, N. A spatial model for multivariate lattice data. **Journal of Econometrics**, Elsevier, v. 140, n. 1, p. 226–259, 2007.

SAIN, S. R.; FURRER, R.; CRESSIE, N. A spatial analysis of multivariate output from regional climate models. **The Annals of Applied Statistics**, JSTOR, p. 150–175, 2011.

SARTÓRIO, V. S.; FONSECA, T. C. Dynamic clustering of time series data. **arXiv preprint arXiv:2002.01890**, 2020.

SCHEUERER, M.; HAMILL, T. M. Variogram-based proper scoring rules for probabilistic forecasts of multivariate quantities. **Monthly Weather Review**, American Meteorological Society, v. 143, n. 4, p. 1321–1334, 2015.

STEIN, M. L. **Interpolation of spatial data: some theory for kriging**. [S.l.]: Springer Science & Business Media, 1999.

STEIN, M. L. Space–time covariance functions. **Journal of the American Statistical Association**, Taylor & Francis, v. 100, n. 469, p. 310–321, 2005.

STEIN, M. L.; CHI, Z.; WELTY, L. J. Approximating likelihoods for large spatial data sets. **Journal of the Royal Statistical Society: Series B (Statistical Methodology)**, Wiley Online Library, v. 66, n. 2, p. 275–296, 2004.

VECCHIA, A. V. Estimation and model identification for continuous spatial processes. **Journal of the Royal Statistical Society: Series B (Methodological)**, Wiley Online Library, v. 50, n. 2, p. 297–312, 1988.

WACKERNAGEL, H. **Multivariate geostatistics: an introduction with applications**. [S.l.]: Springer Science & Business Media, 2003.

WEST, M.; HARRISON, J. **Bayesian forecasting and dynamic models.** 2. ed. [S.l.]: Springer., 1997.

# APÊNDICE A – DISTRIBUIÇÕES CONDICIONAIS COMPLETAS

Neste apêndice são apresentadas as distribuições condicionais completas e distribuições *a Posteriori* dos parâmetros para a implementação do algoritmo MCMC.

- Assumindo $\sigma_j^2 = \sigma^2$, para todo $j$, tem-se que: $\sigma^{-2}|\cdot \sim Gama(\alpha, \lambda)$.

$$\alpha = a + \frac{T \times n \times m}{2};$$
$$\lambda = b + \frac{1}{2}\sum_{t=1}^{T}\sum_{k=1}^{K} I(Z_t = k)\sum_{j=1}^{m}\sum_{i=1}^{n}\left(Y_j^{(t)}(s_i) - \mu_j^{(t)}(s_i) - W_j(s_i)\right)^2$$

Para o processo espacial $\boldsymbol{W}$ e seus hiperparâmetros o índice da mistura $k$ está implícito. Isso também vale para $\Pi_q^{(k)} \equiv \Pi_q$. Pela notação os modelos sem mistura são aqueles onde $K = 1$. Desta forma, as distribuições a seguir podem ser consideradas para todos os modelos. Além disso, defina o conjunto $J_q$ como o conjunto de componentes $\boldsymbol{W}_j$ que contém $\boldsymbol{W}_q$ como uma das condicionantes.

- Para $k = 1, \ldots, K$ e $q = 1, \ldots, m$: $\phi_q^{(k)}, \nu_q^{(k)}, \alpha_q^{(k)}$. Então:

  - $\phi_q$:

$$\phi_q^{-1}|\cdot \sim Gama\left(a_q + \frac{n}{2}; \gamma_q + \frac{1}{2}\left(\boldsymbol{W}_q - \boldsymbol{m}_{0q}\right)^T S_q^{-1}\left(\boldsymbol{W}_q - \boldsymbol{m}_{0q}\right)\right)$$
$$\boldsymbol{m}_{0q} = \sum_{j \in \Pi_q} B_{jq}\boldsymbol{W}_j;$$
$$C_q = \phi_q S_q$$

  - $\nu_q$ e $\alpha_q$:

$$P(v_q, \alpha_q|\cdot) \propto P(\boldsymbol{W}_q|\Pi_q) \times P(v_q) \times P(\alpha_q) \text{ ou}$$
$$P(v_q|\cdot) \propto P(\boldsymbol{W}_q|\Pi_q) \times P(v_q)$$
$$P(\alpha_q|\cdot) \propto P(\boldsymbol{W}_q|\Pi_q) \times P(\alpha_q)$$

- Para $k = 1, \ldots, K$ e $q = 2, \ldots, m$, $j < q$: $A_{jq}^{(k)}, \boldsymbol{\Delta}_{jq}^{(k)}, r_{jq}^{(k)}$ (os dois últimos não conhecidos). Então:

  - Para $A_{jq}$: $A_{jq}|\cdot \sim N\left(m_{A_{jq}}, \Sigma_{A_{jq}}\right)$

$$\Sigma_{A_{jq}} = \left[\left(B'_{jq}\boldsymbol{W}_j\right)^T C_q^{-1}\left(B'_{jq}\boldsymbol{W}_j\right) + \frac{1}{V_{A_{jq}}}\right]^{-1};$$
$$m_{A_{jq}} = \Sigma_{A_{jq}} \times \left[\left(B'_{jq}\boldsymbol{W}_j\right)^T C_q^{-1}\left(\boldsymbol{W}_q - \sum_{r \in \Pi_q - \{j\}} B_{rq}\boldsymbol{W}_r\right)\right];$$
$$B_{jq} = A_{jq}B'_{jq}$$

– Para $\boldsymbol{\Delta}_{jq}$ e $r_{jq}$:

$$P(\boldsymbol{\Delta}_{jq}|\cdot) \propto P(\boldsymbol{W}_q|\Pi_q) \times P(\boldsymbol{\Delta}_{jq})$$
$$P(r_{jq}|\cdot) \propto P(\boldsymbol{W}_q|\Pi_q) \times P(r_{jq})$$

- Para os processos espaciais $\boldsymbol{W}_1^{(k)}, \boldsymbol{W}_2^{(k)}, \ldots, \boldsymbol{W}_m^{(k)}$, para $k = 1, \ldots, K$, tem-se:

  – Para $\boldsymbol{W}_1$: $\boldsymbol{W}_1|\cdot \sim N(\boldsymbol{m}_{w_1}, \Sigma_{w_1})$.

$$\Sigma_{\boldsymbol{W}_1} = \left[\frac{z_k}{\sigma^2} I_n + C_1^{-1} + \sum_{j \in J_1} B_{1j}^T C_j^{-1} B_{1j}\right]^{-1};$$
$$m_{\boldsymbol{W}_1} = \Sigma_{\boldsymbol{W}_1} \times \left[\sum_{t=1}^{T} \frac{I(Z_t = k)}{\sigma^2}\left(\boldsymbol{Y}_1^{(t)} - \boldsymbol{\mu}_1^{(t)}\right) + \sum_{j \in J_1} B_{1j}^T C_j^{-1}\left(\boldsymbol{W}_j - \sum_{r \in \Pi_j - \{1\}} B_{rj}\boldsymbol{W}_r\right)\right]$$

  – Para $\boldsymbol{W}_q$, $q = 2, \ldots, m-1$: $\boldsymbol{W}_q|\cdot \sim N(\boldsymbol{m}_{w_q}, \Sigma_{w_q})$.

$$\Sigma_{\boldsymbol{W}_q} = \left[\frac{z_k}{\sigma^2} I_n + C_q^{-1} + \sum_{j \in J_q} B_{qj}^T C_j^{-1} B_{qj}\right]^{-1};$$
$$m_{\boldsymbol{W}_q} = \Sigma_{\boldsymbol{W}_q} \times \left[\sum_{t=1}^{T} \frac{I(Z_t = k)}{\sigma^2}\left(\boldsymbol{Y}_q^{(t)} - \boldsymbol{\mu}_q^{(t)}\right) + C_q^{-1} m_{0q} + \sum_{j \in J_q} B_{qj}^T C_j^{-1}\left(\boldsymbol{W}_j - \sum_{r \in \Pi_j - \{q\}} B_{rj}\boldsymbol{W}_r\right)\right]$$

  – Para $\boldsymbol{W}_m$: $\boldsymbol{W}_m|\cdot \sim N(\boldsymbol{m}_{w_m}, \Sigma_{w_m})$.

$$\Sigma_{\boldsymbol{W}_m} = \left[\frac{z_k}{\sigma^2} I_n + C_m^{-1}\right]^{-1};$$
$$m_{\boldsymbol{W}_m} = \Sigma_{\boldsymbol{W}_m} \times \left[\sum_{t=1}^{T} \frac{I(Z_t = k)}{\sigma^2}\left(\boldsymbol{Y}_m^{(t)} - \boldsymbol{\mu}_m^{(t)}\right) + C_m^{-1} m_{0m}\right]$$

- No caso do modelo de mistura com pesos estáticos no tempo ($\boldsymbol{\eta}^{(t)} = \boldsymbol{\eta}$), dado uma distribuição *a priori* $\boldsymbol{\eta} \sim Dirichlet(\boldsymbol{c}_0)$, tem-se que: $\boldsymbol{\eta}|\cdot \sim Dirichlet(\boldsymbol{c}_0 + \boldsymbol{z})$, em que $\boldsymbol{z} = (z_1, \ldots, z_K)$, é a contagem de períodos em que cada modelo foi selecionado, $z_k = \sum_{t=1}^{T} I(Z_t = k)$.

- Distribuição de $Z_t$:

$$P(Z_t = k|\cdot) = \frac{\eta_k^{(t)} f_k\left(\boldsymbol{y}_t|\boldsymbol{\mu}_t, \boldsymbol{W}^{(k)}, V\right)}{\sum_{l=1}^{K} \eta_l^{(t)} f_l\left(\boldsymbol{y}_t|\boldsymbol{\mu}_t, \boldsymbol{W}^{(l)}, V\right)}.$$

  No caso do modelo de mistura com pesos estáticos no tempo, o lado direito da equação considera $\eta_k^{(t)} = \eta_k, \forall k$.

# APÊNDICE B – ESTUDOS DE SIMULAÇÃO

Neste apêndice são apresentados resultados complementares dos estudos de simulação realizados no trabalho. São apresentados os hiperparâmetros das funções de covariância e funções de interação usados na geração dos dados em cada cenário de ambos os estudos. Além disso, para o primeiro estudo são disponibilizados gráficos com as estimativas dos hiperparâmetros em cada cenário e para o segundo estudo, os gráficos de interpolação dos cenários 2 e 3 no segundo estudo de simulação, para cada componente, para cada modelo ajustado e para cada estrutura de covariância simulada na geração dos dados. As figuras seguem a mesma estrutura dos gráficos de interpolação do primeiro cenário, que foram apresentados ao longo do texto.

## B.1 ESTUDO DE SIMULAÇÃO 1

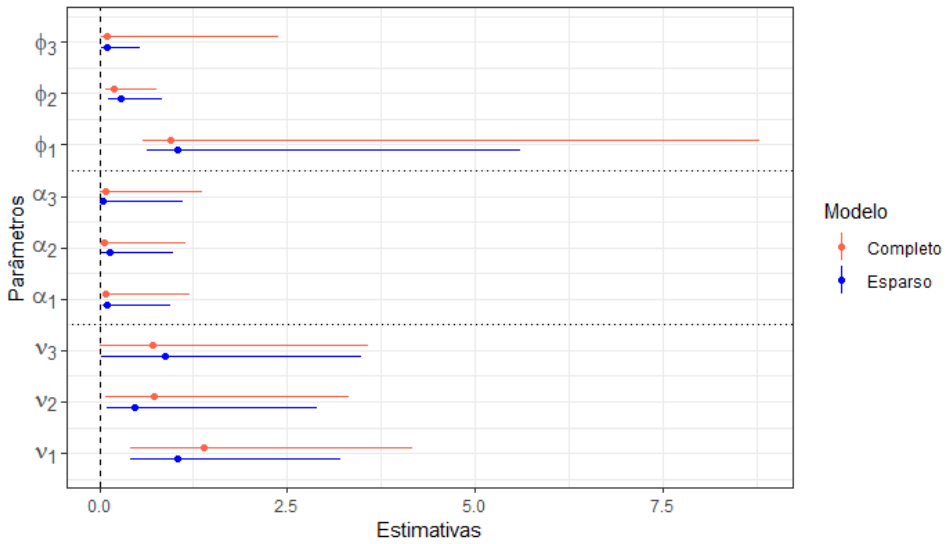
Na Tabela abaixo, seguem os hiperparâmetros usados para a geração dos dados artificiais no primeiro estudo de simulação.

| Hiperparâmetros | Cenário 1 | Cenário 2 | Cenário 3 | Cenário 4 |
|---|---|---|---|---|
| $\phi_1$ | 1 | 1 | 1 | 1 |
| $\phi_2$ | 0.25 | 0.25 | 0.25 | 0.25 |
| $\phi_3$ | 0.1 | 0.1 | 0.1 | 0.1 |
| $\nu_1$ | 1.5 | 1.5 | 1.5 | 1.5 |
| $\nu_2$ | 1 | 1 | 1 | 1 |
| $\nu_3$ | 0.5 | 0.5 | 0.5 | 0.5 |
| $\alpha_1$ | 0.08 | 0.08 | 0.08 | 0.08 |
| $\alpha_2$ | 0.075 | 0.075 | 0.075 | 0.075 |
| $\alpha_3$ | 0.05 | 0.05 | 0.05 | 0.05 |
| $A_{12}$ | 1 | 1 | 1 | 1 |
| $A_{13}$ | 0.25 | 0.25 | 0.25 | 0.5 |
| $A_{23}$ | 0.75 | 0.75 | 0.75 | 0.5 |
| $r_{12}$ | 0.1 | 0.1 | 0.1 | 0.1 |
| $r_{13}$ | 0.05 | 0.05 | 0.05 | 0.1 |
| $r_{23}$ | 0.1 | 0.1 | 0.1 | 0.1 |
| $\Delta_{12}^1$ | 0 | 0 | 0 | 0 |
| $\Delta_{13}^1$ | 0 | -0.1 | 0 | 0.1 |
| $\Delta_{23}^1$ | 0 | 0 | 0.3 | -0.1 |
| $\Delta_{12}^2$ | 0 | 0 | 0 | 0 |
| $\Delta_{13}^2$ | 0 | 0.1 | 0 | -0.1 |
| $\Delta_{23}^2$ | 0 | 0 | -0.1 | 0.1 |

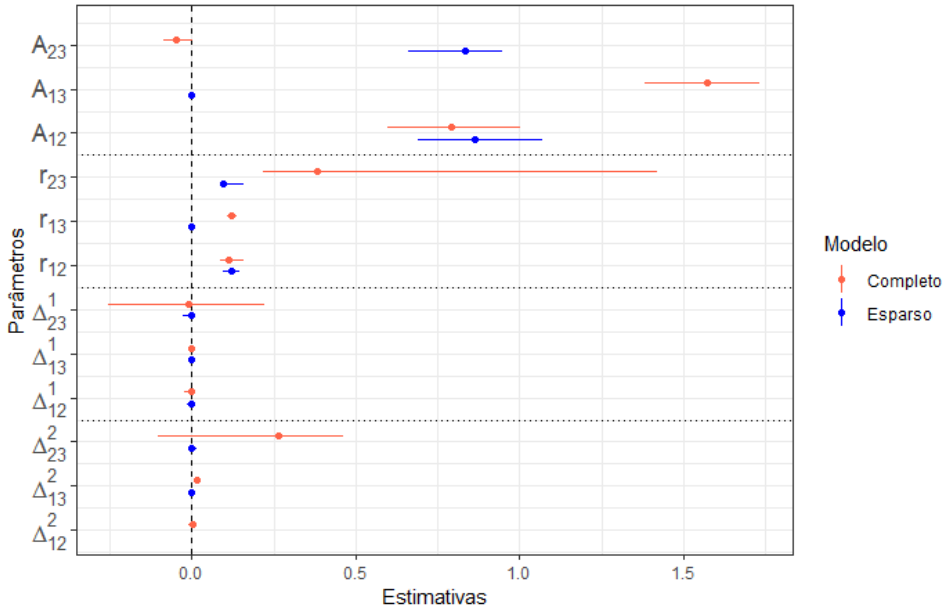
Tabela 12 – Hiperparâmetros utilizados para a geração dos dados no primeiro estudo de simulação em cada cenário.

Seguem as estimativas dos hiperparâmetros dos modelos ajustados no primeiro estudo de simulação.

### B.1.1 Cenário 1



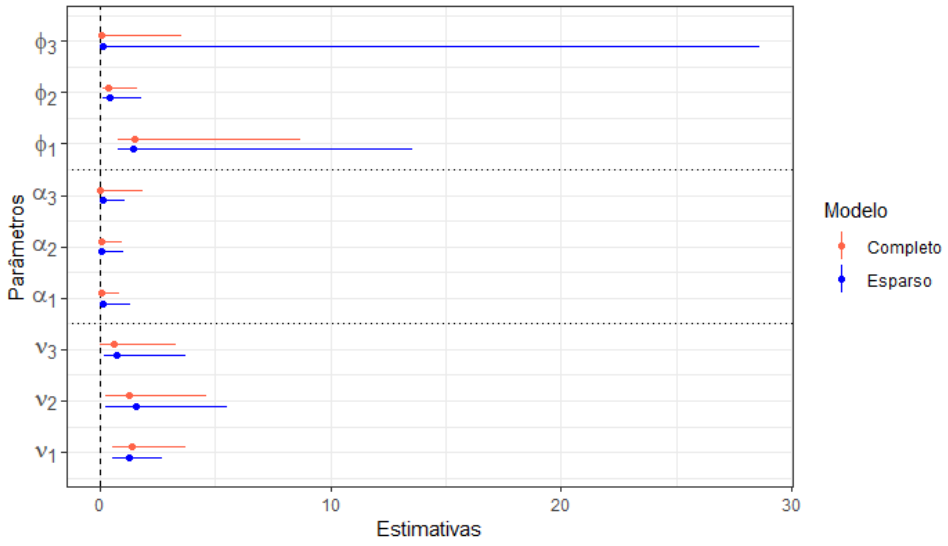

(a) Hiperparâmetros das funções de covariância.



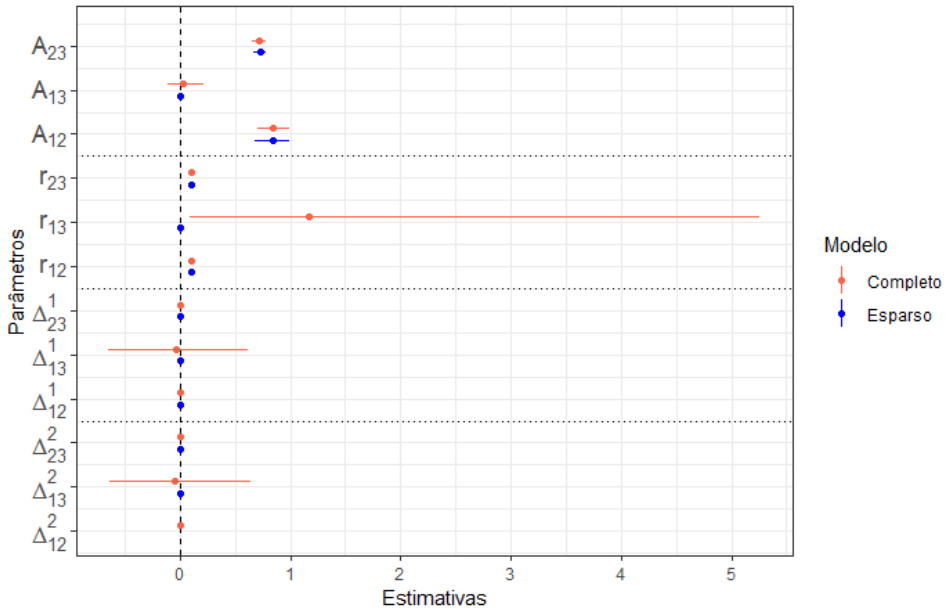

(b) Hiperparâmetros das funções de interação.

Figura 33 – Mediana *a posteriori* e intervalos de credibilidade de 95% dos hiperparâmetros estimados pelos modelos completo e esparso.

### B.1.2 Cenário 2



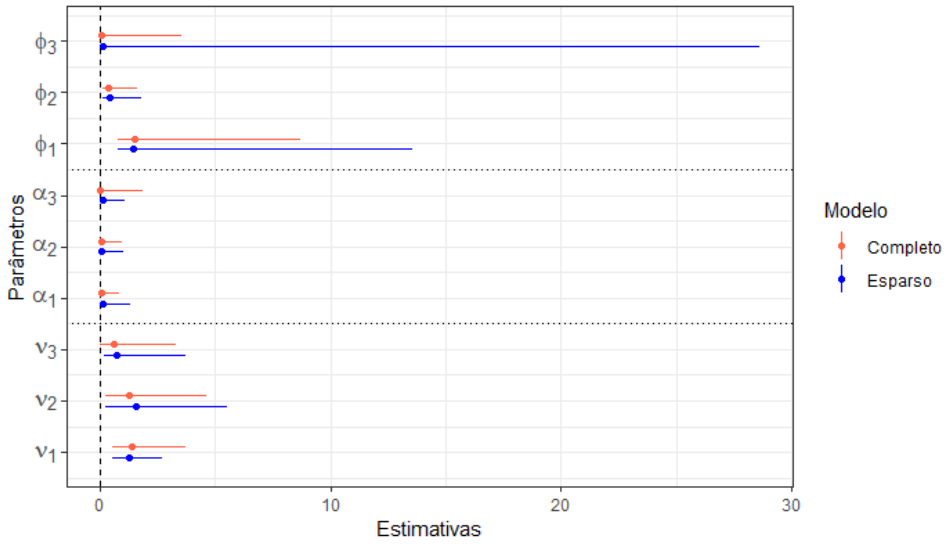

(a) Hiperparâmetros das funções de covariância.



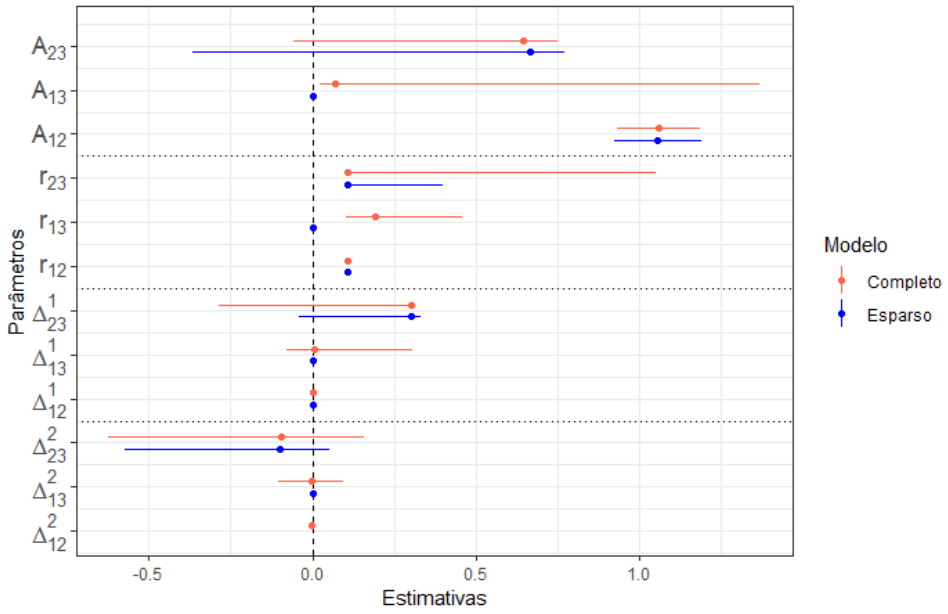

(b) Hiperparâmetros das funções de interação.

Figura 34 – Mediana *a posteriori* e intervalos de credibilidade de 95% dos hiperparâmetros estimados pelos modelos completo e esparso.

### B.1.3 Cenário 3



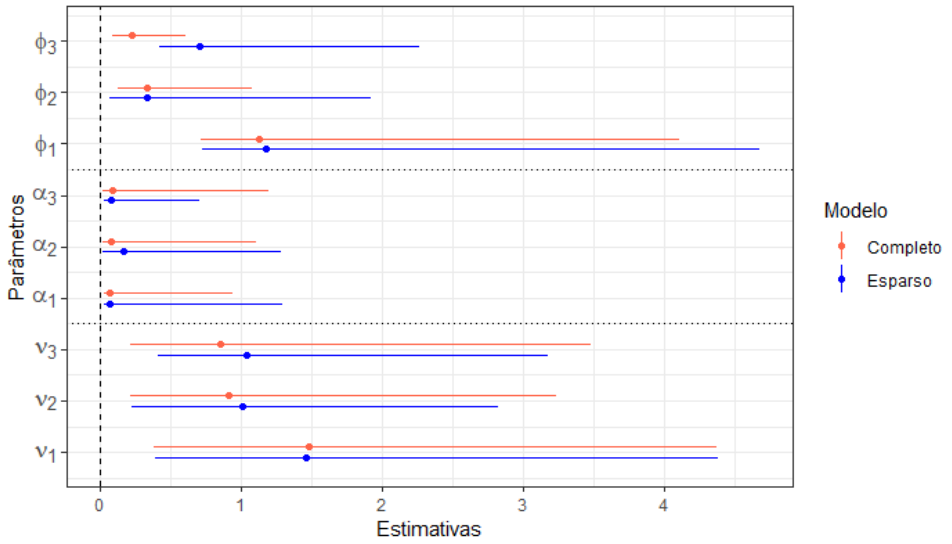

(a) Hiperparâmetros das funções de covariância.



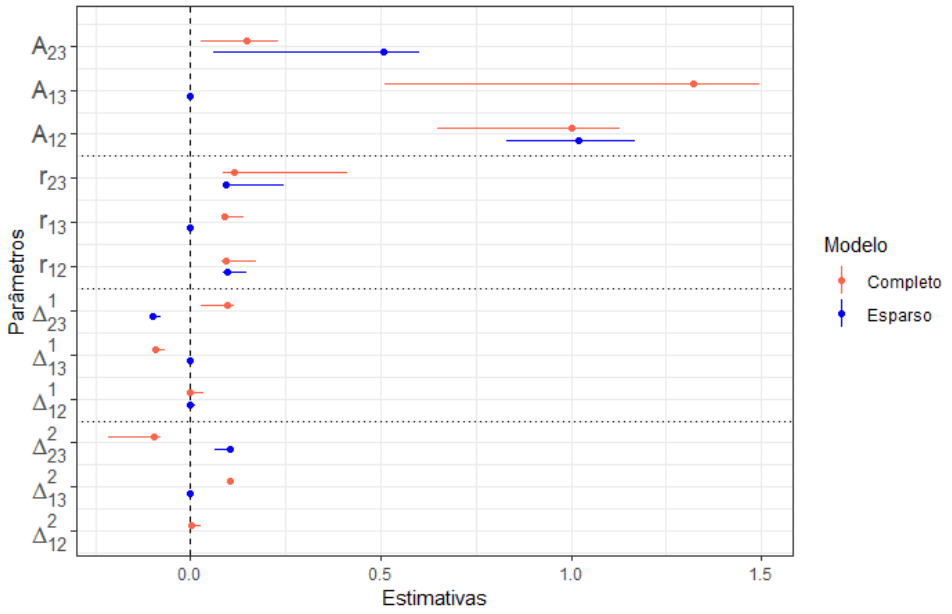

(b) Hiperparâmetros das funções de interação.

Figura 35 – Mediana *a posteriori* e intervalos de credibilidade de 95% dos hiperparâmetros estimados pelos modelos completo e esparso.

### B.1.4 Cenário 4




(a) Hiperparâmetros das funções de covariância.




(b) Hiperparâmetros das funções de interação.

Figura 36 – Mediana *a posteriori* e intervalos de credibilidade de 95% dos hiperparâmetros estimados pelos modelos completo e esparso.

## B.2 ESTUDO DE SIMULAÇÃO 2

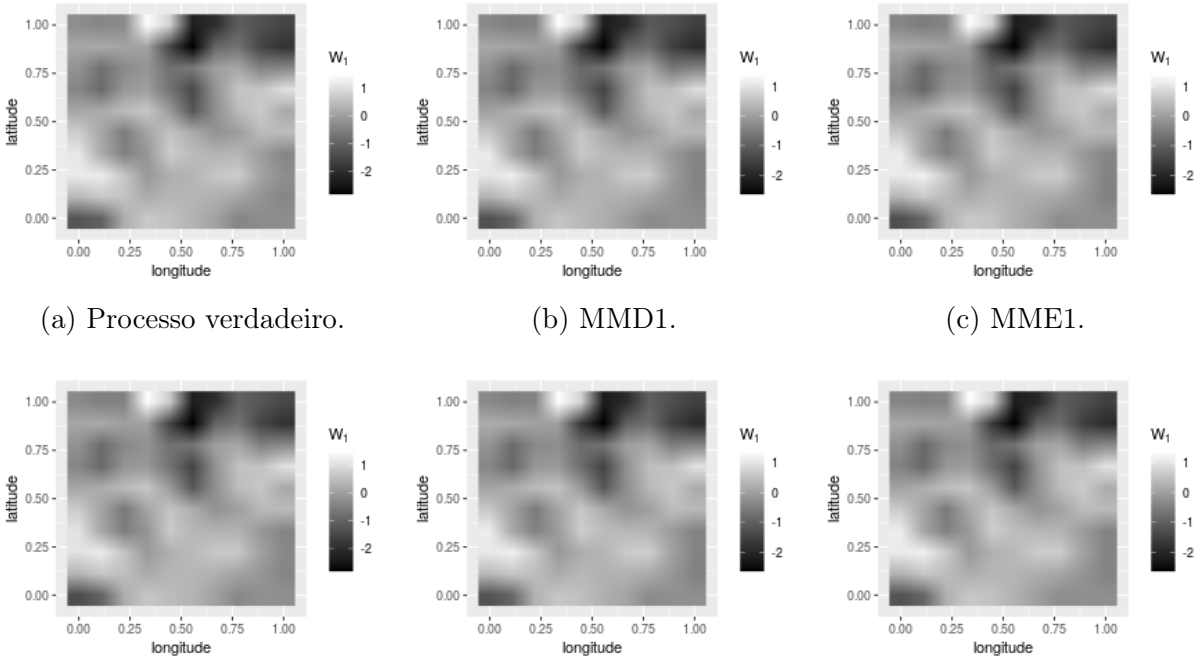
Na Tabela a seguir, são apresentados os hiperparâmetros usados para a geração dos dados artificiais no segundo estudo de simulação.

| | Cenário 1 | | Cenário 2 | | Cenário 3 | |
|---|---|---|---|---|---|---|
| Hiperparâmetros | Estrutura 1 | Estrutura 2 | Estrutura 1 | Estrutura 2 | Estrutura 1 | Estrutura 2 |
| $\phi_1$ | 1 | 1 | 1 | 1 | 1 | 1 |
| $\phi_2$ | 0.25 | 0.5 | 0.25 | 0.25 | 0.25 | 0.25 |
| $\phi_3$ | 0.1 | 0.25 | 0.1 | 0.1 | 0.1 | 0.1 |
| $\nu_1$ | 1.5 | 1.5 | 1.5 | 1.5 | 1.5 | 1.5 |
| $\nu_2$ | 1 | 1.5 | 1 | 1 | 1 | 1 |
| $\nu_3$ | 0.75 | 1.5 | 0.75 | 0.75 | 0.75 | 0.75 |
| $\alpha_1$ | 0.08 | 0.08 | 0.08 | 0.08 | 0.08 | 0.08 |
| $\alpha_2$ | 0.075 | 0.06 | 0.075 | 0.075 | 0.075 | 0.075 |
| $\alpha_3$ | 0.05 | 0.06 | 0.05 | 0.05 | 0.05 | 0.05 |
| $A_{12}$ | 1 | 0.75 | 1 | 0.75 | 0.75 | 1 |
| $A_{13}$ | 0.25 | 0.75 | 0.25 | 0.25 | 0.25 | 0.25 |
| $A_{23}$ | 0.75 | 0.25 | 0.75 | 0.5 | 0.5 | 1 |
| $r_{12}$ | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 |
| $r_{13}$ | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 |
| $r_{23}$ | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 | 0.1 |
| $\Delta^1_{12}$ | 0 | 0 | 0 | 0 | 0 | 0.2 |
| $\Delta^1_{13}$ | 0 | 0 | 0 | 0 | 0 | 0 |
| $\Delta^1_{23}$ | 0.3 | 0.1 | 0.3 | 0.3 | 0 | 0.01 |
| $\Delta^2_{12}$ | 0 | 0 | 0 | 0 | 0 | -0.1 |
| $\Delta^2_{13}$ | 0 | 0 | 0 | 0 | 0 | 0 |
| $\Delta^2_{23}$ | -0.1 | -0.1 | -0.1 | -0.1 | 0 | -0.01 |

Tabela 13 – Hiperparâmetros utilizados para a geração dos dados de cada estrutura de correlação no segundo estudo de simulação nos três cenários.

Seguem os gráficos de interpolação do segundo e do terceiro cenário.

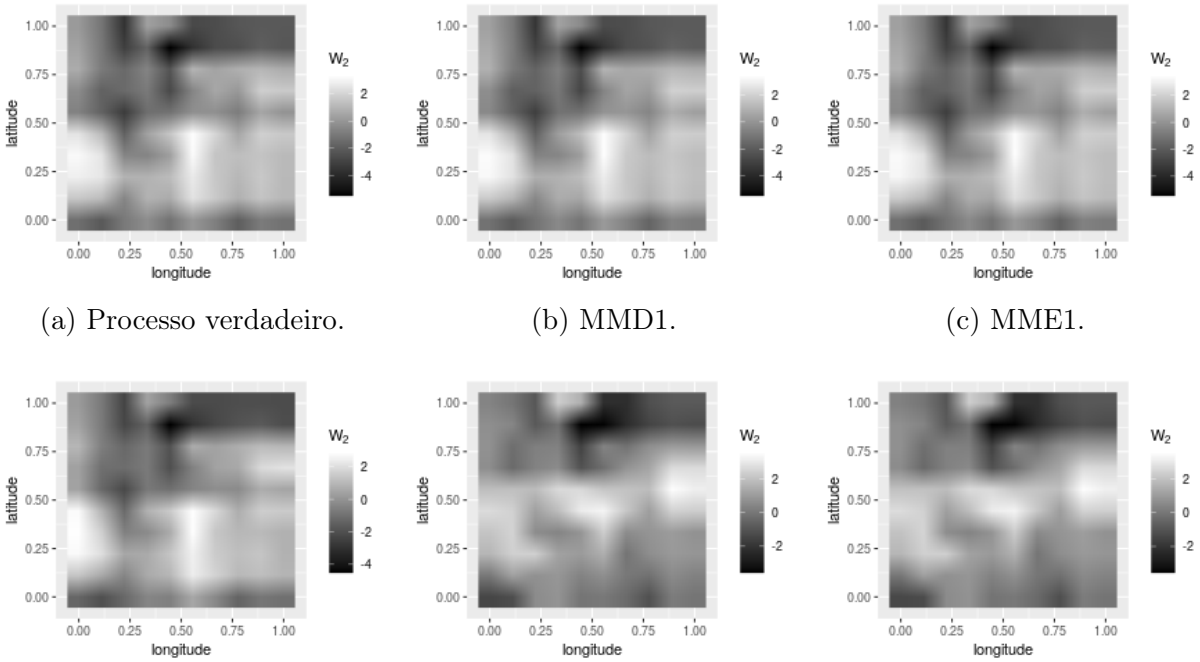
### B.2.1 Cenário 2




(a) Processo verdadeiro. (b) MMD1. (c) MME1.

(d) MC. (e) MMD2. (f) MME2.

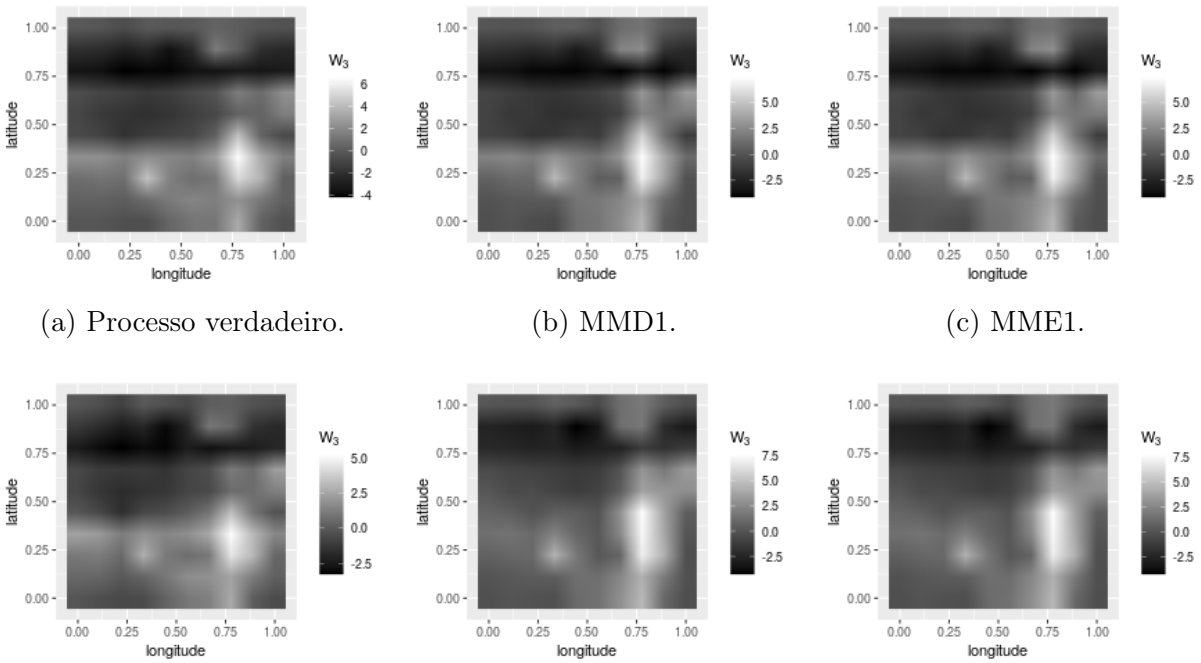
Figura 37 – Interpolação de $\boldsymbol{W}_1$ na primeira estrutura de dependência espacial.




(a) Processo verdadeiro. (b) MMD1. (c) MME1.

(d) MC. (e) MMD2. (f) MME2.

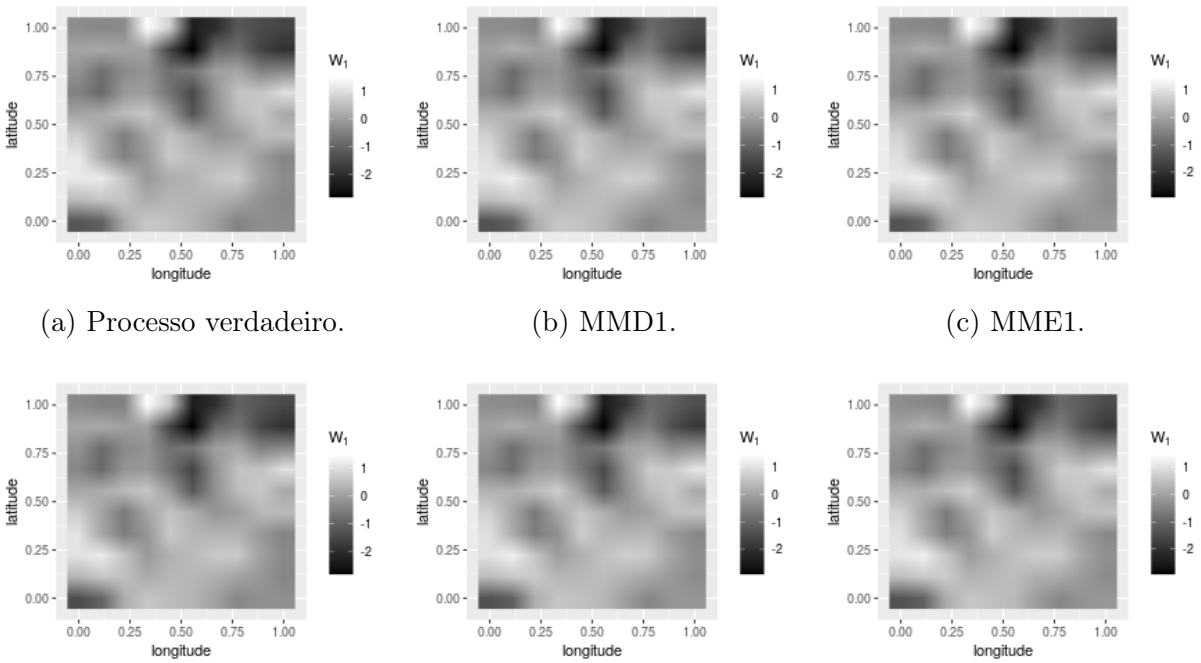
Figura 38 – Interpolação de $\boldsymbol{W}_2$ na primeira estrutura de dependência espacial.




(a) Processo verdadeiro. (b) MMD1. (c) MME1.

(d) MC. (e) MMD2. (f) MME2.

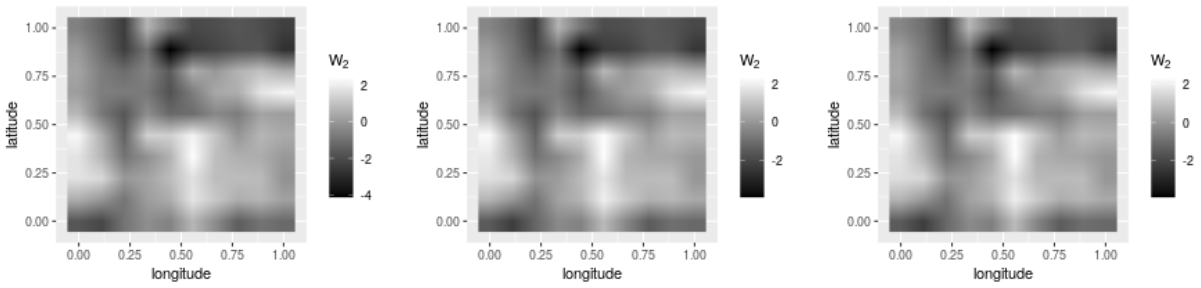
Figura 39 – Interpolação de $\boldsymbol{W}_3$ na primeira estrutura de dependência espacial.



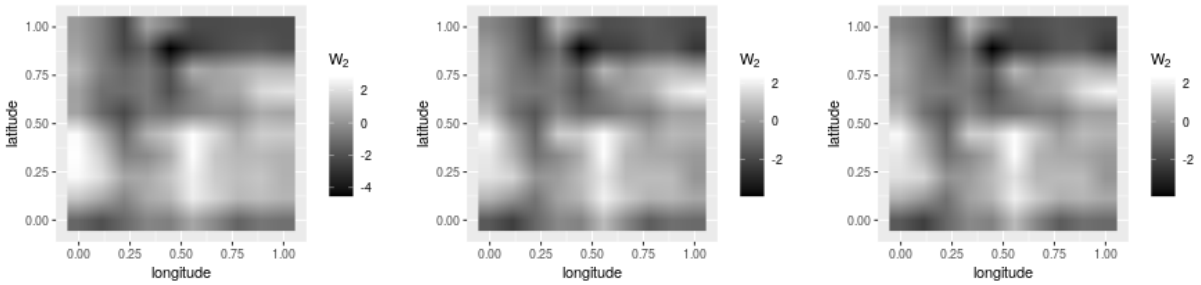

(a) Processo verdadeiro. (b) MMD1. (c) MME1.

(d) MC. (e) MMD2. (f) MME2.

Figura 40 – Interpolação de $\boldsymbol{W}_1$ na segunda estrutura de dependência espacial.

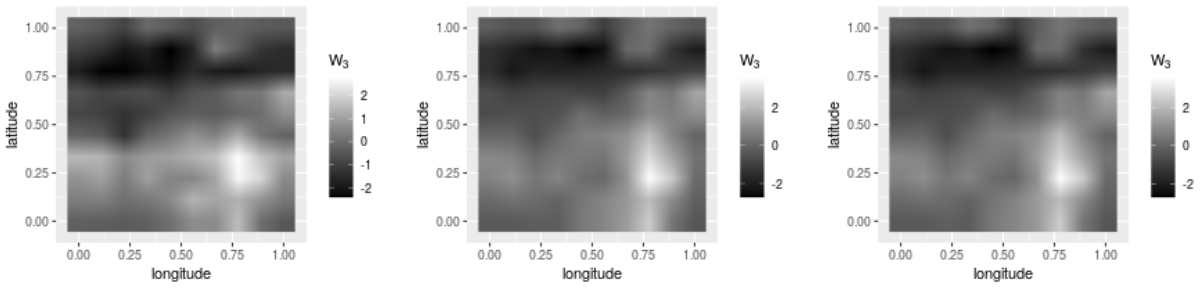



(a) Processo verdadeiro. (b) MMD1. (c) MME1.

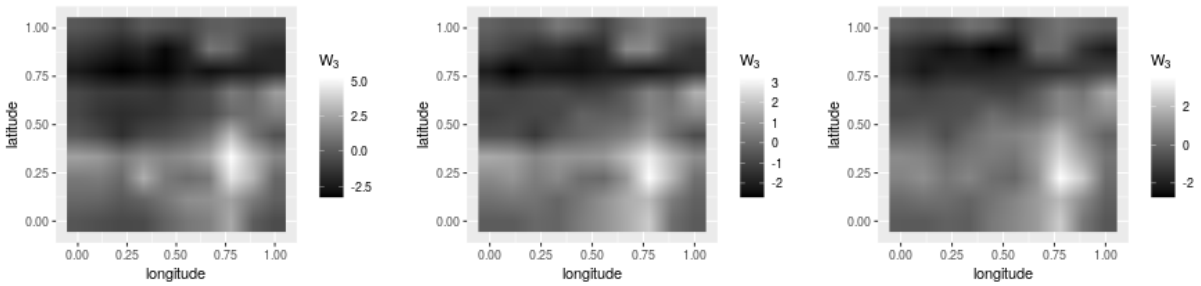



(d) MC. (e) MMD2. (f) MME2.

Figura 41 – Interpolação de $\boldsymbol{W}_2$ na segunda estrutura de dependência espacial.



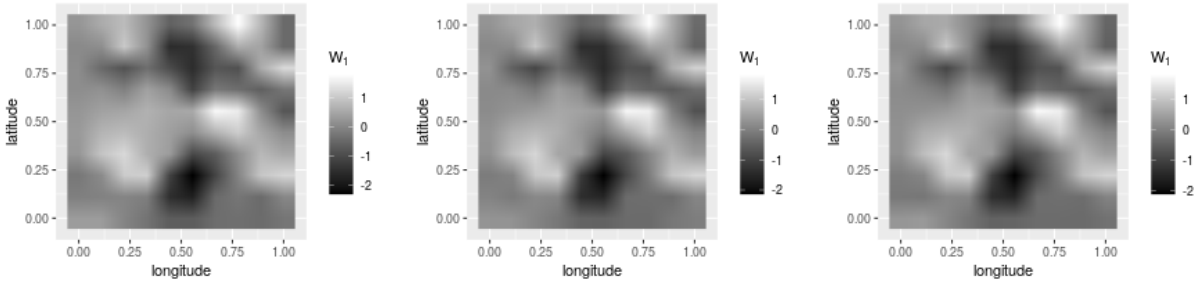

(a) Processo verdadeiro. (b) MMD1. (c) MME1.



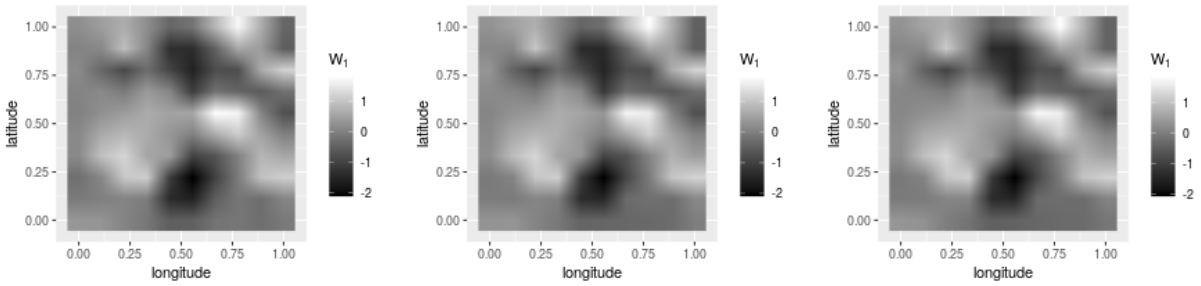

(d) MC. (e) MMD2. (f) MME2.

Figura 42 – Interpolação de $\boldsymbol{W}_3$ na segunda estrutura de dependência espacial.

### B.2.2 Cenário 3

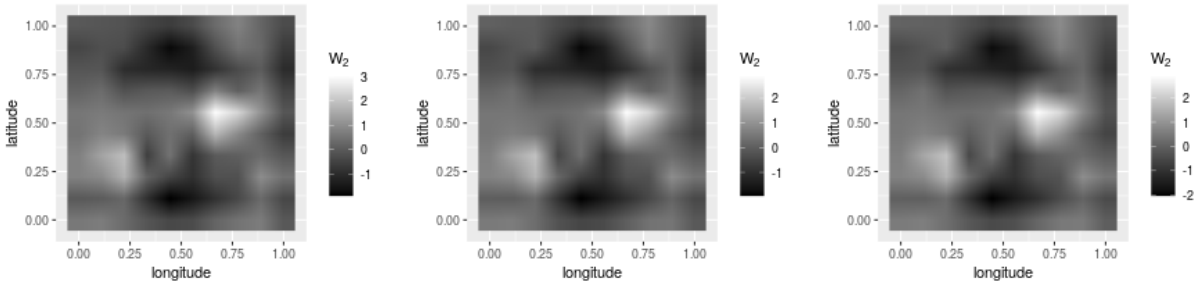



(a) Processo verdadeiro. (b) MMD1. (c) MME1.

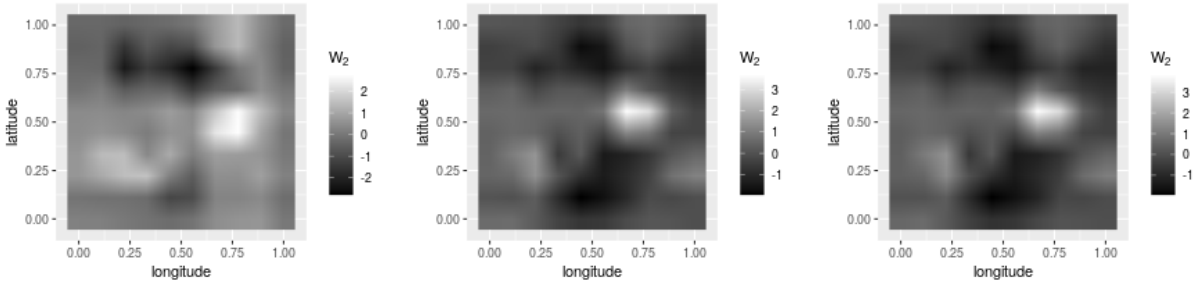



(d) MC. (e) MMD2. (f) MME2.

Figura 43 – Interpolação de $\boldsymbol{W}_1$ na primeira estrutura de dependência espacial.

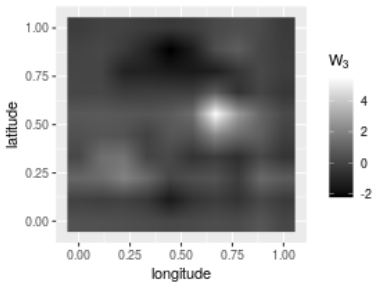



(a) Processo verdadeiro. (b) MMD1. (c) MME1.

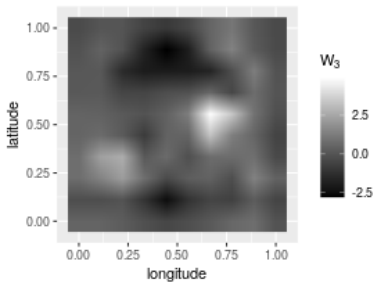



(d) MC. (e) MMD2. (f) MME2.

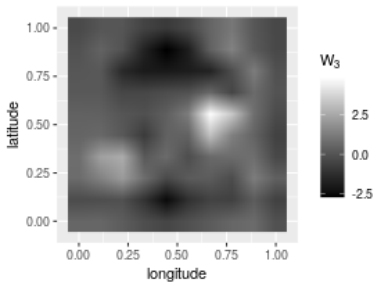
Figura 44 – Interpolação de $\boldsymbol{W}_2$ na primeira estrutura de dependência espacial.



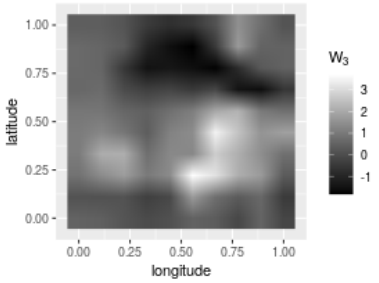

(a) Processo verdadeiro.



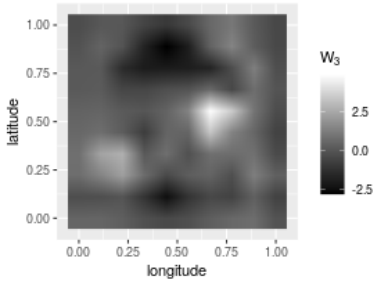

(b) MMD1.



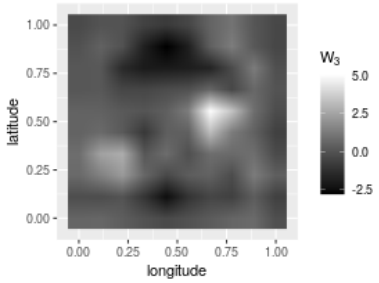

(c) MME1.




(d) MC.

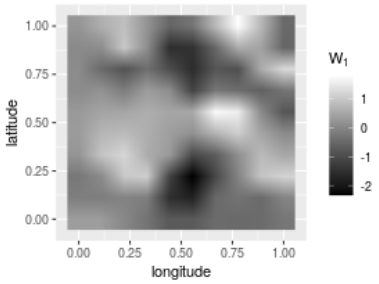



(e) MMD2.

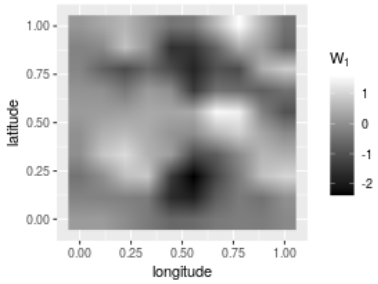



(f) MME2.

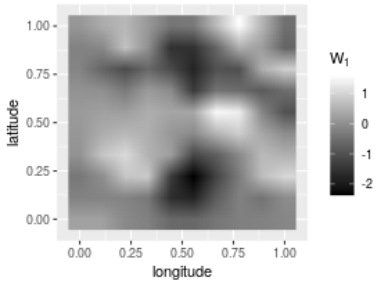
Figura 45 – Interpolação de $\boldsymbol{W}_3$ na primeira estrutura de dependência espacial.



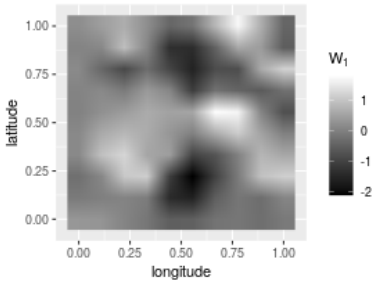

(a) Processo verdadeiro.



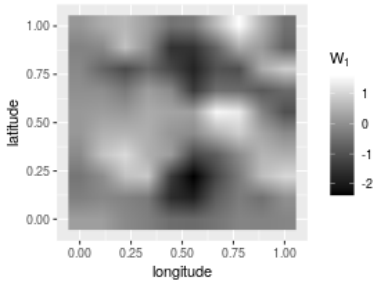

(b) MMD1.



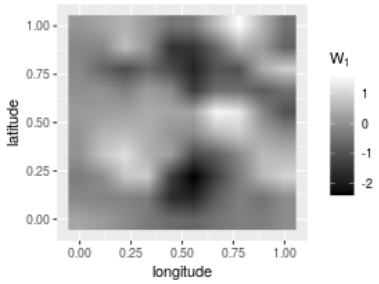

(c) MME1.




(d) MC.

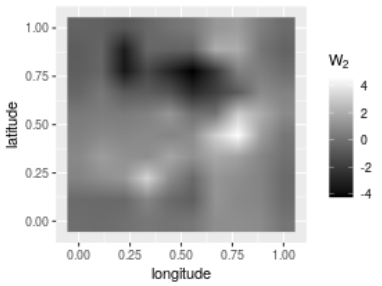



(e) MMD2.

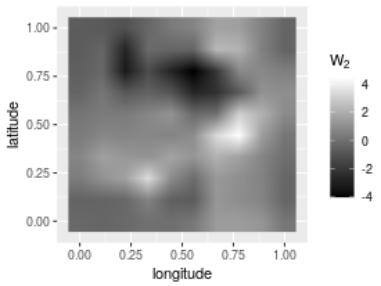



(f) MME2.

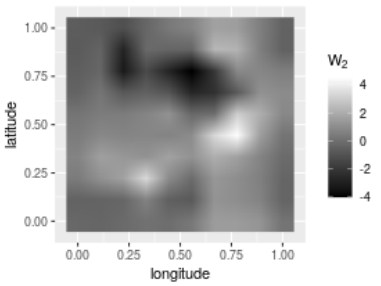
Figura 46 – Interpolação de $\boldsymbol{W}_1$ na segunda estrutura de dependência espacial.



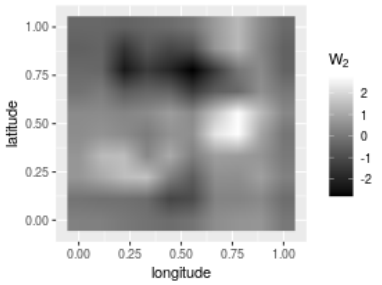

(a) Processo verdadeiro.



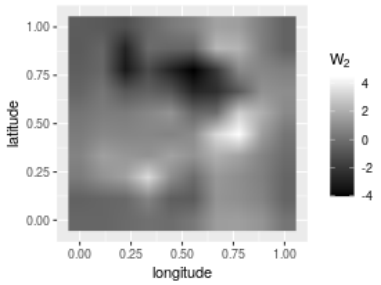

(b) MMD1.



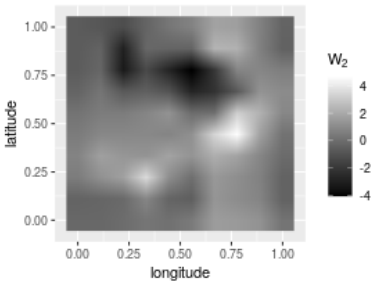

(c) MME1.




(d) MC.

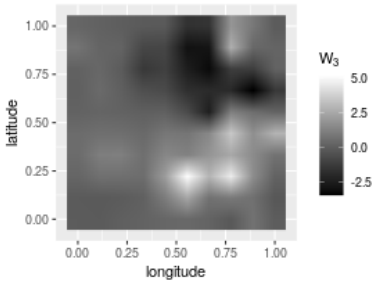



(e) MMD2.

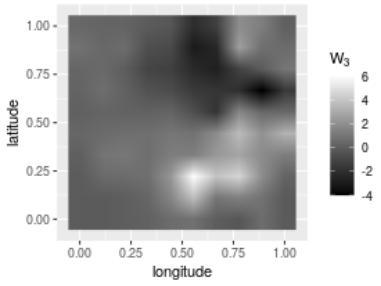



(f) MME2.

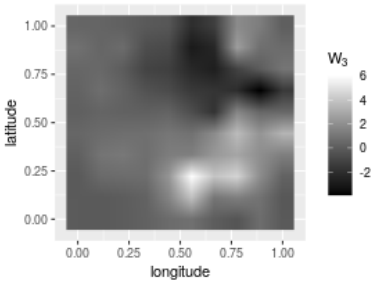
Figura 47 – Interpolação de $\boldsymbol{W}_2$ na segunda estrutura de dependência espacial.



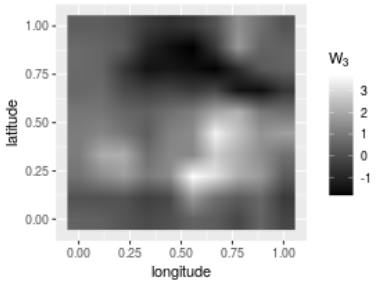

(a) Processo verdadeiro.



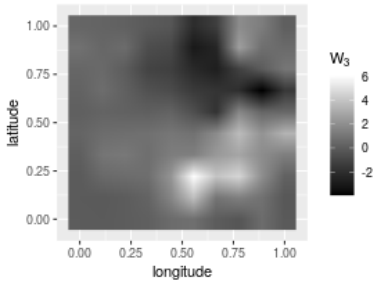

(b) MMD1.



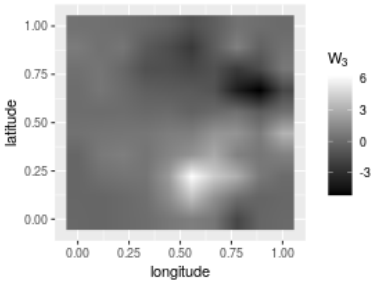

(c) MME1.




(d) MC.




(e) MMD2.




(f) MME2.

Figura 48 – Interpolação de $\boldsymbol{W}_3$ na segunda estrutura de dependência espacial.

# APÊNDICE C – GRÁFICOS DA APLICAÇÃO

Neste apêndice são apresentados alguns gráficos da aplicação que não puderam ser colocados ao longo do texto, entre eles, tem os gráficos das cadeias da log-verossimilhança de cada modelo na Figura 49, gráficos com os efeitos da longitude e da latitude sobre os poluentes ao longo do tempo para cada modelo e gráficos de interpolação dos poluentes atmosféricos.

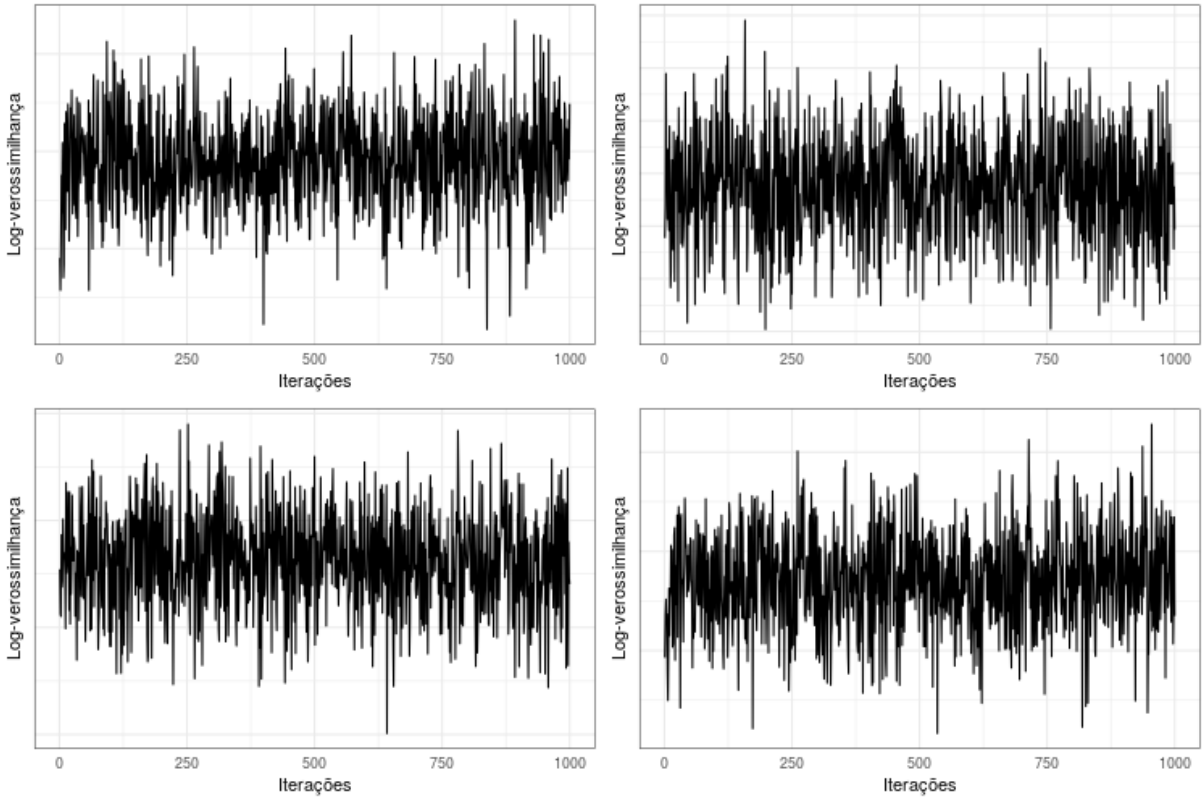
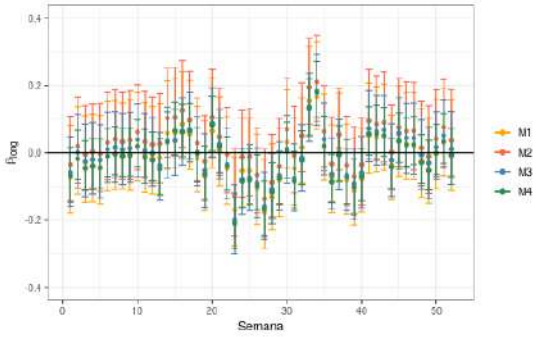
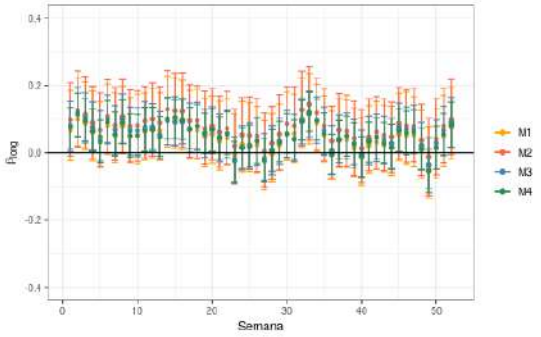


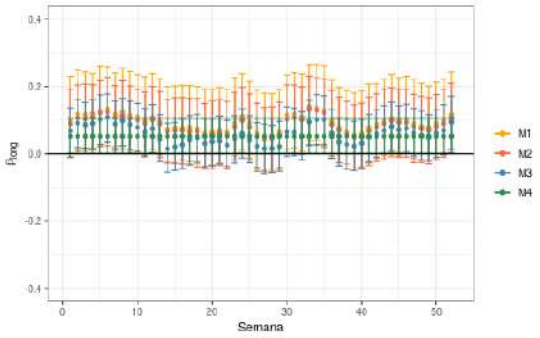

Figura 49 – Cadeia da log-verossimilhança em cada modelo. Na primeira linha, tem a log-verossimilhança dos modelos sem mistura $M1$ e $M2$, respectivamente. Na segunda linha, o primeiro gráfico é referente ao modelo de mistura com pesos evoluindo no tempo de acordo com processo Dirichlet e o último é referente ao modelo de mistura com pesos estáticos no tempo.



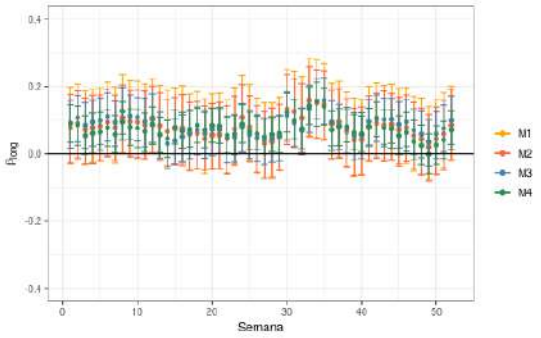

(a) $NO$.




(b) $NO_2$.

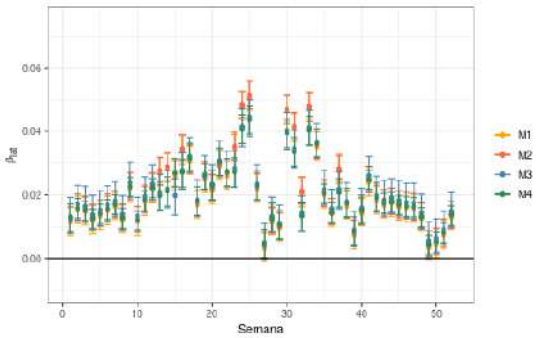



(c) $PM_{10}$.

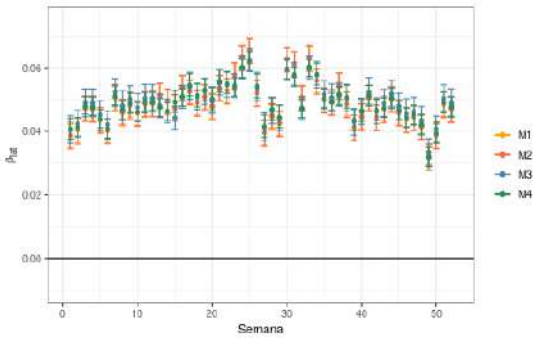



(d) $PM_{2.5}$.

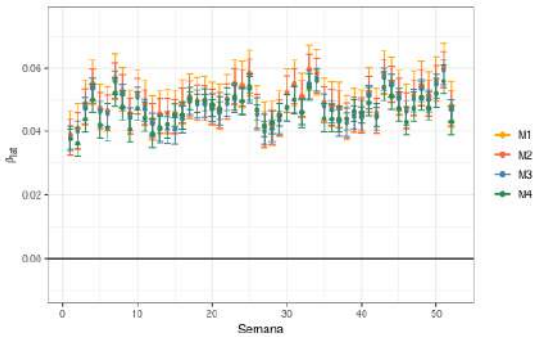
Figura 50 – Efeitos estimados por cada modelo da longitude ao longo das 52 semanas para cada poluente.



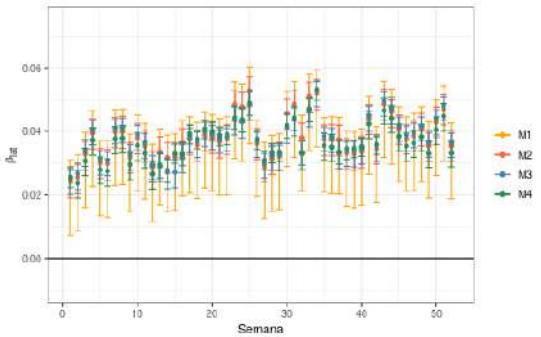

(a) $NO$.




(b) $NO_2$.

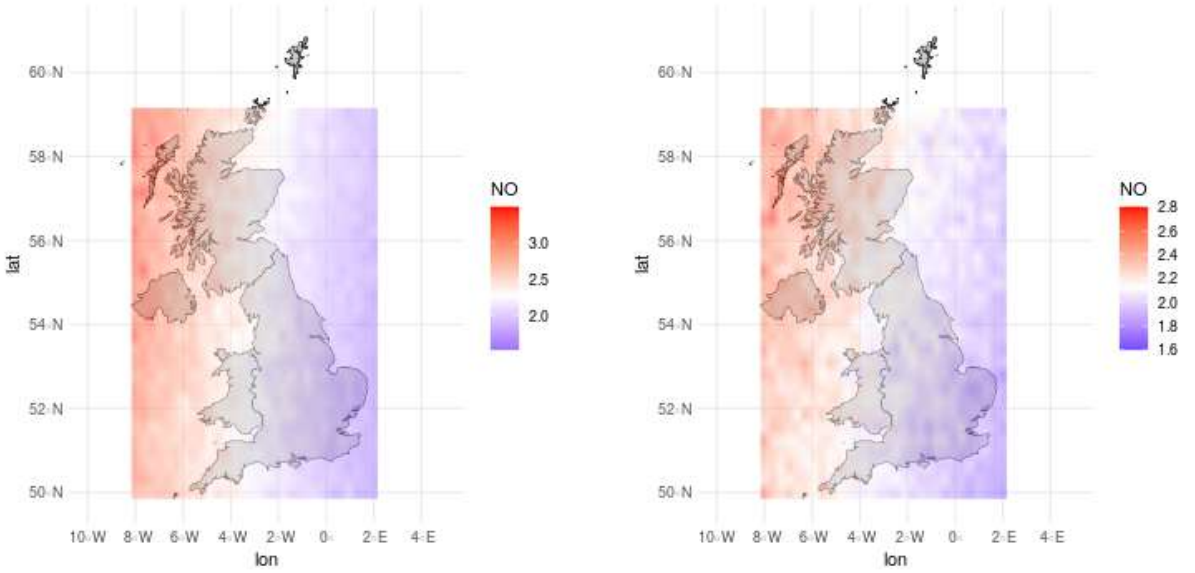



(c) $PM_{10}$.




(d) $PM_{2.5}$.

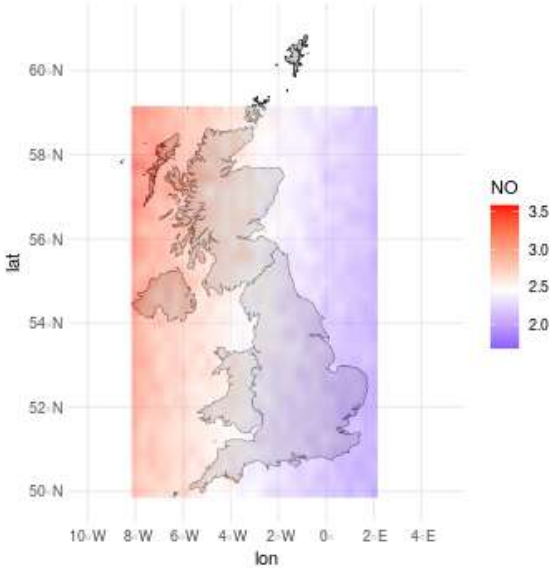
Figura 51 – Efeitos estimados por cada modelo da latitude ao longo das 52 semanas para cada poluente.



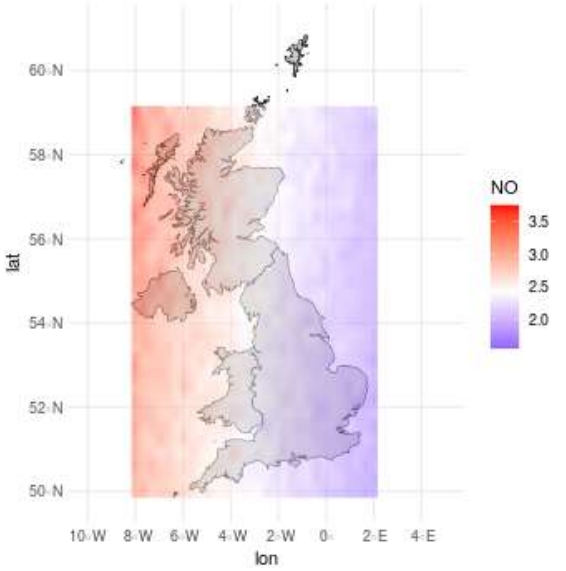

(a) M1.




(b) M2.

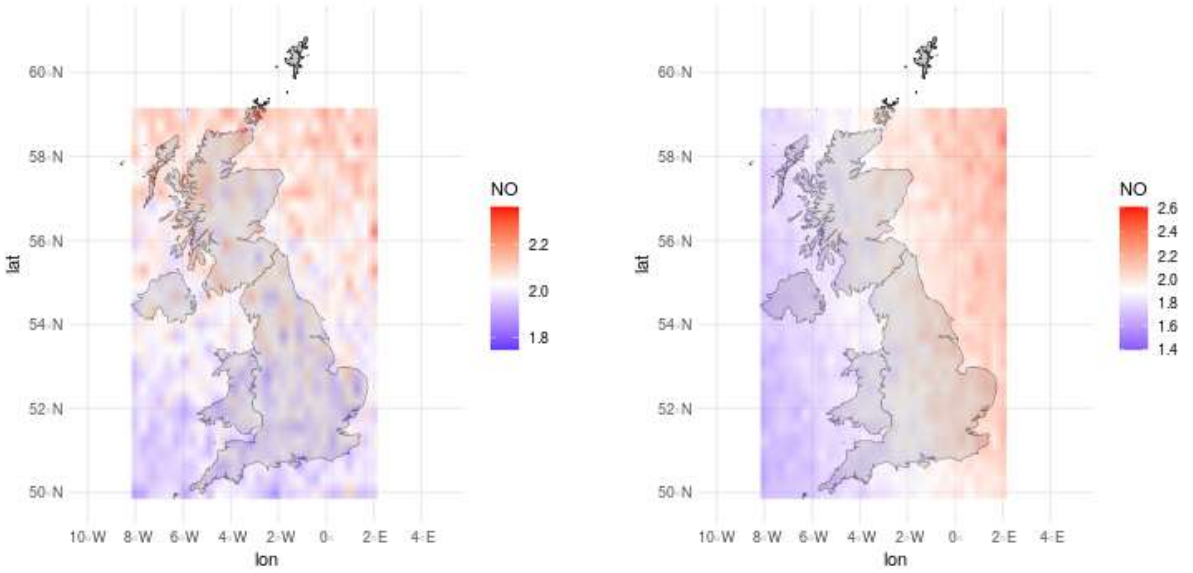



(c) M3.



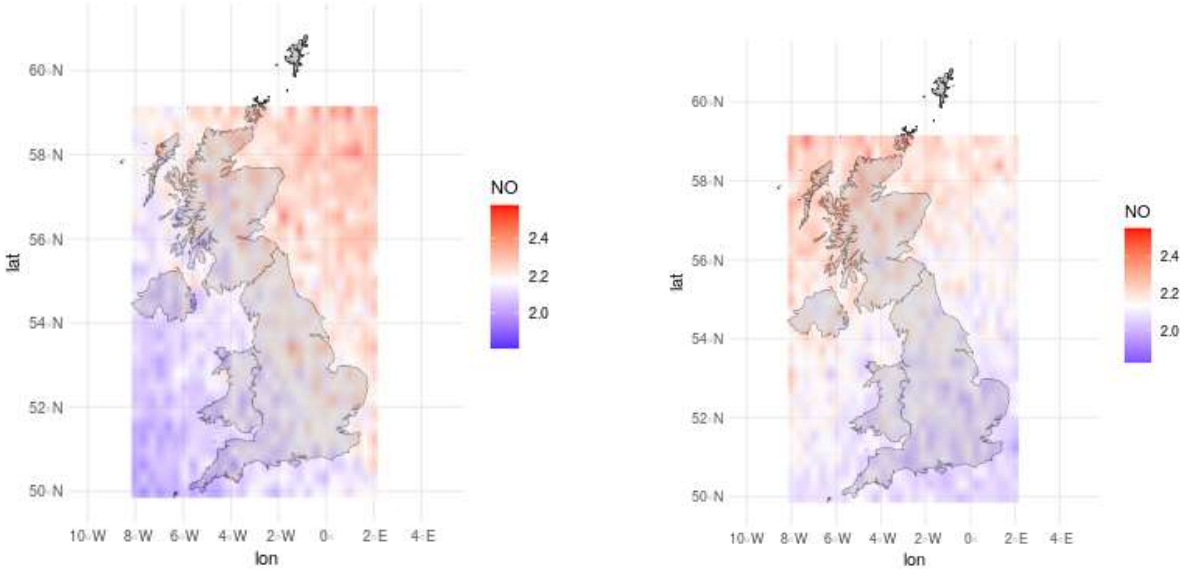

(d) M4.

Figura 52 – Interpolação da concentração de $NO$ ao longo de um grid regular construído ao redor do Reino Unido, na 1ª semana do período analisado. A cor azul indica níveis mais baixos, a cor branca representa níveis intermediários e a cor vermelha, níveis elevados de concentração do poluente.




(a) M1.

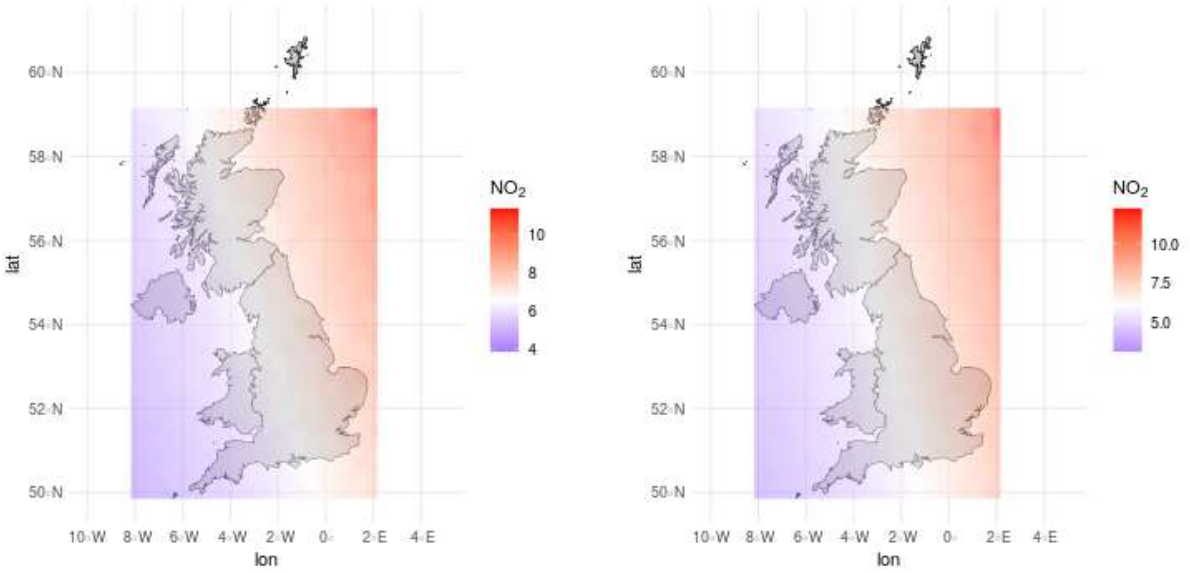
(b) M2.




(c) M3.

(d) M4.

Figura 53 – Interpolação da concentração de $NO$ ao longo de um grid regular construído ao redor do Reino Unido, na 52ª semana do período analisado. A cor azul indica níveis mais baixos, a cor branca representa níveis intermediários e a cor vermelha, níveis elevados de concentração do poluente.



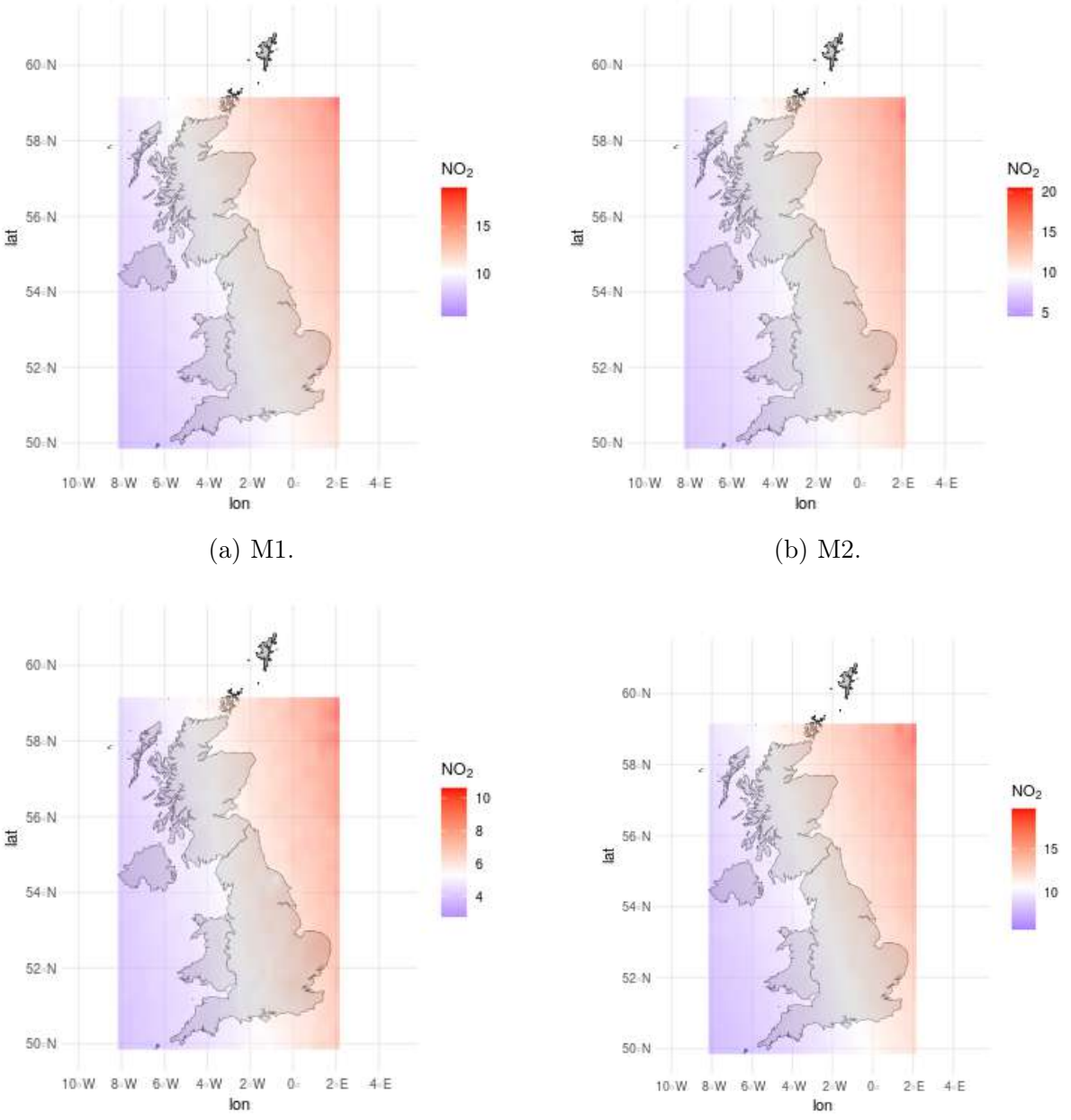

(a) M1.

(b) M2.

(c) M3.

(d) M4.

Figura 54 – Interpolação da concentração de $NO_2$ ao longo de um grid regular construído ao redor do Reino Unido, na 1ª semana do período analisado. A cor azul indica níveis mais baixos, a cor branca representa níveis intermediários e a cor vermelha, níveis elevados de concentração do poluente.



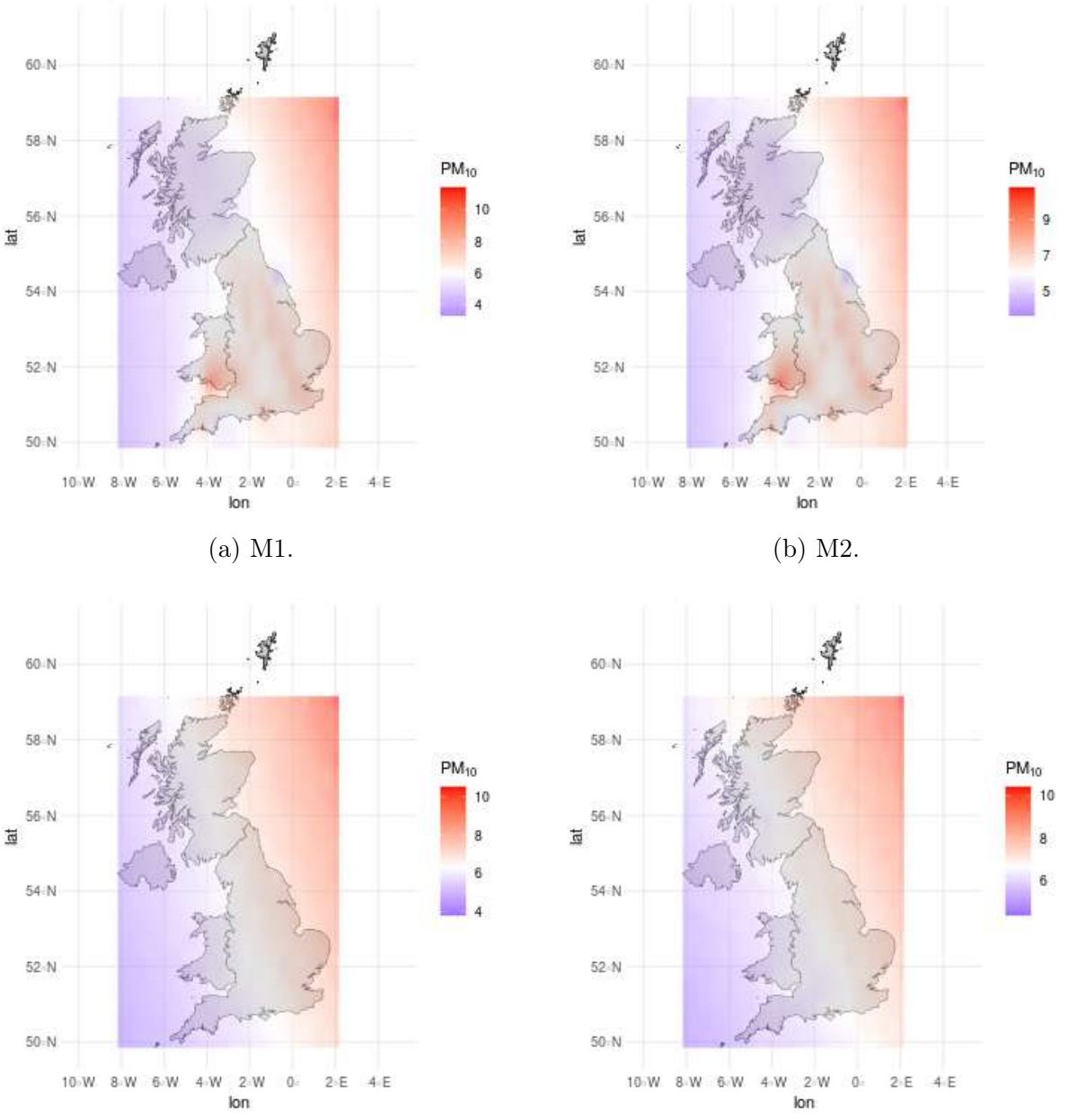

(a) M1.

(b) M2.

(c) M3.

(d) M4.

Figura 55 – Interpolação da concentração de $NO_2$ ao longo de um grid regular construído ao redor do Reino Unido, na 52ª semana do período analisado. A cor azul indica níveis mais baixos, a cor branca representa níveis intermediários e a cor vermelha, níveis elevados de concentração do poluente.



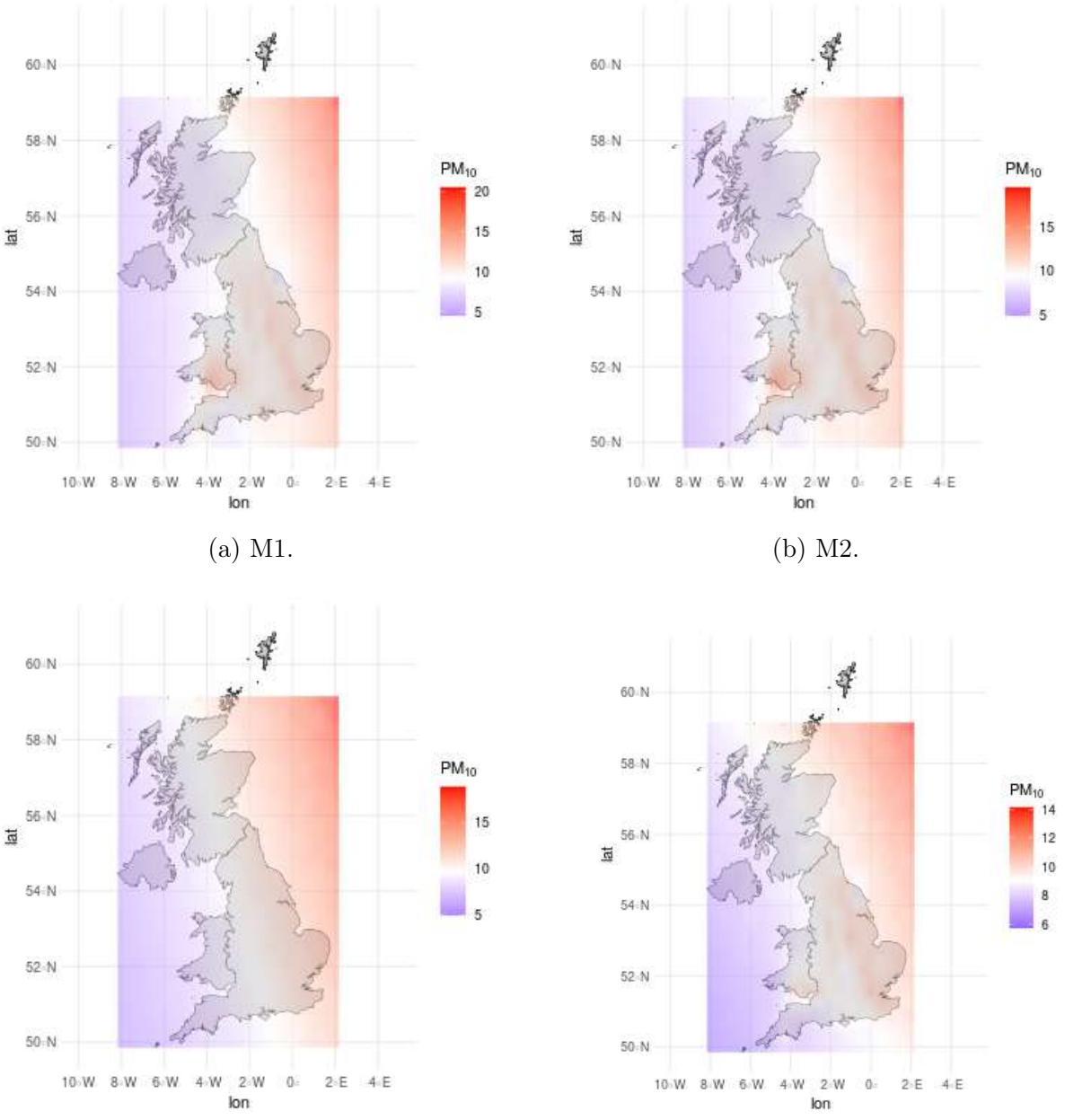

(a) M1.

(b) M2.

(c) M3.

(d) M4.

Figura 56 – Interpolação da concentração de $PM_{10}$ ao longo de um grid regular construído ao redor do Reino Unido, na 1ª semana do período analisado. A cor azul indica níveis mais baixos, a cor branca representa níveis intermediários e a cor vermelha, níveis elevados de concentração do poluente.



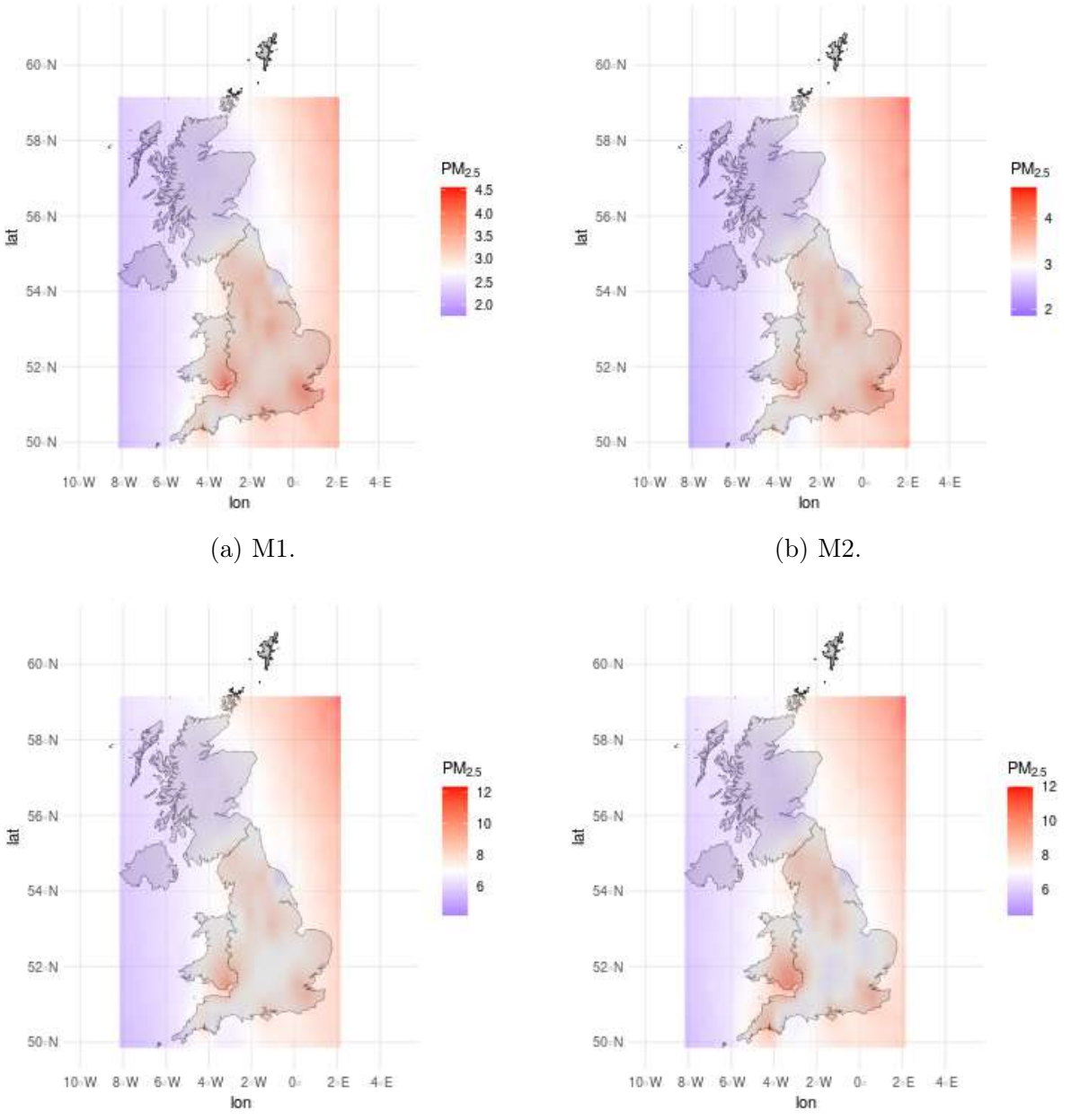

(a) M1.

(b) M2.

(c) M3.

(d) M4.

Figura 57 – Interpolação da concentração de $PM_{10}$ ao longo de um grid regular construído ao redor do Reino Unido, na 52ª semana do período analisado. A cor azul indica níveis mais baixos, a cor branca representa níveis intermediários e a cor vermelha, níveis elevados de concentração do poluente.



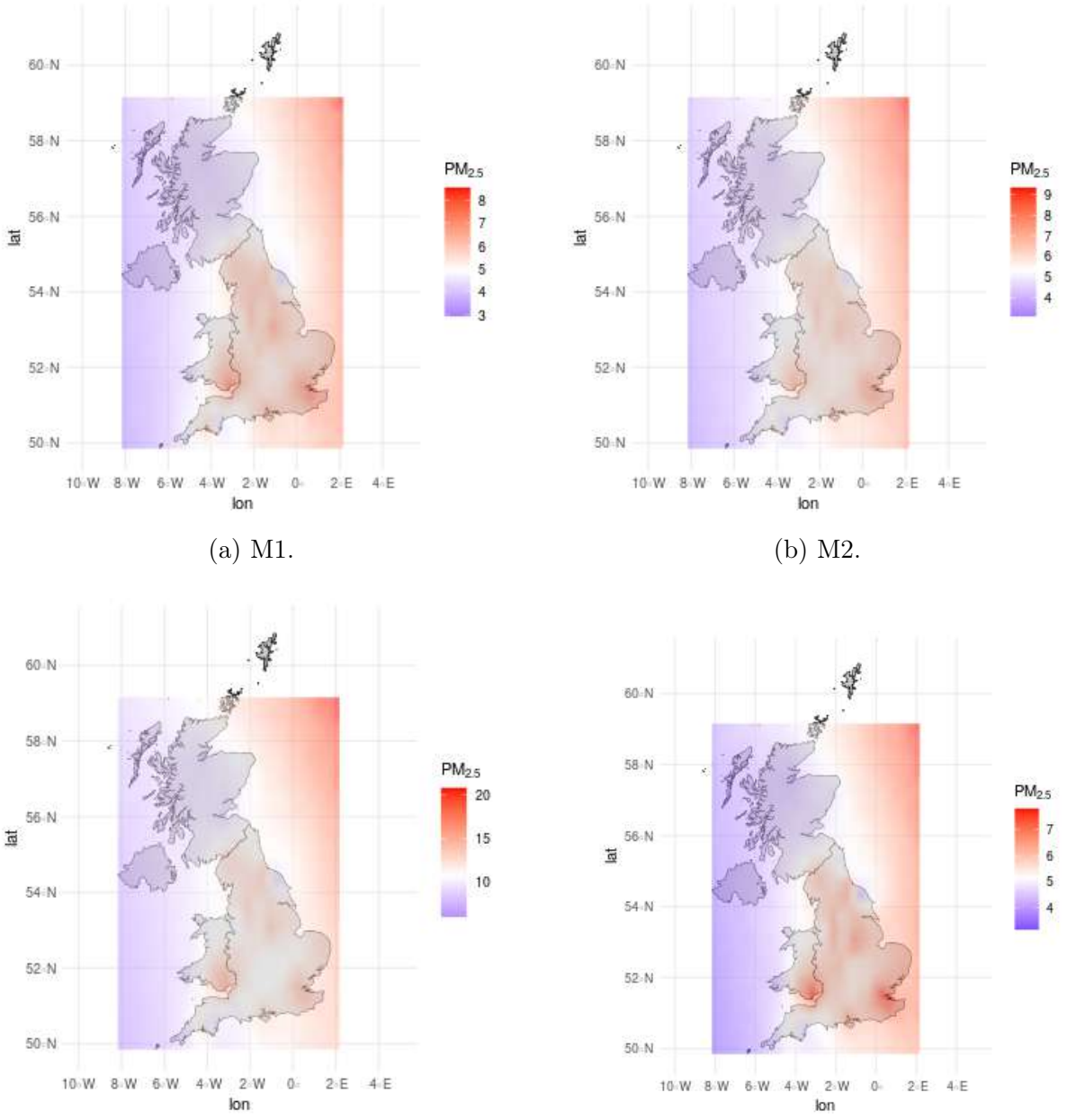

(a) M1.

(b) M2.

(c) M3.

(d) M4.

Figura 58 – Interpolação da concentração de $PM_{2.5}$ ao longo de um grid regular construído ao redor do Reino Unido, na 1ª semana do período analisado. A cor azul indica níveis mais baixos, a cor branca representa níveis intermediários e a cor vermelha, níveis elevados de concentração do poluente.




(a) M1.

(b) M2.

(c) M3.

(d) M4.

Figura 59 – Interpolação da concentração de $PM_{2.5}$ ao longo de um grid regular construído ao redor do Reino Unido, na 52ª semana do período analisado. A cor azul indica níveis mais baixos, a cor branca representa níveis intermediários e a cor vermelha, níveis elevados de concentração do poluente.