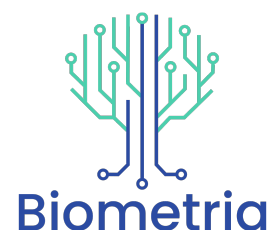




Universidade Estadual Paulista “Júlio de Mesquita Filho”
Instituto de Biociências – Câmpus de Botucatu
Programa de Pós-graduação em Biometria



Comparação no Desempenho dos Modelos Bayesianos Hierárquicos no Estudo da Evolução Temporal do Número de Óbitos por COVID-19 no Estado de São Paulo, 2020-2021

Juliana Aparecida Gualberto

Botucatu
2024

Juliana Aparecida Gualberto

Comparação no Desempenho dos Modelos Bayesianos Hierárquicos no Estudo da Evolução Temporal do Número de Óbitos por COVID-19 no Estado de São Paulo, 2020-2021

Tese de Doutorado apresentada ao Programa de Pós-graduação em Biometria da Universidade Estadual Paulista “Júlio de Mesquita Filho”.

Orientador: Prof. Dr. José Silvio Govone

Botucatu
2024

FICHA CATALOGRÁFICA ELABORADA PELA SEÇÃO TÉC. AQUIS. TRATAMENTO DA INFORM.
DIVISÃO TÉCNICA DE BIBLIOTECA E DOCUMENTAÇÃO - CÂMPUS DE BOTUCATU - UNESP

BIBLIOTECÁRIA RESPONSÁVEL: MARIA CAROLINA A. CRUZ E SANTOS-CRB 8/10188

Gualberto, Juliana Aparecida.

Comparação no desempenho dos modelos Bayesianos hierárquicos no estudo da evolução temporal do número de óbitos por COVID-19 no Estado de São Paulo, 2020-2021 / Juliana Aparecida Gualberto. - Botucatu, 2024

Tese (doutorado) - Universidade Estadual Paulista "Júlio de Mesquita Filho", Instituto de Biociências de Botucatu

Orientador: José Silvio Govone

Capes: 10203001

1. Análise Espaço-Temporal. 2. Autocorrelação (Estatística).
3. Espaço e tempo. 4. COVID-19 (Doença). 5. Análise Multinível.

Palavras-chave: Análises espaço-temporais; Autocorrelação; DIC;
Interação espaço-temporal; Modelos autoregressivos condicionais.

ATA DA DEFESA PÚBLICA DA TESE DE DOUTORADO DE JULIANA APARECIDA GUALBERTO, DISCENTE DO PROGRAMA DE PÓS-GRADUAÇÃO EM BIOMETRIA, DO INSTITUTO DE BIOCIÊNCIAS - CÂMPUS DE BOTUCATU.

Aos 19 dias do mês de fevereiro do ano de 2024, às 08:00 horas, no(a) Sala da Pós-graduação - IB, realizou-se a defesa de TESE DE DOUTORADO de JULIANA APARECIDA GUALBERTO, intitulada **Comparação no Desempenho dos Modelos Bayesianos Hierárquicos no Estudo da Evolução Temporal da Incidência de Mortes por COVID-19 no Estado de São Paulo, 2020-2021**. A Comissão Examinadora foi constituída pelos seguintes membros: Prof. Dr. JOSÉ SILVIO GOVONE (Orientador(a) - Participação Presencial) do(a) Departamento de Estatística Matemática Aplicada e Computação / Instituto de Geociências e Ciências Exatas (IGCE) Câmpus de Rio Claro UNESP, Prof.^a Dr.^a MIRIAM HARUMI TSUNEMI (Participação Presencial) do(a) Departamento de Biodiversidade e Bioestatística / Instituto de Biociências de Botucatu UNESP, Profa. Dra. LICIANA VAZ DE ARRUDA SILVEIRA (Participação Presencial) do(a) Departamento de Biodiversidade e Bioestatística / Instituto de Biociências de Botucatu - UNESP, Profa. Dra. GIOVANA FUMES GHANTOUS (Participação Virtual) do(a) Departamento de Ciências Básicas (ZAB) / Universidade de São Paulo Faculdade de Zootecnia e Engenharia de Alimentos Ciência e inovação da fazenda à mesa (FZEA) - USP Pirassununga, Profa. Dra. SÔNIA MARIA DE STEFANO PIEDADE (Participação Virtual) do(a) Departamento de Ciências Exatas / Escola Superior de Agricultura Luiz de Queiroz / USP. Após a exposição pela doutoranda e arguição pelos membros da Comissão Examinadora que participaram do ato, de forma presencial e/ou virtual, a discente recebeu o conceito final: Aprovada ___. Nada mais havendo, foi lavrada a presente ata, que após lida e aprovada, foi assinada pelo(a) Presidente(a) da Comissão Examinadora.

Prof. Dr. JOSÉ SILVIO GOVONE



Dedico este trabalho aos meus pais, que sempre acreditaram em mim e me apoiaram incondicionalmente ao longo de toda a minha jornada acadêmica. O amor, encorajamento e valores que vocês me transmitiram são a base desta conquista. Ao meu orientador, cuja orientação sábia e apoio constante foram fundamentais para o sucesso deste trabalho. Sua dedicação à excelência acadêmica e sua paixão pela pesquisa serviram como inspiração em cada etapa. E, com profunda tristeza, dedico esta tese às inúmeras vítimas da COVID-19 e às suas famílias. Que suas memórias sejam honradas, e que esta tese possa, de alguma forma, contribuir para a compreensão e enfrentamento de desafios de saúde pública como os que enfrentamos durante a pandemia.

Agradecimentos

A Deus, por ter abençoado todos os dias da minha vida, por iluminar meu caminho e me dar forças para seguir sempre em frente.

Ao Professor Doutor José Silvio Govone, que me acompanhou desde a Iniciação Científica até a conclusão do Doutorado. Sua orientação, amizade e mentoria ao longo desta jornada acadêmica foram inestimáveis. Expresso minha profunda gratidão por sua orientação e apoio, que foram fundamentais para o meu crescimento e sucesso.

Aos meus pais, Luiz e Maria, pelo apoio incondicional, torcida constante e pela confiança que depositaram em mim ao longo desta jornada.

Aos meus amigos, que se tornaram uma parte inseparável da minha vida durante a pós-graduação. Agradeço pelo apoio inabalável, incentivo, pelos momentos felizes e pela capacidade de tornar cada dia na pós-graduação uma experiência gratificante. Foi extremamente enriquecedor conhecer e conviver com cada um de vocês.

A todos os professores, funcionários e colegas do Programa de Pós-Graduação em Biometria da Universidade Estadual Paulista “Julio de Mesquita Filho” em Botucatu. Sua convivência tornou esta jornada verdadeiramente gratificante.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pelo apoio financeiro essencial que tornou este trabalho possível.

E a todas as demais pessoas que, direta ou indiretamente, contribuíram para a elaboração deste trabalho ou participaram da minha vida. Se, porventura, eu tenha me esquecido de agradecer a alguém, saibam que sua influência e apoio são profundamente valorizados e apreciados. Este trabalho não seria possível sem a contribuição de todos vocês. Muito obrigada!

*Nothing in life is to be feared, it is only to be understood. Now is the time to understand more, so
that we may fear less.*
Marie Cure

Resumo

Em dezembro de 2019, o surgimento do novo coronavírus (nCoV-2019 ou COVID-19) em Wuhan, China, desencadeou uma pandemia global que se disseminou rapidamente, afetando milhares de pessoas. A situação atingiu proporções alarmantes, levando a Organização Mundial da Saúde (OMS) a declarar, em 30 de janeiro de 2020, uma emergência de saúde pública de importância internacional, devido ao aumento exponencial de casos em todo o mundo, incluindo na China. O Brasil confirmou seu primeiro caso em 26 de fevereiro de 2020, na cidade de São Paulo, marcando o início da disseminação do vírus no país. Entre março de 2020 e dezembro de 2021, o Brasil registrou mais de 22 milhões de casos confirmados de COVID-19 e lamentáveis 600 mil mortes. No contexto da epidemiologia e análise de dados, a compreensão do comportamento da doença requer uma abordagem espaço-temporal pois permite identificar a configuração espacial da doença e sua evolução ao longo do tempo, facilitando a identificação de áreas de maior risco e embasando decisões em saúde pública. Nesse cenário, os modelos bayesianos hierárquicos têm se destacado como ferramentas valiosas para detectar aglomerados de doenças, desempenhando um papel crucial na identificação de áreas de maior risco. Além disso, a abordagem bayesiana oferece a vantagem de incorporar incertezas nos modelos e avaliar a qualidade das previsões, aspecto crucial em um contexto epidemiológico. Dada a existência de diversos modelos bayesianos hierárquicos, o objetivo deste estudo é avaliar se há um modelo superior para capturar a evolução espaço-temporal do número de óbitos decorrentes da COVID-19 no Estado de São Paulo. Para a modelagem espacial e temporal, foram empregados modelos amplamente utilizados na literatura, incluindo modelos autoregressivos condicionais (CAR) para efeitos espaciais e RW(1), RW(2) e AR(1) para efeitos temporais. Métricas como o Critério de Informação Deviance (DIC) e o Critério de Informação Watanabe-Akaike (WAIC) foram utilizadas para medir o desempenho e realizar comparações entre os modelos, sendo cruciais na avaliação da adequação dos modelos estatísticos e na determinação daqueles que oferecem as melhores previsões e *insights* para a compreensão e gestão da pandemia de COVID-19. Embora não tenha sido identificado um modelo que se destacasse dos demais, foi possível observar como os óbitos aumentaram ou diminuíram em diferentes períodos da pandemia. Este estudo evidencia a importância da análise espaço-temporal e do uso de modelos bayesianos hierárquicos na epidemiologia, contribuindo significativamente para a tomada de decisões em saúde pública e fornecendo uma ferramenta valiosa para enfrentar desafios como a pandemia de COVID-19.

Palavras-chave: Análises espaço-temporais, Autocorrelação, DIC, Interação espaço-temporal, Modelos autoregressivos condicionais, Poisson inflacionado de zero, WAIC.

Abstract

In December 2019, the emergence of the novel coronavirus (nCoV-2019 or COVID-19) in Wuhan, China, triggered a global pandemic that rapidly spread, affecting thousands of people. The situation reached alarming proportions, prompting the World Health Organization (WHO) to declare, on January 30, 2020, a public health emergency of international concern due to the exponential increase in cases worldwide, including in China. Brazil confirmed its first case on February 26, 2020, in the city of São Paulo, marking the onset of the virus's spread in the country. Between March 2020 and December 2021, Brazil recorded over 22 million confirmed cases of COVID-19 and a lamentable 600 thousand deaths. In the context of epidemiology and data analysis, understanding the disease's behavior requires a space-time approach as it allows for the identification of the spatial configuration of the disease and its evolution over time, facilitating the identification of higher-risk areas and informing public health decisions. In this scenario, hierarchical Bayesian models have stood out as valuable tools for detecting clusters of diseases, playing a crucial role in identifying higher-risk areas. Furthermore, the Bayesian approach offers the advantage of incorporating uncertainties into the models and evaluating the quality of predictions, a crucial aspect in an epidemiological context. Given the existence of various hierarchical Bayesian models, the objective of this study is to assess whether there is a superior model to capture the space-time evolution of the number of deaths due to COVID-19 in the State of São Paulo. For spatial and temporal modeling, widely used models in the literature, including conditional autoregressive (CAR) models for spatial effects and RW(1), RW(2), and AR(1) priors for temporal effects, were employed. Metrics such as the Deviance Information Criterion (DIC) and the Watanabe-Akaike Information Criterion (WAIC) were used to measure performance and make comparisons between models, being crucial in evaluating the adequacy of statistical models and determining those that offer the best predictions and insights for understanding and managing the COVID-19 pandemic. Although no model stood out from the others, it was possible to observe how deaths increased or decreased in different periods of the pandemic. This study underscores the importance of space-time analysis and the use of hierarchical Bayesian models in epidemiology, making a significant contribution to public health decision-making and providing a valuable tool to address challenges such as the COVID-19 pandemic.

Keywords: Space-time analysis, Autocorrelation, DIC, Space-time interaction, Conditional autoregressive models, Zero-inflated Poisson, WAIC.

Lista de figuras

Figura 1 – Representação simbólica do modelo de efeito principal. Os círculos denotam a independência a priori, enquanto os arredondados indicam a dependência a priori. Os pontos pequenos azuis indicam as observações no espaço-tempo (CRUZ, 2022).	29
Figura 2 – Interação do Tipo I (CRUZ, 2022).	30
Figura 3 – Interação do Tipo II (CRUZ, 2022).	31
Figura 4 – Interação do Tipo III (CRUZ, 2022).	32
Figura 5 – Interação do Tipo IV (CRUZ, 2022).	33
Figura 6 – Total de óbitos por mês por COVID-19 no período de 2020 e 2021 no Estado de São Paulo.	35
Figura 7 – Distribuição dos óbitos totais por COVID-19 no período de 2020 e 2021 nos municípios do Estado de São Paulo.	36
Figura 8 – Matriz de pesos espaciais por meio do estilo de vizinhança B (ou contiguidade <i>queen</i>) para o Estado de São Paulo.	38
Figura 9 – Matriz de pesos espaciais por meio do estilo de vizinhança M (ou contiguidade <i>rook</i>) para o Estado de São Paulo.	39
Figura 10 – Gráfico dos resíduos padronizados vs valores ajustados do modelo com a combinação de autocorrelação espacial de Besag, autocorrelação temporal RW(1) e interação tipo II	52
Figura 11 – Mapa das médias posteriori dos valores ajustados do número de óbitos por COVID-19 ao longo do tempo do modelo com a combinação de autocorrelação espacial de Besag, autocorrelação temporal RW(1) e interação tipo II. Continuação próxima página.	53
Figura 12 – Mapa das médias posteriori dos valores ajustados do número de óbitos por COVID-19 ao longo do tempo. À esquerda: modelo com autocorrelação espacial de Besag, autocorrelação temporal AR(1) com interação tipo II, covariáveis, e efeitos aleatórios não estruturados espaciais/temporais. À direita: modelo com autocorrelação espacial de Besag, autocorrelação temporal RW(1) com interação tipo II, covariáveis, e efeitos aleatórios não estruturados espaciais/temporais.	56

Lista de tabelas

Tabela 1	– Estatísticas descritivas para o número de óbitos por COVID-19 dos 645 municípios do Estado de São Paulo de 2020 e 2021.	37
Tabela 2	– Restrições de identificabilidade para os modelos espaço-temporais na Equação (2.40) (CRUZ, 2022).	47
Tabela 3	– Avaliação dos 48 modelos espaço-temporais ajustados para o número de óbitos por COVID-19 dos municípios do Estado de São Paulo de 2020 a 2021 considerando o DIC.	49
Tabela 4	– Avaliação dos 48 modelos espaço-temporais ajustados para o número de óbitos por COVID-19 dos municípios do Estado de São Paulo de 2020 a 2021 considerando o WAIC.	50
Tabela 5	– Estimativas e médias a posteriori do modelo ajustado com combinação de autocorrelação espacial de Besag, autocorrelação temporal RW(1) e interação tipo II, com intervalos de credibilidade de 95% para o número de óbitos por COVID-19 nos municípios do Estado de São Paulo de 2020 e 2021.	50

Sumário

	1 INTRODUÇÃO	1
1.1	Objetivos	2
	2 REVISÃO BIBLIOGRÁFICA	4
2.1	Inferência bayesiana	4
2.1.1	Modelos bayesianos hierárquicos	5
2.1.2	Família exponencial	6
2.2	Modelos lineares generalizados	7
2.2.1	Modelo regressão de Poisson	8
2.2.2	Modelos aditivamente generalizados	9
2.3	Metodologia INLA	9
2.4	Campos aleatórios gaussianos	11
2.4.1	Campos aleatórios gaussianos markovianos	13
2.5	Mapeamento espaço-temporal para dados de doenças	14
2.5.1	Dados de séries temporais	14
2.5.2	Dados espaciais	15
2.6	Modelos autoregressivos (AR)	16
2.6.1	Modelo AR(1)	16
2.6.2	Modelo RW(1)	18
2.6.3	Modelo RW(2)	19
2.7	Modelos para dados de área	20
2.8	Modelos autoregressivos condicionais (CAR)	21
2.8.1	Modelagem CAR	21
2.8.2	Modelagem ICAR	23
2.8.3	Modelagem de Leroux	24
2.9	Modelos espaço-temporais	25
2.9.1	Modelo de Bernardinelli	26
2.9.2	Modelo de interação espaço-tempo	28
	3 MATERIAIS E MÉTODOS	34
3.1	Banco de dados	34
3.2	Software utilizado	34
3.3	Análise descritiva	35
3.4	Matriz de vizinhança	38
3.5	Análise da evolução espaço-temporal	39
3.6	Métodos de avaliação dos modelos	47

	4 RESULTADOS	49
4.1	Discussão	55
	5 CONCLUSÃO	57
	Referências	59
	Apêndices	65
	APÊNDICE A – PRODUTO DE KRONECKER	66
	APÊNDICE B – CÓDIGO	67

1 Introdução

No início de dezembro de 2019, um novo agente viral do coronavírus, conhecido como SARS-CoV-2 (também referido como nCoV-2019 ou COVID-19), foi identificado após o registro de casos na cidade de Wuhan, localizada na província de Hubei, China ([WORLD HEALTH ORGANIZATION, 2020a](#)). O COVID-19 é uma infecção respiratória aguda de elevada transmissibilidade, com potencial gravidade. Rapidamente, o vírus se disseminou por outras partes da China e por diversos países do mundo. Em 30 de janeiro de 2020, o Comitê de Emergência da Organização Mundial da Saúde (OMS) declarou estado de emergência de saúde global em resposta ao crescente número de casos notificados na China e em outros países ([WORLD HEALTH ORGANIZATION, 2020b](#)).

O Brasil registrou seu primeiro caso confirmado de COVID-19 em 26 de fevereiro de 2020, na cidade de São Paulo. O paciente, um homem de 61 anos e residente da capital paulista, havia recentemente viajado para a Itália ([SAÚDE, 2020](#)). Desde então, de fevereiro de 2020 até dezembro de 2021, o Brasil registrou mais de 22 milhões de casos confirmados e mais de 600 mil óbitos.

No entanto, mesmo com o elevado número de casos e óbitos confirmados de COVID-19 no Brasil, há uma grande quantidade de subnotificações, conforme apontado por [BASTOS et al. \(2019\)](#): “os atrasos na notificação, especialmente aqueles cuja estrutura varia no tempo, distorcem a relação entre a incidência da doença relatada e a verdadeira incidência da doença”. O rastreamento em tempo real de pandemias é dificultado por atrasos na notificação, que podem ser atribuídos a atrasos na confirmação laboratorial, problemas logísticos, falta de infraestrutura e testes, entre outros fatores.

Diante desse cenário, a capacidade de prever a evolução da pandemia e adotar medidas de controle eficazes se torna crucial. O mapeamento espaço-temporal se apresenta como uma ferramenta analítica valiosa para essa finalidade, permitindo a análise espacial que revela a autocorrelação entre áreas próximas e destaca regiões com maiores incidências.

Os modelos espaço-temporais frequentemente empregam conceitos como modelos autoregressivos condicionais (CAR) para fatores espaciais e modelos autoregressivos como AR(1), modelo de Random Walk, RW(1) ou RW(2) para capturar a dinâmica temporal. Esses modelos são eficazes na identificação de aglomerados de doenças e fornecem informações essenciais para

a tomada de decisões.

No entanto, a escolha adequada do modelo estatístico e a avaliação de seu desempenho são etapas cruciais na análise de dados epidemiológicos. Métricas como o Critério de Informação Deviance (DIC) (SPIEGELHALTER et al., 2002) e o Critério de Informação Watanabe-Akaike (WAIC) (VEHTARI; GELMAN; GABRY, 2017) desempenham um papel fundamental na determinação de quais modelos oferecem as previsões mais precisas e *insights* relevantes para a compreensão e gestão da pandemia de COVID-19.

Nesse contexto, a abordagem bayesiana hierárquica desempenha um papel significativo em modelos espaço-temporais, permitindo a incorporação da variabilidade entre unidades espaciais e temporais. Essa abordagem possibilita a estimativa de hiperparâmetros que modelam a heterogeneidade espacial e temporal, bem como a inclusão de covariáveis que afetam a dinâmica da doença (MORAGA, 2019).

Além de sua utilidade na detecção de aglomerados de doenças, os modelos bayesianos hierárquicos desempenham um papel crucial na identificação de áreas de maior risco e auxiliando nas tomadas de decisões em saúde pública. Além disso, a abordagem bayesiana permite a inclusão de incertezas nos modelos e a avaliação da qualidade das previsões (GELMAN et al., 1995).

1.1 Objetivos

O objetivo principal deste estudo é realizar uma comparação entre modelos bayesianos hierárquicos, visando investigar a evolução espaço-temporal do número de óbitos causados pela COVID-19 no Estado de São Paulo, nos anos de 2020 e 2021. Este propósito será concretizado mediante a utilização de diversos modelos espaço-temporais amplamente empregados na literatura especializada. Os quais incorporam o conceito de Regressão Autoregressiva Condicional (CAR), conforme delineado na obra de BESAG (1974), bem como nas contribuições subsequentes de BESAG, YORK e MOLLIE (1991), LEROUX, LEI e BRESLOW (2000) e SIMPSON et al. (2017). Adicionalmente, serão adotados modelos autoregressivos, a saber, Modelo Autorregressivo de Ordem 1 (AR(1)), Modelo de Random Walk 1 e 2 (RW(1) e RW(2)), para a modelagem da estrutura temporal e diferentes interações espaço-temporais. Todos esses modelos possuirão uma estrutura bayesiana hierárquica.

A conjunção de modelagem bayesiana hierárquica com modelos espaço-temporais reveste-se de significativa relevância, especialmente em cenários que envolvem dados caracterizados tanto pela dimensão espacial quanto pela temporal. A combinação dessas abordagens proporciona uma notável flexibilidade e precisão na captura da dependência existente entre os dados em termos tanto espaciais quanto temporais. Trata-se de uma abordagem particularmente apropriada para analisar a dinâmica da pandemia de COVID-19, uma vez que permite a consideração das influências geográficas e temporais na disseminação da doença.

Para efetuar a comparação entre os modelos propostos e avaliar sua qualidade, serão empregadas as métricas Critério de Informação Deviance (DIC) e o Critério de Informação Watanabe-Akaike (WAIC). Estas métricas proporcionam uma base sólida para a avaliação do ajuste dos modelos aos dados, considerando tanto a capacidade preditiva quanto a complexidade do modelo.

Ao alcançar esses objetivos, espera-se contribuir para uma compreensão mais profunda da evolução da pandemia no contexto estadual, ao mesmo tempo em que orientará na seleção do modelo mais adequado para descrever com precisão os padrões da pandemia.

2 Revisão Bibliográfica

Neste capítulo, exploraremos uma série de tópicos fundamentais que constituem a base analítica deste trabalho, desempenhando um papel central na análise de dados que abrange tanto situações temporais quanto espaciais. Esta revisão bibliográfica estabelecerá o alicerce essencial para a compreensão das abordagens analíticas que serão examinadas nas seções subsequentes deste estudo. As subseções 2.1 a 2.4 abordarão a modelagem e a estimação bayesianas, enquanto a subseção 2.5 tratará do mapeamento espaço-temporal. As subseções 2.6 e 2.7 discutirão modelos para dados em séries temporais e modelos para dados espaciais, respectivamente, culminando na exploração de modelos espaço-temporais na subseção 2.8.

2.1 Inferência bayesiana

A inferência bayesiana constitui um método estatístico fundamentado nas contribuições de Thomas Bayes (1701 - 1761), Pierre-Simon Laplace (1749 - 1827), Harold Jeffreys (1891 - 1989) e Bruno de Finetti (1906 - 1985). Sua abordagem se estrutura em torno da associação de uma distribuição de probabilidade ao parâmetro desconhecido de interesse. O objetivo principal consiste em aprimorar o entendimento dos parâmetros em análise e reduzir a incerteza subjacente a eles, as quais se apresentam por meio de modelos probabilísticos criteriosamente delineados (FIENBERG, 2006).

Partimos da premissa de que o parâmetro desconhecido θ possui informações resumidas, sintetizadas pela distribuição de probabilidade $p(\theta)$, a qual pode ser aprimorada mediante a análise da distribuição amostral $p(\mathbf{y}|\theta)$ que delinea a relação entre a variável aleatória $\mathbf{y}$ e o parâmetro θ . A aplicação do teorema de Bayes (BAYES, 1763) é empregada como uma metodologia para mensurar esse aumento de informação,

$$p(\theta|\mathbf{y}) = \frac{p(\mathbf{y}|\theta)p(\theta)}{p(\mathbf{y})}, \quad (2.1)$$

em que $1/p(\mathbf{y})$, que não depende de θ , assume o papel de constante normalizadora de $p(\theta|\mathbf{y})$, para um valor fixo de $\mathbf{y}$, a função $p(\mathbf{y}|\theta)$ fornece a verossimilhança de cada um dos possíveis valores de θ e $p(\theta)$ é chamada distribuição a priori de θ (GELMAN et al., 1995).

A constante normalizadora ou probabilidade marginal $p(\mathbf{y})$ pode ser obtida pela seguinte integral:

$$p(\mathbf{y}) = \int p(\mathbf{y}, \boldsymbol{\theta}) d(\boldsymbol{\theta}) = \int p(\boldsymbol{\theta}) p(\mathbf{y}|\boldsymbol{\theta}) d(\boldsymbol{\theta}), \quad (2.2)$$

e como este termo não depende de $\boldsymbol{\theta}$ podemos omitir o termo $p(\mathbf{y})$ e a igualdade em (2.1) é substituída por uma proporcionalidade,

$$p(\boldsymbol{\theta}|\mathbf{y}) \propto p(\mathbf{y}|\boldsymbol{\theta}) p(\boldsymbol{\theta}), \quad (2.3)$$

o qual em palavras temos que:

$$\text{distribuição a posteriori} \propto \text{verossimilhança} \times \text{distribuição a priori},$$

ou seja, a densidade a posteriori (evidência atualizada) é proporcional ao produto da verossimilhança (evidência dos dados) pela priori (informação inicial). Esta forma simplificada é útil em problemas envolvendo estimação de parâmetros, uma vez que o denominador é apenas uma constante de normalização (GELMAN et al., 1995).

2.1.1 Modelos bayesianos hierárquicos

Os modelos bayesianos hierárquicos são uma extensão dos modelos bayesianos convencionais, que permitem abordar a heterogeneidade e a estrutura hierárquica presentes nos dados. São particularmente relevantes em situações em que os dados se originam de múltiplas unidades ou grupos, bem como quando existe incerteza quanto à variabilidade entre tais unidades. Esses modelos desempenham um papel crucial ao oferecer uma estrutura analítica que reflete as complexas interações entre unidades, permitindo a incorporação por exemplo de correlações espaciais, temporais e espaço-temporais, além de fornecer uma expressão formal de incerteza nas estimativas de risco (MORAGA, 2019).

Supondo as variáveis aleatórias X , Y e Z , é possível expressar a distribuição conjunta por meio de uma fatoração, ou seja, $p(X, Y, Z) = p(Z|Y, X)p(Y|X)p(X)$. Voltando à questão da inferência, a essência “hierárquica” desses modelos reside sob não conhecimento do parâmetro $\boldsymbol{\psi}$ que refere-se aos hiperparâmetros dos modelos hierárquicos, o que implica a existência de sua própria distribuição a priori, $p(\boldsymbol{\psi})$. Em cenários de modelagem de dados mais complexos, a abordagem visa simplificar o problema por meio da divisão em três níveis primários (SIMÕES, 2013):

- **1º nível:** Modela os dados observados com uma função de verossimilhança (distribuição de probabilidade) e especifica um modelo de processo que pode incluir, por exemplo, efeitos aleatórios espaciais, temporais e espaço-temporais:

$$p(\text{dados}|\text{processo}, \text{hiperparâmetros}) \rightarrow \text{distribuição do dado observado}, p(\mathbf{y}|\boldsymbol{\theta}, \boldsymbol{\psi});$$

- **2º nível:** Atribui distribuições a priori para os parâmetros do modelo:
 $p(\text{processo}|\text{hiperparâmetros}) \rightarrow$ modelo a priori para o conjunto de parâmetros do preditor linear, $p(\boldsymbol{\theta}|\boldsymbol{\psi})$;
- **3º nível:** Especifica hiperprioris para parâmetros de distribuições a priori do segundo nível:
 $p(\text{hiperparâmetros}) \rightarrow$ distribuições a priori dos hiperparâmetros, $p(\boldsymbol{\psi})$.

Normalmente, estamos interessados na distribuição do processo e nos hiperparâmetros, atualizados pela distribuição dos dados (verossimilhança), pela distribuição a posteriori, que é obtida pela regra de Bayes. Cuja a base da análise bayesiana hierárquica é dada por:

$$p(\boldsymbol{\theta}, \boldsymbol{\psi}|\mathbf{y}) \propto p(\mathbf{y}|\boldsymbol{\theta}, \boldsymbol{\psi})p(\boldsymbol{\theta}, \boldsymbol{\psi})p(\boldsymbol{\psi}). \quad (2.4)$$

Em palavras temos que:

$$p(\text{processo}, \text{hiperparâmetros}|\text{dados}) \propto p(\text{dados}|\text{processo}, \text{hiperparâmetros}) \times \\ p(\text{processo}|\text{hiperparâmetros}) \times p(\text{hiperparâmetros}).$$

A maior vantagem de modelar a distribuição condicional de dados, dado o verdadeiro processo e hiperparâmetros, é que ela permite simplificações substanciais da forma do modelo (GELMAN et al., 1995).

2.1.2 Família exponencial

Considere as variáveis aleatórias independentes $Y_1, \dots, Y_n$ cuja distribuição de probabilidade depende dos parâmetros $\boldsymbol{\theta}$ e $\boldsymbol{\phi}$. Segundo CRUZ (2022), “a variável aleatória pertence a família exponencial se puder ser escrita na forma:

$$f(y_i; \theta_i; \boldsymbol{\phi}) = \exp \left\{ \frac{1}{a(\boldsymbol{\phi})} [y_i \theta_i - b(\theta_i)] + c(y_i; \boldsymbol{\phi}) \right\},$$

em que a , b e c são funções que assumem valores em subconjuntos dos reais, o parâmetro $\boldsymbol{\phi}$ é o parâmetro de dispersão e $\boldsymbol{\theta}_i = (\theta_1, \dots, \theta_n)$ é o parâmetro canônico da distribuição. A função $a(\boldsymbol{\phi})$ é geralmente da forma $a(\boldsymbol{\phi}) = \boldsymbol{\phi} \cdot \omega$, em que ω é uma constante conhecida (DOBSON; BARNETT, 2018; MYERS, 2012). A média e a variância são expressas por:

$$E(Y_i) = \mu_i = \frac{db(\boldsymbol{\theta}_i)}{d\boldsymbol{\theta}_i},$$

$$Var(Y_i) = \frac{d^2b(\boldsymbol{\theta}_i)}{d\boldsymbol{\theta}_i^2} \cdot a(\boldsymbol{\phi})."$$

Entre os membros da família exponencial pode-se destacar os seguintes modelos de probabilidade: Bernoulli, Binomial, Geométrica, Poisson, Normal, Gama, Exponencial e t-student. Por exemplo, a distribuição de Poisson para variável aleatória Y é:

$$f(Y = y, \lambda) = \frac{\lambda^y \exp(-\lambda)}{y!},$$

em que y assume os valores $0, 1, 2, \dots$. Isso pode ser reescrito como:

$$f(y, \lambda) = \exp(y \log \lambda - \lambda - \log y!),$$

que está na forma canônica porque $a(y) = y$. Além disso, o parâmetro natural é $\log \lambda$.

2.2 Modelos lineares generalizados

Modelos lineares generalizados (GLM, do inglês *Generalized Linear Models*) são extensões de modelos de regressão simples e múltiplos, que são modelos mais básicos projetados para estabelecer a relação linear entre uma variável Y (variável dependente ou endógena) e uma variável X (variável independente ou “regressora”). A fim de aprimorar as técnicas estatísticas e obter resultados mais confiáveis, NELDER e WEDDERBURN (1972) e MCCULLAGH (1989) em seus trabalhos demonstraram que várias técnicas conhecidas podem ser unificadas em uma classe de modelos de regressão. Assim, os modelos GLM permitem maior flexibilidade na relação funcional entre a média da variável resposta e o preditor linear.

Geralmente em um problema de modelagem envolvendo GLM, deve-se definir (BIASOLI, 2005):

- O comportamento (distribuição) da variável resposta;
- As variáveis explicativas;
- A função de ligação que irá ligar as variáveis explicativas à variável resposta.

Considere que os valores $Y_i, i = 1, \dots, n$, sejam observações de uma variável aleatória que são independentes entre si. Ao incorporar os valores de $p < n$ variáveis explicativas $x_{i1}, \dots, x_{ip}$, temos que um GLM é composto pelos seguintes componentes (FOX, 2008):

- **Componente aleatório:** Em suma, é a variável resposta do modelo. Dado um conjunto de variáveis aleatórias independentes Y_i , onde a distribuição $f(\cdot)$ pertence à família exponencial, com médias μ_i e parâmetro de escala ϕ ;
- **Componente sistemático:** É o preditor linear do modelo, estabelecido entre as covariáveis e vetor de parâmetros, ou seja, as variáveis explicativas determinam o valor médio de Y_i por uma função linear das variáveis predictoras:

$$\eta_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip}.$$

Em forma matricial, $\boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta}$, em que $\mathbf{X} = (x_1, \dots, x_p)$ é uma matriz $n \times p$ e $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)$ é um vetor p -dimensional vetor de coeficientes de regressão;

- **Função de ligação:** Responsável por relacionar o valor esperado da componente aleatória μ_i e a estrutura linear do modelo η_i com a função densidade de probabilidade $f(\cdot)$ por meio de uma função invertível suave chamada função de ligação $g(\cdot)$ tal que $g(\mu_i) = \eta_i$ ou em forma matricial,

$$g(\boldsymbol{\mu}) = \boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta} \Rightarrow g^{-1}(\mathbf{X}\boldsymbol{\beta}) = \boldsymbol{\mu} = E(\mathbf{Y}),$$

em que $\mathbf{X}\boldsymbol{\beta}$ é um vetor n -dimensional das funções lineares do preditor $\boldsymbol{\eta}$. Em situações onde a função de ligação $g(\cdot)$ transforma o valor esperado do componente aleatório (μ_i), no parâmetro canônico (θ_i), ela é denominada, função de ligação canônica, por exemplo, a função de ligação canônica é log para Poisson.

2.2.1 Modelo regressão de Poisson

A regressão de Poisson é uma técnica estatística usada para modelar dados de contagem, onde a variável dependente é o número de ocorrências de um evento em um intervalo de tempo, espaço ou outra unidade de contagem. A distribuição de Poisson é definida como:

$$p(Y = y|\lambda) = \frac{\lambda^y \exp(-\lambda)}{y!}, \quad (2.5)$$

em que $p(Y = y)$ é a probabilidade do número de eventos ser igual a y e λ é o parâmetro que representa o valor esperado de Y . Prova-se que a média é $E(Y) = \lambda$ e a variância é $Var(Y) = \lambda$ (AGRESTI, 2012).

O conjunto de variáveis preditoras $\mathbf{X}$ está relacionado a λ por meio da função:

$$\log(\lambda) = \mathbf{X}'\boldsymbol{\beta}, \quad (2.6)$$

em que $\boldsymbol{\beta}$ é o vetor de parâmetros de regressão a serem estimados a partir dos dados (AGRESTI, 2012).

Sua função de ligação é reconhecida como um dos casos especiais dos GLMs de maior importância, devido sua capacidade de análise da associação existente entre dados qualitativos, mediante observações de duas ou mais variáveis (MCCULLAGH, 1989).

2.2.2 Modelos aditivamente generalizados

O modelo aditivo generalizado (GAM, do inglês *Generalized Additive Model*) é uma extensão do modelo GLM, sendo uma classe de modelos estatísticos que permitem modelar relações complexas entre variáveis dependentes e independentes de maneira flexível e interpretável.

Na especificação de um modelo estatístico, é comum igualar uma função da média ($g(\mu_i)$) da variável resposta ($y_i; i = 1, \dots, n$) a um preditor linear aditivo (η_i). Muitos modelos estatísticos se enquadram nessa classe geral. No entanto, quando essa condição não pode ser satisfeita, é possível usar uma relação apropriada conhecida de variância média para aplicar métodos de quase-verossimilhança. No contexto dos modelos GAM, o preditor linear segue uma estrutura na forma (HASTIE, 2017):

$$g(\mu_i) = \alpha + X_i\beta + f_1(x_{1i}) + f_2(x_{2i}) + f_3(x_{3i}) + \dots, \quad (2.7)$$

em que g é uma função de ligação monótona suave conhecida, α é o intercepto, $X_i\beta$ é a parte paramétrica do preditor linear, x_{ji} são variáveis do preditor e f_j são funções suaves. Ou seja, modelo aditivamente generalizado é semelhante ao modelo de regressão linear múltipla, porém a única diferença é que ele se ajusta a uma função não linear para cada variável preditora, permitindo modelar relacionamentos não lineares.

Além disso, esse modelo permite combinar várias covariáveis em uma única função suavizadora, do seguinte modo (HASTIE, 2017):

$$g(\mu_i) = \alpha + X_i\beta + f_1(x_{1i}) + f_2(x_{2i}) + f_3(x_{3i}, x_{4i}) + \dots$$

Exemplos de funções não lineares são Splines, Polinomiais e Funções de Etapa.

2.3 Metodologia INLA

A Aproximação Integrada de Laplace Aninhada (INLA, do inglês *Integrated Nested Laplace Approximation*) é uma metodologia computacional eficiente para conduzir inferência bayesiana em modelos hierárquicos (RUE; MARTINO; N.CHOPIN, 2009).

A implementação da metodologia INLA é realizada por meio do pacote R-INLA (RUE et al., 2013), desenvolvido em linguagem C, permitindo a especificação e o ajuste de modelos dentro do ambiente R (R CORE TEAM, 2021). Recursos como a documentação, exemplos e um fórum de discussão estão disponíveis para *download* no site <<http://www.r-inla.org/>> (RUE; MARTINO; N.CHOPIN, 2009).

O INLA é um método de inferência bayesiana aproximada que se destaca como uma alternativa computacionalmente valiosa aos métodos tradicionais, como Método de Monte Carlo via Cadeias de Markov (MCMC), devido à sua eficiência e à conveniência do pacote R-INLA. A

abordagem INLA capitaliza as suposições embutidas no modelo para gerar uma aproximação numérica da distribuição de probabilidade posterior de interesse, baseando-se na técnica de Laplace (RUE; MARTINO; N.CHOPIN, 2009).

A metodologia INLA usa uma subclasse especial de modelos de regressão aditivos estruturados propostos por RUE, MARTINO e N.CHOPIN (2009), que constituem uma grande classe de modelos (MORAGA et al., 2017), que é caracterizada por uma variável resposta Y_i com distribuição pertencente à família exponencial, com valor esperado μ_i relacionado com um preditor linear aditivo η_i por meio de uma função de ligação $g(\cdot)$, tal que $g(\mu_i) = \eta_i$. O preditor linear η_i pode ser expresso como:

$$\eta_i = \alpha + \sum_{j=1}^{\eta_\beta} \beta_j z_{ji} + \sum_{k=1}^{\eta_f} f^{(k)}(\mathbf{u}_{ki}) + \epsilon_i; i = 1, \dots, n, \quad (2.8)$$

em que, α é o intercepto, $\beta_j, j = 1, \dots, \eta_\beta$, coeficientes de algumas covariáveis $\{z_j\}_{j=1}^{\eta_\beta}$, funções $f^{(k)}$ definem n efeitos aleatórios em algum vetor de covariáveis $\{\mathbf{u}_k\}_{k=1}^{\eta_f}$ e ϵ_i são os termos não estruturados. Nesse modelo, todas as variáveis latentes, ou seja, $\alpha, \{f^{(k)}(\cdot)\}, \{\beta_j\}$ e $\{\epsilon_i\}$, são consideradas como tendo distribuições Gaussianas.

O principal interesse está na distribuição a posteriori de $(\boldsymbol{\theta}, \boldsymbol{\Psi})$, dado os valores observados $\mathbf{y}$, que pode ser obtida por através do teorema de Bayes:

$$\pi(\boldsymbol{\theta}, \boldsymbol{\Psi} | \mathbf{y}) = \frac{\pi(\mathbf{y} | \boldsymbol{\theta}, \boldsymbol{\Psi}) \pi(\boldsymbol{\theta}, \boldsymbol{\Psi})}{\pi(\mathbf{y})} \propto \pi(\mathbf{y} | \boldsymbol{\theta}, \boldsymbol{\Psi}) \pi(\boldsymbol{\theta}, \boldsymbol{\Psi}), \quad (2.9)$$

em que $\pi(\mathbf{y})$ representa a probabilidade marginal do modelo, que é uma constante de normalização. No entanto, INLA fornece uma aproximação precisa para esta quantidade (HUBIN; STORVIK, 2016).

O $\pi(\mathbf{y} | \boldsymbol{\theta}, \boldsymbol{\Psi})$ representa a verossimilhança e supondo que as observações $(y_1, \dots, y_n)$ são independentes, dados os efeitos latentes $\boldsymbol{\theta}$ e $\boldsymbol{\Psi}$, pode ser reescrita como:

$$\pi(\mathbf{y} | \boldsymbol{\theta}, \boldsymbol{\Psi}) = \prod_{i=1}^n \pi(y_i | \theta_i, \boldsymbol{\Psi}). \quad (2.10)$$

A distribuição conjunta dos efeitos aleatórios e dos hiperparâmetros, $\pi(\boldsymbol{\theta}, \boldsymbol{\Psi})$, pode ser reorganizada como $\pi(\boldsymbol{\theta} | \boldsymbol{\Psi}) \pi(\boldsymbol{\theta})$, com $\pi(\boldsymbol{\Psi})$ a distribuição a priori do conjunto de hiperparâmetros $\boldsymbol{\Psi}$. Como muitos dos hiperparâmetros são considerados como independentes a priori, é provável que a distribuição de $\boldsymbol{\Psi}$ seja o produto de Ψ_i priori (RUE; MARTINO; N.CHOPIN, 2009).

Portanto, supondo que $\boldsymbol{\theta}$ é um campo aleatório de gaussiano markoviano (GMRF, do inglês *Gaussian Markov Random Fields*), a distribuição posteriori dos efeitos latentes é:

$$\pi(\boldsymbol{\theta} | \boldsymbol{\Psi}) \propto |Q(\boldsymbol{\Psi})|^{\frac{1}{2}} \exp \left\{ -\frac{1}{2} \boldsymbol{\theta}^T Q(\boldsymbol{\Psi}) \boldsymbol{\theta} \right\}, \quad (2.11)$$

em que $Q(\Psi)$ é a matriz de precisão.

Considerando todas as distribuições, a distribuição posteriori conjunta dos efeitos latentes e hiperparâmetros pode ser escrita como:

$$\begin{aligned} \pi(\theta, \Psi | \mathbf{y}) &\propto \pi(\Psi) | Q(\Psi) |^{\frac{1}{2}} \exp\left\{-\frac{1}{2} \theta^T Q(\Psi) \theta\right\} \prod_{i=1}^n \pi(y_i | \theta_i, \Psi) = \\ &\pi(\Psi) | Q(\Psi) |^{\frac{1}{2}} \exp\left\{-\frac{1}{2} \theta^T Q(\Psi) \theta + \sum_{i=1}^n \log(\pi(y_i | \theta_i, \Psi))\right\}. \end{aligned} \quad (2.12)$$

A metodologia INLA é especialmente útil em modelos hierárquicos, nos quais há múltiplos níveis de parâmetros e incertezas. Ela lida com a heterogeneidade e a estrutura hierárquica dos dados. A mesma funciona em três níveis (BLANGIARDO; CAMELETTI, 2015):

Primeiro Nível: Modela os dados observados com uma função de verossimilhança que descreve a probabilidade dos dados condicionalmente aos parâmetros do modelo e aos hiperparâmetros.

Segundo Nível: Atribui distribuições a priori para os parâmetros do modelo. Isso representa o conhecimento prévio sobre os parâmetros antes de observar os dados.

Terceiro Nível: Especifica hiperprioris para os parâmetros das distribuições a priori do segundo nível. Isso representa o conhecimento prévio sobre os hiperparâmetros.

A metodologia INLA combina todas essas informações para calcular a distribuição posterior conjunta dos parâmetros e hiperparâmetros, levando em consideração todas as incertezas inerentes ao modelo. Em resumo, é uma abordagem eficiente para a inferência bayesiana em modelos complexos, permitindo que se obtenha estimativas precisas das distribuições de probabilidade posteriores, mesmo em cenários desafiadores. Ela é especialmente útil em modelos hierárquicos e tem sido amplamente adotada em várias disciplinas, incluindo estatística, epidemiologia e ciência de dados (RUE; MARTINO; N.CHOPIN, 2009).

2.4 Campos aleatórios gaussianos

A modelagem da variação espacial e temporal nos dados pode ser abordada de forma flexível por meio de um campo aleatório gaussiano (GRF), conhecido em inglês como *Gaussian Random Field*. Essa abordagem oferece uma maneira poderosa de capturar padrões complexos de variação, permitindo que o modelo se ajuste de maneira mais precisa aos dados observados. O GRF é um processo estocástico definido em um espaço $D \subset \mathbb{R}^k$, em que k é um número inteiro e D é um espaço fixo e contínuo. Ele pode ser representado por $X(s); s \in D$, onde s denota a posição no espaço, por exemplo, coordenadas geográficas de latitude e longitude (x, y) . Aqui, $X(s)$ representa a variável aleatória observada em cada posição s . Esta abordagem oferece

uma estrutura robusta para análise de dados geoespaciais, permitindo uma compreensão mais profunda dos padrões de variação e suas implicações (CRUZ, 2022).

Em aplicações geoestatísticas, é frequente encontrar dados parcialmente observados devido a limitações práticas que impedem a obtenção de observações em todos os pontos do domínio D . Nessas circunstâncias, para uma posição específica s , a variável $X(s)$ é tratada como uma variável aleatória, sujeita às leis probabilísticas estabelecidas (WALLER; GOTWAY, 2004; CRUZ, 2022). Essa abordagem reconhece a realidade da coleta de dados em ambientes geoespaciais, onde a disponibilidade de informações pode ser limitada, e permite a aplicação de métodos estatísticos adequados para análise e modelagem.

Dado um campo aleatório, se

$$E(X(s)) = \mu, \forall s \in D, \quad (2.13)$$

ou seja, a média do processo não depende da localização e

$$\text{Cov}(X(s_i), X(s_j)) = C(s_i - s_j), \forall s_i, s_j \in D, \quad (2.14)$$

em que a covariância depende apenas da diferença entre as localizações s_i e s_j , $X(\cdot)$ é dita estacionária de segunda ordem, e a função $C(\cdot)$ é chamada de função de covariância e é uma medida de autocorrelação espacial. Assim, devido à suposição de estacionaridade, o processo essencialmente se repete no espaço, proporcionando a replicação necessária para estimação e inferência. A função $C(s_i - s_j)$ é isotrópica, pois corresponde somente a distância e não a direção (CRUZ, 2022).

O campo aleatório $X(s)$, $s \in D$ é considerado um campo aleatório gaussiano quando, para todo $n \geq 1$ e quaisquer localizações $s_1, s_2, \dots, s_n \in D$, o vetor aleatório $(X(s_1), X(s_2), \dots, X(s_n))^T$ tem distribuição multivariada com vetor de médias $\mu = (\mu_1(s_1), \mu_2(s_2), \dots, \mu_n(s_n))^T$ e matriz de covariância:

$$\Sigma = \begin{bmatrix} C(s_1, s_1) & C(s_1, s_2) & \cdots & C(s_1, s_n) \\ C(s_2, s_1) & C(s_2, s_2) & \cdots & C(s_2, s_n) \\ \vdots & \vdots & \ddots & \vdots \\ C(s_n, s_1) & C(s_n, s_2) & \cdots & C(s_n, s_n) \end{bmatrix},$$

em que $C(s_i, s_j) = \text{Cov}(X(s_i), X(s_j))$. Assim, cada $X(s_i)$ é uma variável aleatória gaussiana univariada. A função de covariância deve ser positiva definida, ou seja, para qualquer conjunto de locais, a matriz de covariância é definida positiva (CRUZ, 2022).

2.4.1 Campos aleatórios gaussianos markovianos

A análise de campos aleatórios gaussianos markovianos (GMRF, do inglês *Gaussian Markov Random Field*) baseia-se na teoria dos grafos, onde os grafos não direcionados são utilizados para representar a estrutura de dependência entre as variáveis. Essa abordagem permite modelar e estudar a dependência espacial entre as observações de forma eficiente.

Segundo RUE e HELD (2005), um grafo dirigido G é o par $G = (V, E)$, em que V é o conjunto de nós e E é o conjunto de arcos ou arestas i, j , em que $i, j \in V$, e existe uma ligação dirigida entre os nós i e j . Caso contrário não há ligação entre os elementos i e j . Um grafo é completamente conectado se $i, j \in E$ para todo $i, j \in V$ com $i \neq j$.

Um dos fundamentos do INLA são os campos aleatórios gaussianos markovianos, que provém dos processos de Markov, que é o processo de crescimento de variáveis latentes do modelo de Markov oculto (HMM, do inglês *Hidden Markov Model*), e possui uma propriedade de independência condicional muito interessante.

O vetor aleatório $X(\mathbf{s})$, $\mathbf{s} \in D$ é um GMRF com relação ao grafo $G = (V, E)$ com média $\boldsymbol{\mu}$ e matriz de precisão $\mathbf{Q} = \boldsymbol{\Sigma}^{-1} > 0$, em que $\boldsymbol{\Sigma}$ é matriz de covariância, se e somente se sua densidade tem a forma:

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

em que

$$Q_{ij} \neq 0 \Leftrightarrow i, j \in E \quad \forall i \neq j,$$

ou seja, é um vetor normalmente distribuído com a propriedade de independência condicional associada.

Seja $\mathbf{X}$ normalmente distribuído com média $\boldsymbol{\mu}$ e precisão $\mathbf{Q}$. Então para $i \neq j$ temos:

$$X(\mathbf{s}_i) \perp X(\mathbf{s}_j) | X(\mathbf{s}_{-ij}) \Leftrightarrow Q_{ij} = 0,$$

em que $\perp$ significa independência, ou seja, podemos verificar em $\mathbf{Q}$ quando $X(\mathbf{s}_i)$ e $X(\mathbf{s}_j)$ são condicionalmente independentes, simplesmente verificando se a entrada Q_{ij} é zero. Temos que $\mathbf{Q}$ é uma matriz esparsa, ou seja, possui muitos zeros.

Isso significa que $Y(\mathbf{s}_i)$ e $Y(\mathbf{s}_j)$ são condicionalmente independentes e significa que a distribuição condicional da variável observada em algum nó depende apenas de seus vizinhos. Para uma abordagem completa do GMRF, recomendamos a consulta (RUE; HELD, 2005).

2.5 Mapeamento espaço-temporal para dados de doenças

O mapeamento espaço-temporal para dados de doenças é frequentemente usado, pois podem ser desenvolvidas funções para interpolar pontos, linhas e áreas, e usar estatísticas espaciais, temporais e técnicas de análise de rede para encontrar padrões que ajudem a identificar fatores de risco em regiões em determinados períodos de tempo, para que as intervenções de saúde pública possam ser direcionadas para regiões específicas.

De acordo com [ELLIOTT e WARTENBERG \(2004\)](#), os mapas de doenças fornecem resumos visuais rápidos de informações geográficas complexas e podem identificar padrões sutis em dados que são perdidos em representações tabulares. Embora visualmente e intuitivamente atraentes, os mapas de doenças precisam ser interpretados com cautela, pois padrões aparentes podem ser criados ou perdidos artificialmente, dependendo de como os dados mapeados são representados (por exemplo, número e limites das categorias) e escala ou resolução geográfica, mas a produção do mapa complementa a análise formal de dados epidemiológicos espaço-temporais e pode ter um impacto maior nos receptores do que nas estatísticas de acompanhamento ([REZAEIAN et al., 2004](#)).

Os dados espaço-temporais incorporam informações tanto sobre a localização geográfica quanto sobre a dimensão temporal. Em outras palavras, esses dados relacionam observações a lugares específicos em momentos específicos.

2.5.1 Dados de séries temporais

Dados de séries temporais são informações registradas sequencialmente ao longo do tempo, geralmente em intervalos fixos. Esses dados podem ser coletados em diversas áreas, como finanças, economia, meteorologia, medicina, entre outras.

Uma característica importante dos dados de séries temporais é que eles são dependentes do tempo, ou seja, a observação em um determinado momento é influenciada pelas observações anteriores e pode influenciar as observações futuras. Portanto, a sua análise requer técnicas estatísticas específicas que levem em conta essa dependência temporal ([HYNDMAN; ATHANASOPOULOS, 2018](#)).

Algumas das principais técnicas de análise de dados de séries temporais incluem a análise de tendências, que busca identificar se há uma mudança sistemática no padrão de dados ao longo do tempo; a análise de sazonalidade, que busca identificar padrões repetitivos em determinados períodos do ano; e a análise de autocorrelação, que busca identificar a relação entre uma observação e as observações anteriores ([SHUMWAY; STOFFER, 2000](#)).

A previsão de dados também é uma aplicação importante da análise de dados de séries temporais. Isso pode ser feito usando modelos estatísticos que levam em conta a estrutura temporal dos dados, como modelos média móvel integrada autoregressiva (ARIMA, do inglês

AutoRegressive Integrated Moving Average) e modelos de suavização exponencial. A previsão de dados de séries temporais é útil em muitas áreas, como previsão de demanda, previsão de preços e previsão de eventos climáticos (COWPERTWAIT; METCALFE, 2009).

A série temporal para os dados de óbitos por COVID-19, por exemplo, pode conter variações diárias, semanais, mensais e anuais significativas ao longo do período de tempo observado. Uma análise prática de interesse seria não apenas avaliar o período de pico de óbitos, mas também verificar a relação do número de casos com os óbitos em momentos particulares do tempo.

2.5.2 Dados espaciais

Dados espaciais são quaisquer tipos de dados que descrevem fenômenos associados a certas dimensões espaciais. Dados geográficos ou georreferenciados são dados espaciais em que as dimensões espaciais estão associadas à sua posição na superfície da Terra em um determinado momento ou período de tempo (CÂMARA et al., 1996).

Por várias razões, como a confidencialidade do paciente, os dados de saúde pública são frequentemente obtidos pela agregação de dados pontuais para sub-regiões da área de estudo, como municípios, áreas ponderadas, setores censitários, etc. Ao escolher o nível de agregação em que a pesquisa será conduzida, as unidades de análise são definidas de acordo com as questões indicadas na pesquisa: variáveis (como são formuladas), indicadores, propriedades das amostras e como são medidas, procedimentos analíticos, interpretação dos resultados e as possíveis inferências (RESENDES et al., 2006).

Os métodos estatísticos espaciais introduzidos por CRESSIE (1993) exploram diversas aplicações que podem ser abordadas por meio de métodos geoestatísticos. Contudo, nos últimos dez anos, ocorreram avanços significativos na área de computação, especialmente com o uso de métodos como o MCMC ou INLA, que possibilitaram a realização de análises completamente bayesianas de modelos multiníveis mais complexos em dados georreferenciados (BANERJEE; CARLIN; GELFAND, 2014).

Os dados geográficos têm propriedades geométricas e topológicas. As propriedades geométricas são propriedades métricas. Criam relacionamentos métricos a partir de recursos geométricos primitivos que representam a geometria sólida, como pontos, linhas e polígonos. Essas relações representam a medida dos recursos relativos ao sistema de coordenadas. A partir de formas geométricas, estabeleça propriedades geométricas como comprimento, curvatura e orientação de linhas; perímetro e área de superfície de polígonos, volume de sólidos 3D e forma e inclinação de linhas e polígonos (LAURINI; THOMPSON, 1992). As propriedades topológicas (não métricas) são baseadas na posição relativa dos objetos no espaço, como conectividade, direção (de, para), adjacência e contenção.

2.6 Modelos autoregressivos (AR)

São uma classe de modelos estatísticos amplamente utilizados para análise de séries temporais. Esses modelos pressupõem que o valor atual de uma variável depende linearmente de seus valores passados, ou seja, a variável é “autoregressiva”. Essa dependência temporal é modelada por meio de um ou mais termos autoregressivos, que são geralmente representados por “ $AR(p)$ ”, em que “ p ” é o número de valores passados que são usados como regressores.

Dizemos que X_t satisfaz um modelo autoregressivo de ordem p ou simplesmente $AR(p)$ se

$$X_t = \rho_1 X_{t-1} + \rho_2 X_{t-2} + \dots + \rho_p X_{t-p} + \epsilon_t, \quad (2.15)$$

em que X_t é estacionário, $\epsilon_t \sim N(0, \sigma_\epsilon^2)$ é um ruído branco (ou erro residual) e os pesos ρ_i da combinação linear são os parâmetros do modelo e são considerados constantes. Além disso, os $X_{t-1}, X_{t-2}, \dots, X_{t-p}$ são independentes de ϵ_t (HAMILTON, 2020).

Costumamos restringir os modelos autoregressivos a dados estacionários, caso em que algumas restrições sobre os valores dos parâmetros são necessárias (BOX et al., 2015; CRUZ, 2022):

- Para um modelo $AR(1)$: $|\rho| < 1$;
- Para um modelo $AR(2)$: $|\rho_2| < 1$, $\rho_1 + \rho_2 < 1$ e $\rho_2 - \rho_1 < 1$.

Quando $p \geq 3$, as restrições são bastante abstratas, mas o *software* R lida com elas ao estimar um modelo.

2.6.1 Modelo $AR(1)$

O modelo autoregressivo de primeira ordem, conhecido como modelo $AR(1)$, é amplamente utilizado em séries temporais devido a sua estrutura simples e sua ampla aplicação, podendo ser definido a partir da seguinte equação:

$$X_t - \rho X_{t-1} = \epsilon_t, t = 1, 2, \dots, T, \quad (2.16)$$

em que $X_1 = \epsilon_1, \epsilon_t \sim N(0, \sigma_\epsilon^2)$.

Temos que $E(X_t) = 0$ para todo t e a variância é dada por (BOX et al., 2015; CRUZ, 2022):

$$\text{Var}(X_t) = \text{Var}(\rho X_{t-1} + \epsilon_t) = \rho^2 \text{Var}(X_{t-1}) + \tau_\epsilon^{-1} = \rho^2 \text{Var}(X_t) + \sigma_\epsilon^2.$$

Então,

$$\text{Var}(X_t) = \frac{\sigma_\epsilon^2}{1 - \rho^2}, \quad |\rho| < 1.$$

A distribuição marginal desse processo para um conjunto finito de pontos no tempo $\mathbf{X} = X_1, \dots, X_T$ é obtida considerando $X_1 \sim N(0, \frac{\sigma_\epsilon^2}{1 - \rho^2})$ e $X_t | X_{t-1}, \tau_\epsilon \sim N(\rho X_{t-1}, \tau_\epsilon^{-1})$, em que σ_ϵ^2 é a variância do termo de erro, ρ é o parâmetro de autocorrelação e $\tau_\epsilon = 1/\sigma_\epsilon^2$ é a precisão do termo de erro. A densidade conjunta de $\mathbf{X}$ é expressa como (CRUZ, 2022):

$$\begin{aligned} \pi(\mathbf{X}) &= \pi(X_1) \cdot \pi(X_2 | X_1) \cdot \pi(X_3 | X_1, X_2) \cdot \dots \cdot \pi(X_T | X_1, X_2, \dots, X_{T-1}) \\ &= \pi(X_1) \cdot \pi(X_2 | X_1) \cdot \pi(X_3 | X_2) \cdot \dots \cdot \pi(X_T | X_{T-1}) \\ &= (2\pi\tau_\epsilon^{-1})^{-1/2} (1 - \rho^2)^{1/2} \exp\left\{-\frac{\tau_\epsilon}{2} (1 - \rho^2) y_1^2\right\} \cdot (2\pi\tau_\epsilon^{-1})^{-1/2} \exp\left\{-\frac{\tau_\epsilon}{2} (X_2 - \rho X_1)^2\right\} \\ &\quad \cdot (2\pi\tau_\epsilon^{-1})^{-1/2} \exp\left\{-\frac{\tau_\epsilon}{2} (X_3 - \rho X_2)^2\right\} \cdot \dots \cdot (2\pi\tau_\epsilon^{-1})^{-1/2} \exp\left\{-\frac{\tau_\epsilon}{2} (X_T - \rho X_{T-1})^2\right\} \\ &= (2\pi\tau_\epsilon^{-1})^{-T/2} \sqrt{1 - \rho^2} \exp\left\{-\frac{\tau_\epsilon}{2} [(1 - \rho^2) X_1^2 + (X_2 - \rho X_1)^2 + \dots + (X_T - \rho X_{T-1})^2]\right\}. \end{aligned}$$

A soma entre colchetes na exponencial pode ser escrita na forma matricial como (CRUZ, 2022):

$$\begin{aligned} &X_1^2 - \rho^2 X_1^2 + X_2^2 - 2\rho X_1 X_2 + \rho^2 X_1^2 + \dots + X_T^2 - 2\rho X_{T-1} X_T + \rho^2 X_{T-1}^2 \\ &= \sum_{i=1}^T X_i^2 - 2\rho \sum_{i=2}^T X_i X_{i-1} + \rho^2 \sum_{i=2}^{T-1} X_i^2 = \mathbf{X}^T \left[\mathcal{D}_T^{(1,1+\rho^2,1)} + \mathcal{T}_T^{(0,-\rho)} \right] \mathbf{X}, \end{aligned} \quad (2.17)$$

em que

$$\mathcal{D}_T^{(1,1+\rho^2,1)} = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 1 + \rho^2 & 0 & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \dots & 0 & 1 + \rho^2 & 0 \\ 0 & \dots & 0 & 0 & 1 \end{pmatrix}_{T \times T} \quad \text{e} \quad \mathcal{T}_T^{(0,-\rho)} = \begin{pmatrix} 0 & -\rho & 0 & \dots & 0 \\ -\rho & 0 & -\rho & \ddots & 0 \\ 0 & \ddots & \ddots & \ddots & 0 \\ \vdots & 0 & -\rho & 0 & -\rho \\ 0 & \dots & 0 & -\rho & 0 \end{pmatrix}_{T \times T}.$$

A matriz de covariância para uma série de tamanho T é (LINDSEY, 2004; WANG; FU; PAUL, 2022; CRUZ, 2022):

$$\Sigma = \frac{\tau_\epsilon^{-1}}{1 - \rho^2} \begin{pmatrix} 1 & \rho & \dots & \rho^{T-2} & \rho^{T-1} \\ \rho & 1 & \dots & \rho^{T-3} & \rho^{T-2} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \rho^{T-2} & \rho^{T-3} & \dots & 1 & \rho \\ \rho^{T-1} & \rho^{T-2} & \dots & \rho & 1 \end{pmatrix}.$$

A inversa dessa matriz, denominada matriz de precisão, tem uma forma tridiagonal (WADE; QUAAS, 1993; HODGES, 2013; CRUZ, 2022):

$$\mathbf{Q} = \Sigma^{-1} = \tau_\epsilon \begin{pmatrix} 1 & -\rho & 0 & 0 & 0 & \cdots & 0 \\ -\rho & 1 + \rho^2 & -\rho & 0 & 0 & \cdots & 0 \\ 0 & -\rho & 1 + \rho^2 & -\rho & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & -\rho & 1 + \rho^2 & -\rho & 0 \\ 0 & \cdots & 0 & 0 & -\rho & 1 + \rho^2 & -\rho \\ 0 & \cdots & 0 & 0 & 0 & -\rho & 1 \end{pmatrix} = \mathcal{D}_T^{(1, 1 + \rho^2, 1)} + \mathcal{T}_T^{(0, -\rho)},$$

em que $\mathbf{Q}$ possui uma decomposição simples de Cholesky (RUE, 2001), $\Sigma^{-1} = \mathbf{A}^\top \mathbf{A}$, no qual:

$$\mathbf{A} = \frac{1}{\sqrt{\tau_\epsilon}} \begin{pmatrix} \sqrt{1 - \rho^2} & 0 & \cdots & 0 & 0 \\ -\rho & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \\ 0 & 0 & \cdots & -\rho & 1 \end{pmatrix}.$$

A propriedade chave que faz com que $\mathbf{Q}$ seja esparsa está nas condicionais completas (CRUZ, 2022):

$$X_t \mid \mathbf{X}_{-t} \sim N \left(\frac{\rho}{1 - \rho^2} (X_{t-1} + X_{t+1}), \frac{\tau_\epsilon^{-1}}{1 + \rho^2} \right).$$

A dependência condicional entre os pontos no tempo é limitada aos dois pontos temporais mais próximos, o que explica a estrutura tridiagonal da matriz de precisão $\mathbf{Q}$. RUE e HELD (2005) demonstraram que a dispersão da matriz de precisão $\mathbf{Q}$ é preservada em sua decomposição de Cholesky, possibilitando uma avaliação rápida da densidade (linear em relação a T) e geração de números aleatórios. Em contraste, uma decomposição de Cholesky ou auto-decomposição da matriz de covariância (que é densa) possui uma complexidade computacional de ordem T^3 (ALLÉVIUS, 2018; CRUZ, 2022).

2.6.2 Modelo RW(1)

O modelo Random Walk 1 (RW(1)) é um método estatístico que é usado para prever valores futuros de uma série temporal univariada. Ele é chamado de *Random Walk* porque a previsão é baseada na suposição de que o próximo valor da série é igual ao valor atual mais um erro aleatório.

O modelo RW(1) é uma forma simples de modelo ARIMA, em que a parte integrada é definida como zero. Em outras palavras, o modelo não inclui nenhum termo de diferenciação, o

que significa que ele assume que a série é estacionária em torno da média. A equação do modelo autoregressivo RW(1) é dada por (BROCKWELL; DAVIS, 2002):

$$X_t - X_{t-1} = \epsilon_t, t = 1, 2, \dots, T, \quad (2.18)$$

A esperança do modelo RW(1) é dada por:

$$E(X_t) = E(X_{t-1} + \epsilon_t) = E(X_{t-1}) + E(\epsilon_t) = E(X_{t-1}).$$

Isso ocorre porque a média dos erros aleatórios é zero. Portanto, a expectativa de X_t é igual à expectativa de X_{t-1} .

A variância do modelo RW(1) é dada por:

$$Var(X_t) = Var(X_{t-1} + \epsilon_t) = Var(X_{t-1}) + Var(\epsilon_t).$$

Assumindo que o erro aleatório ϵ_t tem variância constante (σ^2), a variância da série temporal no tempo t é igual à variância da série temporal no tempo $t - 1$ mais a variância do erro aleatório.

Portanto, a variância do modelo RW(1) é crescente com o tempo, o que significa que a incerteza em torno da previsão aumenta à medida que avançamos no tempo. Isso torna o modelo RW(1) menos útil para prever valores distantes no futuro.

2.6.3 Modelo RW(2)

O modelo Random Walk 2 (RW(2)) também um método estatístico simples usado em análise de séries temporais, especificamente, se refere a uma série temporal em que cada valor é igual à média dos dois valores anteriores, mais um termo de erro.

A equação geral para um RW(2) é dada por (WEI, 2013):

$$X_t = \frac{X_{t-1} + X_{t-2}}{2} + \epsilon_t, \quad (2.19)$$

em que X_t é o valor da série temporal no tempo t , ϵ_t é um termo de erro estocástico, que representa a variação aleatória em cada ponto no tempo.

A esperança do modelo RW(2) é dada por:

$$E(X_t) = E\left(\frac{X_{t-1} + X_{t-2}}{2}\right) = \frac{1}{2}E(X_{t-1}) + \frac{1}{2}E(X_{t-2}).$$

Para variância do modelo RW(2) assumimos que ϵ_t tem uma variância constante σ^2 , a variância é dada por:

$$\text{Var}(X_t) = \text{Var}\left(\frac{X_{t-1} + X_{t-2}}{2}\right) + \sigma^2.$$

Este modelo é chamado de “ordem 2” porque ele depende dos dois valores anteriores, X_{t-1} e X_{t-2} . É importante notar que, devido à média dos valores anteriores, a série não está realmente “se movendo” no sentido tradicional; em vez disso, ela tende a oscilar em torno de uma média constante, com flutuações aleatórias. A principal diferença entre RW1 e RW2 é a quantidade de valores anteriores que são considerados para prever o próximo valor na série.

O modelo de RW2 é uma simplificação e é frequentemente usado como um modelo de base para comparar com modelos mais complexos em análise de séries temporais.

2.7 Modelos para dados de área

A análise da distribuição de variáveis nas áreas visa revelar padrões, tendências e relações subjacentes aos dados, permitindo uma compreensão mais profunda e informada do fenômeno em estudo. Ela desempenha um papel crucial na tomada de decisões, formulação de políticas, avanço científico e resolução de problemas complexos em diversas áreas do conhecimento.

Diversas abordagens são utilizadas para representar as relações espaciais entre áreas ou pontos fixos em um mapa, tais como a localização de casos ou os centroides de municípios, setores censitários e áreas de ponderação. Nesse contexto, uma matriz de pesos espaciais, denotada por W , desempenha um papel fundamental na modelagem de dados espaciais, pois captura as interações espaciais entre as regiões ou pontos georreferenciados com base em observações. A relação espacial entre essas regiões ou pontos constitui um aspecto crucial na análise dos dados, e a construção da matriz W permite expressar essas relações em termos quantitativos (ANSELIN, 1988). No entanto, ao contrário dos dados convencionais, no caso número de óbitos por exemplo, essa forma de dados frequentemente se apoia em suposições sobre a conectividade entre as regiões.

Vamos reformular o parágrafo:

Considere que i e j representem áreas vizinhas se possuírem uma fronteira em comum, indicada por $i \sim j$. Essa condição parece plausível quando as regiões são de tamanho uniforme e estão dispostas regularmente (WAKEFIELD, 2007).

Geralmente, os elementos da matriz W assumem valores binários, 0 ou 1. Em que w_{ij} é o elemento na linha i e coluna j ($i = 1, \dots, n$ e $j = 1, \dots, n$) e $w_{ij} = 1$, afirmamos que as regiões i e j são adjacentes uma à outra. Se $w_{ij} = 0$, indica que as regiões i e j não são vizinhas. Cada elemento diagonal w_{ii} em W é definido como 0 (para todo i), o que implica que uma região não pode ser adjacente a si mesma. Por exemplo, ao construir uma matriz de vizinhança W de dimensão $n \times n$, em que n representa o número de áreas (como setores censitários, áreas de ponderação, distritos, bairros, etc.), os elementos dessa matriz representam as relações de

dependência entre as áreas. Dessa forma, a representação do conjunto da região e/ou i -ésima região pode ser expressa como M_i . Em geral, temos a seguinte formulação (ANSELIN, 1988):

$$M_i = \{k; k \text{ é vizinho de } i\}, i = 1, \dots, n.$$

2.8 Modelos autoregressivos condicionais (CAR)

Os Modelos Autoregressivos Condicionais (CAR, do inglês *Conditional Autoregressive Models*) são uma extensão dos modelos autoregressivos para dados espaciais, que consideram a estrutura de dependência espacial entre as observações. Enquanto nos modelos AR as dependências são consideradas ao longo do tempo, nos modelos CAR as dependências são consideradas entre as observações em uma área geográfica específica. Esses modelos são usados para prever a variação espacial em uma variável de interesse, levando em conta as correlações espaciais entre as observações.

Os modelos CAR são amplamente utilizados em diversas áreas, como saúde pública, epidemiologia, ecologia, entre outras, para modelar a ocorrência de doenças ou eventos em uma determinada região geográfica. Eles são particularmente úteis quando as observações são influenciadas pelo seu entorno geográfico, como é o caso de dados espaciais (LAWSON, 2018).

Em resumo, os modelos autoregressivos condicionais permitem que os dados espaciais sejam modelados levando em conta as dependências espaciais, tornando-se uma ferramenta importante para a análise de dados espaciais em diversas áreas.

Modelos CAR têm sido amplamente utilizado para identificar e detectar *clusters* com risco de doença. Nesses modelos, o risco de doença em qualquer região é influenciado pelo risco nas regiões vizinhas. BESAG (1974) introduziu o CAR, mas apenas na última década houve um grande aumento no uso, devido a sua aplicação na amostragem de Gibbs e métodos mais gerais de Monte Carlo via cadeia de Markov (MCMC) para ajustar certas classes de modelos espaciais hierárquicos (LAWSON, 2018).

2.8.1 Modelagem CAR

O modelo de regressão espacial com parâmetros globais, conhecido como “modelo de erro espacial” (*spatial error model*) ou “autoregressivo condicional” (*conditional autoregressive model* (CAR)), têm sido amplamente utilizados na análise de dados espaciais em diversas áreas.

Modelos autoregressivos espaciais ou modelos auto-normais são modelos nos quais uma parte é covariável e outra parte é autoregressiva nos resíduos. Assim, tem-se que (BAILEY; GATRELL et al., 1995; CRESSIE, 1993):

$$Z = X\beta + e, \tag{2.20}$$

em que $\mathbf{Z}$ é um vetor das observações da variável dependente ($n \times 1$), $\mathbf{X}$ é a matriz ($n \times p + 1$) das variáveis preditoras, β é um vetor ($p + 1 \times 1$) dos coeficientes de regressão e e é dado por:

$$e = \rho \mathbf{W}, \quad (2.21)$$

sendo a autoregressão do erro ponderada pela matriz de vizinhança $\mathbf{W}$ e o parâmetro de dependência ρ , então a matriz de vizinhança irá “separar” as observações de seus vizinhos.

Seja Z_i e Z_j observações de subáreas, de forma que $i \neq j$, então tem-se que:

$$Z_i | Z_j, \mu, \sigma \sim N(\mu_i + \sum_{j \sim i} W_{ij}(Z_j - \mu_j), \sigma_i^2), \quad (2.22)$$

em que $i \sim j$ indica que as regiões i e j são vizinhas (normalmente definida em termos de adjacência espacial), μ_i representa a variável espacialmente autocorrelacionada, além disso, μ_i é o valor de resposta esperada (tal que $E(\mu_i) = \mu_i$), σ_i^2 é a variância condicional, e W_{ij} são pesos conhecidos ou desconhecidos (com $W_{ii} = 0$) que descreve a proximidade ou vizinhança entre os locais i e j .

Seja $\mathbf{Z}$ o vetor de observações, então a distribuição marginal é dada por:

$$\mathbf{Z} \sim NM(\mathbf{X}\beta, (\mathbf{I} - \rho\mathbf{W})^{-1}\sigma^2\mathbf{I}), \quad (2.23)$$

sendo $NM(\mathbf{X}\beta, (\mathbf{I} - \rho\mathbf{W})^{-1}\sigma^2\mathbf{I})$ uma distribuição normal multivariada.

Logo, tem-se que o logaritmo da verossimilhança marginal é dada por:

$$-\frac{n}{2} \log(2\pi\sigma^2) + \frac{1}{2} |\mathbf{I} - \rho\mathbf{W}| - \frac{1}{2\sigma^2} (\mathbf{Z} - \mu)' (\mathbf{I} - \rho\mathbf{W}) (\mathbf{Z} - \mu). \quad (2.24)$$

Para o modelo bayesiano hierárquico (CAR generalizado) tem-se que a verossimilhança é dada pela equação (LAWSON, 2018):

$$L(\mathbf{Z}_i | \beta, \sigma^2) = \int_e \prod_{i=1}^n f(Z_i | \beta, \sigma^2, e) f(e) de. \quad (2.25)$$

No caso da adição de mais uma variável latente i.i.d., a verossimilhança é dada por:

$$L(\mathbf{Z}_i | \beta, \sigma^2, v^2) = \int_e \int_{\mathbf{b}_k} \prod_{k=1}^n \prod_{i=1}^n f(Z_i | \beta, \sigma^2, \mathbf{b}_k, e) f(\mathbf{b}_k) f(e) d\mathbf{b}_k de. \quad (2.26)$$

Nota-se que $i = k$. Tem-se também que:

$$e \sim NM_n(0, \Sigma_{CAR}), \quad (2.27)$$

$$\mathbf{b}_k \sim N(0, v^2), \quad (2.28)$$

$$\Sigma_{CAR} = (\mathbf{I} - \mathbf{W})^{-1} \sigma^2 \mathbf{I}. \quad (2.29)$$

Seja o estimador bayesiano do modelo dado por:

$$f(\boldsymbol{\beta}, \sigma^2, v^2 | Z_i) = \frac{L(\mathbf{Z}_i | \boldsymbol{\beta}, \sigma^2, v^2) \pi(\boldsymbol{\beta}) \pi(\sigma^2) \pi(v^2)}{C(Z_i)}, \quad (2.30)$$

em que a constante $C(Z_i)$:

$$C(Z_i) = \int_{\boldsymbol{\beta}} \int_{\sigma^2} \int_{v^2} L(\mathbf{Z}_i | \boldsymbol{\beta}, \sigma^2, v^2) \pi(\boldsymbol{\beta}) \pi(\sigma^2) \pi(v^2) dv^2 d\sigma^2 d\boldsymbol{\beta}. \quad (2.31)$$

2.8.2 Modelagem ICAR

O modelo Intrínseco Condicionado Autoregressivo (ICAR, do inglês *Intrinsic Conditional Autoregressive*), proposto por [BESAG, YORK e MOLLIE \(1991\)](#), é um dos modelos Conditional Autoregressive (CAR) mais simples e amplamente utilizados na análise espacial de dados. Diferentemente de muitos modelos CAR, o ICAR possui uma matriz de variâncias que não é definida positiva, tornando-o um modelo impróprio para servir como verossimilhança direta para os dados. Em vez disso, o ICAR é empregado como uma distribuição a priori sobre os parâmetros do modelo. Referências a dois estudos clássicos, [BESAG \(1974\)](#) e [BESAG \(1975\)](#), demonstraram, utilizando o teorema da fatoração, que as distribuições condicionais completas na Equação (2.22), assumindo um campo aleatório markoviano gaussiano multivariado (GMRF), determinam unicamente a distribuição conjunta subsequente:

$$\mathbf{u} | \boldsymbol{\mu}, \boldsymbol{\sigma} \sim NMV \left(\boldsymbol{\mu}, (\mathbf{I}_n - \mathbf{B})^{-1} \mathbf{M} \right), \quad (2.32)$$

em que $\mathbf{B} = [b_{ij}]$ é uma matriz $n \times n$ definida por contiguidade ou distância, $\mathbf{I}$ é uma matriz identidade, e $\mathbf{M}$ é uma matriz diagonal definida por $M_{ii} = \sigma_i^2$. Esta distribuição conjunta é usada, juntamente com a função de verossimilhança do conjunto de dados para estimar as distribuições a posteriores de todos os parâmetros desconhecidos ([CRUZ, 2022](#)).

Para que a distribuição na equação (2.32) seja válida, é necessário que a matriz de covariância seja simétrica e positiva definida. Para atender a essa condição, são impostas restrições nas formas de $\mathbf{B}$ e $\mathbf{M}$. Comumente, assume-se que $\mathbf{B} = \rho \mathbf{W}$ e $\sigma_i^2 = (\sigma^2 / w_{i+})$, em que ρ é conhecido como coeficiente (ou parâmetro) de autocorrelação espacial ou dependência espacial, e $\mathbf{M}$ é a matriz de peso padronizada por linha, ou seja, $\mathbf{W} = [w_{ij}^*]$, com $w_{ij}^* = (w_{ij} / w_{i+})$ e w_{i+} é a soma da i -ésima linha de $\mathbf{W}$. Por construção, os elementos diagonais de $\mathbf{W}$ são todos zeros ([CRESSIE, 1993](#); [CRUZ, 2022](#)).

Essa estrutura de variância significa que, na presença de forte correlação espacial, quanto mais vizinhos a região tiver, mais informações haverá nos dados sobre seus valores de efeito aleatório. Além disso, o parâmetro de variância σ^2 controla a variação entre os efeitos aleatórios

estruturados (BANERJEE; CARLIN; GELFAND, 2014). Essa abordagem permite a ponderação de matrizes com base na continuidade e distância, mas não inclui um método de ponderação para o k -ésimo vizinho mais próximo, uma vez que esses pesos violam a condição de simetria. A abordagem mais comum é definir vizinhos de primeira ordem, em que os pesos são definidos de modo que $w_{ij} = 1$ se i e j compartilham fronteiras comuns, e $w_{ij} = 0$ caso contrário. Além disso, por construção, os elementos diagonais de $\mathbf{W}$ (w_{ii}) são todos zeros (CRESSIE, 1993; CRUZ, 2022).

De acordo com CRESSIE (1993), a matriz $(\mathbf{I}_n - \rho \mathbf{W})^{-1} \rho \mathbf{W}$ é uma matriz de covariância válida se $\xi_{\max}^{-1} < \rho < \xi_{\min}^{-1}$, em que $\xi_{\min}$ e $\xi_{\max}$ representam, respectivamente, os autovalores mínimo e máximo de $\mathbf{W}$.

O modelo ICAR aqui considerado apresenta algumas desvantagens. A força da dependência espacial entre os efeitos aleatórios é sempre considerada máxima (justamente por considerar $\rho = 1$), sendo o modelo adequado na presença de forte correlação espacial. Outro ponto é que o parâmetro de variância σ^2 é utilizado tanto para captar superdispersão quanto dependência espacial (LEROUX; LEI; BRESLOW, 2000).

2.8.3 Modelagem de Leroux

Outro modelo, foi proposto por LEROUX, LEI e BRESLOW (2000) é representado por uma distribuição normal multivariada:

$$\mathbf{u} \mid \boldsymbol{\sigma} \sim NM(\mathbf{0}, \sigma^2[\rho \mathbf{W} + (1 - \rho)\mathbf{I}]^{-1}). \quad (2.33)$$

Nesse caso, a matriz de precisão é $\mathbf{L} = \rho \mathbf{W} + (1 - \rho)\mathbf{I}$, em que $\mathbf{I}$ denota uma matriz identidade de ordem n e a matriz $\mathbf{W}$ é a mesma já definida. É fácil perceber que para $\rho = 0$, $\mathbf{L} = \mathbf{I}$ e tem-se um modelo com efeitos aleatórios independentes.

Já quando $\rho = 1$, $\mathbf{L} = \mathbf{W}$, ou seja, obtém-se o modelo intrínseco. Se $0 \leq \rho < 1$, a distribuição conjunta é própria. A distribuição condicional completa univariada é então dada por

$$u_i \mid u_{-i} \sim N\left(\frac{\rho}{n_i \rho + 1 - \rho} \sum_{j \sim i} u_j, \frac{\sigma^2}{n_i \rho + 1 - \rho}\right). \quad (2.34)$$

RODRIGUES (2011) e BANERJEE, CARLIN e GELFAND (2014) apontaram que os modelos CAR próprios podem induzir pouca correlação entre regiões vizinhas. Essa deficiência pode ser uma das razões pelas quais modelos intrínsecos e de convolução são as escolhas mais populares para acomodar os efeitos aleatórios espaciais.

2.9 Modelos espaço-temporais

Modelos espaço-temporais são modelos estatísticos usados para analisar dados que variam tanto no espaço quanto no tempo. Eles são usados em muitas áreas, incluindo ciências ambientais, epidemiologia, geografia, economia e muitas outras.

Os modelos espaço-temporais geralmente envolvem a combinação de modelos espaciais, que lidam com variações no espaço, e modelos temporais, que lidam com variações no tempo. Eles são capazes de levar em conta a dependência espacial e temporal nos dados, o que significa que eles podem modelar padrões complexos em dados que mudam ao longo do tempo e do espaço. Os mesmos também podem ser usados para analisar a relação entre variáveis em diferentes locais e momentos, bem como para avaliar a influência de fatores externos (CRESSIE; WIKLE, 2015).

Há diferentes modelos espaço-temporais, incluindo modelos autoregressivos integrados de médias móveis (ARIMA, *Autoregressive Integrated Moving Average*), modelos de espaço-tempo de efeitos fixos e aleatórios (STARMA, *Seasonal Autoregressive Moving Average*), modelos de espaço-tempo bayesianos e modelos de espaço-tempo de processos pontuais, entre outros. Cada tipo de modelo tem seus próprios pressupostos e propriedades, e a escolha do modelo certo depende dos dados e do problema específico em questão (CRESSIE; WIKLE, 2015).

Os modelos hierárquicos bayesianos espaço-temporais são uma classe de modelos que combinam técnicas bayesianas, hierárquicas e espaciais-temporais para modelar dados que variam no tempo e no espaço, levando em conta a estrutura hierárquica dos dados. Esses modelos são usados para modelar a dependência espacial e temporal dos dados, bem como a variação nos parâmetros do modelo em diferentes locais e momentos. Eles permitem que diferentes níveis de variação sejam modelados, desde variações espaciais e temporais aleatórias até variações sistemáticas explicadas por covariáveis (VANEM; BITNER-GREGENSEN; WIKLE, 2013).

Em geral, os modelos bayesianos hierárquicos espaço-temporais são compostos por uma estrutura hierárquica de três níveis que modelam diferentes fontes de variação nos dados (BERLINER, 1996). Nesse contexto, o termo hierárquico é usado para descrever a estrutura dos modelos bayesianos: o primeiro nível modela os dados observados com uma função de verossimilhança (distribuição de probabilidade) e especifica um modelo de processo que inclui efeitos aleatórios espaciais, temporais e espaço-temporais; o segundo nível atribui distribuições a priori para os parâmetros do modelo; e o terceiro nível especifica hiperpriors para parâmetros de distribuições a priori de segundo nível (por exemplo, variâncias). Distribuições a priori quantificam a incerteza do pesquisador no parâmetro (GELMAN et al., 2013).

A vantagem dos modelos bayesianos hierárquicos espaço-temporais é que eles permitem incorporar informações prévias sobre os parâmetros do modelo, o que é particularmente importante em cenários onde há poucos dados disponíveis. Eles também fornecem uma estrutura para lidar com incertezas em todos os níveis do modelo, permitindo que sejam feitas inferências mais precisas sobre os dados (VANEM; BITNER-GREGENSEN; WIKLE, 2013).

Por exemplo, o modelo mais utilizado na epidemiologia é Poisson:

$$y_{it} \mid \mu_{it} \sim \text{Poisson}(\mu_{it}),$$

$$\mu_{it} = E_{it}\psi_{it},$$

na região i e no momento t , em que E_{it} é o número esperado e ψ_{it} o risco relativo, em que no Poisson são geralmente usados em análises espaciais e espaço-temporais com dados de contagem (HAINING; LI, 2020).

A estrutura do modelo espaço-temporal separável leva em consideração a variabilidade espaço-temporal observada em dois componentes de modelagem: o componente espacial geral que descreve o padrão espacial, permanecendo relativamente estável ao longo do tempo e o componente temporal geral que captura o padrão geral de tendência.

Dado que tanto a dimensão espacial quanto a temporal possuem atributos únicos que não podem ser completamente explicados por covariáveis, é prudente incorporar efeitos aleatórios latentes para cada área e período no modelo. Essa abordagem visa mitigar a heterogeneidade intrínseca dos dados. Adotando a suposição de dados de Poisson, o modelo espaço-temporal é desdobrado de maneira aditiva em componentes que abrangem espaço, tempo e covariáveis:

$$\eta_{it} = \log(\psi_{it}) = \alpha + \beta_{it}x_{it} + S_i + T_t, \quad (2.35)$$

em que o parâmetro α quantifica a taxa média em todas as áreas e β é o vetor dos coeficientes de regressão $\beta = (\beta_{1it}, \dots, \beta_{pit})'$ associado ao vetor x das covariáveis, $S = (S_1, \dots, S_i)$ é a componente espacial global e T é a componente temporal global (HAINING; LI, 2020; CRUZ, 2022).

2.9.1 Modelo de Bernardinelli

O modelo espaço-temporal de Bernardinelli é utilizado para analisar a distribuição espacial e temporal de eventos, como casos de doenças, acidentes ou outras ocorrências. Ele foi desenvolvido pelo estatístico italiano Alessandro Bernardinelli em 1995. O objetivo é identificar a existência de agrupamentos de eventos em uma determinada região e período de tempo, além de avaliar se esses agrupamentos são estatisticamente significativos.

Este é um modelo paramétrico no qual a função linear (ou quadrática) do tempo é utilizada para determinar o logaritmo do risco em uma determinada área. Os coeficientes nessa função são específicos para cada área e possuem uma estrutura espacial que permite áreas vizinhas evoluírem de forma semelhante (LÓPEZ-QUILEZLEZ; MUNOZ, 2009).

BERNARDINELLI et al. (1995) sugeriu um modelo em que tanto o intercepto específico da área quanto a tendência temporal são modelados como efeitos aleatórios. Essa formulação permite interações espaço-temporais, em que a tendência temporal do risco pode ser diferente

para diferentes localizações espaciais e pode até ter uma estrutura espacial (LAWSON; BROWNE; RODEIRO, 2003; CRUZ, 2022).

Em resumo, este modelo é capaz de modelar as tendências temporais do risco em diferentes áreas de forma personalizada, levando em consideração a estrutura espacial dessas áreas e permitindo a detecção de interações espaço-temporais.

Utilizando o modelo proposto por BERNARDINELLI et al. (1995), e considerando $Y_{it} \sim \text{Poisson}(\mu_{it})$, o preditor linear para o componente temporal pode ser expresso da seguinte forma (CRUZ, 2022):

$$\log(\mu_{it}) = \alpha + u_i + v_i + (\beta + \delta_i)t, \quad (2.36)$$

em que u_i e v_i são respectivamente, efeitos aleatórios espaciais estruturados e não estruturados da área i , em que o efeito não estruturado refere-se a uma componente ou variação espacial que não é modelada explicitamente pelo componente de estrutura espacial, β_t representa o efeito de tendência temporal linear global, enquanto $\delta_i t$ representa a interação entre espaço e tempo, que define a diferença entre β e a tendência temporal específica da área, também conhecida como tendência diferencial da i -ésima área. O intercepto indica a taxa média em todas as áreas de ponderação. Cada unidade espacial apresenta sua própria tendência temporal com um intercepto espacial $\alpha + u_i + v_i$ e uma inclinação $\beta + \delta_i$. Consequentemente, o risco relativo para a i -ésima área e tempo t pode ser determinado pela equação proposta (CRUZ, 2022):

$$\psi_{it} = \exp(\alpha + u_i + v_i + (\beta + \delta_i)t). \quad (2.37)$$

Para a estimação dos parâmetros, consideramos que a função de verossimilhança é:

$$L(\mathbf{y}, \mathbf{E}, \boldsymbol{\psi} \mid \mathbf{u}, \mathbf{v}, \boldsymbol{\delta}, \beta, \tau_u, \tau_v, \tau_\delta), \quad (2.38)$$

em que os parâmetros τ_u , τ_v e τ_δ representam o efeito espacial, temporal e a interação espaço-tempo, respectivamente (CRUZ, 2022).

Portanto, a distribuição a posteriori é dada por:

$$p(\mathbf{u}, \mathbf{v}, \boldsymbol{\delta}, \beta, \tau_u, \tau_v, \tau_\delta \mid \mathbf{y}, \mathbf{E}, \boldsymbol{\psi}) \propto L(\mathbf{y}, \mathbf{E}, \boldsymbol{\psi} \mid \mathbf{u}, \mathbf{v}, \boldsymbol{\delta}, \beta, \tau_u, \tau_v, \tau_\delta) p(\mathbf{u})p(\mathbf{v})p(\boldsymbol{\delta})p(\beta). \quad (2.39)$$

Embora o modelo espaço-temporal de Bernardinelli seja útil para analisar a distribuição espacial e temporal de doenças, é importante ter em mente suas limitações e considerar outras técnicas estatísticas que possam ser mais adequadas para dados específicos. Uma limitação do modelo é a suposição de uma tendência temporal linear em cada unidade espacial, mas esta limitação foi resolvida pelo modelo de KNORR-HELD (2000) descrito a seguir (CRUZ, 2022).

2.9.2 Modelo de interação espaço-tempo

O modelo de interação espaço-tempo (STIM, do inglês *Space-Time Interaction Model*) é uma abordagem estatística para estudar a evolução temporal de um fenômeno em um espaço geográfico. O objetivo principal é entender como as interações entre unidades espaciais adjacentes influenciam a evolução temporal do fenômeno em questão.

O STIM parte da premissa de que as unidades espaciais adjacentes são interdependentes e que as mudanças em uma unidade podem afetar as unidades vizinhas. Essas interações podem ocorrer em diferentes escalas de tempo e espaço, e a abordagem estatística é usada para modelar essas relações. O modelo é baseado em uma série de equações que descrevem como as unidades espaciais interagem entre si ao longo do tempo. Essas equações geralmente levam em consideração fatores como a distância entre as unidades, as características geográficas da região, a densidade populacional, entre outros (KNORR-HELD, 2000).

KNORR-HELD (2000) introduziu este modelo com o intuito de superar a restrição do modelo paramétrico (2.37) proposto por Bernardinelli (1995), que limitava a linearidade na tendência temporal diferencial. Para lidar com essa limitação, Knorr especificou uma tendência temporal dinâmica não paramétrica. Além disso, para melhor explicar as variações nas tendências temporais das doenças entre diferentes áreas geográficas, é necessário expandir o modelo não paramétrico anterior, permitindo uma interação mais flexível entre espaço e tempo. Nessa ampliação, o logaritmo do risco é formulado como (CRUZ, 2022):

$$\log(\psi_{it}) = \alpha + u_i + v_i + \phi_t + \gamma_t + \delta_{it}, \quad (2.40)$$

em que a equação utiliza o parâmetro α para quantificar o logaritmo do risco médio em todas as áreas. Os efeitos espaciais estruturados e não estruturados para cada área i são representados por u_i e v_i , enquanto ϕ_t e γ_t representam os efeitos temporais estruturados e não estruturados para o tempo t . Esses componentes são considerados separadamente para levar em conta os efeitos espaciais e temporais. Além disso, o componente δ_{it} captura a tendência diferencial que explica a interação espaço-tempo entre as tendências temporais em diferentes locais (BLANGIARDO; CA-MELETTI, 2015; CRUZ, 2022). Para fins computacionais, os componentes espaciais/temporais ($u_i, v_i, \phi_t, \gamma_t$ e δ_{it}) são modelados como uma distribuição gaussiana multivariada com média zero e matriz de precisão $\tau \mathbf{R}$, em que τ é um escalar a ser estimado e $\mathbf{R}$ é uma matriz de estrutura conhecida e diferente para cada componente (GILKS; RICHARDSON; SPIEGELHALTER, 1995; KNORR-HELD, 2000; CRUZ, 2022).

Para produzir o efeito espacial estruturado u_i , é utilizada uma matriz $\mathbf{R}_u$ que possui elementos não diagonais iguais a -1 para as áreas vizinhas e elementos diagonais iguais ao número de áreas adjacentes. Todos os outros elementos em $\mathbf{R}_u$ são zero (KNORR-HELD, 2000). Para exemplificar, se considerarmos quatro áreas numeradas de 1 a 4, e apenas as áreas 1 e 4 forem vizinhas, a matriz de estrutura espacial neste caso será (CRUZ, 2022):

$$\mathbf{R}_u = \begin{bmatrix} 1 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 1 \end{bmatrix}. \quad (2.41)$$

Para modelar o efeito aleatório temporal estruturado γ_t , adotamos um processo auto-regressivo de primeira ordem com base em quatro pontos no tempo. Nesse caso, a matriz de estrutura que descreve a dependência temporal é expressa como (CRUZ, 2022):

$$\mathbf{R}_\gamma = \begin{bmatrix} 1 & -\rho & 0 & 0 \\ -\rho & 1 + \rho^2 & -\rho & 0 \\ 0 & -\rho & 1 + \rho^2 & -\rho \\ 0 & 0 & -\rho & 1 \end{bmatrix}. \quad (2.42)$$

As matrizes identidade I são empregadas como matrizes de estrutura para os efeitos aleatórios espaciais e temporais não estruturados v_i e ϕ_t , respectivamente. Essa escolha é baseada na suposição de que tais efeitos são independentes e identicamente distribuídos (i.i.d.). Uma representação visual do modelo de efeitos principais é vista na Figura 1 (CRUZ, 2022).

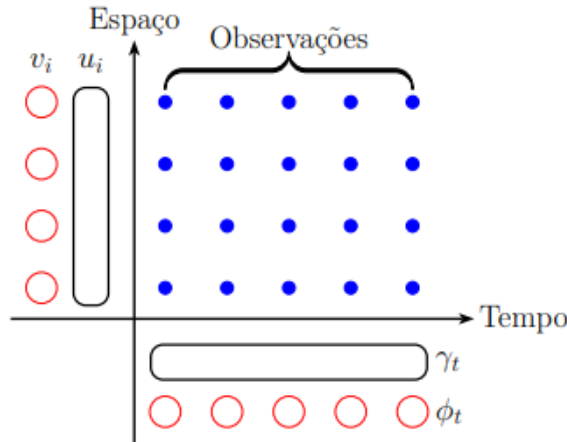


Figura 1 – Representação simbólica do modelo de efeito principal. Os círculos denotam a independência a priori, enquanto os arredondados indicam a dependência a priori. Os pontos pequenos azuis indicam as observações no espaço-tempo (CRUZ, 2022).

Como existem duas componentes tanto no efeito espacial quanto no temporal, há quatro tipos diferentes de interações (KNORR-HELD, 2000):

Interação tipo I

Na interação do tipo I, presume-se que os efeitos principais não estruturados espacial e temporal (v_i e γ_t) interajam, sem gerar qualquer forma de estrutura no espaço ou no tempo, conforme ilustrado na Figura 2. Isso significa que não há padrões discerníveis de variação espacial ou temporal associados à interação entre esses efeitos (KNORR-HELD, 2000).

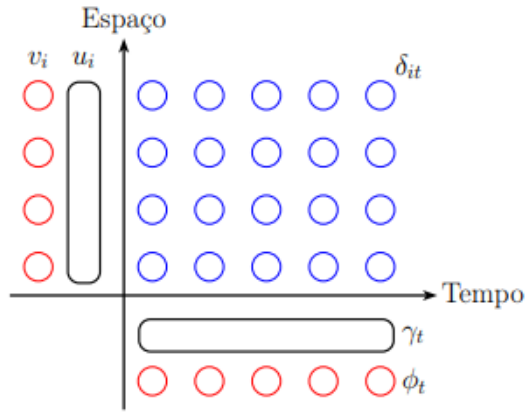


Figura 2 – Interação do Tipo I (CRUZ, 2022).

Na interação do tipo I, cada valor δ_{it} é identicamente e independentemente distribuído com uma distribuição gaussiana a priori com média zero e precisão τ_δ . A matriz de estrutura ou matriz de precisão pode ser representada da seguinte forma:

$$\mathbf{R}_\delta = \mathbf{R}_v \otimes \mathbf{R}_\phi = \mathbf{I}_n \otimes \mathbf{I}_T = \mathbf{I}_{nT},$$

em que $\otimes$ denota o produto de Kronecker (Apêndice A). GILKS, RICHARDSON e SPIEGELHALTER (1995) sugerem que seja especificado $\mathbf{R}_\delta$ como o produto de Kronecker (indicado por $\otimes$) das matrizes de estrutura dos efeitos espaciais e temporais que se interagem. A densidade do componente de interação δ é então (CRUZ, 2022):

$$p(\boldsymbol{\delta} \mid \tau_\delta) \propto \exp \left(-\frac{\tau_\delta}{2} \sum_{i=1}^n \sum_{t=1}^T \delta_{it}^2 \right) = \exp \left(-\frac{1}{2} \boldsymbol{\delta}' \mathbf{Q}_\delta \boldsymbol{\delta} \right).$$

Ao considerarmos a matriz de precisão $\mathbf{Q}_\delta = \tau_\delta \mathbf{R}_\delta$, em que $\boldsymbol{\delta} = (\delta_{11}, \dots, \delta_{nT})'$, podemos interpretar as interações como covariáveis não observadas para cada observação (CRUZ, 2022).

Interação tipo II

A interação do tipo II presume a existência de interações entre o efeito temporal estruturado γ_t e o efeito espacial não estruturado v_i (Figura 3). Dado que γ_t segue uma distribuição AR(1), σ_i também segue essa distribuição, independente de todas as outras. Assim, $\boldsymbol{\delta}_i = (\delta_{i1}, \dots, \delta_{iT})'$, em que $i = 1, 2, \dots, n$, segue um processo autoregressivo de ordem 1 (AR(1)) independente e δ é uma variável aleatória com média zero (CRUZ, 2022). Ou seja, na interação tipo II as tendências temporais são diferentes de região para região, mas não tem nenhuma estrutura no espaço (KNORR-HELD, 2000).

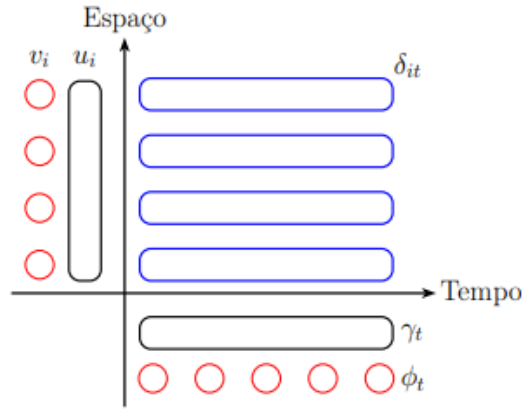


Figura 3 – Interação do Tipo II (CRUZ, 2022).

A matriz de estrutura da interação do tipo II é denotada como $\mathbf{R}_\delta = \mathbf{R}_v \otimes \mathbf{R}_\gamma$, em que $\mathbf{R}_v = \mathbf{I}_n$ e $\mathbf{R}_\gamma$ é a matriz de estrutura temporal AR(1). O modelo de interação do tipo II implica que cada região possui uma tendência temporal única, e essas tendências temporais não possuem estrutura espacial. Dado que γ_t segue uma distribuição AR(1), δ_t segue uma distribuição AR(1) (CRUZ, 2022):

$$\delta_t | \delta_{t-1} \sim \mathcal{N}(\rho_1 \gamma_{t-1}, \tau_\delta^{-1}).$$

A matriz de estrutura $\mathbf{R}_\delta$ tem posto $n(T-1)$, e a distribuição a priori para δ é dada por (KNORR-HELD, 2000):

$$p(\delta | \tau_\delta) \propto \exp \left(-\frac{\tau_\delta}{2} \sum_{i=1}^n \sum_{t=1}^T (\delta_{it} - \rho_1 \delta_{i,t-1})^2 \right).$$

Aqui, δ é definido como o vetor coluna (CRUZ, 2022):

$$\boldsymbol{\delta} = \begin{bmatrix} \delta_1 \\ \delta_2 \\ \vdots \\ \delta_n \end{bmatrix},$$

e $\mathbf{Q}_\delta = \tau_\delta \mathbf{R}_\delta$ é a matriz de precisão. $\mathbf{R}_\delta$ é a matriz de estrutura da tendência temporal.

Interação tipo III

A interação do tipo III pressupõe a presença de efeitos interdependentes entre o efeito espacial estruturado u_i e o efeito temporal não estruturado ϕ_t (Figura 4). Essa especificação será razoável se as tendências espaciais forem diferentes de um ponto temporal para outro sem qualquer estrutura temporal (KNORR-HELD, 2000).

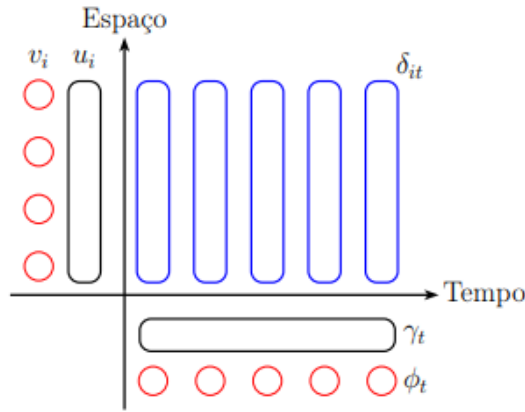


Figura 4 – Interação do Tipo III (CRUZ, 2022).

Tendo em vista que u_i segue uma distribuição ICAR, δ_i é também uma distribuição ICAR e sua definição ocorre da seguinte maneira (CRUZ, 2022):

$$\delta_i \mid \boldsymbol{\delta}_{-i} \sim \text{Normal} \left(\frac{1}{m_i} \sum_{i \sim j} \delta_j, \frac{1}{\tau_\delta m_i} \right). \quad (2.43)$$

O modelo de interação do tipo III descreve a estrutura espacial para cada tempo de forma distinta, utilizando a matriz de estrutura espacial $\mathbf{R}_u$ definida pela especificação a priori ICAR, combinada com a matriz de identidade $\mathbf{R}_\phi = \mathbf{I}_T$. É importante destacar que as estruturas espaciais são independentes entre tempos. A matriz estruturada resultante, $\mathbf{R}_\delta = \mathbf{R}_u \otimes \mathbf{R}_\phi$, tem posto $(n - 1)T$. Além disso, a priori para $\boldsymbol{\delta}$ é fornecida (CRUZ, 2022):

$$p(\boldsymbol{\delta} \mid \tau_\delta) \propto \exp \left(-\frac{\tau_\delta}{2} \sum_{t=2}^T \sum_{i \sim j}^n (\delta_{it} - \delta_{jt})^2 \right) = \exp \left(-\frac{\tau_\delta}{2} \boldsymbol{\delta}' \mathbf{Q}_\delta \boldsymbol{\delta} \right).$$

Interação tipo IV

A interação do tipo IV implica que existe uma conexão entre o efeito espacial estruturado u_i e o efeito temporal estruturado γ_t , conforme representado na Figura 5. Neste cenário, a estrutura temporal de cada região é influenciada pela estrutura temporal das regiões vizinhas, já que u_i segue uma distribuição ICAR e γ_t segue uma distribuição AR(1). Portanto, o vetor $\boldsymbol{\delta} = (\delta_{11}, \dots, \delta_{nt})'$, $i = 1, \dots, n$ e $t = 1, \dots, T$, é completamente dependente tanto do espaço quanto do tempo, e não pode ser completamente explicado apenas por esses dois efeitos estruturados. Nesse sentido, pode-se pensar em δ_{it} como uma mistura parcial dos modelos do tipo II e tipo III (CRUZ, 2022).

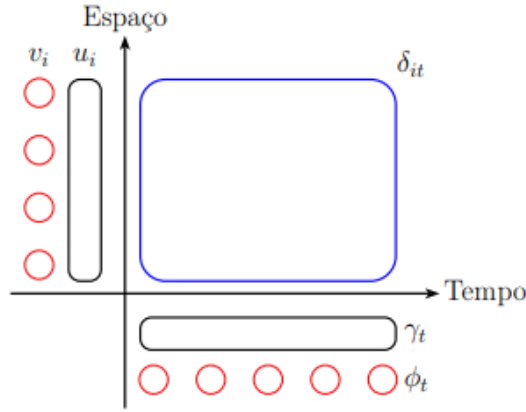


Figura 5 – Interação do Tipo IV (CRUZ, 2022).

A matriz de estrutura da interação do tipo IV é dada por $\mathbf{R}_\delta = \mathbf{R}_u \otimes \mathbf{R}_\gamma$, em que $\mathbf{R}_u$ representa a matriz de estrutura espacial definida pela técnica ICAR, e $\mathbf{R}_\gamma$ é a matriz de estrutura de tendência temporal definida por meio de uma especificação AR(1). A matriz estruturada $\mathbf{R}_\delta$ tem posto $(T - 1)(n - 1)$. A densidade conjunta a priori para $\mathbf{R}_\delta$ é dada por:

$$p(\delta|\tau_\delta) \propto \exp \left(-\frac{\tau_\delta}{2} \sum_{t=1}^T \sum_{i \sim j} (\delta_{i,t} - \delta_{j,t} - \rho_1 \delta_{i,t-1} + \delta_{j,t-1})^2 \right) = \exp \left(-\frac{\tau_\delta}{2} \boldsymbol{\delta}' \mathbf{Q}_\delta \boldsymbol{\delta} \right),$$

em que $\boldsymbol{\delta} = (\delta_{11}, \dots, \delta_{nT})'$, $\mathbf{Q}_\delta = \tau_\delta \mathbf{R}_\delta$ é a matriz de precisão e $\mathbf{R}_\delta$ é a matriz de estrutura de tendência temporal. Esta priori assume a existência de uma estrutura espacial entre as regiões para um determinado tempo, e essa estrutura espacial também é dependente em diferentes tempos. De forma equivalente, pode-se enfatizar que existe uma estrutura temporal entre os tempos para uma determinada região, e essa estrutura temporal também é dependente entre as regiões vizinhas (KNORR-HELD, 2000; CRUZ, 2022).

3 Materiais e métodos

3.1 Banco de dados

O banco de dados empregado na presente pesquisa foi obtido através da plataforma de dados de acesso público do Estado de São Paulo, disponível no seguinte endereço: <<https://www.saopaulo.sp.gov.br/planosp/simi/dadosabertos/>>. A coleta de dados teve início em 26 de fevereiro de 2020, data em que se confirmou o primeiro caso de COVID-19 tanto no Brasil quanto no Estado de São Paulo. Desde essa data, informações diárias relacionadas à COVID-19 em cada um dos 645 municípios do Estado têm sido constantemente atualizadas. Além disso, dados demográficos e informações abrangentes sobre óbitos durante o período de estudo foram obtidos através da plataforma de dados de acesso público fornecida pelo Instituto Brasileiro de Geografia e Estatística (IBGE), acessível pelo endereço: <<https://www.ibge.gov.br/estatisticas/downloads-estatisticas>>.

A presente análise incorporou os dados de óbitos confirmados por COVID-19 acumulados mensalmente até 31 de dezembro de 2021, abrangendo todos os 645 municípios do Estado de São Paulo. Até a data mencionada, os municípios haviam registrado um total de 4.453.789 casos confirmados de COVID-19 e 155.198 óbitos em decorrência da mesma doença. Além disso, foram utilizados dados de óbitos gerais durante o mesmo período e população. Este conjunto de dados abrange um período de vinte e três meses da pandemia no Estado.

3.2 Software utilizado

Os procedimentos analíticos foram conduzidos mediante a implementação dos modelos estatísticos no ambiente de programação estatística R, versão 4.3.1 (R CORE TEAM, 2021), utilizando o pacote R-INLA (RUE; MARTINO; N.CHOPIN, 2009), que adota a abordagem INLA para a estimação de modelos bayesianos. A produção dos mapas temáticos foi realizada por meio da aplicação do pacote ggplot2 (WICKHAM, 2016). A base cartográfica utilizada, obtida a partir do portal do IBGE (disponível em: <<https://portaldemapas.ibge.gov.br/portal.php#homepage>>), foi convertida para o formato *GeoPackage* com auxílio do software QGIS, versão 3.18 (QGIS DEVELOPMENT TEAM, 2010; QGIS DEVELOPMENT TEAM, 2021). Essa conversão foi adotada em virtude da maior versatilidade e robustez oferecida pelo formato *GeoPackage* no que se refere ao armazenamento e gerenciamento de dados geoespaciais. A execução das análises foi

conduzida em um microcomputador com configuração composta por processador Intel-core i5 de 2,71 GHz e 16,0 GB de memória RAM.

3.3 Análise descritiva

Em nossa análise, investigamos um conjunto de dados que abrange um total de 155.198 óbitos relacionados à COVID-19 ocorridos no Estado de São Paulo entre fevereiro de 2020 e dezembro de 2021. Na Figura 6, apresentamos uma representação gráfica do número total de óbitos registrados para cada mês dentro desse período.

É importante ressaltar que os meses de fevereiro a dezembro de 2020 foram analisados, e durante esse período, observamos um pico no número de óbitos em julho, alcançando um total de 8.228 registros. Por outro lado, os meses de janeiro a dezembro de 2021 foram considerados, sendo o mês de abril o de maior destaque com um número significativo de óbitos, totalizando 21.524 casos.

Uma comparação entre os meses de junho de 2020 e junho de 2021 revela uma duplicação no número de óbitos, indicando um aumento substancial da mortalidade ao longo desse período. Contudo, é pertinente mencionar que, independentemente do ano, os meses de novembro demonstraram uma diminuição no número de óbitos quando contrastados com outros períodos do ano.

Essas observações destacam a evolução temporal do número de óbitos por COVID-19 no Estado de São Paulo, fornecendo *insights* importantes sobre os padrões de mortalidade associados à pandemia durante os anos de 2020 e 2021.

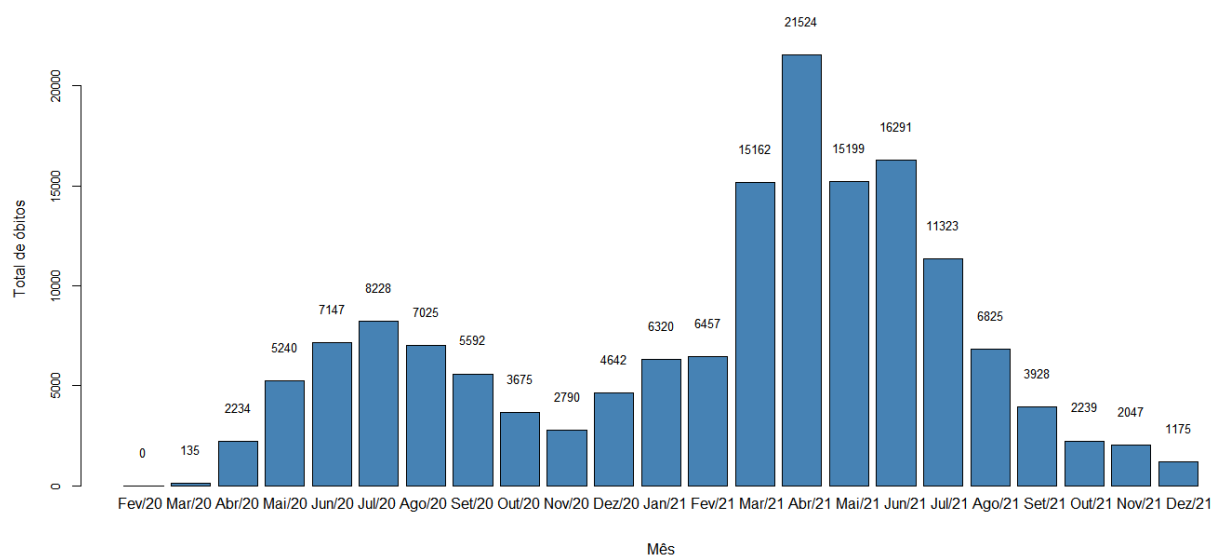


Figura 6 – Total de óbitos por mês por COVID-19 no período de 2020 e 2021 no Estado de São Paulo.

A distribuição dos óbitos totais relacionados à COVID-19 no Estado de São Paulo, durante o período abrangido entre os anos de 2020 e 2021, é apresentada de forma cartográfica na Figura 7. Dado o contexto de disparidade populacional entre os municípios que compõem o Estado, procedeu-se à normalização do número total de óbitos por COVID-19 em relação à população de cada município. Tal procedimento consistiu na divisão do total de óbitos pelo número de habitantes correspondente a cada município, seguido da multiplicação do resultado por 100.000. Desta maneira, a construção do mapa resultou em uma representação que se situa dentro da escala de variação de $[0, 800[$ óbitos por 100.000 habitantes.

De maneira geral, a análise do mapa revela um padrão em que a maioria dos municípios do Estado de São Paulo apresenta taxas de óbitos por COVID-19 variando entre 200 e 400 casos por 100.000 habitantes. No entanto, chama a atenção a região noroeste do estado, onde os índices são mais elevados, oscilando entre 400 e 800 óbitos por 100.000 habitantes. Essa disparidade geográfica nos níveis de mortalidade ressalta a importância de considerar as nuances locais e regionais na análise.

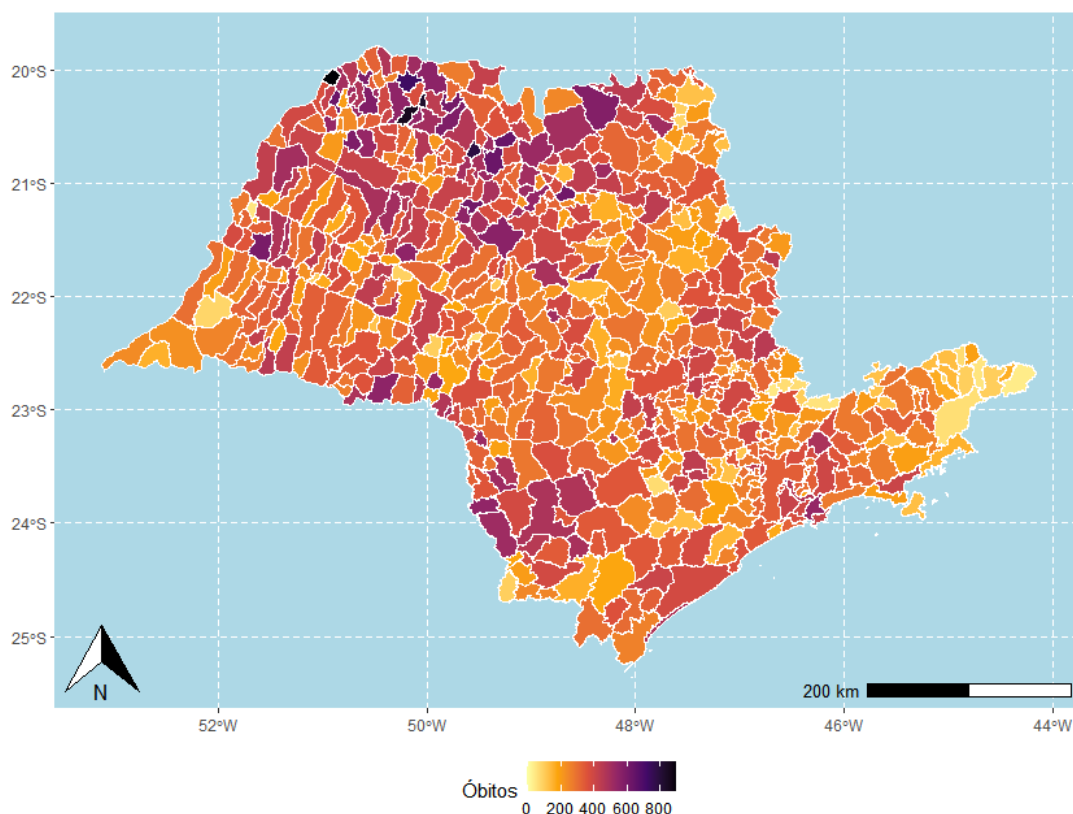


Figura 7 – Distribuição dos óbitos totais por COVID-19 no período de 2020 e 2021 nos municípios do Estado de São Paulo.

A Tabela 1 apresenta informações relacionadas ao número de óbitos por COVID-19 nos 645 municípios do Estado de São Paulo durante o período de 2020 e 2021. Nesta tabela, são disponibilizados os quartis, a média, a mediana, o desvio padrão, os valores máximos e mínimos.

Tabela 1 – Estatísticas descritivas para o número de óbitos por COVID-19 dos 645 municípios do Estado de São Paulo de 2020 e 2021.

Ano	Mês	Mínimo	Máximo	Média	Desvio Padrão	1º Quartil	Mediana	3º Quartil
2020	fev	0	0	0	0	0	0	0
	mar	0	121	0,21	4,77	0	0	0
	abr	0	1401	3,46	55,39	0	0	0
	mai	0	2758	8,12	109,88	0	0	1
	jun	0	2882	11,08	116,26	0	0	2
	jul	0	2387	12,76	98,65	0	1	5
	ago	0	1859	10,89	76,77	0	1	5
	set	0	1322	8,67	55,27	0	1	4
	out	0	867	5,70	35,76	0	1	3
	nov	0	804	4,33	32,73	0	1	2
	dez	0	1278	7,20	52,62	0	1	3
2021	jan	0	1654	9,80	67,89	0	1	5
	fev	0	1309	10,01	55,74	0	2	6
	mar	0	3486	23,51	145,39	1	4	12
	abr	0	5240	33,37	216,47	2	6	17
	mai	0	3228	23,56	134,32	1	4	14
	jun	0	2914	25,26	124,35	1	5	16
	jul	0	1940	17,55	83,81	1	3	11
	ago	0	1655	10,58	67,88	0	2	6
	set	0	1147	6,09	47,05	0	1	3
	out	0	561	3,47	23,36	0	0	2
	nov	0	467	3,17	20,26	0	0	1
	dez	0	281	1,82	11,69	0	0	1

No ano de 2020, observamos que os números de óbitos por COVID-19 começaram a surgir em março, com um máximo de 121 óbitos registrados nesse mês. No entanto, a média mensal manteve-se relativamente baixa, atingindo seu pico em junho com uma média de 11,08 óbitos. O desvio padrão também demonstrou uma considerável variabilidade nos números de óbitos entre os municípios.

No ano de 2021, a situação evoluiu, com os números de óbitos atingindo seu máximo em abril, com 5.240 óbitos registrados. A média mensal para 2021 variou consideravelmente, alcançando um pico em abril com 33,37 óbitos. Mais uma vez, o desvio padrão indicou uma grande dispersão nos números de óbitos, sugerindo uma variação significativa entre os municípios.

É importante notar que em todos os meses, tanto em 2020 quanto em 2021, houve registros de zero óbitos por COVID-19 em alguns municípios, o que reflete a heterogeneidade na incidência da doença ao longo do tempo e entre as regiões do Estado de São Paulo.

A análise dos quartis e seus intervalos na Tabela 1 é fundamental para compreender não apenas o valor central da distribuição dos óbitos por COVID-19, mas também a variabilidade nos dados, destacando a amplitude dos valores e as diferenças entre os meses e anos específicos.

Essas estatísticas descritivas fornecem uma visão abrangente das tendências e variações nos números de óbitos por COVID-19 nos municípios do Estado de São Paulo durante os anos de

2020 e 2021, evidenciando os momentos de maior incidência e a heterogeneidade nos registros ao longo do tempo e entre os municípios.

3.4 Matriz de vizinhança

Em geral, é razoável supor que áreas geograficamente próximas apresentem uma carga de doenças mais semelhante do que áreas distantes. Para definir o que se entende por “próximo”, é necessário estabelecer uma estrutura de vizinhança. Uma suposição comum é considerar as áreas i e j como vizinhas se elas compartilham uma fronteira comum, o que é denotado por $i \sim j$. Essa suposição parece ser apropriada para regiões de dimensões semelhantes dispostas em um padrão regular (WAKEFIELD, 2007). No entanto, se as regiões apresentarem tamanhos e formas irregulares ou com descontinuidades, outras definições de vizinhança podem ser necessárias.

No caso específico do Estado de São Paulo, cada município é associado a um código único, o que permite definir o conjunto de índices $D = 1, \dots, 645$ representando as regiões geográficas do estado. A partir desse conjunto, é possível construir uma estrutura de dependência espacial para as observações, considerando as regiões vizinhas cujas fronteiras físicas são compartilhadas.

Para representar as relações espaciais entre as áreas em um contexto cartográfico, é uma prática comum a utilização de uma matriz de pesos espaciais, frequentemente denotada como matriz W . A construção dessa matriz pode ser efetuada através da aplicação de diferentes estilos de vizinhança, comumente conhecidos como estilos “queen” e “rook”.

No estilo de vizinhança “queen”, as unidades geográficas são consideradas vizinhas se compartilharem uma fronteira comum ou um vértice. Em termos práticos, isso implica que duas unidades são vizinhas se estiverem contíguas na diagonal ou nas faces de um quadrado que circunscreve as unidades (ANSELIN, 1988). Esse estilo de vizinhança é representado graficamente na Figura 8.

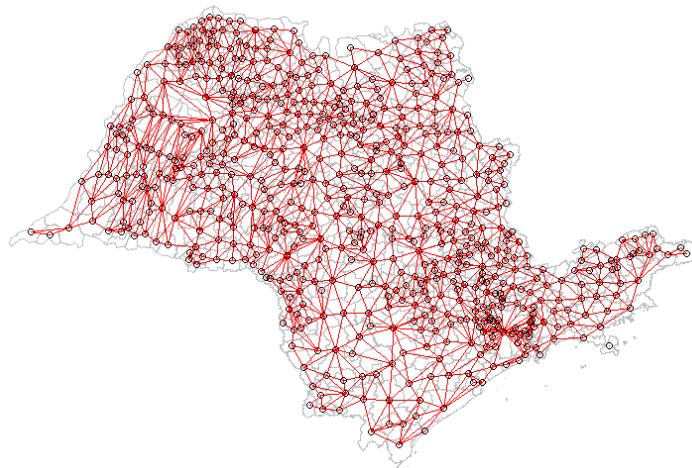


Figura 8 – Matriz de pesos espaciais por meio do estilo de vizinhança B (ou contiguidade *queen*) para o Estado de São Paulo.

Por outro lado, no estilo de vizinhança “*rook*”, as unidades geográficas são consideradas vizinhas apenas se compartilharem uma fronteira comum nas faces do quadrado que envolve as unidades, excluindo-se a diagonal (ANSELIN, 1988). Essa diferenciação na definição de vizinhança é ilustrada na Figura 9.

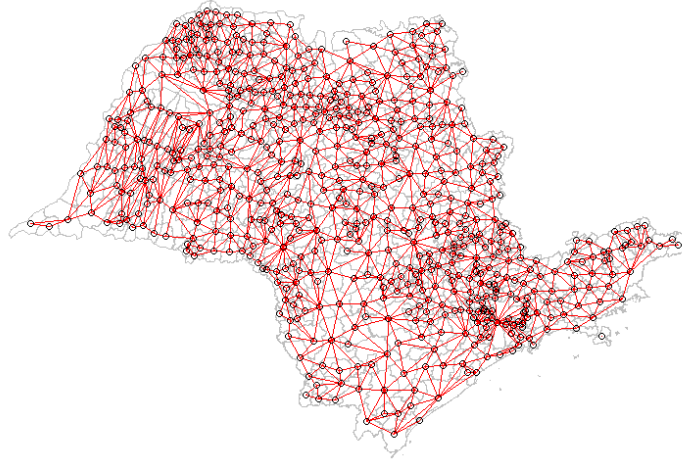


Figura 9 – Matriz de pesos espaciais por meio do estilo de vizinhança M (ou contiguidade *rook*) para o Estado de São Paulo.

A escolha entre o estilo de vizinhança “*queen*” (B) ou “*rook*” (M) para a construção da matriz de pesos espaciais depende do contexto e das necessidades específicas da análise espacial a ser realizada. Cada estilo de vizinhança pode ser mais apropriado para determinados tipos de modelos ou situações geográficas, sendo importante a consideração cuidadosa do estilo mais adequado para a investigação em questão. Neste trabalho foi utilizado a vizinhança “*queen*” (B) para os modelos Besag, BYM e BYM2, e “*rook*” (M) para o modelo Leroux, pois este é aplicado quando há uma estrutura de vizinhança complexa.

3.5 Análise da evolução espaço-temporal

Para realizar a análise da evolução espaço-temporal da COVID-19, foram empregados modelos bayesianos hierárquicos que aderem à estrutura definida pela distribuição de probabilidade dos dados observados. Esses modelos foram construídos de forma a incorporar efeitos aleatórios que abrangem as dimensões espacial, temporal e espaço-temporal. No âmbito de sua estrutura, esses modelos abrangem três níveis hierárquicos e buscam incorporar as covariáveis número de casos de COVID-19, óbitos gerais e população.

No primeiro nível, o objetivo é ter um modelo que represente os dados de óbitos por COVID-19. O preditor linear do modelo, estabelecido entre as covariáveis e vetor de parâmetros, ou seja, as variáveis explicativas que determinam o valor médio de Y_{it} , em que Y_{it} é o valor observado na área i e no tempo t , por uma função linear das variáveis predictoras é dado por:

$$\eta_{it} = \alpha + \mathbf{X}_p \boldsymbol{\beta} + u_i + v_i + \gamma_t + \phi_t + \delta_{it}, \quad (3.1)$$

em que α é um intercepto comum para todas as unidades espaciais e temporais, $\mathbf{X}_p$ são covariáveis, $\boldsymbol{\beta}$ é um vetor de coeficientes associados às covariáveis. Os efeitos espaciais estruturados e não estruturados para cada área i são representados por u_i e v_i , enquanto γ_t e ϕ_t representam os efeitos temporais estruturados e não estruturados para o tempo t . A componente δ_{it} captura a tendência diferencial que explica a interação espaço-tempo entre as tendências temporais em diferentes locais.

A distribuição de Poisson é comumente usada para expressar a distribuição de probabilidade de dados de saúde. Porém há um excesso de zeros nos dados de óbitos por COVID-19 e uma maneira de lidar com essa situação é utilizar o modelo de Poisson inflacionado de zero (ZIP), este modelo é composto por duas partes: uma parte binomial que modela a probabilidade de um zero de excesso, ou seja, ela determina a probabilidade de não ocorrer nenhum evento em uma observação. Isso é tratado por uma distribuição binomial. E uma parte modela a distribuição de Poisson condicional (condicionada ao evento de não ser um zero de excesso). Ela determina a distribuição da contagem de eventos quando pelo menos um evento ocorreu. Então temos que (ATKINS et al., 2013):

Probabilidade da ocorrência de zero inflacionado (Binomial):

$$P(Y_{it} = 0) = \pi + (1 - \pi) \exp(-\lambda),$$

em que Y_{it} é a contagem de óbitos por COVID-19 em uma determinada área i ($i \in 1, \dots, 645$) e tempo t ($t \in 1, \dots, 23$), π é a probabilidade de ser um zero de excesso e λ é o parâmetro de taxa da distribuição Poisson condicional.

Temos que a função de ligação será dada pela equação $\text{logit}(\pi) = \log\left(\frac{\pi}{1-\pi}\right) = \eta_{it}$, em que η_{it} (3.1) é uma combinação linear das covariáveis do modelo.

Probabilidade de Poisson condicional:

$$P(Y_{it} = y_{it} | Y_{it} > 0) = \frac{(1 - \pi)(\lambda^{y_{it}} \exp(-\lambda))}{y_{it}!},$$

em que Y_{it} é a contagem de óbitos por COVID-19 em uma determinada área i ($i \in 1, \dots, 645$) e tempo t ($t \in 1, \dots, 23$), π é a probabilidade de ser um zero de excesso e λ é o parâmetro de taxa da distribuição Poisson condicional. Temos que a função de ligação é dada pela equação $\log(\mu_{it}) = \eta_{it}$, em que μ_{it} é calculado pela fórmula $\mu_{it} = E_{it} \psi_{it}$, em que E_{it} é o número esperado de óbitos e ψ_{it} representa o risco relativo para a área e tempo em questão.

Assim, o objetivo do modelo é identificar as relações entre as covariáveis do modelo e o número de óbitos por COVID-19 em uma determinada área e tempo, utilizando a distribuição ZIP e a função logarítmica que relaciona o número esperado de óbitos e o risco relativo.

No segundo nível do modelo são atribuídas as distribuições a priori aos parâmetros. Essas distribuições representam as crenças iniciais sobre os valores plausíveis dos parâmetros antes da observação dos dados. No modelo em questão, foram atribuídas distribuições a priori para as variáveis independentes, bem como para os parâmetros de precisão espacial e temporal. Tais decisões metodológicas foram fundamentadas na análise crítica da literatura especializada, em particular, nos estudos conduzidos por JAYA, FOLMER e LUNDBERG (2022), JALILIAN e MATEU (2021) e SAHU e BÖHNING (2022), que contribuíram com as diretrizes conceituais e empíricas que embasaram a seleção dessas distribuições a priori.

Para os parâmetros de efeitos fixos, foram especificadas prioris com distribuições Gaussianas não informativas, ou seja, $\alpha \sim N(0; 10^6)$ e $\beta_{it} \sim N(0; 10^{-3})$, para $i = 1, 2, \dots, n$ e $t = 1, 2, \dots, T$.

Para os efeitos espaciais estruturados, as distribuições a priori mais comuns incluem o modelo condicional autoregressivo intrínseco (ICAR) e o modelo condicional autoregressivo Leroux mais geral (LCAR). No caso dos efeitos temporais estruturados, as distribuições a priori mais comuns são o modelo autoregressivo 1 (AR(1)) e os modelos *Random Walk* 1 e 2 (RW(1) e RW(2)). Abaixo estão apresentadas quatro diferentes especificações para as componentes espaciais e três diferentes especificações para as componentes temporais:

- **Modelo de Besag:** O modelo de Besag é um modelo estatístico simples e amplamente utilizado, que se baseia no modelo condicional autoregressivo intrínseco (ICAR) proposto por BESAG, YORK e MOLLIE (1991). O ICAR tem distribuições condicionais para u_i , dadas por $u_i | \mathbf{u}_{-i}, \tau_u \sim N\left(\frac{1}{m_i} \sum_{j \sim i} u_j, \frac{1}{\tau_u m_i}\right)$, em que τ_u é o parâmetro de precisão e $\mathbf{u}_{-i}$ representa os vizinhos da i -ésima área. A média de u_{-i} é a média dos seus vizinhos e a precisão é proporcional ao número de vizinhos que a área possui. A variância de $V(u_i | \mathbf{u}_{-i})$ é inversamente proporcional ao número de vizinhos (m_i). A distribuição a priori atribuída a τ_u e a estrutura do grafo desempenham um papel importante na determinação da distribuição CAR.

A estrutura do grafo é definida com base no pressuposto de independência condicional entre as áreas. Em outras palavras, o modelo assume que os valores de u_i são condicionalmente independentes dada a informação dos valores nas áreas vizinhas u_j , $i \sim j$. Essas suposições são refletidas na estrutura da matriz de precisão da distribuição conjunta para $\mathbf{u}$, dada por:

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

em que τ_u é o parâmetro de precisão, $\mathbf{u}$ é o vetor de valores das áreas, $\mathbf{Q}$ é a matriz de precisão definida por $Q_{ii} = m_i$ (o número de áreas vizinhas da i -ésima área) e $Q_{ij} = -1$ se as áreas i e j são vizinhas e 0 caso contrário. O vetor de vizinhança é denotado por $i \sim j$.

- **Modelo BYM:** O modelo de Besag-York-Mollie - BYM (BESAG; YORK; MOLLIE, 1991), também conhecido como modelo de convolução, é uma extensão do modelo ICAR que incorpora tanto efeitos espaciais quanto efeitos aleatórios não espaciais na análise de dados espaciais. Enquanto o modelo ICAR assume que a variável de interesse é influenciada por efeitos espaciais locais e globais, o modelo BYM decompõe o efeito espacial em um componente espacial estruturado e um componente não estruturado. Essa decomposição é feita de forma a considerar $u_i + v_i$ como independentes a priori, onde u_i representa o efeito espacial estruturado e v_i representa o efeito não estruturado.

Para modelar o efeito espacial estruturado, adota-se o modelo ICAR, que assume que o risco de incidência de uma determinada doença em uma área espacial é derivado de vizinhos geograficamente próximos. Um modelo ICAR (BESAG; KOOPERBERG, 1995) para u_i , que pode ser escrito como $u_i \mid \mathbf{u}_{-i} \sim N\left(\frac{1}{m_i} \sum_{i \sim j} u_j, \frac{u}{\tau_u m_i}\right)$, em que $i \sim j$ indica que as áreas i e j são vizinhas, m_i é o número de áreas que compartilham fronteira com a i -ésima área, $\tau_u = 1/\sigma_u^2$ é o parâmetro de precisão e σ_u^2 , a variância. A dependência espacial em u_i é representada pela priori ICAR, que estende o modelo Besag com uma distribuição gaussiana e implica que cada u_i é condicional ao vizinho u_j , com a variância dependente do número de áreas vizinhas m_i da i -ésima unidade espacial (CRUZ, 2022). Essa dependência espacial estruturada é conhecida como autocorrelação espacial.

Em resumo, o modelo BYM é uma generalização do modelo ICAR que permite a inclusão de efeitos aleatórios não espaciais na análise de dados espaciais. A dependência espacial estruturada é modelada pelo ICAR, que assume que a incidência de uma doença em uma área espacial é influenciada por vizinhos geograficamente próximos. Essa abordagem é útil para entender e prever a ocorrência de doenças em áreas espaciais.

Uma priori permutável é usada para modelar o efeito aleatório espacialmente não estruturado v_i :

$$v_i \mid \tau_v \sim N(0, \tau_v^{-1}),$$

em que τ_v é o parâmetro de precisão.

A matriz de covariância resultante de $\mathbf{u} + \mathbf{v}$ é:

$$\text{Var}(\mathbf{u} + \mathbf{v} \mid \tau_u, \tau_v) = \tau_v^{-1} \mathbf{I} + \tau_u^{-1} \mathbf{Q}^-,$$

em que $\mathbf{Q}^-$ denota a matriz inversa generalizada de Moore-Penrose (BAKSALARY; TRENKLER, 2021; CRUZ, 2022).

Uma restrição de soma zero $\sum_{i=1}^n (u_i + v_i) = 0$ é imposta a $u_i + v_i$ para garantir a identificabilidade do intercepto α , pois os componentes de variância no modelo BYM não são identificáveis a partir dos dados (CRUZ, 2022).

- **Modelo BYM2:** O modelo BYM2 é uma extensão do modelo BYM original, incorporando a possibilidade de modelar a variação temporal dos dados. Dessa forma, ele pode ser utilizado para análise de dados longitudinais, que acompanham a evolução do fenômeno ao longo do tempo.

Uma das vantagens do modelo BYM2 é que ele permite a obtenção de estimativas robustas e precisas das taxas de incidência/prevalência, levando em conta a estrutura espacial e temporal dos dados (BLANGIARDO; CAMELETTI, 2015). Isso pode ser especialmente útil em áreas como a saúde pública e a segurança pública, onde a compreensão da distribuição geográfica e temporal de um fenômeno é fundamental para o desenvolvimento de políticas eficazes de prevenção e controle.

Recentemente, SIMPSON et al. (2017) sugeriram que os componentes u e v não devem ser considerados independentes no modelo de Besag, York e Mollié. Eles propuseram uma reparametrização do modelo BYM original, que incorpora apenas um parâmetro de precisão nos componentes (τ) e um parâmetro de mistura (λ) para mensurar a variação espacial e não-espacial. O novo modelo, chamado BYM2, inclui um componente estruturado espacialmente u_* e um componente não estruturado v_* , permitindo uma melhor compreensão da variação espacial e não-espacial:

$$\mathbf{b} = \frac{1}{\sqrt{\tau_b}} \left(\sqrt{1 - \lambda} \mathbf{v}_* + \sqrt{\lambda} \mathbf{u}_* \right). \quad (3.2)$$

A matriz de covariância é dada por:

$$\text{Var}(\mathbf{b} \mid \tau_b, \lambda) = \tau_b^{-1} \left((1 - \lambda) \mathbf{I} + \lambda \mathbf{Q}_*^- \right), \quad (3.3)$$

em que $\mathbf{Q}_*^-$ denota a matriz inversa generalizada de Moore-Penrose (BAKSALARY; TRENKLER, 2021). A Equação (3.2) é a reparametrização do modelo BYM, em que $\tau_u^{-1} = \tau_b^{-1} \lambda$ e $\tau_v^{-1} = \tau_b^{-1} (1 - \lambda)$ (MACNAB, 2011). Tal como ocorreu no modelo anterior, é necessário impor uma restrição de soma zero $\sum_{i=1}^n b_i = 0$ para prevenir problemas de identificabilidade (CRUZ, 2022).

- **Modelo de Leroux:** O modelo Leroux (LEROUX; LEI; BRESLOW, 2000) é também uma extensão do modelo BYM que permite a inclusão de covariáveis para explicar as diferenças na estrutura espacial entre as unidades geográficas. O modelo Leroux também introduz um parâmetro adicional para modelar a variação na intensidade de dependência espacial. Este parâmetro controla a escala da variação espacial e, portanto, permite que o modelo se adapte a diferentes padrões de dependência espacial em diferentes regiões.

Além disso, o modelo Leroux apresenta uma abordagem inovadora para a especificação da estrutura de dependência espacial. Ele usa um processo de Markov para modelar a variação espacial, que é então integrado ao modelo bayesiano. Essa abordagem permite

que o modelo capture padrões complexos de dependência espacial, como a presença de *clusters*, sem exigir uma especificação explícita da estrutura de dependência espacial.

Enquanto o modelo BYM contém dois efeitos aleatórios espaciais, o modelo Leroux possui um único efeito aleatório espacial, u_i , que ajusta a força da autocorrelação espacial da vizinhança local por uma constante λ (CRUZ, 2022). O modelo é dado por (LEE, 2011):

$$u_i \mid \mathbf{u}_{-i} \sim N \left(\frac{\lambda \sum_{i \sim j} u_j}{\lambda m_i + 1 - \lambda}, \frac{\tau_u^{-1}}{\lambda m_i + 1 - \lambda} \right). \quad (3.4)$$

A distribuição conjunta a priori do vetor $\mathbf{u}$ segue uma distribuição normal com média zero e matriz de covariância:

$$\text{Var}(\mathbf{u} \mid \tau_u, \lambda) = \frac{1}{\tau_u} [\lambda \mathbf{Q} + (1 - \lambda) \mathbf{I}_n]^{-1}, \quad (3.5)$$

em que a diagonal q_{ii} da matriz $\mathbf{Q}$, contém o número de vizinhos da área i , e $q_{ij} = -1$ se $i \sim j$, e 0 caso contrário. O parâmetro λ , $0 \leq \lambda \leq 1$ denota um parâmetro de dependência espacial, tal que $\lambda = 0$ define um modelo não espacial (modelo independente) e o nível de dependência espacial aumenta com λ . Quando $\lambda = 1$, o modelo de Leroux se reduz ao modelo ICAR em que a suavização dos efeitos aleatórios vizinhos é totalmente correlacionada (CRUZ, 2022). Por outro lado, quando λ é igual a 0, a suavização é totalmente não correlacionada. O parâmetro λ pondera a intensidade da correlação na suavização dos efeitos aleatórios vizinhos, permitindo ajustar a quantidade de correlação desejada no modelo (LEE; MITCHELL, 2012).

No modelo Leroux, assim como no modelo BYM, é necessário restringir os efeitos aleatórios para evitar problemas de identificabilidade com o intercepto. Quando o intercepto é incluído como um termo separado no segundo estágio da hierarquia, os efeitos aleatórios devem ser restritos a somar zero, $\sum_{i=1}^n u_i = 0$, de modo a permitir uma estimação única e precisa do intercepto. Portanto, a restrição é fundamental para garantir a coerência e a estabilidade do modelo.

- **Modelo autoregressivo de ordem 1:** O Modelo Autoregressivo de ordem 1, também conhecido como AR(1), é um modelo estatístico que descreve a relação entre uma variável aleatória e seus valores anteriores, com base na ideia de que os valores passados da variável podem ser usados para prever seus valores futuros. O modelo assume que a variável aleatória é uma combinação linear dos seus valores anteriores e um erro aleatório (BOX et al., 2015):

$$\gamma_t = \rho \gamma_{t-1} + \epsilon_t, |\rho| < 1 \text{ e } \epsilon_t \sim N(0, \tau_\gamma^{-1}), \quad (3.6)$$

em que ϵ_t é um processo de ruído branco com média zero e variância constante τ_γ^2 e ρ é o parâmetro de autocorrelação temporal, que introduz uma estrutura de correlação temporal em γ_t .

O AR(1) é especialmente útil para modelar séries temporais que exibem autocorrelação significativa, ou seja, quando os valores atuais da série estão correlacionados com seus valores passados. O modelo pode ser estendido para ordens mais altas, como AR(p), para incorporar dependências de ordem superior (BROCKWELL; DAVIS, 2002).

- **Modelo *Random Walk 1*:** O modelo *Random Walk 1* (RW(1)) é um modelo simples de séries temporais que assume que o valor atual da série é igual ao valor anterior mais um erro aleatório. Matematicamente, o modelo RW(1) é definido como (BROCKWELL; DAVIS, 2002):

$$\gamma_t = \gamma_{t-1} + \epsilon_t, \text{ e } \epsilon_t \sim N(0, \tau_\gamma^{-1}), \quad (3.7)$$

em que γ_t é o valor da série temporal no tempo t , γ_{t-1} é o valor da série temporal no tempo $t - 1$ e ϵ_t é um erro aleatório com média zero e variância constante.

O modelo RW(1) é frequentemente usado como uma linha de base para comparar a precisão de outros modelos de previsão. Ele é simples de implementar e pode ser útil para prever séries temporais altamente voláteis. No entanto, o modelo RW(1) assume que os valores da série são imprevisíveis e não levam em conta a dependência temporal, o que pode limitar sua utilidade em prever valores distantes no futuro. Além disso, é uma ferramenta importante na análise de séries temporais e é frequentemente usado como um modelo de referência para avaliar outros modelos mais complexos.

- **Modelo *Random Walk 2*:** O Modelo *Random Walk 2* (RW(2)) é uma extensão do Modelo RW(1). Assim como o RW(1), ele é usado para modelar séries temporais, mas inclui uma dependência temporal maior, levando em conta os dois valores anteriores da série temporal. Matematicamente, o modelo RW(2) é definido como:

$$\gamma_t = \gamma_{t-1} + \gamma_{t-2} + \epsilon_t, \text{ e } \epsilon_t \sim N(0, \tau_\gamma^{-1}), \quad (3.8)$$

em que γ_t representa o valor da série temporal no tempo t , γ_{t-1} é o valor no tempo $t - 1$, γ_{t-2} é o valor no tempo $t - 2$, e ϵ_t é um erro aleatório com média zero e variância constante.

O modelo RW(2) captura uma maior dependência temporal em comparação com o RW(1) e, portanto, pode ser mais adequado para séries temporais que exibem padrões mais complexos. Ele também é usado como uma referência em previsões de séries temporais, especialmente quando a dependência de dois períodos anteriores é relevante. No entanto, assim como o RW(1), o RW(2) não considera tendências de longo prazo ou sazonalidade, e sua utilidade pode ser limitada em séries com essas características distintas. Em geral, o RW(2) é uma ferramenta importante na análise de séries temporais, servindo como base para comparações com modelos mais sofisticados.

Para representar o efeito espacial não estruturado, adotamos uma distribuição a priori Normal permutável, o que implica que os parâmetros seguem a mesma distribuição Normal, com média e variância idênticas. Nesse cenário, temos:

$$v_i \mid \tau_v \sim N(0, \tau_v),$$

em que τ_v é o parâmetro de variância de v_i .

Para o efeito temporal não estruturado, também adotamos uma distribuição a priori Normal permutável e nesse cenário, temos:

$$\phi_t \mid \tau_\phi \sim N(0, \tau_\phi),$$

em que τ_ϕ é o parâmetro de variância de ϕ_t .

Em complemento às especificações delineadas, procedeu-se à incorporação das quatro modalidades de interações espaço-temporais mencionadas na Subseção 2.9.2, as quais foram empregadas para caracterizar o termo δ_{it} na Equação (3.1). Nesse contexto, as prioris adotadas são as mesmas que aquelas dos efeitos aleatórios espaciais estruturados e efeitos aleatórios temporais estruturados que interagem entre si.

No terceiro nível do modelo, especificamos as hiperprioris para os parâmetros das distribuições a priori do segundo nível baseado nos artigos de JAYA, FOLMER e LUNDBERG (2022) e JALILIAN e MATEU (2021). As hiperprioris são distribuições a priori para os parâmetros dessas distribuições a priori e representam as crenças iniciais sobre os valores plausíveis desses parâmetros antes de observar os dados. Neste modelo, as hiperprioris foram definidas para os hiperparâmetros de variância, incluindo α , τ_u , τ_v , τ_γ , τ_ϕ , e δ . Uma escolha comum é a distribuição gama inversa não informativa (IG) com parâmetros de forma e escala próximos a zero. Portanto, os hiperparâmetros, α , τ_u , τ_v , τ_γ , τ_ϕ , e δ , seguem uma distribuição $IG \sim (0.001, 0.001)$.

O modelo (3.1) apresenta desafios de identificabilidade, uma vez que o nível geral pode ser absorvido tanto pelos efeitos espaciais quanto pelos efeitos temporais, e o termo de interação pode ser confundido com os efeitos principais. Para contornar esse problema, é prática comum impor restrições de soma zero aos efeitos aleatórios do modelo (conforme mostrado na Tabela 2). Essas restrições asseguram a identificabilidade do modelo (CRUZ, 2022).

Em resumo, o modelo bayesiano hierárquico utilizado permite incorporar incertezas em múltiplos níveis nas estimativas dos parâmetros do modelo. Essa abordagem hierárquica é comumente usada em modelos bayesianos e é particularmente útil quando há informações limitadas sobre os parâmetros do modelo e/ou quando os dados são complexos ou incompletos.

Ao longo deste estudo, foram desenvolvidos um total de 48 modelos, todos eles fundamentados nos mesmos elementos componentes presentes na Equação (3.1), incorporando informações das covariáveis no contexto espaço-temporal.

Tabela 2 – Restrições de identificabilidade para os modelos espaço-temporais na Equação (2.40) (CRUZ, 2022).

Interação	Espacial		Temporal
Tipo I	BESAG	$\sum_{i=1}^n u_i = 0$	$\sum_{t=1}^T \gamma_t = 0, \sum_{i=1}^n \sum_{t=1}^T \delta_{it} = 0$
	BYM	$\sum_{i=1}^n (u_i + v_i) = 0$	
	BYM2	$\sum_{i=1}^n b_i = 0$	
	LEROUX	$\sum_{i=1}^n u_i = 0$	
Tipo II	BESAG	$\sum_{i=1}^n u_i = 0$	$\sum_{t=1}^T \gamma_t = 0, \sum_{t=1}^T \delta_{it} = 0, \text{ para } i = 1, \dots, n$
	BYM	$\sum_{i=1}^n (u_i + v_i) = 0$	
	BYM2	$\sum_{i=1}^n b_i = 0$	
	LEROUX	$\sum_{i=1}^n u_i = 0$	
Tipo III	BESAG	$\sum_{i=1}^n u_i = 0$	$\sum_{t=1}^T \gamma_t = 0, \sum_{i=1}^n \delta_{it} = 0, \text{ para } t = 1, \dots, T$
	BYM	$\sum_{i=1}^n (u_i + v_i) = 0$	
	BYM2	$\sum_{i=1}^n b_i = 0$	
	LEROUX	$\sum_{i=1}^n u_i = 0$	
Tipo IV	BESAG	$\sum_{i=1}^n u_i = 0$	$\sum_{t=1}^T \gamma_t = 0$ $\sum_{t=1}^T \delta_{it} = 0, \text{ para } i = 1, \dots, n$ $\sum_{i=1}^n \delta_{it} = 0, \text{ para } t = 1, \dots, T$
	BYM	$\sum_{i=1}^n (u_i + v_i) = 0$	
	BYM2	$\sum_{i=1}^n b_i = 0$	
	LEROUX	$\sum_{i=1}^n u_i = 0$	

3.6 Métodos de avaliação dos modelos

A modelagem bayesiana requer uma avaliação cuidadosa da plausibilidade e ajuste do modelo, conhecida como crítica do modelo. Isso envolve tomar decisões importantes, como a seleção de variáveis, suposições sobre parâmetros e verossimilhança, além da atribuição de distribuições a priori para os parâmetros e hiperparâmetros.

Para avaliar, comparar o ajuste e a complexidade dos modelos, um método amplamente utilizado é o Critério de Informação de Deviance (DIC, do inglês *deviance information criterion*), que é expresso pela soma da média a posteriori da deviance (uma medida de ajuste do modelo) e o número efetivo de parâmetros (uma medida de complexidade do modelo), frequentemente menor que o número total de parâmetros. A penalização P_D refere-se a um termo adicionado à deviance do modelo para levar em consideração a complexidade do modelo, com o objetivo de evitar o *overfitting*. O DIC é dado por:

$$DIC = \bar{D} + P_D \quad (3.9)$$

em que $\bar{D}$ é a média a posteriori da deviance e P_D é o número efetivo de parâmetros (SPIEGELHALTER et al., 2002).

Valores menores de DIC e P_D indicam um melhor ajuste do modelo e uma melhor compensação entre ajuste e complexidade do modelo. SPIEGELHALTER et al. (2002) recomendam descartar o modelo com o DIC mais alto quando a diferença entre os modelos é maior que 10. Diferenças entre 5 e 10 são consideradas substanciais, mas se a diferença de DIC for menor que 5 e os modelos apresentarem muitas diferenças nas inferências, considerar apenas o modelo com o menor DIC pode ser enganoso.

Um outro método é o critério de informação de Watanabe-Akaike (WAIC) é uma medida de comparação de modelos estatísticos. Ele foi desenvolvido para avaliar a qualidade dos modelos de regressão ajustados a dados.

O WAIC é uma alternativa ao Critério de Informação de Akaike (AIC), que é amplamente utilizado em estatística para comparar modelos. Enquanto o AIC é baseado na máxima verossimilhança e penaliza a complexidade do modelo, o WAIC é baseado em uma abordagem bayesiana que leva em consideração a distribuição posterior dos parâmetros do modelo (VEHTARI; GELMAN; GABRY, 2017).

O WAIC fornece uma medida da capacidade preditiva do modelo, ajustando-se à adequação do modelo aos dados e a complexidade do mesmo. Quanto menor o valor do WAIC, melhor é a capacidade preditiva do modelo. O WAIC também pode ser usado para selecionar o melhor modelo entre um conjunto de modelos, onde o modelo com o menor valor do WAIC é geralmente escolhido. No entanto, o cálculo do WAIC pode ser computacionalmente intensivo e requer a amostragem de cadeias de Markov Monte Carlo (MCMC) para estimar a distribuição posterior dos parâmetros do modelo. Além disso, o WAIC pode ser sensível ao número de amostras e à escolha da distribuição priori (VEHTARI; GELMAN; GABRY, 2017).

4 Resultados

Na estimativa dos óbitos por COVID-19, o presente estudo adotou uma abordagem fundamentada na modelagem bayesiana hierárquica utilizando a distribuição de Poisson inflacionado de zeros. Nesse contexto, foram contemplados quatro distintos conjuntos de prioris espaciais, a saber, as abordagens de Besag, BYM, BYM2 e Leroux, a fim de incorporar efeitos espaciais estruturados. Além disso, prioris AR(1), RW(1) e RW(2) foram empregadas para modelar o efeito temporal estruturado, e foram considerados quatro tipos de interações espaço-temporais.

O estudo conduziu a estimativa mediante a implementação de um conjunto total de 48 modelos. A comparação entre esses modelos foi conduzida com base na avaliação de métricas pertinentes, incluindo o Critério de Informação de Deviance (DIC) e Watanabe-Akaike (WAIC). Os resultados detalhados desta avaliação encontram-se apresentados nas Tabelas 3 e 4 do presente trabalho.

Tabela 3 – Avaliação dos 48 modelos espaço-temporais ajustados para o número de óbitos por COVID-19 dos municípios do Estado de São Paulo de 2020 a 2021 considerando o DIC.

Modelo	Interação tipo I	Interação tipo II	Interação tipo III	Interação tipo IV
Besag+AR(1)	57006,05	56521,88	57090,28	57103,94
Besag+RW(1)	56997,89	56518,54	57107,38	57102,70
Besag+RW(2)	56987,18	56530,49	57106,95	57106,80
BYM+AR(1)	57009,64	56537,63	56989,66	57086,33
BYM+RW(1)	57004,06	56533,65	57069,71	57104,57
BYM+RW(2)	57017,45	56538,98	57105,97	57093,43
BYM2+AR(1)	57001,42	56542,16	57102,74	57133,28
BYM2+RW(1)	57005,22	56533,25	57131,51	57124,29
BYM2+RW(2)	57005,33	56533,25	57109,84	57118,51
Leroux+AR(1)	57038,06	56576,26	57102,44	57202,17
Leroux+RW(1)	57045,39	56577,75	57124,56	57175,24
Leroux+RW(2)	57048,81	56595,18	57115,63	57181,03

Tabela 4 – Avaliação dos 48 modelos espaço-temporais ajustados para o número de óbitos por COVID-19 dos municípios do Estado de São Paulo de 2020 a 2021 considerando o WAIC.

Modelo	Interação tipo I	Interação tipo II	Interação tipo III	Interação tipo IV
Besag+AR(1)	56823,83	56376,61	56690,38	56683,17
Besag+RW(1)	56817,19	56350,98	56689,33	56683,21
Besag+RW(2)	56813,73	56354,58	56693,95	56685,34
BYM+AR(1)	56832,59	56359,10	56633,38	56675,29
BYM+RW(1)	56820,82	56357,18	56627,40	56677,12
BYM+RW(2)	56829,05	56370,85	56633,91	56679,01
BYM2+AR(1)	56829,28	56371,89	56641,74	56688,13
BYM2+RW(1)	56826,27	56364,78	56641,09	56684,92
BYM2+RW(2)	56821,47	56360,07	56642,01	56686,34
Leroux+AR(1)	56892,36	56428,88	56710,03	56724,32
Leroux+RW(1)	56896,07	56432,23	56714,58	56773,61
Leroux+RW(2)	56889,48	56391,60	56715,67	56775,25

Dentre os modelos analisados, aquele que incorpora a autocorrelação espacial de Besag, a autocorrelação temporal RW(1) com a interação tipo II, além da inclusão de todas as covariáveis e efeitos aleatórios não estruturados espacial e temporal, destacou-se como o mais bem ajustado em comparação aos demais modelos. Essa conclusão se baseia nos menores valores de DIC e WAIC alcançados por esse modelo.

Com base nessas avaliações, o modelo final selecionado foi aquele que combina a autocorrelação espacial de Besag com a autocorrelação temporal RW(1) e a interação tipo II. A seguir, vamos explorar como esse modelo se ajustou aos dados.

Tabela 5 – Estimativas e médias a posteriori do modelo ajustado com combinação de autocorrelação espacial de Besag, autocorrelação temporal RW(1) e interação tipo II, com intervalos de credibilidade de 95% para o número de óbitos por COVID-19 nos municípios do Estado de São Paulo de 2020 e 2021.

Efeitos fixos	Média	Desvio Padrão	1º Quartil	Mediana	3º Quartil	Moda
Intercepto	1,5323	6,2757e-02	0,7010	1,5323	2,4100	1,5325
Óbitos gerais	1,0369	1,0337e-04	0,9880	1,0369	1,0761	1,0369
Casos	1,0098	7,4230e-06	0,9995	1,0098	1,0198	1,0098
População	1,0052	1,0955e-07	1,0032	1,0052	1,0072	1,0052
Efeitos aleatórios						
λ	0,6309	1,9520e-02	0,5931	0,6306	0,6699	0,6289
ρ	0,9999	4,9887e-05	0,9999	0,9999	0,9999	0,9998
α	0,4418	4,2267e-03	0,4331	0,4419	0,4498	0,4425
τ_u	2,4315	4,9533e-01	1,5339	2,4073	3,4737	2,3914
τ_v	0,9007	6,9494e-02	0,7645	0,9009	1,0375	0,9064
τ_γ	3,4075	1,2839e+00	1,3309	3,2813	6,2494	2,9504
τ_ϕ	307,6006	9,6867e+02	12,1207	102,0379	1870,2007	27,8681
τ_δ	2,4517	1,9520e-02	2,2246	2,4552	2,6621	2,4735

Os resultados apresentados na Tabela 5 refletem as estimativas e médias a posteriori do modelo hierárquico que possui autocorrelação espacial de Besag, autocorrelação temporal RW(1) e interação tipo II utilizado para analisar os óbitos por COVID-19 nos municípios do Estado de São Paulo de 2020 a 2021. A linha correspondente ao intercepto mostra que a média a posteriori é 1.5323. Isso significa que, considerando o modelo ajustado com os efeitos fixos e aleatórios especificados, a média estimada para o número de óbitos por COVID-19 nos municípios é de aproximadamente 1.5323. Além disso, os intervalos de credibilidade de 95% para os efeitos fixos, como intercepto, óbitos gerais, casos e população, fornecem faixas confiáveis que permitem afirmar, com 95% de confiança, os verdadeiros efeitos dessas variáveis. Notavelmente, constata-se que o aumento nos óbitos gerais, nos casos e na população está associado a um aumento no número de óbitos por COVID-19, pois há uma associação positiva entre essas variáveis e o número de óbitos por COVID-19 nos municípios do Estado de São Paulo.

Em relação aos efeitos aleatórios, a análise temporal revelou uma significativa estacionariedade, indicada pelo coeficiente ρ do modelo RW(1). O valor de 0.99 para este coeficiente sugere uma forte dependência temporal nos dados, ressaltando a importância da temporalidade na evolução dos óbitos por COVID-19. Adicionalmente, a média a posteriori do parâmetro de suavização espacial λ da distribuição de Besag para o efeito aleatório da área de ponderação é 0.63, apontando para uma moderada dependência espacial entre as áreas. Este resultado sugere uma correlação espacial reduzida entre regiões vizinhas, o que contribui para uma compreensão mais aprofundada da variabilidade espacial dos óbitos por COVID-19 no Estado de São Paulo. Essa baixa correlação pode ser atribuída, em parte, às políticas públicas implementadas durante a pandemia, como o isolamento social. A efetividade dessas medidas pode ter influenciado a disseminação do vírus de forma mais heterogênea, resultando em padrões menos correlacionados entre as áreas adjacentes e, assim, moldando a dinâmica da distribuição geográfica dos óbitos. Em conjunto, esses achados destacam a complexidade e a interconexão de fatores temporais e espaciais na dinâmica da pandemia na região.

A Figura 10 apresenta o gráfico dos resíduos padronizados vs valores ajustados do modelo ajustado com combinação de autocorrelação espacial de Besag, autocorrelação temporal RW(1) e interação tipo II, em que resíduos padronizados representam as diferenças entre os valores observados e os valores ajustados, divididos pelo desvio padrão dos resíduos. Este tipo de gráfico ajuda a identificar padrões sistemáticos nos resíduos, o que pode indicar problemas com a modelagem. Observamos que não há resíduos elevados negativamente. Talvez o modelo tenda a subestimar um pouco os valores.

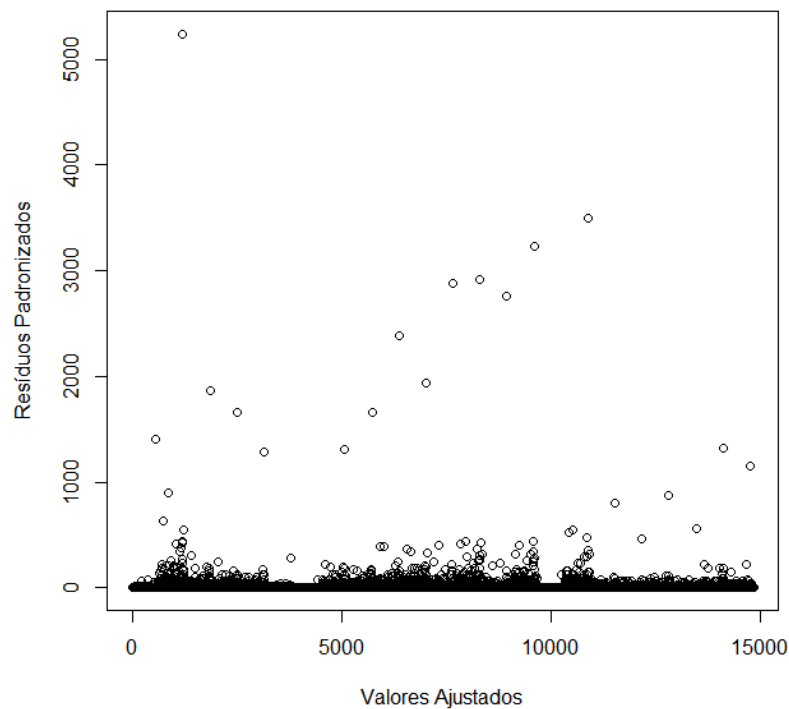


Figura 10 – Gráfico dos resíduos padronizados vs valores ajustados do modelo com a combinação de autocorrelação espacial de Besag, autocorrelação temporal RW(1) e interação tipo II .

A Figura 11 apresenta uma representação gráfica dos mapas que exibem as médias a posteriori dos valores ajustados do número de óbitos por COVID-19 em diferentes regiões geográficas ao longo do tempo. Nesses mapas, as cores mais escuras indicam áreas de maior risco, ou seja, aquelas com um maior número de óbitos.

É importante notar que as escalas de cores utilizadas nos mapas variam devido à ampla gama de valores observados ao longo do tempo. Isso reflete a significativa variação no número de óbitos ocorridos em diferentes períodos. Além disso, esses mapas permitem identificar o potencial impacto das regiões vizinhas nos valores de óbitos observados em uma determinada área.

Através dessa representação visual, é possível obter *insights* sobre as tendências e padrões espaciais do número de óbitos por COVID-19, bem como identificar áreas com maior vulnerabilidade. Essas informações são relevantes para orientar a implementação de medidas de controle e prevenção, direcionando recursos e esforços para regiões com maior necessidade.

Em resumo, os mapas com as médias a posteriori dos valores ajustados pelo modelo com a combinação de autocorrelação espacial de Besag, autocorrelação temporal RW(1) e interação tipo II proporcionam uma visão abrangente e espacialmente explícita da evolução dos óbitos por COVID-19, permitindo uma análise mais aprofundada das disparidades regionais e o impacto

das regiões vizinhas nesses valores.

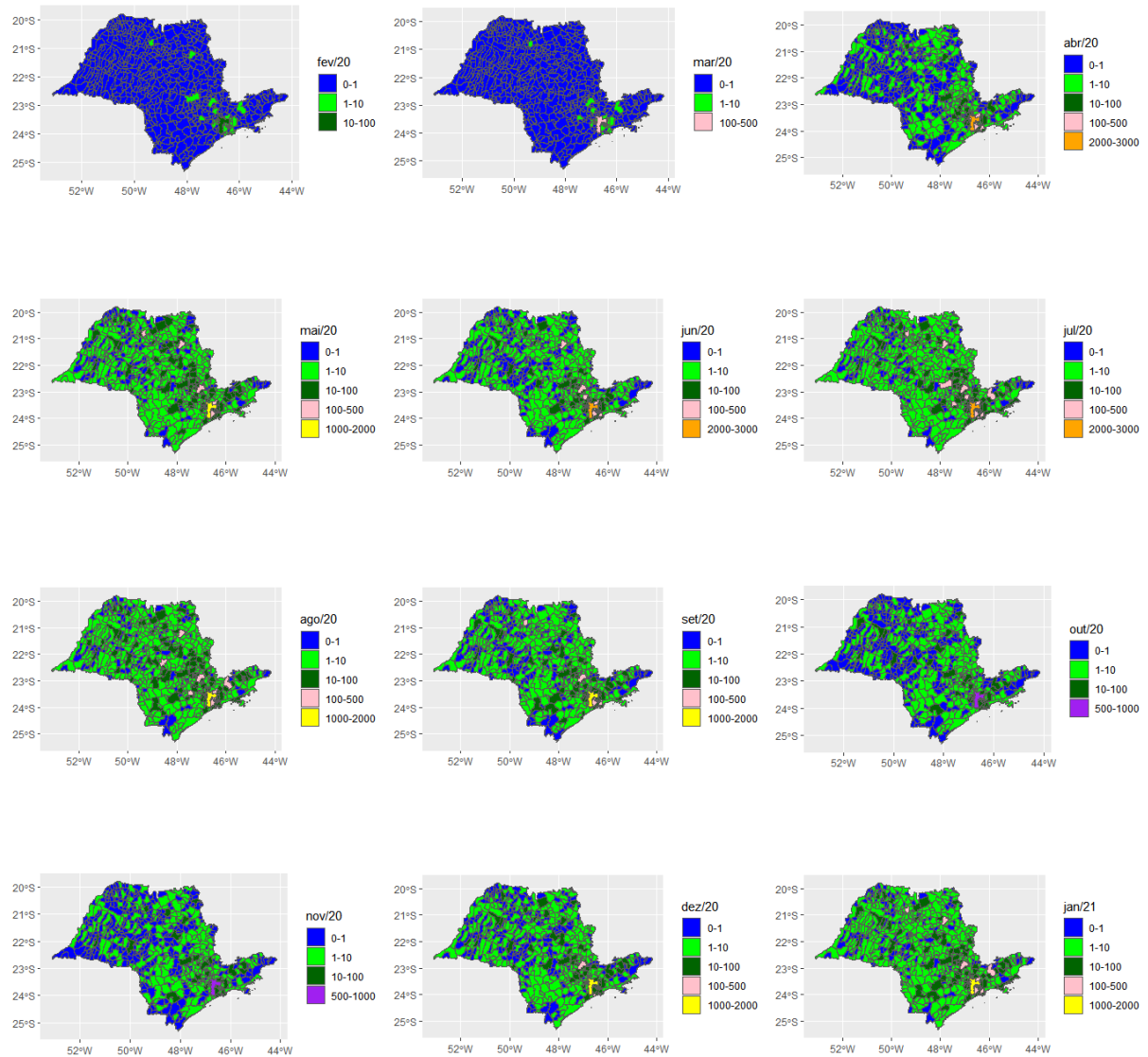
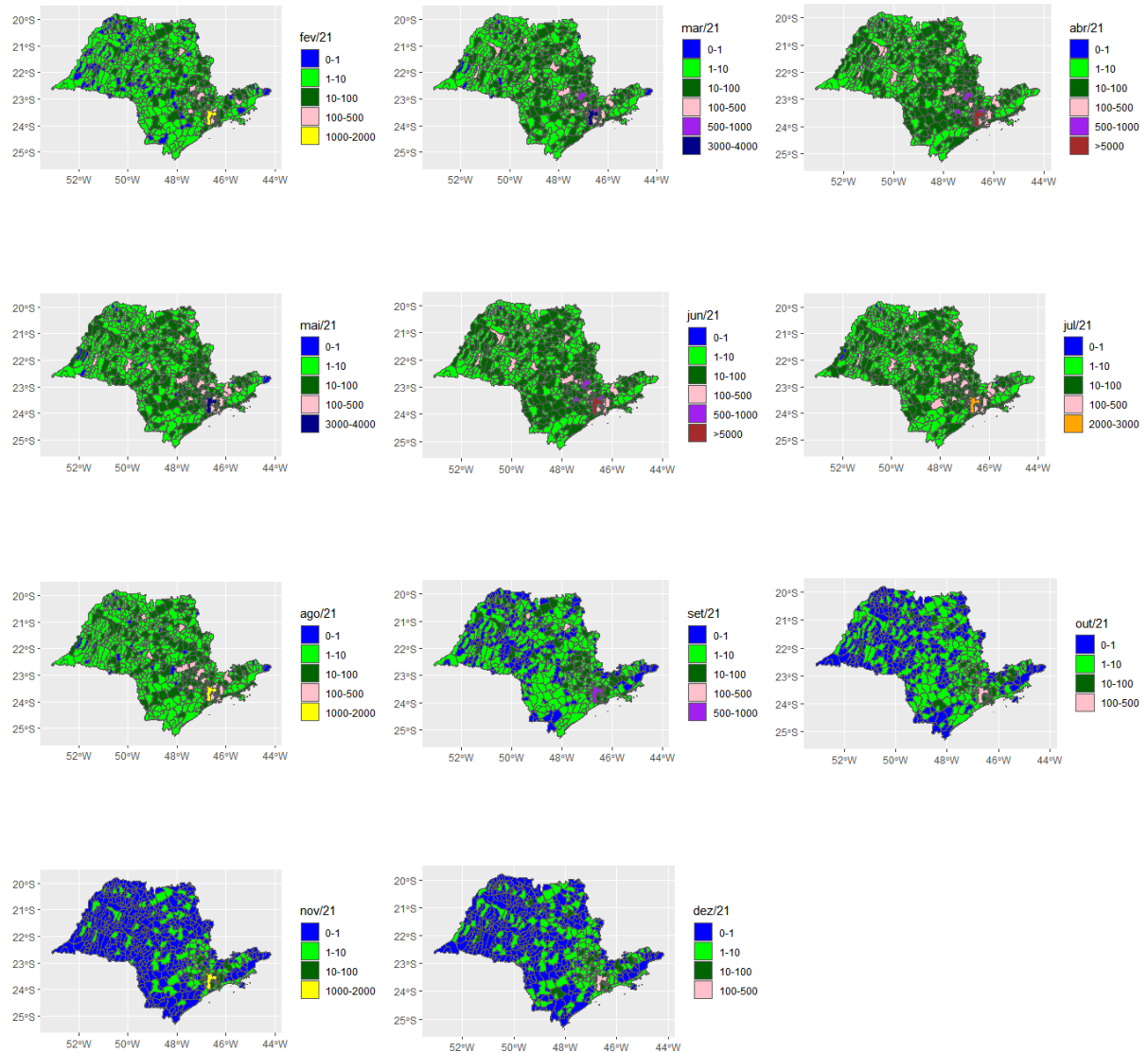


Figura 11 – Mapa das médias posteriori dos valores ajustados do número de óbitos por COVID-19 ao longo do tempo do modelo com a combinação de autocorrelação espacial de Besag, autocorrelação temporal RW(1) e interação tipo II. Continuação próxima página.



Na Figura 11, destaca-se a maior incidência de óbitos no Estado de São Paulo e sua região metropolitana. Além disso, é evidente que o modelo conseguiu capturar com precisão os picos da pandemia, alinhando-se com as análises descritivas dos dados. A capacidade do modelo em refletir de maneira precisa a intensidade dos eventos adversos em regiões específicas, como os picos de mortalidade, fortalece sua utilidade na análise da dinâmica temporal e geográfica dos óbitos por COVID-19. Através dela, é possível acompanhar o início da pandemia nos meses de fevereiro e março de 2020. Em abril de 2020, observa-se um aumento no número de cidades capturadas pelo modelo. Nos meses seguintes, há um crescimento no número de óbitos até julho de 2020. A partir de agosto até novembro, o número de óbitos diminui devido ao isolamento social adotado pelos municípios. Com as flexibilizações a partir de dezembro, no mês abril de 2021, ocorreu o pico no número de óbitos e novamente foi aplicado o isolamento social, e as cidades iniciaram as vacinações, resultando na diminuição do número de óbitos até dezembro de 2021.

4.1 Discussão

Este estudo teve como objetivo a comparação de diferentes modelos bayesianos hierárquicos na análise da evolução temporal do número de óbitos por COVID-19 no Estado de São Paulo durante os anos de 2020 e 2021.

A abordagem de conjuntos de dados espaço-temporais por meio de métodos bayesianos com algoritmos de Cadeias de Markov Monte Carlo (MCMC) é reconhecida por sua intensidade computacional devido à evolução lenta das distribuições a posteriori. Essa demanda por recursos computacionais requer um monitoramento cuidadoso das estimativas e verificações de convergência. Em contraste, o método INLA (Aproximação Laplace Aninhada Integrada) calcula as distribuições a posteriori por meio de integração numérica, eliminando a necessidade de análise de convergência (CARROLL, 2015). No entanto, as vantagens do modelo INLA são mais evidentes em conjuntos de dados espaço-temporais de maior escala. Neste estudo, foram analisadas 14.835 unidades espaço-tempo, correspondendo a 645 localizações e 23 meses.

Ao comparar os Critérios de Informação de Deviance (DIC) e de Watanabe-Akaike (WAIC) de 48 modelos, notou-se uma notável proximidade entre esses valores. Isso sugere que os modelos apresentam desempenho semelhante na análise dos dados, indicando comparabilidade em suas complexidades. A Figura 12 ilustra a comparação dos mapas das médias posteriores dos valores ajustados para o número de óbitos por COVID-19 nos meses de abril de 2020 e 2021, representando o início e o pico da pandemia, respectivamente. Os mapas à esquerda correspondem ao modelo que incorpora a autocorrelação espacial de Besag, a autocorrelação temporal AR(1) com a interação tipo II, e inclui todas as covariáveis, bem como efeitos aleatórios não estruturados espaciais e temporais, e os mapas à direita correspondem ao modelo que incorpora a autocorrelação espacial de Besag, a autocorrelação temporal RW(1) com a interação tipo II. Esses modelos, dentre os avaliados, apresentaram os valores de DIC mais baixos.

Notamos uma sutil disparidade nos mapas referentes a abril de 2020, especialmente em uma região específica do Estado. Essa discrepância pode ser atribuída ao baixo número de óbitos registrado nessa época, sendo a influência da vizinhança um fator provável que contribuiu para esse fenômeno. No entanto, contrastando com essa situação, em abril de 2021, período caracterizado pelo pico da pandemia, não observamos uma variação semelhante. Este fenômeno pode ser explicado pela maior quantidade de notificações nesse período, indicando que o aumento substancial de casos pode ter minimizado a influência da vizinhança na variação dos mapas.

Por meio de um modelo bayesiano hierárquico, foram obtidas estimativas mais suavizadas da incidência de óbitos por COVID-19 ao longo do tempo, contribuindo para atenuar as variações nos dados. Utilizando modelos espaço-temporais, observou-se um aumento no número de óbitos por COVID-19 ao longo do período considerado. Além disso, esses modelos identificaram áreas geográficas específicas com uma incidência persistente e notável de óbitos relacionados à COVID-19 ao longo do tempo.

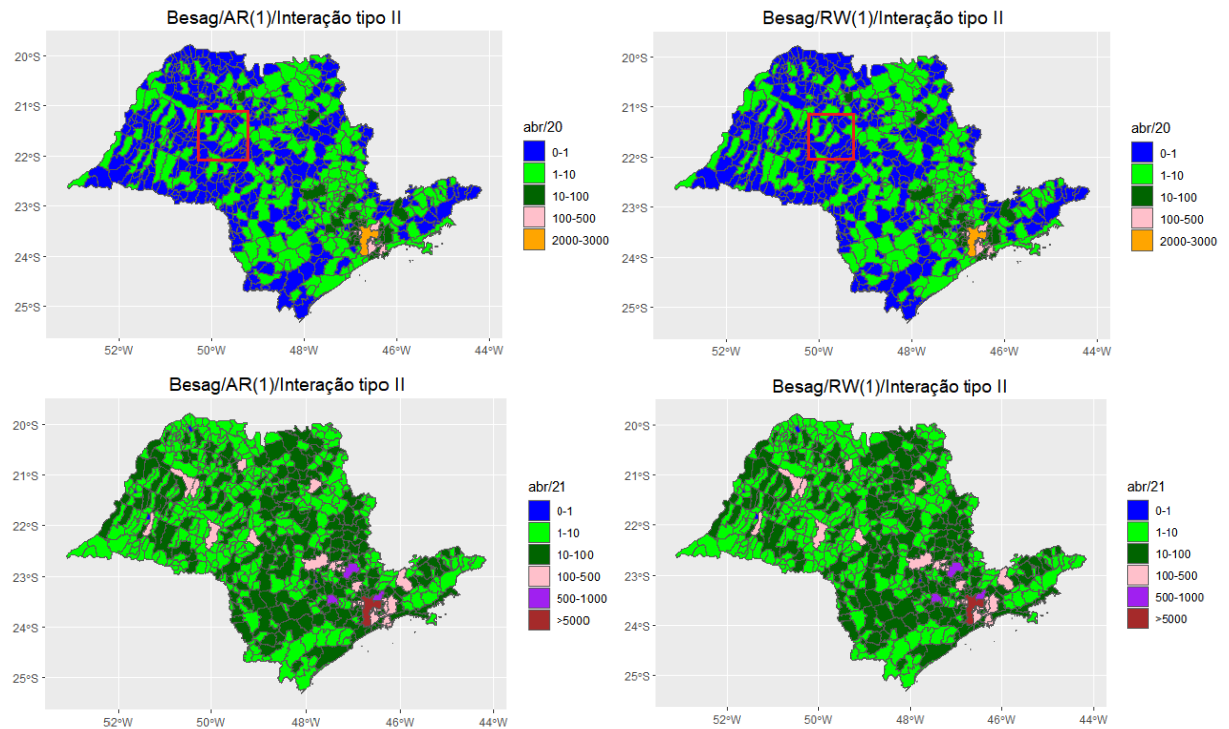


Figura 12 – Mapa das médias posteriori dos valores ajustados do número de óbitos por COVID-19 ao longo do tempo. À esquerda: modelo com autocorrelação espacial de Besag, autocorrelação temporal AR(1) com interação tipo II, covariáveis, e efeitos aleatórios não estruturados espaciais/temporais. À direita: modelo com autocorrelação espacial de Besag, autocorrelação temporal RW(1) com interação tipo II, covariáveis, e efeitos aleatórios não estruturados espaciais/temporais.

Entretanto, é crucial mencionar as limitações deste estudo, uma vez que se baseou em dados secundários na análise. Isso pode introduzir incertezas devido à potencial presença de incompletude, sub-registro, subnotificação, erros de classificação e preenchimento nos dados. Portanto, é fundamental reconhecer que as limitações associadas à qualidade das informações coletadas devem ser consideradas com rigor na interpretação dos resultados deste estudo.

Não foram encontrados artigos que realizam uma comparação direta entre os modelos mencionados. No entanto, é válido observar que existe um estudo relevante conduzido por [WAH, AHERN e EARNEST \(2020\)](#), o qual realiza uma revisão abrangente de modelos bayesianos espaço-temporais aplicados à incidência e mortalidade por câncer. A maioria dos modelos analisados nesse trabalho compartilha uma estrutura semelhante à utilizada neste estudo. O referido artigo conclui que a maioria dos modelos empregados em pesquisas dessa natureza foi categorizada com uma qualidade muito alta, evidenciando a eficácia e robustez dessas abordagens na análise de dados espaço-temporais relacionados à incidência e mortalidade por câncer.

5 Conclusão

O presente estudo teve como objetivo realizar uma comparação entre diversos modelos bayesianos hierárquicos, os quais foram empregados para investigar a evolução espaço-temporal do número de óbitos relacionados à COVID-19 no Estado de São Paulo, durante o período de 2020 e 2021. Neste trabalho, procedeu-se à adaptação de modelos bayesianos hierárquicos espaço-temporais, a fim de analisar a evolução espaço-temporal dos óbitos causados pela COVID-19 no Estado de São Paulo, no mencionado intervalo de tempo, visando à avaliação do desempenho desses modelos.

No âmbito desses modelos, incluímos a modelagem conhecida como CAR, que foi utilizada para conduzir a análise espacial dos dados relativos aos óbitos por COVID-19, os quais foram agregados em áreas ponderadas, de acordo com a divisão geográfica do estado. A escolha da modelagem CAR se fundamentou na natureza dos dados, que consistem em contagens agrupadas por unidades geográficas, e na capacidade intrínseca desses modelos de considerar a estrutura de vizinhança que existe entre essas unidades. Na análise temporal, empregamos a modelagem autoregressiva, incluindo AR(1), RW(1) e RW(2), que são os mais comumente utilizados na literatura. Contudo, a estrutura hierárquica inerente a esses modelos nos desafiou do ponto de vista computacional, desafio este que foi mitigado pela aplicação do método INLA para a inferência.

No contexto da análise estatística voltada para a investigação de doenças por meio de mapeamento, os modelos bayesianos hierárquicos têm adquirido ampla aceitação, dada sua capacidade de acomodar diferentes estruturas, incorporando efeitos aleatórios que permitem considerar informações espaciais, temporais e interações espaço-temporais. Contudo, é fundamental ressaltar que, à medida que se incorporam tais efeitos, a complexidade do modelo aumenta, o que requer uma abordagem cuidadosa no processo de desenvolvimento e implementação do modelo.

É importante observar que a escolha de um “melhor” modelo não implica que ele seja o mais apropriado para todas as situações, mas sim que auxilia na obtenção de uma compreensão robusta dos dados. A abordagem estatística empregada neste estudo constitui uma contribuição significativa para o campo da pesquisa, sendo particularmente indicada para cenários que envolvem a integração de dados provenientes de múltiplas fontes, em uma ampla escala geográfica e temporal. Essa abordagem aprimora a capacidade de realizar inferências precisas e oferece uma estrutura sólida para a análise de dados complexos.

Ao aplicar essa abordagem em outros contextos, pesquisadores e profissionais têm a oportunidade de explorar os benefícios de uma análise integrativa, incorporando informações provenientes de diversas fontes para obter uma compreensão mais abrangente e precisa dos fenômenos em estudo. Ademais, a utilização de escalas geográficas e temporais abrangentes possibilita a captura da variabilidade e das relações entre as variáveis ao longo do tempo e do espaço, contribuindo para uma melhor capacidade de inferência e tomada de decisões embasadas.

Como sugestão para trabalhos futuros, recomenda-se a realização de simulações dos modelos propostos. Essa abordagem permitirá verificar se o comportamento nos ambientes de dados controlados é congruente com as situações observadas nos dados reais. A execução de simulações proporcionará uma análise mais abrangente, possibilitando avaliar a robustez e a generalização dos modelos em diferentes contextos. Essa investigação adicional contribuirá para fortalecer a validade e a aplicabilidade dos modelos propostos em cenários práticos e fornecerá *insights* valiosos sobre seu desempenho sob condições diversas. Um exemplo da estrutura dos modelos utilizados está no Apêndice [B](#).

Referências

- AGRESTI, A. *Categorical data analysis*. [S.l.]: John Wiley & Sons, 2012. v. 792. 8
- ALLÉVIUS, B. On the precision matrix of an irregularly sampled ar(1) process. *arXiv preprint arXiv:1801.03791*, 2018. 18
- ANSELIN, L. *Spatial econometrics: methods and models*. [S.l.]: Springer Science & Business Media, 1988. v. 4. 20, 21
- ANSELIN, L. *Spatial econometrics: methods and models*. [S.l.]: Springer Science & Business Media, 1988. v. 4. 38, 39
- ATKINS, D. C. et al. A tutorial on count regression and zero-altered count models for longitudinal substance use data. *Psychology of Addictive Behaviors*, American Psychological Association, v. 27, n. 1, p. 166, 2013. 40
- BAILEY, T. C.; GATRELL, A. C. et al. *Interactive spatial data analysis*. [S.l.]: Longman Scientific & Technical Essex, 1995. v. 413. 21
- BAKSALARY, O. M.; TRENKLER, G. The moore-penrose inverse: a hundred years on a frontline of physics research. *The European Physical Journal H*, Springer, v. 46, p. 1–10, 2021. 42, 43
- BANERJEE, S.; CARLIN, B. P.; GELFAND, A. E. *Hierarchical modeling and analysis for spatial data*. [S.l.]: CRC press, 2014. 15, 24, 66
- BASTOS, L. S. et al. A modelling approach for correcting reporting delays in disease surveillance data. *Statistics in medicine*, Wiley Online Library, v. 38, n. 22, p. 4363–4377, 2019. 1
- BAYES, T. Lii. an essay towards solving a problem in the doctrine of chances. by the late rev. mr. bayes, frs communicated by mr. price, in a letter to john canton, amfr s. *Philosophical transactions of the Royal Society of London*, v. 53, p. 370–418, 1763. 4
- BERLINER, L. M. Hierarchical bayesian time series models. In: SPRINGER. *Maximum Entropy and Bayesian Methods: Santa Fe, New Mexico, USA, 1995 Proceedings of the Fifteenth International Workshop on Maximum Entropy and Bayesian Methods*. [S.l.], 1996. p. 15–22. 25
- BERNARDINELLI, L. et al. Bayesian analysis of space—time variation in disease risk. *Statistics in medicine*, Wiley Online Library, v. 14, n. 21-22, p. 2433–2443, 1995. 26, 27
- BESAG, J. Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society: Series B (Methodological)*, Wiley Online Library, v. 36, n. 2, p. 192–225, 1974. 2, 21, 23
- BESAG, J. Statistical analysis of non-lattice data. *Journal of the Royal Statistical Society: Series D (The Statistician)*, Wiley Online Library, v. 24, n. 3, p. 179–195, 1975. 23
- BESAG, J.; KOOPERBERG, C. On conditional and intrinsic autoregressions. *Biometrika*, Oxford University Press, v. 82, n. 4, p. 733–746, 1995. 42

- BESAG, J.; YORK, J.; MOLLIE, A. Bayesian image restoration, with two applications in spatial statistics. *Annals of the institute of statistical mathematics*, Springer, v. 43, p. 1–20, 1991. 2, 23, 41, 42
- BIASOLI, P. K. *Modelagem conjunta de média e variância em experimentos fracionados sem repetição utilizando GLM*. Dissertação (Mestrado), 2005. 7
- BLANGIARDO, M.; CAMELETTI, M. *Spatial and spatio-temporal Bayesian models with R-INLA*. [S.l.]: John Wiley & Sons, 2015. 11, 28, 43
- BOX, G. E. P. et al. *Time series analysis: forecasting and control*. [S.l.]: Princeton university press, 2015. 16, 44
- BOX, G. E. P.; TIAO, G. C. *Bayesian inference in statistical analysis*. [S.l.]: John Wiley & Sons, 2011. 66
- BROCKWELL, P. J.; DAVIS, R. A. *Introduction to time series and forecasting*. [S.l.]: Springer, 2002. 19, 45
- BROXSON, B. J. The kronecker product. UNF Digital Commons, 2006. 66
- CÂMARA, G. et al. *Anatomia de sistemas de informação geográfica-campinas: Unicamp. Instituto de Computação*, 1996. 15
- CARROLL, R. Comparing inla and openbugs for hierarchical poisson modeling in disease mapping. *Spatial and Spatio-temporal Epidemiology*, Elsevier, v. 14, p. 45–54, 2015. 55
- COWPERTWAIT, P. S. P.; METCALFE, A. V. *Introductory time series with R*. [S.l.]: Springer, 2009. 15
- CRESSIE, N.; WIKLE, C. K. *Statistics for spatio-temporal data*. [S.l.]: John Wiley & Sons, 2015. 25
- CRESSIE, N. A. C. *Statistics for spatial data*. Wiley, New York, NY, 1993. 15, 21, 23, 24
- CRUZ, R. F. Modelos espaço-temporais para mapeamento do risco de sífilis congênita e sua relação com as condições socioeconômicas e ambientais no município de são paulo de 2010 a 2016. Universidade Estadual Paulista (Unesp), 2022. ix, x, 6, 12, 16, 17, 18, 23, 24, 26, 27, 28, 29, 30, 31, 32, 33, 42, 43, 44, 46, 47
- DOBSON, A. J.; BARNETT, A. G. *An introduction to generalized linear models*. [S.l.]: CRC press, 2018. 6
- ELLIOTT, P.; WARTENBERG, D. Spatial epidemiology: current approaches and future challenges. *Environmental health perspectives*, v. 112, p. 998–1006, 2004. 14
- FIENBERG, S. E. When did bayesian inference become "bayesian". *Chance*, v. 19, n. 2, p. 49–52, 2006. 4
- FOX, J. *Applied regression, generalized linear models, and related methods*. Beverly Hills, Calif, USA: Sage Publications, 2008. 7
- GELMAN, A. et al. *Bayesian data analysis*. [S.l.]: Chapman and Hall/CRC, 1995. 2, 4, 5, 6
- GELMAN, A. et al. *Bayesian data analysis*. [S.l.]: CRC Press, 2013. 25

- GILKS, W. R.; RICHARDSON, S.; SPIEGELHALTER, D. *Markov chain Monte Carlo in practice*. [S.l.]: CRC press, 1995. 28, 30
- HAINING, R. P.; LI, G. *Regression Modelling With Spatial and Spatial-Temporal Data: A Bayesian Approach*. [S.l.]: CRC Press, 2020. 26
- HAMILTON, J. D. *Time series analysis*. [S.l.]: Princeton university press, 2020. 16, 66
- HASTIE, T. J. Generalized additive models. In: *Statistical models in S*. [S.l.]: Routledge, 2017. p. 249–307. 9
- HODGES, J. S. *Richly parameterized linear models: additive, time series, and spatial models using random effects*. [S.l.]: CRC Press, 2013. 18
- HUBIN, A.; STORVIK, G. Estimating the marginal likelihood with integrated nested laplace approximation (inla). *arXiv preprint arXiv:1611.01450*, 2016. 10
- HYNDMAN, R. J.; ATHANASOPOULOS, G. *Forecasting: principles and practice*. [S.l.]: OTexts, 2018. 14
- JALILIAN, A.; MATEU, J. A hierarchical spatio-temporal model to analyze relative risk variations of covid-19: a focus on spain, italy and germany. *Stochastic Environmental Research and Risk Assessment*, Springer, v. 35, p. 797–812, 2021. 41, 46
- JAYA, I. G. N. M.; FOLMER, H.; LUNDBERG, J. A joint bayesian spatiotemporal risk prediction model of covid-19 incidence, ic admission, and death with application to sweden. *The Annals of Regional Science*, Springer, p. 1–34, 2022. 41, 46
- 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. 27, 28, 29, 30, 31, 33
- LAURINI, R.; THOMPSON, D. *Fundamentals of spatial information systems*. [S.l.]: Academic press, 1992. 15
- LAWSON, A. B. *Bayesian disease mapping: hierarchical modeling in spatial epidemiology*. [S.l.]: CRC press, 2018. 21, 22
- LAWSON, A. B.; BROWNE, W. J.; RODEIRO, C. L. V. *Disease mapping with WinBUGS and MLwiN*. [S.l.]: John Wiley & Sons, 2003. v. 11. 27
- LEE, D. A comparison of conditional autoregressive models used in bayesian disease mapping. *Spatial and spatio-temporal epidemiology*, Elsevier, v. 2, n. 2, p. 79–89, 2011. 44
- LEE, D.; MITCHELL, R. Boundary detection in disease mapping studies. *Biostatistics*, Oxford University Press, v. 13, n. 3, p. 415–426, 2012. 44
- LEROUX, B. G.; LEI, X.; BRESLOW, N. Estimation of disease rates in small areas: a new mixed model for spatial dependence. In: SPRINGER. *Statistical models in epidemiology, the environment, and clinical trials*. [S.l.], 2000. p. 179–191. 2, 24, 43
- LINDSEY, J. K. *Statistical analysis of stochastic processes in time*. [S.l.]: Cambridge University Press, 2004. 17

- LÓPEZ-QUILEZLEZ, A.; MUNOZ, F. Review of spatio-temporal models for disease mapping. *Final Rep. EUROHEIS*, v. 2, p. 1–19, 2009. 26
- LUTKEPOHL, H. *New introduction to multiple time series analysis*. [S.l.]: Springer Science & Business Media, 2005. 66
- MACNAB, Y. C. On gaussian markov random fields and bayesian disease mapping. *Statistical Methods in Medical Research*, SAGE Publications Sage UK: London, England, v. 20, n. 1, p. 49–68, 2011. 43
- MCCULLAGH, P. *Generalized linear models*. New York, USA: [S.l.]: Chapman Hall, 1989. 7, 8
- MORAGA, P. *Geospatial health data: Modeling and visualization with R-INLA and shiny*. [S.l.]: CRC Press, 2019. 2, 5
- MORAGA, P. et al. A geostatistical model for combined analysis of point-level and area-level data using inla and spde. *Spatial Statistics*, v. 21, p. 27–41, 2017. 10
- MYERS, R. H. *Generalized linear models: with applications in engineering and the sciences*. [S.l.]: John Wiley & Sons, 2012. 6
- NELDER, J. A.; WEDDERBURN, R. W. Generalized linear models. *Journal of the Royal Statistical Society: Series A (General)*, v. 135, p. 370–384, 1972. 7
- QGIS DEVELOPMENT TEAM. QGIS: A Comprehensive Open Source GIS. v. 24, n. 12, p. 1785–1802, 2010. 34
- QGIS DEVELOPMENT TEAM. *QGIS Geographic Information System*. Chicago, 2021. Disponível em: <<https://qgis.org/>>. 34
- R CORE TEAM. *R: A Language and Environment for Statistical Computing*. Vienna, Austria, 2021. Disponível em: <<https://www.R-project.org/>>. 9, 34
- RESENDES, A. et al. *Abordagens espaciais na Saúde Pública-Capacitação e atualização em geoprocessamento em saúde*. Brasília: Editora MS, 2006. 139 p. 15
- REZAEIAN, M. et al. The production and interpretation of disease maps: A methodological case-study. *Social psychiatry and psychiatric epidemiology*, v. 39, p. 947–954, 2004. 14
- RODRIGUES, E. C. *Inferindo a estrutura de vizinhança em modelos bayesianos espaciais*. Dissertação (Mestrado) — Universidade Federal de Minas Gerais, 2011. 24
- RUE, H. Fast sampling of gaussian markov random fields. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, Wiley Online Library, v. 63, n. 2, p. 325–338, 2001. 18
- RUE, H.; HELD, L. *Gaussian Markov random fields: theory and applications*. [S.l.]: CRC press, 2005. 13, 18
- RUE, H. et al. R-inla: approximate bayesian inference using integrated nested laplace approximations. *Trondheim*, v. 23, p. 115, 2013. 9
- RUE, H.; MARTINO, S.; N.CHOPIN. Approximate bayesian inference for latent gaussian models by using integrated nested laplace approximations. *Journal of the royal statistical society: Series b (statistical methodology)*, v. 71, p. 319–392, 2009. 9, 10, 11, 34

- SAÚDE, M. D. Primeiro caso de coronavírus (covid-19) no brasil. Fevereiro 2020. Acessado em: 17 de maio de 2022. Disponível em: <<https://www.saude.gov.br/noticias/agencia-saude/46435-coronavirus-brasil-confirma-primeiro-caso-da-doencas>>. 1
- SAHU, S. K.; BÖHNING, D. Bayesian spatio-temporal joint disease mapping of covid-19 cases and deaths in local authorities of england. *Spatial Statistics*, Elsevier, v. 49, p. 100519, 2022. 41
- SHUMWAY, R. H.; STOFFER, D. S. *Time series analysis and its applications*. [S.l.]: Springer, 2000. v. 3. 14
- SIMÕES, T. C. *Modelagem da dinâmica espaço-temporal da abundância do Aedes aegypti, obtida com armadilhas de captura de mosquitos adultos*. Tese (Doutorado) — BVS, Rio de Janeiro, 2013. 5
- SIMPSON, D. et al. Penalising model component complexity: A principled, practical approach to constructing priors. *Statistical Science*, v. 32, n. 1, p. 1–28, 2017. 2, 43
- SPIEGELHALTER, D. J. et al. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, Oxford University Press, v. 64, n. 4, p. 583–639, 2002. 2, 47
- VANEM, E.; BITNER-GREGERSEN, E. M.; WIKLE, C. K. *Bayesian hierarchical space-time models with application to significant wave height*. [S.l.]: Springer, 2013. 25
- VEHTARI, A.; GELMAN, A.; GABRY, J. Practical bayesian model evaluation using leave-one-out cross-validation and waic. *Statistics and computing*, Springer, v. 27, p. 1413–1432, 2017. 2, 48
- WADE, K. M.; QUAAS, R. L. Solutions to a system of equations involving a first-order autoregressive process. *Journal of dairy science*, v. 76, n. 10, p. 3026–3032, 1993. 18
- WAH, W.; AHERN, S.; EARNEST, A. A systematic review of bayesian spatial–temporal models on cancer incidence and mortality. *International Journal of Public Health*, Springer, v. 65, p. 673–682, 2020. 56
- WAKEFIELD, J. Disease mapping and spatial regression with count data. *Biostatistics*, Oxford University Press, v. 8, p. 158–183, 2007. 20, 38
- WALLER, L. A.; GOTWAY, C. A. *Applied spatial statistics for public health data*. [S.l.]: John Wiley & Sons, 2004. 12
- WANG, Y.; FU, L.; PAUL, S. *Analysis of Longitudinal Data with Examples*. [S.l.]: CRC Press, 2022. 17
- WEI, W. W. S. *Time series analysis*. Pearson Addison Wesley, v. 33, 2013. 19
- WICKHAM, H. *ggplot2: Elegant Graphics for Data Analysis*. [S.l.]: Springer-Verlag New York, 2016. 34
- WORLD HEALTH ORGANIZATION. Novel coronavirus (2019-ncov): Situation report, 1. jan. 2020. Acessado em: 17 de maio de 2022. Disponível em: <https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200121-sitrep-1-2019-ncov.pdf?sfvrsn=20a99c10_4>. 1

WORLD HEALTH ORGANIZATION. *Statement on the second meeting of the International Health Regulations (2005) Emergency Committee regarding the outbreak of novel coronavirus (2019-nCoV)*. 2020. <[https://www.who.int/news/item/30-01-2020-statement-on-the-second-meeting-of-the-international-health-regulations-\(2005\)-emergency-committee-regarding-the-outbreak-of-novel-coronavirus-\(2019-ncov\)](https://www.who.int/news/item/30-01-2020-statement-on-the-second-meeting-of-the-international-health-regulations-(2005)-emergency-committee-regarding-the-outbreak-of-novel-coronavirus-(2019-ncov))>. Acessado em: 17 de maio de 2022. 1

Apêndices

A Produto de Kronecker

O produto de Kronecker é uma operação matemática que leva dois vetores ou matrizes e produz uma nova matriz cujas entradas são o produto de cada entrada da primeira matriz com a segunda matriz.

Seja A uma matriz $m \times n$ e B uma matriz $p \times q$. O produto de Kronecker de A e B , denotado por $A \otimes B$, é definido como a matriz $(mp) \times (nq)$ cujas entradas são dadas por:

$$(A \otimes B)_{ij} = A_{ij}B_{kl}$$

onde i, j variam de 1 a m e n , respectivamente, e k, l variam de 1 a p e q , respectivamente (BROXSON, 2006).

O produto de Kronecker é uma operação comutativa e associativa, ou seja, $A \otimes B = B \otimes A$ e $(A \otimes B) \otimes C = A \otimes (B \otimes C)$. Além disso, o produto de Kronecker também distribui sobre a adição, ou seja, $(A + B) \otimes C = A \otimes C + B \otimes C$ e $A \otimes (B + C) = A \otimes B + A \otimes C$ (BROXSON, 2006).

O produto de Kronecker é muito utilizado em estatística e processamento de sinais, especialmente na modelagem de séries temporais e na análise espectral. É uma ferramenta poderosa para a manipulação e análise de dados estruturados em forma de matrizes ou tensores.

A definição e as propriedades do produto de Kronecker podem ser encontradas em LUTKEPOHL (2005), BOX e TIAO (2011), BANERJEE, CARLIN e GELFAND (2014), HAMILTON (2020).

B Código

Aqui está um exemplo do código utilizado para implementar o modelo, que incorpora a autocorrelação espacial de Besag, a autocorrelação temporal AR(1) com interação tipo II, além da inclusão de covariáveis, efeitos aleatórios não estruturados espaciais e temporais.

Passo 1: Especificar a Estrutura do Modelo

```
modelo <- Yi ~ X1 + X2 + X3 +
f(variavel_espacial, model = "besag", graph = matriz_pesos_espaciais) +
f(variavel_tempo, model = "ar1") +
f(variavel_espacial, model = "iid") +
f(variavel_tempo, model = "iid") +
f(variavel_espacial, model = "iid", group = variavel_tempo,
control.group = list(model = "ar1"))
```

Neste passo, estamos especificando a estrutura do modelo. A formulação da regressão inclui variáveis preditoras (X_1, X_2, X_3), componentes espaciais (autocorrelação espacial, efeitos aleatórios espaciais), componentes temporais (autocorrelação temporal, efeitos aleatórios temporais) e a interação espacial e temporal.

Passo 2: Definir Priors no Segundo Nível

```
prior <- list(
list(family = "gaussian", param = c(0, 1e6)), # Prior para o intercepto
list(family = "gaussian", param = c(0, 1e-3)), # Prior para coef. preditores
list(family = "gaussian", param = c(1.5, 1e6)) # Prior para o parâmetro do ZIP
```

Aqui, estamos definindo priors Bayesianas para os parâmetros do modelo, usando distribuições gaussianas.

Passo 3: Definir Hiperpriors para os Parâmetros das Priors no Terceiro Nível

```
hyperprior <- list(  
  list(prior = "gamma", param = c(0.001, 0.001)), # Hiperprior para o  
  shape da distribuição log-gama  
  list(prior = "gamma", param = c(0.001, 0.001)), # Hiperprior para a  
  taxa da distribuição gama  
  list(prior = "gamma", param = c(0.001, 0.001)) # Hiperprior para o  
  shape da distribuição gaussiana  
)
```

Aqui, estamos definindo hiperpriors para os parâmetros das priors, usando distribuições gamma.

Passo 4: Ajustar o Modelo Hierárquico usando INLA

```
ajuste_hierarquico0 <- inla(modelo, data = banco_dados, family =  
  "zeroinflatedpoisson0",  
  control.compute = list(cpo = TRUE, dic = TRUE, waic = TRUE),  
  control.family = list(hyper = hyperprior),  
  control.predictor = list(hyper = prior),  
  verbose = TRUE)
```

Aqui, estamos ajustando o modelo hierárquico aos dados usando o pacote INLA. O modelo é ajustado aos dados contidos em `banco_dados`, usando a família de distribuição ZIP e configurações específicas para computação, família e preditores.