



MARCEL IRVING PEREIRA MELO

**DISTRIBUIÇÃO ESPAÇO-TEMPORAL DO NÚMERO E DO
VALOR DE AUTOS DE INFRAÇÃO AMBIENTAL EM MINAS
GERAIS DE 2008 A 2016**

LAVRAS – MG

2020

MARCEL IRVING PEREIRA MELO

**DISTRIBUIÇÃO ESPAÇO-TEMPORAL DO NÚMERO E DO VALOR DE AUTOS DE
INFRAÇÃO AMBIENTAL EM MINAS GERAIS DE 2008 A 2016**

Tese apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Estatística e Experimentação Agropecuária, para a obtenção do título de Doutor.

Professor Júlio Sílvio de Sousa Bueno Filho
Orientador

LAVRAS – MG

2020

**Ficha catalográfica elaborada pelo Sistema de Geração de Ficha Catalográfica da Biblioteca
Universitária da UFLA, com dados informados pelo(a) próprio(a) autor(a).**

Melo, Marcel Irving Pereira

Distribuição espaço-temporal do número e do valor de
autos de infração ambiental em Minas Gerais de 2008 a 2016 /
Marcel Irving Pereira Melo. – 2020.

137 p. :

Orientador(a): Júlio Sílvio de Sousa Bueno Filho.

Tese (doutorado)–Universidade Federal de Lavras, 2020.
Bibliografia.

1. Distribuições conjuntas. 2.Inferência bayesiana 3.INLA.
I. Bueno Filho, Júlio Sílvio de Sousa. II. Título.

MARCEL IRVING PEREIRA MELO

**DISTRIBUIÇÃO ESPAÇO-TEMPORAL DO NÚMERO E DO VALOR DE AUTOS DE
INFRAÇÃO AMBIENTAL EM MINAS GERAIS DE 2008 A 2016
SPATIO-TEMPORAL DISTRIBUTION OF THE NUMBER AND VALUE OF
ENVIRONMENTAL INFRACTION NOTICES IN MINAS GERAIS FROM 2008 TO
2016**

Tese apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Estatística e Experimentação Agropecuária, para a obtenção do título de Doutor.

APROVADA em 12 de Novembro de 2020.

Prof. DSc. Elias Teixeira Krainski	UFPR
Prof. DSc. Danilo Machado Pires	UNIFAL-MG
Prof ^a . DSc. Camila Marques Barroso	UFLA
Prof. DSc. Renato Ribeiro de Lima	UFLA

Professor Júlio Sílvio de Sousa Bueno Filho
Orientador

**LAVRAS – MG
2020**

AGRADECIMENTOS

À minha mãe Neusa e meu pai Osvaldo, sempre dedicados na minha educação e formação humana, serei eternamente grato pelo amor, carinho e esforço que sempre tiveram por mim.

À minha querida Renata, pelo amor, carinho, dedicação, companheirismo, respeito e por todos esses anos que compartilhamos juntos. Pelas sugestões, críticas e auxílios para a realização do trabalho.

Aos meus irmãos Rafael e Roger pela amizade e respeito.

À minha segunda família, meu sogros Sebastião e Maria de Fátima e meus cunhados Renan e Rosana, pela amizade, respeito e apoio.

Ao professor Júlio Sílvio de Sousa Bueno Filho, pela orientação, pelos ensinamentos, incentivos e confiança.

Aos membros da banca, pelos ensinamentos, críticas e sugestões.

Aos colegas, professores e servidores do Programa de Pós-Graduação em Estatística e Experimentação Agropecuária.

Muito obrigado!

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

*Encumbered forever by desire and ambition
There's a hunger still unsatisfied
Our weary eyes still stray to the horizon
Though down this road we've been so many times
(David Jon Gilmour / Polly Anne Samson)*

RESUMO

O objetivo desta tese é sugerir uma modelagem estatística para o número e valor agregados dos autos de infração ambiental ocorridos nos municípios de Minas Gerais entre 2008 e 2016. Dados de tal importância, tomados em base municipal, careciam de forma integrada de análise. Foram implementados modelos bayesianos espaço-temporais com diferentes funções de verossimilhanças para modelar aspectos de autos de infração: binomiais para presença ou ausência, Poisson para número de ocorrências e gama para o valor das ocorrências. Inicialmente, foi testado o ajuste individual dessas e de outras distribuições, sendo escolhido o modelo com os melhores ajustes para montar a verossimilhança conjunta. O ajuste dos modelos foi possível com a utilização do INLA (e o pacote *R-INLA*). Isso permitiu que análises conjuntas complexas fossem realizadas em, aproximadamente, 10 horas de computação, utilizando uma máquina com 24 núcleos. A tendência de crescimento temporal tanto do número, quanto do valor dos autos de infrações ambientais é indício da pouca eficácia da repressão de ações lesivas ao meio ambiente em Minas Gerais. Este modelo pode ser adaptado para diversas situações semelhantes, como, por exemplo, nas áreas de ecologia, economia e em epidemiologia.

Palavras-chave: Infrações ambientais. Distribuição Poisson. Distribuição gama. Distribuições conjuntas. Inferência bayesiana. INLA.

ABSTRACT

The aim of this thesis is to present statistical modelling for number and aggregate value of environmental infraction notices that happen in Minas Gerais from 2008 to 2016. Such important data at municipality basis lack integrated analysis. We implemented Bayesian space-time modelling with different likelihood functions: binomial for presence/absence of notices, poisson for the number of notices in the year and gamma for the aggregate value of the fines in the year. This and other candidate likelihood functions were initially fitted to one-dimensional data and joint likelihood was built with best fitting models. Fitting was carried out using INLA (R-INLA package). This allowed complex joint modelling was performed in about 10 hours of computing time using a 24 cores machine. Time trends of increasing number of notices and fine values are evidence of lack of efficiency in law enforcement on environmental infractions in MG. This model can be adapted to similar cases. For example, in the fields of science: ecological, economic and epidemiological.

Keywords: Bayesian Inference. Joint distribution. Environmental infraction. Gamma distribution. Poisson distribution. INLA.

LISTA DE FIGURAS

Figura 2.1 – Representação de independência condicional em um grafo não orientado	20
Figura 2.2 – Estrutura de vizinhança: vizinhos de primeira ordem (à esquerda) e vizinhos de primeira e segunda ordem (à direita)	31
Figura 3.1 – Mapa das regiões intermediárias do estado de Minas Gerais	51
Figura 3.2 – Mapa das regiões imediatas do estado de Minas Gerais	52
Figura 3.3 – Mapa do estado de Minas Gerais com sua respectiva estrutura de vizinhança entre seus municípios. Os limites dos municípios são descritos pela borda (cor branca). A definição de vizinhança entre municípios se dá por aqueles que compartilham de uma mesma borda. A estrutura de vizinhança é informada pelo grafo (cor azul); os nós localizados nos centroides de cada município representam o próprio município enquanto que as arestas que conectam os nós dos municípios representam que estes são vizinhos	53
Figura 3.4 – Distribuição espacial dos biomas presentes nos municípios do estado de Minas Gerais	54
Figura 4.1 – Número de autos de infração observados nos municípios de Minas Gerais de 2008 a 2016 (em escala raiz quadrada)	62
Figura 4.2 – Média a posteriori do efeito aleatório espacial para os municípios de Minas Gerais para modelo final com verossimilhança Poisson e preditor linear com a configuração V para o número de autos de infração em Minas Gerais no período de 2008 a 2016	66
Figura 4.3 – Média a posteriori (pontos), intervalo de credibilidade de 95% (faixas) e ajuste via LOESS (curva) do efeito temporal do número de autos de infrações ambientais dos municípios de MG entre os anos de 2008 e 2016	67
Figura 4.4 – Interações espaços-temporais para o número de autos	68
Figura 4.5 – Valores preditos <i>versus</i> observados do número de autos de infração ambientais do modelo final com verossimilhança Poisson e preditor linear na configuração V	69
Figura 4.6 – Valores preditos do número de autos de infração ambientais do modelo final com verossimilhança Poisson e preditor linear na configuração V (na escala raiz quadrada)	69

Figura 4.7 – Ocorrência de autos de infração observados nos municípios de Minas Gerais de 2008 a 2016	70
Figura 4.8 – Valor dos autos de infração observados nos municípios de Minas Gerais de 2008 a 2016 (em escala logarítmica)	71
Figura 4.9 – Média a posteriori do efeito aleatório espacial municipal referente ao modelo final <i>hurdle</i> -gama para a verossimilhança binomial e preditor linear com a configuração VIII para a ocorrência ou não de autos de infração em Minas Gerais no período de 2008 a 2016	75
Figura 4.10 – Média a posteriori do efeito aleatório espacial municipal do modelo final <i>hurdle</i> -gama para a verossimilhança gama e preditor linear com a configuração VII para o valor de autos de infração em Minas Gerais no período de 2008 a 2016	76
Figura 4.11 – Média a posteriori (pontos), intervalo de credibilidade de 95% (faixas) e ajuste via LOESS (curva) do efeito temporal da ocorrência (esquerda) e valor (direita) de autos de infrações ambientais dos municípios de MG entre os anos de 2008 e 2016	76
Figura 4.12 – Média a posteriori da interação para a ocorrência de autos	77
Figura 4.13 – Valores preditos da probabilidade de ocorrência de autos de infração ambientais do modelo final <i>hurdle</i> -Gama com verossimilhança binomial e preditor linear na configuração VIII e verossimilhança gama e preditor linear VII	78
Figura 4.14 – Média a posteriori da interação para o valor de autos	79
Figura 4.15 – Preditos <i>versus</i> observados para o valor dos autos de infrações ambientais ocorridos nos municípios de MG entre os anos de 2008 e 2016	80
Figura 4.16 – Valores preditos dos valores agregados dos autos de infração ambientais do modelo final <i>hurdle</i> -Gama com verossimilhança binomial e preditor linear na configuração VIII e verossimilhança gama e preditor linear VII (escala logarítmica)	81
Figura 4.17 – Média a posteriori do efeito aleatório espacial municipais do modelo final conjunto para ocorrência de autos de infração ambientais em Minas Gerais no período de 2008 a 2016 – efeito espacial para número e valor possuem mapas similares, sendo escalados por λ_{21} e λ_{31} , respectivamente	86

Figura 4.18 – Média a posteriori (pontos), intervalo de credibilidade de 95% (faixas) e ajuste via LOESS (curva) do efeito temporal do modelo conjunto para os autos de infrações ambientais dos municípios de MG entre os anos de 2008 e 2016 – efeito temporal para número e valor possuem mapas similares, sendo escalados por λ_{22} e λ_{32} , respectivamente	87
Figura 1 – Distribuição marginal a posteriori das covariáveis do modelo final poisson para o número de autos de infração em Minas Gerais no período de 2008 a 2016	97
Figura 2 – Distribuição marginal a posteriori dos hiperparâmetros do modelo final poisson para o número de autos de infração em Minas Gerais no período de 2008 a 2016	98
Figura 3 – Distribuição marginal a posteriori das covariáveis do modelo final <i>hurdle</i> -gama para o valor dos autos de infração em Minas Gerais no período de 2008 a 2016	99
Figura 4 – Distribuição marginal a posteriori dos hiperparâmetros referente ao preditor da verossimilhança binomial do modelo final <i>hurdle</i> -gama para o valor de autos de infração em Minas Gerais no período de 2008 a 2016	100
Figura 5 – Distribuição marginal a posteriori dos hiperparâmetros referente ao preditor da verossimilhança gama do modelo final <i>hurdle</i> -gama para o valor de autos de infração em Minas Gerais no período de 2008 a 2016	101
Figura 6 – Média a posteriori da interação para a ocorrência de autos do modelo final conjunto para ocorrência, número e valor dos autos de infração ambientais em Minas Gerais no período de 2008 a 2016	102
Figura 7 – Média a posteriori da interação para o número de autos do modelo final conjunto para ocorrência, número e valor dos autos de infração ambientais em Minas Gerais no período de 2008 a 2016	103
Figura 8 – Média a posteriori da interação para o valor de autos do modelo final conjunto para ocorrência, número e valor dos autos de infração ambientais em Minas Gerais no período de 2008 a 2016	104
Figura 9 – Distribuições marginais a posteriori das covariáveis do modelo final conjunto para o número e valor de autos de infração em Minas Gerais no período de 2008 a 2016	105

Figura 10 – Distribuições marginais a posteriori dos hiperparâmetros do modelo final conjunto para o número e valor de autos de infração em Minas Gerais no período de 2008 a 2016	106
Figura 11 – Valores preditos da probabilidade de ocorrência dos autos de infração ambientais do modelo final conjunto com verossimilhanças binomial, Poisson e gama	107
Figura 12 – Valores preditos do número de autos de infração ambientais do modelo final conjunto com verossimilhanças binomial, Poisson e gama (escala raiz quadrada)	108
Figura 13 – Valores preditos do valor agregado de autos de infração ambientais do modelo final conjunto com verossimilhanças binomial, Poisson e gama (escala logarítmica)	109

LISTA DE TABELAS

Tabela 3.1 – Tipo de Intervenção ocorrida em Áreas de Preservação Permanente e Reserva Legal entre 2008 e 2016, no estado de Minas Gerais	55
Tabela 3.2 – Configurações das componentes espaciais, temporais e espaço-temporais consideradas para o ajuste dos modelos univariados para a resposta número e valor de autuações ambientais. Todos os modelos utilizaram $f(s) = bym2$	59
Tabela 4.1 – Valores WAIC e RMSE para os modelos testados para o número de autos de infrações ambientais nos municípios de Minas Gerais para o período de 2008 a 2016 (em negrito está a combinação com melhor desempenho) . . .	63
Tabela 4.2 – Média, desvio padrão e intervalos de credibilidade a posteriori (95%) para as covariáveis e hiperparâmetros do modelo final com verossimilhança Poisson e preditor linear com a configuração espaço-temporal V	64
Tabela 4.3 – Valores das medidas WAIC, Brier Score e LogScore dos modelos ajustados para ocorrência de autos de infração nos municípios de Minas Gerais, considerando preditores cujas configurações estão descritas na Tabela 3.2 . . .	72
Tabela 4.4 – Valores WAIC e RMSE para os modelos testados para o valor de autos de infrações ambientais nos municípios de Minas Gerais para o período de 2008 a 2016 (todas as combinações dos modelos utilizaram o preditor VIII para a verossimilhança binomial. Em negrito encontra-se a combinação com melhor desempenho)	72
Tabela 4.5 – Média, desvio padrão e intervalos de credibilidade a posteriori (95%) para as covariáveis e hiperparâmetros do modelo final <i>hurdle</i> -Gama, com respectivos preditores lineares com a configuração espaço-temporal <i>VIII</i> e <i>VII</i>	73
Tabela 4.6 – Modelos considerados para a modelagem conjunta, de ocorrência, número e valor de autos de infrações em Minas Gerais para os anos de 2008 a 2016.	82
Tabela 4.7 – Comparação das medidas WAIC, RMSE e LogScore entre os modelos finais Poisson, <i>hurdle</i> -Gama e conjunto	83
Tabela 4.8 – Média, desvio padrão e intervalos de credibilidade a posteriori (95%) para as covariáveis e hiperparâmetros do modelo final conjunto para ocorrência, número e valor de autos ambientais (<i>VI</i>)	84

SUMÁRIO

1	INTRODUÇÃO	13
2	REFERENCIAL TEÓRICO	16
2.1	Campos Aleatórios Markovianos Gaussianos	16
2.1.1	Vetores e matrizes	16
2.1.2	Normal Multivariada	18
2.1.3	Independência Condicional	20
2.2	Modelos Latentes Gaussianos	23
2.2.1	Integrated Nested Laplace Approximation (INLA)	26
2.3	Modelos Espaciais Latentes Gaussianos	28
2.3.1	Modelos espaciais para dados de área	30
2.3.2	Modelo Besag	31
2.3.3	Modelo BYM	33
2.3.4	Modelo de Leroux	33
2.3.5	Modelo de Dean	34
2.3.6	Modelo BYM2	34
2.4	Modelos temporais	37
2.4.1	Modelo de efeitos aleatórios autorregressivos (AR)	37
2.4.2	Passeio Aleatório (RW)	39
2.5	Modelos espaços-temporais	40
2.5.1	Interação Tipo I	41
2.5.2	Interação Tipo II	41
2.5.3	Interação Tipo III	41
2.5.4	Interação Tipo IV	42
2.6	Análise Conjunta Combinada	42
2.7	Prioris de Complexidade Penalizada (PC <i>priors</i>)	44
2.7.1	Prioris de complexidade penalizada para a precisão de um efeito aleatório gaussiano	46
2.7.2	Priori de complexidade penalizada para o parâmetro de mistura ϕ no modelo BYM2	47
2.7.3	Priori de complexidade penalizada para o parâmetro de correlação ρ de um modelo AR(1) com modelo base $\rho = 0$	48

3	MATERIAL E MÉTODOS	50
3.1	Área de Estudo e Dados	50
3.2	Modelagem espaço-temporal bayesiana para o número e para o valor dos autos de infrações ambientais	55
3.2.1	Número de autos	55
3.2.2	Valor dos autos	56
3.3	Seleção de modelos para número e valor de autos de infrações ambientais	58
3.4	Modelagem Espaço-Temporal Bayesiana Conjunta	60
4	RESULTADOS	62
4.1	Modelo espaço-temporal bayesiano final para o número de autos	62
4.2	Modelo hurdle-Gama para o valor de autos	70
4.3	Modelagem bayesiana conjunta para ocorrência, número e valor de autos	81
5	CONCLUSÃO	89
	REFERÊNCIAS	90
	APENDICE A – Divergência de Kullback-Leibler para Distribuições Gaussianas	94
	APENDICE B – Figuras adicionais	97
	APENDICE C – Rotinas R	110

1 INTRODUÇÃO

A preservação do meio ambiente é um assunto de relevância global. A ocorrência de diversas degradações ambientais tais como queimadas, lançamento de dejetos em lagoas, supressão da vegetação, assoreamentos e tantas outras que são acometidas em áreas que deveriam ser permanentemente preservadas trazem prejuízos para a vida no planeta e o bem estar social. A perda de biodiversidade é um desses prejuízos, pois é prejudicial em aspectos evolutivos, históricos e filogenéticos, assim como a mudança climática que aflige cada vez mais a saúde da população, como aumento da frequência e da intensidade de ondas de calor, secas, chuvas extremas, ciclones (WHO et al., 2020).

De acordo com SILVA, Florestal et al. (2012), o Brasil é o país com maior diversidade biológica mundial, isso porque, dispõe de ao menos 20% das espécies do planeta, com altas taxas de endemismo para diferentes grupos taxonômicos. Da mesma forma que isso gera oportunidades econômicas no sentido da criação e desenvolvimento de novos alimentos, fármacos, bioterápicos, madeiras e fibras, tecnologias biomiméticas e o turismo ecológico, gera também uma grande responsabilidade na preservação destes recursos para as futuras gerações.

Os mesmos autores salientam que o primeiro passo para preservação dos recursos naturais e sustentabilidade na agricultura ocorre pelo uso adequado das terras. No entanto, ainda que houveram avanços na agricultura conservacionistas, o processo histórico de ocupação do território brasileiro resultou, em alguns casos, no aumento das pressões sobre o meio ambiente, em processos erosivos, perda de biodiversidade, na contaminação ambiental e em desequilíbrios sociais. Desse modo, o desperdício dos recursos naturais decorrente do uso inadequado das terras é uma realidade a ser enfrentada. Para estancar esse quadro, os autores salientam que as Áreas de Preservação Permanente (APP) e Reservas Legais (RL) deveriam ser consideradas como parte fundamental do planejamento agrícola conservacionista das propriedades.

No Brasil, a Lei 12651/2012, denominada de Lei de Proteção de Vegetação Nativa (LPVN) e, popularmente, conhecida como Novo Código Florestal, define sobre como a norma para a proteção nativa, em especial, Áreas de Preservação Permanente, Reserva Legal e Uso Restrito e estabelece que os estados brasileiros deverão implementar políticas públicas para regularização ambiental, estabelecendo programas que cumpram com essa finalidade.

É declarado que na LPVN, Áreas de Preservação Permanente, cobertas ou não por vegetação nativa possuem função ambiental de preservar recursos hídricos, a paisagem, a estabi-

lidade geológica e a biodiversidade, facilitar o fluxo gênico de fauna e flora, proteger o solo e assegurar o bem estar das populações.

Dois aspectos refletem a importância das APP: parte física do sistema agrícola e relacionada aos serviços ecológicos fornecidos pela flora existente, incluindo todas as suas conexões, tais como os componentes biológicos e não biológicos do ecossistema agrícola (BORGES; REZENDE, 2012; TUNDISI; TUNDISI, 2010). Percebe-se que as APP constituem o equilíbrio dos processos ecológicos naturais, sendo assim, de extrema importância. De acordo com ODUM (2001) o fluxo de troca de materiais entre a parte biológica e não biológica e a biodiversidade claramente definida fazem com que o ecossistema seja uma unidade funcional ecológica básica. Mudanças na unidade ecológica, que conectam seres vivos por meio da interação de efeitos complexos que afetam as propriedades dos componentes um dos outros, irão prejudicar sistematicamente a qualidade do meio ambiente. Contudo, mesmo diante de tamanha importância das APP e implicações devidas à intervenção humana nessas áreas, as regras do novo Código Florestal se tornaram mais brandas.

Haja vista a importância das APPs e RL, somada ao cenário de redução da obrigação legal de preservação e recuperação dessas áreas, é notável a necessidade de monitorar e avaliar as infrações ambientais ocorridas nesses locais. Apesar disso, existem poucos trabalhos na literatura abordando este aspecto. Sobretudo, se considerar a utilização de modelagem estatística, a existência de estudos relacionados é escassa.

Dessa forma, o intuito maior do presente trabalho é incorporar uma modelagem estatística para analisar as infrações ambientais sobre as APP e RL, utilizando dados do número e valor agregados dos autos de infração ocorridos nos municípios de Minas Gerais durante o período de 2008 a 2016.

Dentro da abordagem estatística, análises espaciais, espaço-temporais de dados de frequência e/ou severidade, como por exemplo, os dados de autos de infração, são comumente realizadas através de formulações hierárquicas bayesianas. Alguns trabalhos nas áreas de epidemiologia e ecologia foram propostos, tais como (BESAG; YORK; MOLLÍÉ, 1991; BERNARDINELLI; CLAYTON; MONTOMOLI, 1995; KNORR-HELD, 2000; LEROUX; LEI; BRESLOW, 2000). Para fazer a inferência bayesiana desses modelos, a abordagem *Integrated Nested Laplace Approximation* (INLA) tem se mostrado uma abordagem computacional mais eficiente que aquelas baseadas em métodos de Monte Carlo via Cadeias de Markov (MCMC) (RUE; MARTINO; CHOPIN, 2009). A partir de então, trabalhos como os de Schrödle e Held (2011),

Blangiardo et al. (2013), Cosandey-Godin et al. (2015) e Goicoa et al. (2018), utilizando o INLA, foram publicados.

Embora a análise univariada de modelos espaços-temporais forneça ao pesquisador várias soluções para suas questões científicas, questões relativas às interações entre variáveis respostas ainda ficam sem solução e requerem uma análise conjunta. Alguns trabalhos propuseram a utilização de modelagem conjunta espaço-temporal com verossimilhanças Bernoulli (NTI-RAMPEBA; NEEMA; KAZEMBE, 2017), verossimilhanças Poisson inflacionadas de zero (ASMARIAN et al., 2019) e verossimilhanças binomial e gama (SADYKOVA et al., 2017). Forlani et al. (2020) propuseram uma abordagem espaço-temporal conjunta para dados espacialmente desalinhados representando campos espacialmente contínuos com modelos de Equações Diferenciais Parciais Estocásticas (SPDE) utilizando verossimilhanças normais, através da metodologia INLA, que trouxeram melhor qualidade nos resultados, que se tornaram mais precisos (LINDGREN; RUE; LINDSTRÖM, 2011; KRAINSKI et al., 2018). No entanto, não foram encontrados estudos que modelam conjuntamente verossimilhanças binomial, Poisson e gama no contexto de modelos espaço-temporais, na área de ecologia.

Portanto, o objetivo deste trabalho é avaliar metodologias para:

- 1) implementar soluções INLA para analisar cada uma das variáveis (presença, número e valor de autos de infração ambiental) que possam ser enquadradas em campos latentes gaussianos;
- 2) implementar um modelo para a distribuição conjunta das variáveis em estudo, considerando distribuições discretas e contínuas que se enquadram nas características das variáveis;
- 3) avaliar as propriedades descritivas dos modelos resultantes e verificar se há vantagens na modelagem conjunta, que certamente é mais complexa.
- 4) apreciar, ao menos inicialmente, conclusões práticas decorrentes desta análise que possam subsidiar discussão técnica com especialistas da área de ciências ambientais.

2 REFERENCIAL TEÓRICO

Neste capítulo apresentaremos a base teórica e o estado da arte da modelagem bayesiana utilizando a aproximação de Laplace aninhada integrada (INLA, do inglês *Integrated Nested Laplace Approximation*) que serão implementadas na tese.

2.1 Campos Aleatórios Markovianos Gaussianos

Resumidamente, um Campo Aleatório Markoviano Gaussiano (GMRF, do inglês *Gaussian Markov Random Fields*) consiste em um vetor aleatório normal multivariado que assume algumas propriedades de independência condicional (RUE; HELD, 2005).

Antes de abordarmos o tema de Campos Aleatórios Markovianos Gaussianos, vamos apresentar alguns conceitos iniciais que julgamos dar maior fluidez na leitura do texto.

2.1.1 Vetores e matrizes

Denotamos por $\mathbf{x} = [x_1, \dots, x_n]^\top$ o vetor n -dimensional, tal que x_i é o i -ésimo elemento, $\mathbf{x}_A$ é o vetor cujos elementos x_i são tais que i pertence ao conjunto A , isto é, $\mathbf{x}_A = \{x_i ; i \in A\}$. Analogamente, $\mathbf{x}_{-A} = \{x_i ; i \notin A\}$. Temos ainda que, $\mathbf{x}_{i:j} = [x_i, x_{i+1}, \dots, x_j]^\top$ para $j \geq i$.

Uma matriz $\mathbf{A}$ de ordem $(n \times p)$ é o arranjo de números dado por

$$\mathbf{A} = \begin{bmatrix} A_{11} & A_{12} & \dots & A_{1p} \\ A_{21} & A_{22} & \dots & A_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ A_{n1} & A_{n2} & \dots & A_{np} \end{bmatrix} = [A_{ij}],$$

tal que A_{ij} é o elemento da i -ésima linha e j -ésima coluna da matriz $\mathbf{A}$. Denota-se por $\mathbf{A}^\top$ a matriz transposta de $\mathbf{A}$, de tal modo que, $\mathbf{A}^\top = [A_{ji}]$. No caso em que $\mathbf{A} = \mathbf{A}^\top$, dizemos que a matriz $\mathbf{A}$ é simétrica.

Dadas duas matrizes $\mathbf{A}$ ($m \times n$) e $\mathbf{B}$ ($p \times q$), tem-se que o Produto de Kronecker das matrizes $\mathbf{A}$ e $\mathbf{B}$ resulta em uma matriz $\mathbf{C}$ ($mp \times nq$) de tal forma que

$$\begin{aligned} \mathbf{C} &= \mathbf{A} \otimes \mathbf{B} \\ &= \begin{bmatrix} A_{11}\mathbf{B} & A_{12}\mathbf{B} & \dots & A_{1n}\mathbf{B} \\ A_{21}\mathbf{B} & A_{22}\mathbf{B} & \dots & A_{2n}\mathbf{B} \\ \vdots & \vdots & \ddots & \vdots \\ A_{m1}\mathbf{B} & A_{m2}\mathbf{B} & \dots & A_{mn}\mathbf{B} \end{bmatrix}. \end{aligned}$$

Dada duas matrizes $\mathbf{A}$ e $\mathbf{B}$ de mesma ordem, o Produto de Hadamard é dado por uma matriz $\mathbf{D}$, de tal forma que

$$\begin{aligned} \mathbf{D} &= \mathbf{A} \odot \mathbf{B} \\ &= \begin{bmatrix} A_{11}B_{11} & A_{12}B_{12} & \dots & A_{1n}B_{1n} \\ A_{21}B_{21} & A_{22}B_{22} & \dots & A_{2n}B_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ A_{m1}B_{m1} & A_{m2}B_{m2} & \dots & A_{mn}B_{mn} \end{bmatrix}. \end{aligned}$$

Dado o vetor n -dimensional $\mathbf{x}$ e a matriz quadrada $\mathbf{A}$ de ordem n , definimos $\text{diag}(\mathbf{x})$ e $\text{diag}(\mathbf{A})$ como

$$\text{diag}(\mathbf{x}) = \begin{bmatrix} x_1 & & & \\ & x_2 & & \\ & & \ddots & \\ & & & x_n \end{bmatrix} \text{ e } \text{diag}(\mathbf{A}) = \begin{bmatrix} A_{11} & & & \\ & A_{22} & & \\ & & \ddots & \\ & & & A_{nn} \end{bmatrix},$$

de modo que os elementos fora da diagonal principal são iguais a zero.

A matriz $\mathbf{I}$ é denominada matriz identidade e é uma matriz diagonal cujos elementos são iguais a 1. Uma matriz $\mathbf{A}$ é dita triangular inferior se $A_{ij} = 0$ sempre que $i < j$ e triangular superior se $A_{ij} = 0$ sempre que $i > j$.

Dado um vetor n -dimensional $\mathbf{x} \neq \mathbf{0}$ e uma matriz $\mathbf{A}_{(n \times n)}$, classifica-se a matriz $\mathbf{A}$ de acordo com sua forma quadrática dada por $\mathbf{x}^\top \mathbf{A} \mathbf{x}$,

$$\mathbf{A} = \begin{cases} \text{positiva definida,} & \mathbf{x}^\top \mathbf{A} \mathbf{x} > 0 \\ \text{semipositiva definida,} & \mathbf{x}^\top \mathbf{A} \mathbf{x} \geq 0 \\ \text{seminegativa definida,} & \mathbf{x}^\top \mathbf{A} \mathbf{x} \leq 0 \\ \text{negativa definida,} & \mathbf{x}^\top \mathbf{A} \mathbf{x} < 0 \end{cases}$$

De acordo com [Rue e Held \(2005\)](#), se uma matriz $\mathbf{A}_{(n \times n)}$ é simétrica e positiva definida (SPD), então $\mathbf{A}$ tem as seguintes propriedades:

1. O posto da matriz $\mathbf{A}$ é igual a n , isto é, $\text{rank}(\mathbf{A}) = n$.
2. O determinante de $\mathbf{A}$ é maior que zero, isto é, $|\mathbf{A}| > 0$.
3. $A_{ii} > 0$.
4. $A_{ii}A_{jj} - A_{ij}^2 > 0$, para $i \neq j$.
5. $A_{ii} + A_{jj} - 2|A_{ij}| > 0$, para $i \neq j$.
6. $\max(A_{ii}) > \max_{i \neq j} |A_{ij}|$.
7. A inversa da matriz $\mathbf{A}$, denotada por $\mathbf{A}^{-1}$, é SPD.
8. Os autovalores $\lambda_1, \lambda_2, \dots, \lambda_n$ da matriz $\mathbf{A}$ são estritamente positivos.
9. Existe uma matriz $\mathbf{L}$ tal que $\mathbf{A} = \mathbf{L}\mathbf{L}^\top$. Se a matriz $\mathbf{L}$ é triangular inferior, então $\mathbf{L}$ é chamada fator de Cholesky da matriz $\mathbf{A}$.

Mais detalhes e propriedades do tratamento matricial aplicados na estatística podem ser encontrados em [Searle \(1982\)](#) e [Ferreira \(2011\)](#).

2.1.2 Normal Multivariada

Cabe ressaltar aqui, que utilizamos letras maiúsculas em negrito para representar matrizes e letras minúsculas em negrito para representar vetores. Por simplicidade, não há distinção entre notação de variável aleatória (ou vetor aleatório) e uma específica realização da variável ou vetor aleatório. Utiliza-se a notação $\pi(\cdot)$ para representar a densidade de seu argumento, isto é, $\pi(\mathbf{x})$ é a densidade de $\mathbf{x}$ e $\pi(\mathbf{x}_A | \mathbf{x}_{-A})$ é a densidade de $\mathbf{x}_A$ condicionada a realização de $\mathbf{x}_{-A}$.

O símbolo “ $\sim$ ”, significa “é distribuído como”, “tem/segue distribuição”, assim, dessa forma, $\mathbf{x} \sim \mathcal{L}$ significa que o vetor aleatório $\mathbf{x}$ é distribuído de acordo com a lei $\mathcal{L}$.

Isso posto, um vetor finito aleatório $\mathbf{x} = [x_1, x_2, \dots, x_n]^\top$ tem distribuição normal multivariada com vetor de médias $\boldsymbol{\mu}$ e matriz de covariâncias $\boldsymbol{\Sigma}$, se e somente se sua função de densidade de probabilidade é dada por

$$\pi(\mathbf{x}) = (2\pi)^{-\frac{n}{2}} |\boldsymbol{\Sigma}|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu}) \right\}, \mathbf{x} \in \mathbb{R}^n,$$

em que,

$$\boldsymbol{\mu} = [\mu_1, \mu_2, \dots, \mu_n]^\top \text{ e } \boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \dots & \sigma_{1n} \\ \sigma_{21} & \sigma_2^2 & \dots & \sigma_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{n1} & \sigma_{n2} & \dots & \sigma_n^2 \end{bmatrix}.$$

Podemos escrever isto como $\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Temos ainda que $E(x_i) = \mu_i$, $Var(x_i) = \sigma_i^2$ e $Cov(x_i, x_j) = \sigma_{ij}$.

Se decomposmos o vetor $\mathbf{x}$ de modo tal que $\mathbf{x} = (\mathbf{x}_A^\top, \mathbf{x}_B^\top)^\top$ e, consequentemente, em conformidade, particionarmos o vetor de médias e a matriz de covariâncias na seguinte forma:

$$\boldsymbol{\mu} = \begin{bmatrix} \boldsymbol{\mu}_A \\ \boldsymbol{\mu}_B \end{bmatrix} \text{ e } \boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{AA} & \boldsymbol{\Sigma}_{AB} \\ \boldsymbol{\Sigma}_{BA} & \boldsymbol{\Sigma}_{BB} \end{bmatrix}.$$

Temos as seguintes propriedades:

1. $\mathbf{x}_A \sim \mathcal{N}(\boldsymbol{\mu}_A, \boldsymbol{\Sigma}_{AA})$.
2. $\boldsymbol{\Sigma}_{AB} = \mathbf{0}$ se, e somente se $\mathbf{x}_A$ e $\mathbf{x}_B$ são independentes.
3. A distribuição condicional de $\pi(\mathbf{x}_A | \mathbf{x}_B)$ é $\mathcal{N}(\boldsymbol{\mu}_{A|B}, \boldsymbol{\Sigma}_{A|B})$, em que

$$\boldsymbol{\mu}_{A|B} = \boldsymbol{\mu}_A + \boldsymbol{\Sigma}_{AB} \boldsymbol{\Sigma}_{BB}^{-1} (\mathbf{x}_B - \boldsymbol{\mu}_B)$$

$$\boldsymbol{\Sigma}_{A|B} = \boldsymbol{\Sigma}_{AA} - \boldsymbol{\Sigma}_{AB} \boldsymbol{\Sigma}_{BB}^{-1} \boldsymbol{\Sigma}_{BA}$$

4. Se $\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ e $\mathbf{x}' \sim \mathcal{N}(\boldsymbol{\mu}', \boldsymbol{\Sigma}')$ são independentes, então $\mathbf{x} + \mathbf{x}' \sim \mathcal{N}(\boldsymbol{\mu} + \boldsymbol{\mu}', \boldsymbol{\Sigma} + \boldsymbol{\Sigma}')$.

2.1.3 Independência Condicional

Um conceito primordial para o entendimento de GMRF é o de independência condicional. Naturalmente, duas variáveis aleatórias x_i e x_j são independentes se, e somente se $\pi(x_i, x_j) = \pi(x_i)\pi(x_j)$, que vamos denotar por $x_i \perp x_j$. Analogamente, x_i e x_j são condicionalmente independentes dado $\mathbf{x}_{-\{i,j\}}$ se, e somente se, $\pi(x_i, x_j | \mathbf{x}_{-\{i,j\}}) = \pi(x_i | \mathbf{x}_{-\{i,j\}})\pi(x_j | \mathbf{x}_{-\{i,j\}})$ e denotamos por $x_i \perp x_j | \mathbf{x}_{-\{i,j\}}$.

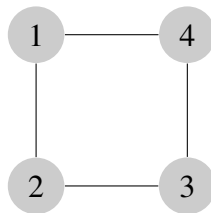
Uma maneira de representar independência condicional é através de grafos não direcionados. Um grafo não direcionado G é um par de conjuntos V e E , isto é, $G = (V, E)$ tal que $V = \{1, 2, \dots, n\}$ é o conjunto formado pelos vértices no grafo e $E = \{\{i, j\} | i, j \in V\}$ é o conjunto formado pelas arestas no grafo, que são pares ordenados $\{i, j\}$ de tal modo que $i, j \in V$ e $i \neq j$. Se o par ordenado $\{i, j\} \in E$, então existe uma aresta não orientada que liga o vértice i ao vértice j . Dizemos que o grafo é completamente conectado se o par ordenado $\{i, j\} \in E$, para todo $i, j \in V$, para $i \neq j$. Note que o conjunto formado pelos vizinhos do vértice i são todos os vértices que pertencem ao grafo G que possuem arestas com i , isto é,

$$ne(i) = \{j \in V | \{i, j\} \in E\}.$$

Escrevemos $i \stackrel{G}{\sim} j$ se os vértices i e j são vizinhos no grafo G , ou apenas, $i \sim j$ quando o grafo G está implícito.

Na Figura 2.1, consideramos um exemplo de representação de independência condicional em grafos de tal forma que $V = \{1, 2, 3, 4\}$ e $E = \{\{1, 2\}, \{2, 3\}, \{3, 4\}, \{4, 1\}\}$.

Figura 2.1 – Representação de independência condicional em um grafo não orientado



Note que através do grafo da Figura 2.1, deduzimos que $x_2 \perp x_4 | \mathbf{x}_{\{1,3\}}$ e $x_1 \perp x_3 | \mathbf{x}_{\{2,4\}}$. Observe também que $ne(1) = \{2, 4\}$, $ne(2) = \{1, 3\}$, $ne(3) = \{2, 4\}$ e $ne(4) = \{1, 3\}$.

Antes de definirmos formalmente um GMRF apresenta-se um resultado que relaciona o grafo G e os parâmetros de uma distribuição normal multivariada. Isto é, queremos especificar um GMRF $\mathbf{x}$ com propriedades de independência condicional de acordo com algum grafo G . Veremos que a propriedade de independência condicional está relacionada à matriz de covariâncias $\mathbf{\Sigma}$, mais especificamente na inversa da matriz de covariâncias, isto é, a matriz de precisão $\mathbf{Q} = \mathbf{\Sigma}^{-1}$.

Teorema 1. *Seja $\mathbf{x} \in \mathbb{R}^n$ um vetor aleatório com distribuição normal multivariada com vetor de médias $\boldsymbol{\mu}$ e matriz de precisão SPD $\mathbf{Q}$, então para $i \neq j$,*

$$x_i \perp x_j | \mathbf{x}_{-ij} \iff Q_{ij} = 0.$$

Prova 1. (RUE; HELD, 2005)

Pelo Teorema 1, sabemos que o padrão de termos não nulos da matriz de precisão $\mathbf{Q}$ determina um associado grafo G , isto é, através de $\mathbf{Q}$ podemos verificar se x_i e x_j são condicionalmente independentes. Por outro lado, o conhecimento de um determinado grafo G é suficiente para determinar os termos não nulos de $\mathbf{Q}$. Segundo Rue e Held (2005), isto pode ser utilizado para fornecer uma parametrização de $\mathbf{Q}$, sabendo que $\mathbf{Q}$ é SPD.

Definição . *Um vetor aleatório $\mathbf{x} = (x_1, \dots, x_n)^\top \in \mathbb{R}^n$ é denominado Campo Aleatório Markoviano Gaussiano (GMRF) associado ao grafo $G = (V, E)$ com vetor de médias $\boldsymbol{\mu}$ e matriz de precisão SPD $\mathbf{Q}$ se, e somente se sua densidade tem forma*

$$\pi(\mathbf{x}) = (2\pi)^{-n/2} |\mathbf{Q}|^{1/2} \exp \left(-\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})^\top \mathbf{Q} (\mathbf{x} - \boldsymbol{\mu}) \right) \quad (2.1)$$

tais que

$$Q_{ij} \neq 0 \iff \{i, j\} \in E, \forall i \neq j.$$

Para o caso em que a matriz de precisão $\mathbf{Q}$ é singular, o GMRF ainda é obtido com forma explícita para sua densidade conjunta, entretanto, a densidade conjunta é imprópria. Tal especificação não pode ser utilizada para modelar os dados, no entanto, pode ser utilizada como informação a priori, caso produza uma posteriori adequada (RUE et al., 2017).

Se a matriz de precisão $\mathbf{Q}$ é completamente densa, então o grafo G é completamente conectado. Isso implica que qualquer distribuição normal multivariada com matriz de covari-

âncias SPD é também um GMRF e vice-versa. Todavia, para que as propriedades do GMRF sejam realmente úteis, a matriz de precisão $\mathbf{Q}$ deve ser esparsa (RUE; HELD, 2005). Vejamos o seguinte resultado que discorre sobre as propriedades dos elementos de $\mathbf{Q}$.

Teorema 2. *Seja $\mathbf{x}$ um GMRF com respectivo grafo $G = (V, E)$ com vetor de médias $\boldsymbol{\mu}$ e matriz de precisão SPD $\mathbf{Q}$, então*

$$E(x_i | \mathbf{x}_{-i}) = \mu_i - \frac{1}{Q_{ii}} \sum_{j: j \sim i} Q_{ij}(x_j - \mu_j), \quad (2.2)$$

$$\text{Prec}(x_i | \mathbf{x}_i) = Q_{ii} \quad e \quad (2.3)$$

$$\text{Corr}(x_i, x_j | \mathbf{x}_{-ij}) = -\frac{Q_{ij}}{\sqrt{Q_{ii}Q_{jj}}}, \quad i \neq j. \quad (2.4)$$

Prova 2. (RUE; HELD, 2005)

Assim, temos que os elementos da diagonal de $\mathbf{Q}$ são as precisões condicionais de x_i , dado $\mathbf{x}_{-i}$, enquanto que os elementos fora da diagonal, em uma escala adequada, fornecem informação sobre a correlação condicional entre x_i e x_j , dado $\mathbf{x}_{-ij}$.

Rue e Held (2005) afirmam que esses resultados devem ser comparados com a interpretação da matriz de covariâncias $\boldsymbol{\Sigma}$, uma vez que a mesma carrega a informação marginal da variância de x_i e da correlação marginal entre x_i e x_j e é intuitiva e diretamente informativa, além de reduzir a interpretação de uma distribuição n -dimensional para de uma distribuição unidimensional ou bidimensional.

Além do mais, interpretar marginalmente a informação trazida por $\mathbf{Q}$ é muito difícil, devemos integrar em $\mathbf{x}_{-i}$ e em $\mathbf{x}_{-ij}$ da distribuição conjunta parametrizada em termos de $\mathbf{Q}$. No entanto, em termos matriciais, isto é imediato, uma vez que por definição $\mathbf{Q} = \boldsymbol{\Sigma}^{-1}$ e $\Sigma_{ii} = \sigma_i^2$ depende de todos os elementos de $\mathbf{Q}$ e vice-versa (RUE; HELD, 2005).

Vimos que, para um determinado grafo G , podemos checar se $x_i \perp x_j | \mathbf{x}_{-ij}$ ou não. Do mesmo modo, pelo Teorema 1, podemos fazer essa mesma verificação utilizando a matriz de precisão $\mathbf{Q}$, mais precisamente, verificando se $Q_{ij} = 0$ ou não. Desse modo, o grafo G é construído por meio dos termos não nulos de $\mathbf{Q}$.

No entanto, propriedades mais gerais de independência condicional de um GMRF podem ser extraídas de um grafo associado G .

Portanto, se $\mathbf{x}$ é um GMRF com respectivo grafo $G = (V, E)$, então, $\mathbf{x}$ obedece as seguintes propriedades de Markov:

1. Pareada:

$$x_i \perp x_j | \mathbf{x}_{-ij}, \quad \text{se } \{i, j\} \notin E \quad e \quad i \neq j$$

2. Local:

$$x_i \perp \mathbf{x}_{\{i, ne(i)\}} | \mathbf{x}_{ne(i)} \quad \forall i \in V$$

3. Global:

$$\mathbf{x}_A \perp \mathbf{x}_B | \mathbf{x}_C$$

Para A , B e C disjuntos e C separa o conjunto A e B , em que A e B são não vazios.

2.2 Modelos Latentes Gaussianos

Segundo Rue, Martino e Chopin (2009), os modelos latentes gaussianos (LGMs) são úteis, pois incorporam diversos modelos estatísticos em apenas uma grande classe de modelos, de modo que o processo de inferência estatística desses modelos pode ser unificado para toda classe.

Os LGMs são obtidos utilizando uma formulação hierárquica de 3 estágios, em que o primeiro estágio é dado pelos valores observados $\mathbf{y}$, usualmente, assumindo independência condicional ao campo aleatório latente gaussiano $\mathbf{x}$ e, possivelmente, hiperparâmetros $\boldsymbol{\theta}_1$, isto é,

$$\pi(\mathbf{y} | \mathbf{x}, \boldsymbol{\theta}_1) = \prod_{i \in I} \pi(y_i | x_i, \boldsymbol{\theta}_1). \quad (2.5)$$

O campo latente $\mathbf{x}$ constitui o segundo estágio do nosso modelo hierárquico e é modelado como um campo aleatório markoviano gaussiano (GMRF) com matriz de precisão $\mathbf{Q}$ que é função de alguns hiperparâmetros $\boldsymbol{\theta}_2$. Portanto,

$$\pi(\mathbf{x} | \boldsymbol{\theta}_2) \propto \exp \left\{ -\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu}(\boldsymbol{\theta}_2))^T \mathbf{Q}(\boldsymbol{\theta}_2) (\mathbf{x} - \boldsymbol{\mu}(\boldsymbol{\theta}_2)) \right\}. \quad (2.6)$$

O terceiro estágio do modelo é determinado pela distribuição a priori dos hiperparâmetros $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$. Os hiperparâmetros $\boldsymbol{\theta}$ controlam o campo latente gaussiano e/ou a verossimilhança dos dados, dessa forma a posteriori é dada por

$$\pi(\mathbf{x}, \boldsymbol{\theta} | \mathbf{y}) \propto \pi(\boldsymbol{\theta}) \pi(\mathbf{x} | \boldsymbol{\theta}) \prod_{i \in I} \pi(y_i | x_i, \boldsymbol{\theta}) \quad (2.7)$$

De acordo com Rue, Martino e Chopin (2009), por razões computacionais, devemos assumir um pequeno número de hiperparâmetros, isto é, $|\boldsymbol{\theta}|$, usualmente, deve estar entre 2 e 5, mas não deve exceder 20; a distribuição de $\mathbf{x}|\boldsymbol{\theta}$ além de gaussiana, deve ser, ao menos aproximadamente, um campo aleatório markoviano gaussiano quando a dimensão n é alta ($10^3, 10^5$); as observações $\mathbf{y}$ são mutuamente condicionalmente independentes de $\mathbf{x}$ e $\boldsymbol{\theta}$, ou seja, cada observação y_i dependerá somente de um componente do campo latente, por exemplo, x_i .

Os modelos latentes gaussianos se relacionam com uma classe de modelos bem ampla, os modelos aditivos, isso é possível, pois podemos interpretar $\{x_i, i \in I\}$ como o preditor η_i , que por si só é aditivo com respeito aos outros efeitos. Desse modo, temos

$$\eta_i = \mu + \sum_j \beta_j z_{ij} + \sum_k f_{k,j_k(i)}, \quad (2.8)$$

em que μ é o intercepto geral, β_j são os efeitos “fixos” das covariáveis z_{ij} e os termos f_k representam específicos processos gaussianos.

Exemplos de modelo das componentes f_k são modelos auto-regressivos de séries temporais, modelos estocásticos *spline*, modelos para suavização (*smoothing models*), modelos para erro de medição, modelos (generalizados) mistos com efeitos aleatórios com diferente tipos de correlação e modelos espaciais (RUE et al., 2017).

Em suma, temos que a estrutura dos modelos latentes gaussianos, como anteriormente dito, pode ser definida com base em um modelo de estrutura hierárquica constituído por uma verossimilhança para os dados no 1º nível de hierarquia, um campo latente gaussiano para o segundo nível de hierarquia e de prioris para os hiperparâmetros no terceiro nível de hierarquia.

A verossimilhança univariada captura a distribuição marginal dos dados; frequentemente o modelo escolhido, mas não necessário que seja, é um membro da família exponencial (Normal, Binomial, Poisson, Gama, Binomial Negativa, etc.) similar ao que ocorre nos modelos lineares generalizados. A média da verossimilhança – ou algum outro parâmetro relacionado às medidas de tendência central – se relaciona com o preditor latente gaussiano através de uma função de ligação tal que $E(\mathbf{y}|\boldsymbol{\eta}) = h^{-1}(\boldsymbol{\eta})$, sendo $\mathbf{y}$ as observações, $\boldsymbol{\eta}$ um preditor gaussiano e h uma função de ligação apropriadamente escolhida (OPITZ, 2017).

Equivalentemente à equação 2.7, a densidade a posteriori conjunta do campo latente $\mathbf{x}$ e hiperparâmetros $\boldsymbol{\theta}$ é dada por

$$\pi(\mathbf{x}, \boldsymbol{\theta} | \mathbf{y}) \propto \exp \left\{ -\frac{1}{2} \mathbf{x}^\top \mathbf{Q}(\boldsymbol{\theta}) \mathbf{x} + \sum_i \log \pi(y_i | \eta_i(\mathbf{x}), \boldsymbol{\theta}) + \log \pi(\boldsymbol{\theta}) \right\}. \quad (2.9)$$

Assumimos, a priori, que as componentes do modelo são independentes. A priori dos efeitos fixos $(\mu, \boldsymbol{\beta})$ é conjuntamente normal e assumida independente das outras componentes do modelo. Sendo assim, a estrutura em (2.8) e os modelos latentes gaussianos associam a mesma classe de modelos quando se assume priori gaussiana para o intercepto e os outros parâmetros dos efeitos fixos. Consequentemente, o campo latente na formulação hierárquica dos modelos latentes gaussianos é $\mathbf{x} = (\boldsymbol{\eta}, \mu, \boldsymbol{\beta}, \mathbf{f}_1, \mathbf{f}_2, \dots)$ e tem distribuição normal multivariada e também não singular, se acrescentarmos um pequeno termo de ruído na equação 2.8. Isso produz um campo latente $\mathbf{x}$ na formulação hierárquica de LGM. A $\dim(\mathbf{x}) = n$ pode facilmente ficar grande, devido ser igual ao número de observações, mais intercepto(s) e efeitos fixos, mais a soma da dimensão de todos as componentes do modelo (RUE et al., 2017).

Os hiperparâmetros $\boldsymbol{\theta}$ englobam os parâmetros da verossimilhança e das componentes do modelo. Geralmente, a verossimilhança e cada componente do modelo possuem entre zero e dois hiperparâmetros cada. Esses parâmetros, normalmente, incluem algum tipo de variância, escala ou parâmetros de correlação. O número de hiperparâmetros é tipicamente pequeno, e ainda, não depende da dimensão do campo latente ou do número de observações. Isto é primordial para a eficiência computacional, mesmo para grandes conjuntos de dados o número de hiperparâmetros permanece constante (RUE et al., 2017).

Na construção de modelos aditivos, incluindo GMRF, segue que $\mathbf{x} = (\boldsymbol{\eta}, \mu, \boldsymbol{\beta}, \mathbf{f}_1, \mathbf{f}_2, \dots)$ é também um GMRF com matriz de precisão, sendo a soma das matrizes de precisão dos efeitos fixos e das outras componentes do modelo. Para tanto, vamos ver um exemplo.

Exemplo 1. (Rue et al. (2017)) Considere o seguinte preditor linear:

$$\eta_i = \mu + \beta z_i + f_{1j_1(i)} + f_{2j_2(i)} + \varepsilon_i, \quad i = 1, \dots, n, \quad (2.10)$$

em que foi adicionado um pequeno ruído ε_i . As componentes $f_{1j_1(i)}$ e $f_{2j_2(i)}$ possuem matrizes de precisão esparsas $\mathbf{Q}_1(\boldsymbol{\theta})$ e $\mathbf{Q}_2(\boldsymbol{\theta})$ de ordem m_1 e m_2 , respectivamente. Seja τ_μ e τ_β a priori da precisão de μ e β . Similarmente, podemos escrever a equação 2.10 na forma matricial

$$\boldsymbol{\eta} = \mu \mathbf{1} + \beta \mathbf{z} + \mathbf{A}_1 \mathbf{f}_1 + \mathbf{A}_2 \mathbf{f}_2 + \boldsymbol{\varepsilon}, \quad (2.11)$$

em que $\mathbf{A}_1$ e $\mathbf{A}_2$ são matrizes esparsas de ordem $n \times m_1$ e $n \times m_2$, respectivamente. Tais matrizes são esparsas, pois possuem apenas um elemento não nulo (o número 1) em cada linha. A matriz de precisão conjunta de $(\boldsymbol{\eta}, \mathbf{f}_1, \mathbf{f}_2, \boldsymbol{\beta}, \mu)$ pode ser obtida reescrevendo

$$\exp \left\{ -\frac{\tau_\epsilon}{2} (\boldsymbol{\eta} - (\mu \mathbf{1} + \boldsymbol{\beta} \mathbf{z} + \mathbf{A}_1 \mathbf{f}_1 + \mathbf{A}_2 \mathbf{f}_2))^\top (\boldsymbol{\eta} - (\mu \mathbf{1} + \boldsymbol{\beta} \mathbf{z} + \mathbf{A}_1 \mathbf{f}_1 + \mathbf{A}_2 \mathbf{f}_2)) \right. \\ \left. - \frac{\tau_\mu}{2} \mu^2 - \frac{\tau_\beta}{2} \boldsymbol{\beta}^2 - \frac{1}{2} \mathbf{f}_1^\top \mathbf{Q}_1(\boldsymbol{\theta}) \mathbf{f}_1 - \frac{1}{2} \mathbf{f}_2^\top \mathbf{Q}_2(\boldsymbol{\theta}) \mathbf{f}_2 \right\}$$

como

$$\exp \left\{ -\frac{1}{2} (\boldsymbol{\eta}, \mathbf{f}_1, \mathbf{f}_2, \boldsymbol{\beta}, \mu)^\top \mathbf{Q}_{conjunta}(\boldsymbol{\theta}) (\boldsymbol{\eta}, \mathbf{f}_1, \mathbf{f}_2, \boldsymbol{\beta}, \mu) \right\} \quad (2.12)$$

em que

$$\mathbf{Q}_{conjunta}(\boldsymbol{\theta}) = \begin{bmatrix} \tau_\epsilon \mathbf{I} & \tau_\epsilon \mathbf{A}_1 & \tau_\epsilon \mathbf{A}_2 & \tau_\epsilon \mathbf{I} \mathbf{z} & \tau_\epsilon \mathbf{I} \mathbf{1} \\ & \mathbf{Q}_1(\boldsymbol{\theta}) + \tau_\epsilon \mathbf{A}_1 \mathbf{A}_1^\top & \tau_\epsilon \mathbf{A}_1 \mathbf{A}_2^\top & \tau_\epsilon \mathbf{A}_1 \mathbf{z} & \tau_\epsilon \mathbf{A}_1 \mathbf{1} \\ & & \mathbf{Q}_2(\boldsymbol{\theta}) + \tau_\epsilon \mathbf{A}_2 \mathbf{A}_2^\top & \tau_\epsilon \mathbf{A}_2 \mathbf{z} & \tau_\epsilon \mathbf{A}_2 \mathbf{1} \\ & \text{simétrica} & & \tau_\beta + \tau_\epsilon \mathbf{z}^\top \mathbf{z} & \tau_\epsilon \mathbf{z}^\top \mathbf{1} \\ & & & & \tau_\mu + \tau_\epsilon \mathbf{1}^\top \mathbf{1} \end{bmatrix}$$

A dimensão da matriz de precisão conjunta é $n + m_1 + m_2 + 2$. A esparsidade de $\mathbf{Q}_{conjunta}(\boldsymbol{\theta})$ ilustra como o modelo facilita a eficiência computacional, uma vez que a dependência de $\boldsymbol{\theta}$ ocorre através das matrizes $\mathbf{Q}_1(\boldsymbol{\theta})$ e $\mathbf{Q}_2(\boldsymbol{\theta})$, e como $\boldsymbol{\theta}$ se altera, o custo computacional para recalcular $\mathbf{Q}_{conjunta}(\boldsymbol{\theta})$ é desprezível. Por simplicidade, suponha que $n = m_1 = m_2$, e denote por e_1 e e_2 o número médio de elementos não nulos nas linhas de $\mathbf{Q}_1(\boldsymbol{\theta})$ e $\mathbf{Q}_2(\boldsymbol{\theta})$, respectivamente. Assim, um limite superior para o número de elementos não nulos em $\mathbf{Q}_{conjunta}(\boldsymbol{\theta})$ é $n(19 + e_1 + e_2) + 4$. Em média, temos, aproximadamente $(19 + e_1 + e_2)/3$ termos não nulos para cada linha de $\mathbf{Q}_{conjunta}(\boldsymbol{\theta})$, que a caracteriza esparsa.

2.2.1 Integrated Nested Laplace Approximation (INLA)

Proposta por Rue, Martino e Chopin (2009), *Integrated Nested Laplace Approximation* (INLA) é uma abordagem computacional para fazer inferência bayesiana em modelos da forma

LGMs. A metodologia utiliza aninhadas aproximações de Laplace, uma aproximação determinística de integrais utilizada no escopo da abordagem INLA para aproximar as densidades marginais a posteriori do campo latente e dos hiperparâmetros.

A posteriori conjunta descrita em (2.9) é uma densidade de alta dimensão de difícil interpretação. No entanto, frequentemente, o principal interesse está nas marginais a posteriori do campo latente $\pi(x_i|\mathbf{y})$ e/ou nas marginais a posteriori dos hiperparâmetros $\pi(\theta_k|\mathbf{y})$. A abordagem INLA fornece aproximações para tais densidades marginais a posteriori, que podem ser utilizadas para calcular, via aproximações determinísticas, sumários estatísticos como médias, variâncias, quantis, entre outros (RUE et al., 2017).

A Aproximação de Laplace é o núcleo da metodologia INLA, sucintamente, pode ser obtida quando se tem o interesse em aproximar a integral de uma função da forma:

$$I_n = \int \exp(nf(x))dx, \quad \text{quando } n \rightarrow \infty.$$

Se definirmos x_0 um ponto em que $f(x)$ possui um máximo e aplicarmos uma aproximação de Taylor de 2º ordem, temos que:

$$\begin{aligned} I_n &\approx \int_x \exp \left(n \left(f(x_0) + \frac{1}{2}(x-x_0)^2 f''(x_0) \right) \right) dx \\ &= \exp(nf(x_0)) \sqrt{\frac{2\pi}{-nf''(x_0)}} \\ &= \tilde{I}_n. \end{aligned}$$

A ideia é aproximar a função alvo com uma distribuição normal, fazendo com que coincidam a moda e a curvatura na moda. Rue et al. (2017) afirma que se interpretar $nf(x)$ como a soma de log-verossimilhanças e x como um parâmetro desconhecido, pelo Teorema Central do Limite, a aproximação Gaussiana é exata quando $n \rightarrow \infty$.

Nosso interesse é aproximar as posteriores marginais $\pi(x_i|\mathbf{y})$ e $\pi(\theta_k|\mathbf{y})$. A metodologia INLA tem por objetivo construir aproximações aninhadas da forma

$$\tilde{\pi}(x_i|\mathbf{y}) = \int \tilde{\pi}(x_i|\boldsymbol{\theta}, \mathbf{y}) \tilde{\pi}(\boldsymbol{\theta}|\mathbf{y}) d\boldsymbol{\theta} \quad (2.13)$$

$$\tilde{\pi}(\theta_k|\mathbf{y}) = \int \tilde{\pi}(\boldsymbol{\theta}|\mathbf{y}) d\boldsymbol{\theta}_{-k}. \quad (2.14)$$

Dessa forma, deve-se calcular as posteriores $\pi(\boldsymbol{\theta}|\mathbf{y})$, o que implica que as marginais $\pi(\theta_k|\mathbf{y})$ podem ser obtidas e $\pi(x_i|\boldsymbol{\theta}, \mathbf{y})$, necessária para a obtenção de $\pi(x_i|\mathbf{y})$.

A aproximação de Laplace para $\tilde{\pi}(\boldsymbol{\theta}|\mathbf{y})$ é

$$\tilde{\pi}(\boldsymbol{\theta}|\mathbf{y}) \propto \frac{\pi(\mathbf{y}|\mathbf{x}, \boldsymbol{\theta})\pi(\mathbf{x}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})}{\tilde{\pi}_G(\mathbf{x}|\boldsymbol{\theta}, \mathbf{y})} \Big|_{\mathbf{x}=\boldsymbol{\mu}^*(\boldsymbol{\theta})}, \quad (2.15)$$

em que $\tilde{\pi}_G(\mathbf{x}|\boldsymbol{\theta}, \mathbf{y})$ é a aproximação gaussiana da condicional completa de $\mathbf{x}|\boldsymbol{\theta}, \mathbf{y}$, dada pela Aproximação de Laplace, coincidindo com a moda e a curvatura na moda, e $\boldsymbol{\mu}^*(\boldsymbol{\theta})$ é a moda para um dado $\boldsymbol{\theta}$.

A aproximação para $\tilde{\pi}(x_i|\boldsymbol{\theta}, \mathbf{y})$ é realizada através da expressão

$$\tilde{\pi}(x_i|\boldsymbol{\theta}, \mathbf{y}) \propto \frac{\pi(\mathbf{y}|\mathbf{x}, \boldsymbol{\theta})\pi(\mathbf{x}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})}{\tilde{\pi}(\mathbf{x}_{-i}|\mathbf{x}_i, \boldsymbol{\theta}, \mathbf{y})}. \quad (2.16)$$

Em que a estratégia de aproximação realizada no denominador de (2.16) pode ser de três tipos, denominadas de gaussiana, laplace simplificada e laplace, sendo a laplace simplificada, a aproximação mais comumente utilizada, por apresentar melhor custo-benefício com respeito a custo computacional e precisão da aproximação.

Dados $\tilde{\pi}(\boldsymbol{\theta}|\mathbf{y})$ e alguma aproximação de $\tilde{\pi}(x_i|\boldsymbol{\theta}, \mathbf{y})$ (Gaussiana, Laplace Simplificada ou Laplace Completa) a integral da equação (2.13) pode ser obtida numericamente

$$\tilde{\pi}(x_i|\mathbf{y}) = \sum_k \tilde{\pi}(x_i|\boldsymbol{\theta}^{(k)}, \mathbf{y}) \tilde{\pi}(\boldsymbol{\theta}^{(k)}|\mathbf{y}) \Delta\boldsymbol{\theta}^{(k)}, \quad (2.17)$$

em que $\{\tilde{\pi}(\boldsymbol{\theta}^{(k)}|\mathbf{y})\}$ são valores da densidade calculados durante exploração da grade, planejamento composto central ou *empirical bayes* em $\tilde{\pi}(\boldsymbol{\theta}|\mathbf{y})$.

Um detalhamento da metodologia INLA pode ser encontrada em [Rue, Martino e Chopin \(2009\)](#) e [Rue et al. \(2017\)](#).

2.3 Modelos Espaciais Latentes Gaussianos

Diversos modelos podem ser reescritos na forma geral de campos gaussianos latentes. Dentre estes, destacamos os principais que serão utilizados isoladamente ou em conjunto para construir modelos neste trabalho.

Dados espaciais são definidos como realizações de um processo estocástico indexado pelo espaço

$$Y(\mathbf{s}) \equiv \{y(\mathbf{s}), \quad \mathbf{s} \in D\},$$

em que D é um subconjunto (fixo) d -dimensional dos números reais, isto é, $D \subset \mathbb{R}^d$.

De acordo com Cressie (1993), dados espaciais, resumidamente, se enquadram em 3 classes:

1. **Dados de área ou látice:** neste caso, $y(\mathbf{s})$ é um valor agregado aleatório sob cada unidade de área $\mathbf{s}$ com limites/fronteiras/bordas bem definidas em D , que por sua vez é definido como um conjunto enumerável de unidades espaciais d -dimensionais. A diferença entre dados de área e látice se resume no fato de que látice possui cada unidade de área $\mathbf{s}$ caracterizada por ter bordas que formam uma grade regular, enquanto que os de área se caracterizam por terem esses limites irregulares, normalmente são bordas que se referem a fronteiras administrativas, tais como: bairros, cidades, mesorregiões, países e etc.

Para o tratamento desses dados, frequentemente, tem-se o interesse em obter uma suavização ou um mapeamento da resposta y em função das unidades de área $\mathbf{s}$ no conjunto D .

2. **Dados de superfície contínua (georreferenciados ou geoestatísticos):** neste caso $y(\mathbf{s})$ é uma resposta aleatória em uma localização espacial específica indexada por $\mathbf{s}$ que varia continuamente no domínio fixo D . De maneira geral, a localização $\mathbf{s}$ é um vetor bidimensional contendo as informações de latitude e longitude. Em algumas aplicações, a localização $\mathbf{s}$ é um vetor tridimensional, incluindo a altitude.

Um conjunto típico de dados de superfície contínua é representado por $\mathbf{y} = \{y(\mathbf{s}_1), \dots, y(\mathbf{s}_n)\}$, em que o conjunto $\{\mathbf{s}_1, \dots, \mathbf{s}_n\}$ indicam as localidades $\mathbf{s}_i$, $i = 1, \dots, n$ em que as medidas da resposta $y(\mathbf{s}_i)$ foram tomadas. Note, portanto, que as localidades em que as medidas (que são aleatórias) serão tomadas, são previamente estabelecidas, isto porque o objetivo principal para análise desse tipo de dados é prever a resposta y em uma localização não observada em D . Dessa forma, pode-se, por exemplo, “reconstruir” toda superfície definida em D das respostas preditas y^* .

3. **Padrões pontuais espaciais:** nesta situação, $y(\mathbf{s})$ normalmente representa a ocorrência de um evento na localização $\mathbf{s}$. Dessa maneira, a localização $\mathbf{s} \in \mathbb{R}^d$ da ocorrência do evento é aleatória. O domínio espacial D é um conjunto de pontos em $\mathbb{R}^d$ em que algum evento ocorreu. Neste caso, $y(\mathbf{s})$ assume valores 0 ou 1 que representam a não ocorrência ou ocorrência do evento em estudo, respectivamente. Para as situações em que há a possibilidade de se obter informações adicionais do fenômeno em estudo através de uma covariável, descrevemos este tipo de análise como *processos pontuais marcados*. Em processos pontuais, costumeiramente, o interesse é avaliar, no espaço, a existência de possíveis comportamentos de agrupamentos (ou repelimentos) entre as observações.

Quando nos deparamos com dados espaciais, sua tendência espacial precisa ser considerada no processo de inferência, uma vez que a informação contida nessa tendência que pode vir a gerar algum conhecimento. Desconsiderar tal informação possivelmente nos levará a cometer viés nas estimativas dos parâmetros.

Se os dados estão disponíveis a nível de área e consistem em contagens agregadas da variável resposta e/ou das covariáveis, geralmente, podem ser especificados modelos de mapeamento de doenças e de regressão ecológica (WIKLE, 2003; CLARK; GELFAND, 2006; LAWSON, 2018).

Sob tais circunstâncias, Dunson (2001) afirma que a abordagem bayesiana é mais efetiva para o tratamento de dados espaciais. De acordo com Gerber, Furrer et al. (2015) mapas de dados espaciais de contagem agregadas são difíceis de serem interpretados e para corrigir este problema, modelos hierárquicos bayesianos são, frequentemente utilizados, com intuito de identificar padrões suaves explicados por meio de covariáveis e fatores espaciais.

A seguir, apresentaremos os principais modelos espaciais de dados de área.

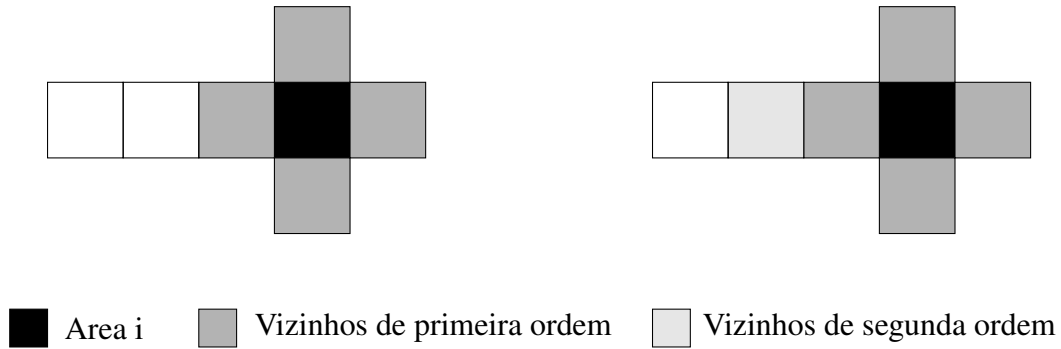
2.3.1 Modelos espaciais para dados de área

Quando trabalhamos com dados de área, geralmente, assumimos que áreas mais próximas no espaço são mais similares com relação às características avaliadas do que em áreas não tão próximas. Neste contexto, “áreas próximas” significam que há a necessidade de se construir uma estrutura de vizinhança. Podemos simplificar a notação para as áreas $\mathbf{s} = \{s_1, \dots, s_n\}$ como $\mathbf{s} = \{1, \dots, n\}$. Para definir a estrutura de vizinhança, é de costume supor que áreas são classificadas como vizinhas se as mesmas compartilham bordas/limites. Em notação matemática, para representar que a área i é vizinha da área j com $i = 1, \dots, n$; $j = 1, \dots, n$; e $i \neq j$, utilizamos

a notação $i \sim j$. Também definimos o conjunto dos vizinhos da área i como δ_i e o número de vizinhos n_{δ_i} como cardinalidade do conjunto δ_i , isto é, $n_{\delta_i} = \#\delta_i$.

De maneira mais detalhada, Blangiardo e Cameletti (2015) trazem o conceito de vizinhança de primeira e de segunda ordem. Na de primeira ordem, são consideradas áreas vizinhas de uma área i específica, apenas as áreas que compartilham de mesma borda. Já na vizinhança de segunda ordem, além dos vizinhos de primeira ordem da área i , são também considerados vizinhos da área i os vizinhos (de 1º ordem) dos vizinhos de primeira ordem da área i .

Figura 2.2 – Estrutura de vizinhança: vizinhos de primeira ordem (à esquerda) e vizinhos de primeira e segunda ordem (à direita)



Fonte: elaborado pelo próprio autor. Adaptado em (BLANGIARDO; CAMELETTI, 2015)

2.3.2 Modelo Besag

Um campo aleatório gaussiano markoviano intrínseco, referido como Modelo de Besag (BESAG; YORK; MOLLIÉ, 1991), também conhecido como um modelo auto-regressivo condicional intrínseco (iCAR) é uma das abordagens mais populares para a modelagem de correlação espacial e modela valores u_i numa coleção de i regiões, condicionados às regiões vizinhas, $i = 1, 2, \dots, n$. Aqui, vamos definir duas regiões vizinhas quando estas compartilham de uma mesma fronteira, ou seja, vizinhança de primeira ordem. A distribuição condicional para u_i é dada por:

$$u_i | \mathbf{u}_{-i}, \tau_u \sim \mathcal{N} \left(\frac{1}{n_{\delta_i}} \sum_{j \in \delta_i} u_j, \frac{1}{n_{\delta_i} \tau_u} \right)$$

em que τ_u é o parâmetro de precisão, n_{δ_i} é o número de vizinhos da região u_i e $\mathbf{u}_{-i} = (u_1, u_2, \dots, u_{i-1}, u_{i+1}, \dots, u_n)^\top$.

A média de u_i é igual a média dos efeitos de seus vizinhos e sua precisão é proporcional ao seu número de vizinhos. A distribuição conjunta de $\mathbf{u}$ é dada por

$$\pi(\mathbf{u}|\tau_u) \propto \exp\left(-\frac{\tau_u}{2} \sum_{i \sim j} (u_i - u_j)^2\right) \propto \exp\left(-\frac{\tau_u}{2} \mathbf{u}^\top \mathbf{Q} \mathbf{u}\right), \quad (2.18)$$

em que a matriz de precisão $\mathbf{Q}$ possui entradas dada por:

$$Q_{ij} = \begin{cases} n_{\delta_i}, & i = j \\ -1, & i \sim j \\ 0, & \text{caso contrário} \end{cases}$$

Intuitivamente, nota-se que a precisão τ_u controla a similaridade de u_i com suas áreas vizinhas, no sentido de que uma alta precisão τ_u nos remete a similaridades entre os vizinhos, enquanto que uma baixa precisão τ_u permite dissimilaridades entre os vizinhos.

Como apresentado por Rue e Held (2005), o modelo é intrínseco devido à singularidade da matriz de precisão $\mathbf{Q}$, ou seja, possui um espaço nulo não vazio $\mathbf{V}$. Adicionalmente, tem-se que para qualquer que seja a estrutura de vizinhança, o vetor $\mathbf{1}_n$ é um autovetor de $\mathbf{Q}$ associado ao autovalor 0, portanto, $\mathbf{Q}\mathbf{1}_n = 0$ e a densidade de (2.18) é *flat* (imprópria) ao longo de $\mathbf{u} = \alpha \mathbf{1}_n$, $\alpha \in \mathbb{R}$. Dessa maneira, a densidade é invariante à adição de uma constante (RIEBLER et al., 2016). Se a estrutura espacial possui ilhas, então a definição de espaço nulo é mais complexa e depende de como as ilhas são definidas, por exemplo, se elas formam um novo grafo desconectado ou estão ligadas à região principal. A deficiência do posto em um modelo Besag é igual ao número de subgrafos conectados. A deficiência do posto é igual a 1, se toda região é um grafo conectado (para mais detalhes ver Hodges (2013)). Assim, a densidade conjunta para o Modelo Besag é dada por:

$$\pi(\mathbf{u}|\tau_u) = K \tau_u^{(n-I)} \exp\left(-\frac{\tau_u}{2} \mathbf{u}^\top \mathbf{Q} \mathbf{u}\right)$$

em que I denota o número de subgrafos conectados e K é uma constante (HODGES, 2013). Para evitar o confundimento com o intercepto, restrições de que os efeitos somem zero são necessárias em cada subgrafo conectado.

2.3.3 Modelo BYM

Bakka et al. (2018) recomenda combinar o modelo de Besag, acrescentando um efeito aleatório não estruturado espacial, uma vez que o Modelo de Besag somente assume uma componente estruturada espacial. Portanto, esta formulação de modelo nos leva a modelagem de possíveis erros aleatórios não estruturados, ou, por exemplo, superdispersão pura dentro de uma área i por meio de uma correlação espacial, ocasionando estimativas de parâmetros equivocadas (BRESLOW; LEROUX; PLATT, 1998).

Denotando por $\mathbf{b}$ o efeito espacial regional, o modelo que decompõe $\mathbf{b}$ como a soma de duas componentes espaciais, sendo uma não estruturada e outra estruturada, tal que $\mathbf{b} = \mathbf{v} + \mathbf{u}$ é conhecido como Modelo BYM e foi proposto por Besag, York e Mollié (1991) para corrigir possíveis estimativas enganosas dos parâmetros ao se considerar apenas o Modelo de Besag.

Aqui, a componente espacial não estruturada $\mathbf{v} \sim \mathcal{N}(\mathbf{0}, \tau_v^{-1} \mathbf{I})$ e descreve a superdispersão pura, enquanto que a componente espacial estruturada $\mathbf{u} \sim \mathcal{N}(\mathbf{0}, \tau_u^{-1} \mathbf{Q}^-)$ é o modelo de Besag, em que $\mathbf{Q}^-$ é a inversa generalizada da matriz de precisão $\mathbf{Q}$. A matriz de covariâncias de $\mathbf{b}$ é dada por:

$$Var(\mathbf{b} | \tau_u, \tau_v) = \tau_v^{-1} \mathbf{I} + \tau_u^{-1} \mathbf{Q}^-$$

2.3.4 Modelo de Leroux

No modelo BYM, as componentes estruturada e não estruturada não podem ser analisadas independentemente uma da outra, sendo, portanto, não identificáveis (MACNAB, 2011). Bernardinelli, Clayton e Montomoli (1995), bem como Wakefield (2007) descrevem que a escolha das hiperprioris no Modelo BYM é um desafio adicional, uma vez que essa dificuldade se dá pelo fato de que τ_v e τ_u não representam variabilidade na mesma escala, uma vez que τ_v^{-1} representa a variância marginal do efeito aleatório não estruturado, enquanto que τ_u^{-1} representa a variabilidade da componente estruturada u_i , condicionada aos efeitos das áreas vizinhas.

Isto posto, Leroux, Lei e Breslow (2000) propuseram um novo modelo que é baseado em uma diferente especificação da inversa generalizada da matriz de covariâncias, de maneira que torna mais explícita a relação entre a variabilidade da dependência espacial estruturada e não estruturada.

Nesta nova especificação, é assumido que $\mathbf{b}$ segue distribuição normal multivariada com vetor de médias $\mathbf{0}$ e matriz de covariâncias dada por:

$$\text{Var}(\mathbf{b}|\tau_b, \phi) = \tau_b^{-1}((1 - \phi)\mathbf{I} + \phi\mathbf{Q})^{-1}, \quad (2.19)$$

em que $\phi \in [0, 1]$ denota um parâmetro de mistura.

Se observarmos a matriz de covariâncias do Modelo de Leroux apresentada em (2.19), percebemos que o modelo se reduz ao modelo de superdispersão pura se $\phi = 0$, ou se reduz ao modelo de Besag se $\phi = 1$. Portanto, a esperança condicional de b_i , dado todos os outros efeitos aleatórios resulta da média ponderada do modelo não estruturado com média zero e o valor médio do Modelo de Besag, enquanto que a variância condicional é a média ponderada de τ_b^{-1} e τ_b^{-1}/n_{δ_i} .

2.3.5 Modelo de Dean

O modelo proposto por Dean, Ugarte e Militino (2001) consiste em uma reparametrização do Modelo BYM descrito na seção 2.3.3 de tal modo que,

$$\mathbf{b} = \frac{1}{\sqrt{\tau_b}} \left(\sqrt{1 - \phi} \mathbf{v} + \sqrt{\phi} \mathbf{u} \right)$$

com matriz de covariâncias dada por:

$$\text{Var}(\mathbf{b}|\tau_b, \phi) = \tau_b^{-1} ((1 - \phi)\mathbf{I} + \phi\mathbf{Q}^-).$$

A reparametrização é obtida fazendo $\tau_u^{-1} = \tau_b^{-1}\phi$ e $\tau_v^{-1} = \tau_b^{-1}(1 - \phi)$.

Riebler et al. (2016) declara que no Modelo de Dean a decomposição aditiva da variância está na escala do logaritmo do risco relativo, enquanto que no Modelo de Leroux a decomposição aditiva da variância está na escala do logaritmo do risco relativo, condicionado em b_j , $j \in \delta_i$.

2.3.6 Modelo BYM2

Uma questão inerente a todos os modelos descritos até então (seção 2.3.1) é que a componente estruturada espacialmente não é escalada. Sørbye e Rue (2014) declaram que há a necessidade de escalar a componente espacial estruturada para facilitar a definição da hiperpriori e garantir que a hiperpriori definida tenha a mesma interpretação quando utilizada em uma outra aplicação.

O modelo de Besag penaliza desvios locais do seu espaço nulo, que é um nível constante para o caso de um componente conectado (RUE; HELD, 2005). A hiperpriori irá controlar este desvio local e, assim sendo, influencia na suavidade dos efeitos espaciais estimados. Se o campo estimado é muito suave, a precisão é grande e a potencial existência da variação espacial pode ser mascarada. Por outro lado, se a precisão é muito pequena o modelo pode sofrer um superajuste devido à grande variabilidade local (SØRBYE; RUE, 2014).

Sørbye e Rue (2014) mostraram que a variância generalizada para duas estruturas de grafos diferentes não são iguais, mesmo quando estes tem o mesmo número de regiões. Os autores mostraram que essa diferença pode ser impactante a medida que o número de regiões cresce, condicionado a uma mesma precisão τ_b .

Neste sentido, a fim de unificar a interpretação de uma priori definida para τ_b e fazer esta ser transferível para outras aplicações, Sørbye e Rue (2014) propuseram que os efeitos estruturados devam ser escalados de tal modo que a variância generalizada – computada como a média geométrica – de um efeito aleatório estruturado $\mathbf{b}$ seja igual ao inverso de sua precisão, isto é, $\sigma_{VG}^2(\mathbf{b}) = 1/\tau_b$. Isso implica que τ_b representa a precisão do desvio (marginal) de um nível constante, independente do grafo subjacente.

Dessa forma, Simpson et al. (2017) propuseram uma modificação do Modelo de Dean que aborda tanto a identificabilidade, quanto a questão da escala do Modelo BYM. O modelo utiliza uma componente estruturada escalada $\mathbf{u}_*$, em que $\mathbf{Q}_*$ é a matriz de precisão do modelo de Besag, escalada de tal modo que $\sigma_{VG}^2(\mathbf{b}) = 1/\tau_b$. Isso fornece uma versão modificada do efeito aleatório, tal que:

$$\mathbf{b} = \frac{1}{\sqrt{\tau_b}} \left(\sqrt{1-\phi} \mathbf{v} + \sqrt{\phi} \mathbf{u}_* \right),$$

com matriz de covariâncias

$$\text{Var}(\mathbf{b} | \tau_b, \phi) = \tau_b^{-1} ((1-\phi) \mathbf{I} + \phi \mathbf{Q}_*^{-1}).$$

Similarmente ao modelo de Leroux e Dean, o modelo BYM2 enfatiza a flexibilidade entre superdispersão pura e correlação estruturada espacialmente, através do parâmetro de mistura ϕ , tal que $0 \leq \phi \leq 1$ e mensura a proporção da variância marginal explicada pela componente estruturada espacial. O modelo se reduz à superdispersão pura quando $\phi = 0$ e se reduz ao modelo de Besag se $\phi = 1$.

De acordo com Riebler et al. (2016), sem escala adequada, os hiperparâmetros não possuem significado claro e podem ser falsamente interpretados. Por exemplo, nos modelos de Leroux e Dean, o parâmetro de precisão tende a ser falsamente interpretado como a precisão marginal controlando o desvio de um nível de risco constante ao longo do espaço. No entanto, o parâmetro depende da estrutura de vizinhança do grafo subjacente e está confundida com o parâmetro de mistura se a componente estruturada espacial não está adequadamente escalada.

Além disso, Riebler et al. (2016) afirmam que ainda não está claro como se deve escolher uma priori para o parâmetro de precisão, pois, devido a falta de escala, uma hiperpriori fixa para a precisão e as hiperprioris comumente utilizadas induzem a um super ajuste do modelo e não permite que o modelo se reduza a um modelo com erros não correlacionados ao longo do espaço.

No modelo BYM2 o parâmetro de precisão representa a precisão marginal e controla a variabilidade explicada pelo efeito espacial, enquanto que o parâmetro de mistura distribui a variabilidade existente entre componente estruturada e não estruturada.

Uma vez que no modelo BYM2 a componente espacial é escalada, a tarefa da definição da priori se torna mais simples. Riebler et al. (2016) sugerem as prioris PC, utilizadas para favorecer modelos mais simples até que um modelo mais complexo seja suportado.

Breslow, Leroux e Platt (1998) criticaram o modelo BYM no sentido de que ao ser acrescido ao modelo de Besag uma componente não estruturada as interpretações da média condicional e variância condicional tornaram-se confusas. De acordo com Riebler et al. (2016), ao utilizar essa nova parametrização do modelo BYM2, $Var(\mathbf{b}|\tau_b, \phi)$ possui interpretação clara e intuitiva e a interpretação em termos da distribuição condicional de b_i dado b_j para $j \in \delta_i$ são evitadas.

Riebler et al. (2016) declaram que como o efeito estruturado $\mathbf{u}_*$ é escalado, o efeito aleatório $\mathbf{b}$ também é escalado, assim, a priori definida para τ_b possui a mesma interpretação, sendo transferível entre grafos. Além disso, os hiperparâmetros τ_b e ϕ não estão mais confundidos e, agora, são interpretáveis.

Por fim, diferentemente do modelo de Dean, o modelo de Leroux não pode ser escalado, pois, por construção a escala depende do valor de ϕ .

2.4 Modelos temporais

Uma série temporal é qualquer conjunto de dados ordenados no tempo. Nesse sentido, a análise de séries temporais se refere a análise de dados de uma série temporal. O tempo pode ser indexado tanto no domínio discreto quanto no domínio contínuo. Para séries temporais definidas no domínio discreto, apresentaremos os modelos auto-regressivos, enquanto que para as séries definidas no domínio contínuo abordaremos métodos de suavização tais como os modelos de passeio aleatório (*random walk*).

2.4.1 Modelo de efeitos aleatórios autorregressivos (AR)

Modelos autorregressivos são uma maneira conveniente de modelar os dados quando o tempo é indexado sob um domínio discreto. Um caso particular de um GMRF é um processo auto-regressivo de ordem $p = 1$, denotado por AR(1) e com erros tendo distribuição normal com média zero e precisão τ . O modelo é escrito da seguinte forma:

$$x_t = \rho x_{t-1} + \varepsilon_t, \quad \varepsilon_t \stackrel{iid}{\sim} \mathcal{N}(0, \tau_x^{-1}), \quad |\rho| < 1, \quad (2.20)$$

em que o índice t representa o tempo. A suposição de independência condicional não está explicitamente declarada em (2.20), no entanto, se reescrevermos o modelo da seguinte maneira

$$x_t | x_1, \dots, x_{t-1} \sim \mathcal{N}(\rho x_{t-1}, \tau_x^{-1}), \quad t = 2, \dots, n \quad (2.21)$$

perceberemos que o modelo possui independência condicional representada da seguinte forma: dado x_s e x_t com $1 \leq s < t \leq n$, então $x_s \perp x_t | x_{s+1}, \dots, x_{t-1}$ se $t - s > 1$. Além de (2.21), assumimos que a distribuição marginal de x_1 é normal com média zero e variância $(\tau(1 - \rho^2))^{-1}$, que é simplesmente a distribuição estacionária deste processo. Consequentemente, a distribuição conjunta de $\mathbf{x}$ é dada por:

$$\begin{aligned} \pi(\mathbf{x}) &= \pi(x_1) \pi(x_2 | x_1) \cdots \pi(x_n | x_{n-1}) \\ &= \frac{1}{(2\pi)^{n/2}} |\mathbf{Q}|^{1/2} \exp \left\{ -\frac{1}{2} \mathbf{x}^\top \mathbf{Q} \mathbf{x} \right\}, \end{aligned} \quad (2.22)$$

em que a matriz de precisão $\mathbf{Q}$ é uma matriz tridiagonal dada por:

$$\mathbf{Q} = \tau_x \begin{pmatrix} 1 & -\rho & & & \\ -\rho & 1 + \rho^2 & -\rho & & \\ & & \ddots & \ddots & \ddots \\ & & & -\rho & 1 + \rho^2 & -\rho \\ & & & & -\rho & 1 \end{pmatrix} \quad (2.23)$$

em que os elementos fora de sua diagonal principal e das diagonais imediatamente acima e abaixo da principal são nulos.

Simplificações devido a independência condicional não aparecem apenas nas distribuições condicionais diretas (2.21), mas também nas distribuições condicionais completas (indiretas) $\pi(x_t | \mathbf{x}_{-t})$, apresentadas a seguir:

$$x_t | \mathbf{x}_{-t} \sim \begin{cases} \mathcal{N}(\rho x_{t+1}, \tau^{-1}) & t = 1, \\ \mathcal{N}\left(\frac{\rho}{1 + \rho^2}(x_{t-1} + x_{t+1}), \frac{\tau_x^{-1} \rho}{1 + \rho^2}\right) & 1 < t < n, \\ \mathcal{N}(\rho x_{n-1}, \tau^{-1}) & t = n. \end{cases}$$

A especificação do modelo AR(1) por meio das condicionais completas é uma maneira de especificação alternativa e equivalente da representação da densidade conjunta (RUE; HELD, 2005).

De maneira geral, um modelo auto-regressivo de ordem p (AR(p)) para um vetor normal $\mathbf{x} = (x_1, \dots, x_n)$ é definido similarmente ao AR(1), mas agora com *lag* de ordem p . Assim,

$$x_t = \rho_1 x_{t-1} + \rho_2 x_{t-2} + \dots + \rho_p x_{t-p} + \varepsilon_t, \quad t = p, \dots, n.$$

O modelo AR(p) possui uma parametrização mais complexa pois existem várias restrições não lineares nos parâmetros ρ_t para definir um modelo não estacionário. Uma solução, consiste em reparametrizar o modelo através da função de autocorrelação parcial, cujo domínio é o intervalo aberto $(-1, 1)$. Para mais detalhes sobre essa reparametrização ver Rue e Held (2005) e Brockwell e Davis (2009).

2.4.2 Passeio Aleatório (RW)

Um modelo passeio aleatório de ordem 1 (RW1) para um vetor normal $\mathbf{x} = (x_1, \dots, x_n)^\top$ é construído assumindo incrementos independentes, isto é,

$$\Delta x_i = x_{i+1} - x_i \sim \mathcal{N}(0, \tau^{-1}), \quad i = 1, \dots, n.$$

A densidade de $\mathbf{x}$ derivada dos $n - 1$ incrementos é

$$\begin{aligned} \pi(\mathbf{x}|\tau) &\propto \tau^{(n-1)/2} \exp \left\{ -\frac{\tau}{2} \sum_{i=1}^{n-1} (\Delta x_i)^2 \right\} \\ &= \tau^{(n-1)/2} \exp \left\{ -\frac{\tau}{2} \sum_{i=1}^{n-1} (x_{i+1} - x_i)^2 \right\} \\ &= \tau^{(n-1)/2} \exp \left\{ -\frac{\tau}{2} \mathbf{x}^\top \mathbf{Q} \mathbf{x} \right\} \end{aligned}$$

em que $\mathbf{Q} = \tau \mathbf{R}$ e $\mathbf{R}$ é uma matriz estruturada de acordo com a estrutura de vizinhança do modelo da forma:

$$\mathbf{R} = \begin{bmatrix} 1 & -1 & & & \\ -1 & 2 & -1 & & \\ & -1 & 2 & -1 & \\ & & \ddots & \ddots & \ddots \\ & & & -1 & 2 & -1 \\ & & & & -1 & 2 & -1 \\ & & & & & -1 & 1 \end{bmatrix} \quad (2.24)$$

Um modelo de passeio aleatório de ordem 2 (RW2) é construído sob incrementos de segunda ordem, dado por:

$$\Delta^2 x_i = \Delta(\Delta x_i) = x_i - 2x_{i+1} + x_{i+2} \sim \mathcal{N}(0, \tau^{-1}), \quad i = 1, \dots, n-2. \quad (2.25)$$

A densidade conjunta de $\mathbf{x}$ é dada por

$$\pi(\mathbf{x}) \propto \tau_x^{(n-2)/2} \exp \left(-\frac{\tau_x}{2} \sum_{i=1}^{n-2} (x_i - 2x_{i+1} + x_{i+2})^2 \right) \quad (2.26)$$

$$= \tau_x^{(n-2)/2} \exp \left(-\frac{1}{2} \mathbf{x}^\top \mathbf{Q} \mathbf{x} \right), \quad (2.27)$$

em que a matriz de precisão $\mathbf{Q}$ é dada por:

$$\mathbf{Q} = \tau_x \begin{pmatrix} 1 & -2 & 1 & & & & & & \\ -2 & 5 & -4 & 1 & & & & & \\ 1 & -4 & 6 & -4 & 1 & & & & \\ & 1 & -4 & 6 & -4 & 1 & & & \\ & & \ddots & \ddots & \ddots & \ddots & \ddots & & \\ & & & 1 & -4 & 6 & -4 & 1 & \\ & & & & 1 & -4 & 6 & -4 & 1 \\ & & & & & 1 & -4 & 5 & -2 \\ & & & & & & 1 & -2 & 1 \end{pmatrix}$$

2.5 Modelos espaços-temporais

Em formulações de modelos cujo o preditor possui a presença de efeitos principais tanto de espaço, quanto de tempo, [Knorr-Held \(2000\)](#) sugere acrescentar interações espaço-tempo. O modelo do preditor linear proposto pelo autor é dado por:

$$\eta_{it} = \mu + \alpha_t + \gamma_t + \theta_i + \phi_i + \delta_{it}, \quad i = 1, \dots, I \text{ e } t = 1, \dots, T. \quad (2.28)$$

em que,

- μ é o intercepto geral, geralmente, assumindo uma priori não informativa.
- α_t é um efeito temporalmente estruturado a priori, podendo, por exemplo, ser modelado através de um RW1.
- γ_t é um efeito temporalmente não estruturado, tal que $\gamma_t \sim \mathcal{N}(0, \tau_\gamma^{-1})$
- $\theta_i + \phi_i$ são, respectivamente, efeitos estruturados e não estruturados espacialmente e juntos formam o modelo BYM.

- δ_{it} é a interação entre os efeitos principais de espaço e tempo.

Clayton (1996) sugere especificar a matriz $\mathbf{R}_\delta$ como o produto de Kronecker das matrizes estruturadas, cujos efeitos principais interagem, por suposição. Como a formulação do modelo 2.28 possui 2 efeitos principais de tempo e 2 efeitos principais de espaço, temos $2 \times 2 = 4$ combinações de interações espaço-temporais possíveis, sendo que cada uma das combinações implicam diferentes inter-relações entre os δ_{it} (KNORR-HELD, 2000). Vamos fazer uma breve descrição de cada uma das interações:

2.5.1 Interação Tipo I

Se esperamos que os dois efeitos principais não estruturados γ e ϕ se interajam, temos, pela Regra de Clayton, que $\mathbf{R}_\delta = \mathbf{R}_\gamma \otimes \mathbf{R}_\phi = \mathbf{I} \otimes \mathbf{I} = \mathbf{I}$, portanto, todos δ_{it} são a priori independentes, de modo que sua densidade é

$$\pi(\boldsymbol{\delta}|\tau_\delta) \propto \exp \left\{ -\frac{\tau_\delta}{2} \sum_{i=1}^I \sum_{t=1}^T (\delta_{it})^2 \right\}. \quad (2.29)$$

De acordo com Knorr-Held (2000), δ_{it} pode ser considerado como covariáveis não observadas para cada ponto (i, t) , que não possuem nenhuma estrutura no espaço-tempo.

2.5.2 Interação Tipo II

É obtida combinando o efeito estruturado principal de tempo α (RW1 nesse caso) com o efeito não estruturado de espaço ϕ . Assim, pela Regra de Clayton, cada $\boldsymbol{\delta}_i = (\delta_{i1}, \dots, \delta_{iT})^\top, i = 1, \dots, I$ segue um modelo de passeio aleatório, independente de todas as outras regiões. A matriz $\mathbf{R}_\delta = \mathbf{R}_\alpha \otimes \mathbf{I}$ tem posto $I(T-1)$ e a densidade a priori de $\boldsymbol{\delta}$ pode ser escrita por

$$\pi(\boldsymbol{\delta}|\tau_\delta) \propto \exp \left\{ -\frac{\tau_\delta}{2} \sum_{i=1}^I \sum_{t=2}^T (\delta_{it} - \delta_{i,t-1})^2 \right\}. \quad (2.30)$$

O modelo 2.30 é adequado para o caso em que as tendências temporais são diferentes de região para região, mas não possui estrutura espacial (KNORR-HELD, 2000).

2.5.3 Interação Tipo III

Assume interação entre os efeitos principais de γ (não estruturado temporalmente) e θ (estruturado espacialmente). Portanto, cada $\boldsymbol{\delta}_t = (\delta_{1t}, \dots, \delta_{It})^\top, t = 1, \dots, T$ segue uma estrutura

de vizinhança modelada através do (independentes) modelo CAR. O posto $\mathbf{R}_\delta$ é $(I - 1)T$ e a priori pode ser escrita como

$$\pi(\boldsymbol{\delta}|\tau_\delta) \propto \exp \left\{ -\frac{\tau_\delta}{2} \sum_{t=1}^T \sum_{i \sim j} (\delta_{it} - \delta_{jt})^2 \right\}. \quad (2.31)$$

A especificação 2.31 é adequada se para cada ponto no tempo as tendências espaciais são diferentes, sem qualquer estrutura temporal.

2.5.4 Interação Tipo IV

É o tipo de interação com estrutura mais complexa, assumindo que os efeitos espacialmente e temporalmente estruturados θ e α , se interagem. Knorr-Held (2000) afirma que essa interação, do ponto de vista teórico é a mais interessante, devido ao fato de ser resultado de dois efeitos principais dependentes de tal maneira que δ é completamente dependente do espaço e do tempo, não permitindo-o ser fatorado em blocos independentes. A densidade a priori para δ nesse caso é da forma

$$\pi(\boldsymbol{\delta}|\tau_\delta) \propto \exp \left\{ -\frac{\tau_\delta}{2} \sum_{t=2}^T \sum_{i \sim j} (\delta_{it} - \delta_{jt} - \delta_{i,t-1} + \delta_{j,t-1})^2 \right\}. \quad (2.32)$$

com independentes contrastes $\delta_{it} - \delta_{jt} - \delta_{i,t-1} + \delta_{j,t-1}$. Note que $\mathbf{R}_\delta = \mathbf{R}_\alpha \otimes \mathbf{R}_\theta$ tem posto igual $(I - 1)(T - 1)$.

2.6 Análise Conjunta Combinada

Embora a análise univariada de modelos espaços-temporais forneça ao pesquisador várias soluções para suas questões científicas, questões relativas às interações entre variáveis respondidas ainda ficam sem solução e requerem uma análise conjunta.

Ntirampeba, Neema e Kazembe (2017) propuseram uma análise bivariada combinando dois modelos univariados de epidemiologia, através de uma componente espacial compartilhada entre os dois modelos, a informação obtida no modelo da primeira variável resposta pode ser acrescentada para melhorar a predição da segunda variável resposta. Sendo que os dois modelos possuem também covariáveis separadas. O modelo proposto tem seguinte estrutura:

Considere um modelo bivariado, que combina dois conjuntos de dados, sendo $\mathbf{y} = [\mathbf{y}_{i1}^\top \mathbf{y}_{i2}^\top]^\top$ um vetor contendo as variáveis respondidas 1 e 2. Os preditores, são da forma 2.8 e dados por:

$$\eta_{i1} = \mu_1 + \sum_j \beta_{j1} z_{1ij} + f_1(g_{i1}) + u_1(s_i) \quad (2.33)$$

e

$$\eta_{i2} = \mu_2 + \sum_j \beta_{j2} z_{2ij} + f_2(g_{i1}) + u_2(s_i) + \delta u_1(s_i) \quad (2.34)$$

em que,

- μ_1 e μ_2 são os interceptos referentes às respostas 1 e 2, respectivamente.
- β_{j1} e β_{j2} são os efeitos fixos relacionados as covariáveis $\mathbf{z}$ de acordo com sua respectiva variável resposta.
- $f_1(g_{i1})$ e $f_2(g_{i1})$ são referentes aos anos e são modelados como RW1, para cada variável resposta.
- u_1 e u_2 são efeitos espaciais referentes as regiões em cada variável resposta, modelados através de uma estrutura espacial GMRF. No entanto, u_1 é a componente compartilhada entre as duas variáveis respostas, através da interação δ .

O parâmetro δ faz a ligação entre as duas variáveis respostas e descreve quanto a estrutura capturada em u_1 é também inerente a η_{ij} .

Kifle, Hens e Faes (2017), apresentaram, em uma de suas formulações de modelos, um modelo espaço-temporal estacionário multivariado, em que a densidade conjunta $\pi(\mathbf{z}(s, t))$ pode ser formulada através de um produto de campos aleatórios univariados condicionais, isto é:

$$\pi(\mathbf{z}(s, t)) = \pi(z_1(s, t))\pi(z_2(s, t)|z_1(s, t))\pi(z_3(s, t)|z_1(s, t), z_2(s, t)).$$

Após apresentar uma reparametrização para implementação da abordagem utilizando o INLA, a conexão entre a especificação conjunta e condicionais é dada pela seguinte especificação incondicional:

$$\begin{aligned} \eta_1(\mathbf{s}, t) &= \beta_{01} + f_1(\mathbf{s}, t) \\ \eta_2(\mathbf{s}, t) &= \beta_{02} + \lambda_1 f_1(\mathbf{s}, t) + f_2(\mathbf{s}, t) \\ \eta_3(\mathbf{s}, t) &= \beta_{03} + \lambda_2 f_1(\mathbf{s}, t) + \lambda_3 f_2(\mathbf{s}, t) + f_3(\mathbf{s}, t) \end{aligned} \quad (2.35)$$

em que as componentes $f_i(\mathbf{s}, t)$, $i = 1, 2, 3$ são independentes entre os diferentes preditores e λ_i , $i = 1, 2, 3$ são coeficientes do modelo e β_{0i} , $i = 1, 2, 3$ são os respectivos interceptos de cada preditor.

Forlani et al. (2020) utilizam uma pequena variação da estrutura apresentada em (2.35) para modelar a poluição do ar, acrescentando uma covariável desalinhada. No trabalho em questão, os autores ajustaram modelos separados, para depois utilizarem uma estrutura compartilhada entre os preditores, que são da forma:

$$\begin{aligned}\eta_1(\mathbf{s}, t) &= \alpha_1 + f_1(\mathbf{s}) \\ \eta_2(\mathbf{s}, t) &= \alpha_2 + \lambda_{1,2}f_1(\mathbf{s}) + f_2(t) \\ \eta_3(\mathbf{s}, t) &= \alpha_3 + \beta_{k_s} + \lambda_{1,3}f_1(\mathbf{s}) + \lambda_{2,3}f_2(t) + f_3(t, k_s)\end{aligned}\tag{2.36}$$

sendo α_i os interceptos, β_{k_s} é um efeitos fixo de uma variável categórica (tipo de local), $\lambda_{i,j}$ os coeficientes de escala das componentes compartilhadas entre os preditores η_i e η_j , f_1 e f_2 são efeitos aleatórios compartilhados, enquanto f_3 é um termo de interação da diferença residual da tendência temporal para cada tipo de local.

2.7 Prioris de Complexidade Penalizada (PC priors)

Em Simpson et al. (2017) é proposto um novo conceito de construção de prioris, denominadas de prioris de complexidade penalizada (PC). Como o próprio nome remete, tais prioris penalizam o aumento da complexidade de um modelo - definido aqui como modelo flexível - em detrimento da utilização de um modelo mais simples, definido como modelo base.

Fuglstad et al. (2019) declaram que a inclusão de um Campo Aleatório Gaussiano (GRF) em um modelo pode acarretar em um superajuste, como por exemplo, estimando tendências espaciais ou tendências temporais espúrias.

Simpson et al. (2017) sugere tratar a questão do superajuste - assim como a inclusão de (GRF) em um modelo - tratando as componentes do modelo como extensões flexíveis de componentes mais simples, isto é, tratando *modelos flexíveis* como extensões ou generalizações do *modelos base* que são menos flexíveis e, em seguida, desenvolver prioris que encolham tais componentes flexíveis na direção do modelo base. Um exemplo seria enxergar um efeito

aleatório com variância maior que zero como uma extensão de um efeito aleatório com variância igual a zero e construir uma priori que encolhe a variância do efeito aleatório na direção do zero.

Dessa forma, as prioris PC são construídas obedecendo 4 princípios:

- P1 (Navalha de Occam): também conhecida como princípio da parcimônia, declara que modelos mais simples são preferíveis, até que haja justificativa suficiente, para a utilização de um modelo mais complexo. Portanto, deve-se construir uma priori que penalize desvios do modelo base.
- P2 (Medida de complexidade): utilizada para medir o aumento de complexidade comparada a um modelo base. A medida de complexidade se baseia na Divergência de Kullback-Leibler (KLD), medida apropriada para medir a informação perdida quando se utiliza um modelo base para aproximar um modelo flexível. A KLD é definida como:

$$KLD(\xi) = KLD(\pi(\mathbf{x}|\xi) || \pi(\mathbf{x}|\xi = 0)) = \int \pi(\mathbf{x}|\xi) \log \left(\frac{\pi(\mathbf{x}|\xi)}{\pi(\mathbf{x}|\xi = 0)} \right) d\mathbf{x},$$

em que ξ é um parâmetro de flexibilidade. Para o caso de $\xi = 0$, tem-se que $\pi(\mathbf{x}|\xi = 0)$ corresponde ao modelo base do modelo flexível $\pi(\mathbf{x}|\xi)$.

De acordo com Simpson et al. (2017), para fazer com que KLD seja uma medida interpretável como uma “distância” é necessário fazer uma correção de escala de modo que a nova medida que é uma função da KLD é uma distância unidirecional entre dois modelos f, g com densidades $\pi(\mathbf{x}|\xi)$ e $\pi(\mathbf{x}|\xi = 0)$, respectivamente, definida por

$$d(\xi) = \sqrt{2KLD(\xi)}.$$

Desse modo, consideramos d como a medida de complexidade do modelo f quando comparado ao modelo g .

- P3 (Taxa de penalização constante): a penalização para desvios do modelo base, parametrizado com distância d , deve ser feita através de uma taxa de penalização constante r , tal que a priori satisfaça a propriedade de perda de memória:

$$\frac{\pi_d(d + \delta)}{\pi_d(d)} = r^\delta, d, \delta \geq 0$$

para alguma constante $0 < r < 1$.

A distribuição contínua que satisfaz essa propriedade é a distribuição exponencial, definida por:

$$\pi(d|\lambda) = \lambda e^{-\lambda d}, d > 0$$

com taxa de penalização constante $r = e^{-\lambda}$. Portanto, a mudança relativa na priori quando a distância cresce δ não depende da distância atual d . Isso corresponde a seguinte priori no seu espaço original:

$$\pi(\xi|\lambda) = \lambda e^{-\lambda d(\xi)} \left| \frac{\partial d(\xi)}{\partial(\xi)} \right|$$

P4 (Escala definida pelo usuário): este princípio declara que o usuário tem uma ideia do tamanho do parâmetro de interesse ou de uma propriedade do modelo correspondente. Sendo assim, podemos escolher λ e controlar a massa na cauda da priori, através da relação:

$$P(Q(\xi) > U) = \alpha,$$

em que Q é uma transformação interpretável do parâmetro de flexibilidade ξ , U é um *sensible*, isto é, um valor razoável de limite superior definido pelo usuário que especifica o que ele pensa sobre como é um evento na cauda e α é o peso que ele atribui a este evento.

2.7.1 Prioris de complexidade penalizada para a precisão de um efeito aleatório gaussiano

Seja $\mathbf{x}$, um efeito aleatório gaussiano da forma $\mathbf{x} \sim \mathcal{N}(\mathbf{0}, \tau^{-1} \mathbf{R}^{-1})$, com $R \geq 0$ conhecida. A matriz $\mathbf{R}$ definida dessa forma geral, permite que esta derivação da priori PC também seja válida para as precisões dos modelos apresentados nas seções (2.3.6), (2.4.1) e (2.4.2), como salientado por [Simpson et al. \(2017\)](#).

Sabe-se que a Divergência de Kullback-Leibler entre duas normais p -variadas com vetor de médias e matriz de covariâncias dadas por $\boldsymbol{\mu}_1$, $\boldsymbol{\mu}_0$, $\boldsymbol{\Sigma}_1$ e $\boldsymbol{\Sigma}_0$, respectivamente é dada por (ver Apêndice (A)):

$$KLD(\mathcal{N}_p(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1) || \mathcal{N}_p(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)) = \frac{1}{2} \left\{ tr(\boldsymbol{\Sigma}_0^{-1} \boldsymbol{\Sigma}_1) + (\boldsymbol{\mu}_0 - \boldsymbol{\mu}_1)^\top \boldsymbol{\Sigma}_0^{-1} (\boldsymbol{\mu}_0 - \boldsymbol{\mu}_1) - p - \log \left(\frac{|\boldsymbol{\Sigma}_1|}{|\boldsymbol{\Sigma}_0|} \right) \right\}$$

Sendo assim, para obter a priori para a precisão τ , consideramos a matriz de covariâncias $\boldsymbol{\Sigma}_1 = \tau^{-1} \mathbf{R}$ para o modelo flexível e $\boldsymbol{\Sigma}_0 = \mathbf{0}$ para o modelo base. Desse modo, temos que estudar o comportamento limitante de $\boldsymbol{\Sigma}_0 = \tau_0^{-1} \mathbf{R}$ para um valor fixo grande de τ_0^{-1} . Sendo assim, temos que a Divergência de Kullback - Leibler para o parâmetro τ é:

$$KLD(\tau) = \frac{p}{2} \frac{\tau_0}{\tau} \left(1 + \frac{\tau}{\tau_0} \log \left(\frac{\tau}{\tau_0} \right) - \frac{\tau}{\tau_0} \right) \longrightarrow \frac{p}{2} \frac{\tau_0}{\tau}, \text{ quando } \tau \ll \tau_0.$$

Portanto, $d(\tau) = \sqrt{p\tau_0/\tau}$. Como $d \sim Exp(\lambda_0)$, obtém-se a priori PC de τ pelo Teorema da transformação, assim, tem-se que $\tau \sim \text{Gumbel tipo-2}(1/2, \lambda)$, com $\lambda = \lambda_0 p^{1/2} \tau_0^{1/2}$, portanto:

$$\pi(\tau) = \frac{\lambda}{2} \tau^{-\frac{3}{2}} \exp \left(-\lambda \tau^{-1/2} \right), \tau > 0, \lambda > 0.$$

A forma que o usuário determina a escala é definida pela função $Q(\tau) = \frac{1}{\sqrt{\tau}} = \sigma$, uma vez que esta priori corresponde a uma distribuição exponencial com taxa λ para o desvio padrão σ . Portanto, o usuário pode determinar λ impondo sua noção de escala através da especificação de (U, α) tal que $P(1/\sqrt{\tau} > U) = \alpha$, para $U > 0$, o que implica que $\lambda = -\frac{\log(\alpha)}{U}$. O parâmetro λ determina a magnitude da penalidade para desvios do modelo base, sendo maior a penalidade, quanto maior for o seu valor (SIMPSON et al., 2017).

2.7.2 Priori de complexidade penalizada para o parâmetro de mistura ϕ no modelo BYM2

Para derivar a priori PC para o parâmetro de mistura ϕ do modelo BYM2, consideramos que o modelo BYM2 $\mathbf{x} \sim \mathcal{N}_p(\mathbf{0}, \boldsymbol{\Sigma}_{flex}(\phi))$ é uma extensão do modelo base $\mathbf{x} \sim \mathcal{N}_p(\mathbf{0}, \boldsymbol{\Sigma}_{base})$ em que $\boldsymbol{\Sigma}_{flex}(\phi) = (1 - \phi)\mathbf{I} + \phi \mathbf{Q}_\star^-$ e $\boldsymbol{\Sigma}_{base} = \mathbf{I}$.

À luz do resultado da seção (2.7.1), temos que a Divergência de Kullback-Leibler é dada por:

$$KLD(\mathcal{N}_p(\mathbf{0}, \mathbf{\Sigma}_{flex}(\phi)) || \mathcal{N}_p(\mathbf{0}, \mathbf{\Sigma}_{base})) = \frac{1}{2} \{tr(\mathbf{\Sigma}_{flex}(\phi)) - p - \log(|\mathbf{\Sigma}_{flex}(\phi)|)\}$$

Do mesmo modo realizado para $d(\tau)$, assume-se priori exponencial com parâmetro λ_0 para $d(\phi) = \sqrt{2KLD(\phi)}$, dada por:

$$d(\phi) = \sqrt{p\phi \left(\frac{1}{p} tr(\mathbf{Q}_*^-) - 1 - \sum_{i=1}^p \log(1 - \phi + \phi \tilde{\gamma}_i) \right)},$$

em que foi utilizado o resultado apresentado em [Simpson et al. \(2017\)](#) de que

$$|\mathbf{\Sigma}_{flex}(\phi)| = |(1 - \phi)\mathbf{I} + \phi \mathbf{Q}_*^-| = \prod_{i=1}^p (1 - \phi + \phi \tilde{\gamma}_i),$$

com γ_i sendo os autovalores da matriz $\mathbf{Q}_*$. Sendo assim, definidos como:

$$\tilde{\gamma}_i = \begin{cases} \frac{1}{\gamma_i}, & \text{se } \gamma_i > 0 \\ 0, & \text{se } \gamma_i = 0. \end{cases}$$

Diferentemente da priori PC para a precisão de um efeito aleatório gaussiano, a priori PC para o parâmetro de mistura não possui forma fechada, portanto, ela é obtida computacionalmente. Uma implementação da mesma está disponível no software R ([R CORE TEAM, 2020](#)), através da função `inla.pc.bym.phi()` do pacote R-INLA ([RUE; MARTINO; CHOPIN, 2009; RUE et al., 2017](#)).

A declaração probabilística feita pelo usuário para especificar completamente a priori PC para ϕ também é feita de maneira similar, devendo o usuário estabelecer um limite superior U para ϕ e uma probabilidade α , isto é, $P(\phi < U) = \alpha$.

2.7.3 Priori de complexidade penalizada para o parâmetro de correlação ρ de um modelo AR(1) com modelo base $\rho = 0$

De acordo com [Simpson et al. \(2017\)](#), quando consideramos uma componente aleatória $\mathbf{x} \sim \mathcal{N}_p(\mathbf{0}, \mathbf{R})$, sendo $\mathbf{R}$ uma matriz de correlação de ordem $p \times p$, temos que a função da distância para um modelo base $\mathbf{x} \sim \mathcal{N}_p(\mathbf{0}, \mathbf{I})$ é dada pela função:

$$d(\mathbf{R}) = \sqrt{-\log|\mathbf{R}|}.$$

Para quando a componente $\mathbf{x}$ segue um modelo auto-regressivo de ordem 1, temos que a matriz $\mathbf{R}$ é da forma 2.23 e, portanto, seu determinante equivale a $(1 - \rho^2)$. Assim, tem-se que a função da distância para um modelo base $\rho = 0$ é dada por:

$$d(\rho) = \sqrt{-\log(1 - \rho^2)}.$$

A priori para $d(\rho)$ é uma distribuição exponencial com parâmetro λ , portanto, pelo Teorema da transformação, obtém-se que a priori PC para ρ é:

$$\pi(\rho) = \lambda \exp(-\lambda d(\rho)) \frac{|\rho|}{d(\rho)(1 - \rho^2)}.$$

A maneira pela qual o usuário define o parâmetro λ é definindo (U, α) tal que:

$$P(|\rho| > U) = \alpha, \quad 0 \leq U < 1, \quad 0 < \alpha < 1,$$

em que $\lambda = -\frac{\alpha}{d(U)}$.

3 MATERIAL E MÉTODOS

3.1 Área de Estudo e Dados

O estudo foca nos autos de infração ambientais ocorridos nos municípios do estado de Minas Gerais, durante o período de 2008 a 2016. O estado de Minas Gerais, pertencente à região Sudeste do Brasil fica localizado entre os paralelos $14^{\circ}13'58''$ e $22^{\circ}54'00''$ de latitude sul e os meridianos de $39^{\circ}51'32''$ e $51^{\circ}02'35''$ a oeste de Greenwich. Sendo assim, correspondendo a um fuso horário -3h em relação a Greenwich. Possui extensão territorial de $586.521.123\text{km}^2$, Produto Interno Bruto (PIB) estimado em aproximadamente R\$580 bilhões e taxa de urbanização de 85,3%.

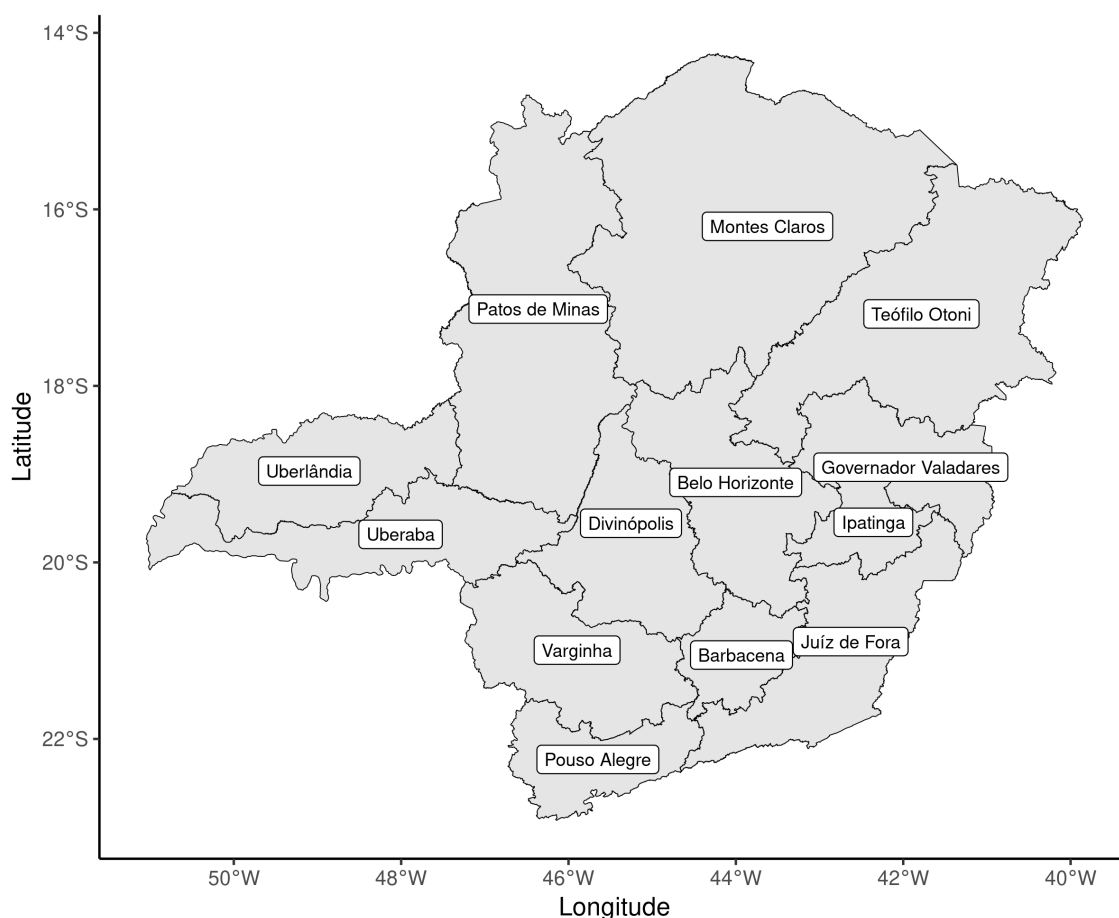
O estado possui um território bastante heterogêneo, sendo subdividido em 13 regiões geográficas intermediárias, 70 regiões geográficas imediatas e 853 municípios, possuindo a décima quarta densidade demográfica do país, sendo de $33,41\text{hab}/\text{km}^2$. Ainda de acordo com o órgão, tais subdivisões do território do estado são úteis para fins de criações de políticas públicas, subsídio ao sistema de decisões com respeito à localização de atividades econômicas, sociais e tributárias assim como, para atividades de planejamento, estudos e identificação das estruturas espaciais de regiões metropolitanas e outras formas de aglomerações urbanas e rurais.

As Figuras 3.1 e 3.2 apresentam o mapa geográfico do estado de Minas Gerais de acordo com os limites das regiões imediatas e intermediárias do estado. Enquanto que na Figura 3.3 apresenta o mapa geográfico do estado de Minas Gerais com limites municipais (bordas da cor branca), bem como a estrutura de vizinhança entre seus municípios representada pelo grafo (arestas azuis).

O meio ambiente de Minas Gerais pode ser resumido por 3 biomas de características distintas, sendo eles, a Mata Atlântica, o Cerrado e a Caatinga. A Figura 3.4 apresenta a distribuição espacial dos biomas presentes no estado de Minas Gerais.

O Cerrado corresponde a cerca de 57% do território do estado mineiro, o clima predominante é o tropical sazonal com duas estações bem definidas: inverno seco e verão chuvoso. A temperatura média anual é de 23°C , enquanto que as máximas correspondem a 40°C . Os meses mais frios do ano são junho e julho. A vegetação desse bioma é bastante heterogênea, composta por campos limpos desprovidos de vegetação lenhosa, matas ciliares, veredas, cerradão, campos rupestres e etc. Os solos são, em sua maioria, antigos, fortemente intemperizados, ácidos, profundos e bem drenados, tendo baixa fertilidade natural e elevadas concentrações de metais.

Figura 3.1 – Mapa das regiões intermediárias do estado de Minas Gerais

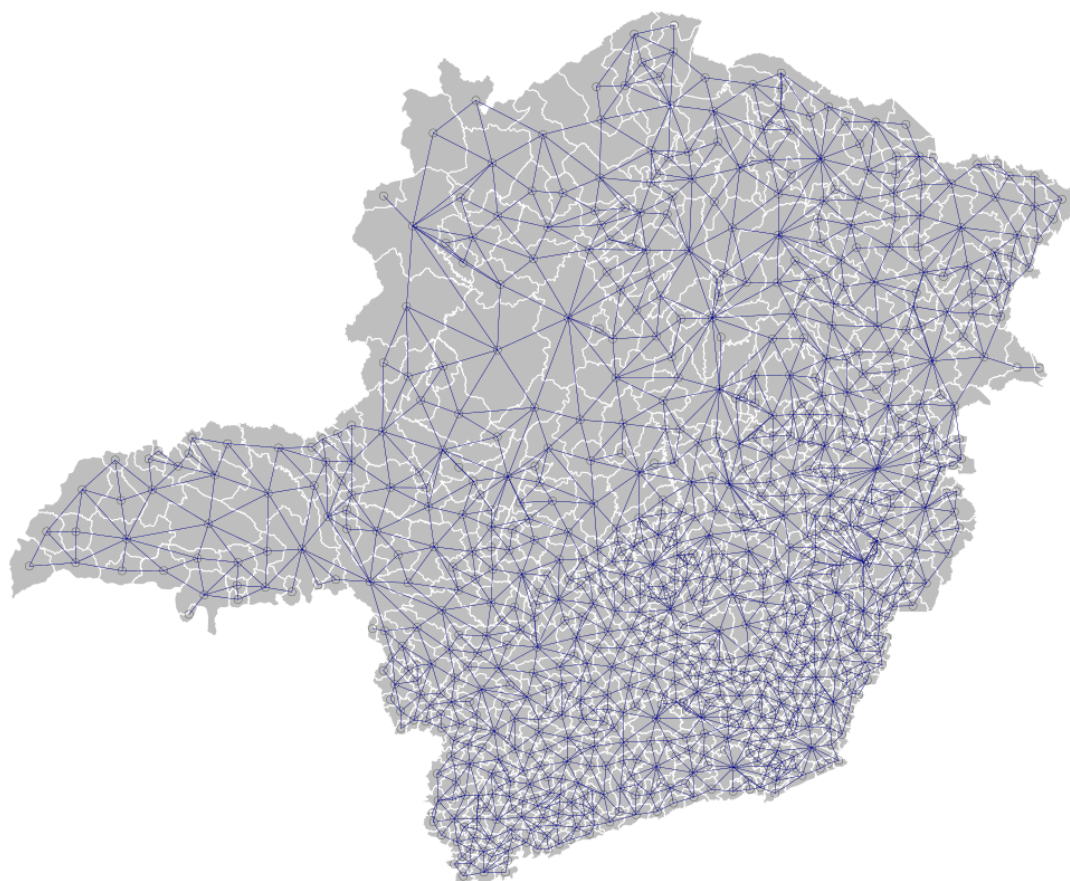


FONTE: Elaborado pelo autor.

A Mata Atlântica localiza-se sobre serras montanhosas que se estendem ao longo da costa brasileira e está presente em 17 estados brasileiros. Esse fato é uma das razões pelas quais este bioma possui um alto índice pluviométrico, fazendo com que a Mata Atlântica ocupe a posição de uma das florestas pluviais do planeta. Todas as regiões deste bioma são extremamente úmidas e apresentam temperaturas médias elevadas durante o ano todo, variando em torno de 25°C . A vegetação é densa e constantemente verde. O solo possui características de raso, ácido, pouco oxigenado, sempre úmido e pouco iluminado, em contrapartida, é extremamente fértil, devido a uma camada de restos de vegetação que cobre a superfície do solo, denominada de serapilheira, nessa camada encontram-se matérias orgânicas como folhas, caules, cascas de frutos e entre outras. É nesta volumosa decomposição de matéria orgânica em que ocorre grande parte da reciclagem dos nutrientes deste bioma.

A Caatinga é um bioma exclusivamente brasileiro situa-se, predominantemente, na região Nordeste do Brasil e tem uma pequena ocupação na região Norte de Minas Gerais com

Figura 3.3 – Mapa do estado de Minas Gerais com sua respectiva estrutura de vizinhança entre seus municípios. Os limites dos municípios são descritos pela borda (cor branca). A definição de vizinhança entre municípios se dá por aqueles que compartilham de uma mesma borda. A estrutura de vizinhança é informada pelo grafo (cor azul); os nós localizados nos centroides de cada município representam o próprio município enquanto que as arestas que conectam os nós dos municípios representam que estes são vizinhos



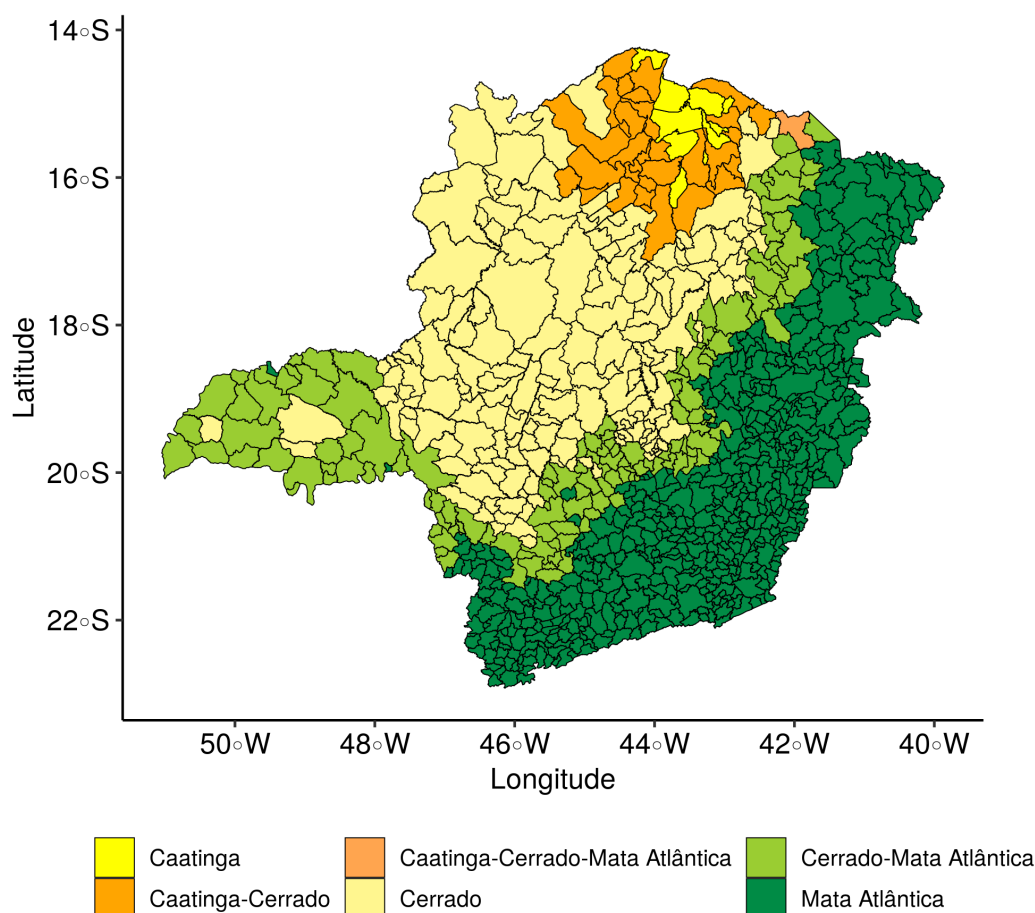
FONTE: Elaborado pelo autor.

ocorrência, números de notificações e valor das mesmas é uma opção metodológica que abre novas frentes de estudo, mas dificilmente este estudo poderia ser feito com dados georreferenciados. A coleta do dado é baseada em relatórios por municipalidade.

Consideramos para a escolha do recorte temporal o ano limite para flexibilização do uso consolidado de áreas protegidas (2008) até a disponibilidade do sistema para época de coleta de dados (2016).

Ao todo, foram avaliados 4.518 autos de infração, contendo as seguintes variáveis: contagem e valores agregados dos autos de infrações ambientais por ano e município, número de ocorrências de cada tipo de intervenção, classificados em 23 tipos. No banco de dados também

Figura 3.4 – Distribuição espacial dos biomas presentes nos municípios do estado de Minas Gerais



FONTE: Elaborado pelo autor a partir de dados do IBGE, 2019.

estão relacionados os números de autos de infrações em APP e RL como nascentes, lagoas, brejos, veredas, reservatórios artificiais e não especificadas. Adicionalmente, foram incluídos no banco de dados do Índice de Desenvolvimento Humano Municipal (IDHM). A descrição das covariáveis de acordo com o tipo de intervenção está descrita na Tabela 3.1.

O software R (R CORE TEAM, 2020), bem como os pacotes stats (R CORE TEAM, 2020), sp (BIVAND; PEBESMA; GOMEZ-RUBIO, 2013), maptools (BIVAND; LEWIN-KOH, 2017), spdep (BIVAND; HAUKE; KOSSOWSKI, 2013), tidyverse (WICKHAM et al., 2019) foram utilizados para a descrição dos dados e pacote R-INLA (RUE; MARTINO; CHOPIN, 2009; RUE et al., 2017) para o ajuste dos modelos apresentados.

Na seção 3.2, iremos apresentar a modelagem para cada uma das variáveis resposta, ou seja para o número e para o valor de autos de infrações ambientais ocorridos nos municípios de Minas Gerais. Na seção 3.3, trataremos de apresentar como iremos realizar a seleção de cada um dos modelos com respeito à família de verossimilhanças e quais os modelos gaussianos iremos

Tabela 3.1 – Tipo de Intervenção ocorrida em Áreas de Preservação Permanente e Reserva Legal entre 2008 e 2016, no estado de Minas Gerais

Covariável	Tipo de intervenção
T1	Instalação de fornos de carvão
T2	Supressão da vegetação
T3	Práticas de queimadas
T4	Construção de Estação de Tratamento de Esgoto (ETE) ou lagoas para tratamento de dejetos
T5	Construção de baias, curral ou bebedouros
T6	Construções de infraestrutura como casa, muro, piscina, passarela e rampas
T7	Construção de açudes, represas, reservatórios, tanques para peixes
T8	Construção de fossas, cisternas, cacimbas
T9	Construções para captação de água como drenos, introdução de manilhas, passagem de adutora, lançamento de tubulação
T10	Construção de barragens ou diques
T11	Abertura de valas
T12	Movimentação de terra como assoreamento, aterro, soterramento, revolvimento, terraplenagem, remoção do solo, depósito de terra ou areia, aragem, gradagem
T13	Intervenção com uso de maquinários
T14	Uso de defensivos
T15	Criação de animais bovinos e equinos
T16	Extração de areia, cascalho ou solo
T17	Depósito de resíduos sólidos de construção
T18	Silvicultura
T19	Lavoura
T20	Parcelamento do solo
T21	Abertura de estradas
T22	Queima de lixos
T23	Impedimento da regeneração natural

utilizar para compor a estrutura do preditor linear. Por fim, a seção 3.4 descreve a modelagem conjunta das variáveis resposta.

3.2 Modelagem espaço-temporal bayesiana para o número e para o valor dos autos de infrações ambientais

Inicialmente iremos modelar o número de autos e o valor dos autos separadamente.

3.2.1 Número de autos

Quando a variável resposta é uma contagem, frequentemente atribui-se à verossimilhança uma família de distribuições Poisson. Dessa maneira, iremos atribuir uma verossimi-

lança Poisson para cada observação do número de autos de infrações ambientais ocorridos no município s no ano t e denotaremos por $y_i(s, t)$. Além disso, iremos utilizar a função de ligação logarítmica para relacionar o preditor linear à média da distribuição. Dessa forma, podemos fazer esta representação através da seguinte formulação:

$$y_i(s, t) | \eta_i(s, t) \sim \text{Poisson}(\exp(\eta_i(s, t))). \quad (3.1)$$

O preditor linear $\eta(s, t)$ pode ser descrito de maneira geral como um intercepto somado a um conjunto de covariáveis acrescidos de componentes aditivas espaciais, temporais e uma componente de interação espaço-temporal. Portanto, temos:

$$\eta(s, t) = \beta_0 + \sum_{k=1}^K \beta_k x_k(s, t) + f_1(s) + f_2(t) + f_3(s, t) \quad (3.2)$$

em que,

- β_0 é o intercepto, com priori gaussiana com média 0 e precisão 0
- β_k é o efeito da k -ésima covariável x ocorrida no município s no ano t , com priori gaussiana com média 0 e precisão 0,001.
- $f_1(s)$ é uma componente espacial, modelada através de um modelo BYM2
- $f_2(t)$ é uma componente temporal, modelada através de um modelo IID, AR1, RW1 ou RW2
- $f_3(s, t)$ é uma componente de interação espaço-temporal, podendo ser do Tipo I, Tipo II, Tipo III e Tipo IV, a depender das componentes temporais e espaciais.

3.2.2 Valor dos autos

Para o caso em que o domínio da função de distribuição de probabilidade da variável resposta é o conjunto dos reais positivos $\mathbb{R}_+$, uma escolha inicial para a verossimilhança é frequentemente feita por uma distribuição gama. Todavia, a distribuição de probabilidade dos valores agregados para os autos de infrações ambientais nos municípios de Minas Gerais deve ser uma função cujo domínio é o conjunto dos números reais positivos unido ao número zero, em símbolos, $\mathbb{R}_+ \cup \{0\}$, uma vez que para que haja algum valor positivo de autuação ambiental em um determinado município e ano é necessária que haja a ocorrência de uma autuação ambiental.

Neste sentido, podemos construir um modelo que adapta-se tanto a resposta zero, quanto a resposta positiva do valor da autuação ambiental ao combinar duas verossimilhanças. Sendo a primeira verossimilhança responsável pela modelagem da ocorrência da autuação e a segunda verossimilhança responsável por modelar o valor da autuação. Uma possível formulação seria uma verossimilhança Binomial para modelar a ocorrência e uma verossimilhança Gama para modelar o valor da autuação. Neste caso, a grande vantagem de ter essas duas verossimilhanças é que podemos modelar a probabilidade de ocorrência de autuação ambiental e o valor da autuação ambiental separadamente ou ainda poder os preditores de cada verossimilhança possuir alguns de seus termos compartilhados entre os preditores.

Desse modo, iremos atribuir à variável $z_i(s, t)$ a ocorrência ou não de autuação ambiental no município s no ano t e $y_i(s, t)$ o valor da autuação no município s e ano t funções tais que:

$$z(s, t) = \begin{cases} 1, & \text{se ocorreu autuação ambiental no município } s \text{ no ano } t \\ 0, & \text{caso contrário} \end{cases}$$

e o valor da autuação como

$$y(s, t) = \begin{cases} NA, & \text{se } z(s, t) = 0 \\ \text{valor da autuação ambiental no município } s \text{ no ano } t, & \text{se } z(s, t) = 1. \end{cases}$$

Em seguida, definimos uma verossimilhança para cada variável resposta. Iremos, inicialmente, atribuir uma distribuição binomial para a variável $z_i(s, t) \sim \text{binomial}(1, \pi(s, t))$ e uma distribuição gama para a variável $y_i(s, t) \sim \text{gama}(\tau, \tau/\mu(s, t))$

Esta configuração é equivalente a um modelo *hurdle*-Gama, sendo a parte do *hurdle* definida em $y_i = 0$. Neste caso, o valor esperado da autuação ambiental é dada por $\pi(s, t)\mu(s, t)$, sendo $\mu(s, t)$ o valor esperado da verossimilhança gama. A parte $\pi(s, t)$ é interpretada como a probabilidade de se “atravessar” o *hurdle*, sendo, portanto, interpretado para o nosso caso como a probabilidade de ocorrer ao menos uma autuação ambiental no município s no ano t .

A maneira pela qual os preditores lineares $\eta_1(s, t)$ e $\eta_2(s, t)$ se relacionam com as respectivas médias das verossimilhanças são através das funções de ligação $\text{logit}(\cdot)$ e $\log(\cdot)$, respectivamente. Portanto,

$$\text{logit}(\pi(s,t)) = \eta_1(s,t) \text{ e } \log(\mu(s,t)) = \eta_2(s,t)$$

Os preditores lineares $\eta_1(s,t)$ e $\eta_2(s,t)$ tem a seguinte forma:

$$\begin{aligned}\eta_1(s,t) &= \beta_0 1 + f_1(s) + f_2(t) + f_3(s,t) \\ \eta_2(s,t) &= \beta_0 2 + \sum_{k=1}^K \beta_k x_k(s,t) + f_1(s) + f_2(t) + f_3(s,t)\end{aligned}$$

3.3 Seleção de modelos para número e valor de autos de infrações ambientais

Na busca de encontrar o modelo que melhor explica as variáveis respostas para o número e valor de autuações ambientais, iremos testar algumas configurações distintas de combinações das componentes espaciais, temporais e interações espaço-temporais nos preditores lineares definidos em (3.2) e (3.3). Tais configurações estão apresentadas na Tabela 3.2.

Com a mesma finalidade descrita anteriormente, iremos ajustar os modelos com as componentes descritas na Tabela 3.2 para outras verossimilhanças, a saber, Poisson inflacionada de zeros do tipo 0 (ZIP-0) e Poisson inflacionada de zeros do tipo 1 (ZIP-1) para o número de autuações ambientais, enquanto que para o valor das autuações também serão ajustadas as distribuições lognormal e Weibull substituindo a parte da verossimilhança gama no modelo da seção 3.2.2.

As verossimilhanças ZIP-0 e ZIP-1 são definidas da seguinte forma:

$$\text{ZIP-0: } \pi(y) = p \times I_{[y=0]} + (1-p) \times \text{Poisson}(y|y > 0).$$

$$\text{ZIP-1: } \pi(y) = p \times I_{[y=0]} + (1-p) \times \text{Poisson}(y).$$

Para ambos casos, p é um hiperparâmetro e é modelado por:

$$p = \frac{\exp(\theta)}{1 + \exp(\theta)},$$

em que a priori é definida para θ .

As distribuições a priori utilizadas para os hiperparâmetros são dadas da seguinte forma:

- Para os modelos espaciais BYM2, foram utilizadas prioris PC tanto para a precisão, quanto para o parâmetro de mistura. Os valores de (U, α) foram definidos da seguinte forma:

$$- \text{ Para o número de autos: } P(1/\sqrt{\tau_{bym2}} > sd(y_{num})) = 0,01.$$

Tabela 3.2 – Configurações das componentes espaciais, temporais e espaço-temporais consideradas para o ajuste dos modelos univariados para a resposta número e valor de autuações ambientais. Todos os modelos utilizaram $f(s) = bym2$.

Configurações	$f(t)$	$f(s, t)$
I	iid	ausente
II	ar1	ausente
III	rw1	ausente
IV	rw2	ausente
V	iid	tipo I
VI	iid	tipo III
VII	ar1	tipo II
VIII	ar1	tipo IV
IX	rw1	tipo II
X	rw1	tipo IV
XI	rw2	tipo II
XII	rw2	tipo IV

- Para a ocorrência ou não de autos: $P(1/\sqrt{\tau_{bym2}} > sd(z_{ocor})) = 0,01$.
- Para o valor de autos: $P(1/\sqrt{\tau_{bym2}} > sd(y_{valor})) = 0,01$.
- Para o parâmetro de mistura ϕ : $P(\phi < 0,50) = 0,50$.
- Para a precisão dos modelos temporais AR1, RW1 e RW2 foram utilizadas prioris PC de acordo com a recomendação apresentada no repositório padrão do pacote R-INLA *R-INLA project* (www.r-inla.org).
 - Para o número de autos: $P(1/\sqrt{\tau} > 1) = 0,01$.
 - Para a ocorrência ou não de autos com ligação $logit(\cdot)$: $P(1/\sqrt{\tau} > 0,50) = 0,01$.
 - Para o valor de autos, com ligação log : $P(1/\sqrt{\tau} > 1) = 0,01$.
- Para as precisões das interações do tipo I,II,III e IV foram utilizadas prioris PC $P(1/\sqrt{\tau_{iid}} > sd(y)) = 0,01$.
- Para precisão da verossimilhança gama, foram utilizadas prioris gama, que é definida para o log da precisão, temos que $\log(\tau) \sim loggama(1, 5 \times 10^{-5})$.

A seleção dos modelos será feita segundo o Critério de Informação de Watanabe (WAIC) (WATANABE, 2010), também não deixando de observar a raiz quadrada do erro quadrático médio (RMSE) e Brier Score para modelos com verossimilhança binomial (BRIER, 1950).

Quando necessário, observaremos a média do score logarítmico de validação cruzada (LogScore) para também medir a qualidade preditiva do modelo (GNEITING; RAFTERY, 2007). Por fim, caso haja mais de um modelo adequado para o número ou valor de autuação ambiental, iremos comparar as posteriores das precisões de cada componentes dos modelos para escolhermos o modelo final para cada resposta. Caso haja uma ou mais componentes com alta precisão com respeito a média e desvio padrão a posteriori optaremos pela exclusão da componente quando possível ou pela opção de outro modelo em que isto não ocorra, pois altas precisões indicam que indicam uma baixa variabilidade da componente, que na maioria das vezes possuem média igual a zero, portanto, pouco influenciam a variável resposta. Sendo assim, se a variância da componente é muito baixa, significa que seus valores estão muito próximos da média, que é igual a 0, por definição dos mesmos.

3.4 Modelagem Espaço-Temporal Bayesiana Conjunta

O modelo a ser descrito é conjunto no sentido de que será especificado uma verossimilhança para cada variável resposta, bem como, um preditor linear para cada verossimilhança, no qual os preditores conterão componentes compartilhadas entre si. O modelo é conjunto, no sentido de que para realizar o ajuste dos preditores, serão utilizadas todas os dados das variáveis resposta.

Os preditores podem ser definidos, de maneira geral, da seguinte forma:

$$\eta_1(s, t) = \beta_{01} + f_1(s) + f_2(t) + f_3(s, t) \quad (3.3)$$

$$\eta_2(s, t) = \beta_{02} + \sum_{k=1}^K \beta_{k2} x_k(s, t) + \lambda_{2,1} f_1(s) + \lambda_{2,2} f_2(t) + \lambda_{2,3} f_3(s, t) \quad (3.4)$$

$$\eta_3(s, t) = \beta_{03} + \sum_{k=1}^K \beta_{k3} x_k(s, t) + \lambda_{3,1} f_1(s) + \lambda_{3,2} f_2(t) + \lambda_{3,3} f_3(s, t), \quad (3.5)$$

em que:

- $\eta_1(s, t)$, $\eta_2(s, t)$ e $\eta_3(s, t)$ são os preditores para a ocorrência, número e valor de autos.
- β_{01} , β_{02} e β_{03} são os respectivos interceptos com prioris gaussianas com média 0 e precisão 0

- β_{k2} e β_{k3} são os coeficientes de efeitos fixos com prioris gaussianas com média zero e precisão 0,001 para o número e valor de autos de infrações ambientais, respectivamente.
- $f_1(s)$, $f_2(t)$ e $f_3(s, t)$ são as componentes espaciais, temporais e interações espaço-temporais, respectivamente. Tais interações são compartilhadas entre as respostas, devendo as mesmas, possuírem o mesmo modelo para cada resposta.
- $\lambda_{2,1}$, $\lambda_{2,2}$, $\lambda_{2,3}$, $\lambda_{3,1}$, $\lambda_{3,2}$, $\lambda_{3,3}$, são coeficientes de escalas para os suas respectivas componentes espaciais, temporais e espaço-temporais com cada uma, tendo priori gaussiana com média 1 e precisão 0,01.

Além disso, foram excluídas as respostas para número de autos de infração igual a zero para evitar uma dupla contagem, uma vez que essa informação está contida na verossimilhança binomial.

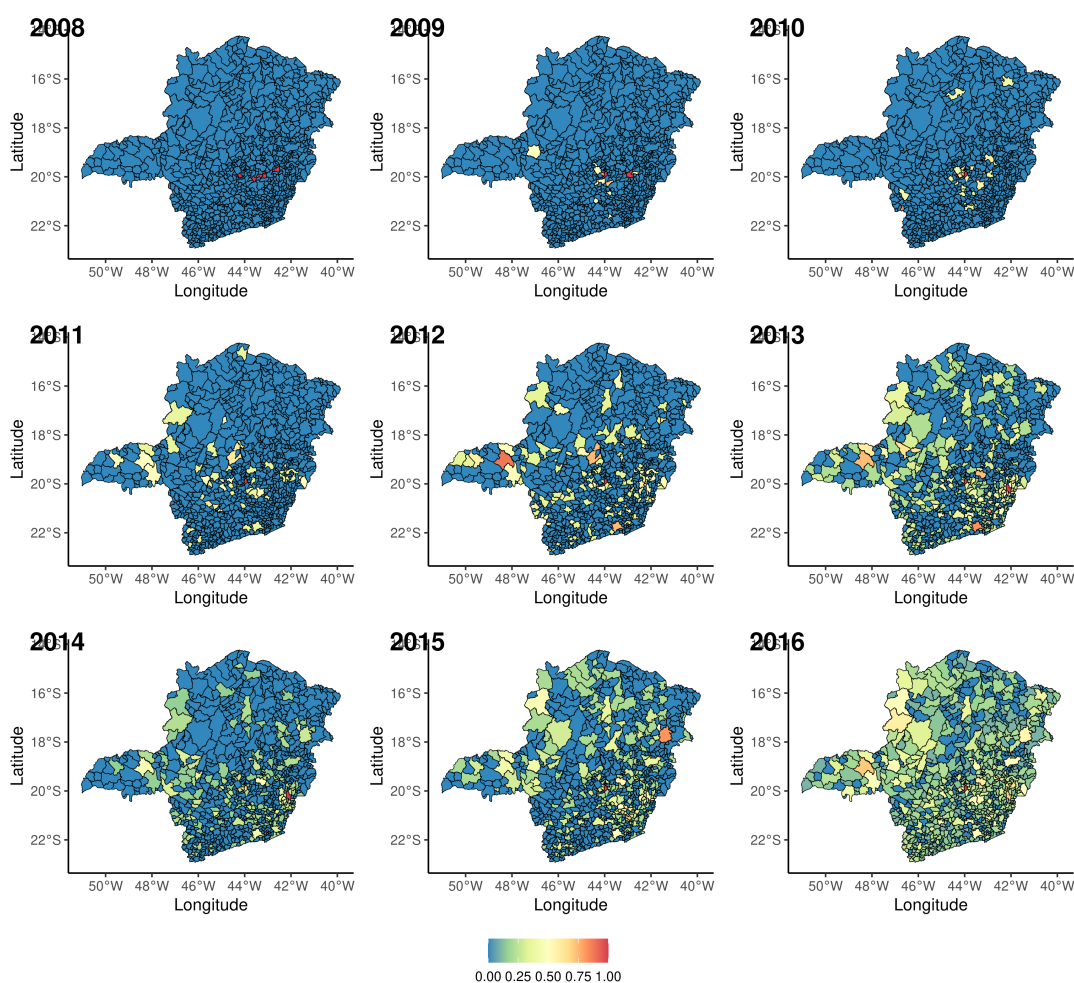
4 RESULTADOS

Para realizar a análise dos autos de infrações ambientais nos municípios de Minas Gerais de 2008 e 2016, foram ajustados modelos com diferentes estruturas espaciais, temporais, bem como, estes modelos, terem ou não alguma componente de interação espaço-tempo. Conjuntamente com tais estruturas, foram incluídas covariáveis, cujas características foram julgadas serem potenciais variáveis explicativas que poderiam explicar a resposta.

4.1 Modelo espaço-temporal bayesiano final para o número de autos

Os valores observados do número e valores dos autos de infrações ambientais estão descritos nas Figuras 4.1 indicam um aumento do número de autos ao longo dos anos, bem como, um aumento do número de municípios que tiveram autos de infrações nos anos de 2008 a 2016.

Figura 4.1 – Número de autos de infração observados nos municípios de Minas Gerais de 2008 a 2016 (em escala raiz quadrada)



A Tabela 4.1 dispõe os valores de WAIC e RMSE dos modelos ajustados para o número de infrações ambientais nos municípios de Minas Gerais entre os anos de 2008 a 2016.

Tabela 4.1 – Valores WAIC e RMSE para os modelos testados para o número de autos de infrações ambientais nos municípios de Minas Gerais para o período de 2008 a 2016 (em negrito está a combinação com melhor desempenho)

Preditor Linear	WAIC			RMSE		
	Poisson	ZIP tipo 0	ZIP tipo I	Poisson	ZIP tipo 0	ZIP tipo I
I	6242,03	11805,59	6244,70	0,93	0,78	0,93
II	6242,20	11804,74	6244,11	0,93	0,78	0,93
III	6241,24	11804,77	6243,60	0,93	0,77	0,94
IV	6239,89	11804,11	6242,60	0,93	0,78	0,93
V	5885,10	11549,21	5939,06	0,54	0,39	0,54
VI	6549,30	11807,27	6552,29	0,74	0,91	0,74
VII	5888,78	11543,89	5942,19	0,54	0,39	0,54
VIII	5955,28	11568,58	5995,86	0,64	0,46	0,57
IX	7213,85	13758,47	7212,11	0,67	3,21	0,67
X	NA	NA	NA	NA	NA	NA
XI	20135,42	83444,21	NA	0,55	115,75	NA
XII	NA	81685,49	NA	NA	122,60	NA

Pelos valores descritos na Tabela 4.1, se estratificarmos pela verossimilhança, observa-se que os modelos que se sobressaem perante os demais são aqueles que possuem os preditores lineares com as configurações V, VII e VIII. São os preditores que modelam a dependência temporal por meio de uma forma não estruturada ou os que modelam a dependência temporal através de uma estrutura auto-regressiva de ordem 1. Essas estruturas, sobretudo, a autorregressiva, tiveram melhores desempenhos em relação aos modelos com estruturas temporais *random walk*. Observe ainda que para cada verossimilhança estas 3 configurações apresentam menores valores de WAIC e também menores valores de RMSE. Porém, dentre todos os modelos comparados, aquele cuja verossimilhança é dada pela distribuição Poisson e o preditor linear é descrito pela configuração V apresentou menor valor de WAIC que os demais modelos. Sendo assim, este modelo é o escolhido para descrever o números de autos de infração ocorridos nos municípios de Minas Gerais, durante os anos de 2008 a 2016. Um resumo dos efeitos fixos e hiperparâmetros do modelo Poisson com a configuração V é apresentado na Tabela 4.2

Na Tabela 4.2, percebemos que a variável Índice de Desenvolvimento Humano Municipal possui efeito proporcional no número de autos de infração ambiental, sugerindo que quanto maior o IDHM do município, maior será o seu número de autos. Além disso, o efeito da variável IDHM é o de maior média a posteriori dentre as variáveis explicativas de efeitos fixos,

Tabela 4.2 – Média, desvio padrão e intervalos de credibilidade a posteriori (95%) para as covariáveis e hiperparâmetros do modelo final com verossimilhança Poisson e preditor linear com a configuração espaço-temporal V

Covariáveis	Média	Desvio padrão	$q_{0.025}$	$q_{0.975}$
Intercepto	-6.46	0.35	-7.16	-5.77
IDHM	4.53	0.50	3.55	5.52
T2	0.62	0.06	0.51	0.73
T3	0.27	0.12	0.03	0.51
T6	0.22	0.08	0.06	0.38
T7	0.29	0.12	0.04	0.53
T12	0.46	0.08	0.31	0.62
T14	1.25	0.47	0.29	2.13
T15	0.46	0.14	0.18	0.74
T19	0.45	0.17	0.13	0.78
Curso D'água	0.43	0.07	0.29	0.57
APP NE	2.56	0.14	2.24	2.79
Intervenções NE	1.32	0.12	1.09	1.55
Hiperparâmetros	Média	Desvio padrão	$q_{0.025}$	$q_{0.975}$
τ_{bym2}	12.94	3.38	8.08	21.15
ϕ_{bym2}	0.73	0.14	0.41	0.95
τ_{iid}	4.38	2.94	0.98	12.00
τ_{int}	6.55	0.88	4.89	8.37

indicando que essa variável, é a que mais possui relevância para explicar o número de autos ambientais, dentre as variáveis estudadas.

Para as variáveis com respeito ao tipo de intervenção ocorrida em Áreas de Preservação Permanente e Reserva Legal, temos que as variáveis T2 (supressão da vegetação), T3 (práticas de queimadas), T6 (construções de infraestrutura como casa, piscina,...), T7 (construção de açudes, tanque para peixes), T12 (movimentação de terra), T14 (Uso de defensivos), T15 (criação de animais bovinos e equinos), T19 (lavoura) e intervenção não especificada, foram significativamente positivas para explicar número de autos. Indicando que esses tipos de intervenções foram as mais frequentes nas APP.

Seguindo este mesmo raciocínio, as APP de cursos d'água e as não especificadas, foram as que tiveram efeitos médios a posteriori significativos. Indicando que tais APP foram as áreas de preservação permanentes que, frequentemente, sofreram mais intervenções e, portanto, maior número de autuações ambientais.

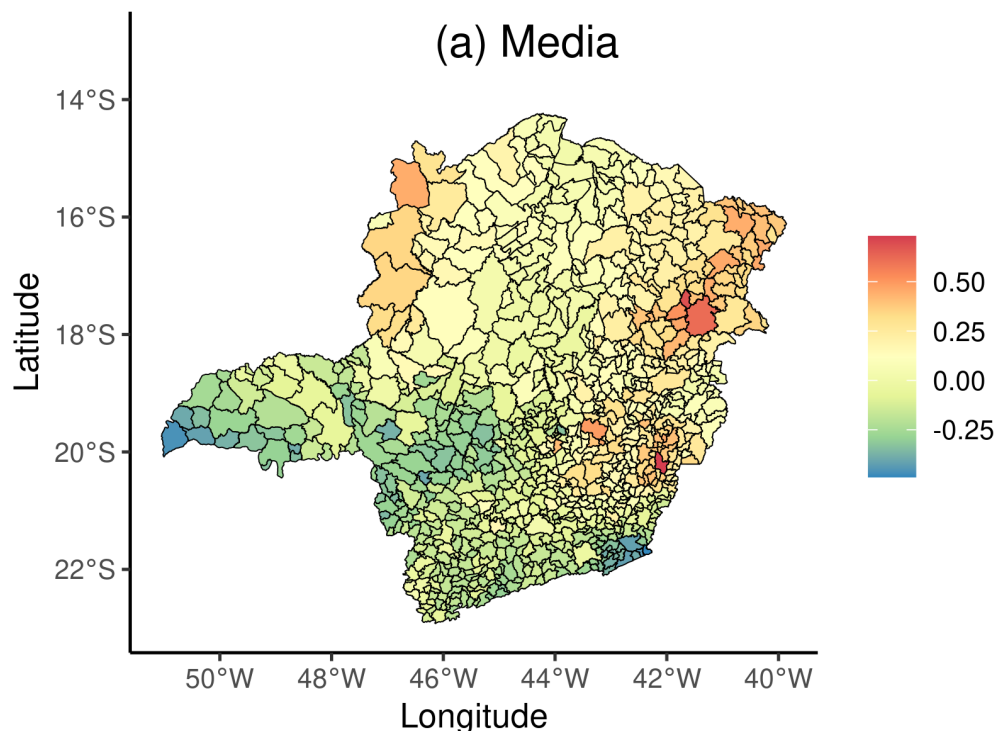
Com respeito aos hiperparâmetros, podemos ver que a precisão marginal espacial τ_{bym2} é 73% explicada, em média, pela sua componente espacial estruturada (ϕ_{bym2}), sendo o restante devido a superdispersão. A variabilidade espacial, pode ser observada na Figura 4.2. Para

a variável temporal anual, temos um efeito aleatório não estruturado no tempo cuja média a posteriori da precisão τ_{iid} é 4,38, o que significa que há a presença de variabilidade temporal anual suficiente para que haja um comportamento de crescimento anual dos números de autos de infração presente nos dados (ver Figura 4.3). No mesmo sentido, a precisão da interação espaço-tempo τ_{int} é do tipo I, isto é, supõe independência temporal e espacial para cada município em determinado ano. Dessa forma, a média a posteriori da precisão sendo pequena, significa uma alta variabilidade, portanto, além do efeito principal temporal comum a todos os municípios, há ainda, adicionalmente, outro efeito aleatório espaço-temporal atuante para explicar a variabilidade entre os números de autos para diferentes combinações de município e ano. Tal variabilidade, pode ser observado na Figura 4.4.

O gráfico das funções das marginais a posteriori das covariáveis e dos hiperparâmetros podem ser encontradas no Apêndice B.

Para interpretar a variabilidade do modelo, deve-se ter em mente que o modelo selecionado assume que a variação presente nos valores observados do número de autos de infração se deve a soma das componentes dos efeitos aleatórios espaciais e temporais globais acrescido da componente de interação espaço-tempo. Essa interação, mensura variações ou desvios dos efeitos espaciais e temporais globais para determinados municípios em determinados anos, ocasionando uma maior qualidade do ajuste. As médias da posteriori dos efeitos espaciais dos municípios para o modelo final com verossimilhança Poisson e preditor linear com a configuração V para o número de autos de infrações ambientais ocorridos em Minas Gerais dentre os anos de 2008 a 2016, são apresentadas na Figura 4.2.

Figura 4.2 – Média a posteriori do efeito aleatório espacial para os municípios de Minas Gerais para modelo final com verossimilhança Poisson e preditor linear com a configuração V para o número de autos de infração em Minas Gerais no período de 2008 a 2016

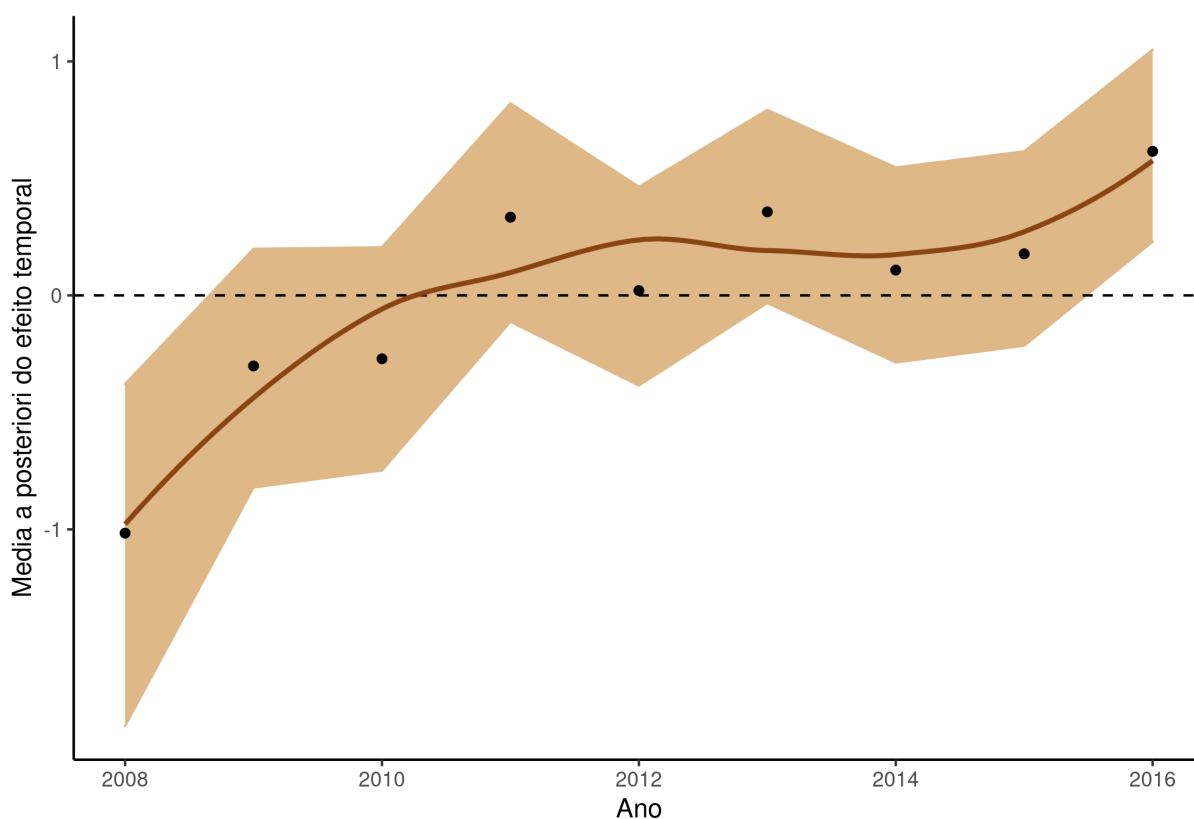


Na Figura 4.2 os municípios com médias a posteriori significativamente positivas para o efeito espacial são, principalmente, aqueles municípios pertencentes às regiões intermediárias de Ipatinga - tais como os municípios Vermelho Novo, Raul Soares e João Monlevade-, Juiz de Fora - com destaques para os municípios de Manhuaçu, Caputira, Matipó, Pedra Bonita, Brás Pires e Presidente Bernardes-, Barbacena - destacando-se os municípios de Senhora de Oliveira, Rio Espera, Piranga, Cipotânea, Alto Rio Doce e Desterro do Melo-, Teófilo Otoni - para os quais destacaram Itaipé, Ladainha, Setubinha, Poté e Palmópolis- e Montes Claros - em que Bonito de Minas, Cônego Marinho e Januária. Enquanto que os municípios com médias a posteriori significativamente negativas e estão situados, principalmente, nas regiões intermediárias de Varginha - com destaques para os municípios de São Tomás de Aquino, Itaú de Minas, Capetinga, Fama, Pratápolis, Poço Fundo, Cássia, Cabo Verde, Divisa Nova, Serrania e Fortaleza de Minas - e Uberaba - destacando os municípios Fronteira, Água Comprida, Carneirinho, Itapagipe, Planura, Campo Florido, Comendador Gomes, União de Minas e Pirajuba.

Na Figura 4.3 são apresentadas a média, quantis de 2,5% e 97,5% a posteriori, bem como ajuste de uma curva polinomial pelo método de Regressão Local (LOESS). Como pode

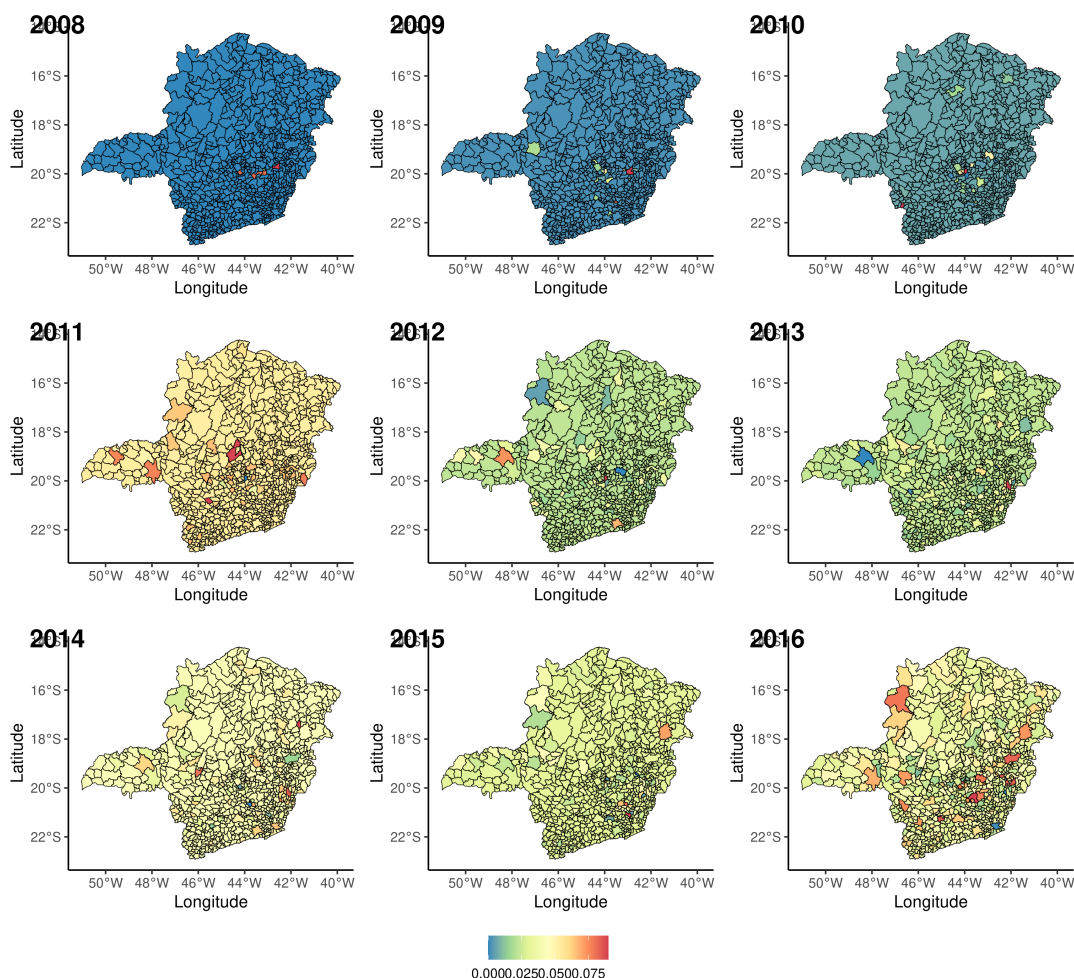
ser observado, há uma tendência temporal crescente, indicando um aumento geral dos números de autos de decorrer dos anos.

Figura 4.3 – Média a posteriori (pontos), intervalo de credibilidade de 95% (faixas) e ajuste via LOESS (curva) do efeito temporal do número de autos de infrações ambientais dos municípios de MG entre os anos de 2008 e 2016



De modo geral, as médias a posteriori dos efeitos das interações espaço-tempo apresentados na Figura 4.4 evidencia a tendência crescente nas médias a posteriori dos efeitos dos municípios ao longo dos anos. Indicando que a interação espaço-tempo foi capaz de captar uma tendência de crescimento global dos autos de infrações ambientais, além da melhora na qualidade do ajuste do modelo.

Figura 4.4 – Interações espaciais-temporais para o número de autos



Na Figura 4.5 é apresentado o gráfico de dispersão dos valores preditos e observados do modelo final Poisson para o número de autos de infração ambientais nos municípios de Minas Gerais, em que pode ser observado que o modelo foi capaz de prever razoavelmente os autos de infrações, haja vista que o coeficiente R^2 ajustado foi de 0,89. De modo equivalente, observa-se, também, o bom ajuste do modelo através dos valores preditos na Figura ??, se comparado aos valores observados do número de autos, apresentados na Figura 4.1.

Figura 4.5 – Valores preditos *versus* observados do número de autos de infração ambientais do modelo final com verossimilhança Poisson e preditor linear na configuração V

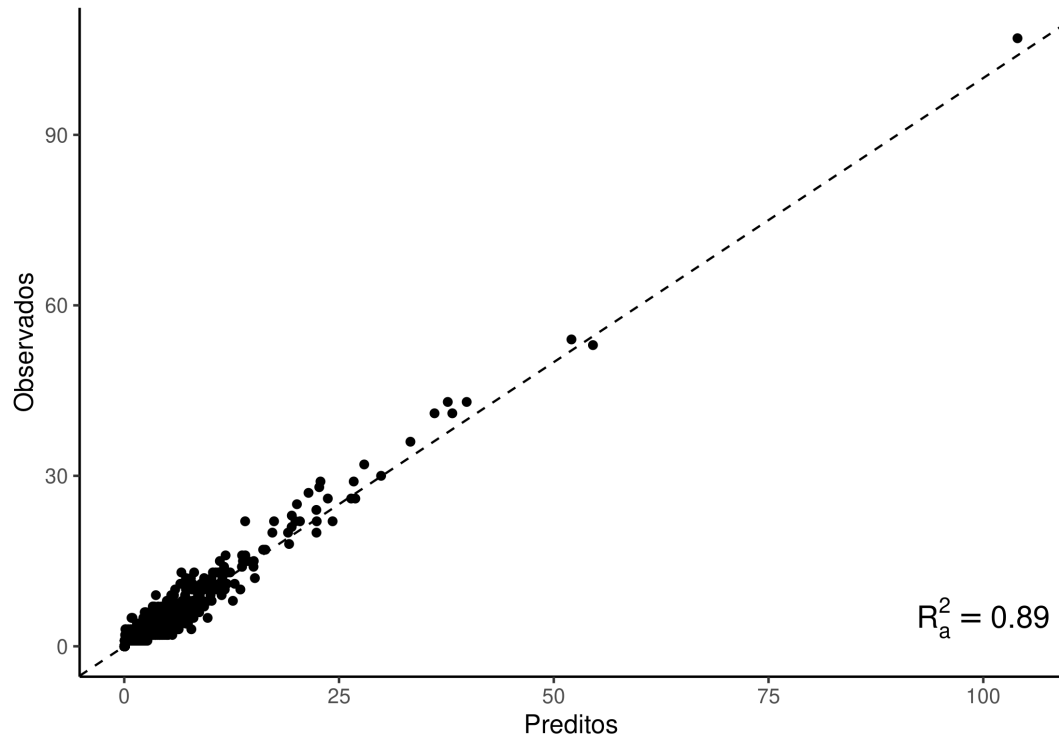
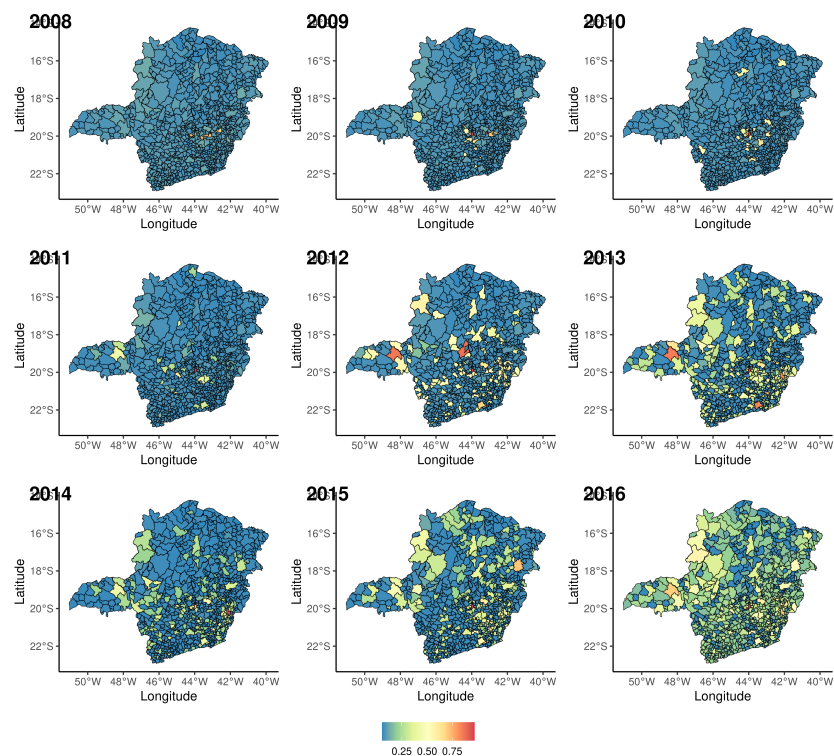


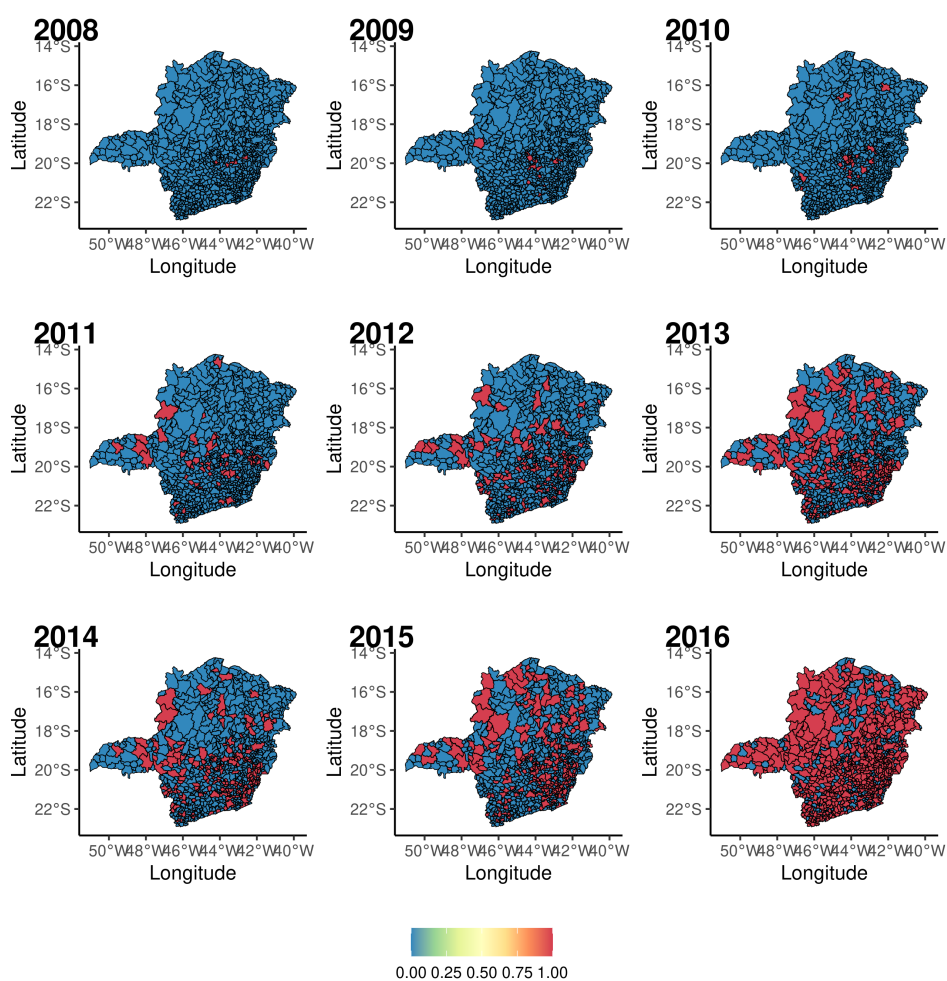
Figura 4.6 – Valores preditos do número de autos de infração ambientais do modelo final com verossimilhança Poisson e preditor linear na configuração V (na escala raiz quadrada)



4.2 Modelo hurdle-Gama para o valor de autos

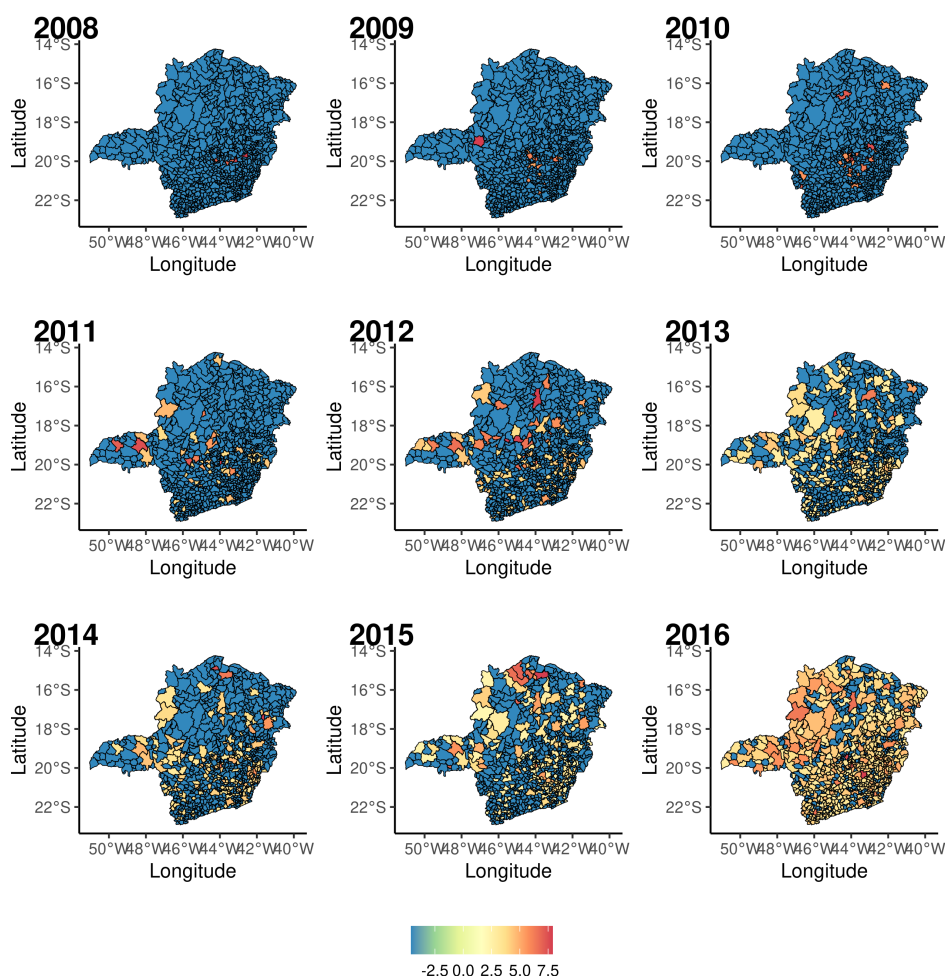
A ocorrência observada dos autos de infrações ambientais está descrita na Figura 4.7 indica um aumento de municípios no decorrer dos anos que tiveram em suas áreas a ocorrência de autos de infração nos anos de 2008 a 2016.

Figura 4.7 – Ocorrência de autos de infração observados nos municípios de Minas Gerais de 2008 a 2016



O valor observado dos autos de infrações ambientais está descrito na Figura 4.8 indicam um aumento dos valores agregados dos autos nos municípios no decorrer dos anos.

Figura 4.8 – Valor dos autos de infração observados nos municípios de Minas Gerais de 2008 a 2016 (em escala logarítmica)



Com a finalidade de simplificar a seleção dos modelos *hurdle*, iremos, primeiramente, selecionar a estrutura mais adequada para modelar a ocorrência dos autos de infração nos municípios de Minas Gerais. Sendo assim, todos os modelos *hurdle* ajustados, terão a estrutura referente a verossimilhança binomial, dado pelo modelo final desta seção.

A Tabela 4.3 dispõe os valores de WAIC, Brier Score e LogScore dos modelos ajustados para a resposta ocorrência de autos de infração nos municípios de Minas Gerais.

Observa-se na Tabela 4.3 que a maioria das configurações do preditor linear apresentaram desempenhos similares com respeito a predição. Enquanto que pelos valores de WAIC, a configuração VIII obteve melhor desempenho. Portanto, seleciona-se o preditor na configuração VIII como preditor para explicar a ocorrência de autos de infração.

Uma vez escolhida a configuração do preditor linear da verossimilhança binomial para modelar a ocorrência de autos, faremos a seleção da melhor combinação de preditor linear e

Tabela 4.3 – Valores das medidas WAIC, Brier Score e LogScore dos modelos ajustados para ocorrência de autos de infração nos municípios de Minas Gerais, considerando preditores cujas configurações estão descritas na Tabela 3.2

Preditor Linear	WAIC	Brier Score	LogScore
I	4771,93	0,28	0,31
II	4773,36	0,28	0,31
III	4773,84	0,28	0,31
IV	4772,29	0,28	0,31
V	4773,49	0,28	0,31
VI	5540,15	0,32	0,35
VII	4772,42	0,28	0,31
VIII	4668,49	0,27	0,31
IX	5427,12	0,29	0,35
X	55717,39	0,30	33,53
XI	5985,16	0,27	0,40
XII	144863,68	NA	29,22

verossimilhança para modelar o valor dos autos de infração em Minas Gerais. Na Tabela 4.4 apresentamos os valores de WAIC e RMSE dos modelos *hurdle*-Gama utilizando o preditor VIII para a verossimilhança binomial, para as configurações de preditores descritos na Tabela 3.2.

Tabela 4.4 – Valores WAIC e RMSE para os modelos testados para o valor de autos de infrações ambientais nos municípios de Minas Gerais para o período de 2008 a 2016 (todas as combinações dos modelos utilizaram o preditor VIII para a verossimilhança binomial. Em negrito encontra-se a combinação com melhor desempenho)

Preditor Linear	WAIC			RMSE	
	Total	Binomial	Gama	Binomial	Gama
I	33319,09	4675,51	28643,58	0,27	171854,4
II	33304,63	4664,84	28639,79	0,27	171812,6
III	33313,47	4670,57	28642,90	0,27	172092,5
IV	33301,78	4663,17	28638,62	0,27	121989,9
V	NA	NA	NA	NA	NA
VI	32044,17	4658,09	27386,08	0,27	111537,8
VII	26661,66	4670,26	21991,40	0,27	6918,57
VIII	30821,44	4681,78	26139,66	0,27	69242,56
IX	32155,38	4639,46	27515,93	0,27	NA
X	NA	NA	NA	NA	NA
XI	NA	NA	NA	NA	NA
XII	NA	NA	NA	NA	NA

Nota-se na Tabela 4.4 que apenas o modelo *hurdle*-Gama com preditor VII apresentou resultados satisfatórios para explicar os valores dos autos de infração, uma vez que obteve me-

nores valores de WAIC e RMSE do que as demais configurações. Também foram ajustados para as mesmas configurações de preditores, modelos com verossimilhança lognormal e weibull. Porém, não apresentamos tais resultados, pois o desempenho com essas verossimilhanças foram abaixo do desempenho dos modelos com verossimilhança gama.

Assim sendo, selecionamos como modelo final para a modelagem dos valores dos autos de infração ambientais o modelo *hurdle*-Gama com verossimilhança binomial e preditor com a configuração VIII e verossimilhança Gama com preditor VII. Um resumo deste modelo é apresentado na Tabela 4.5

Tabela 4.5 – Média, desvio padrão e intervalos de credibilidade a posteriori (95%) para as covariáveis e hiperparâmetros do modelo final *hurdle*-Gama, com respectivos preditores lineares com a configuração espaço-temporal VIII e VII

Covariáveis	Média	Desvio padrão	$q_{0.025}$	$q_{0.975}$
Intercepto (Binomial)	-2,77	0,11	-2,99	-2,58
Intercepto (Gama)	4,95	0,40	4,15	5,74
IDHM	3,22	0,68	1,89	4,55
T1	1,01	0,39	0,25	1,76
T2	0,89	0,08	0,74	1,04
T3	0,83	0,20	0,44	1,21
T6	0,56	0,13	0,31	0,81
T12	0,35	0,11	0,14	0,55
T18	1,40	0,48	0,45	2,35
T22	-2,79	1,16	-5,08	-0,51
T23	2,44	0,76	0,94	3,94
Intervenções NE	1,01	0,10	0,92	1,29
Hiper(Binomial)	Média	Desvio padrão	$q_{0.025}$	$q_{0.975}$
τ_{bym2}	0,89	0,15	0,61	1,20
ϕ_{bym2}	0,07	0,06	0,01	0,24
τ_{ar1}	0,73	0,21	0,40	1,21
ρ_{ar1}	0,56	0,15	0,20	0,58
τ_{int}	1,31	0,37	0,78	2,21
ρ_{int}	0,80	0,07	0,64	0,90
Hiper(Gama)	Média	Desvio padrão	$q_{0.025}$	$q_{0.975}$
τ_{gama}	190,36	120,30	47,74	501,54
τ_{bym2}	5,02	1,23	2,94	7,74
ϕ_{bym2}	0,66	0,14	0,35	0,89
τ_{ar1}	23,23	17,33	4,42	68,79
ρ_{ar1}	0,10	0,33	-0,54	0,70
τ_{int}	0,95	0,05	0,86	1,05
ρ_{int}	0,05	0,06	-0,06	0,16

Na Tabela 4.5, as covariáveis que foram importantes para explicar o valor de autos de infração relacionadas à verossimilhança gama foram o IDHM, e os tipos de intervenções como

a não especificada, instalação de fornos de carvão (T1), supressão da vegetação (T2), práticas de queimadas (T3), construções de infraestrutura como casas, piscinas..(T6), movimentação de terra (T12), silvicultura (T18), queima de lixo (T22) e impedimento da regeneração natural (T23). Dentre essas variáveis, apenas queima de lixo (T22) possui efeito negativo nos valores agregados dos autos de infração ambientais. As demais variáveis relacionadas ao tipo de intervenção ocorridas em APP, relacionam-se proporcionalmente com os valores agregados de autos de infração, indicando que os municípios que possuem infrações com estes tipos de intervenção, são aqueles que possuem maior montante financeiro devido a essas multas.

Os resultados apontam que os municípios com maiores IDHM são também os que contêm maiores valores agregados referentes as essas infrações ambientais, além disso, o efeito da variável IDHM é o de maior média a posteriori dentre as variáveis explicativas de efeitos fixos, indicando que essa variável, é a que mais possui relevância para explicar o valor agregado de autos ambientais, dentre as variáveis estudadas.

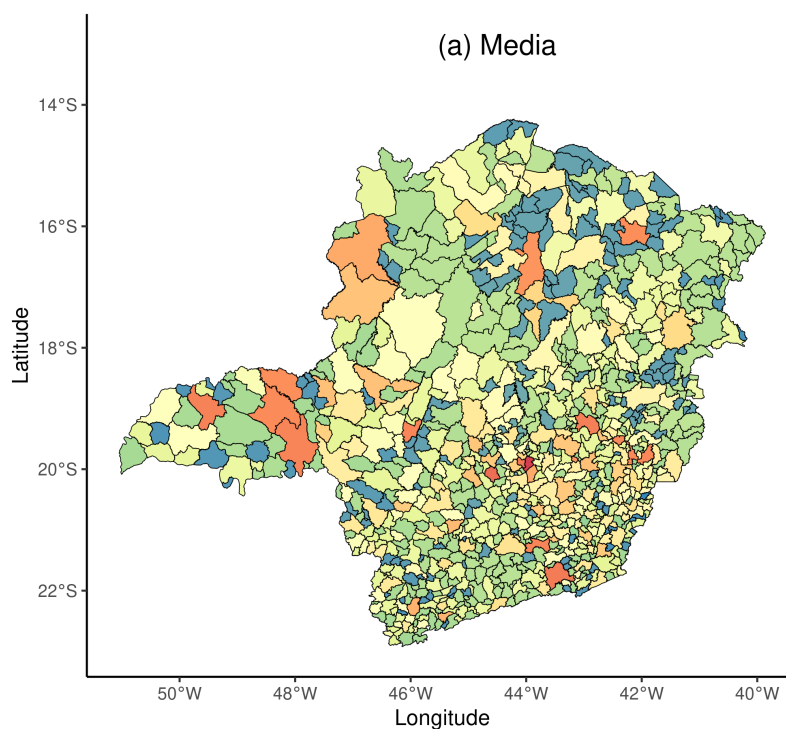
A parte do modelo com respeito a verossimilhança binomial modela a probabilidade de que ocorra ao menos um auto de infração ambiental para um determinado município e determinado ano. Dessa forma, os efeitos principais espaciais de municípios para a parte da verossimilhança binomial indicam quais são os municípios que possuem maiores probabilidades de terem um ou mais autuações ambientais para o período e estão apresentados na Figura 4.9.

Na Figura 4.9 apresenta-se a média a posteriori do efeito espacial de ocorrência de autos em nível municipal. Nos municípios com maiores efeitos espaciais é mais provável que ocorra ao menos uma infração ambiental. Alguns dos municípios mineiros com maiores efeitos espaciais foram Belo Horizonte, Contagem, Itaúna, Ferros, Montes Claros, Nova Lima, Barbacena, João Monlevade, Caratinga, Uberlândia, Ituiutaba, Juiz de Fora, Mariana e entre outros.

Já os efeitos principais espaciais de municípios para a parte da verossimilhança gama, quanto maiores tais efeitos, implicam em municípios com maiores valores de autos de infrações ambientais. O mapa dos efeitos espaciais médios são apresentados na Figura 4.10.

Pode-se observar que na Figura 4.10 os municípios com maiores efeitos aleatórios espaciais se situam na região intermediária de Montes Claros, com destaques para as cidades (em ordem decrescente): São João das Missões, Bonito de Minas, Jaíba, Itacarambi, Pirapora, Januária, Montalvânia, Juvenília, Ponto Chique, Cônego Marinho, Marivânia, Manga, Montes Claros, Varzelândia e entre outros municípios. Isso indica que na região intermediária de Montes Claros ocorreram as infrações ambientais mais dispendiosas e, possivelmente, as infrações

Figura 4.9 – Média a posteriori do efeito aleatório espacial municipal referente ao modelo final *hurdle-gama* para a verossimilhança binomial e preditor linear com a configuração VIII para a ocorrência ou não de autos de infração em Minas Gerais no período de 2008 a 2016



mais graves ao meio ambiente, uma vez que não dispomos da informação do tamanho da área degradada dos autos de infrações ambientais. Os municípios com menores médias a posteriori dos efeitos espaciais se situam na região intermediária de Pouso Alegre, em que os municípios que se destacaram com as menores médias foram (em ordem crescente): Itanhandu, Itamonte, Virgínia, Aiuruoca, Cristina, Marmelópolis, Passa Quatro, Pouso Alto, Dom Viçoso e Pedralva. Também na lista incluem os municípios de Ibituruna e São Gonçalo do Sapucaí situados na região intermediária de Varginha, Astolfo Dutra presente na região intermediária de Juiz de Fora e o município de Barbacena.

Os efeitos principais temporais anuais referente à verossimilhança binomial (à esquerda) e à verossimilhança gama (à direita) para todos os municípios, são apresentados na Figura 4.11.

Figura 4.10 – Média a posteriori do efeito aleatório espacial municipal do modelo final *hurdle*-gama para a verossimilhança gama e preditor linear com a configuração VII para o valor de autos de infração em Minas Gerais no período de 2008 a 2016

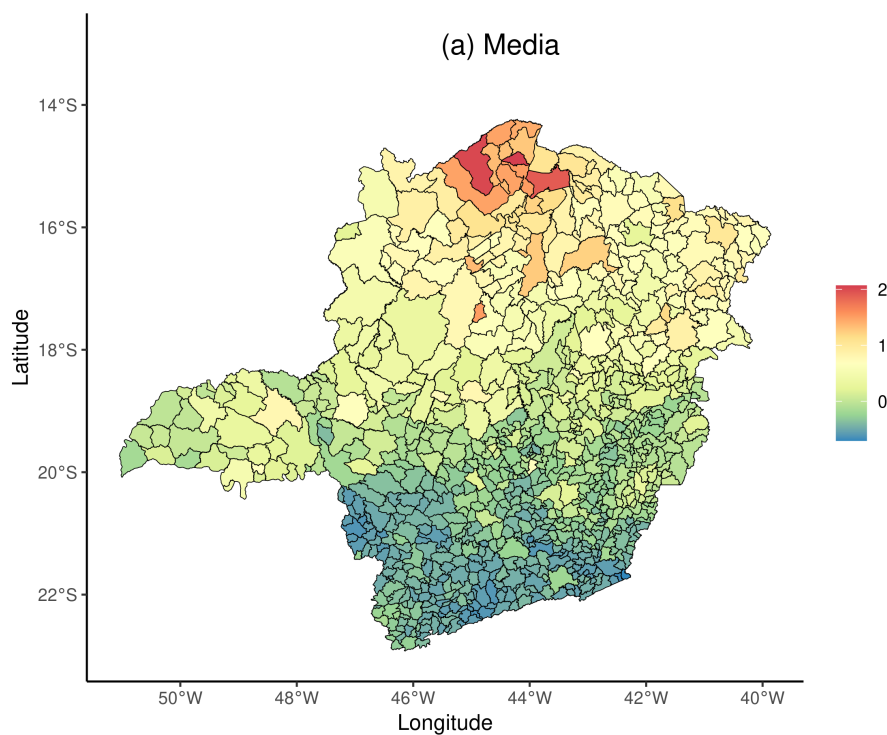
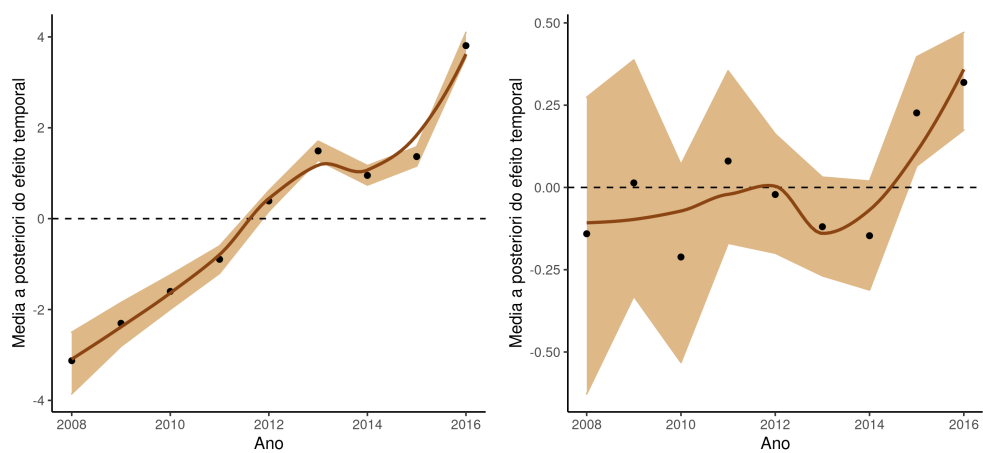


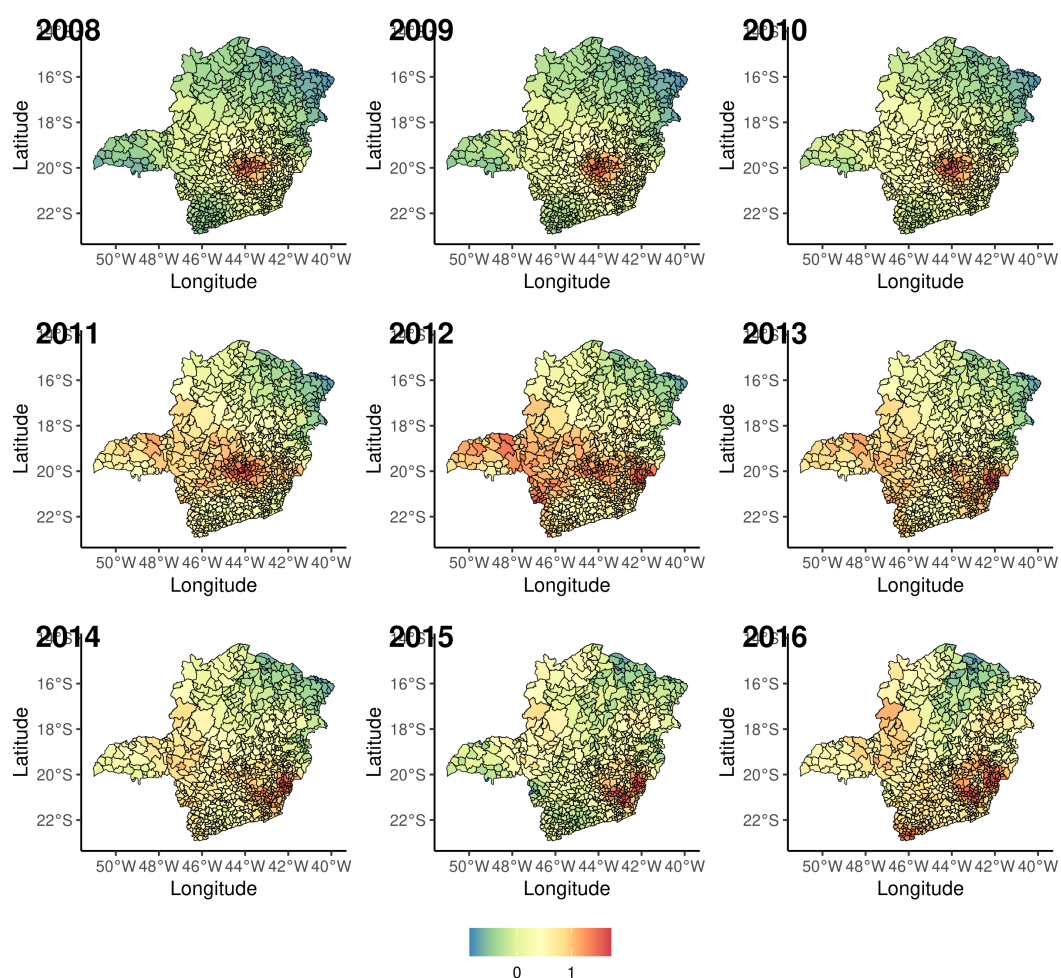
Figura 4.11 – Média a posteriori (pontos), intervalo de credibilidade de 95% (faixas) e ajuste via LOESS (curva) do efeito temporal da ocorrência (esquerda) e valor (direita) de autos de infrações ambientais dos municípios de MG entre os anos de 2008 e 2016



No gráfico à esquerda da Figura 4.11, referente à verossimilhança binomial, verifica-se que existe uma tendência crescente no efeito temporal, indicando que com o decorrer dos anos a probabilidade de que os municípios tenham 1 ou mais ocorrências de autos de infração foi aumentando gradativamente. Já no gráfico à direita da Figura 4.11, referente à verossimilhança gama, indicam que os valores dos autos permaneceram estáveis até o ano de 2014, tendo um crescimento abrupto para os anos de 2015 e 2016.

A média a posteriori do efeito espaço-temporal para a ocorrência e para o valor dos autos de infração dos municípios de Minas Gerais são apresentadas nas Figuras 4.12 e 4.14.

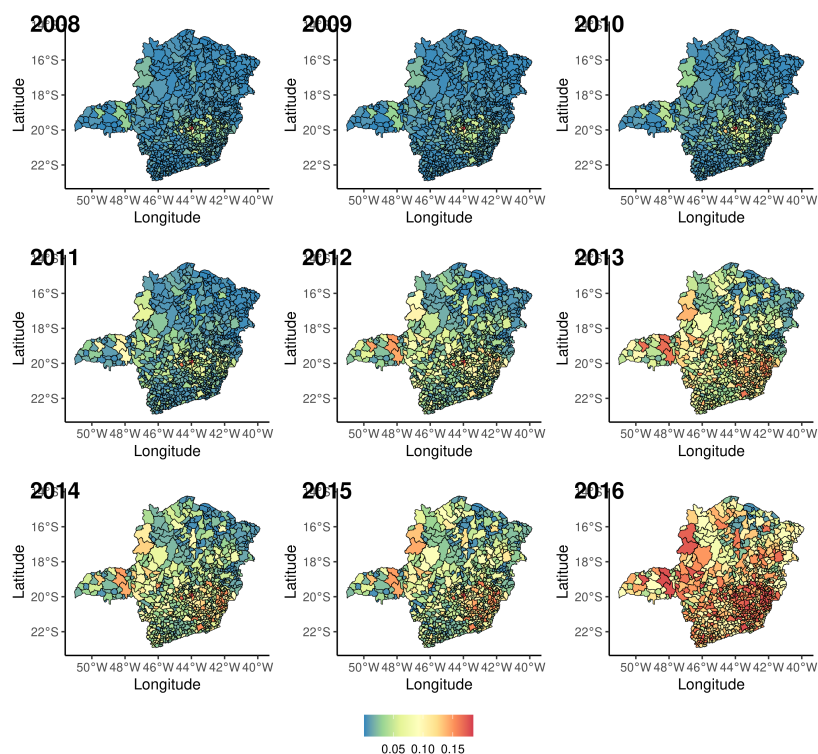
Figura 4.12 – Média a posteriori da interação para a ocorrência de autos



Na Figura 4.12 há, de maneira geral, um padrão de aumento dos valores da média a posteriori da interação para os municípios de Minas Gerais no decorrer dos anos. Aliado ao efeito temporal crescente, esses resultados corroboram com o aumento da probabilidade de ocorrência de autos nos municípios de Minas Gerais no decorrer dos anos. Na Figura 4.13 pode-se observar

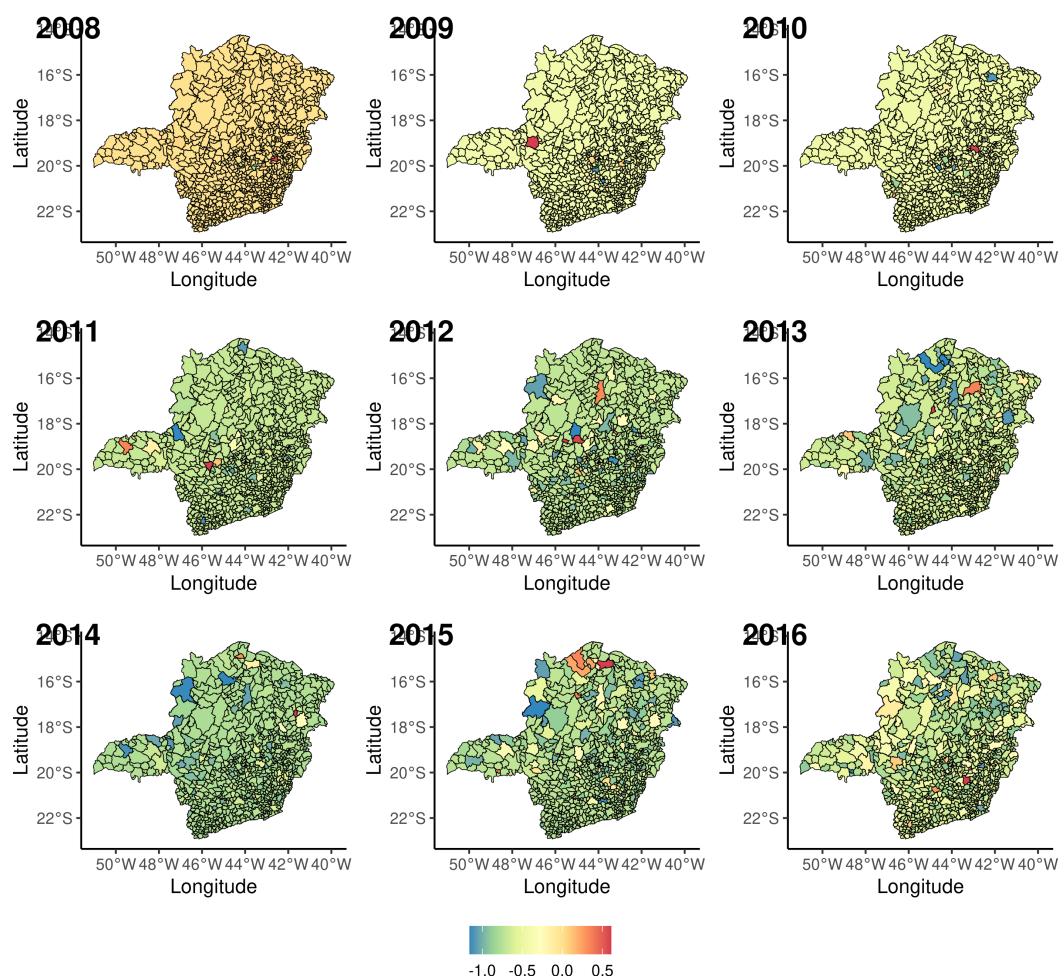
o crescente aumento das probabilidade de ocorrência de autos de infração ao longo dos anos, sobretudo, nos municípios pertencentes às regiões imediatas de Belo Horizonte, Divinópolis, Juiz de Fora, Ipatinga, Governador Valadares, Patos de Minas, Uberaba, Uberlândia.

Figura 4.13 – Valores preditos da probabilidade de ocorrência de autos de infração ambientais do modelo final *hurdle*-Gama com verossimilhança binomial e predictor linear na configuração VIII e verossimilhança gama e predictor linear VII



Tendo em vista que a precisão da interação espaço-tempo para a verossimilhança gama foi a menor das dentre as componentes espaciais, temporais e espaço-temporais, temos, portanto, que a maior variabilidade dos valores dos autos de infração é captada pela componente de interação espaço-temporal. Dada que ela é assumida ser do tipo II, indica que cada município possui sua estrutura temporal modelada pelo modelo AR1 sem que haja estrutura espacial entre os municípios. No entanto, uma vez que o hiperparâmetro de correlação da componente de interação tem alta probabilidade na região próxima do zero, significando que o comportamento anual dos municípios também oscilou de forma independente. Isto indica que a componente de interação não capta um efeito global e sim anomalias locais indicando alguns poucos municípios em que o modelo não foi capaz de capturar toda a variabilidade utilizando apenas os efeitos globais espaciais e temporais.

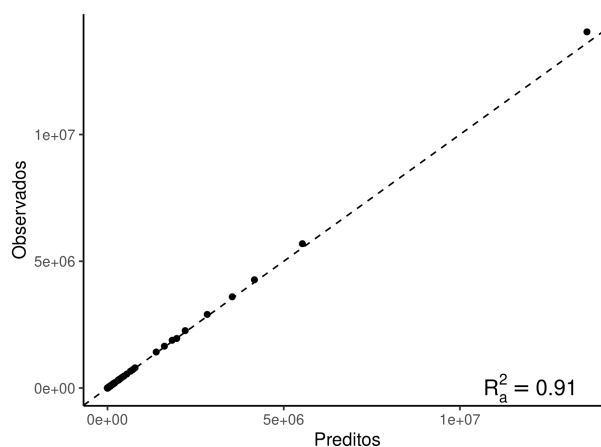
Figura 4.14 – Média a posteriori da interação para o valor de autos



Como destaque nos mapas dos valores de autos de infração corrigidos pela interação, observa-se que média a posteriori do efeito de interação espaço-temporal para o município de Jaíba é muito elevada para os anos de 2014 e 2015. Nestes anos ocorreram no município um auto de infração em cada ano, sendo que cada uma das infrações tiveram multas no valor de R\$ 137.473,30 e R\$ 14.048.645,25, respectivamente. Ainda em 2014, Itaipé, município pertencente a região imediata de Teófilo Otoni e São João das Missões, município pertencente à região intermediária de Montes Claros, tiveram infrações ambientais que somadas resultaram em multas nos valores de R\$ 1.420.305,62 e R\$ 797.141,80. Para o ano de 2015, além do ocorrido no município de Jaíba, houveram outros únicos autos nas cidades de Ponto Chique, Pirajuba, Bonito de Minas, Januária também da ordem de milhões de reais. Em 2016, uma das maiores médias a posteriori captadas pela interação ocorreu no município de Mariana em que houve o desastroso rompimento das barragens. Para o ano em questão, foram 22 infrações ambientais cujo montante do valor das multas é de R\$ 5.692.878,93.

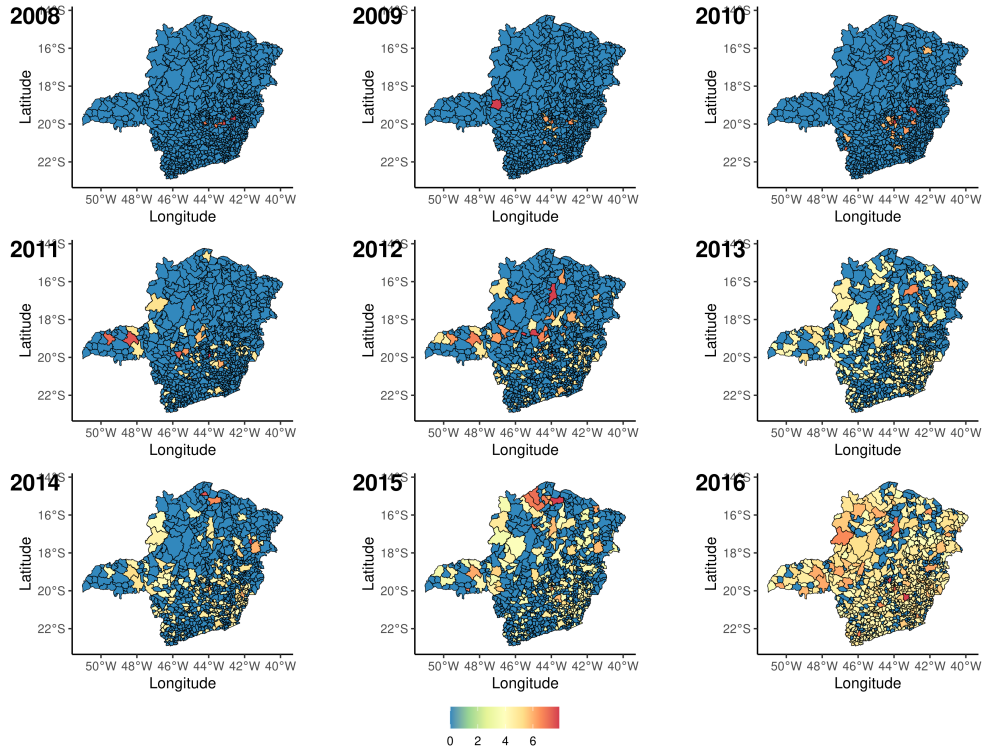
Na Figura 4.15, é apresentado o gráfico de dispersão entre valores preditos versus observados referentes ao valor dos autos de infração ambientais ocorridos em Minas Gerais, cujo coeficiente de determinação ajustado (R_a^2) foi igual a 91%.

Figura 4.15 – Preditos *versus* observados para o valor dos autos de infrações ambientais ocorridos nos municípios de MG entre os anos de 2008 e 2016



Na Figura 4.16 pode-se observar que o modelo prediz o crescente aumento dos valores agregados dos autos de infração ocorridos nos municípios ao longo dos anos, corroborando com os valores observados apresentados na Figura 4.8.

Figura 4.16 – Valores preditos dos valores agregados dos autos de infração ambientais do modelo final *hurdle*-Gama com verossimilhança binomial e preditor linear na configuração VIII e verossimilhança gama e preditor linear VII (escala logarítmica)



4.3 Modelagem bayesiana conjunta para ocorrência, número e valor de autos

Uma vez selecionados separadamente os modelos para número e valor dos autos de infração ambientais em Minas Gerais, iremos apresentar os resultados da modelagem conjunta para ambas variáveis.

A Tabela 4.6 apresenta 6 distintos modelos de acordo com as configurações para os preditores para ocorrência (η_1), número (η_2) e valor (η_3). As componentes $f_1(s)$ e $f_2(t)$ são modeladas como modelo BYM2 e modelo AR1, respectivamente, enquanto que a componente $f_3(s, t)$ representa interações espaço-temporais do tipo IV, isto é, segue distribuição $\mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_{BESAG} \otimes \mathbf{\Sigma}_{AR1})$. A presença de tais componentes em mais de um preditor do mesmo modelo, representa que essas componentes foram compartilhadas naquele modelo. Enquanto que, as componentes $f_{TipoI}(s, t)$ e $f_{TipoII}(s, t)$ não serão compartilhadas entre preditores, sendo específicas, apenas a verossimilhança do seu respectivo preditor. As componentes $f_{TipoI}(s, t)$ e $f_{TipoII}(s, t)$ correspondem a interação espaço-temporal do tipo I, isto é, $f_{TipoI}(s, t) \sim \mathcal{N}(\mathbf{0}, \mathbf{I} \otimes \mathbf{I})$ e do tipo II, isto é, segue distribuição $\mathcal{N}(\mathbf{0}, \mathbf{I} \otimes \mathbf{\Sigma}_{AR1})$, respectivamente.

Tabela 4.6 – Modelos considerados para a modelagem conjunta, de ocorrência, número e valor de autos de infrações em Minas Gerais para os anos de 2008 a 2016.

	Modelo	WAIC
(I)	$\eta_1 = \beta_{01} + f_1(s) + f_2(t) + f_3(s, t)$ $\eta_2 = \beta_{02} + cov + \lambda_{21}f_1(s) + \lambda_{22}f_2(t) + \lambda_{23}f_3(s, t)$ $\eta_3 = \beta_{03} + cov + \lambda_{31}f_1(s) + \lambda_{32}f_2(t) + \lambda_{33}f_3(s, t)$	38696,71
(II)	$\eta_1 = \beta_{01} + f_1(s) + f_2(t) + f_3(s, t)$ $\eta_2 = \beta_{02} + cov + \lambda_{21}f_1(s) + \lambda_{22}f_2(t) + f_{tipoII}(s, t)$ $\eta_3 = \beta_{03} + cov + \lambda_{31}f_1(s) + \lambda_{32}f_2(t) + f_{tipoII}(s, t)$	31698,14
(III)	$\eta_1 = \beta_{01} + f_1(s) + f_2(t) + f_3(s, t)$ $\eta_2 = \beta_{02} + cov + \lambda_{21}f_1(s) + \lambda_{22}f_2(t) + f_{tipoI}(s, t)$ $\eta_3 = \beta_{03} + cov + \lambda_{31}f_1(s) + \lambda_{32}f_2(t) + f_{tipoII}(s, t)$	NA
(IV)	$\eta_1 = \beta_{01} + f_1(s) + f_2(t) + f_3(s, t)$ $\eta_2 = \beta_{02} + cov + \lambda_{21}f_1(s) + \lambda_{22}f_2(t) + f_{tipoI}(s, t)$ $\eta_3 = \beta_{03} + cov + \lambda_{31}f_1(s) + \lambda_{32}f_2(t) + f_{tipoI}(s, t)$	38682,74
(V)	$\eta_1 = \beta_{01} + f_1(s) + f_2(t)$ $\eta_2 = \beta_{02} + cov + \lambda_{21}f_1(s) + \lambda_{22}f_2(t) + f_{tipoII}(s, t)$ $\eta_3 = \beta_{03} + cov + \lambda_{31}f_1(s) + \lambda_{32}f_2(t) + f_{tipoII}(s, t)$	31936,26
(VI)	$\eta_1 = \beta_{01} + f_1(s) + f_2(t) + f_3(s, t)$ $\eta_2 = \beta_{02} + cov + \lambda_{21}f_1(s) + f_{iid}(t) + f_{tipoI}(s, t)$ $\eta_3 = \beta_{03} + cov + \lambda_{31}f_1(s) + \lambda_{32}f_2(t) + f_{tipoII}(s, t)$	31973,11

Dentre os modelos ajustados, o modelo conjunto com os preditores na configuração (II) foi o que apresentou o menor valor de WAIC. O modelo (II) possui as componentes espaciais (BYM2) e temporais (AR1) compartilhadas e escaladas pelos parâmetros λ_{ij} . Adicionalmente, cada preditor do modelo possuem suas próprias covariáveis (caso dos preditores para as respostas número e valor de autos) e interações espaço-temporais, sendo do Tipo IV para a resposta ocorrência e do Tipo II para as respostas número e valor. Portanto, é selecionado o modelo (II) como modelo final conjunto para as respostas ocorrência, número e valor de autos de infração ambientais.

Pela Tabela 4.7, temos a comparação dos WAIC, RMSE e *LogScore* para os modelos finais univariado Poisson, *hurdle*-Gama e conjunto. Pode-se observar que a modelagem conjunta melhora o ajuste referente às verossimilhanças Poisson e Gama, uma vez que produziu menor WAIC se comparada ao WAIC modelo final poisson e WAIC referente a verossimilhança gama no modelo hurdle-gama. Pelo LogScore, a modelagem conjunta melhora a capacidade preditiva do valor dos autos de infração. Também houve maior qualidade de ajuste pelo aumento do coeficiente de determinação ajustado $R_a^2 = 0,92$ para o preditor da resposta gama no modelo conjunto.

Tabela 4.7 – Comparação das medidas WAIC, RMSE e LogScore entre os modelos finais Poisson, *hurdle*-Gama e conjunto

Verossimilhança	WAIC		RMSE		LogScore	
	Conjunta	Univariada	Conjunta	Univariada	Conjunta	Univariada
Binomial	4756,81	4670,26	0,28	0,27	0,31	0,31
Poisson	5041,06	5885,10	0,62	0,54	1,78	0,40
Gama	21900,26	21991,40	7429,61	6918,57	16,45	16,64

As médias, desvio padrões, limites dos intervalos de 95% credibilidade das marginais a posteriori dos efeitos fixos e dos hiperparâmetros do modelo conjunto final (II) são apresentadas na Tabela 4.8. Cada preditor possui suas próprias covariáveis, sendo que para o preditor referente ao número de autos, observa-se que municípios com maiores IDHM se associam com maiores números de autos de infração e, também, é a covariável que maior influencia o número de autos de infrações ambientais. No mesmo sentido as covariáveis T2, T3, T6, T7, T12, T15, T19 e não específicas que representam os tipos de intervenção ocorridas em áreas de preservação permanente se associam a maiores números de autos de infração. O mesmo vale para as covariáveis do tipo de APP cursos d'água e não especificadas.

Para a resposta valor de autos de infração, a covariável IDHM foi também uma das mais relevantes e teve efeito médio positivo, sugerindo que municípios com maiores IDHM se associam a maiores valores de autos de infração. Dentre os tipos de intervenção, destacam-se a covariável T23 impedimento da regeneração natural foi a mais relevante para explicar autos valores de infração ambiental, enquanto que a covariável T22 queima de lixos teve efeito médio negativo e está associada a menores valores de autos de infração.

Quanto aos hiperparâmetros, foram obtidas baixas médias a posteriori para as precisões espaciais e temporais, sugerindo grandes variabilidades espaciais, temporais e de interações, sobretudo, para os efeitos de tempo e de interações. O parâmetro de mistura ϕ do modelo BYM2 apresentou média a posteriori de 0,29, o que representa que 29% da variância espacial é explicada pela componente estruturada, enquanto que restante é devido a superdispersão. O parâmetro de correlação do efeito temporal obteve média a posteriori de 0,54, o que indica que o efeito temporal de um ano se correlaciona em 0,54 com seu ano anterior. Os parâmetros λ_{ij} servem para reescalar os efeitos espaciais e temporais globais. Sendo assim, para o preditor com verossimilhança Poisson, a média do efeito espacial para os municípios são, em média, 1,26 vezes o efeito espacial do preditor com verossimilhança binomial, no mesmo sentido, a média do efeito espacial do preditor com verossimilhança gama são, em média, 2,44 vezes o

Tabela 4.8 – Média, desvio padrão e intervalos de credibilidade a posteriori (95%) para as covariáveis e hiperparâmetros do modelo final conjunto para ocorrência, número e valor de autos ambientais (VI)

Modelo	Covariáveis	Média	Desvio padrão	$q_{0.025}$	$q_{0.975}$
Binomial	Intercepto	-2,65	0,09	-2,83	-2,49
	Intercepto	-2,94	0,28	-3,48	-2,39
Poisson	IDHM	3,25	0,45	2,35	4,13
	T2	0,51	0,05	0,41	0,61
	T3	0,43	0,11	0,22	0,64
	T6	0,30	0,07	0,16	0,44
	T7	0,33	0,11	0,12	0,54
	T12	0,33	0,07	0,20	0,46
	T15	0,27	0,12	0,03	0,51
	T19	0,30	0,14	0,02	0,57
	APP NE	0,29	0,11	0,06	0,51
	Cursos d'água	0,23	0,06	0,11	0,35
	Intervenção NE	0,86	0,09	0,67	1,04
Gama	Intercepto	7,57	0,15	7,27	7,87
	IDHM	2,12	0,74	0,65	3,54
	T1	1,00	0,38	0,27	1,75
	T2	0,81	0,07	0,65	3,54
	T3	0,81	0,19	0,44	1,19
	T6	0,59	0,12	0,35	0,83
	T15	0,55	0,22	0,12	0,97
	T18	1,40	0,47	0,49	2,31
	T22	-2,75	1,10	-4,91	-0,59
	T23	2,18	0,73	0,76	3,61
	Intervenção NE	0,90	0,09	0,73	1,07
Geral	Hiperparâmetros	Média	Desvio padrão	$q_{0.025}$	$q_{0.975}$
	τ_{gama}	199,77	133,79	49,16	550,49
	τ_{bym2}	16,37	4,86	8,01	26,73
	ϕ_{bym2}	0,29	0,10	0,15	0,52
	τ_{ar1}	0,76	0,22	0,42	1,27
	ρ_{ar1}	0,54	0,16	0,19	0,80
	λ_{21}	1,26	0,21	0,80	1,63
	λ_{31}	2,44	0,40	1,58	3,12
	λ_{22}	0,18	0,02	0,15	0,21
	λ_{32}	0,20	0,02	0,16	0,23
	$\tau_{int-ocor}$	0,46	0,07	0,34	0,61
	$\rho_{int-ocor}$	0,95	0,02	0,91	0,97
	$\tau_{int-num}$	9,41	1,37	7,24	12,56
	$\rho_{int-num}$	-0,06	0,15	-0,36	0,21
	$\tau_{int-val}$	1,10	0,04	1,01	1,18
	$\rho_{int-val}$	-0,04	0,06	-0,13	0,08

efeito espacial do preditor binomial. Para os efeitos temporais anuais, o raciocínio é semelhante, as médias a posteriori dos efeitos anuais para o preditor com resposta Poisson são, em média,

0,18 vezes as médias a posteriori do efeito anual para verossimilhança binomial, enquanto que o efeito temporal para a verossimilhança gama corresponde, em média, a 0,20 vezes o efeito temporal da verossimilhança binomial.

Na Figura 4.17, são apresentadas as médias a posteriori dos efeitos espaciais municipais para os autos de infração nos municípios de Minas Gerais no período de 2008 a 2016. Os municípios com maiores médias a posteriori para os efeitos espaciais, possuem maiores probabilidades de ocorrências de infrações ambientais, maiores números de autos de infrações ambientais e maiores valores agregados de autos de infrações ambientais, uma vez que as médias a posteriori dos parâmetros λ_{21} e λ_{31} são positivas.

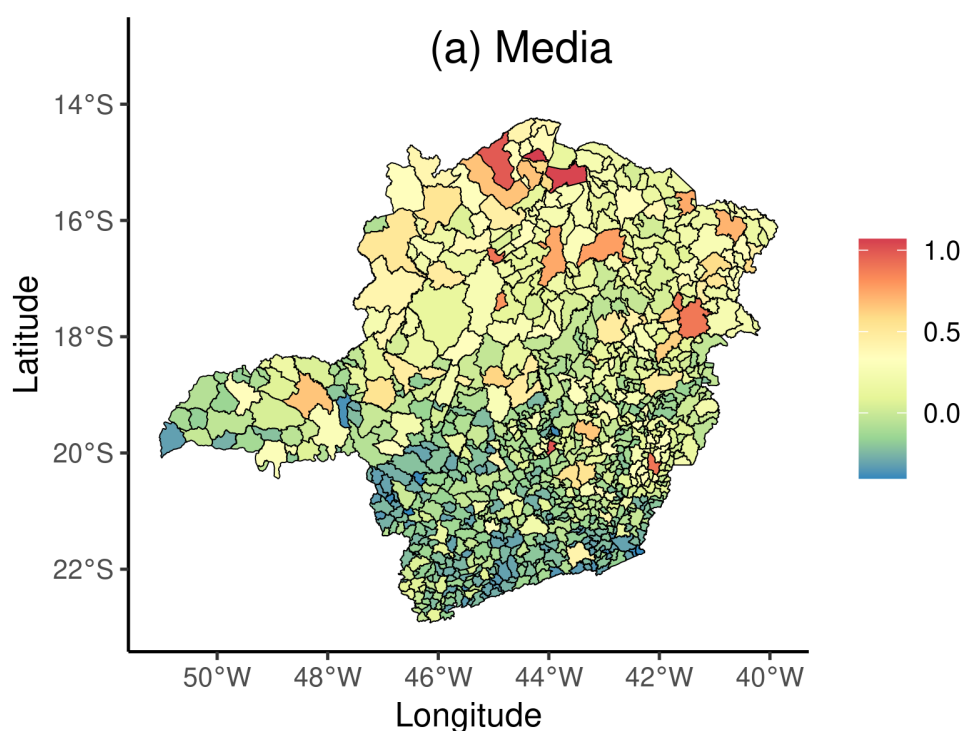
Dentre os municípios com maiores efeitos espaciais destacaram-se os municípios da região intermediária de Belo Horizonte - Belo Horizonte, Contagem, Betim, Mariana, Sete Lagoas e Nova Lima -, da região intermediária de Montes Claros - Montes Claros, Pirapora, Jaíba -, região intermediária de Patos de Minas - Patos de Minas e Patrocínio -, região intermediária de Juiz de Fora - Juiz de Fora, Abre Campo, Matipó, Ubá, Caputira, Sete Lagoas e Nova Lima -, região intermediária de Uberlândia - Araguari, Ituiutaba, Uberlândia, Uberaba -, região intermediária de Ipatinga - João Monlevade e Raul Soares -, região intermediária de Divinópolis - Divinópolis, Itaúna, Nova Serrana -, região intermediária de Barbacena - Barbacena e Piranga e Campo Belo, pertencente a região intermediária de Varginha.

Na Figura 4.18 são apresentados resumos dos efeitos principais temporais anuais, bem como, um ajuste de regressão local (lowess) para as médias a posteriori do efeito temporal em função do ano. Claramente, há uma tendência crescente dos efeitos temporais ao longo dos anos e, uma vez que as médias a posteriori dos parâmetros λ_{22} e λ_{32} são positivas, representa, que, de modo geral, as probabilidades de ocorrência de autos de infração ambientais, os números de autos de infração e o valor agregados anual dos autos de infração aumentaram ao longo dos anos.

A média a posteriori dos efeitos de interação espaço-temporal para a ocorrência, número dos autos de infração e valor dos municípios de Minas Gerais são, similares aos efeitos de interação para os seus respectivos modelos: Poisson univariado e *hurdle*-Gama. Os mapas dessas interações são apresentadas no Apêndice B, Figuras 6, 7 e 8.

De modo equivalente, o mesmo padrão de semelhança se repete nos valores preditos para as respostas ocorrência, número e valor de autos. Os gráficos desses valores preditos também se encontram no Apêndice B nas Figuras 11, 12 e 13.

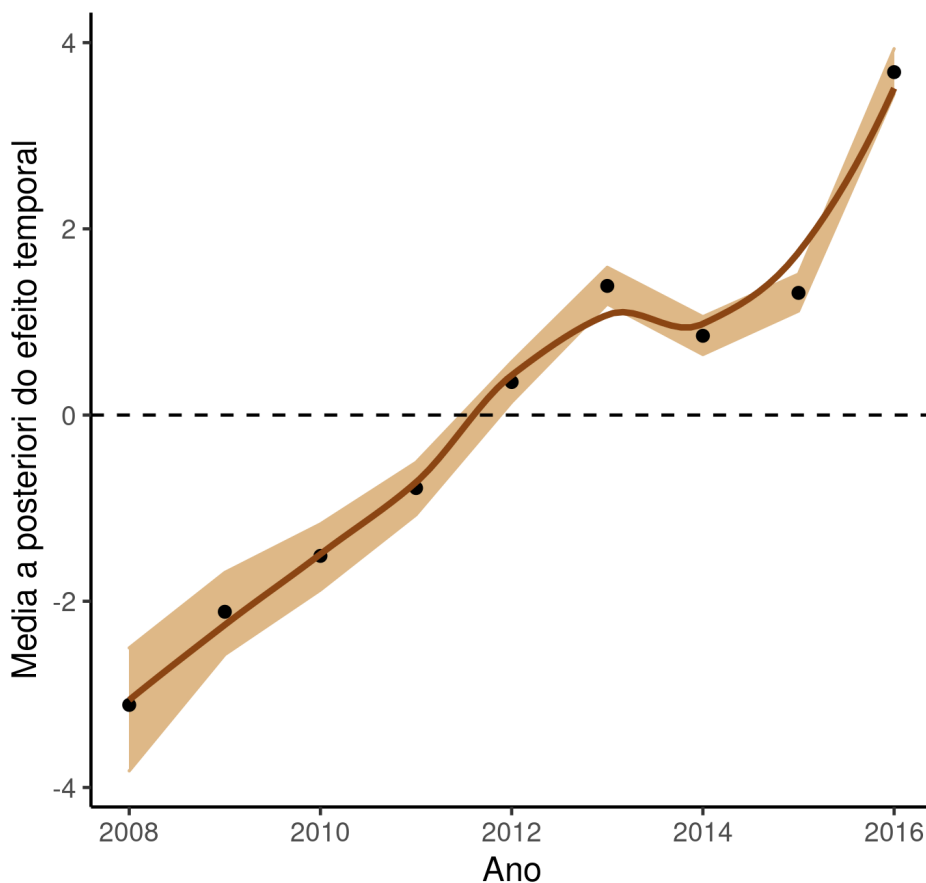
Figura 4.17 – Média a posteriori do efeito aleatório espacial municipais do modelo final conjunto para ocorrência de autos de infração ambientais em Minas Gerais no período de 2008 a 2016 – efeito espacial para número e valor possuem mapas similares, sendo escalados por λ_{21} e λ_{31} , respectivamente



De modo geral, o modelo conjunto obteve melhor qualidade do ajuste e capacidade preditiva, obtendo estimativas com maiores precisões, sobretudo, para os número e valores dos autos de infrações ambientais.

Ademais, efeitos principais temporais combinados foram estimados com maior precisão comparados aos modelos individuais para o número e valor, o mesmo ocorreu para os efeitos espaciais combinando as três verossimilhanças, que além de serem estimados com maior precisão, foi capaz de distinguir os municípios com maiores probabilidades de ocorrência de autos,

Figura 4.18 – Média a posteriori (pontos), intervalo de credibilidade de 95% (faixas) e ajuste via LOESS (curva) do efeito temporal do modelo conjunto para os autos de infrações ambientais dos municípios de MG entre os anos de 2008 e 2016 – efeito temporal para número e valor possuem mapas similares, sendo escalados por λ_{22} e λ_{32} , respectivamente



maiores números e valores das autuações ambientais, incorporando características observadas tanto para o modelo final para o número, quanto para o modelo final para o valor dos autos de infração ambientais.

Além da modelagem conjunta, utilizando verossimilhanças binomial, Poisson e gama. Outro ponto de destaque se dá pela inédita utilização de priors de complexidade penalizada, para modelar os hiperparâmetros dos efeitos espaciais, temporais e espaço-temporais Nos modelos de autos de infrações ambientais. As vantagens são diversas. Para o modelo BYM2, a priori PC é transferível mediante a realização da escala da matriz de precisão, isto é, após a matriz de precisão ser escalada pelo o inverso da sua variância generalizada, fixada a priori, o parâmetro de precisão fornece uma mesma quantidade de suavização se há mudança no grafo. As interpretações dos parâmetros de precisão e mistura são realizadas de forma simples. Por fim, ao se definir priors PC para a precisão, tais priors tem a capacidade de encolherem para

modelos mais simples. Enquanto que a definição de prioris gama tendem a ter um superajuste para modelos mais flexíveis, mesmo quando o modelo base é o modelo correto.

Por exemplo, no modelo conjunto o parâmetro de correlação as interações espaço-temporais para os preditores cuja resposta são o número e valor reduziram à priori PC, cujo modelo base é $\rho = 0$.

Os resultados claramente indicam um aumento do número e dos valores dos autos de infrações ambientais no período entre os anos de 2008 e 2016 nos municípios de Minas Gerais, bem como, detectou municípios em que o número e valores destes autos de infrações ocorreram, destacando-se, principalmente, os municípios das regiões intermediária de Montes Claros e Belo Horizonte. Estes resultados vão ao encontro com a análise feita por Araujo e Nogueira (2020), em que por meio de revisão de referências bibliográficas e documentos oficiais brasileiros, concluíram que as multas ambientais tem sido um mecanismo ineficaz para o controle e a repressão de ações lesivas ao meio ambiente.

O tempo computacional para os modelos foram satisfatórios e permitiram comparar diversas estruturas de modelos. O tempo para os ajustes dos modelos finais foram 635 segundos para o modelo Poisson para o número de autos de infrações ambientais, 473 segundos para o modelo final *hurdle*-gama para o valor dos autos de infrações ambientais e 7435 segundos para o modelo final conjunto. Houve uma tentativa de ajuste de um modelo univariado Poisson para o número de autos de infração via MCMC, através do pacote *brms* do programa R, que é um pacote baseado na plataforma Stan de modelagem e computação estatística. No entanto, o tempo para realização de um passo do algoritmo foi de, aproximadamente, 20h, o que torna o ajuste desse modelo inviável para a maioria das situações rotineiras. Por fim, todos os ajustes foram realizados utilizando um computador com 24 núcleos.

5 CONCLUSÃO

Neste trabalho ajustamos modelos bayesianos espaço-temporais conjuntos, para analisar o número e valor agregados dos autos de infração ambiental ocorridos nos municípios de Minas Gerais entre 2008 e 2016, combinando as distribuições binomial, Poisson e gama na montagem das verossimilhanças. Embora os ajustes individuais destas distribuições algumas vezes fossem bons, a análise conjunta melhorou a descrição, em especial para a variável contínua (valor dos autos).

A proposição de prioris de complexidade penalizada para os parâmetros de precisão, mistura e correlação para modelos espaciais, temporais e espaço-temporais auxiliaram nas convergências dos modelos e evitou que os mesmos apresentassem *overfit*. Sendo, sua aplicação, inédita na área de Direito Ambiental.

Para a obtenção de distribuições marginais a posteriori, o uso do INLA permitiu que análises conjuntas complexas fossem realizadas em horas (ao invés de dias) de computação, utilizando uma máquina com 24 núcleos. Utilizando MCMC, o problema seria resolvido em meses para cada modelo considerado, com o mesmo equipamento.

Recomenda-se fortemente o uso do INLA se formos tentar ajustar modelos conjuntos de respostas com diferentes distribuições que possam ser enquadradas em campos latentes gaussianos. O modelo utilizado neste estudo pode ser adaptado para diversas situações semelhantes, como, por exemplo, em ecologia e epidemiologia.

Por fim, a tendência de crescimento dos autos de infrações ambientais ao longo do tempo, tanto em relação ao número, quanto ao seu valor agregado, sugerem que a aplicação dessas autuações não têm sido eficazes no controle e repressão de ações lesivas ao meio ambiente.

REFERÊNCIAS

- ARAUJO, R. C. de; NOGUEIRA, J. M. **A ineficácia da multa ambiental em Direito dos Desastres: meio ambiente natural, cultural e artificial**. [S.l.]: UniCEUB:ICPD, 2020. ISBN 978-65-990501-9-0.
- ASMARIAN, N. et al. Bayesian spatial joint model for disease mapping of zero-inflated data with r-inla: A simulation study and an application to male breast cancer in iran. **International journal of environmental research and public health**, Multidisciplinary Digital Publishing Institute, v. 16, n. 22, p. 4460, 2019.
- BAKKA, H. et al. Spatial modelling with r-inla: A review. **arXiv preprint arXiv:1802.06350**, 2018.
- BERNARDINELLI, L.; CLAYTON, D.; MONTOMOLI, C. Bayesian estimates of disease maps: how important are priors? **Statistics in medicine**, Wiley Online Library, v. 14, n. 21-22, p. 2411–2431, 1995.
- BESAG, J.; YORK, J.; MOLLIÉ, A. Bayesian image restoration, with two applications in spatial statistics. **Annals of the institute of statistical mathematics**, Springer, v. 43, n. 1, p. 1–20, 1991.
- BIVAND, R.; HAUKE, J.; KOSSOWSKI, T. Computing the jacobian in gaussian spatial autoregressive models: An illustrated comparison of available methods. **Geographical Analysis**, v. 45, n. 2, p. 150–179, 2013. Disponível em: <<http://www.jstatsoft.org/v63/i18/>>.
- BIVAND, R.; LEWIN-KOH, N. **maptools: Tools for Reading and Handling Spatial Objects**. [S.l.], 2017. R package version 0.9-2. Disponível em: <<https://CRAN.R-project.org/package=maptools>>.
- BIVAND, R. S.; PEBESMA, E.; GOMEZ-RUBIO, V. **Applied spatial data analysis with R, Second edition**. Springer, NY, 2013. Disponível em: <<http://www.asdar-book.org/>>.
- BLANGIARDO, M.; CAMELETTI, M. **Spatial and spatio-temporal Bayesian models with R-INLA**. [S.l.]: John Wiley & Sons, 2015.
- BLANGIARDO, M. et al. Spatial and spatio-temporal models with r-inla. **Spatial and spatio-temporal epidemiology**, Elsevier, v. 4, p. 33–49, 2013.
- BORGES, L. A. C.; REZENDE, J. L. P. Áreas protegidas no interior de propriedades rurais: a questão das app e rl. **Floresta e Ambiente**, Instituto de Florestas/UFRRJ, v. 18, n. 2, p. 210–222, 2012.
- BRESLOW, N.; LEROUX, B.; PLATT, R. Approximate hierarchical modelling of discrete data in epidemiology. **Statistical Methods in Medical Research**, Sage Publications Sage CA: Thousand Oaks, CA, v. 7, n. 1, p. 49–62, 1998.
- BRIER, G. W. Verification of forecasts expressed in terms of probability. **Monthly weather review**, v. 78, n. 1, p. 1–3, 1950.
- BROCKWELL, P. J.; DAVIS, R. A. **Time Series: Theory and Methods, Second Edition (Springer Series in Statistics)**. 2nd ed. 1991. 2nd printing. ed. Springer, 2009. ISBN 1441903194,9781441903198. Disponível em: <<http://gen.lib.rus.ec/book/index.php?md5=a7050c98e54f74341373675d192a22fe>>.

CLARK, J. S.; GELFAND, A. E. **Hierarchical modelling for the environmental sciences: statistical methods and applications**. [S.l.]: Oxford University Press on Demand, 2006.

CLAYTON, D. G. Generalized linear mixed models. **Markov chain Monte Carlo in practice**, v. 1, p. 275–302, 1996.

COSANDEY-GODIN, A. et al. Applying bayesian spatiotemporal models to fisheries bycatch in the canadian arctic. **Canadian Journal of Fisheries and Aquatic Sciences**, v. 72, n. 2, p. 186–197, 2015. Disponível em: <<https://doi.org/10.1139/cjfas-2014-0159>>.

CRESSIE, N. A. C. **Statistics for spatial data**. [S.l.]: Wiley Online Library, 1993.

DEAN, C.; UGARTE, M.; MILITINO, A. Detecting interaction between random region and fixed age effects in disease mapping. **Biometrics**, Wiley Online Library, v. 57, n. 1, p. 197–202, 2001.

DUNSON, D. B. Commentary: practical advantages of bayesian analysis of epidemiologic data. **American journal of Epidemiology**, Oxford University Press, v. 153, n. 12, p. 1222–1226, 2001.

FERREIRA, D. F. **Estatística multivariada**. [S.l.]: Editora UFLA, 2011.

FORLANI, C. et al. A joint bayesian space–time model to integrate spatially misaligned air pollution data in r-inla. **Environmetrics**, Wiley Online Library, p. e2644, 2020.

FUGLSTAD, G.-A. et al. Constructing priors that penalize the complexity of gaussian random fields. **Journal of the American Statistical Association**, Taylor & Francis, v. 114, n. 525, p. 445–452, 2019.

GERBER, F.; FURRER, R. et al. Pitfalls in the implementation of bayesian hierarchical modeling of areal count data: An illustration using bym and leroux models. **Journal of Statistical Software, Code Snippets**, v. 63, n. 1, p. 1–32, 2015.

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.

GOICOA, T. et al. In spatio-temporal disease mapping models, identifiability constraints affect pql and inla results. **Stochastic Environmental Research and Risk Assessment**, Springer, v. 32, n. 3, p. 749–770, 2018.

HODGES, J. S. **Richly parameterized linear models: additive, time series, and spatial models using random effects**. [S.l.]: CRC Press, 2013.

KIFLE, Y. W.; HENS, N.; FAES, C. Cross-covariance functions for additive and coupled joint spatiotemporal spde models in r-inla. **Environmental and Ecological Statistics**, Springer, v. 24, n. 4, p. 551–586, 2017.

KNORR-HELD, L. Bayesian modelling of inseparable space-time variation in disease risk. **Statistics in medicine**, Wiley Online Library, v. 19, n. 17-18, p. 2555–2567, 2000.

KRAINSKI, E. T. et al. **Advanced spatial modeling with stochastic partial differential equations using R and INLA**. [S.l.]: CRC Press, 2018.

LAWSON, A. B. **Bayesian Disease Mapping : Hierarchical Modeling in Spatial Epidemiology, Third Edition**. Third edition. CRC Press, 2018. (Interdisciplinary statistics). ISBN 9781351271745,1351271741,9781351271752,135127175X,9781351271769,1351271768.

Disponível em: <<http://gen.lib.rus.ec/book/index.php?md5=569de1ce0f221af1a7fee025cf9b1191>>.

LEROUX, B. G.; LEI, X.; BRESLOW, N. Estimation of disease rates in small areas: a new mixed model for spatial dependence. **Statistical models in epidemiology, the environment, and clinical trials**, Springer, p. 179–191, 2000.

LINDGREN, F.; RUE, H.; LINDSTRÖM, J. An explicit link between gaussian fields and gaussian markov random fields: the stochastic partial differential equation approach. **Journal of the Royal Statistical Society: Series B (Statistical Methodology)**, Wiley Online Library, v. 73, n. 4, p. 423–498, 2011.

MACNAB, Y. C. On gaussian markov random fields and bayesian disease mapping. **Statistical Methods in Medical Research**, v. 20, n. 1, p. 49–68, 2011. PMID: 20547586. Disponível em: <<https://doi.org/10.1177/0962280210371561>>.

NTIRAMPEBA, D.; NEEMA, I.; KAZEMBE, L. Joint spatial modelling of disease risk using multiple sources: an application on hiv prevalence from antenatal sentinel and demographic and health surveys in namibia. **Global health research and policy**, BioMed Central, v. 2, n. 1, p. 22, 2017.

ODUM, E. **Fundamentos de ecología**. [S.l.]: Lisboa. Fundação Kalouste Gulbenkian, 927p., 2001.

OPITZ, T. Latent gaussian modeling and inla: A review with focus on space-time applications to appear in the journal of the french statistical society (special issue on space-time statistics). 2017.

R CORE TEAM. **R: A Language and Environment for Statistical Computing**. Vienna, Austria, 2020. Disponível em: <<https://www.R-project.org/>>.

RIEBLER, A. et al. An intuitive bayesian spatial model for disease mapping that accounts for scaling. **Statistical methods in medical research**, Sage Publications Sage UK: London, England, v. 25, n. 4, p. 1145–1165, 2016.

RUE, H.; HELD, L. **Gaussian Markov Random Fields: Theory and Applications**. [S.l.]: CRC press, 2005.

RUE, H.; MARTINO, S.; CHOPIN, N. Approximate bayesian inference for latent gaussian models by using integrated nested laplace approximations. **Journal of the royal statistical society: Series B (statistical methodology)**, Wiley Online Library, v. 71, n. 2, p. 319–392, Apr. 2009.

RUE, H. et al. Bayesian computing with inla: a review. **Annual Review of Statistics and Its Application**, Annual Reviews, v. 4, p. 395–421, 2017.

SADYKOVA, D. et al. Bayesian joint models with inla exploring marine mobile predator–prey and competitor species habitat overlap. **Ecology and Evolution**, Wiley Online Library, v. 7, n. 14, p. 5212–5226, 2017.

SCHRÖDLE, B.; HELD, L. A primer on disease mapping and ecological regression using inla. **Computational statistics**, Springer, v. 26, n. 2, p. 241–258, 2011.

SEARLE, S. R. Matrix algebra useful for statistics (wiley series in probability and statistics). Wiley-Interscience, 1982.

SILVA, J. A. A. d. C.; FLORESTAL, G. d. T. do C. et al. O código florestal e a ciência: contribuições para o diálogo. **SBPC**, SBPC, 2012.

SIMPSON, D. et al. Penalising model component complexity: A principled, practical approach to constructing priors. **Statistical science**, Institute of Mathematical Statistics, v. 32, n. 1, p. 1–28, 2017.

SØRBYE, S. H.; RUE, H. Scaling intrinsic gaussian markov random field priors in spatial modelling. **Spatial Statistics**, Elsevier, v. 8, p. 39–51, 2014.

TUNDISI, J. G.; TUNDISI, T. M. Impactos potenciais das alterações do código florestal nos recursos hídricos. **Biota Neotropica**, SciELO Brasil, v. 10, n. 4, p. 67–75, 2010.

WAKEFIELD, J. Disease mapping and spatial regression with count data. **Biostatistics**, Oxford University Press, v. 8, n. 2, p. 158–183, 2007.

WATANABE, S. Asymptotic equivalence of bayes cross validation and widely applicable information criterion in singular learning theory. **Journal of machine learning research**, v. 11, n. 12, 2010.

WHO et al. Who global strategy on health, environment and climate change: the transformation needed to improve lives and wellbeing sustainably through healthy environments. World Health Organization, 2020.

WICKHAM, H. et al. Welcome to the tidyverse. **Journal of Open Source Software**, v. 4, n. 43, p. 1686, 2019.

WIKLE, C. K. Hierarchical models in environmental science. **International Statistical Review**, Wiley Online Library, v. 71, n. 2, p. 181–199, 2003.

APÊNDICE A – Divergência de Kullback-Leibler para Distribuições Gaussianas

Seja $\mathcal{N}^{(p)}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ uma distribuição normal multivariada com dimensão p . A divergência KL entre duas distribuições $\mathcal{N}_1^{(p)}$ e $\mathcal{N}_0^{(p)}$ é definida por

$$\text{KLD} \left(\mathcal{N}_1^{(p)} \parallel \mathcal{N}_0^{(p)} \right) = \mathbb{E}_{\mathcal{N}_1^{(p)}} \left[\log \frac{\mathcal{N}_1^{(p)}}{\mathcal{N}_0^{(p)}} \right]$$

Para $\mathcal{N}_1^{(p)}(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_1)$ e $\mathcal{N}_0^{(p)}(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$, temos que

$$\begin{aligned} \text{KLD} \left(\mathcal{N}_1^{(p)} \parallel \mathcal{N}_0^{(p)} \right) &= \mathbb{E}_{\mathcal{N}_1^{(p)}} \left[\log \frac{\mathcal{N}_1^{(p)}}{\mathcal{N}_0^{(p)}} \right] \\ &= \mathbb{E}_{\mathcal{N}_1^{(p)}} \left[\log \mathcal{N}_1^{(p)} - \log \mathcal{N}_0^{(p)} \right] \\ &= \frac{1}{2} \mathbb{E}_{\mathcal{N}_1^{(p)}} \left[-\log(|\boldsymbol{\Sigma}_1|) - (\mathbf{x} - \boldsymbol{\mu}_1)^\top \boldsymbol{\Sigma}_1^{-1} (\mathbf{x} - \boldsymbol{\mu}_1) \right. \\ &\quad \left. + \log(|\boldsymbol{\Sigma}_0|) + (\mathbf{x} - \boldsymbol{\mu}_0)^\top \boldsymbol{\Sigma}_0^{-1} (\mathbf{x} - \boldsymbol{\mu}_0) \right] \\ &= \frac{1}{2} \log \left(\frac{|\boldsymbol{\Sigma}_0|}{|\boldsymbol{\Sigma}_1|} \right) + \frac{1}{2} \mathbb{E}_{\mathcal{N}_1^{(p)}} \left[-(\mathbf{x} - \boldsymbol{\mu}_1)^\top \boldsymbol{\Sigma}_1^{-1} (\mathbf{x} - \boldsymbol{\mu}_1) \right. \\ &\quad \left. + (\mathbf{x} - \boldsymbol{\mu}_0)^\top \boldsymbol{\Sigma}_0^{-1} (\mathbf{x} - \boldsymbol{\mu}_0) \right] \\ &= \frac{1}{2} \log \left(\frac{|\boldsymbol{\Sigma}_0|}{|\boldsymbol{\Sigma}_1|} \right) + \frac{1}{2} \mathbb{E}_{\mathcal{N}_1^{(p)}} \left[-\text{tr} \left(\boldsymbol{\Sigma}_1^{-1} (\mathbf{x} - \boldsymbol{\mu}_1)^\top (\mathbf{x} - \boldsymbol{\mu}_1) \right) \right. \\ &\quad \left. + \text{tr} \left(\boldsymbol{\Sigma}_0^{-1} (\mathbf{x} - \boldsymbol{\mu}_0)^\top (\mathbf{x} - \boldsymbol{\mu}_0) \right) \right] \\ &= \frac{1}{2} \log \left(\frac{|\boldsymbol{\Sigma}_0|}{|\boldsymbol{\Sigma}_1|} \right) + \frac{1}{2} \mathbb{E}_{\mathcal{N}_1^{(p)}} \left[-\text{tr}(\boldsymbol{\Sigma}_1^{-1} \boldsymbol{\Sigma}_1) \right. \\ &\quad \left. + \text{tr} \left(\boldsymbol{\Sigma}_0^{-1} (\mathbf{x} \mathbf{x}^\top - 2\mathbf{x} \boldsymbol{\mu}_0^\top + \boldsymbol{\mu}_0 \boldsymbol{\mu}_0^\top) \right) \right] \\ &= \frac{1}{2} \log \left(\frac{|\boldsymbol{\Sigma}_0|}{|\boldsymbol{\Sigma}_1|} \right) - \frac{1}{2} p \\ &\quad + \frac{1}{2} \text{tr} \left(\boldsymbol{\Sigma}_0^{-1} (\boldsymbol{\Sigma}_1 + \boldsymbol{\mu}_1 \boldsymbol{\mu}_1^\top - 2\boldsymbol{\mu}_0 \boldsymbol{\mu}_1^\top + \boldsymbol{\mu}_0 \boldsymbol{\mu}_0^\top) \right) \\ &= \frac{1}{2} \left(\log \left(\frac{|\boldsymbol{\Sigma}_0|}{|\boldsymbol{\Sigma}_1|} \right) - p + \text{tr}(\boldsymbol{\Sigma}_0^{-1} \boldsymbol{\Sigma}_1) \right) \\ &\quad + \frac{1}{2} \text{tr} \left(\boldsymbol{\mu}_1 \boldsymbol{\Sigma}_0^{-1} \boldsymbol{\mu}_1^\top - 2\boldsymbol{\mu}_0 \boldsymbol{\Sigma}_0^{-1} \boldsymbol{\mu}_1^\top + \boldsymbol{\mu}_0 \boldsymbol{\Sigma}_0^{-1} \boldsymbol{\mu}_0^\top \right) \\ &= \frac{1}{2} \left(\log \left(\frac{|\boldsymbol{\Sigma}_0|}{|\boldsymbol{\Sigma}_1|} \right) - p + \text{tr}(\boldsymbol{\Sigma}_0^{-1} \boldsymbol{\Sigma}_1) \right) \\ &\quad + \frac{1}{2} (\boldsymbol{\mu}_0 - \boldsymbol{\mu}_1)^\top \boldsymbol{\Sigma}_0^{-1} (\boldsymbol{\mu}_0 - \boldsymbol{\mu}_1) \end{aligned}$$

$$\text{KLD} \left(\mathcal{N}_1^{(p)} \parallel \mathcal{N}_0^{(p)} \right)$$

$$\text{KLD} = \frac{1}{2} \left\{ \text{tr}(\mathbf{\Sigma}_0^{-1} \mathbf{\Sigma}_1) + (\boldsymbol{\mu}_0 - \boldsymbol{\mu}_1)^\top \mathbf{\Sigma}_0^{-1} (\boldsymbol{\mu}_0 - \boldsymbol{\mu}_1) - p - \log \left(\frac{|\mathbf{\Sigma}_1|}{|\mathbf{\Sigma}_0|} \right) \right\}.$$

Para derivar a priori PC para τ , em que $\mathbf{\Sigma}_1 = \frac{\mathbf{R}}{\tau}$, $\mathbf{R}$ uma matriz fixa de posto completo e $\mathbf{\Sigma}_0 = \mathbf{0}$. Para estudar comportamento limitante quando $\mathbf{\Sigma}_0 = \frac{\mathbf{R}}{\tau_0}$ para alto valor fixado de τ_0 , pode-se realizar o limite de KLD quando $\tau_0 \rightarrow \infty$.

Desse modo, temos que

$$\text{tr}(\mathbf{\Sigma}_0^{-1} \mathbf{\Sigma}_1) = \text{tr} \left(\frac{\mathbf{R}^{-1} \mathbf{R}}{\tau_0^{-1} \tau} \right) = \text{tr} \left(\frac{\tau_0}{\tau} \mathbf{I} \right) = \frac{\tau_0}{\tau} p \quad (1)$$

e

$$\begin{aligned} \log \left(\frac{|\mathbf{\Sigma}_1|}{|\mathbf{\Sigma}_0|} \right) &= \log \left(\frac{\left| \frac{\mathbf{R}}{\tau} \right|}{\left| \frac{\mathbf{R}}{\tau_0} \right|} \right) = \log \left(\frac{\frac{1}{\tau^p} |\mathbf{R}|}{\frac{1}{\tau_0^p} |\mathbf{R}|} \right) = \log \left(\frac{\tau_0}{\tau} \right)^p = \\ &= p \log \left(\frac{\tau_0}{\tau} \right) = -p \log \left(\frac{\tau}{\tau_0} \right) \end{aligned} \quad (2)$$

De (1) e (2), podemos reescrever

$$\text{KLD} = \frac{1}{2} \left\{ \frac{\tau_0}{\tau} p + (\boldsymbol{\mu}_0 - \boldsymbol{\mu}_1)^\top \left(\frac{\mathbf{R}}{\tau_0} \right)^{-1} (\boldsymbol{\mu}_0 - \boldsymbol{\mu}_1) - p + p \log \left(\frac{\tau}{\tau_0} \right) \right\}. \quad (3)$$

Aplicando o limite em (3), temos que

$$\begin{aligned} \lim_{\tau_0 \rightarrow \infty} (3) &= \lim_{\tau_0 \rightarrow \infty} \frac{1}{2} (\boldsymbol{\mu}_0 - \boldsymbol{\mu}_1)^\top \left(\frac{\mathbf{R}}{\tau_0} \right)^{-1} (\boldsymbol{\mu}_0 - \boldsymbol{\mu}_1) \\ &\quad + \lim_{\tau_0 \rightarrow \infty} \frac{p}{2} \frac{\tau_0}{\tau} \left(1 + \frac{\tau}{\tau_0} \log \left(\frac{\tau}{\tau_0} \right) - \frac{\tau}{\tau_0} \right) \end{aligned} \quad (4)$$

Mas,

$$\lim_{\tau_0 \rightarrow \infty} \frac{1}{2} (\boldsymbol{\mu}_0 - \boldsymbol{\mu}_1)^\top \left(\frac{\mathbf{R}}{\tau_0} \right)^{-1} (\boldsymbol{\mu}_0 - \boldsymbol{\mu}_1) = 0$$

e

$$\begin{aligned} \lim_{\tau_0 \rightarrow \infty} \frac{p}{2} \frac{\tau_0}{\tau} \left(1 + \frac{\tau}{\tau_0} \log \left(\frac{\tau}{\tau_0} \right) - \frac{\tau}{\tau_0} \right) &= \lim_{\tau_0 \rightarrow \infty} \frac{p}{2} \frac{\tau_0}{\tau} \\ &\times \lim_{\tau_0 \rightarrow \infty} \left(1 + \frac{\tau}{\tau_0} \log \left(\frac{\tau}{\tau_0} \right) - \frac{\tau}{\tau_0} \right). \end{aligned}$$

Porém,

$$\lim_{\tau_0 \rightarrow \infty} \frac{p}{2} \frac{\tau_0}{\tau} \left(1 + \frac{\tau}{\tau_0} \log \left(\frac{\tau}{\tau_0} \right) - \frac{\tau}{\tau_0} \right) = 1,$$

pois,

$$\lim_{\tau_0 \rightarrow \infty} \frac{\tau}{\tau_0} \log \left(\frac{\tau}{\tau_0} \right) = \lim_{\tau_0 \rightarrow \infty} \frac{\log \left(\frac{\tau}{\tau_0} \right)}{\frac{\tau_0}{\tau}} \stackrel{L'H}{=} \lim_{\tau_0 \rightarrow \infty} \frac{\frac{\tau_0}{\tau} \left(-\frac{\tau}{\tau_0^2} \right)}{\frac{1}{\tau}} = \lim_{\tau_0 \rightarrow \infty} -\frac{\tau}{\tau_0} = 0.$$

Portanto,

$$\lim_{\tau_0 \rightarrow \infty} (\text{KLD}) = \lim_{\tau_0 \rightarrow \infty} \frac{p}{2} \frac{\tau_0}{\tau} \rightarrow \frac{p}{2} \frac{\tau_0}{\tau},$$

quando $\tau \ll \tau_0$.

APÊNDICE B – Figuras adicionais

Figura 1 – Distribuição marginal a posteriori das covariáveis do modelo final poisson para o número de autos de infração em Minas Gerais no período de 2008 a 2016

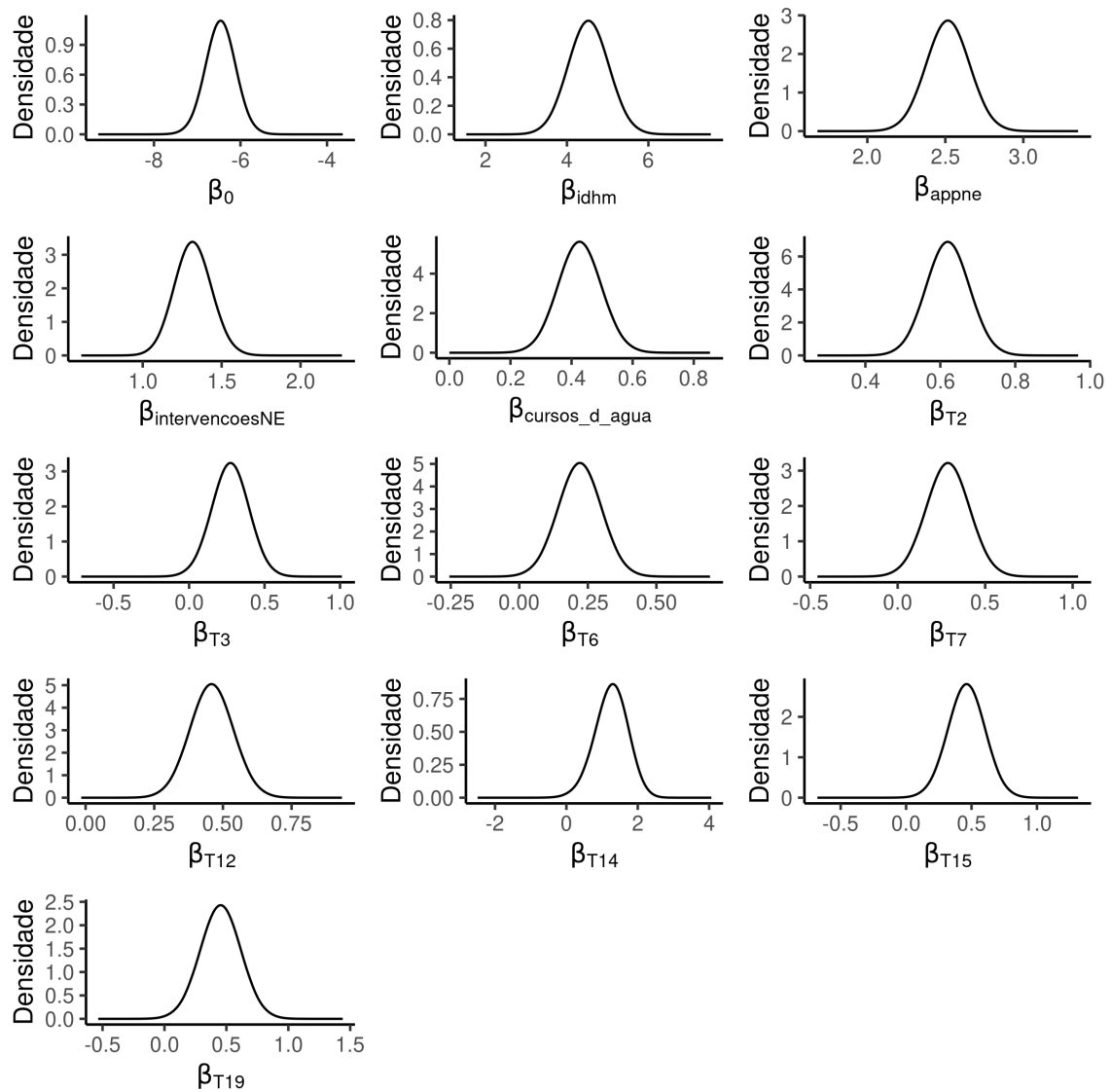


Figura 2 – Distribuição marginal a posteriori dos hiperparâmetros do modelo final poisson para o número de autos de infração em Minas Gerais no período de 2008 a 2016

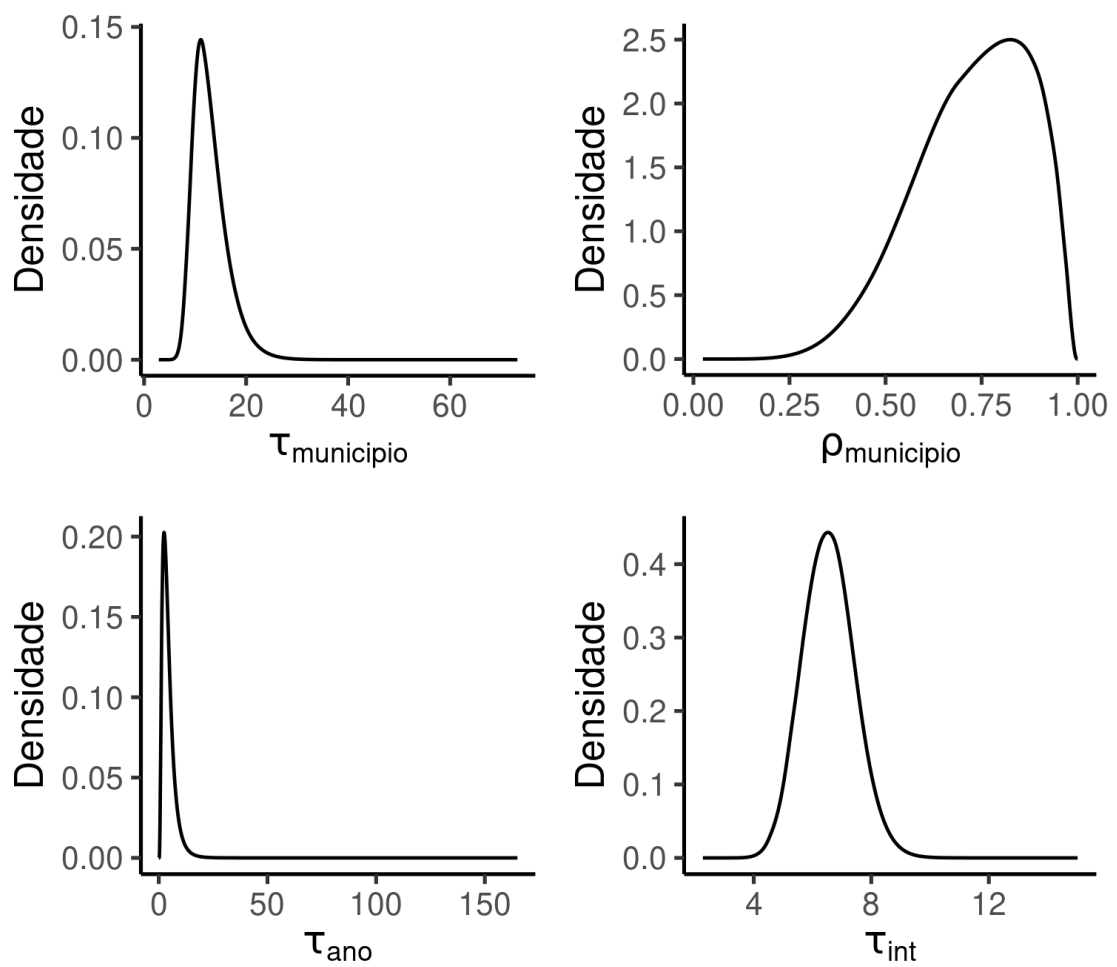


Figura 3 – Distribuição marginal a posteriori das covariáveis do modelo final *hurdle*-gama para o valor dos autos de infração em Minas Gerais no período de 2008 a 2016

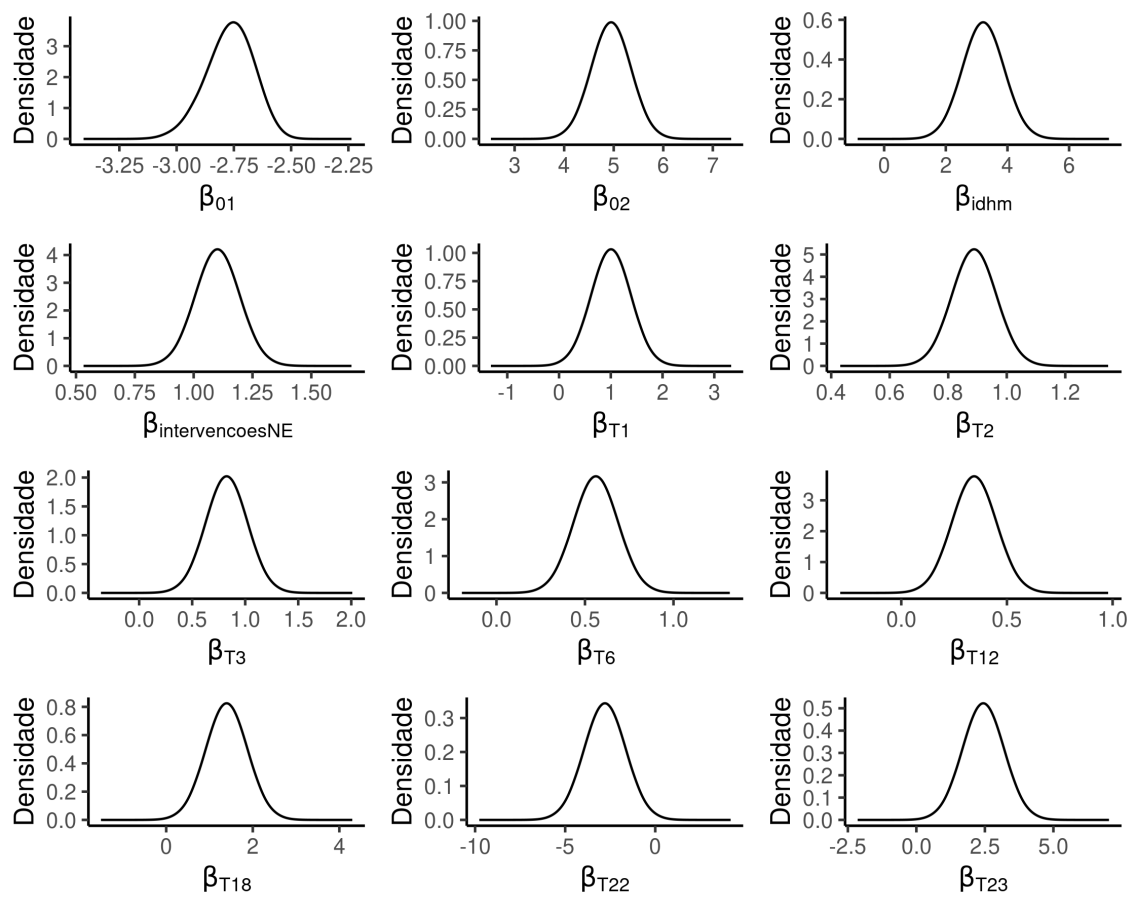


Figura 4 – Distribuição marginal a posteriori dos hiperparâmetros referente ao preditor da verossimilhança binomial do modelo final *hurdle*-gama para o valor de autos de infração em Minas Gerais no período de 2008 a 2016

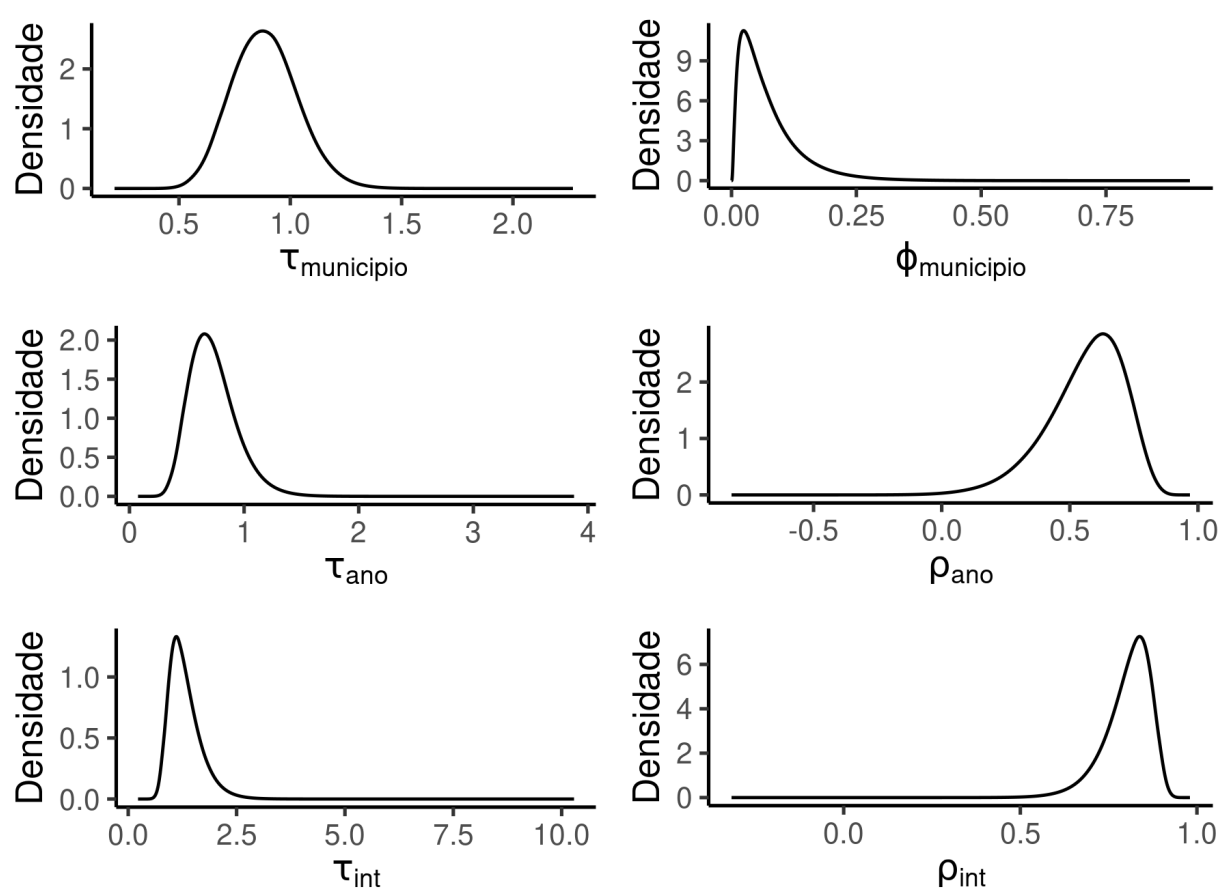


Figura 5 – Distribuição marginal a posteriori dos hiperparâmetros referente ao preditor da verossimilhança gama do modelo final *hurdle*-gama para o valor de autos de infração em Minas Gerais no período de 2008 a 2016

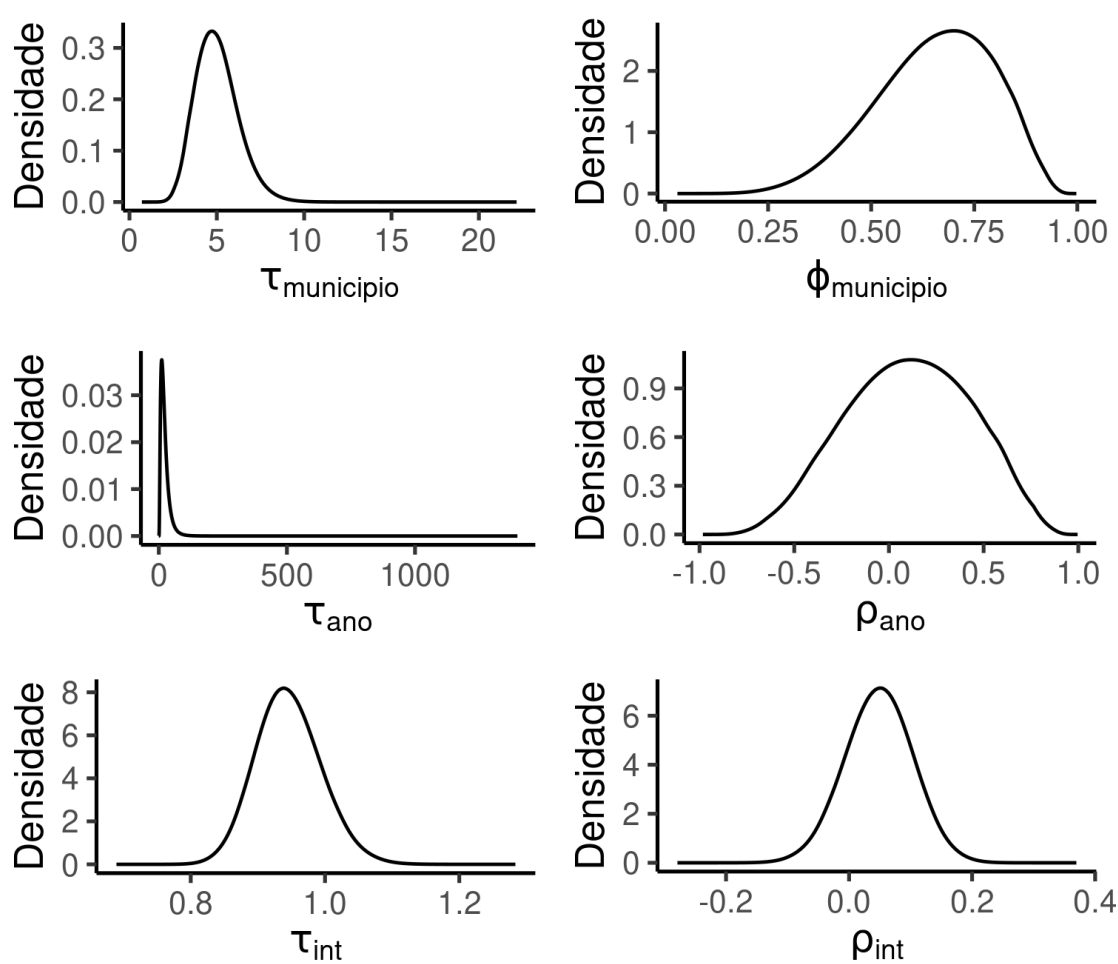


Figura 6 – Média a posteriori da interação para a ocorrência de autos do modelo final conjunto para ocorrência, número e valor dos autos de infração ambientais em Minas Gerais no período de 2008 a 2016

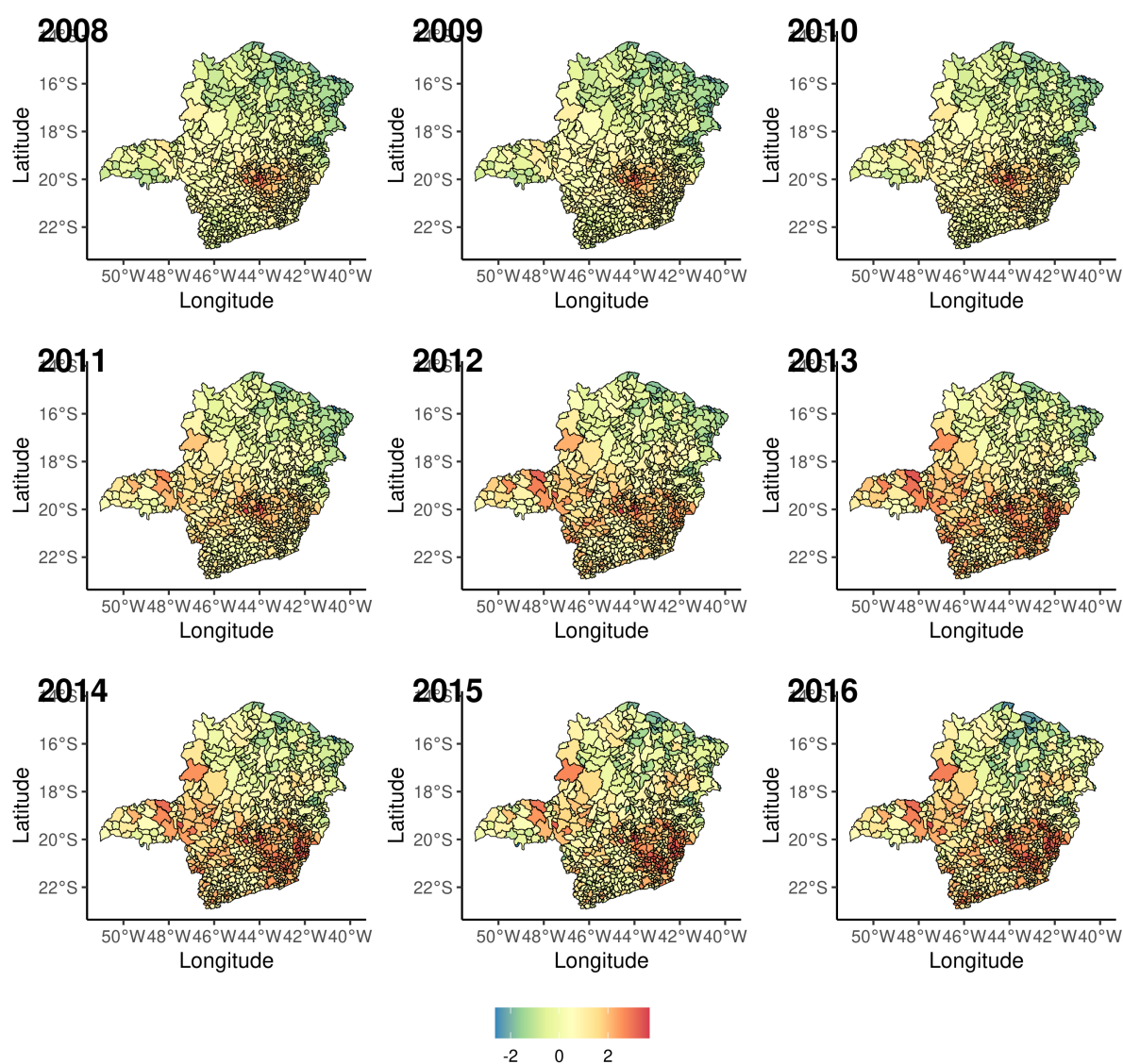


Figura 7 – Média a posteriori da interação para o número de autos do modelo final conjunto para ocorrência, número e valor dos autos de infração ambientais em Minas Gerais no período de 2008 a 2016

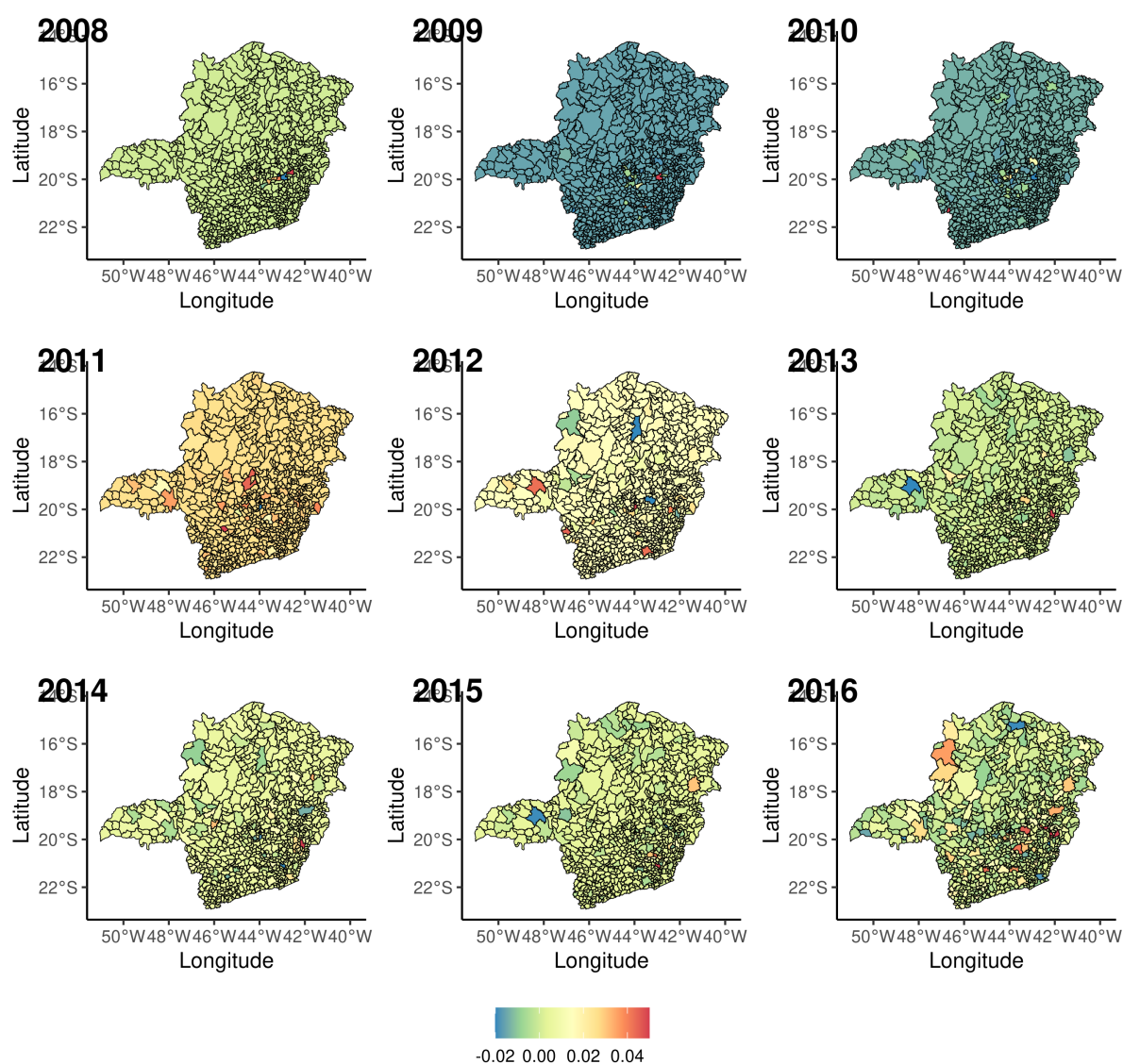


Figura 8 – Média a posteriori da interação para o valor de autos do modelo final conjunto para ocorrência, número e valor dos autos de infração ambientais em Minas Gerais no período de 2008 a 2016

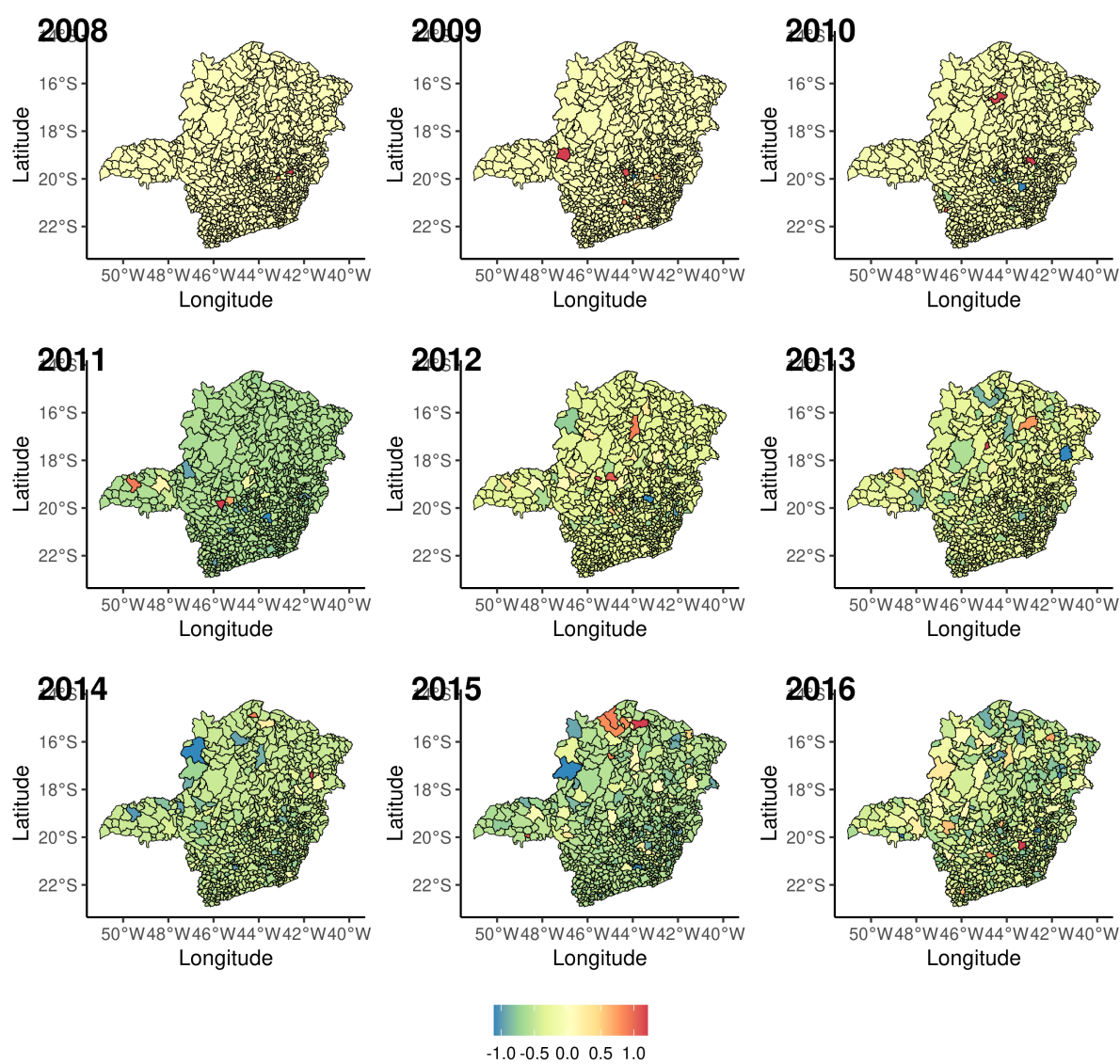


Figura 9 – Distribuições marginais a posteriori das covariáveis do modelo final conjunto para o número e valor de autos de infração em Minas Gerais no período de 2008 a 2016

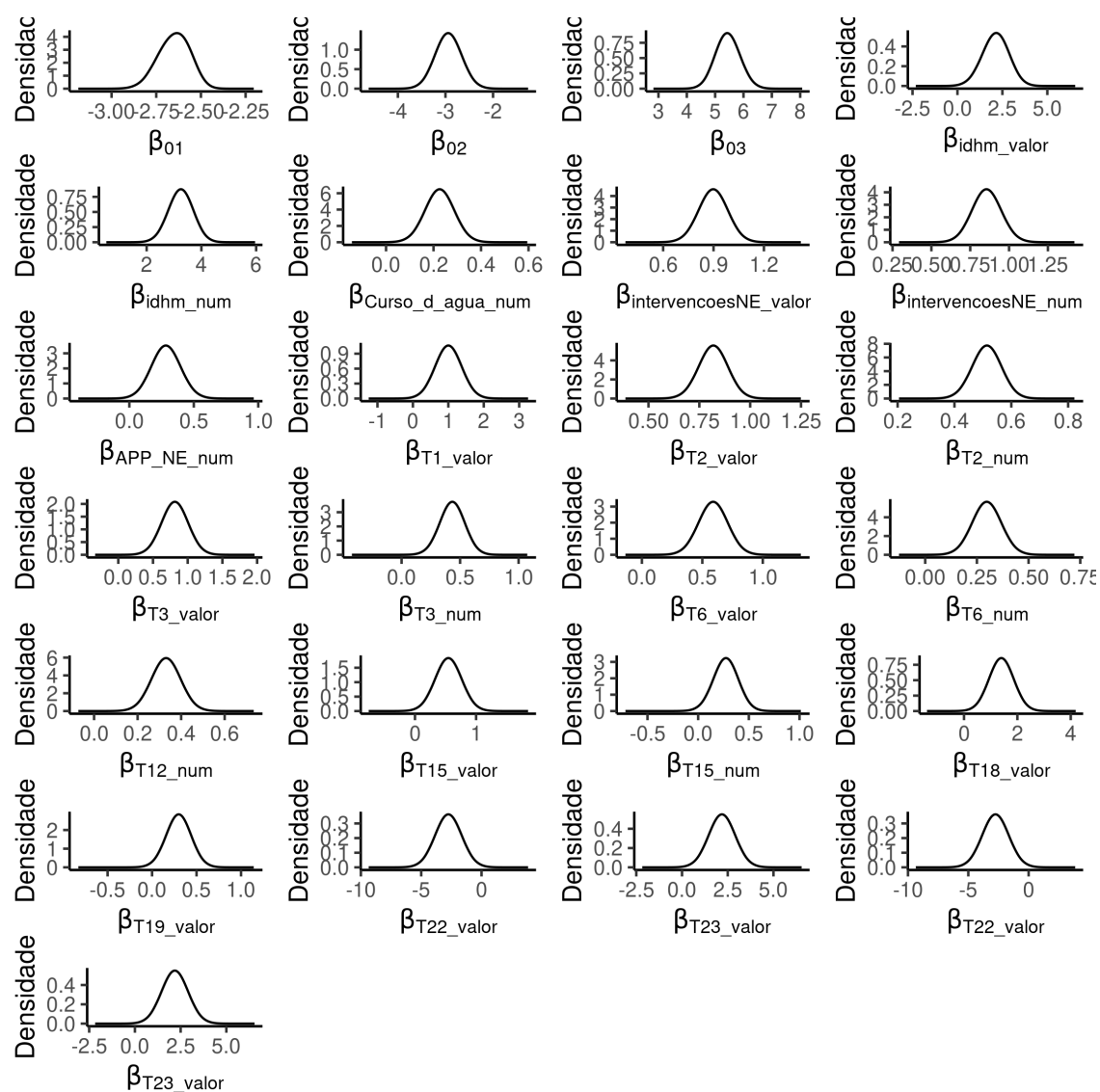


Figura 10 – Distribuições marginais a posteriori dos hiperparâmetros do modelo final conjunto para o número e valor de autos de infração em Minas Gerais no período de 2008 a 2016

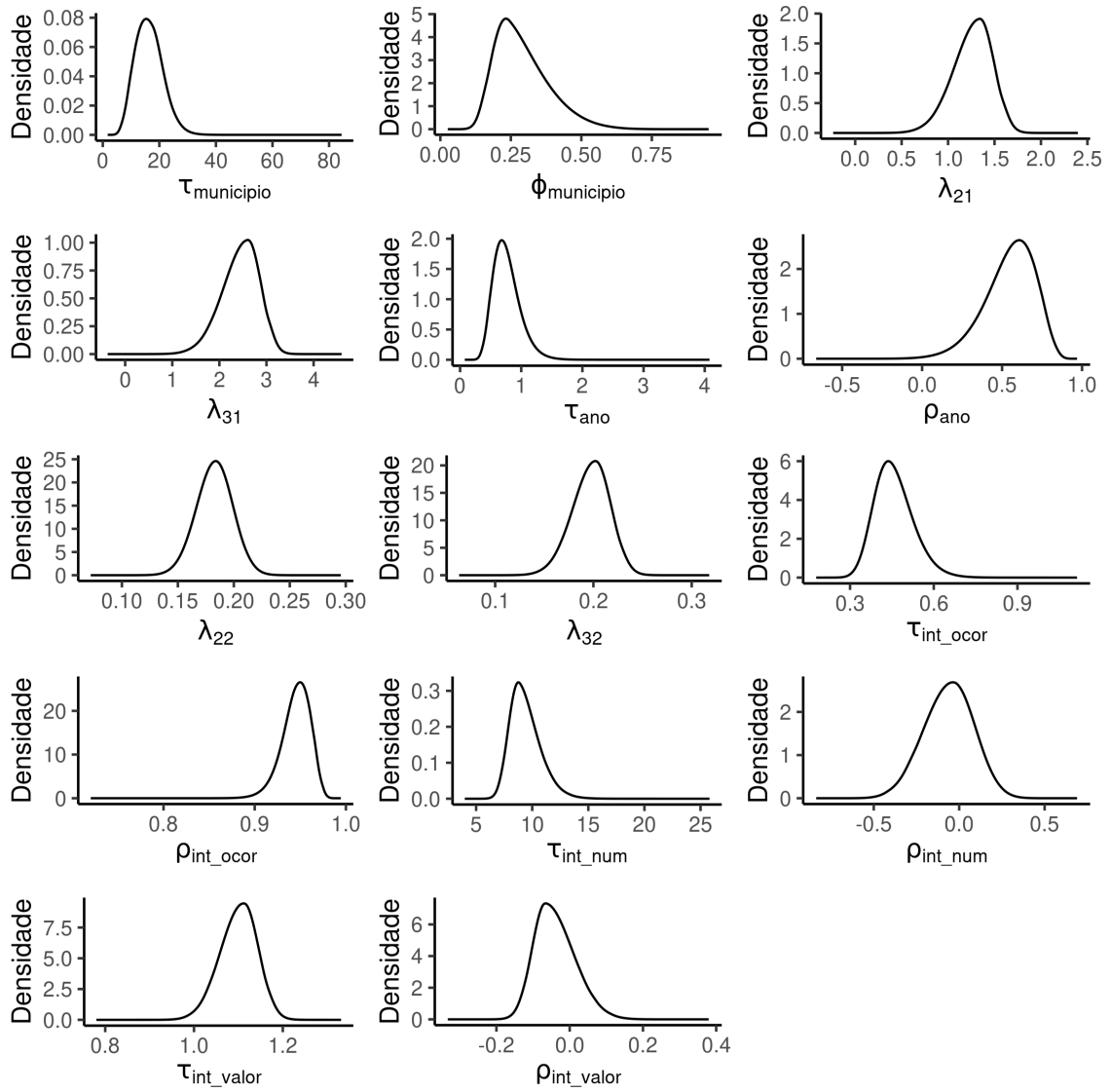


Figura 11 – Valores preditos da probabilidade de ocorrência dos autos de infração ambientais do modelo final conjunto com verossimilhanças binomial, Poisson e gama

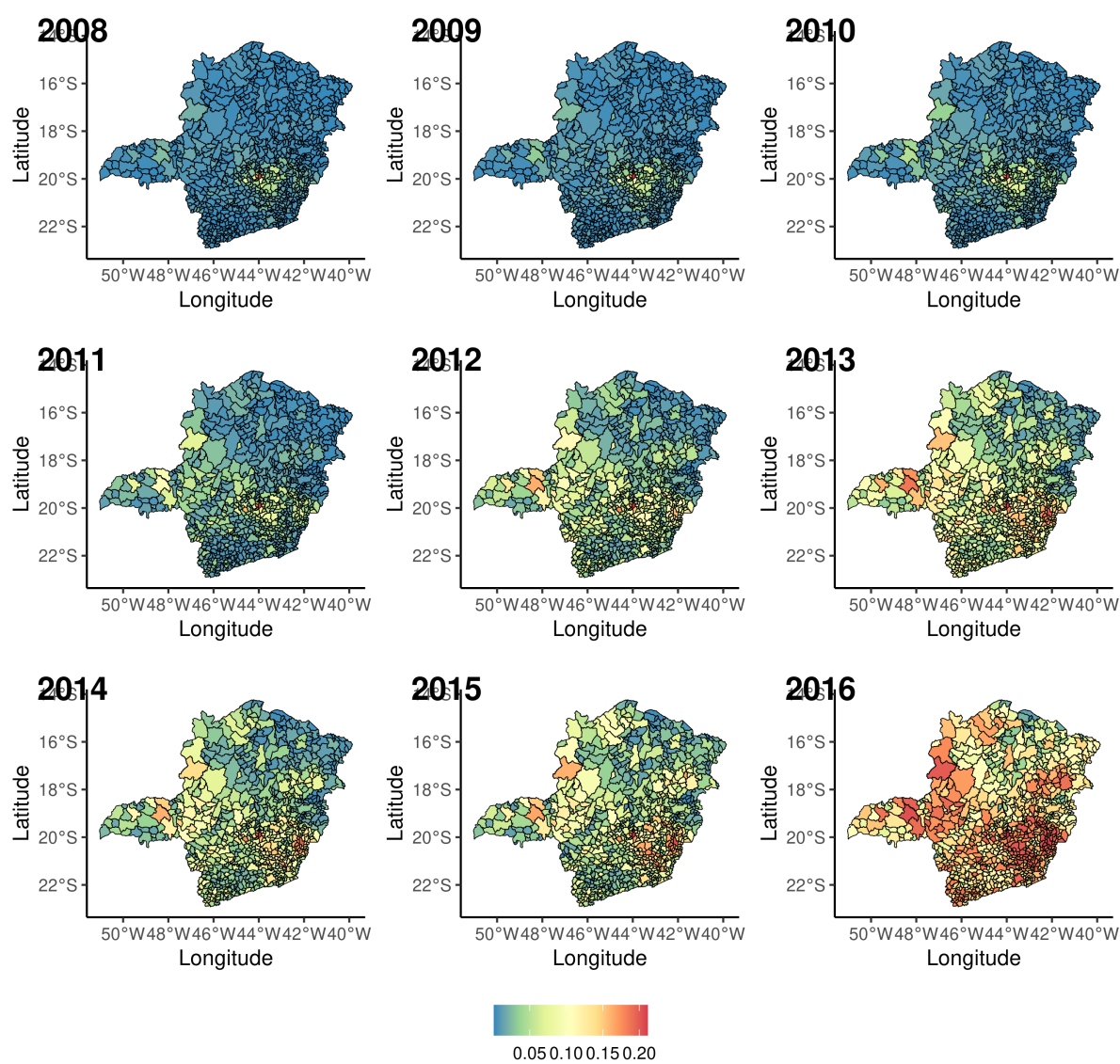


Figura 12 – Valores preditos do número de autos de infração ambientais do modelo final conjunto com verossimilhanças binomial, Poisson e gama (escala raiz quadrada)

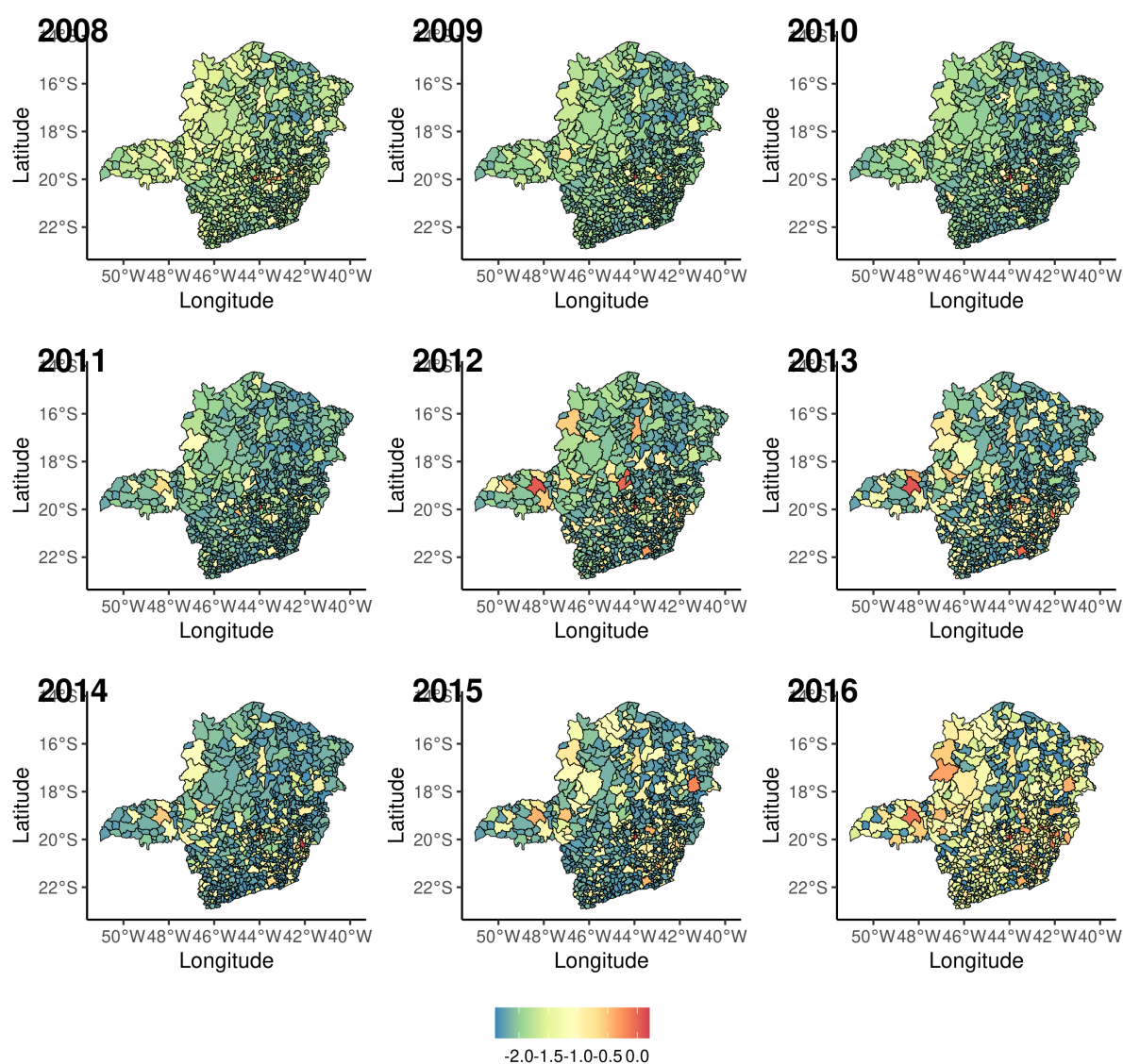
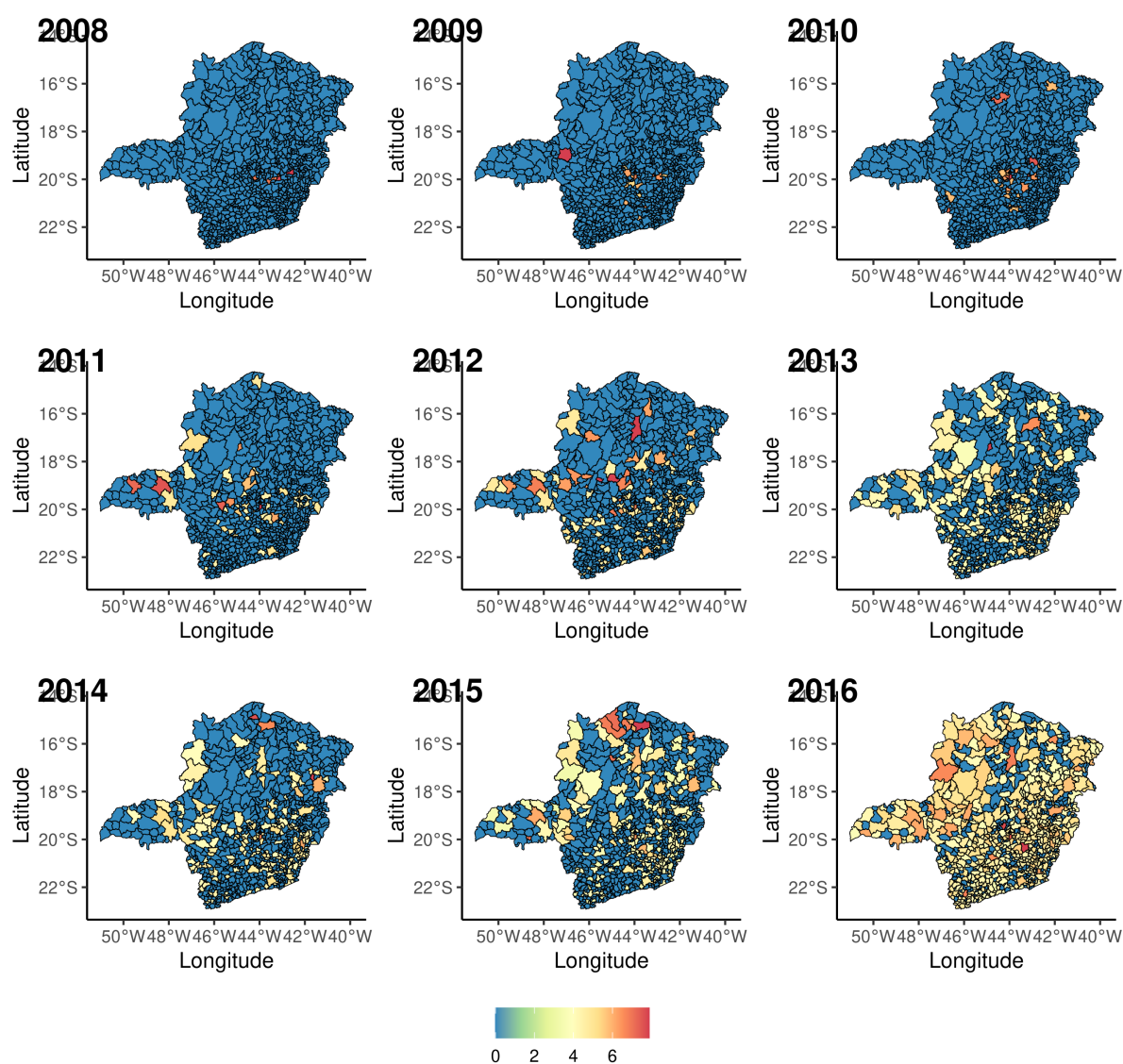


Figura 13 – Valores preditos do valor agregado de autos de infração ambientais do modelo final conjunto com verossimilhanças binomial, Poisson e gama (escala logarítmica)




```

33         phi = list(prior = "pc",
34                    param = c(0.5, 0.5)))
35
36 hyperprior_rw <- list(prec = list(prior = "pc.prec",
37                                   param = c(u = 1, alpha = 0.01)))
38
39 hyperprior_ar1 <- list(theta1 = list(prior = "pc.prec",
40                                     param = c(u = 1, alpha = 0.01)),
41                        theta2 = list(prior = "pc.cor0",
42                                     param = c(u = 0.5, alpha = 0.5)))
43
44 ## MODELO FINAL PARA O NUMERO DE AUTOS DE INFRACAO (POISSON VII) ##
45
46 # predictor
47 f <- n ~ idhm.r + idhm.e +
48     appne + nascente + cursodagua +
49     resrv_artf + t2 + t15 + t23 +
50     f(cod, model = "bym2", graph = adj,
51       constr = TRUE, scale.model = TRUE,
52       hyper = hyperprior_bym2) +
53     f(year, model = "ar1",
54       hyper = hyperprior_ar1) +
55     f(cod.copy, model = "iid",
56       group = year.copy,
57       control.group = list(model = "ar1"))
58
59 # Ajuste do modelo
60 m7 <- inla(formula = f,
61           family = "poisson", data = data,
62           control.inla = list(h = 0.001,
63                               strategy = "simplified.laplace",
64                               int.strategy = "ccd"),
65           control.predictor = list(link = 1, compute = TRUE),
66           control.compute = list(waic = TRUE, cpo = TRUE,

```



```

67     openmp.strategy="huge"))
68
69 # resumo do modelo (efeitos fixos e hiperparametros)
70
71 summary(m7)
72
73 # medidas de criterios de selecao
74
75 # Preditos
76 y_star <- m7$summary.fitted.values$mean
77
78 # log - score
79 (lm7 <- -mean(log(m7$cpo$cpo), na.rm = TRUE))
80
81 # WAIC
82 (wm7 <- m7$waic$waic)
83
84 # RMSE
85 (rm7 <- sqrt(1/N * sum((y - y_star)^2)))
86
87 save.image(file = "P7.RData")
88
89 # FIGURAS DOS EFEITOS ALEATORIOS
90 load('P7.RData')
91
92 # estatisticas do efeito espacial dos municipios
93 mean.spatial <- m7$summary.random$cod$mean[1:Nareas]
94 sd.spatial <- m7$summary.random$cod$sd[1:Nareas]
95 q2.5 <- m7$summary.random$cod$`0.025quant`[1:Nareas]
96 q97.5 <- m7$summary.random$cod$`0.975quant`[1:Nareas]
97 summary(mean.spatial)
98 summary(sd.spatial)
99
100

```

```

101 # figuras do modelo bym2
102
103 # adicionando informacoes ao shape file para a construcao dos graficos
104 shp <- read_sf('31MUE250GC_SIR.shp')
105 shp$mean.spatial <- mean.spatial
106 shp$sd.spatial <- sd.spatial
107 breaks.mean <- seq(-1.5,1.5,0.5)
108 breaks.sd <- seq(0.2,0.8,0.10)
109
110 # construindo o grafico da media a posteriori dos efeitos espaciais
    municipais
111 media <- ggplot(shp) +
112   geom_sf(aes(fill = mean.spatial), size = 0.2, color = "black") +
113   xlab("Longitude") + ylab("Latitude") + labs(fill = " ") +
114   annotate(geom = "text", label = "(a) Media", x = -44.5, y = -13,
115     color = "black", size = 5.5) +
116   scale_fill_viridis_c(option = "cividis") +
117   theme_classic2()
118   print(media)
119
120 # salvando o grafico da media a posteriori dos efeitos espaciais
    municipais
121 ggsave("num_ms.png", dpi = "retina")
122
123 # construcao do grafico do desvio padrao a posteriori dos efeitos
    espaciais municipais
124
125 desv <- ggplot(shp) +
126   geom_sf(aes(fill = sd.spatial), size = 0.2, color = "black") +
127   xlab("Longitude") + ylab("Latitude") + labs(fill = " ") +
128   annotate(geom = "text", label = "(b) Desvio Padrao", x = -44.5, y =
    -13,
129     color = "black", size = 5.5) +
130   scale_fill_viridis_c(option = "cividis") +

```

```

131 theme_classic2()
132
133 print(desv)
134
135 # salvando o grafico do desvio padrao a posteriori
136 ggsave("num_sds.png", dpi = 320)
137
138 #####
139 #figura do efeito temporal
140
141 # Resumo do Efeito temporal (AR1)
142 mean.temporal <- m7$summary.random$year$mean
143 sd.temporal <- m7$summary.random$year$sd
144 q2.5temporal <- m7$summary.random$year$`0.025quant`
145 q97.5temporal <- m7$summary.random$year$`0.975quant`
146
147 data.times.effect <- tibble(mean.temporal,
148                               sd.temporal,
149                               q2.5temporal,
150                               q97.5temporal,
151                               year = 2008:2016)
152
153 # Grafico do efeito temporal
154 ggplot(data.times.effect, aes(year, mean.temporal)) +
155   geom_ribbon(aes(ymin=q2.5temporal, ymax=q97.5temporal),
156             fill= "burlywood",
157             color= "burlywood")+
158   geom_point()+
159   geom_hline(yintercept = 0, linetype = "dashed")+
160   labs(y = "Media a posteriori do efeito temporal", x = "Ano")+
161   stat_smooth(method = "loess", col = "chocolate4", se = FALSE) +
162   theme_classic()
163 #
164

```

```

165 # salvando o grafico do efeito temporal anual
166 ggsave("temp_effect_num.png", dpi = "retina")
167
168 #####
169 #grafico do efeito de interacao espaco-tempo
170
171 # Estatisticas da interacao espaco-temporal
172 mean.interaction <- m7$summary.random$cod.copy$mean
173 sd.interaction   <- m7$summary.random$cod.copy$sd
174 q2.5interaction <- m7$summary.random$cod.copy$`0.025quant`
175 q97.5interaction <- m7$summary.random$cod.copy$`0.975quant`
176
177 # adicionando informacoes da media a posteriori das interacoes ao
    shape file
178 vetor <- round(m7$summary.random$cod.copy$mean,2)
179 summary(vetor)
180
181 i <- 853
182 # all.equal(data$count[data$id.date == 38] , data$count[(37 * i + 1)
    :(38*i)])
183
184 shp$ano2008 <- vetor[1:i]
185 shp$ano2009 <- vetor[(i+1):(2*i)]
186 shp$ano2010 <- vetor[(2*i+1):(3*i)]
187 shp$ano2011 <- vetor[(3*i+1):(4*i)]
188 shp$ano2012 <- vetor[(4*i+1):(5*i)]
189 shp$ano2013 <- vetor[(5*i+1):(6*i)]
190 shp$ano2014 <- vetor[(6*i+1):(7*i)]
191 shp$ano2015 <- vetor[(7*i+1):(8*i)]
192 shp$ano2016 <- vetor[(8*i+1):(9*i)]
193
194 # contruindo os mapas das interacoes para cada ano
195 int_2008 <- ggplot(shp) + geom_sf(aes(fill = ano2008),
196 size = 0.2, color = "black") +

```

```

197 xlab("Longitude") + ylab("Latitude") + labs(fill = " ") +
198 scale_fill_viridis_c(option = "cividis") +
199 theme_classic2()
200
201 int_2009 <- ggplot(shp) + geom_sf(aes(fill = ano2009),
202 size = 0.2, color = "black") +
203 xlab("Longitude") + ylab("Latitude") + labs(fill = " ") +
204 scale_fill_viridis_c(option = "cividis")+
205 theme_classic2()
206
207 int_2010 <- ggplot(shp) +
208     geom_sf(aes(fill = ano2010),
209             size = 0.2, color = "black") +
210     xlab("Longitude") + ylab("Latitude") +
211     labs(fill = " ") +
212     scale_fill_viridis_c(option = "cividis")+
213     theme_classic2()
214
215 int_2011 <- ggplot(shp) +
216     geom_sf(aes(fill = ano2011),
217             size = 0.2, color = "black") +
218     xlab("Longitude") + ylab("Latitude") +
219     labs(fill = " ") +
220     scale_fill_viridis_c(option = "cividis")+
221     theme_classic2()
222
223 int_2012 <- ggplot(shp) +
224     geom_sf(aes(fill = ano2012),
225             size = 0.2, color = "black") +
226     xlab("Longitude") + ylab("Latitude") +
227     labs(fill = " ") +
228     scale_fill_viridis_c(option = "cividis")+
229     theme_classic2()
230

```

```

231 int_2013 <- ggplot(shp) +
232     geom_sf(aes(fill = ano2013),
233     size = 0.2, color = "black") +
234     xlab("Longitude") + ylab("Latitude") +
235     labs(fill = " ") +
236     scale_fill_viridis_c(option = "cividis")+
237     theme_classic2()
238
239 int_2014 <- ggplot(shp) +
240     geom_sf(aes(fill = ano2014),
241     size = 0.2, color = "black") +
242     xlab("Longitude") + ylab("Latitude") +
243     labs(fill = " ") +
244     scale_fill_viridis_c(option = "cividis")+
245     theme_classic2()
246
247 int_2015 <- ggplot(shp) +
248     geom_sf(aes(fill = ano2015),
249     size = 0.2, color = "black") +
250     xlab("Longitude") + ylab("Latitude") +
251     labs(fill = " ") +
252     scale_fill_viridis_c(option = "cividis")+
253     theme_classic2()
254
255 int_2016 <- ggplot(shp) +
256     geom_sf(aes(fill = ano2016),
257     size = 0.2, color = "black") +
258     xlab("Longitude") + ylab("Latitude") +
259     labs(fill = " ") +
260     scale_fill_viridis_c(option = "cividis")+
261     theme_classic2()
262
263 # plotando todos os graficos de interacoes em um mesmo quadro, sob
    uma mesma legenda

```

```

264
265 Grafico <- ggarrange(int_2008, int_2009,
266                     int_2010,int_2011,
267                     int_2012,int_2013,
268                     int_2014,int_2015,
269                     int_2016, ncol=3,
270                     nrow=3,
271                     common.legend = TRUE,
272                     labels = as.character(c(2008:2016)),
273                     legend="bottom",
274                     font.label = list(size = 18))
275 print(Grafico)
276
277 # salvando o grafico das medias a posteriori das interacoes
278
279 ggsave("Interacao_N.png",
280 plot = Grafico, scale = 2.0,
281 dpi = "retina")
282
283 #####
284
285 # grafico das densidades a posteriori marginais dos efeitos fixos e
    hiperparametros
286
287 # intercepto
288 b0 <- inla.s marginal(m7$marginals.fixed$(Intercept))
289 beta0 <- tibble(x = b0$x, y = b0$y)
290
291 beta0.fig <- ggplot(beta0) +
292 geom_line(aes(x = x, y = y)) +
293 labs(x = expression(beta[0]), y = "Densidade") +
294 theme_classic2()
295
296 # idhm renda

```

```

297 bIDHMr <- inla.sMarginal(m7$marginals.fixed$idhm.r)
298 betaIDHMr <- tibble(x = bIDHMr$x, y = bIDHMr$y)
299
300 betabetaIDHMr.fig <- ggplot(betaIDHMr) +
301   geom_line(aes(x = x, y = y)) +
302   labs(x = expression(beta[idhm_renda]), y = "Densidade") +
303   theme_classic2()
304
305 # idhm educacao
306 bIDHMe <- inla.sMarginal(m7$marginals.fixed$idhm.e)
307 betaIDHMe <- tibble(x = bIDHMe$x, y = bIDHMe$y)
308
309 betabetaIDHMe.fig <- ggplot(betaIDHMe) +
310   geom_line(aes(x = x, y = y)) +
311   labs(x = expression(beta[idhm_educacao]), y = "Densidade") +
312   theme_classic2()
313
314 # area de preservacao permanente nao identificada
315 bappne <- inla.sMarginal(m7$marginals.fixed$appne)
316 betaappne <- tibble(x = bappne$x, y = bappne$y)
317
318 betaappne.fig <- ggplot(betaappne) +
319   geom_line(aes(x = x, y = y)) +
320   labs(x = expression(beta[appne]), y = "Densidade") +
321   theme_classic2()
322 betaappne.fig
323
324 # nascentes
325 bnascente <- inla.sMarginal(m7$marginals.fixed$nascente)
326 betanascente <- tibble(x = bnascente$x, y = bnascente$y)
327
328 betanascente.fig <- ggplot(betascente) +
329   geom_line(aes(x = x, y = y)) +
330   labs(x = expression(beta[nascente]), y = "Densidade") +

```



```

331 theme_classic2()
332
333 # cursos d agua
334
335 bcda <- inla.s marginal(m7$marginals.fixed$cursodagua)
336 betacda <- tibble(x = bcda$x, y = bcda$y)
337
338 betacda.fig <- ggplot(betacda) +
339   geom_line(aes(x = x, y = y)) +
340   labs(x = expression(beta[cursos_d_agua]), y = "Densidade") +
341   theme_classic2()
342
343 # reservas artificiais
344
345 bra <- inla.s marginal(m7$marginals.fixed$resrv_artf)
346 betara <- tibble(x = bra$x, y = bra$y)
347
348 betara.fig <- ggplot(betara) +
349   geom_line(aes(x = x, y = y)) +
350   labs(x = expression(beta[reserva_artificial]), y = "Densidade") +
351   theme_classic2()
352
353 # supressao da vegetacao
354
355 bt2 <- inla.s marginal(m7$marginals.fixed$t2)
356 betat2 <- tibble(x = bt2$x, y = bt2$y)
357
358 betat2.fig <- ggplot(betat2) +
359   geom_line(aes(x = x, y = y)) +
360   labs(x = expression(beta[T2]), y = "Densidade") +
361   theme_classic2()
362
363 # criacao de animais bovinos e equinos
364

```

```

365 bt15 <- inla.s marginal(m7$marginals.fixed$t15)
366 betat15 <- tibble(x = bt15$x, y = bt15$y)
367
368 betat15.fig <- ggplot(betat15) +
369   geom_line(aes(x = x, y = y)) +
370   labs(x = expression(beta[T15]), y = "Densidade") +
371   theme_classic2()
372
373 # impedimento de regeneracao natural
374
375 bt23 <- inla.s marginal(m7$marginals.fixed$t23)
376 betat23 <- tibble(x = bt23$x, y = bt23$y)
377
378 betat23.fig <- ggplot(betat23) +
379   geom_line(aes(x = x, y = y)) +
380   labs(x = expression(beta[T23]), y = "Densidade") +
381   theme_classic2()
382
383 # plotando todos os graficos em um mesmo plot
384 fixedfig <- ggarrange(beta0.fig, betabetaIDHMe.fig,
385                       betabetaIDHMr.fig, betaappne.fig,
386                       betanascente.fig, betacda.fig,
387                       betara.fig, betat2.fig,
388                       betat15.fig, betat23.fig,
389                       nrow = 5, ncol = 2)
390
391 # exportando o grafico
392 ggsave("fixed_dist_N.png", plot = fixedfig,
393         width = 7, height = 7, dpi = "retina")
394
395
396 # grafico dos hiperparametros
397
398 # precisao marginal espacial

```

```

399 tby2 <- inla.sMarginal(m7$marginals.hyperpar$`Precision for cod`)
400 taub2 <- tibble(x = tby2$x, y = tby2$y)
401
402 taub2.fig <- ggplot(taub2) +
403   geom_line(aes(x = x, y = y)) +
404   labs(x = expression(tau[b2]), y = "Densidade") +
405   theme_classic2()
406
407 # proporcao espacial estruturada
408 pb2 <- inla.sMarginal(m7$marginals.hyperpar$`Phi for cod`)
409 phib2 <- tibble(x = pb2$x, y = pb2$y)
410 phib2.fig <- ggplot(phib2) +
411   geom_line(aes(x = x, y = y)) +
412   labs(x = expression(tau[b2]), y = "Densidade") +
413   theme_classic2()
414
415 # precisao temporal
416 tar1 <- inla.sMarginal(m7$marginals.hyperpar$`Precision for year`)
417 tauar1 <- tibble(x = tar1$x, y = tar1$y)
418
419 tauar1.fig <- ggplot(tauar1) +
420   geom_line(aes(x = x, y = y)) +
421   labs(x = expression(tau[ar1]), y = "Densidade") +
422   theme_classic2()
423
424 # correlacao temporal AR1
425 rar1 <- inla.sMarginal(m7$marginals.hyperpar$`Rho for year`)
426 rhoar1 <- tibble(x = rar1$x, y = rar1$y)
427
428 rhoar1.fig <- ggplot(rhoar1) +
429   geom_line(aes(x = x, y = y)) +
430   labs(x = expression(tau[ar1]), y = "Densidade") +
431   theme_classic2()
432

```

```

433 #
434
435 tint <- inla.sMarginal(m7$marginals.hyperpar$`Precision for cod.copy`
  `)
436 tauint <- tibble(x = tint$x, y = tint$y)
437
438 tauint.fig <- ggplot(tauint) +
439 geom_line(aes(x = x, y = y)) +
440 labs(x = expression(tau[int]), y = "Densidade") +
441 theme_classic2()
442 tauint.fig
443
444 rint <- inla.sMarginal(m7$marginals.hyperpar$`GroupRho for cod.copy`)
445 rhoint <- tibble(x = rint$x, y = rint$y)
446
447 rhoint.fig <- ggplot(rhoint) +
448 geom_line(aes(x = x, y = y)) +
449 labs(x = expression(rho[int]), y = "Densidade") +
450 theme_classic2()
451 rhoint.fig
452 #####
453
454 hiperfig <- ggarrange(taubym2.fig, phibym2.fig,
455 tauar1.fig, rhoar1.fig,
456 tauint.fig, rhoint.fig,
457 nrow = 3, ncol = 2)
458 ggsave("hiper_dist_N.png",
459 plot = hiperfig,
460 dpi = "retina")
461
462 # construcao do grafico de dispersao dos valores preditos versus
  observados
463
464 pxo <- ggplot(data.frame(y, y_star)) +

```

```

465     geom_point(aes(x = y_star, y = y)) +
466     geom_abline(slope = 1, intercept = 0,
467     linetype = "dashed") +
468     labs(x = "Preditos", y = "Observados") +
469     theme_classic2()
470
471 ggsave("pxo_num.png",
472     plot = pxo,
473     dpi = "retina")
474
475 #####
476 #           Modelo Hurdle Binomial - Gama           #
477 #####
478 # "Empilhando" os dados, construindo a matriz de respostas y
479
480 # z = 1 se ocorreu ao menos um auto de infração
481 # z = 0 caso contrário
482 z <- ifelse(data$n > 0, 1, 0)
483
484 # y = NA se z = 0 ou y = valor do auto se z = 1
485 y <- data$valor
486
487 (sdy <- sd(y))
488 y[data$valor == 0] <- NA
489 n <- length(z)
490
491 # matriz de repostas (duas colunas)
492 # coluna 1 corresponde a z, coluna 2 corresponde a y.
493 # para coluna 1, as n primeiras linhas recebe z, enquanto que
494 # as linhas de n+1 a 2n recebem NA. Para a segunda coluna,
495 # as linhas de 1 a n recebem NA, enquanto que as linhas de n+1 a 2n,
496 # recebem y. Dessa forma que o INLA diferencia quais observações da
497 # primeira verossimilhança (binomial correspondente a coluna 1) e da
498 # segunda verossimilhança (gama correspondente a coluna 2)

```

```

499 tmp <- rep(NA, 4 * n)
500 Y <- matrix(tmp, ncol = 2)
501 Y[1:n, 1] <- z
502 Y[(n + 1):(2*n), 2] <- y
503
504 # Para efeitos fixos e aleatorios, utiliza raciocinio semelhante.
505
506 # Intercepto
507 mu.z <- rep(1:0, c(n, n))
508 mu.y <- rep(0:1, c(n, n))
509 c(rep(0, n),
510 dados <- tibble(Y = Y,
511                 idhm.r.z = c(data$idhm.r.z, rep(0, n)),
512                 idhm.l.z = c(data$idhm.l.z, rep(0, n)),
513                 idhm.e.z = c(data$idhm.e.z, rep(0, n)),
514                 t1.z = c(data$t1, rep(0, n)),
515                 t2.z = c(data$t2, rep(0, n)),
516                 t3.z = c(data$t3, rep(0, n)),
517                 t4.z = c(data$t4, rep(0, n)),
518                 t5.z = c(data$t5, rep(0, n)),
519                 t6.z = c(data$t6, rep(0, n)),
520                 t7.z = c(data$t7, rep(0, n)),
521                 t8.z = c(data$t8, rep(0, n)),
522                 t9.z = c(data$t9, rep(0, n)),
523                 t10.z = c(data$t10, rep(0, n)),
524                 t11.z = c(data$t11, rep(0, n)),
525                 t12.z = c(data$t12, rep(0, n)),
526                 t13.z = c(data$t13, rep(0, n)),
527                 t14.z = c(data$t14, rep(0, n)),
528                 t15.z = c(data$t15, rep(0, n)),
529                 t16.z = c(data$t16, rep(0, n)),
530                 t17.z = c(data$t17, rep(0, n)),
531                 t18.z = c(data$t18, rep(0, n)),
532                 t19.z = c(data$t19, rep(0, n)),

```

```

533     t20.z= c(data$t20, rep(0,n)),
534     t21.z= c(data$t21, rep(0,n)),
535     t22.z = c(data$t22, rep(0,n)),
536     t23.z = c(data$t23, rep(0,n)),
537     nascente.z = c(data$nascente, rep(0,n)),
538     cursodagua.z = c(data$cursodagua, rep(0,n)),
539     lagoas.z      = c(data$lagoas, rep(0,n)),
540     veredas.z      = c(data$veredas, rep(0,n)),
541     resrv_artf.z = c(data$resrv_artf, rep(0,n)),
542     brejos.z       = c(data$brejos, rep(0,n)),
543     appne.z        = c(data$appne, rep(0,n)),
544     mu.z           = mu.z,
545     idhm.r.y = c(rep(0,n), data$idhm.r),
546     idhm.l.y = c(rep(0,n), data$idhm.l),
547     idhm.e.y = c(rep(0,n), data$idhm.e),
548     t1.y= c(rep(0,n), data$t1),
549     t2.y= c(rep(0,n), data$t2),
550     t3.y= c(rep(0,n), data$t3),
551     t4.y= c(rep(0,n), data$t4),
552     t5.y= c(rep(0,n), data$t5),
553     t6.y= c(rep(0,n), data$t6),
554     t7.y= c(rep(0,n), data$t7),
555     t8.y= c(rep(0,n), data$t8),
556     t9.y= c(rep(0,n), data$t9),
557     t10.y= c(rep(0,n), data$t10),
558     t11.y= c(rep(0,n), data$t11),
559     t12.y= c(rep(0,n), data$t12),
560     t13.y= c(rep(0,n), data$t13),
561     t14.y= c(rep(0,n), data$t14),
562     t15.y= c(rep(0,n), data$t15),
563     t16.y= c(rep(0,n), data$t16),
564     t17.y= c(rep(0,n), data$t17),
565     t18.y= c(rep(0,n), data$t18),
566     t19.y= c(rep(0,n), data$t19),

```



```

601         alpha = 0.01)),
602     phi = list(prior = "pc",
603               param = c(0.5, 0.5)))
604
605 hyperprior_ar1 <- list(theta1 = list(prior = "pc.prec",
606                                     param = c(u = 1,
607                                               alpha = 0.01))),
608                       theta2 = list(prior = "pc.cor0",
609                                     param = c(u = 0.5,
610                                               alpha = 0.5)))
611
612 # Definindo o preditor
613
614 f <- Y ~ 0 + mu.z + mu.y +
615     idhm.e.y + appne.y + t2.y +
616     t18.y + t23.y +
617     f(cod.z, model = "bym2", graph = adj,
618       constr = TRUE, scale.model = TRUE,
619       hyper = hyperprior_bym2) +
620     f(year.z, model = "ar1", hyper = hyperprior_ar1) +
621     f(cod.y, model = "bym2", graph = adj,
622       constr = TRUE, scale.model = TRUE,
623       hyper = hyperprior_bym2_gamma) +
624     f(year.y, model = "ar1",
625       hyper = hyperprior_ar1) +
626     f(cod.copy.y, model = "iid",
627       group = year.copy.y,
628       control.group = list(model = "ar1"))
629
630 # Ajuste do modelo
631 m15 <- inla(f, data = dados, family = c("binomial", "gamma"),
632            control.family = list(list(link = "logit"),
633                                  list(link = "log")),
634            control.inla = list(h = 0.0001,

```

```

635         strategy = "simplified.laplace",
636         int.strategy = "ccd"),
637         control.predictor = list(link = 1, compute = TRUE),
638         control.compute = list(waic = TRUE, cpo = TRUE,
639                                openmp.strategy="huge"))
640 # resumo do modelo
641 summary(m15)
642
643 # Media do LogScore
644 (lm15 <- -mean(log(m15$cpo$cpo), na.rm = TRUE))
645 # WAIC
646 (wm15 <- m15$waic$waic)
647 # RMSE gama
648 y_star <- m15$summary.fitted.values$mean[(n+1):(2*n)]
649 y      <- data$valor
650 (rmse <- sqrt(1/n * sum((y - y_star)^2)))
651
652 # R2
653 (R2 <- sum((mean(y) - y_star)^2) / sum((mean(y) - y)^2))
654
655 # R2 ajustado
656 k <- 1907.44 # n de parametros esperados
657 (R2a <- 1 - (n-1)/(n-k)*(1 - R2))
658
659 #
660 #####
661 #
662 #
663 #####
664
665 # Definindo as variaveis para representar as verossimilhanças

```

```

664 # w = binomial (ocorrencia)
665 # z = poisson  (numero de autos)
666 # y = gama      (valor dos autos)
667
668 w <- ifelse(data$n > 0, 1, 0)
669 y <- data$valor
670 sdy <- sd(y)
671 # para evitar dupla contagem, nas posicoes em que w = 1, z = y = NA
672 y[data$valor == 0] <- NA
673 z[data$valor == 0] <- NA
674 n <- length(z)
675
676 # "empilhando os dados" matriz 3*n x 3
677 # coluna 1, os primeiros n valores sao binomiais 0 ou 1, enquanto que
678 # de n+1 ate 3n, recebem NA.
679 # Coluna 2 recebe NA de 1 ate n, recebe o numero de autos de n+1 ate
        2n e
680 # recebe NA de 2n+1 ate 3n
681 # Coluna 3 recebe NA de 1 ate 2n e de 2n+1 ate 3n recebe o valor dos
        autos
682
683 tmp <- rep(NA, 3 * (3 * n))
684 Y <- matrix(tmp, ncol = 3)
685 dim(Y)
686 Y[1:n, 1] <- w
687 Y[(n + 1):(2*n), 2] <- z
688 Y[(2*n + 1):(3*n), 3] <- y
689
690 # para efeitos fixos e aleatorios, possuem raciocinios semelhantes
691 # Intercepto
692 mu.w <- rep(1:0, c(n, 2*n))
693 mu.z <- rep(c(0,1,0), each = n)
694 mu.y <- rep(0:1, c(2*n, n))
695

```



```

730 veredas.w = c(data$veredas,o,o),
731 resrv_artf.w = c(data$resrv_artf,o,o),
732 brejos.w = c(data$brejos,o,o),
733 appne.w = c(data$appne,o,o),
734 mu.w = mu.w,
735 idhm.r.z = c(o,data$idhm.r, o),
736 idhm.l.z = c(o,data$idhm.l,o),
737 idhm.e.z = c(o,data$idhm.e,o),
738 t1.z = c(o,data$t1,o),
739 t2.z = c(o,data$t2,o),
740 t3.z = c(o,data$t3,o),
741 t4.z = c(o,data$t4,o),
742 t5.z = c(o,data$t5,o),
743 t6.z = c(o,data$t6,o),
744 t7.z = c(o,data$t7,o),
745 t8.z = c(o,data$t8,o),
746 t9.z = c(o,data$t9,o),
747 t10.z = c(o,data$t10,o),
748 t11.z = c(o,data$t11,o),
749 t12.z = c(o,data$t12,o),
750 t13.z = c(o,data$t13,o),
751 t14.z = c(o,data$t14,o),
752 t15.z = c(o,data$t15,o),
753 t16.z = c(o,data$t16,o),
754 t17.z = c(o,data$t17,o),
755 t18.z = c(o,data$t18,o),
756 t19.z = c(o,data$t19,o),
757 t20.z = c(o,data$t20,o),
758 t21.z = c(o,data$t21,o),
759 t22.z = c(o,data$t22,o),
760 t23.z = c(o,data$t23,o),
761 nascente.z = c(o,data$nascente,o),
762 cursodagua.z = c(o,data$cursodagua,o),
763 lagoas.z = c(o,data$lagoas,o),

```

```

764 veredas.z = c(o, data$veredas, o),
765 resrv_artf.z = c(o, data$resrv_artf, o),
766 brejos.z = c(o, data$brejos, o),
767 appne.z = c(o, data$appne, o),
768 mu.z = mu.z,
769 idhm.r.y = c(o, o, data$idhm.r),
770 idhm.l.y = c(o, o, data$idhm.l),
771 idhm.e.y = c(o, o, data$idhm.e),
772 t1.y = c(o, o, data$t1),
773 t2.y = c(o, o, data$t2),
774 t3.y = c(o, o, data$t3),
775 t4.y = c(o, o, data$t4),
776 t5.y = c(o, o, data$t5),
777 t6.y = c(o, o, data$t6),
778 t7.y = c(o, o, data$t7),
779 t8.y = c(o, o, data$t8),
780 t9.y = c(o, o, data$t9),
781 t10.y = c(o, o, data$t10),
782 t11.y = c(o, o, data$t11),
783 t12.y = c(o, o, data$t12),
784 t13.y = c(o, o, data$t13),
785 t14.y = c(o, o, data$t14),
786 t15.y = c(o, o, data$t15),
787 t16.y = c(o, o, data$t16),
788 t17.y = c(o, o, data$t17),
789 t18.y = c(o, o, data$t18),
790 t19.y = c(o, o, data$t19),
791 t20.y = c(o, o, data$t20),
792 t21.y = c(o, o, data$t21),
793 t22.y = c(o, o, data$t22),
794 t23.y = c(o, o, data$t23),
795 nascente.y = c(o, o, data$nascente),
796 cursodagua.y = c(o, o, data$cursodagua),
797 lagoas.y = c(o, o, data$lagoas),

```

```

798 veredas.y = c(o,o,data$veredas),
799 resrv_artf.y = c(o,o,data$resrv_artf),
800 brejos.y = c(o,o,data$brejos),
801 appne.y = c(o,o,data$appne),
802 mu.y = mu.y,
803 cod.w      = c(data$cod, nd,nd),
804 cod.copy.w = c(data$cod, nd,nd),
805 year.w     = c(data$year, nd,nd),
806 year.copy.w = c(data$year, nd, nd),
807 cod.z      = c(nd, data$cod, nd),
808 cod.copy.z = c(nd, data$cod, nd),
809 year.z     = c(nd, data$year, nd),
810 year.copy.z = c(nd, data$year, nd),
811 cod.y      = c(nd, nd, data$cod),
812 cod.copy.y = c(nd, nd, data$cod),
813 year.y     = c(nd, nd, data$year),
814 year.copy.y = c(nd, nd, data$year),
815 id.w       = c(data$id,nd,nd),
816 id.z       = c(nd, data$id, nd),
817 id.y       = c(nd, nd, data$id),
818 id.wzy     = rep(data$id,3),
819 id         = 1:(3*n),
820 cod        = rep(data$cod,3),
821 year       = rep(data$year,3),
822 cod.copy1  = rep(data$cod,3),
823 year.copy1 = rep(data$year,3),
824 cod.copy2  = rep(data$cod,3),
825 year.copy2 = rep(data$year,3),
826 cod.copy3  = c(nd, rep(data$cod,2)),
827 year.copy3 = c(nd, rep(data$year,2)),
828 cod.copy4  = c(nd, rep(data$cod,2)),
829 r         = rep(1:3, each = n),
830 dd        = rep(data$dd,3)
831

```



```

866 # Definindo o preditor linear do modelo VI
867 f <- Y ~ -1 + rf +
868   idhm.r.z + idhm.e.z + t2.z + t15.z + t23.z +
869   nascente.z + cursodagua.z + appne.z + resrv_artf.z +
870   idhm.e.y + appne.y + t2.y + t18.y + t23.y +
871   f(cod.w, model = "bym2", graph = adj,
872     constr = TRUE, scale.model = TRUE,
873     hyper = hyperprior_bym2) +
874   f(cod.z, copy="cod.w", fixed = FALSE, hyper=lambda21) +
875   f(cod.y, copy = "cod.w", fixed = FALSE, hyper = lambda31) +
876   f(year.w, model = "ar1", hyper = hyperprior_ar1, constr = TRUE) +
877   f(year.z, copy = "year.w", fixed = FALSE, hyper = lambda22) +
878   f(year.y, copy = "year.w", fixed = FALSE, hyper = lambda32) +
879   f(cod.copy.z, model = "iid", group = year.copy.z,
880     control.group = list(model = "ar1")) +
881   f(cod.copy.y, model = "iid", group = year.copy.y,
882     control.group = list(model = "ar1"))
883
884
885 # Ajuste do modelo
886 m6 <- inla(f, data = dados, family = c("binomial",
887                                         "poisson",
888                                         "gamma"),
889           control.family = list(list(link = "logit"),
890                                 list(link = "log"),
891                                 list(link = "log")),
892           control.inla = list(h = 0.0001,
893                               strategy = "simplified.laplace",
894                               int.strategy = "ccd"),
895           #num.threads=20,
896           control.predictor = list(link = 1, compute = TRUE),
897           control.compute = list(waic = TRUE, cpo = TRUE,
898                                 openmp.strategy="huge"))
899 # resumo do modelo

```

```

900 summary(m6)
901 # media do log score
902 (lm6 <- -mean(log(m6$cpo$cpo), na.rm = TRUE))
903 #WAIC
904 (wm6 <- m6$waic$waic)
905
906 # RMSE p/ cada verossimilhanca
907 y_v <- data$valor
908 y_star_v <- m6$summary.fitted.values$mean[((2*n)+1):(3*n)]
909 y_p <- data$n
910 y_star_p <- m6$summary.fitted.values$mean[(n+1):(2*n)]
911 yb_v <- mean(y_v)
912 yb_p <- mean(y_p)
913 (pois_rmse <- sqrt(1/n * sum((y_star_p - y_p)^2)))
914 (gamma_rmse <- sqrt(1/n * sum((y_star_v - y_v)^2)))
915
916 # R2 e R2 ajustado numero
917 (R2_n <- sum((yb_p - y_star_p)^2) / sum((yb_p - y_p)^2))
918 k <- 2007.73
919 (R2a_n <- 1 - (n-1)/(n-k)*(1 - R2_n))
920
921 # R2 e R2 ajustado valor
922 (R2_v <- sum((yb_v - y_star_v)^2) / sum((yb_v - y_v)^2))
923 k <- 2007.73
924 (R2a_v <- 1 - (n-1)/(n-k)*(1 - R2_v))

```