



MATHEUS FERES FREITAS

**MATRIZES DE VIZINHANÇA NÃO ESPACIAIS EM
MODELOS ESPAÇO-TEMPORAIS DA CLASSE STARMA:
UM ESTUDO DE CASO APLICADO A DADOS
EPIDEMIOLÓGICOS.**

LAVRAS- MG

2024

MATHEUS FERES FREITAS

**MATRIZES DE VIZINHANÇA NÃO ESPACIAIS EM MODELOS ESPAÇO-
TEMPORAIS DA CLASSE STARMA: UM ESTUDO DE CASO APLICADO A DADOS
EPIDEMIOLÓGICOS.**

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

Prof. Dr. Renato Ribeiro de Lima
Orientador

Prof. Dr. Thelma Sáfydi
Coorientadora

**LAVRAS- MG
2024**

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

Freitas, Matheus Feres.

Matrizes de vizinhança não espaciais em modelos espaço-temporais da classe STARMA: : um estudo de caso aplicado a dados epidemiológicos / Matheus Feres Freitas. - 2024.

96 p.

Orientador(a): Renato Ribeiro de Lima.

Coorientador(a): Thelma Sáfadi.

Tese (doutorado) - Universidade Federal de Lavras, 2024.

Bibliografia.

1. Matrizes de vizinhança não espaciais. 2. índice socioeconômico municipal. 3. modelos STARMA. I. Lima, Renato Ribeiro de. II. Sáfadi, Thelma. III. Título.

MATHEUS FERES FREITAS

MATRIZES DE VIZINHANÇA NÃO ESPACIAIS EM MODELOS ESPAÇO-TEMPORAIS DA CLASSE STARMA: UM ESTUDO DE CASO APLICADO A DADOS EPIDEMIOLÓGICOS.

CONSIDERING NON-SPATIAL NEIGHBORHOOD MATRICES IN SPACE-TEMPORAL MODELS OF THE STARMA CLASS: A CASE STUDY APPLIED TO EPIDEMIOLOGICAL DATA.

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

APROVADA em 25 de março de 2024.

Dr. Renato Ribeiro de Lima - UFLA

Dr. Luiz Otavio de Oliveira Pala - UFLA

Dr. Paulo Henrique Sales Guimarães - UFLA

Dr. Elias Silva de Medeiros- UFGD

Dr. Denismar Alves Nogueira- UNIFAL

Prof. Dr. Renato Ribeiro de Lima
Orientador

Prof. Dr. Thelma Sáfadi
Coorientadora

**LAVRAS
2024**

AGRADECIMENTOS

Agradeço primeiramente a Deus pela oportunidade de ter realizado mais um sonho. Foram muitos momentos memoráveis e algumas pessoas especiais que, conforme a vontade de Deus, colaboraram para que eu pudesse evoluir não apenas academicamente.

Aos meus pais, expresso minha profunda gratidão, o apoio e esforço de vocês foram essenciais em todas as fases, desde a graduação até o doutorado. O fato de estarem presente pela terceira vez, apoiando e acreditando em mais um dos meus sonhos, durante essa jornada acadêmica que já ultrapassa uma década, me enche de felicidade. Dedico de maneira especial este doutorado ao meu pai, que sempre sonhou ter um filho doutor.

Igualmente importante foi minha irmã, meu verdadeiro anjo guardião que ofereceu suporte em todas as dimensões possíveis - emocional, mental, física e, acima de tudo, espiritual! Sua tamanha fé e paciência foi fonte inesgotável de palavras de incentivo e de amor, sempre me reerguendo e sustentando de pé.

Mel, meu amor, muito obrigado pela companhia constante e pela paciência inesgotável. Você foi muito importante durante esta caminhada sempre me animando, incentivando e acreditando na minha capacidade (muitas das vezes até mais que eu mesmo).

Ao lado da minha amada, Deus me abençoou com meu filho, a quem dedico esta tese com todo o carinho do mundo. Bento, meu querido filho, sei que ainda não compreende completamente o significado do doutorado para o papai e também não entende por que, muitas vezes, troquei nossos preciosos momentos de interação por uma tela de notebook. No entanto, quero aproveitar esta oportunidade para explicar que o doutorado, juntamente com os dois empregos que precisei acumular (assim como todas as minhas escolhas desde o seu nascimento), foi sempre visando o melhor para nossa família, e principalmente para você. Pequeno, saiba que o papai te ama mais que tudo e espera que, um dia, compreenda e sinta orgulho desta longa jornada acadêmica. Esta tese é dedicada a você, representando todo o meu esforço e amor!

Minha gratidão pela paciência, pelos ensinamentos e direcionamentos do meu orientador Renato Ribeiro de Lima e também de minha coorientadora professora Thelma Sáfadi. O aprendizado se torna notavelmente mais fluido e agradável quando se pode contar com professores com tamanha didática e dedicação, profissionais esse que nem os desafios impostos pelo Covid-19, foram capazes de afastar de suas funções de orientadores. Estendo

também meus agradecimentos aos demais professores e técnicos do Departamento Estatística.

Grato à Antonio, Denize, Haiany e ao Lucas pela amizade advinda desde o mestrado. Foram extremamente importantes nessa caminhada, sempre me ajudando com dúvidas e sendo ombro amigo. Agradeço também aos meus velhos e grandes amigos que sempre me incentivaram: Cristhian, Josi, Jeyson, Gleice, Vanessa, Kamilla, Iara, Richard, Lais Gislaine, Rodrigo, Raquel, Rafaella.

Agradeço a todos meus familiares que sempre me incentivaram com palavras de carinho e confiança, renovando meu animo. Obrigado aos meus primos e tios, especialmente Paulo, Petryssa, Renan, Mitchel, Flavinho, Rita, Tia Andreza e Tio Marcelo, Tio Zezé e Tia Zeli, Tio Vitorio, Tio Sergio, Tia Jane e Tio Giovani.

Na Secretaria Municipal de Saúde de Campo Belo - MG (SMS), direciono minhas palavras de gratidão ao secretário Dr. José Assunção e à minha coordenadora imediata e amiga, Renata Guimarães. Profissionais exemplares, com tamanha sensibilidade, generosidade e coragem, a ponto de oferecer ao jovem engenheiro/estatístico, seu primeiro emprego formal. Trabalho esse que, na verdade, representou a oportunidade de, mesmo fora de minha área de formação, me sentir um profissional capaz e competente. Sempre solícitos às necessidades minhas e de minha família. Sem a compreensão e o apoio de vocês, não teria sido possível!

Minha trajetória na SMS me presenteou com amizades que sempre me confortaram com palavras de ânimo e atitudes que aquecem o coração! Profissionais que, aos poucos, me mostraram que, apesar das dificuldades enfrentadas pela saúde pública, vale sempre a pena se esforçar pelo próximo! Portanto, não posso deixar de expressar meu carinho a essas pessoas que sempre ajudaram a mim e a tantos outros: Percília, Roselaine, Izabella, Arielle, Lucas e Josi. Das 16 UBS também surgiram bons amigos, desde médicos, enfermeiros, técnicos de enfermagem e agentes de saúde, carinhosamente agradeço a Mayra, Leidiane, Livia, Luciano, André e Adriano.

Além dos amigos, a SMS trouxe também minha segunda família! Ana Paula, Carla, Paola, Karina, Leticia, Cinthia e Fabi. Ana Paula, você é muito mais que uma prima, somos irmãos de alma! Carla e Paola foram meus anjos, os melhores presentes da APS. Apesar das muitas horas extras dedicadas aos indicadores, adoro vocês, e teria sido muito mais difícil sem o suporte emocional que vocês e os outros enfermeiros da APS sempre me proporcionaram. Leticia e Karina, muito obrigado pela sinceridade e o companheirismo ao longo desses 3 anos, são amigas que levarei para toda minha vida! Cinthia e Fabi são as minhas irmãs da SMS, tornando o expediente mais leve e divertido. Sempre me auxiliando em tudo, dividindo nossas

tarefas de trabalho e também nossos fardos cotidianos. Grato a todos vocês, profissionais APS que fizeram renascer um novo Matheus, com olhar mais atento e humanizado.

Por fim, direciono meus agradecimentos à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pelo apoio financeiro, por meio da concessão da bolsa de doutorado, da qual eu pude desfrutar e que por muito tempo foi minha única fonte de renda. O presente trabalho só foi possível graças a esse subsídio.

RESUMO

Neste trabalho, foi estudada a viabilidade de se usar matrizes de vizinhança (\mathbf{W}) com base em critérios não espaciais, em modelos espaço-temporais da classe autorregressiva e de médias móveis (STARMA). Os dados utilizados consistem em uma série espaço-temporal composta por nove séries temporais que medem a incidência de tuberculose, observadas mensalmente entre 2002 e 2022, nas cidades mineiras Belo Horizonte, Betim, Contagem, Ibitaré, Nova Lima, Ribeirão das Neves, Sabará, Santa Luzia e Vespasiano. Para avaliar o impacto da matriz \mathbf{W} no ajuste do modelo, utilizou-se a matriz de contiguidade e outras cinco matrizes construídas por critérios não espaciais, buscando assim descrever não somente as interações entre áreas mas também intra-áreas. Essas matrizes foram geradas por um Índice Municipal Socioeconômico (IMS) derivado de combinações lineares de duas variáveis socioeconômicas: o Índice de Desenvolvimento Humano (IDH) municipal mais recente e a média das três avaliações de 2021 a 2023 do Previne Brasil, programa que avalia a qualidade do serviço prestado pela Atenção Primária à Saúde (APS) municipal. Seis modelos STARMA foram ajustados com as matrizes de vizinhança definidas. O ajuste do modelo foi realizado em três etapas: identificação, estimação e diagnóstico. Para a seleção do modelo, utilizou-se o Critério Bayesiano de Schwarz ou Critério de Informação Bayesiano (BIC). Concluiu-se que o melhor modelo foi obtido com uma \mathbf{W} não espacial, fortemente correlacionada à qualidade da saúde básica municipal. Nas previsões, o erro absoluto percentual médio (MAPE) foi utilizado como critério, observando-se que o modelo ajustado com a matriz de contiguidade apresentou aproximadamente 5% menos erro em comparação com o modelo que melhor se adequou aos dados. Este trabalho também demonstrou a necessidade de mais estudos no que tange ao uso de matrizes não espaciais, para responder questionamentos como: as matrizes \mathbf{W} não espaciais são pertinentes apenas para modelos espaço-temporais da classe STARMA? Os modelos aos quais esse tipo de matriz for adequado, são adequados a todos os tipos de dados? Qual a maneira de criação do índice que otimiza a construção da matriz \mathbf{W} não espacial?

Palavras-chave: índice socioeconômico municipal; matriz de vizinhança socioeconômica; tuberculose.

ABSTRACT

In this work, the feasibility of using neighborhood matrices (W) based on non-spatial criteria in space-time models of the autoregressive and moving average class (STARMA) was studied. The data used consist of a space-time series composed of nine temporal series measuring the incidence of tuberculosis, observed monthly between 2002 and 2022, in the cities of Belo Horizonte, Betim, Contagem, Ibirité, Nova Lima, Ribeirão das Neves, Sabará, Santa Luzia, and Vespasiano in the state of Minas Gerais, Brazil. To evaluate the impact of the W matrix on model fitting, the contiguity matrix and five other matrices constructed by non-spatial criteria were used, aiming to describe not only interactions between areas but also within areas. These matrices were generated by a Municipal Socioeconomic Index (IMS) derived from linear combinations of two socioeconomic variables: the most recent municipal Human Development Index (HDI) and the average of the 2021 to 2023 assessments from Previne Brasil, a program that evaluates the quality of service provided by municipal Primary Health Care (PHC). Six STARMA models were fitted with the defined neighborhood matrices. Model fitting was carried out in three stages: identification, estimation, and diagnosis. The Bayesian Information Criterion (BIC) was used for model selection. It was concluded that the best model was obtained with a non-spatial W , strongly correlated with the quality of municipal primary health care. In predictions, the mean absolute percentage error (MAPE) was used as a criterion, observing that the model fitted with the contiguity matrix had approximately 5% less error compared to the model that best fitted the data. This work also demonstrated the need for further studies regarding the use of non-spatial matrices to address questions such as: are non-spatial W matrices relevant only for STARMA-class space-time models? Are the models to which this type of matrix is suitable suitable for all types of data? What is the optimal way to create the index that optimizes the construction of the non-spatial W matrix?

Keywords: municipal socioeconomic index; socioeconomic neighborhood matrix. tuberculosis.

INDICADORES DE IMPACTO

Na tese são propostos modelos espaço-temporais inéditos, construídos a partir de matrizes de vizinhança não espaciais. Modelos da classe STARMA foram ajustados, utilizando-se matrizes de vizinhança que, ao invés de basear-se em critérios geográficos como proximidade física ou fronteiras, basearam-se em variáveis socioeconômicas, oferecendo uma nova perspectiva na modelagem e análise de dados. Esse enfoque inovador permitiu observar padrões e tendências, por uma perspectiva não fundamentada na análise espacial, mas em outras variáveis relevantes para o estudo. Este trabalho está diretamente relacionado aos Objetivos de Desenvolvimento Sustentável (ODS) da Organização das Nações Unidas (ONU), em particular o ODS 4, que trata da educação de qualidade. A pesquisa traz à tona um tema ainda pouco explorado na bibliografia acadêmica, implementado de maneira inédita. Com isso, não só amplia o conhecimento em áreas onde a análise espaço-temporal é relevante, mas também abre caminho para novas abordagens no ensino e na aplicação prática desses modelos. A originalidade e a relevância social do estudo proporcionam uma nova base de aprendizado e investigação para futuras gerações de acadêmicos e profissionais, contribuindo para o desenvolvimento de uma educação mais inclusiva e abrangente. Além da educação, a tese também está intimamente associada ao ODS 3, que se refere à saúde e bem-estar. Os modelos propostos foram aplicados em dados epidemiológicos, destacando o seu potencial na análise de informações relacionadas à saúde pública. A inovação desses modelos vai além da simples interpretação dos dados, oferecendo ferramentas mais robustas para a gestão e planejamento de políticas públicas. Ao considerar critérios socioeconômicos, a modelagem espaço-temporal proposta pode revelar desigualdades na distribuição de doenças e no acesso a serviços de saúde, o que é essencial para a formulação de estratégias mais eficazes de controle e prevenção de surtos. Com este trabalho, a comunidade acadêmica ganha uma nova vertente de pesquisa, que expande as fronteiras do conhecimento sobre a modelagem de dados complexos. A partir desta pesquisa, novos estudos poderão explorar e aprimorar os modelos propostos, promovendo avanços tanto na teoria quanto na prática. O uso desses modelos na análise e gestão de dados em saúde pública pode oferecer soluções inovadoras para desafios globais, como a prevenção de epidemias e a redução de desigualdades no acesso à saúde. Assim, ao integrar inovação metodológica com temas de alta relevância social, esta tese não só contribui para o avanço do conhecimento científico, mas também fortalece o potencial de impacto positivo em toda a sociedade, colocando em pauta a temática 6, voltada à Saúde, da Política Nacional de Extensão.

IMPACT INDICATORS

The thesis proposes unprecedented spatiotemporal models, constructed from non-spatial neighborhood matrices. STARMA class models were fitted using neighborhood matrices that, instead of relying on geographic criteria such as physical proximity or borders, were based on socioeconomic variables, offering a new perspective in data modeling and analysis. This innovative approach allowed the observation of patterns and trends from a perspective not grounded in spatial analysis but in other variables relevant to the study. This work is directly related to the United Nations (UN) Sustainable Development Goals (SDGs), particularly SDG 4, which addresses quality education. The research brings to light a topic still underexplored in the academic literature, implemented in a novel way. In doing so, it not only expands knowledge in areas where spatiotemporal analysis is relevant but also paves the way for new approaches in teaching and practical application of these models. The originality and social relevance of the study provide a new foundation for learning and investigation for future generations of scholars and professionals, contributing to the development of a more inclusive and comprehensive education. Beyond education, the thesis is also closely associated with SDG 3, which relates to health and well-being. The proposed models were applied to epidemiological data, highlighting their potential in the analysis of public health-related information. The innovation of these models goes beyond simple data interpretation, offering more robust tools for public policy management and planning. By considering socioeconomic criteria, the proposed spatiotemporal modeling can reveal inequalities in the distribution of diseases and access to health services, which is essential for formulating more effective outbreak control and prevention strategies. With this work, the academic community gains a new research avenue that expands the boundaries of knowledge about complex data modeling. From this research, new studies can explore and improve the proposed models, promoting advances in both theory and practice. The use of these models in the analysis and management of public health data may offer innovative solutions to global challenges, such as epidemic prevention and reducing inequalities in access to healthcare. Thus, by integrating methodological innovation with highly socially relevant topics, this thesis not only contributes to the advancement of scientific knowledge but also strengthens the potential for positive societal impact, bringing attention to Topic 6, Health, of the National Extension Policy.

LISTA DE ILUSTRAÇÕES

Figura 1- Gráfico de uma série temporal antes e após a aplicação de uma diferença simples e sazonal.	27
Figura 2- Mapa da grande Belo Horizonte- MG, enfatizando os nove municípios de interesse neste trabalho	38
Figura 3- Gráfico da série temporal das nove cidades de interesse: Taxa de incidência de TB entre janeiro de 2002 e junho de 2022.	44
Figura 4- Gráfico da série temporal das nove cidades de interesse: Taxa de incidência de TB entre janeiro de 2002 e junho de 2022 após diferenciação simples	45
Figura 5- Gráficos da stacf para cada W : da esquerda para a direita e de cima para baixo são exibidos os gráficos obtidos, respectivamente, para W_G , W_{IMS1} , W_{IMS2} , W_{IMS3} , W_{IMS4} e W_{IMS5} (sem defasagem espacial ‘slag 0’ e com uma defasagem espacial ‘slag 1’).	52
Figura 6- Gráficos da stpacf para cada W : da esquerda para a direita e de cima para baixo são exibidos os gráficos obtidos, respectivamente, para W_G , W_{IMS1} , W_{IMS2} , W_{IMS3} , W_{IMS4} e W_{IMS5} (sem defasagem espacial ‘slag 0’ e com uma defasagem espacial ‘slag 1’).	53
Figura 7- Gráficos da stacf residual: da esquerda para a direita e de cima para baixo são exibidos os gráficos obtidos, respectivamente, para W_G , W_{IMS1} , W_{IMS2} , W_{IMS3} , W_{IMS4} e W_{IMS5}	60

LISTA DE TABELAS

Tabela 1- Valores das variáveis socioeconômicas de interesse.	47
Tabela 2- Valores dos índices municipais socioeconômicos.	48
Tabela 3- Resultado do ajuste para o modelo STARMA ($8_1, 3_1$) considerando-se W_G	54
Tabela 4- Resultado do ajuste para o modelo STARMA ($5_1, 7_1$) considerando-se W_{IMS1} . .	55
Tabela 5- Resultado do ajuste para o modelo STARMA ($5_1, 7_1$) considerando-se W_{IMS2} . .	55
Tabela 6- Resultado do ajuste para o modelo STARMA ($5_1, 7_1$) considerando-se W_{IMS3} . .	56
Tabela 7- Resultado do ajuste para o modelo STARMA ($5_1, 7_1$) considerando-se W_{IMS4} . .	56
Tabela 8- Resultado do ajuste para o modelo STARMA ($5_1, 7_1$) considerando-se W_{IMS5} . .	57
Tabela 9- Parâmetros de comparação dos modelos STARMA ajustados	61

LISTA DE SIGLAS

APS	Atenção Primária à Saúde
AR	Modelo autorregressivo
ARMA	Modelo autorregressivo e de médias móveis
ARIMA	Modelo autorregressivo integrado e de médias móveis
BCG	Vacina contra o bacilo Calmette-Guérin
BH	Belo Horizonte
BIC	Critério de Informação Bayesiano
CL	Combinação Linear
DATASUS	Departamento de Informática do Sistema Único de Saúde
FPM	Fundo de Participação dos Municípios
GSTAR	Modelo espaço-temporal generalizado autorregressivo
GSTAR-X	Modelo espaço-temporal generalizado autorregressivo com intervenções
IBGE	Instituto Brasileiro de Geografia e Estatística
IDH	Índice de Desenvolvimento Humano
IMS	Índice Municipal Socioeconômico
LSTAR	Modelo espaço-temporal autorregressivo localizado
MA	Modelo de médias móveis
MAPE	Erro Absoluto Percentual Médio
PNCT	Programa Nacional de Controle de Tuberculose
SARIMA	Modelo sazonal autorregressivo integrado e de médias móveis
SMS	Secretaria Municipal de Saúde
stacf	Função de autocorrelação espaço-temporal
STAR	Modelo espaço-temporal autorregressivo
STARMA	Modelo espaço-temporal autorregressivo e de médias móveis
STARIMA	Modelo espaço-temporal autorregressivo integrado e de médias móveis
STCAR	Modelo espaço-temporal autorregressivo bayesiano
STMA	Modelo espaço-temporal de médias móveis

stpacf	Função de autocorrelação parcial espaço-temporal
SUS	Sistema Único de Saúde
TB	Tuberculose
TCU	Tribunal de contas da União
VARMA	Modelo autorregressivo e de médias móveis vetorial

SUMÁRIO

1	INTRODUÇÃO.	16
2	OBJETIVOS.	18
3	REVISÃO BIBLIOGRÁFICA.	19
4	REFERÊNCIAL TEÓRICO.	23
4.1	Tuberculose: contaminação, disseminação e tratamento	23
4.2	Séries temporais: principais conceitos e modelos.	25
4.3	Séries espaço-temporais: principais conceitos	28
4.4	Matriz de vizinhança (W).	29
4.4.1	Matriz de ponderação espacial baseada critérios geográficos	31
4.4.2	Matriz de ponderação espacial baseada critérios não espaciais	32
4.5	Modelos espaço-temporal STARMA e suas variações	34
5	MATERIAIS E METODOS	37
5.1	Dados	37
5.2	Métodos	38
5.2.1	Análises iniciais	39
5.2.2	Definição dos índices municipais socioeconômicos	39
5.2.3	Matrizes de vizinhança	40
5.2.4	Identificação, estimação e diagnóstico dos modelos STARMA.	40
6	RESULTADOS E DISCUSSÕES	43
6.1	Resultado das análises iniciais	43
6.2	Índice municipal socioeconômico	45
6.3	Obtenção das matrizes de vizinhança	48
6.4	Modelos STARMA	50
7	CONCLUSÕES.	63
8	CONSIDERAÇÕES FINAIS.	64
	REFERÊNCIAS.	65
	APÊNDICES.	70

1 INTRODUÇÃO

A análise e o ajuste de séries temporais desempenham um papel fundamental ao proporcionar informações que contribuem significativamente para tomadas de decisões nas mais diversas áreas do conhecimento. Como exemplo, temos: economia, epidemiologia, sismologia, psicologia, bem como estudos em agropecuária, sociais, hídricos e pluviométricos.

Os modelos espaço-temporais, que são mais recentes e um pouco mais complexos, apresentam dentre outras semelhanças com as séries temporais, a capacidade de explicar e prever os mais diversos fenômenos, concomitantemente ao longo do tempo e do espaço. Os modelos espaço-temporais são aplicáveis em diversas situações, como por exemplo: dados de criminalidade, econômicos, agrícolas, climáticos, ambientais, epidemiológicos e até mesmo para precificação de imóveis e madeira.

A distinção primordial entre modelos temporais e espaço-temporais está no número de variáveis correlacionadas. Enquanto nas séries temporais observa-se apenas correlação temporal, nos modelos espaço-temporais, o conjunto de dados apresenta correlação espacial e temporal.

De maneira concisa, compreende-se que um modelo espaço-temporal se ajusta a dados de uma variável específica, observada ao longo do tempo em diferentes localidades. Logo, é essencial compreender e modelar a interação entre cada uma das áreas onde a variável de interesse será mensurada.

As relações entre os locais de interesse são explicitadas pela matriz de ponderação espacial (\mathbf{W}), também conhecida como matriz de vizinhança. Essa matriz, que reflete o grau de similaridade entre as regiões de interesse, pode ser construída de diversas maneiras. Para os modelos espaço-temporais da classe autorregressiva e de médias móveis (STARMA), foco deste estudo, a obtenção de uma matriz \mathbf{W} apropriada é crucial e influenciará diretamente no processo de ajuste do modelo, bem como na qualidade das previsões dele advinda.

Geralmente, as matrizes de vizinhança são elaboradas com base em critérios geográficos, como a existência ou inexistência de fronteiras, o tamanho da fronteira compartilhada ou ainda as distâncias observadas entre as áreas na região de interesse.

Embora menos comum, também é possível obter matrizes \mathbf{W} com base em critérios socioeconômicos. Nesse caso, o grau de similaridade ou de dissimilaridade entre duas áreas na região de interesse é determinado por uma mesma variável socioeconômica mensurada nessas

áreas. As bibliografias e artigos sobre o assunto, sempre destacam o quão importante e subjetiva é a criação de uma matriz \mathbf{W} (MARTINS, BIZ e OZAKI, 2014). Em especial, as matrizes baseadas em critérios não espaciais são pouco discutidas e aplicadas, causando uma enorme necessidade de estudos mais aprofundados sobre as inúmeras possibilidades de obtenção de matrizes \mathbf{W} baseadas em critérios não geográficos, bem como onde e quando são mais efetivas que as matrizes \mathbf{W} tradicionais (baseadas em critérios espaciais).

Neste estudo investiga-se a pertinência do uso de matrizes de vizinhança não espaciais em modelos espaço-temporais da classe STARMA, aplicados a um conjunto de dados epidemiológicos específico. Apesar de possível, existem poucos exemplos e aplicações na literatura de matrizes socioeconômicas aplicadas em modelos STARMA ou ainda em modelos espaciais autorregressivos, que apesar de não serem assunto desse trabalho, podem ser citados como exemplos os modelos espaciais SAR e CAR.

Também é válido ressaltar, que nesse trabalho, as matrizes de vizinhança não espaciais foram construídas de maneira inovadora, por meio de um índice que consiste na combinação linear de duas variáveis socioeconômicas. Mostrando também mais uma nova possibilidade metodológica para criação de matrizes de vizinhança baseadas em critérios não espaciais. Assunto esse, que como dito anteriormente, requer ainda muito estudo e atenção da comunidade acadêmica.

2 OBJETIVOS

O principal objetivo deste estudo consistiu em avaliar o emprego de matrizes de vizinhança não espaciais no ajuste de modelos da classe STARMA, os quais foram aplicados a dados epidemiológicos.

Entretanto, para atingir satisfatoriamente tal objetivo, foi necessário o alcance de objetivos específicos, dentre os quais destacam-se os seguintes:

- a) desenvolver distintas matrizes de vizinhança (\mathbf{W}) como: matriz de contiguidade e outras cinco matrizes \mathbf{W} baseadas em combinações lineares de duas variáveis socioeconômicas;
- b) ajustar modelos espaço-temporais da classe STARMA utilizando as matrizes \mathbf{W} obtidas;
- c) validar o ajuste dos modelos por meio da extensão espaço temporal do teste de Box-Pierce;
- d) comparar os modelos ajustados por meio de diferentes critérios e concluir se a metodologia proposta no trabalho é pertinente ou não.

3 REVISÃO BIBLIOGRÁFICA

Os modelos para séries temporais, sejam univariados ou multivariados, são amplamente aplicados em diversas áreas de pesquisa. Esses modelos, comumente denominados séries temporais, destacam-se pela sua relativa simplicidade e eficácia preditiva. Apesar da sua simplicidade, os modelos univariados possuem muitas aplicações, fornecendo resultados satisfatórios na explicação e também na previsão de diversos fenômenos de interesse da comunidade acadêmica. Entre os modelos univariados, destacam-se o Autorregressivo Integrado de Médias Móveis (ARIMA) e o Sazonal Autorregressivo Integrado de Médias Móveis (SARIMA).

No contexto dos modelos estatísticos aplicados à saúde pública, as séries temporais desempenham um papel crucial. Isso é evidenciado não apenas pela considerável quantidade de publicações que utilizam esses modelos, mas também pelo interesse de revistas científicas, como a *Epidemiologia e Serviço de Saúde* e a *Revista Brasileira de Epidemiologia*, em publicar trabalhos que buscam elucidar conceitos e a aplicar metodologias de séries temporais. Exemplos desse tipo de publicação são os trabalhos Antunes e Cardoso (2015) e Latorre *et al.* (2001).

Com muitas características semelhantes às séries temporais, existem os modelos espaço-temporais. Os modelos espaço-temporais da classe STARMA, são a extensão espaço-temporal dos modelos ARMA. De acordo com Jin (2017), a classe STARMA pode ser compreendida, de maneira simplificada, como um caso específico dos modelos Vetoriais Autorregressivos e de Médias Móveis (VARMA). Como simplificações dos modelos STARMA tem-se os modelos Espaço-Temporal Autorregressivo (STAR) e Espaço-Temporal de Médias Móveis (STMA).

Os modelos espaço-temporais, especialmente da classe STARMA, são mais atuais em comparação com os modelos de séries temporais. Os primeiros trabalhos com modelos STARMA datam da década de 80, realizados por Pfeifer e Deutch (1980a). No entanto, a análise e ajuste de modelos espaço-temporais tem se mostrado útil em diversos estudos, diversas pesquisas buscam compreender tendências e comportamentos espaço-temporais de fenômenos em várias áreas do conhecimento.

Como demonstrado nos parágrafos subsequentes, a maior parte dos trabalhos científicos relevantes na área de saúde pública, não realizaram efetivamente o ajuste de modelos espaço-temporais, limitando-se apenas a abordar a tendência e/ou a distribuição espaço-temporal de epidemias. Isso porque tais estudos, em sua maioria, visam analisar a taxa de incidência de uma

determinada patologia e relacioná-la com indicadores econômicos, culturais e ambientais da região estudada.

O interesse por padrões e tendências espaço-temporais na saúde pública surgiu principalmente na segunda metade da década de 90. Um dos primeiros estudos dessa área foi Zwarewald, Bastos e Castilho (1998), que realizou uma análise espaço-temporal da epidemia de AIDS no Brasil entre 1987 e 1995, porém não foi ajustando nenhum modelo.

O estudo de aspectos espaço-temporais na propagação de epidemias tem particular relevância para doenças como HIV e tuberculose. Porém, inclui outras enfermidades, como malária e obesidade. Nobre, Schmidt e Lopes (2005) analisaram padrões espaço-temporais de incidência de malária no Pará. Enquanto Chaves *et al.* (2010) estudaram a evolução do sobrepeso e da obesidade em adolescentes do sexo masculino entre 1980 e 2005 em todos os estados brasileiros.

Padrões e tendências espaço-temporais da tuberculose foram abordados por Hino, Santos e Villa (2005), Melo *et al.* (2012) e Santos Neto *et al.* (2015). O primeiro trabalho limitou-se aos casos de tuberculose em Ribeirão Preto – SP, entre 1998 e 2002. O segundo considerou casos de tuberculose de 1997 a 2006 em Rondônia e na Amazônia Ocidental. Santos Neto e colaboradores analisaram casos confirmados de tuberculose em São Luiz – MA, entre 2008 e 2012.

Guimarães *et al.* (2012), em um estudo abrangente, também investigaram padrões espaço-temporais de tuberculose, com foco na tendência temporal da doença em diversas regiões. Os autores exploraram a tendência espaço-temporal da tuberculose, do HIV e da pobreza, visando estudar a relação entre as patologias e o nível econômico do Brasil, das Américas e do mundo.

A disseminação de tuberculose também foi de interesse de Amsalu *et al.* (2019), que ajustou modelos temporais, espaciais e espaço-temporais à disseminação de tuberculose em idosos acima de 65 anos na China. Esses autores utilizaram modelos autorregressivos espaciais, modelos de séries temporais e modelos espaço-temporais autorregressivos. Dentre as conclusões, os autores destacam que é possível encontrar agrupamentos espaciais onde a incidência de tuberculose se mostra muito elevada e que tais regiões apresentam baixo nível de apoio social. Na questão temporal observou-se sazonalidade da incidência da tuberculose, observando-se picos na primavera.

A disseminação do vírus HIV foi objeto de investigação por Rodrigues-Junior e Castilho (2004), os quais descreveram uma análise espaço-temporal considerando diagnósticos realizados em todo o território nacional entre 1991 e 2000. Taquette *et al.* (2011) abordaram o

mesmo assunto, concentrando-se nos casos ocorridos na cidade do Rio de Janeiro com adolescentes entre 13 e 19 anos. Rodrigues-Junior e Castilho (2009) aprofundaram a análise dos padrões espaço-temporais da epidemia de AIDS nas fronteiras do Brasil, contudo, limitaram-se ao ajuste de modelos de séries temporais, assim como todos os demais trabalhos supracitados, em que o ajuste de modelos espaço-temporais não foi feito.

Prado *et al.* (2009) apresentou um modelo espaço-temporal ajustado ao número de casos diagnosticados de AIDS em São Paulo entre 1990 e 2005, sendo ajustado o modelo STAR, fundamentado em uma abordagem bayesiana. Szwarcwald *et al.* (2001) ajustou um modelo espaço-temporais a dados epidemiológicos de HIV, apresentando um modelo da classe STAR ajustado ao número de casos confirmados na cidade do Rio de Janeiro entre 1988 e 1996.

Assim como as séries temporais, a análise e o ajuste de modelos espaço-temporais da classe STARMA são aplicáveis a uma enorme gama de áreas além da epidemiologia e saúde pública, sendo inclusive mais comum em outras aplicações. Como exemplo, citam-se Martins, Biz e Ozaki (2014) que utilizaram modelos da classe espaço-temporal autorregressivo e de médias móveis (STARMA) para ajustar dados climáticos oriundos de oito estações meteorológica do Paraná. O mesmo foi feito por Rao e Antunes (2004) para dados climáticos de nove regiões do Reino Unido.

Kamarianakis e Prastacos (2005) utilizaram modelos da classe STARMA buscando descrever o fluxo de tráfego na cidade de Atenas. A mesma classe de modelo foi utilizada no intuito de prever o preço trimestral de madeira na serraria de pinho em onze estados do Estados Unidos (ZHOU E BOUNGIORNO, 2006). Modelos STARMA também foram utilizados na previsão da taxa de crimes violentos em nove regiões da Flórida (Shoesmith, 2013), na previsão de enchentes em Pequim (Zheng *et al.*, 2014) e na precificação de imóveis em seis cidades do Estado de São Paulo (Jin, 2017).

Lee, Rushworth, e Napier (2018) ajustaram modelos espaço-temporais bayesianos autorregressivos (da classe STCAR) utilizando-se o pacote *CARBayesST* do software R (LEE, 2013). Além disso, o trabalho apresenta três exemplos: um estudo de simulação e duas aplicações a dados obtidos junto ao governo escocês (avaliação do mercado imobiliário no país e impacto da poluição do ar na saúde dos escoceses).

Dentre as possíveis variações dos modelos STAR estão os modelos espaço-temporais generalizados autorregressivos (GSTAR). Nesse contexto destacam-se dois trabalhos: Nurhayati, Pasaribu e Neswan (2012), que utilizou modelos da classe GSTAR na análise de dados do Produto Interno Bruto (PIB) de países do oeste europeu. Já Astuti, Ruchjana e

Soemartini (2017) usaram modelos GSTAR e variações para a predição de volume de óleo bruto de palma (CPO) exportado de algumas regiões da Indonésia.

Na saúde pública, o modelo GSTAR foi utilizado recentemente para análise de dados da pandemia do Covid-19. Ip *et al.* (2021) utilizaram um modelo GSTAR-X com intervenções, para analisar o número cumulativo diário de casos COVID-19 confirmados em oito estados da Austrália, entre os dias 25 de janeiro de 2020 a 12 de setembro de 2020. Os autores encontraram um grande número de intervenções (pontos atípicos) e concluíram que o modelo foi adequado para o conjunto de dados.

Chen *et al.* (2018) utilizou uma modificação do modelo STAR de maneira que o modelo se adequasse ao monitoramento em tempo real do fluxo de automóveis em vias públicas. O modelo denominado espaço-temporal autorregressivo localizado (LSTAR) é obtido utilizando-se uma matriz de vizinhança (\mathbf{W}) dinâmica que reflete as taxas de conversão de tráfego atuais observadas em cada interseção de estrada.

A aplicação de modelos STARMA e suas variações, seja em dados epidemiológicos ou de qualquer outra natureza, é ainda pouco explorada, quando comparados a estudos com séries temporais. Ademais o uso de STARMA com aplicação de matrizes \mathbf{W} não espaciais é um assunto ainda menos explorado. Os poucos trabalhos existentes são mais aplicados a modelos espaciais e de econometria. Como exemplos citam-se: Freitas (2019) e Eff (2008). O primeiro tenta, sem sucesso, ajustar modelos da classe STARMA com \mathbf{W} socioeconômicas, enquanto que o segundo trabalho usa matrizes \mathbf{W} baseadas em proximidade cultural, buscando explicar a filogenia linguística por meio de modelos bipartidos com defasagem espacial.

4 REFERENCIAL TEÓRICO

No referencial teórico, inicialmente será abordado as principais características e temas pertinentes a tuberculose, uma vez que a taxa de incidência de tal patologia é tema da presente pesquisa.

Buscando mais clareza e por julgar ser mais didático, na subseção 4.2 serão apresentados fundamentos e conceitos básicos além dos principais modelos para séries temporais. Para que em seguida, na subseção 4.3 sejam apresentados por analogia, os conceitos e fundamentos dos modelos espaço-temporais.

Antes da apresentação das equações que descrevem os modelos espaço-temporais da classe STARMA, será tratado na subseção 4.4, a obtenção da matriz de vizinhança (\mathbf{W}). Uma vez que tal matriz é parte integrante do modelo. É importante defini-la e apresentar as possíveis maneiras de construí-la.

O referencial teórico será finalizado apresentando o modelo STARMA e suas variantes STAR e STMA, em que ficará ainda mais notório o impacto da matriz \mathbf{W} no ajuste do modelo.

4.1 Tuberculose: contaminação, disseminação e tratamento

A tuberculose (TB) é uma doença transmissível causada por micobactérias aeróbicas, com destaque para o *Micobacterium tuberculosis*, responsável pela tuberculose em humanos, e outras variantes como *M. africanum*, *M. bovis*, *M. caprae*, *M. pinnipedii* e *M. microti*, que afetam mamíferos silvestres e domesticados (GAGNEUX, 2018).

Inicialmente conhecida como "Consumo", devido a sua alta capacidade destrutiva do infectado, a TB geralmente manifesta-se com ciclos persistentes de febre, tosse sangrenta, dores no peito e perda significativa de massa muscular. No século XVII, foi apelidada de "capitão de todos esses homens de morte" devido à sua devastação. Mesmo com esforços das autoridades de saúde, a TB dominou como infecção crônica durante a primeira metade do século 20, incluindo a Segunda Guerra Mundial (MISRAH; SUROLIA, 2018).

A transmissão da TB ocorre exclusivamente pelas vias respiratórias, quando gotículas contendo o bacilo são expelidas por pessoas infectadas e inaladas por não infectadas. Embora o pulmão seja o órgão mais afetado, a TB extrapulmonar contribui com 10-20% dos casos, podendo atingir até 60% na coinfeção com HIV (BETHLEM, 2012). Manifestações extrapulmonares incluem pleural, linfonodal, urogenital e, raramente, comprometimento do sistema nervoso central (RIBEIRO *et al.*, 1995; BETHLEM, 2012). A sintomatização e

agudização do quadro clínico causado pela TB decorre do uso das reservas nutricionais do hospedeiro, levando a déficits funcionais significativos e, em casos extremos, óbito (MISRAH e SUROLIA, 2018).

Diagnosticar a tuberculose foi historicamente sempre desafiador, mas avanços tecnológicos, como a tomografia computadorizada, melhoraram a precisão. Métodos de diagnóstico atuais incluem baciloscopia, teste molecular rápido e cultura para micobactéria. O Programa Nacional de Controle da Tuberculose (PNCT) no Brasil teve sucessos notáveis, reduzindo significativamente taxas de incidência e mortalidade (BRASIL, 2011). Em 2016, 10,4 milhões de pessoas foram infectadas com 1,8 milhão de óbitos relacionados à tuberculose (PELLISSARI; DIAZ-QUIJANO, 2017). Apesar dos avanços, a tuberculose permanece uma preocupação de saúde pública, evidenciando a necessidade de esforços contínuos para seu controle e principalmente para sua prevenção.

A prevenção da tuberculose é crucial devido à sua alta transmissibilidade e ao impacto devastador na saúde individual e coletiva. A estratégia de prevenção não apenas reduz a incidência da doença, mas também desempenha um papel fundamental na minimização da resistência antimicrobiana, que representa uma ameaça crescente no controle da tuberculose (WHO, 2022). Além disso, investir em prevenção não só preserva a saúde dos indivíduos, mas também alivia a carga econômica associada ao tratamento prolongado da TB na rede pública de saúde. A seguir são elencadas as principais formas de prevenção da TB, sendo importante destacar que todas elas são desempenhadas quase que exclusivamente na Atenção Primária à Saúde (APS), setor esse do Sistema Único de Saúde (SUS) que tem caráter essencialmente preventivo e eletivo e que, neste trabalho será utilizado como um dos parâmetros para construção das matrizes **W**.

- a) vacinação: a vacina BCG (Bacilo Calmette-Guérin) é uma ferramenta eficaz na prevenção da tuberculose, especialmente em crianças. Apesar de sua eficácia variável em diferentes populações, a vacinação continua sendo uma estratégia essencial (Colditz *et al.*, 1994);
- b) identificação e tratamento precoce e adequado: o rastreamento ativo de casos e diagnóstico precoce são fundamentais para interromper a cadeia de transmissão, dando especial atenção a grupos com maior vulnerabilidade à TB, como por exemplo, portadores do HIV ou pacientes com vulnerabilidade socioeconômica como, por exemplo, moradores de rua. Além disso, garantir o tratamento completo e adequado dos casos identificados é essencial para prevenir a

progressão da doença e a ocorrência de cepas resistentes aos medicamentos (WHO, 2022);

- c) educação e conscientização: Campanhas de conscientização pública sobre os sintomas da TB, sua transmissão e a importância do tratamento adequado, desempenham um papel essencial na prevenção, reduzindo o estigma e promovendo a busca por cuidados de saúde (BRASIL, 2011).

Em 2005, a tuberculose era a principal causa de morbi-mortalidade mundial por doenças infecciosas, representando um quarto das mortes evitáveis em adultos (COELHO *et al.*, 2010). A incidência da TB varia globalmente, com 21% dos casos em países desenvolvidos e 79% em países em desenvolvimento (COELHO *et al.*, 2010). Embora os índices de infecção tenham diminuído, a tuberculose ainda é classificada como uma doença negligenciada pela Organização Mundial da Saúde (BRASIL, 2010).

A tuberculose, ocasionada principalmente pelo *M. tuberculosis*, pode ter sua predisposição aumentada por fatores genéticos e ambientais (FERREIRA *et al.*, 2005). Entre os fatores predisponentes destacam-se a coinfeção por HIV, a pobreza e o crescimento populacional desordenado (GUIMARÃES *et al.*, 2012; HIJJAR 2001). Guimarães (2012) acrescenta ainda que a gravidade e o impacto negativo da doença variam na razão inversa do índice de desenvolvimento humano (IDH). A relação evidenciada por esses estudos entre a TB e fatores socioeconômicos sinaliza a pertinência e necessidade desse e de outros estudos que buscam o ajuste de modelos matemáticos que considerem os fatores mais preponderantes na disseminação de TB e de tantas outras patologias.

4.2 Séries temporais: principais conceitos e modelos

Uma das diversas abordagens para realizar previsões envolve a aplicação de análise e ajuste de séries temporais. O termo "séries temporais" refere-se a conjuntos de dados organizados temporalmente, sugerindo, assim, a presença de autocorrelação temporal nesse conjunto de dados (MORETTIN E TOLOI, 2006).

A exploração de séries temporais é uma prática disseminada em pesquisas (BOX; JENKINS; REINSEL, 2008) de diversas áreas do conhecimento, como Economia, Epidemiologia, Medicina, Sismologia, Geofísica, Meteorologia, Geodesia (NASCIMENTO *et al.*, 2017; SHUMWAY; STOFFER 2006), além de Ciências Sociais e Políticas, Sociologia, História e Psicologia (SÁFADI, 2004).

Uma das características essenciais de séries temporais é a presença de dependência entre cada observação e pelo menos uma das observações que a sucede ou antecede. Se houver uma relação de dependência entre duas ou mais variáveis de um conjunto de dados, na qual uma pode ser expressa em função da outra, diz-se que essas variáveis estão correlacionadas. Analogamente, para o caso de séries temporais, tem-se que a relação de dependência é observada de uma variável em relação a ela mesma, mas em tempos ou localizações diferentes, e por isso, essa relação de dependência é denominada autocorrelação (MORETTIN; TOLOI, 2006).

Quando pretende-se analisar o comportamento de um conjunto de dados e realizar previsões por meio de séries temporais, é crucial considerar a existência de autocorrelação temporal em tal conjunto de dados. Embora amplamente utilizado para previsões, a análise de séries temporais vai além da capacidade preditiva. Os pesquisadores podem ter interesse em compreender o comportamento da série, estudar a tendência temporal ou identificar a presença de sazonalidade na série.

A estacionariedade de uma série temporal ocorre quando ela não apresenta nem tendência e nem sazonalidade, refletindo um desenvolvimento aleatório em torno de um valor constante. A componente de tendência é percebida pelo comportamento crescente e/ou decrescente e pode ser visualizada graficamente. No entanto, Brockwell, Davis e Calder (2002) e Morettin e Toloi (2006) ressaltam a importância de testes apropriados como, por exemplo, o teste de Cox-Stuart, como complementação da análise gráfica, afim de concluir a existência ou inexistência de tendência temporal na série.

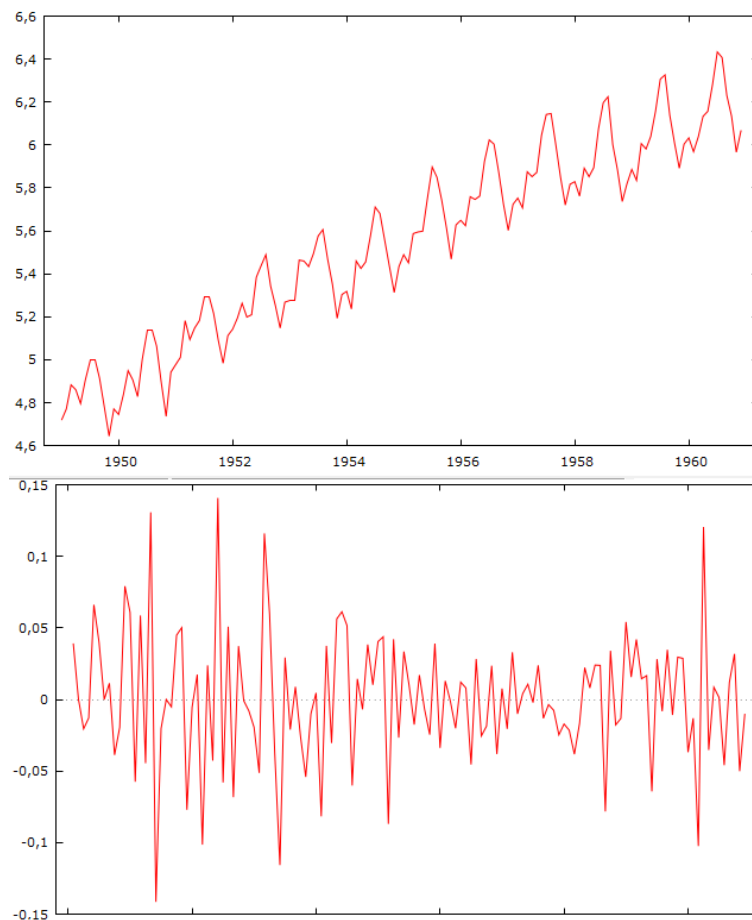
A presença de sazonalidade torna a série temporal periódica, podendo ser semanal, mensal, anual ou com qualquer outra periodicidade. Testes paramétricos e não paramétricos, como o teste F e Teste de Kruskal-Wallis podem ser empregados para verificar a existência de sazonalidade (MORETTIN ; TOLOI, 2006).

Na prática, é raro encontrar séries temporais estacionárias, sendo que a maioria apresenta tendência e/ou sazonalidade. Os modelos autorregressivos e de médias móveis, normalmente denominados, modelos de Box e Jenkins, são adequados apenas para séries estacionárias. Assim, para algumas séries temporais, é necessário estimar e eliminar tendência e sazonalidade antes do ajuste da série.

A eliminação da tendência e sazonalidade pode ser feita por meio da diferenciação, denotada por Δ . Para a eliminação da tendência aplica-se a diferenciação simples, denotada por Δ^d , sendo d a quantidade de diferenciações tomadas. Para eliminar a sazonalidade aplica-se a

ou diferenciação sazonal, denotada por Δ^D , em que D é a quantidade de diferenças sazonais tomadas da série. É importante destacar a importância de se evitar a tomada de excessivas diferenças, a fim de evitar a perda de uma grande quantidade de dados. Na Figura 1, no gráfico superior, é exibido uma série temporal não estacionária com tendência e sazonalidade. No gráfico inferior da mesma figura, a série torna-se estacionária após aplicar diferença simples e diferença sazonal.

Figura 1-Gráfico de série temporal antes e após a aplicação de diferença simples e sazonal.



Fonte: Do autor (2024)

Também denominados modelos de Box e Jenkins, os modelos Autorregressivos e de Médias Móveis (ARMA) e suas variações são apropriados exclusivamente ao ajuste de séries estacionárias. Portanto, as componentes de tendência e sazonalidade devem ser eliminadas da série original, caso elas existam (MORETTIN ; TOLOI, 2006).

O modelo ARMA(p,q) se origina da junção do modelo autorregressivo (AR), de ordem 'p', ao modelo de médias móveis (MA), de ordem 'q'. O modelo Autorregressivo Integrado de Médias Móveis, ARIMA(p,d,q), trata-se de uma variação do modelo ARMA, aplicável às séries não estacionárias, das quais são tomadas 'd' diferenciações simples para que a série se torne

estacionária, sem tendência. Por fim, ao aplicar-se uma ou mais diferenças sazonais, tem-se o modelo da classe Sazonal Autorregressivo Integrado de Médias Móveis, SARIMA(p,d,q)(P,D,Q), representado na Equação 1. Note que, se ‘D’ = 0 tem-se o modelo ARIMA (p,d,q). De modo análogo se ‘d’ = 0 temos o modelo ARMA (p,q).

$$\Phi(B^s)\phi(B)\Delta^d\Delta^D Z_t = \theta(B^s)\theta(B)\epsilon_t, \quad (1)$$

sendo $\Delta^D Z_t$ um processo estacionário; $\phi(B)$ o polinômio autorregressivo, dado por $\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$; $\theta(B)$ o polinômio de médias móveis, dado por $\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q$; Δ^d o operador de diferença simples; $\Phi(B^s)$ o polinômio sazonal autorregressivo, dado por $(1 - \Phi_1 B^s - \Phi_2 B^{2s} - \dots - \Phi_P B^{Ps})$; $\Theta(B^s)$ o polinômio sazonal de médias móveis, dado por $(1 - \theta_1 B^s - \theta_2 B^{2s} - \dots - \theta_P B^{Ps})$; s é o período sazonal; ϵ_t a componente aleatória não correlacionada do modelo, ou simplesmente um ruído branco, com média nula e variância constante (BROCKWELL; DAVIS; CALDER, 2002; MORETTIN; TOLOI, 2006).

Ainda da equação 1, um modelo AR (p) é obtido quando ‘D’ = ‘d’ = 0 e $\theta(B) = 1$. DE maneira semelhante, um modelo MA é obtido quando ‘D’ = ‘d’ = 0 e $\phi(B) = 1$.

É crucial, após o ajuste de modelos, seja da classe ARIMA ou SARIMA, realizar uma verificação da adequação do modelo obtido. Isso ocorre quando os resíduos do modelo comportam-se como um ruído branco, o que implica que esses resíduos são variáveis aleatórias independentes e identicamente distribuídas, com média zero e variância constante (MORETTIN; TOLOI, 2006). Outra abordagem para essa avaliação é a aplicação do teste de Box e Pierce (1970), também conhecido como teste de estatística Q.

4.3 Séries espaço-temporais: principais conceitos

Os modelos espaço-temporais têm a característica de se ajustarem a conjuntos de dados que apresentam não apenas autocorrelação temporal, mas também autocorrelação espacial. Em termos simples, a autocorrelação temporal ocorre quando os elementos do conjunto se distribuem ao longo do tempo de maneira interdependente. Analogamente, afirmar que um conjunto de dados apresenta autocorrelação espacial significa que há um nível de associação espacial entre seus elementos, exigindo que suas disposições e interações com os vizinhos sejam consideradas durante o processo de ajuste do modelo (JIN, 2017; PFEIFER; DEUTRCH, 1980a e 1980b).

Para ilustrar o conceito de autocorrelação espaço-temporal, considere a seguinte situação: uma variável de interesse é monitorada simultaneamente em M diferentes locais durante T instantes de tempo. Em cada instante $t \in T$, existem M observações dessa variável tomadas em diferentes locais. Esse conjunto de dados exibirá correlação espaço-temporal se, em um determinado instante de tempo, escolhermos aleatoriamente uma das M observações, e essa observação estiver relacionada simultaneamente com as observações em sua vizinhança e com as observações anteriores e/ou posteriores naquele local.

Outro conceito relevante na abordagem espaço-temporal é a estacionariedade. Ao ajustar modelos da classe STARMA, Pfeifer e Deutch (1980a) afirmam que é pressuposto não apenas a estacionariedade temporal, mas também a estacionariedade espacial. Isso significa que a dependência observada entre uma observação e suas observações vizinhas deve permanecer a mesma para todas as N regiões de interesse.

4.4 Matriz de vizinhança (W)

Também conhecida como matriz de ponderação espacial, matriz de proximidade espacial ou matriz de pesos espaciais, trata-se de uma matriz que busca estimar e descrever a variabilidade espacial dos dados (MONTEIRO *et al.*, 2004). A matriz de vizinhança é denotada por W e, de acordo com Almeida (2012), trata-se de uma matriz quadrada com dimensão $N \times N$, em que N é o número de regiões ou localizações avaliadas.

De maneira resumida, pode-se dizer que a matriz de vizinhança tem por objetivo explicitar o grau de proximidade entre duas sub-áreas contidas na área de interesse, dado um critério apropriado. Almeida (2012) afirma que o objetivo da matriz W é:

Refletir um determinado arranjo espacial das interações resultantes do fenômeno a ser estudado. Regiões mais conectadas entre si interagem mais do que regiões menos conectadas. Cada conexão entre duas regiões é representada por uma célula desta matriz, sendo denominada de peso espacial (ALMEIDA, 2012, p.75).

Dessa forma, dada uma região de interesse constituída por N diferentes áreas, A_1, A_2, \dots, A_N , a matriz de ponderação espacial de l -ésima ordem será representada de maneira semelhante à definida na Equação 2, no qual cada peso espacial $w_{i,j}$ expressa numericamente o grau de proximidade entre a área A_i e a área A_j para todo $i \neq j$.

$$W^l = \begin{bmatrix} w_{1,1} & w_{1,2} & \dots & w_{1,N} \\ w_{2,1} & w_{2,2} & \dots & w_{2,N} \\ \vdots & \vdots & \ddots & \vdots \\ w_{N,1} & w_{N,2} & \dots & w_{N,N} \end{bmatrix}, \quad (2)$$

em que l é a ordem de vizinhança e $w_{i,j}$ (com $i = 1, 2, 3, \dots, N$ e $j = 1, 2, 3, \dots, N$) são os pesos espaciais.

A ordem da matriz de vizinhança indica uma hierarquia de proximidade entre uma área e seus vizinhos, de forma que os vizinhos de segunda ordem, de determinada área, estão mais distantes que os vizinhos de primeira ordem e mais próximos que os vizinhos de terceira ordem da referida área. Apesar do presente trabalho ter interesse exclusivo de matrizes \mathbf{W} de primeira ordem, é importante salientar que dada uma área de interesse, é possível construir matrizes de vizinhança de diferentes ordens e que a ordem de vizinhança é denotada por ' l '.

É necessário que, após a sua construção, a matriz de proximidade espacial seja normalizada. Tal normalização é realizada em todas as linhas da matriz, de modo que o somatório dos componentes de cada linha seja igual a um, isto é, $\sum_j^N w_{ij} = 1$.

A construção de uma matriz de vizinhança não é trivial e nem única. A definição dos pesos $w_{ij}^{(l)}$ é uma decisão subjetiva do pesquisador, sendo que este deve escolher pesos que reflitam a distribuição espacial do sistema em estudo (PFEIFER ; DEUTRCH, 1980a; MARTINS; BIZ ; OZAKI, 2014; HOEF *et al.*, 2018). Rao e Antunes (2004), Smith (2013) e Biz (2014) são citados em muitos trabalhos, como importantes estudos dedicados à construção de \mathbf{W} e afirmam que a ponderação dos pesos espaciais da matriz \mathbf{W} dependerá da área de interesse do estudo e do tipo de pesquisa.

Visto que a matriz de vizinhança pode ter inúmeras construções, um dos intuítos deste trabalho é averiguar a pertinência da utilização de matrizes de vizinhança baseadas em critérios não geográficos, no ajuste de modelos espaço-temporais da classe STARMA, aplicados a dados epidemiológicos. Ademais, pretende-se verificar a pertinência do método proposto, dar destaque à necessidade de investigar quando \mathbf{W} não espaciais são mais eficientes que as convencionais, além de também evidenciar a necessidade de estudos que procurem descrever qual seria a metodologia ótima de construção das matrizes de vizinhança não espaciais.

A seguir serão descritas diferentes maneiras de construir a matriz de ponderação espacial \mathbf{W} . Primeiramente, serão apresentadas as principais matrizes de vizinhança criadas segundo critérios geográficos, ou seja, considerando-se distâncias ou fronteiras. Em seguida, será apresentado o conceito de matrizes de proximidade espacial considerando-se parâmetros não geográficos, além de apresentar algumas exemplificações geograficamente distantes (ALMEIDA, 2012; JIN, 2017).

4.4.1 Matriz de ponderação espacial baseada em critérios geográficos

As matrizes de vizinhança baseadas em critérios geográficos podem ser construídas considerando-se distâncias, dada uma métrica ou considerando-se as fronteiras das áreas em estudo. As matrizes de vizinhança baseadas em fronteiras podem ser, antes de sua normalização, de natureza binária ou não. Duas formas de atribuir pesos espaciais, tendo em vista as fronteiras das áreas de interesse, são descritas a seguir. A primeira consiste em atribuir pesos espaciais apenas verificando se duas subáreas compartilham ou não fronteiras em comum. Em uma segunda maneira, os pesos espaciais são atribuídos considerando-se as extensões das fronteiras compartilhada entre as subáreas contidas na área de interesse.

Dentre as matrizes de vizinhança obtidas considerando-se fronteiras destaca-se a Matriz de Contiguidade. Essa é uma das matrizes de vizinhança mais utilizada e se destaca por sua simplicidade. Para uma área de interesse, dividida em um total de N subáreas denotadas por A_1, A_2, \dots, A_N , a Matriz de Contiguidade será uma matriz binária de dimensão $N \times N$. Na Equação 3 é definido o critério de ponderação dos pesos espaciais para a Matriz de Contiguidade, para duas subáreas A_i e A_j com $i = 1, 2, 3, \dots, N, j = 1, 2, 3, \dots, N$ e $i \neq j$:

$$w_{i,j} = \begin{cases} 1, & \text{se as áreas } A_i \text{ e } A_j \text{ são contíguas} \\ 0, & \text{caso contrário.} \end{cases} \quad (3)$$

Também é possível a construção da \mathbf{W} baseada na extensão das fronteiras compartilhadas pelas subáreas de interesse. Por não ser de interesse desse trabalho, tal metodologia não será discutida aqui, mas caso o assunto seja de interesse do leitor, recomenda-se a leitura de Almeida (2012). Contudo, cita-se de antemão, como principal desvantagem de tal abordagem, a dificuldade de mensurar o comprimento das fronteiras e dos perímetros de cada área (SMITH, 2013). Outra desvantagem a ser considerada é que, em casos com áreas de tamanhos muito heterogêneos, será obtida uma matriz \mathbf{W} desbalanceada, no sentido de que as áreas com fronteiras maiores tenderão a apresentar um número maior de vizinhos (ALMEIDA, 2012; SMITH 2013).

A obtenção de matrizes de vizinhança com base em distâncias fundamenta-se na premissa de que duas áreas geograficamente próximas apresentarão uma interação espacial mais significativa que áreas geograficamente distantes (ALMEIDA, 2012). Existem várias abordagens para a construção da matriz \mathbf{W} , levando em consideração a distância geográfica.

Algumas delas incluem a utilização de exponencial da distância, da distância inversa e a de k-vizinhos (VIEIRA, 2009; ALMEIDA, 2012).

Possivelmente, a matriz de vizinhança mais comum entre aquelas construídas com base em distâncias, é a matriz exponencial da distância (ALMEIDA, 2012). No entanto, conforme observado pelo autor, a determinação dos pesos espaciais para essa matriz depende de um parâmetro de amortecimento, muitas vezes definido de maneira arbitrária e independente das demais análises.

Apesar de não serem relevantes neste trabalho, caso seja de interesse do leitor, recomenda-se a leitura de Almeida (2012). Contudo, também de antemão, pontua-se que uma desvantagem ao atribuir pesos espaciais com base em distâncias é a escolha da métrica a ser empregada. Embora a distância Euclidiana seja a opção mais comum, a utilização da Distância do Grande Círculo é mais apropriada para situações em que as regiões de interesse estão geograficamente distantes (ALMEIDA, 2012; JIN, 2017).

4.4.2 Matriz de ponderação espacial baseada em critérios não espaciais

De especial interesse neste trabalho, as matrizes de vizinhança baseadas em critérios não geográficos é um assunto recente, com pouca discussão e material bibliográfico. Tal tipo de matriz de vizinhança tem aplicabilidade não somente para os modelos STARMA e suas variações, mas para uma grande gama de modelos.

Apesar das ressalvas em relação à utilização desse tipo de matriz de vizinhança, é necessário ter em mente que por vezes pode ser mais pertinente medir o grau de similaridade entre duas áreas de interesse por meio de variáveis não geográficas. Pode-se, por exemplo, inferir o grau de proximidade de duas cidades por meio do grau de escolaridade de suas populações, de índices de desenvolvimentos financeiros ou ainda por meio da qualidade dos serviços básicos prestados à população tais como: educação, saúde, segurança, lazer e etc.

A utilização de critérios não geográficos para a obtenção de \mathbf{W} , fundamenta-se no pensamento de que áreas (ou parcelas, para o caso de experimentos) com características iguais têm maior proximidade, ainda que geograficamente distantes (ALMEIDA, 2012). Dessa maneira a disposição e as iterações espaciais podem ser melhores representadas por meio de critérios não geográficos do que a partir de critérios geográficos.

Segundo Vieira (2009), o conceito de matriz de vizinhança não geográfica foi introduzido por Case, Rosen e Hines (1993). No entanto, ainda não existe na literatura muitas instruções para a construção de matrizes de ponderação espacial não geográficas, menos ainda

equações que norteiem a ponderação dos pesos espaciais para esse tipo de matriz. De acordo com Almeida (2012), uma das maneiras de se construir uma matriz de vizinhança não geográfica é baseando-se no grau de similaridade entre duas ‘áreas’, considerando critérios adequados.

Em uma matriz de vizinhança não geográfica de similaridade, os pesos espaciais ($w_{i,j}$) são atribuídos a partir do grau de semelhança existente entre duas áreas, dada uma ou mais variáveis de interesse. Neste trabalho, em específico e de modo inédito, os pesos serão atribuídos considerando-se uma combinação linear de duas variáveis socioeconômicas.

No entanto, a ponderação dos pesos espaciais pode ser feita de muitas maneiras distintas, como ilustra o seguinte exemplo. Imagine um estudo espacial ou espaço-temporal no qual o interesse está na região sudeste do Brasil e o objetivo é estudar o nível de alfabetização. Considerando-se o Índice de Desenvolvimento Humano (IDH) dos quatro Estados da região Sudeste, a ponderação dos pesos espaciais para tal matriz socioeconômica de similaridade, pode ser feita, dentre muitas outras maneiras, como nas exemplificações que seguem:

$$a) \ w_{i,j} = \begin{cases} 1, & \text{se } \frac{IDH_i + IDH_j}{2} \geq IDH_m \\ 0, & \text{caso contrário} \end{cases} \quad (4)$$

$$b) \ w_{i,j} = e^{-|IDH_i - IDH_j|} \quad (5)$$

$$c) \ w_{i,j} = \frac{1}{|IDH_i - IDH_j|}, \quad (6)$$

sendo respectivamente, IDH_i e IDH_j os Índice de Desenvolvimento Humano (IDH) dos Estados i e j , respectivamente e IDH_m é o IDH médio da região Sudeste. Observe que que nessa exemplificação foi utilizado o IDH como variável socioeconômica, mas poderiam ter sido utilizadas outras variáveis, como provas estaduais ou federais aplicadas anualmente ou qualquer outro indicador disponibilizado pelo Ministério da Educação.

Existem ainda matrizes de vizinhança baseadas em critérios não geográficos que utilizam conceitos de dissimilaridade e de fluxo. Maiores esclarecimentos podem ser encontrados em Hernández-Murillo *et al.* (2003) e Almeida (2012).

Neste trabalho as variáveis socioeconômicas consideradas para obtenção das matrizes não espaciais serão: o IDH municipal mais atual e a média do Previde Brasil dos anos de 2021 a 2023. A opção pelo IDH municipal se deve a esse dado relacionar cada município com suas

respectivas características sociodemográficas, buscando com isso refletir no modelo o contraste social, demográfico e econômico dos municípios em questão.

Quanto ao Previne Brasil, a segunda variável não espacial escolhida para construção da matriz **W**, trata-se de um programa de avaliação quadrimestral da qualidade do serviço prestado pela Atenção Primária a Saúde (parte do SUS que se destina essencialmente a prevenção e aos atendimentos eletivos). Como já mencionada, é majoritariamente na APS em que é feito a prevenção, o diagnóstico e o tratamento da TB. Portanto, espera-se que municípios bem avaliados no Previne Brasil, tenham menos casos notificados de TB. Com a escolha dessa variável, buscou-se refletir no ajuste do modelo a efetividade da atuação da saúde pública dos municípios em questão, com ações que contemplem desde a prevenção até o tratamento da TB.

Na seção de metodologia desse trabalho, será abordada de forma mais detalhada e clara a ponderação dos pesos espaciais que salienta, de modo sucinto e preliminar, que de maneira inédita, serão obtidos por meio de cinco combinações lineares do IDH municipal com o Previne Brasil.

4.5 Modelos espaço-temporal STARMA e suas variações

Dentre os modelos espaço-temporais, encontra-se uma categoria conhecida como modelos autorregressivos de médias móveis espaço-temporais (STARMA) que, segundo Jin (2017), podem ser compreendidos como uma série temporal da classe de modelos autorregressivos e de médias móveis vetoriais (VARMA).

Os estudos relacionados aos modelos STARMA tiveram seu início em 1975 com as contribuições de Cliff e Ord (1975) e Martin e Oeppen (1975). No entanto, foi somente após os trabalhos de Pfeifer e Deutsch (1980a, 1980b, 1980c) que essa classe de modelos ganhou destaque. Devido às limitações computacionais da década de 80, a implementação completa desse modelo era extremamente desafiadora, levando com que tais modelos caíssem em esquecimento por algum tempo. Com os avanços tecnológicos, especialmente no campo computacional, cerca de duas décadas depois, os modelos STARMA voltaram a ser objeto de interesse na comunidade científica. Desde então, vêm sendo cada vez mais estudados e empregados por pesquisadores de todo o mundo.

Essa classe de modelos é empregada para descrever dados de séries temporais associadas a diferentes localizações espaciais, caracterizando-se pela dependência linear ao longo do tempo e do espaço (PFEIFER; DEUTSCH, 1980a, 1980b; MARTINS; BIZ; OZAKI, 2014).

Vale destacar que os modelos STARMA, assim como os modelos temporais da classe ARMA, são aplicáveis apenas a processos estacionários. Isso significa que a estrutura de covariância de $Z(t)$ deve permanecer inalterada ao longo do tempo e do espaço (PFEIFER E DEUTSCH, 1980a; MARTINS; BIZ; OZAKI, 2014). Também de maneira semelhante aos modelos de Box e Jenkins, os modelos STARMA são derivados por meio de um procedimento iterativo de três estágios, os quais compreendem a identificação do modelo, a estimação dos parâmetros e o diagnóstico do ajuste (CHEYSSON, 2016). Informações detalhadas sobre os três estágios do processo iterativo propostos por Box e Jenkins, estendidos para modelos espaço-temporais STARMA, podem ser encontradas em Box, Jenkins e Reinsel (2008) e também em Pfeifer e Deutsch (1980a).

O modelo STARMA é apresentado na Equação 7. Note que trata-se de sua apresentação vetorial, devido ao interesse deste estudo na aplicação de matrizes de vizinhança não tradicionais, sendo então mais interessante e didático que na equação seja explicitada a matriz \mathbf{W} . No entanto, ressalta-se que a equação do modelo pode ser reescrita de maneira diferente, utilizando-se do operador de defasagem espacial como apresentado em Jin (2017) e Martins; Biz; Ozaki (2014).

$$\mathbf{Z}(t) = \sum_{k=1}^p \sum_{l=0}^{\lambda_k} \phi_{kl} \mathbf{W}^l \mathbf{Z}(t-k) - \sum_{k=1}^q \sum_{l=0}^{m_k} \theta_{kl} \mathbf{W}^l \boldsymbol{\epsilon}(t-k) + \boldsymbol{\epsilon}(t), \quad (7)$$

em que $\mathbf{Z}(t)$ denota a série espaço-temporal no instante t ; p é a ordem autorregressiva e q é a ordem de médias móveis do modelo; λ_k é a ordem máxima de vizinhança do k -ésimo termo autorregressivo e m_k é a ordem máxima espacial do k -ésimo termo de médias móveis, ϕ_{kl} e θ_{kl} são os parâmetros a serem estimados do modelo, \mathbf{W}^l é a matriz de vizinhança de l -ésima ordem (sendo nesse trabalho, l exclusivamente igual a 1) e $\boldsymbol{\epsilon}(t)$ é o vetor com os $\epsilon_i(t)$ erros normais e aleatórios associados ao modelo, com $E[\epsilon_i(t)] = 0$ para todo i e $E[\epsilon_i(t)\epsilon_j(t+s)'] = \begin{cases} \sigma^2, i = j \text{ e } s = 0 \\ 0, \text{ caso contrário} \end{cases}$.

De modo análogo às séries temporais, os modelos STARMA apresentam variações, sendo elas: STAR, STMA e STARIMA. As duas primeiras variantes referem-se a simplificações do modelo STARMA, no qual são inexistentes a parte de médias móveis (STAR- Equação 8) e a autorregressiva (STMA- Equação 9), respectivamente.

$$\mathbf{Z}(t) = \sum_{k=1}^p \sum_{l=0}^{\lambda_k} \phi_{kl} \mathbf{W}^l \mathbf{Z}(t-k) + \boldsymbol{\epsilon}(t) \quad (8)$$

$$\mathbf{Z}(t) = \boldsymbol{\epsilon}(t) - \sum_{k=1}^q \sum_{l=0}^{m_k} \theta_{kl} \mathbf{W}^l \boldsymbol{\epsilon}(t-k) \quad (9)$$

Quanto ao modelo $\text{STARIMA}(p_\lambda, d, q_m)(P_\lambda, D, Q_M)$, trata-se da variação dos modelos STARMA, para o caso de séries espaço-temporais não estacionárias, nas quais são necessárias a aplicação de diferenciação simples e/ou sazonal, conforme apresentado na Equações 10 e 11. Note que o modelo apresentado na Equação 10, trata-se de um STARIMA do qual foram tomados d diferenças simples e na Equação 11 está representado um modelo STARIMA do qual foram tomadas d diferenciações simples e D diferenças sazonais.

$$\nabla^d \mathbf{Y}(t) = \sum_{k=1}^p \sum_{l=0}^{\lambda_k} \phi_{kl} \mathbf{W}^l \nabla^d \mathbf{Y}(t-k) - \sum_{k=1}^q \sum_{l=0}^{m_k} \theta_{kl} \mathbf{W}^l \boldsymbol{\epsilon}(t-k) + \boldsymbol{\epsilon}(t), \quad (10)$$

$$(I - \sum_{k=1}^P \sum_{l=0}^{\Lambda_k} \Phi_{lk} \mathbf{W}^l B^{kS}) \phi_{p,\lambda}(B) \nabla_S^D \nabla^d \mathbf{Y}(t) = (I - \sum_{k=1}^Q \sum_{l=0}^{M_k} \Theta_{kl} \mathbf{W}^l B^{kS}) \theta_{q,m}(B) + \boldsymbol{\epsilon}(t) \quad (11)$$

em que I é a matriz identidade, s é o período sazonal e $\mathbf{Y}(t)$ representa a série espaço temporal não estacionária a ser ajustada.

5 MATERIAIS E MÉTODOS

Esta seção iniciará expondo o conjunto de dados utilizado e suas análises iniciais. Em sequência serão apresentados os métodos utilizados no decorrer do trabalho. Primeiramente será explicitado o método empregado na obtenção das matrizes de vizinhança construídas. Por último, serão esclarecidos os métodos usados na identificação, estimativa e no diagnóstico dos modelos STARMA.

5.1 Dados

Os dados utilizados neste trabalho foram obtidos do site do Departamento de Informática do Sistema Único de Saúde (DATASUS) e quantificam casos confirmados de tuberculose em nove cidades mineiras, conforme ilustrado na Figura 2. São elas: Belo Horizonte, Betim, Contagem, Ibirité, Nova Lima, Ribeirão das Neves, Sabará, Santa Luzia e Vespasiano. Esse conjunto de dados trata-se de observações mensais, coletadas entre janeiro de 2002 e dezembro de 2022 e está disponível em: <http://tabnet.datasus.gov.br/cgi/tabcgi.exe?sinannet/cnv/tubercbr.def>.

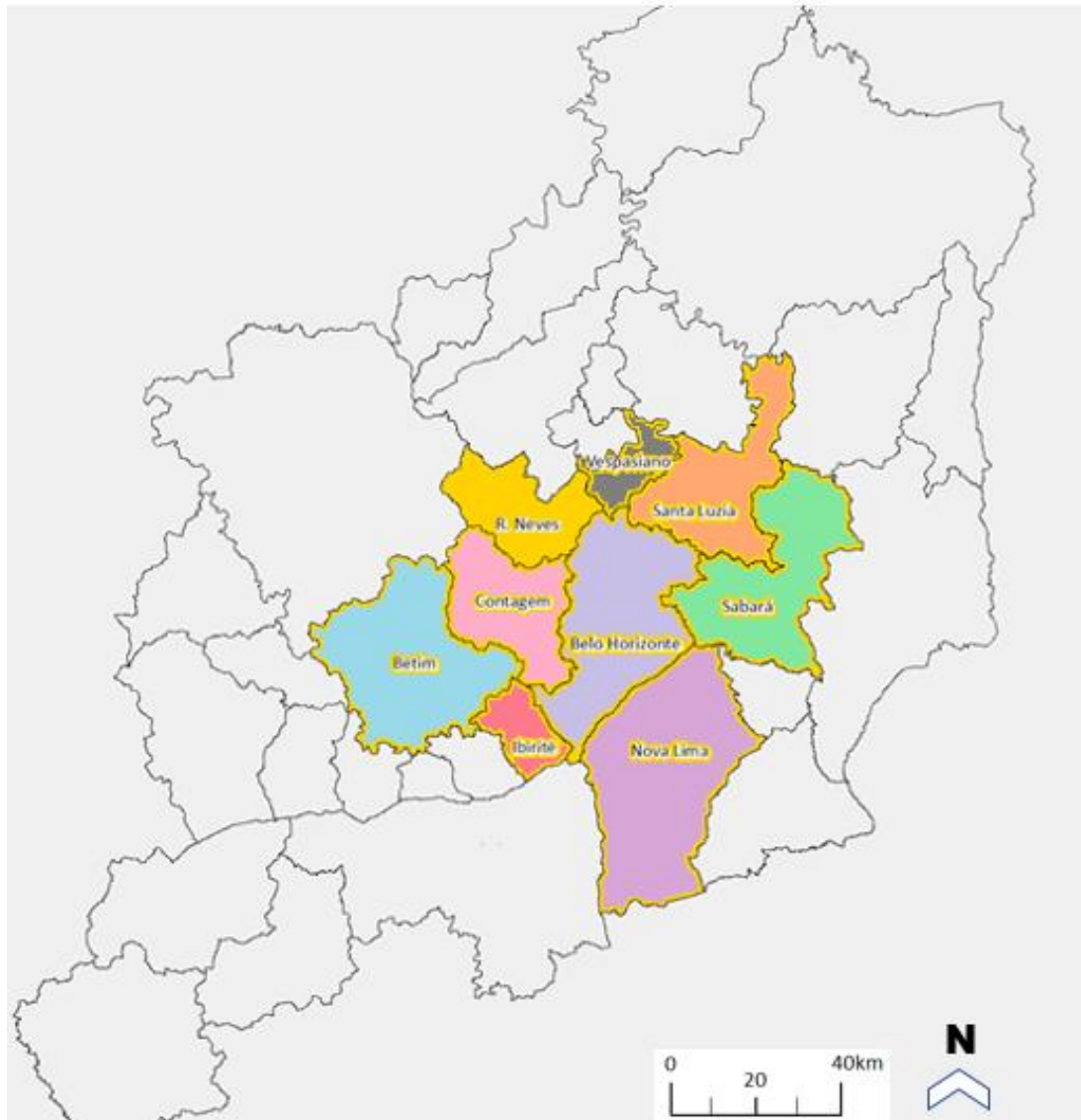
Em seguida, foram calculadas as taxas de incidência mensal de tuberculose para cada município. Para isso foram usadas as estimativas populacionais do IBGE para anos intercensitários, informadas ao Tribunal de Contas da União (TCU) para determinação das cotas do Fundo de Participação dos Municípios (FPM), disponível em <http://tabnet.datasus.gob.br>. As observações mensais de TB foram divididas por suas respectivas populações municipais de cada ano, obtendo-se assim a taxa de incidência mensal de TB para cada um dos nove municípios em estudo.

Por fim, as nove séries temporais foram agrupadas em uma única planilha eletrônica, de modo a formar a série espaço-temporal, formado por 9 colunas e 252 linhas. A fim de aferir a qualidade da previsão realizada pelos modelos que serão ajustados no trabalho, optou-se por reservar as 6 últimas observações (janeiro de 2002 a junho de 2022) de cada uma das nove cidades, de tal maneira que o conjunto de dados passasse a ser formado por 9 colunas e 246 linhas.

Quanto ao IDH municipal, utilizado na construção das matrizes W não espaciais, foi obtido do site do Instituto Brasileiro de Geografia e Estatística (IBGE), disponível em <https://www.ibge.gov.br/>. Também utilizado na construção das matrizes W não espaciais, o resultado do Previne Brasil de cada um dos 9 municípios, nos anos de 2021 a 2023, foi obtido

no site do do Sistema de Informação em Saúde para a Atenção Básica (SISAB), do Ministério da Saúde, disponível no link <https://sisab.saude.gov.br/>.

Figura 2- Mapa da grande Belo Horizonte, MG, enfatizando os nove municípios de interesse nesse trabalho



Fonte: Do autor (2024)

5.2 Métodos

A seguir serão apresentados os métodos utilizados nas análises iniciais do conjunto de dados, bem como na obtenção dos índices municipais utilizados para criação das matrizes W não espaciais. Por fim serão descritos os métodos para o ajuste dos modelos STARMA.

5.2.1 Análises iniciais

Após extração dos dados foi verificado se o conjunto de dados era estacionário. Inicialmente, pela análise gráfica, foi possível perceber a existência de tendência em quase todas as séries temporais, sendo menos evidente apenas para Nova Lima, Sabará e Vespasiano. A existência de tendência foi constatada com a aplicação do teste de Cox-Stuart, popularmente conhecido como Teste do Sinal, que sinalizou a existência de tendência em todas as séries temporais. Buscando eliminar a tendência do conjunto de dados foi tomada uma diferença simples da série espaço-temporal.

A existência de sazonalidade foi verificada por meio do teste F para as variáveis *dummies* sazonais, que nada mais são do que variáveis temporais, que representam cada um dos doze meses.

5.2.2 Definição dos índices municipais socioeconômicos

Como já antecipado, a criação das matrizes **W** não espaciais, foi feita considerando uma combinação linear de duas variáveis socioeconômicas: o IDH municipal (IDH_M) e a média anual de 2021 a 2023 do Previde Brasil (PB_M).

Serão consideradas cinco combinações lineares, denominadas de IMS_1 , IMS_2 , IMS_3 , IMS_4 , IMS_5 , as quais estão definidas nas Equações 12, 13, 14, 15 e 16 e serão, a partir de agora, denominadas nesse trabalho como índice municipal socioeconômico (IMS).

$$a) \quad IMS_1 = 0,00.PB + 1,00.IDH_M; \quad (12)$$

$$b) \quad IMS_2 = 0,25.PB + 0,75.IDH_M; \quad (13)$$

$$c) \quad IMS_3 = 0,50.PB + 0,50.IDH_M; \quad (14)$$

$$d) \quad IMS_4 = 0,75.PB + 0,25.IDH_M; \quad (15)$$

$$e) \quad IMS_5 = 1,00.PB + 0,00.IDH_M. \quad (16)$$

Sendo ‘M’ cada um dos nove municípios de interesse desse trabalho. Observe que os índices foram definidos com diferentes pesos associados às duas variáveis socioeconômicas consideradas.

5.2.3 Matrizes de vizinhança

Das seis matrizes de vizinhança construídas neste trabalho, apenas uma será baseado em critérios espaciais, sendo o critério de fronteiras o utilizado. Tal matriz, denominada a partir de agora por \mathbf{W}_G , representa a matriz de contiguidade da área de interesse do trabalho e foi obtida através da Equação 3, novamente repetida abaixo.

$$w_{i,j} = \{1, \text{ se as áreas } A_i \text{ e } A_j \text{ são contíguas } 0, \quad \text{ caso contrário.}$$

Para a obtenção dos pesos para as cinco matrizes não espaciais (\mathbf{W}_{IMS1} , \mathbf{W}_{IMS2} , \mathbf{W}_{IMS3} , \mathbf{W}_{IMS4} , \mathbf{W}_{IMS5}), foi usada a Equação 17, dada por:

$$w_{K,i,j} = \frac{1}{|IMS_{K,i} - IMS_{K,j}|}, \quad (17)$$

sendo $w_{K,i,j}$ o peso atribuído considerando-se os municípios ‘i’ e ‘j’ (com $i \neq j$) e seus respectivos K -ésimo IMS com $K = 1, 2, 3, 4$ e 5 .

5.2.4 Identificação, estimação e diagnóstico dos modelos STARMA

Para cada uma das seis matrizes de vizinhança construídas, foi ajustado um modelo da classe STARMA. O ajuste dos modelos foi realizado utilizando o software R (R Core Team 2023), mais especificamente o pacote denominado *starma* (CHEYSSON, 2016). A obtenção de modelos STARMA é subdividida em três fases: identificação, estimação e diagnóstico do modelo. A seguir são descritas separadamente cada uma das fases de obtenção dos modelos e no Apêndice A segue o código computacional implementado em R (R CORE TEAM, 2023) utilizado.

a) Identificação:

Por meio das funções *stacf*() e *stpacf*(), foram gerados, respectivamente, o gráfico da função de autocorrelação espaço-temporal (*stacf*) e o gráfico da função de autocorrelação parcial espaço-temporal (*stpacf*). Vale destacar que, assim como a função *stcor.teste*(), as funções *stacf*() e *stpacf*() recebem a matriz \mathbf{W} , dentre outros argumentos. Dessa forma, para cada \mathbf{W} foram gerados seus respectivos gráficos de *stacf* e *stpacf*.

Esses gráficos desempenham um papel crucial na identificação da ordem do modelo e, de acordo com Jin (2017), a identificação da ordem dos modelos STARMA é feita de maneira análoga aos modelos de séries temporais. De acordo com a autora, um modelo espaço-temporal da classe $\text{STAR}(p_{\lambda_0}, p_{\lambda_1}, \dots, p_{\lambda_p})$ possui um gráfico da *stacf* que decai gradativamente e um

gráfico da *stpacf* que é igual a zero a partir do p -ésimo lag no tempo e o λ_p -ésimo lag no espaço. De maneira análoga um modelo STMA $(q_{m_0}, q_{m_1}, \dots, q_{m_q})$ possui um gráfico da *stpacf* que decai gradativamente e um gráfico da *stacf* que é igual a zero no q -ésimo lag temporal e no m_q -ésimo termo espacial. Para o caso de modelos STARMA observa-se o decaimento gradativo tanto no gráfico da *stacf* quanto no gráfico da *stpacf*.

b) Estimação:

A metodologia de estimação proposto por Pfeifer e Deutsch (1980a) baseia-se no método da máxima verossimilhança e é demonstrada matematicamente por Jin (2017). Esse método, que pressupõe a normalidade dos resíduos, era pouco utilizado na década de 80, por ser computacionalmente oneroso.

A estimação dos parâmetros realizada pelo pacote *starma* é realizada por meio do filtro de Kalman (CHEYSSON, 2016). Esse filtro, utilizando equações matriciais e calcula recursivamente a estimativa dos parâmetros, sendo tais estimativas consideradas ótimas pelo princípio dos mínimos quadrados (JIN, 2017).

Além de proporcionar uma considerável otimização computacional, a estimação dos parâmetros por meio do filtro de Kalman é altamente eficiente ao lidar com grandes conjuntos de dados (CHEYSSON, 2016). Importante notar que, apesar de ser o método mais comum, não é necessário assumir normalidade residual ao utilizar o filtro de Kalman para obter estimativas dos parâmetros de modelos STARMA (JIN, 2017).

No pacote *starma*, a estimação dos parâmetros é realizada utilizando a função *starma*. Vale ressaltar que, antes de estimar os parâmetros, é aconselhável centralizar o conjunto de dados e isso pode ser feito por meio da função *stcenter()*. Tal função faz com que a série espaço-temporal passe a ter média nula e erro padrão unitário (CHEYSSON, 2016).

c) Diagnóstico do ajuste:

Assim como no ajuste de modelos para séries temporais, após a estimação dos parâmetros, é essencial avaliar a adequação do modelo. Também de maneira análoga às séries temporais, os modelos STARMA são considerados satisfatoriamente ajustados quando seus resíduos exibem um comportamento semelhante ao de um ruído branco, isto é, o gráfico da função de autocorrelação residual apresentará o máximo 5% de seu valor fora do intervalo de confiança, indicando assim que os resíduos são independentes e identicamente distribuídos.

No contexto do pacote *starma*, a verificação da adequação do modelo é realizada utilizando-se a função *stcot.test()*. Essa função constitui uma extensão espaço-temporal do teste

de Box-Pierce, no qual a hipótese nula de não correlação residual é avaliada. O teste é feito considerando a estatística apresentada na Equação 18 (CHEYSSON, 2016).

$$N \sum (T - s)[\hat{\rho}_l(s)]^2 \sim \chi^2(slag; tlag), \quad (18)$$

em que N é o número de locais considerados, T número de observações obtidas ao longo do tempo, em cada um dos N locais de interesse, s é a defasagem temporal considerada, $\hat{\rho}_l(s)$ a autocorrelação espaço-temporal para l -ésima ordem de vizinhança e s -ésima defasagem temporal, $slag$ é a maior ordem de vizinhança considerada e $tlag$ é o número de defasagem temporal utilizado na construção do gráfico da função de autocorrelação espaço-temporal, que pode ser obtido pela função *stacf()*.

Após a obtenção dos seis modelos adequadamente ajustados, um critério que possibilita a comparação destes é o Critério de Informação Bayesiano (BIC). (JIN, 2017; PFEIFER, DEUTRCH (1980a)). Na Equação 19, é representado o BIC

$$BIC = N.T.ln(\sigma^2) + 2.b.ln(T), \quad (19)$$

sendo σ^2 a estimativa da variância residual; ‘b’ a quantidade de parâmetros do modelo, N o número de regiões de interesse e ‘T’ o tamanho da série.

Uma vez que muitas vezes o ajuste de modelos se presta a realizar previsões, também torna-se pertinente comparar os modelos segundo suas previsões. Para isso será usado como critério o erro absoluto percentual médio (MAPE), apresentado na Equação 20.

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - p_i}{y_i} \right| \quad (20)$$

Em que ‘n’ é o número total de previsões a serem feitas; y_i valor real da da i -ésima observação e p_i o valor da i -ésima previsão. Cabe ressaltar que o MAPE de cada um dos modelos foi obtido por meio da função *mape* da biblioteca *metrics* do R (2023).

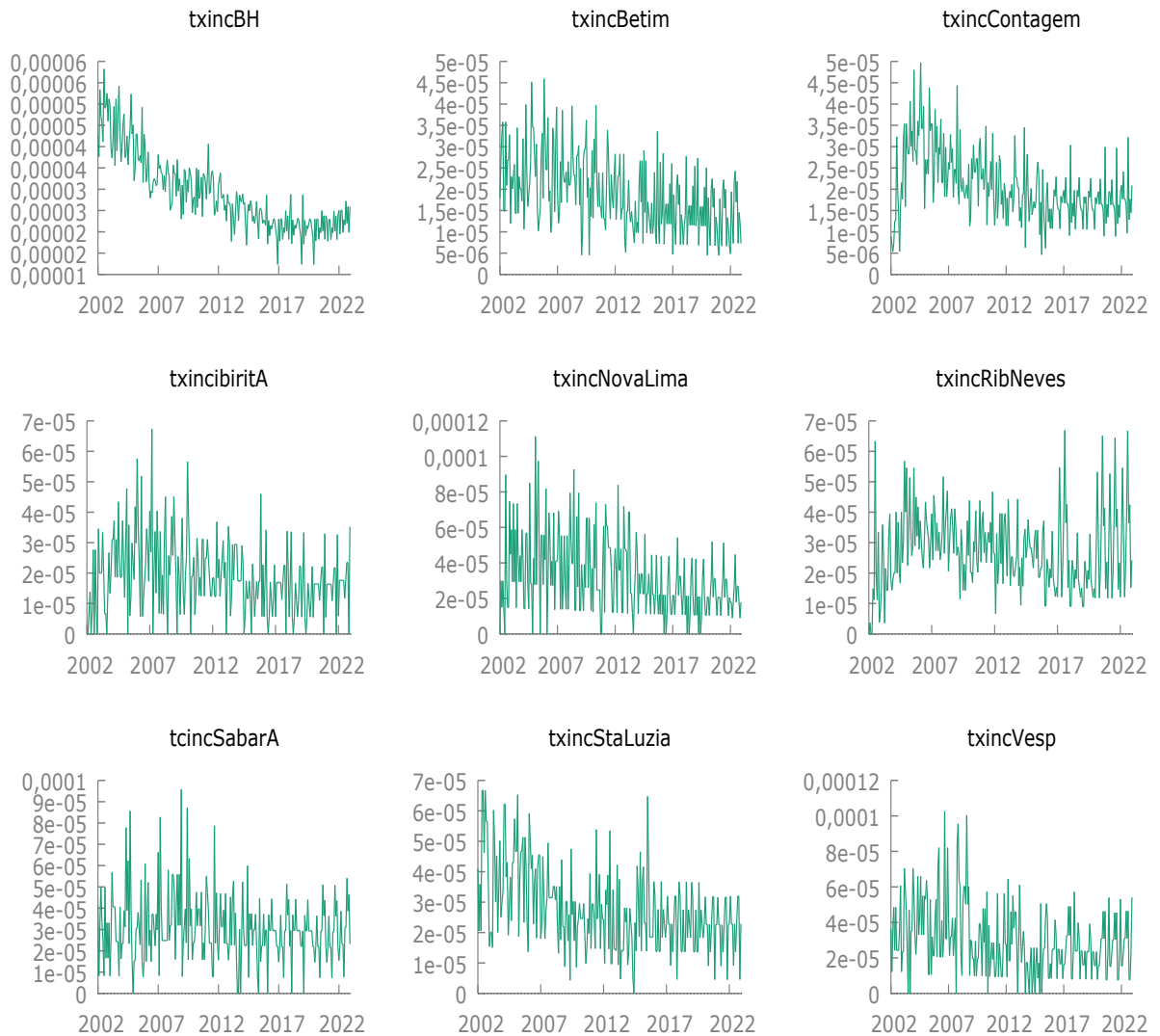
6 RESULTADOS E DISCUSSÕES

Nessa parte do trabalho serão apresentados e discutidos os resultados alcançados com a metodologia proposta. Iniciará com a apresentação dos valores do IMS obtidos bem como sua disposição no mapa da área de interesse. Em seguida serão apresentadas as seis matrizes de vizinhança construídas. Por fim serão tratados a identificação, estimação e diagnóstico dos modelos STARMA.

6.1 Resultados das análises iniciais

Na Figura 3, é exibida a representação gráfica das nove séries temporais. Estes gráficos ilustram a taxa de incidência de TB, observada entre janeiro de 2002 e junho de 2022, em cada um dos nove municípios de interesse. Após aplicação teste de Cox-Stuart, verificou-se a existência de tendência em todas as séries temporais. Foi tomada uma diferença simples da série espaço temporal e em seguida foram refeitos os gráficos das séries temporais de cada município. O resultado após diferenciação é exibido na Figura 4.

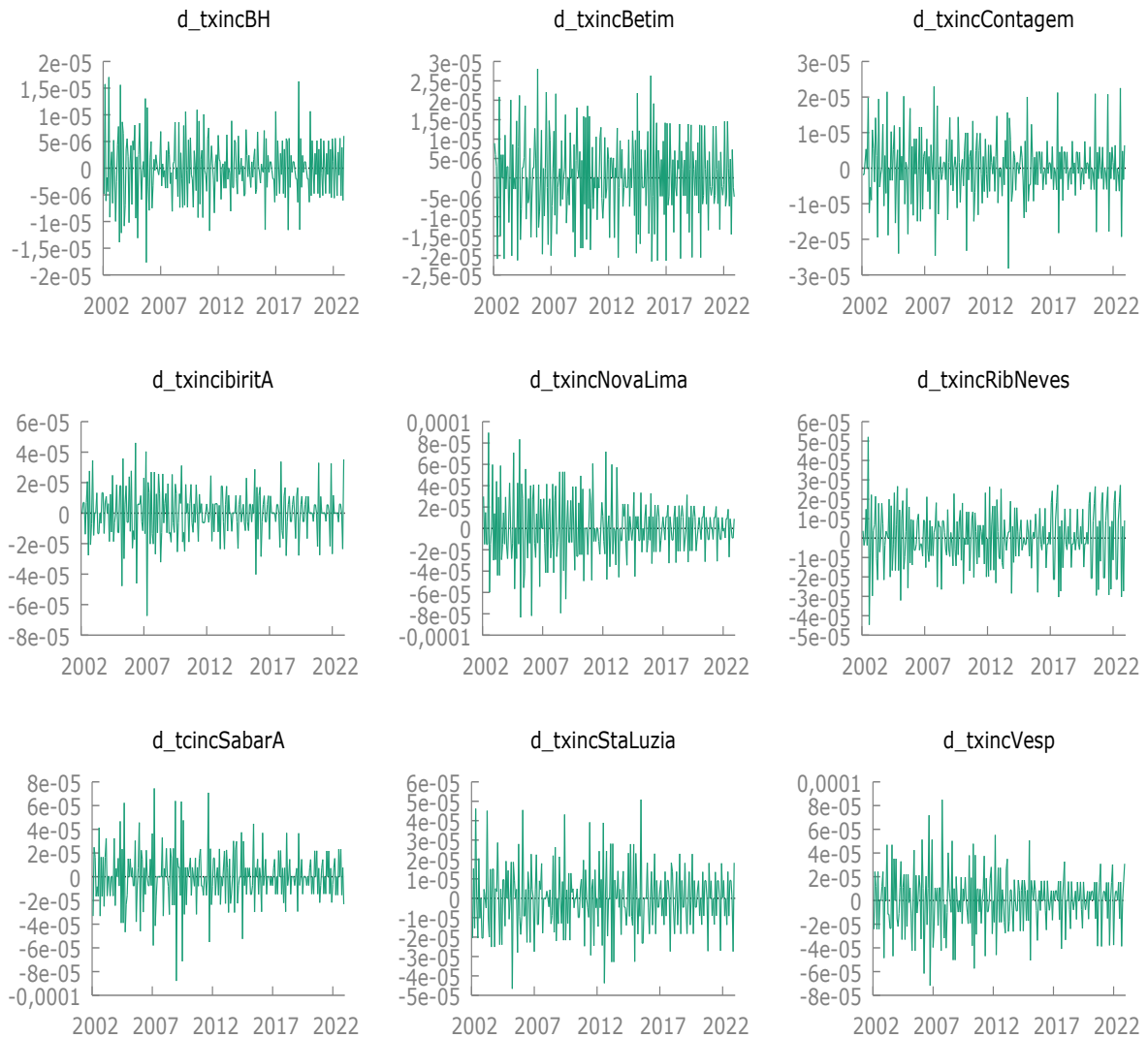
Figura 3- Gráfico da série temporal das nove cidades de interesse: Taxa de incidência de TB entre janeiro de 2002 e junho de 2022.



Fonte: Do autor (2024)

Por meio do teste F para as variáveis Dummies Sazonais, constatou-se que apenas a série de Belo Horizonte apresentava sazonalidade e com nível de significância próximo ao limiar mínimo desejável. Por isso, optou-se por não tomar diferença sazonal, preservando assim as 246 observações que compõem a série espaço-temporal.

Figura 4- Gráfico da série temporal das nove cidades de interesse: Taxa de incidência de TB entre janeiro de 2002 e junho de 2022 após diferenciação simples.



Fonte: Do autor (2024)

6.2 Índice municipal socioeconômico

Na Tabela 1 são exibidos os valores numéricos para as variáveis socioeconômicas de interesse: IDH municipal (IDH_M) e média do Previnde Brasil (PB_M) obtida dos anos 2021 a 2023. Para o IDH_M foi utilizado o valor mais atual do dado, uma vez que tal variável é feita a cada década, junto ao censo populacional do IBGE. Vale ressaltar que devido a Pandemia de Covid-19 o censo populacional de 2020 não foi feito impossibilitando a divulgação do IDH_M , que seria o valor mais atual do dado.

Quanto ao Previne Brasil, foram usadas as últimas 09 avaliações de cada um dos municípios em questão, isso porque as avaliações são quadrimestrais e foram utilizados os resultados dos anos do 2021, 2022 e 2023. A opção por utilizar os últimos 3 anos do programa de financiamento da APS, se deve ao fato de que é um programa que foi implementado recentemente, em 2019, e que devido a pandemia do Covid-19, o Ministério da Saúde, permitiu até o início de 2021, que os municípios brasileiros recebessem integralmente o valor que recebiam anteriormente. A opção pelos anos de 2021, 2022 e 2023 é a busca de representar o estado inicial e a evolução dos municípios em estudo, no que se relaciona à qualidade do atendimento e acompanhamento da APS.

As verbas destinadas a APS, fossem progressivamente repassadas aos municípios, conforme seu desempenho nos 7 indicadores do Previne Brasil. Do final de 2021 até agosto de 2022, passaram a ser cobras progressivamente as metas relacionadas a todos os 7 indicadores do Previne Brasil: primeiro passou-se a considerar a título de pagamento o desempenho nos indicadores 1, 2 e 3 (metas relacionadas aos atendimentos de pré-natal e acompanhamento odontológico da gestante), em seguida indicadores 4 e 5 (metas relacionadas a prevenção do câncer de colo de útero e a vacinação infantil). Por fim, em agosto de 2022 o repasse municipal destinado a APS era proporcional ao desempenho do setor considerando os 7 indicadores do programa nacional de financiamento da APS, englobando também os indicadores 6 e 7 (acompanhamento preventivo de pacientes com condições crônicas: diabéticos e hipertensos).

Tabela 1- Valores das variáveis socioeconômicas dos municípios de interesse.

MUNICÍPIOS	MÉDIA PREVINE BRASIL ANO DE 2021 a 2023 (PB_M)	IDH MUNICIPAL (IDH_M)
Belo Horizonte	0,471	0,81
Betim	0,541	0,749
Contagem	0,4195	0,756
Ibirité	0,5546	0,704
Nova Lima	0,5096	0,813
Ribeirão das Neves	0,3625	0,684
Sabará	0,5475	0,731
Santa Luzia	0,4604	0,715
Vespasiano	0,3258	0,688

Fonte: Do autor (2024)

Após substituição dos valores da Tabela 1 nas combinações lineares, expostas nas Equações 12 a 16, foram calculados os Índices Municipais Socioeconômicos (IMS₁, IMS₂, IMS₃, IMS₄ e IMS₅) para cada um dos nove municípios. Note que para cada município são obtidos 5 valores de IMS, conforme exibido na Tabela 2.

Tabela 2- Valores dos índices municipais socioeconomicos

MUNICIPIOS	IMS ₁	IMS ₂	IMS ₃	IMS ₄	IMS ₅
Belo Horizonte	0,8100	0,7253	0,6405	0,5558	0,4710
Betim	0,7490	0,6970	0,6450	0,5930	0,5410
Contagem	0,7560	0,6719	0,5878	0,5036	0,4195
Ibirité	0,7040	0,6667	0,6293	0,5920	0,5546
Nova Lima	0,8130	0,7372	0,6613	0,5855	0,5096
Ribeirão das Neves	0,6840	0,6036	0,5233	0,4429	0,3625
Sabará	0,7310	0,6851	0,6393	0,5934	0,5475
Santa Luzia	0,7150	0,6514	0,5877	0,5241	0,4604
Vespasiano	0,6880	0,5975	0,5069	0,4164	0,3258

Fonte: Do autor (2024)

As informações da tabela acima, servirão para a construção das matrizes **W** não espaciais tratadas na subseção seguinte.

6.3 Obtenção das matrizes de vizinhança

Inicialmente foi construída a única **W** baseada em critérios espaciais que será usada neste trabalho, a matriz de contiguidade, denominada por **W_G**. Nas Equações 21 e 22 são apresentadas **W_G** antes e após a normalização de suas linhas.

$$W_{cont} = \begin{bmatrix} 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \end{bmatrix} \quad (21)$$

$$W_G = \begin{bmatrix} 0,0000 & 0,0000 & 0,1429 & 0,1429 & 0,1429 & 0,1429 & 0,1429 & 0,1429 & 0,1429 \\ 0,0000 & 0,0000 & 0,5000 & 0,5000 & 0,0000 & 0,0000 & 0,0000 & 0,0000 & 0,0000 \\ 0,2500 & 0,2500 & 0,0000 & 0,2500 & 0,0000 & 0,2500 & 0,0000 & 0,0000 & 0,0000 \\ 0,3333 & 0,3333 & 0,3333 & 0,0000 & 0,0000 & 0,0000 & 0,0000 & 0,0000 & 0,0000 \\ 0,5000 & 0,0000 & 0,0000 & 0,0000 & 0,0000 & 0,0000 & 0,5000 & 0,0000 & 0,0000 \\ 0,3333 & 0,0000 & 0,3333 & 0,0000 & 0,0000 & 0,0000 & 0,0000 & 0,0000 & 0,3333 \\ 0,3333 & 0,0000 & 0,0000 & 0,0000 & 0,3333 & 0,0000 & 0,0000 & 0,3333 & 0,0000 \\ 0,3333 & 0,0000 & 0,0000 & 0,0000 & 0,0000 & 0,0000 & 0,3333 & 0,0000 & 0,3333 \\ 0,3333 & 0,0000 & 0,0000 & 0,0000 & 0,0000 & 0,3333 & 0,0000 & 0,3333 & 0,0000 \end{bmatrix} \quad (22)$$

Em seguida utilizando os valores do IMS exibidos na Tabela 2, foram obtidos os pesos para as cinco matrizes não espaciais (W_{IMS1} , W_{IMS2} , W_{IMS3} , W_{IMS4} , W_{IMS5}), conforme definido na Equação 17 e repetida abaixo

$$w_{K,i,j} = \frac{1}{|IMS_{K,i} - IMS_{K,j}|}$$

Vale ressaltar que, por exemplo, a matriz W_{IMS1} foi obtida considerando-se o IMS_1 e assim sucessivamente. As matrizes obtidas são exibidas nas Equações 23 a 27.

$$W_{IMS1} = \begin{bmatrix} 0,00000 & 0,03931 & 0,04441 & 0,02262 & 0,79937 & 0,01903 & 0,03036 & 0,02524 & 0,01966 \\ 0,05223 & 0,00000 & 0,45519 & 0,07081 & 0,04979 & 0,04902 & 0,17702 & 0,09371 & 0,05223 \\ 0,06361 & 0,49069 & 0,00000 & 0,06605 & 0,06026 & 0,04771 & 0,13739 & 0,08378 & 0,05051 \\ 0,03139 & 0,07395 & 0,06399 & 0,00000 & 0,03053 & 0,16639 & 0,12325 & 0,30252 & 0,20798 \\ 0,80549 & 0,03776 & 0,04239 & 0,02217 & 0,00000 & 0,01873 & 0,02947 & 0,02466 & 0,01933 \\ 0,01992 & 0,03861 & 0,03485 & 0,12547 & 0,01945 & 0,00000 & 0,05339 & 0,08095 & 0,62736 \\ 0,04786 & 0,21006 & 0,15124 & 0,14004 & 0,04611 & 0,08045 & 0,00000 & 0,23631 & 0,08793 \\ 0,03541 & 0,09895 & 0,08206 & 0,30585 & 0,03433 & 0,10853 & 0,21027 & 0,00000 & 0,12460 \\ 0,01951 & 0,03902 & 0,03501 & 0,14878 & 0,01904 & 0,59511 & 0,05536 & 0,08816 & 0,00000 \end{bmatrix} \quad (23)$$

$$W_{IMS2} = \begin{bmatrix} 0,00000 & 0,16878 & 0,08933 & 0,08136 & 0,40067 & 0,03920 & 0,011883 & 0,06452 & 0,03731 \\ 0,13619 & 0,00000 & 0,15312 & 0,12676 & 0,09582 & 0,04120 & 0,32398 & 0,08428 & 0,03865 \\ 0,04487 & 0,09533 & 0,00000 & 0,45839 & 0,03669 & 0,03509 & 0,18076 & 0,11669 & 0,03218 \\ 0,04209 & 0,08128 & 0,47211 & 0,00000 & 0,03499 & 0,03914 & 0,13352 & 0,16123 & 0,03565 \\ 0,45678 & 0,13539 & 0,08327 & 0,07710 & 0,00000 & 0,04071 & 0,10448 & 0,06335 & 0,03891 \\ 0,03261 & 0,04248 & 0,05812 & 0,06294 & 0,02971 & 0,00000 & 0,04867 & 0,08311 & 0,64236 \\ 0,08007 & 0,27057 & 0,24249 & 0,17391 & 0,06176 & 0,03942 & 0,00000 & 0,09513 & 0,03665 \\ 0,05876 & 0,09512 & 0,21157 & 0,28382 & 0,05061 & 0,09099 & 0,12857 & 0,00000 & 0,08056 \\ 0,03196 & 0,04103 & 0,05488 & 0,05903 & 0,02924 & 0,066149 & 0,04659 & 0,07578 & 0,00000 \end{bmatrix} \quad (24)$$

$$W_{IMS3} = \begin{bmatrix} 0,00000 & 0,18313 & 0,01562 & 0,07358 & 0,03962 & 0,00703 & 0,065925 & 0,01561 & 0,00617 \\ 0,38880 & 0,00000 & 0,03056 & 0,11144 & 0,10734 & 0,01437 & 0,30428 & 0,03053 & 0,01267 \\ 0,00094 & 0,00087 & 0,00000 & 0,00120 & 0,00068 & 0,00077 & 0,00097 & 0,99397 & 0,00061 \\ 0,25478 & 0,18176 & 0,06868 & 0,00000 & 0,08917 & 0,02691 & 0,28679 & 0,06860 & 0,02331 \\ 0,21186 & 0,27034 & 0,05991 & 0,13771 & 0,00000 & 0,03192 & 0,19985 & 0,05987 & 0,02854 \\ 0,06354 & 0,06120 & 0,11551 & 0,07026 & 0,1203 & 0,00000 & 0,06423 & 0,11560 & 0,45569 \\ 0,68099 & 0,14804 & 0,01653 & 0,08555 & 0,0780 & 0,000734 & 0,00000 & 0,01651 & 0,00643 \\ 0,00094 & 0,00087 & 0,99397 & 0,00119 & 0,0075 & 0,00077 & 0,00096 & 0,00000 & 0,00062 \\ 0,06094 & 0,05895 & 0,10069 & 0,06651 & 0,1163 & 0,49792 & 0,06151 & 0,10075 & 0,00000 \end{bmatrix} \quad (25)$$

$$W_{IMS4} = \begin{bmatrix} 0,00000 & 0,14793 & 0,10571 & 0,15222 & 0,18553 & 0,04882 & 0,14645 & 0,17382 & 0,03953 \\ 0,00703 & 0,00000 & 0,00293 & 0,24955 & 0,03471 & 0,00175 & 0,69875 & 0,00380 & 0,00148 \\ 0,13516 & 0,07883 & 0,00000 & 0,07977 & 0,08610 & 0,11597 & 0,07850 & 0,34494 & 0,08073 \\ 0,01474 & 0,50819 & 0,00604 & 0,00000 & 0,08209 & 0,00359 & 0,37466 & 0,00786 & 0,00304 \\ 0,06905 & 0,27165 & 0,02506 & 0,31553 & 0,00000 & 0,01438 & 0,25879 & 0,03340 & 0,01213 \\ 0,08654 & 0,06507 & 0,16080 & 0,06553 & 0,06852 & 0,00000 & 0,06491 & 0,120345 & 0,36829 \\ 0,00747 & 0,74927 & 0,00313 & 0,19718 & 0,03545 & 0,00187 & 0,00000 & 0,00405 & 0,00159 \\ 0,19466 & 0,08950 & 0,30212 & 0,09088 & 0,10050 & 0,07602 & 0,08901 & 0,00000 & 0,05730 \\ 0,08103 & 0,06394 & 0,12942 & 0,06432 & 0,06679 & 0,42582 & 0,06380 & 0,10487 & 0,00000 \end{bmatrix} \quad (26)$$

$$W_{IMS5} = \begin{bmatrix} 0,00000 & 0,07323 & 0,09953 & 0,06131 & 0,13280 & 0,04724 & 0,06701 & 0,48358 & 0,03530 \\ 0,04693 & 0,00000 & 0,02704 & 0,24156 & 0,10462 & 0,01840 & 0,50542 & 0,04076 & 0,01527 \\ 0,18211 & 0,07719 & 0,00000 & 0,06942 & 0,10409 & 0,16453 & 0,07327 & 0,22930 & 0,10009 \\ 0,04332 & 0,26626 & 0,02680 & 0,00000 & 0,08047 & 0,01885 & 0,51003 & 0,03844 & 0,01583 \\ 0,17268 & 0,21228 & 0,07398 & 0,14812 & 0,00000 & 0,04531 & 0,17587 & 0,13548 & 0,03627 \\ 0,10565 & 0,06422 & 0,20111 & 0,05967 & 0,07793 & 0,00000 & 0,06196 & 0,11709 & 0,31235 \\ 0,03598 & 0,42340 & 0,02150 & 0,38762 & 0,07261 & 0,01488 & 0,00000 & 0,03160 & 0,01241 \\ 0,49325 & 0,06487 & 0,12783 & 0,05550 & 0,10627 & 0,05341 & 0,06003 & 0,00000 & 0,03884 \\ 0,09672 & 0,06526 & 0,14988 & 0,06138 & 0,07641 & 0,38267 & 0,06335 & 0,10434 & 0,00000 \end{bmatrix} \quad (27)$$

De posse das seis matrizes \mathbf{W} é possível observar que a matriz de contiguidade apresenta muitos zeros quando comparadas as demais. Por outro lado, comparando apenas as matrizes não espaciais, é possível perceber que entre seus pesos há diferenças extremamente sutis, na grande maioria de milésimos. Antecipadamente, vale ressaltar que as diferenças, apesar de serem mínimas, refletem no ajuste dos modelos e suas respectivas previsões. Assim, demonstra-se o quão sensível os modelos STARMA (e possivelmente outros modelos espaciais e espaço-temporais) são ao uso de diferentes critérios na definição da matriz \mathbf{W} e o quão relevante e necessário são estudos que aprofundem na investigação de construção e escolha de matrizes \mathbf{W} adequadas.

6.4 Modelos STARMA

Conforme já mencionado, a obtenção de modelos STARMA usando o pacote do R(2023) denominado *starma* e envolve a execução das seguintes fases: identificação, estimação e diagnóstico do modelo. A seguir são apresentadas separadamente os resultados obtidos em cada uma dessas fases.

a) Identificação:

Antes de determinar a ordem dos modelos, foi aplicada a extensão do teste de Box-Pierce para séries espaço-temporais. Esse teste é realizado por meio da função *stcor.teste()* do pacote *starma* (CHEYSSON, 2016) e é empregado para examinar a presença de autocorrelação

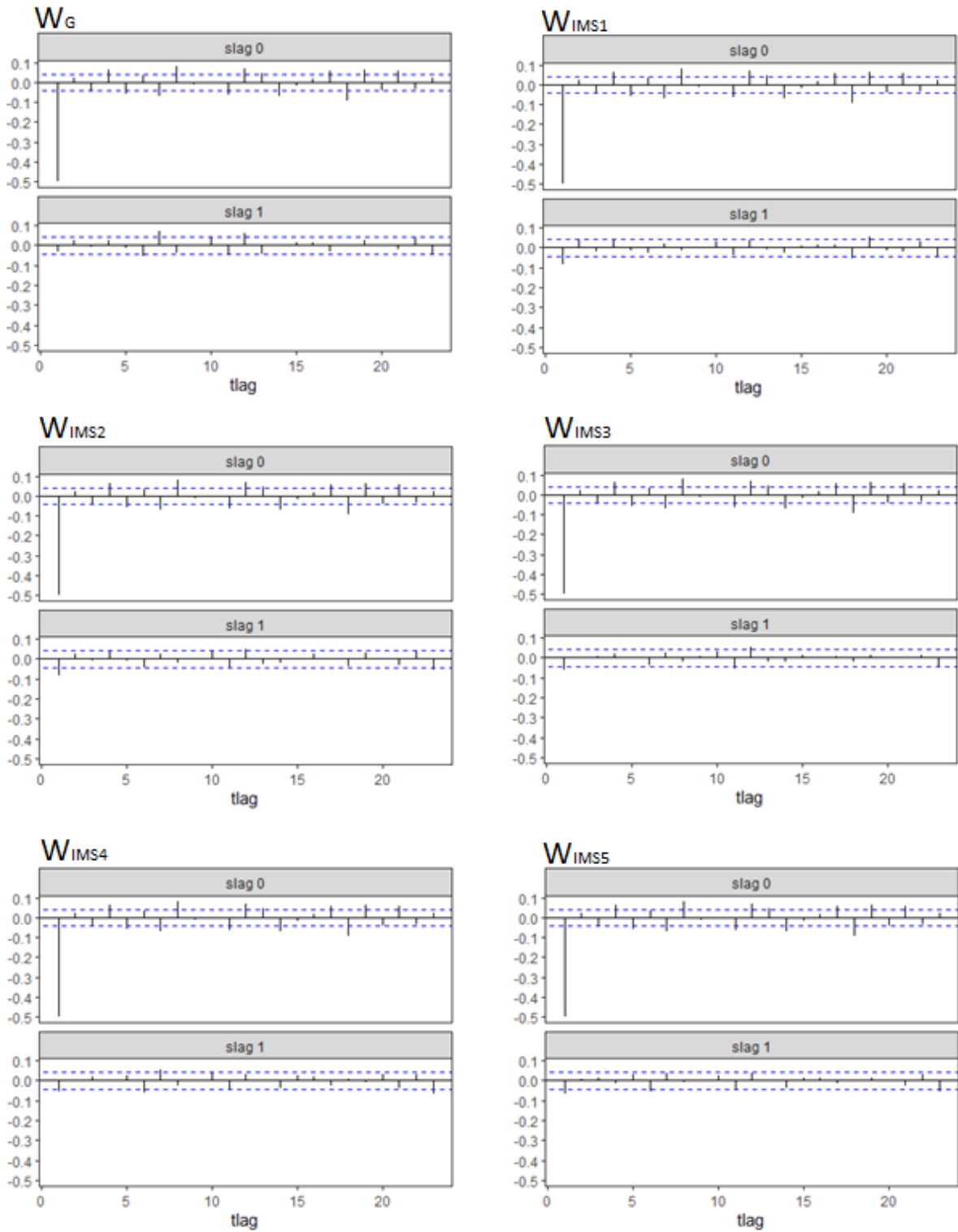
espaço-temporal no conjunto de dados, bem como para avaliar os resíduos dos modelos ajustados.

Para todas as seis matrizes de vizinhança consideradas, o resultado do teste evidenciou que a hipótese de não correlação deve ser rejeitada, concluindo-se que o conjunto de dados apresenta correlação espaço-temporal. Ressalta-se que esse teste representa uma extensão prática do teste de Box-Pierce (CHEYSSON, 2016). Vale lembrar que a série espaço-temporal a qual se aplicou toda metodologia, trata-se da junção das nove séries temporais já sem tendência.

Por meio das funções $stacf()$ e $stpacf()$ foram gerados, respectivamente, o seis gráfico da função de autocorrelação espaço-temporal ($stacf$) e os seis gráfico da função de autocorrelação parcial espaço-temporal ($stpacf$).

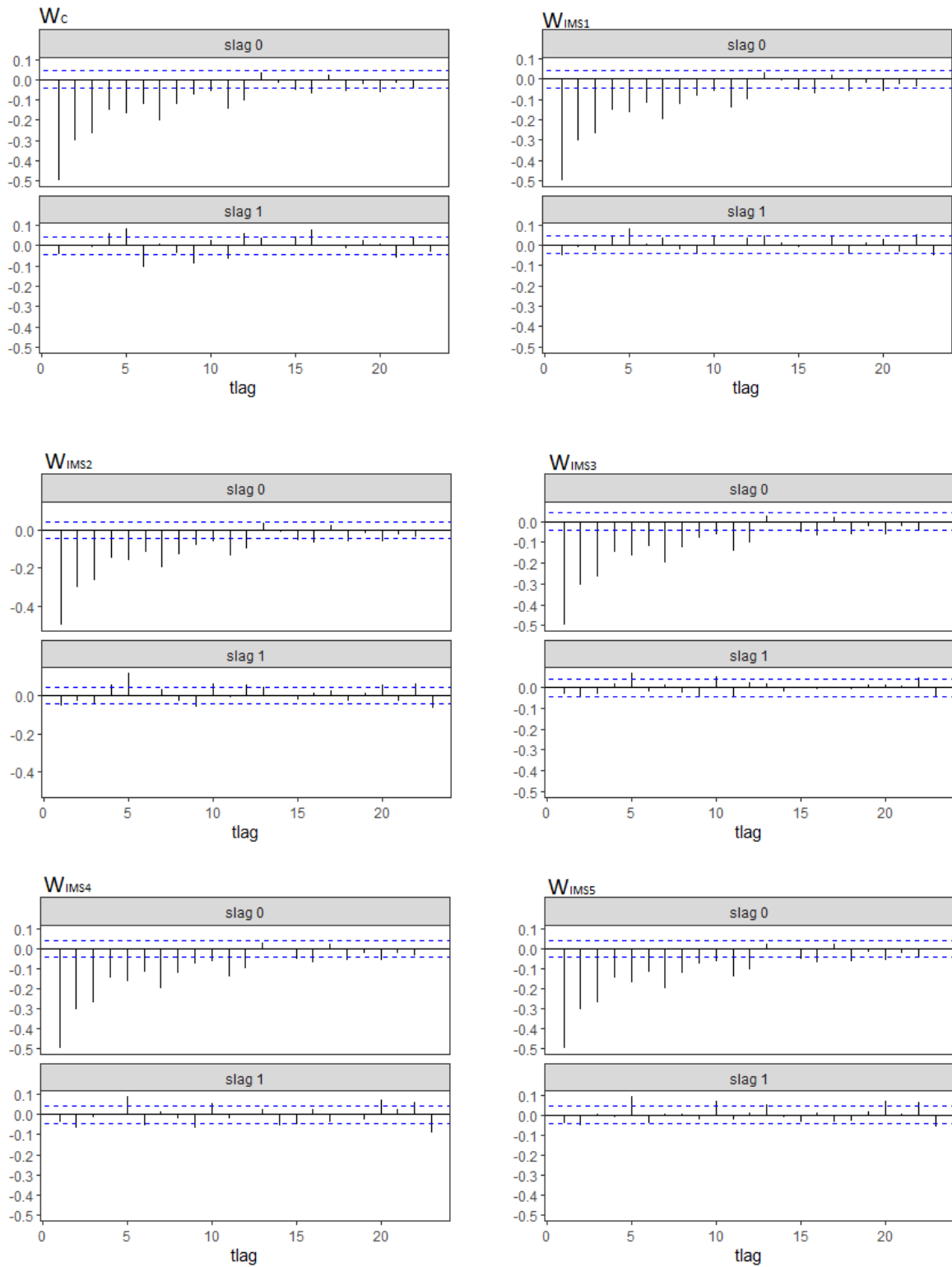
Então, considerando-se os gráficos da $stacf$ e $stpacf$, apresentados respectivamente nas Figuras 5 e 6, conclui-se que, inicialmente, os melhores candidatos seriam da classe STMA. No entanto, o comportamento decrescente da $stpacf$ é observado apenas para o $slag 0$, ou seja, para a série temporal. Para a primeira defasagem espacial ($slag 1$), tal comportamento não se repete. Por esse motivo, caso os modelos da classe STMA não obtenham êxito, é apropriado a tentativa de modelos da classe STARMA.

Figura 5- Gráficos da stacf para cada \mathbf{W} : da esquerda para a direita e de cima para baixo são exibidos os gráficos, respectivamente, para \mathbf{W}_G , \mathbf{W}_{IMS1} , \mathbf{W}_{IMS2} , \mathbf{W}_{IMS3} , \mathbf{W}_{IMS4} e \mathbf{W}_{IMS5} (sem defasagem espacial 'slag0' e com uma defasagem espacial lag1').



Fonte: Do autor (2024)

Figura 6- Gráficos da $stpacf$ para cada W : da esquerda para a direita e de cima para baixo são exibidos os gráficos, respectivamente, para W_G , W_{IMS1} , W_{IMS2} , W_{IMS3} , W_{IMS4} e W_{IMS5} (sem defasagem espacial 'slag0' e com uma defasagem espacial lag1').



Fonte: Do autor (2024)

b) Estimação:

Por meio da função ‘*starma()*’ foram ajustados inicialmente os seis modelos da classe STMA, em conformidade com o observado na fase de identificação dos modelos. Foram ajustados 6 modelos STMA, ou seja, um para cada **W**. Depois de muitas tentativas sem sucesso, optou-se por extrapolar os modelos sugeridos na fase de identificação, buscando modelos com mais parâmetros.

Assim como os modelos STMA, iniciou-se o ajuste sempre procurando modelos mais parcimoniosos possível. Visto que a quantidade reduzida de parâmetro, não estava captando efetivamente toda a correlação espaço-temporal, optou-se pelo ajuste de modelos da classe STARMA. Após algumas tentativas e a inclusão de alguns parâmetros, foram ajustados com sucesso seis modelos STARMA. Nas Tabela 3, 4, 5, 6, 7 e 8 são exibidas as estimativas dos parâmetros de cada um dos modelos, com seus respectivos erros padrão e valor-p.

Tabela 3- Parâmetros estimados do modelo STARMA (8₁, 3₁), considerando-se a matriz **W_G**

PARÂMETRO	ESTIMATIVA	ERRO PADRÃO	valor-p
$\phi_{1,0}$	-0,882892	0,023971	$< 2,2 \times 10^{-16}$
$\phi_{2,0}$	-0,773638	0,031928	$< 2,2 \times 10^{-16}$
$\phi_{3,0}$	-0,713891	0,036634	$< 2,2 \times 10^{-16}$
$\phi_{4,0}$	-0,617834	0,039865	$< 2,2 \times 10^{-16}$
$\phi_{5,0}$	-0,594380	0,041657	$< 2,2 \times 10^{-16}$
$\phi_{6,1}$	-0,102638	0,033783	0,0024084
$\phi_{7,0}$	-0,050737	0,017747	0,0042902
$\phi_{13,0}$	0,112902	0,017921	$3,584 \times 10^{-10}$
$\theta_{6,0}$	-0,531313	0,045392	$< 2,2 \times 10^{-16}$
$\theta_{12,0}$	0,085113	0,024209	0,0004475
$\theta_{12,1}$	0,139749	0,046511	0,0026891

Fonte: Do autor (2024)

A matriz de contiguidade foi a única que produziu um modelo STARMA de ordem autorregressiva maior que ordem de médias móveis. Os demais modelos, ajustados com matrizes não espaciais, apresentaram ordem de médias móveis maior que a ordem autorregressiva.

Tabela 4- Parâmetros estimados do modelo STARMA (5₁, 6₁), considerando-se W_{IMS1}

PARÂMETRO	ESTIMATIVA	ERRO PADRÃO	valor-p
$\phi_{1,0}$	-0,474527	0,069788	$1,349 \times 10^{-11}$
$\phi_{2,0}$	-0,398836	0,063031	$3,010 \times 10^{-10}$
$\phi_{6,1}$	-0,055454	0,026645	0,0375281
$\phi_{17,0}$	0,060673	0,016720	0,0002914
$\phi_{19,0}$	0,053959	0,016828	0,0013629
$\theta_{1,0}$	-0,411553	0,072806	$1,784 \times 10^{-08}$
$\theta_{3,0}$	-0,382648	0,059542	$1,597 \times 10^{-10}$
$\theta_{5,0}$	-0,085927	0,023499	0,0002616
$\theta_{8,0}$	0,111602	0,023916	$3,249 \times 10^{-06}$
$\theta_{14,0}$	-0,084277	0,023134	$3,249 \times 10^{-06}$
$\theta_{21,0}$	0,072946	0,022570	0,0012478

Fonte: Do autor (2024)

Tabela 5- Parâmetros estimados do modelo STARMA (6₁, 7₁), considerando-se W_{IMS2}

PARÂMETRO	ESTIMATIVA	ERRO PADRÃO	valor-p
$\phi_{1,0}$	-0,502350	0,071362	$2,572 \times 10^{-12}$
$\phi_{2,0}$	-0,419225	0,064650	$1,097 \times 10^{-10}$
$\phi_{3,1}$	0,095248	0,040966	0,0201608
$\phi_{6,1}$	-0,100074	0,031721	0,0016275
$\phi_{17,0}$	0,059229	0,016748	0,0004140
$\phi_{19,0}$	0,050781	0,016843	0,0026005
$\theta_{1,0}$	-0,384451	0,074897	$3,101 \times 10^{-07}$
$\theta_{3,0}$	-0,406950	0,060342	$1,964 \times 10^{-11}$
$\theta_{4,1}$	0,155979	0,056441	0,0057652
$\theta_{5,0}$	-0,083533	0,023596	0,0004083
$\theta_{8,0}$	0,107680	0,023870	$6,792 \times 10^{-06}$
$\theta_{14,0}$	-0,079597	0,023169	0,0006026
$\theta_{21,0}$	0,072888	0,022632	0,0012983

Fonte: Do autor (2024)

Tabela 6- Parâmetros estimados do modelo STARMA (5₁, 6₁) considerando-se W_{IMS3}

PARÂMETRO	ESTIMATIVA	ERRO PADRÃO	valor-p
$\phi_{1,0}$	-0,472258	0,068459	$6,856 \times 10^{-12}$
$\phi_{2,0}$	-0,397105	0,062161	$2,041 \times 10^{-10}$
$\phi_{12,1}$	0,067630	0,029314	0,0211432
$\phi_{17,0}$	0,063593	0,016833	0,0001623
$\phi_{19,0}$	0,053298	0,016896	0,0016299
$\theta_{1,0}$	-0,412886	0,071778	$1,004 \times 10^{-08}$
$\theta_{3,0}$	-0,379336	0,059165	$1,758 \times 10^{-10}$
$\theta_{5,0}$	-0,081325	0,023634	0,0005904
$\theta_{8,0}$	0,114197	0,023996	$2,073 \times 10^{-06}$
$\theta_{14,0}$	-0,084099	0,023294	0,0003127
$\theta_{21,0}$	0,072200	0,022736	0,0015165

Fonte: Do autor (2024)

Tabela 7- Parâmetros estimados do modelo STARMA (5₁, 7₁) considerando-se W_{IMS4}

PARÂMETRO	ESTIMATIVA	ERRO PADRÃO	valor-p
$\phi_{1,0}$	-0,484493	0,070434	$7,849 \times 10^{-12}$
$\phi_{2,0}$	-0,405465	0,063181	$1,692 \times 10^{-10}$
$\phi_{10,1}$	0,074428	0,033304	0,0255316
$\phi_{17,0}$	0,061747	0,017768	0,0005204
$\phi_{19,0}$	0,053297	0,017936	0,0029963
$\theta_{1,0}$	-0,402303	0,073662	$5,256 \times 10^{-08}$
$\theta_{3,0}$	-0,394470	0,059169	$3,296 \times 10^{-11}$
$\theta_{5,0}$	-0,092133	0,025039	0,0002393
$\theta_{8,0}$	0,107737	0,025310	$2,162 \times 10^{-05}$
$\theta_{14,0}$	-0,086845	0,024607	0,0004252
$\theta_{21,0}$	0,075687	0,024009	0,0016414

Fonte: Do autor (2024)

Tabela 8- Parâmetros estimados do STARMA (5₁, 7₁) considerando-se WIMS

PARÂMETRO	ESTIMATIVA	ERRO PADRÃO	valor-p
$\emptyset_{1,0}$	-0,466437	0,073739	3,049x10 ⁻¹⁰
$\emptyset_{2,0}$	-0,393419	0,066589	3,999x10 ⁻⁰⁹
$\emptyset_{6,1}$	-0,081542	0,034840	0,0193496
$\emptyset_{17,0}$	0,060969	0,017321	0,0004403
$\emptyset_{19,0}$	0,052716	0,017390	0,0024632
$\theta_{1,0}$	-0,417912	0,077040	6,447x10 ⁻⁰⁸
$\theta_{3,0}$	-0,376180	0,062852	2,519x10 ⁻⁰⁹
$\theta_{5,0}$	-0,083955	0,024287	0,0005572
$\theta_{8,0}$	0,112249	0,024647	5,544x10 ⁻⁰⁶
$\theta_{14,0}$	-0,084057	0,023931	0,0004531
$\theta_{21,0}$	0,070676	0,023344	0,0024940

Fonte: Do autor (2024)

Considerando os parâmetros estimados, os modelos STARMA ajustados são apresentados nas Equações 28 a 33, considerando, respectivamente, as matrizes W_G , W_{IMS1} , W_{IMS2} , W_{IMS3} , W_{IMS4} e W_{IMS5} .

$$\begin{aligned}
\mathbf{Z}(t) = & -0,882892 \mathbf{I} \mathbf{Z}(t-1) - 0,773638 \mathbf{I} \mathbf{Z}(t-2) - 0,713891 \mathbf{I} \mathbf{Z}(t-3) \\
& - 0,617834 \mathbf{I} \mathbf{Z}(t-4) - 0,594380 \mathbf{I} \mathbf{Z}(t-5) - 0,102638 W_G \mathbf{Z}(t-6) \\
& - 0,050737 \mathbf{I} \mathbf{Z}(t-7) + 0,112902 \mathbf{I} \mathbf{Z}(t-13) - 0,531313 \mathbf{I} \epsilon(t-6) \\
& + 0,085113 \mathbf{I} \epsilon(t-12) + 0,139749 W_G \epsilon(t-12) + \epsilon(t)
\end{aligned}
\tag{28}$$

$$\begin{aligned}
\mathbf{Z}(t) = & -0,474527 \mathbf{I} \mathbf{Z}(t-1) - 0,398836 \mathbf{I} \mathbf{Z}(t-2) - 0,055454 W_{IMS1} \mathbf{Z}(t-6) \\
& + 0,060673 \mathbf{I} \mathbf{Z}(t-17) + 0,053959 \mathbf{I} \mathbf{Z}(t-19) - 0,411553 \mathbf{I} \epsilon(t-1) \\
& - 0,382648 \mathbf{I} \epsilon(t-3) - 0,085927 \mathbf{I} \epsilon(t-5) + 0,111602 \mathbf{I} \epsilon(t-8) \\
& - 0,084277 \mathbf{I} \epsilon(t-14) + 0,072946 \mathbf{I} \epsilon(t-21) + \epsilon(t)
\end{aligned}
\tag{29}$$

$$\begin{aligned}
\mathbf{Z}(t) = & -0,502350 \mathbf{I} \mathbf{Z}(t-1) - 0,419225 \mathbf{I} \mathbf{Z}(t-2) + 0,095248 \mathbf{W}_{IMS2} \mathbf{Z}(t-3) \\
& - 0,100074 \mathbf{W}_{IMS2} \mathbf{Z}(t-6) + 0,059229 \mathbf{I} \mathbf{Z}(t-17) \\
& + 0,050781 \mathbf{I} \mathbf{Z}(t-19) - 0,384451 \mathbf{I} \epsilon(t-1) - 0,406950 \epsilon \mathbf{I}(t-3) \\
& + 0,155979 \mathbf{W}_{IMS2} \epsilon(t-4) - 0,083533 \epsilon \mathbf{I}(t-5) + 0,107680 \epsilon \mathbf{I}(t-8) \\
& - 0,079597 \mathbf{I} \epsilon(t-14) + 0,072888 \mathbf{I} \epsilon(t-21) + \epsilon(t)
\end{aligned} \tag{30}$$

$$\begin{aligned}
\mathbf{Z}(t) = & -0,472258 \mathbf{I} \mathbf{Z}(t-1) - 0,397105 \mathbf{I} \mathbf{Z}(t-2) + 0,067630 \mathbf{W}_{IMS3} \mathbf{Z}(t-12) \\
& + 0,063593 \mathbf{I} \mathbf{Z}(t-17) + 0,053298 \mathbf{I} \mathbf{Z}(t-19) - 0,412886 \mathbf{I} \epsilon(t-1) \\
& - 0,379336 \epsilon \mathbf{I}(t-3) - 0,081325 \mathbf{I} \epsilon(t-5) + 0,114197 \mathbf{I} \epsilon(t-8) \\
& - 0,084099 \mathbf{I} \epsilon(t-14) + 0,072200 \mathbf{I} \epsilon(t-21) + \epsilon(t)
\end{aligned} \tag{31}$$

$$\begin{aligned}
\mathbf{Z}(t) = & -0,484493 \mathbf{I} \mathbf{Z}(t-1) - 0,405465 \mathbf{I} \mathbf{Z}(t-2) + 0,074428 \mathbf{W}_{IMS4} \mathbf{Z}(t-10) \\
& + 0,061474 \mathbf{I} \mathbf{Z}(t-17) + 0,053297 \mathbf{I} \mathbf{Z}(t-19) - 0,402303 \mathbf{I} \epsilon(t-1) \\
& - 0,394470 \epsilon \mathbf{I}(t-3) - 0,092133 \mathbf{I} \epsilon(t-5) + 0,107737 \mathbf{I} \epsilon(t-8) \\
& - 0,086845 \mathbf{I} \epsilon(t-14) + 0,075687 \mathbf{I} \epsilon(t-21) + \epsilon(t)
\end{aligned} \tag{32}$$

$$\begin{aligned}
\mathbf{Z}(t) = & -0,466437 \mathbf{I} \mathbf{Z}(t-1) - 0,393419 \mathbf{I} \mathbf{Z}(t-2) - 0,081542 \mathbf{W}_{IMS5} \mathbf{Z}(t-6) \\
& + 0,060969 \mathbf{I} \mathbf{Z}(t-17) + 0,052716 \mathbf{I} \mathbf{Z}(t-19) - 0,417912 \mathbf{I} \epsilon(t-1) \\
& - 0,376180 \epsilon \mathbf{I}(t-3) - 0,083955 \mathbf{I} \epsilon(t-5) + 0,112249 \mathbf{I} \epsilon(t-8) \\
& - 0,084057 \mathbf{I} \epsilon(t-14) + 0,070676 \mathbf{I} \epsilon(t-21) + \epsilon(t)
\end{aligned} \tag{33}$$

c) Diagnóstico do ajuste:

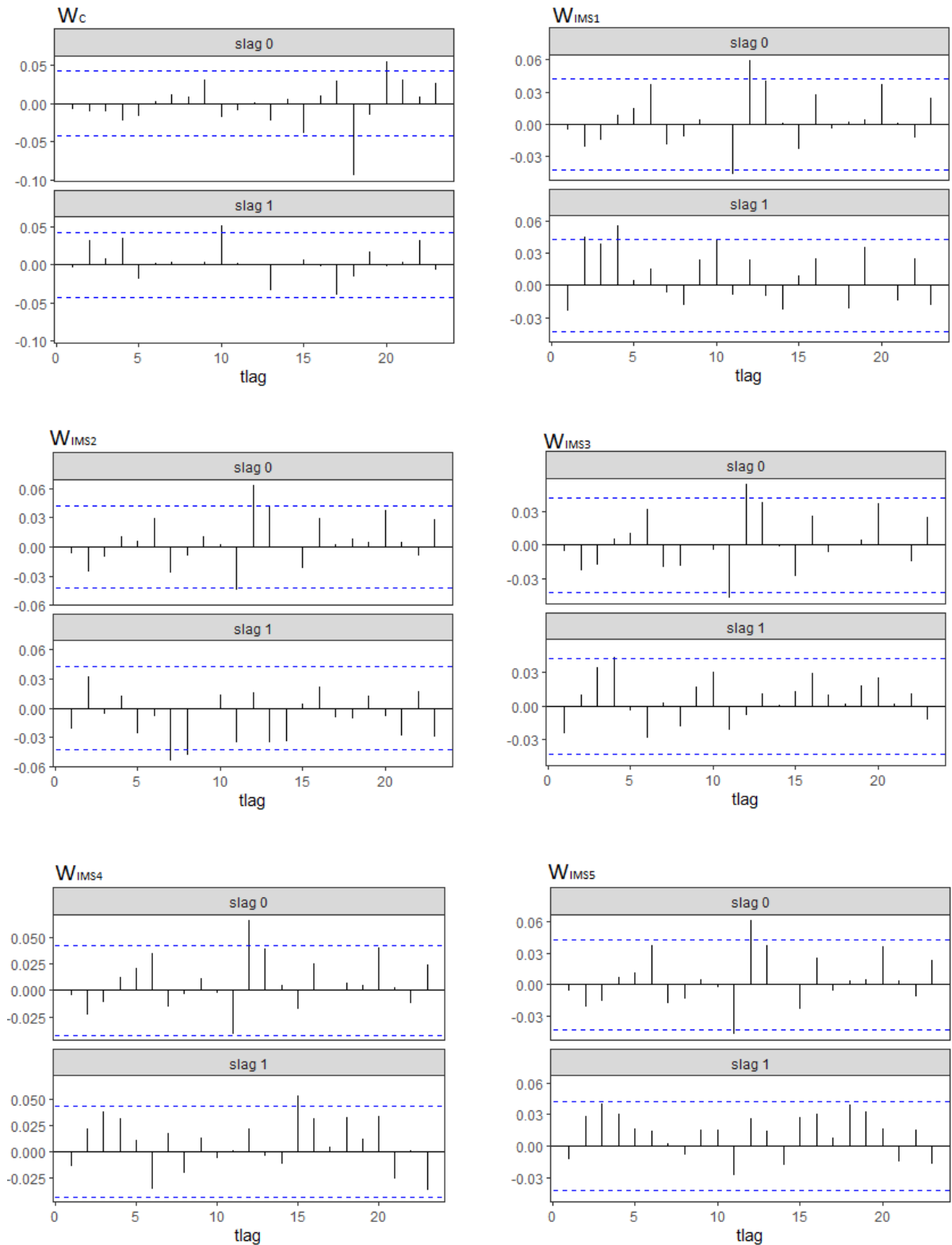
O diagnóstico de adequação de cada modelo consiste em verificar se os resíduos do modelo ajustado apresentam comportamento semelhante ao de um ruído branco, ou seja, se são independentes e têm distribuição gaussiana multivariada com média nula e variância constante.

Essa verificação pode ser realizada de duas maneiras: a primeira consiste em aplicar a extensão espaço-temporal do teste de Box-Pierce, já apresentado na Equação 18. A segunda maneira trata-se da análise visual do gráfico da *stacf*, verificando se apenas 5% dos valores de autocorrelação espaço-temporal (popularmente denominados *lags*) estão fora do intervalo de confiança do gráfico.

Para os seis modelos ajustados, o resultado da extensão espaço-temporal do teste indicou a não rejeição (aceitação) da hipótese nula de não correlação residual. Isso demonstra que a

correlação espaço-temporal, inicialmente presente nos dados, foi totalmente ajustada pelos modelos considerados. Na Figura 7 têm-se os gráficos da stacf residual de cada um dos seis modelos.

Figura 7- Gráficos da stacf residual: da esquerda para a direita e de cima para baixo são exibidos os gráficos obtidos, respectivamente, para W_G , W_{IMS1} , W_{IMS2} , W_{IMS3} , W_{IMS4} e W_{IMS5}



Fonte: Do autor (2024)

Tabela 9- Parâmetro de comparação dos modelos STARMA ajustados.

W UTILIZADA NO AJUSTE	BIC	MAPE
W_G	5035,040	11,06%
W_{IMS1}	5010,153	22,28%
W_{IMS2}	5018,702	1928,12%
W_{IMS3}	5007,762	20,24%
W_{IMS4}	5038,389	20,25%
W_{IMS5}	5019,085	21,96%

Fonte: Do autor (2024)

Apesar de não incorporar todas as cidades da grande BH, a análise dos resultados apresentados na Tabela 9, sinaliza resultados interessante e animadores para estudos futuros. Da Tabela 9, pode-se concluir que o modelo que melhor se ajustou aos dados foi o modelo com a matriz **W_{IMS3}**. Ou seja, o modelo que melhor se ajustou aos dados é aquele que foi ajustado com a matriz não espacial com igual ponderação das variáveis socioeconômicas (05.IDH municipal + 0,5. Previne Brasil).

Esse resultado evidencia a relevância e a necessidade de ainda mais pesquisas que proponham métodos de criação de matrizes não espaciais aplicados a modelos espaço-temporais. Os valores de BIC demonstram que quase todos os modelos construídos com **W** não espaciais, se ajustaram melhor aos dados quando comparado ao modelo construído com a matriz de contiguidade. Outra importante conclusão depreendida da Tabela 9, é a corroboração de que a disseminação da TB é fortemente relacionada a fatores socioeconômicos. Isso reforça ainda mais que provavelmente os fatores socioeconômicos expliquem melhor a disseminação de TB do que as interações do espaço na qual ela se dissemina. Além disso, tal resultado estimula a investigação se o mesmo não ocorreria com outros dados epidemiológicos como, por exemplo, de Covid-19, HIV, malária e outras patologias.

Os resultados da Tabela 9, também sugere que além dos muitos fatores socioeconômicos já sabidamente relacionados a incidência de TB, e destacados, por exemplo, em Bethlem, 2012; Coelho *et al.*, 2010; Guimarães *et al.*, 2012; Hijjar, 2001; Modrá *et al.*, 2019; Ploubidis *et al.*, 2012 é igualmente impactante na disseminação da TB a qualidade dos serviços prestados pela APS, avaliados pelo Previne Brasil.

Nesse trabalho optou-se pela obtenção do IMS por meio de diferentes combinações lineares que o peso de cada combinação linear variasse de 0,25 para mais ou para menos. Formando, assim, um *grid* em que a soma dos pesos de cada uma das combinações lineares

somasse um. É uma pretensão futura trabalhar ainda com esse grid de soma unitária vareando os pesos em 0,20 e 0,10.

Quanto às previsões, o modelo que obteve melhor resultado foi o modelo obtido com a matriz de contiguidade (\mathbf{W}_G), provavelmente por ser o único modelo de ordem predominantemente autorregressiva. Para algumas previsões, é preciso que tal modelo considere valores estimados de previsões de meses anteriores, o que pode ocasionar erros.

No entanto, os demais modelos, com ordem majoritária de médias móveis, além de depender de previsões anteriores, tem sua previsão em um determinado instante 't' influenciada também pelos erros das previsões estimadas anteriores. Portanto, é provável que isso tenha ocasionado valores maiores de MAPE para os modelos baseados em \mathbf{W} não espaciais, inclusive para o modelo que melhor se ajustou aos dados.

7 CONCLUSÕES

O presente trabalho alcançou de maneira positiva todos os objetivos elencados na seção 2, tanto os específicos quanto o principal. Portanto, após o ajuste dos seis modelos STARMA com as matrizes propostas, foi possível concluir que a metodologia proposta foi adequada ao analisar o conjunto de dados em questão.

Embora seja necessário realizar estudos mais aprofundados para que as conclusões possam ser tomadas de maneira mais generalizada, este trabalho nos permite afirmar que algumas vezes as matrizes de vizinhança baseadas em critérios não espaciais podem ser mais adequadas do que as matrizes de vizinhança tradicionalmente tratadas nas bibliografias de Geoestatística e Estatística Espacial. Assim, é prudente, para quase todo estudo, dependendo do conjunto de dados e do modelo a ser ajustado, avaliar o uso de matrizes não espaciais.

Outra singularidade deste trabalho foi a utilização bem-sucedida e eficiente do que, no texto, foi denominado IMS (Índice Municipal Socioeconômico) na construção das matrizes **W** não espaciais. Tal feito reforça a necessidade de estudos aprofundados, não apenas sobre a construção de matrizes **W** baseadas em variáveis não espaciais, mas também sobre a criação de matrizes com combinações das variáveis socioeconômicas consideradas.

8 CONSIDERAÇÕES FINAIS

Buscando o aprimoramento do que foi proposto e realizado neste trabalho, pretende-se no futuro cumprir as seguintes propostas:

- a) estender o conjunto de dados, englobando todas as cidades que compõem a grande BH e verificar se a metodologia empregada será pertinente;
- b) testar para os nove municípios desse trabalho e para a grande BH se a metodologia é pertinente a outros conjuntos de dados. Primeiros dados epidemiológicos, especificamente disseminação e/ou óbitos por Covid-19. E posteriormente para outros tipos de dados como, por exemplo, relacionados a criminalidade;
- c) validar a metodologia por meio de simulação computacional, buscando verificar se é possível obter generalizações sobre quando e como tal metodologia é mais eficiente do que a abordagem tradicional.

REFERÊNCIAS

- ALMEIDA, E. **Econometria espacial aplicada**. Campinas–SP. Alínea, 2012.
- AMSALU, E. *et al.* Spatial-temporal analysis of tuberculosis in the geriatric population of China: An analysis based on the Bayesian conditional autoregressive model. **Archives of gerontology and geriatrics**, v. 83, p. 328-337, 2019.
- ANTUNES, J. L. F.; CARDOSO, M. R. A. Uso da análise de séries temporais em estudos epidemiológicos. **Epidemiologia e Serviços de Saúde**, v. 24, p. 565-576, 2015.
- ASTUTI, D.; RUCHJANA B. N.; SOEMARTINI. Generalized space time autoregressive with exogenous variable model and its application. In: **Journal of Physics: Conference Series**, 2017.bethl
- BETHLEM, E. P. Manifestações clínicas da tuberculose pleural, ganglionar, geniturinária e do sistema nervoso central. **Pulmão RJ**, v. 21, n. 1, p. 19-22, 2012.
- BIZ, G. **Simulações de pesos espaciais para o modelo STARMA e aplicações**. Tese de Doutorado. Universidade de São Paulo, 2014.
- BOX, G. E. P.; JENKINS, G. M.; REINSEL, G. C. **Time Series Analysis**. New Jersey: John Wiley & Sons, 2008.
- BOX, G. E. P.; PIERCE, D. **Distribution of residual autocorrelations in autoregressive-integrated moving average time series models**. Journal of the American Statistical Association, v 65, p. 1509-1526, 1970.
- BRASIL. Doenças negligenciadas: estratégias do Ministério da Saúde. **Rev Saúde Pública** v. 44, n. 1, 2010.
- BRASIL. Ministério da Saúde. Secretaria de Vigilância em Saúde. Departamento de Vigilância Epidemiológica. Manual de recomendações para o controle da tuberculose no Brasil. Brasília; 2011. Disponível em: < http://www.cve.saude.sp.gov.br/htm/TB/mat_tec/manuais/MS11_Manual_Recom.pdf >.
- BROCKWELL, P. J.; DAVIS, R. A.; CALDER, M. V. **Introduction to time series and forecasting**. New York: Springer, 2002.
- CASE, A. C.; ROSEN, H. S.; HINES JR, James R. Budget spillovers and fiscal policy interdependence: Evidence from the states. **Journal of public economics**, v. 52, n. 3, p. 285-307, 1993.
- CHAVES, V. L. V. et al. Evolução espaço-temporal do sobrepeso e da obesidade em adolescentes masculinos brasileiros, 1980 a 2005. **Cadernos de Saúde Pública**, v. 26, p. 1303-1313, 2010.
- CHEN, J. *et al.* Localized space-time autoregressive parameters estimation for traffic flow prediction in urban road networks. **Applied Sciences**, v. 8, n. 2, p. 277, 2018.

- CHEYSSON, F. (2016). Starma: Modelling Space Time Auto Regressive Moving Average (STARMA). Processes. R package version 1.3. <https://CRAN.R-project.org/package=starma>
- CLIFF, A; ORD, J. K. **Spatial process, models and applications**. Londres; Pion, 1981.
- COELHO, D. M. M. et al. Perfil epidemiológico da tuberculose no Município de Teresina-PI, no período de 1999 a 2005. **Epidemiologia e Serviços de Saúde**, v. 19, n. 1, p. 34-43, 2010
- COLDITZ, G. A. et al. Efficacy of BCG vaccine in the prevention of tuberculosis: meta-analysis of the published literature. **JAMA**, 271(9), 698-702.1994
- EFF, E. Anthon. Weight matrices for cultural proximity: Deriving weights from a language phylogeny. **Structure and Dynamics**, v. 3, n. 2, 2008.
- FERREIRA, A. de A. *et al.* Os fatores associados à tuberculose pulmonar e a baciloscopia: uma contribuição ao diagnóstico nos serviços de saúde pública. **Revista Brasileira de Epidemiologia**, v. 8, p. 142-149, 2005.
- FREITAS, M. F. *et al.* Utilização de matrizes de vizinhança socioeconômicas em modelos da classe STARMA aplicados a dados epidemiológicos. **Sigmae**, v. 8, n. 2, p. 29-35, 2019.
- GAGNEUX, S. Ecology and evolution of Mycobacterium tuberculosis. **Nature Reviews Microbiology**, v. 16, n. 4, p. 202, 2018.
- GUIMARÃES, R. M. *et al.* Tuberculose, HIV e pobreza: tendência temporal no Brasil, Américas e mundo. **Jornal Brasileiro de Pneumologia**, v. 38, n. 4, p. 518-525, 2012.
- HERNÁNDEZ-MURILLO, R. *et al.* Interação estratégica nas políticas tributárias entre os estados. **Review-Federal Reserve Bank de Saint Louis** , v. 85, n. 3, pág. 47-56, 2003.
- HIJJAR, M. A.; OLIVEIRA, M. J. P. R. de; TEIXEIRA, G. M. A tuberculose no Brasil e no mundo. **Boletim de Pneumologia Sanitária**, v. 9, n. 2, p. 9-16, 2001.
- HINO, P.; SANTOS, C. B. dos; VILLA, T. C. S. Evolução espaço-temporal dos casos de tuberculose em Ribeirão Preto (SP), nos anos de 1998 a 2002. **J BrasPneumol**, v. 31, n. 6, p. 523-7, 2005.
- HOEF, J. M. V. *et al.* Spatial autoregressive models for statistical inference from ecological data. **Ecological Monographs**, v. 88, n. 1, p. 36-59, 2018.
- IP, R. HL. *et al.* Evaluation of COVID-19 Mitigation Policies in Australia Using Generalised Space-Time Autoregressive Intervention Models. **International Journal of Environmental Research and Public Health**, v. 18, n. 14, p. 7474, 2021.
- JIN, E. Y. **Estrutura de vizinhanças espaciais nos modelos autorregressivos e de médias móveis espaço-temporais STARMA**. Dissertação de Mestrado. Universidade de São Paulo, 2017.
- KAMARIANAKIS, Y.; PRASTACOS, P. Modelagem espaço-temporal do fluxo de tráfego. **Computadores e Geociências**, v. 31, n. 2, pág. 119-133, 2005.

LATORRE, M. do R. D. de *et al.* Análise de séries temporais em epidemiologia: uma introdução sobre os aspectos metodológicos. **Revista Brasileira de Epidemiologia**, v. 4, p. 145-152, 2001.

LEE, D.; RUSHWORTH, A.; NAPIER, G. Spatio-temporal areal unit modeling in R with conditional autoregressive priors using the CARBayesST package. **Journal of Statistical Software**, v. 84, p. 1-39, 2018.

LEE D (2013). "CARBayes: An R Package for Bayesian Spatial Modeling with Conditional Autoregressive Priors." *Journal of Statistical Software*. Disponível em : <https://www.jstatsoft.org/htaccess.php?volume=55&type=i&issue=13>.

MARTIN, R.L.; OEPPEN, J.E. The identification of regional forecasting models using space-time correlation functions. *Nova York*, n. 66, p. 95-118, 1975.

MARTINS, N. S.; BIZ, G. I. Z.; OZAKI, V.A. Modelo auto regressivo de média móvel espaço-temporal aplicado em dados de temperaturas mínimas médias mensais. *Rev. Bras. Biom*, v. 32, n. 1, p. 158-169, 2014.

MELO, T. E. M. de P. *et al.* Distribuição espacial e temporal da tuberculose em indígenas e não indígenas de Rondônia, Amazônia Ocidental, Brasil. **Cadernos de Saúde Pública**, v. 28, p. 267-280, 2012.

MISHRA, A.; SUROLIA, A. Mycobacterium tuberculosis: surviving and indulging in an unwelcoming host. **IUBMB life**, 2018.

MODRÁ, H. *et al.* Socio-economic and environmental factors related to spatial differences in human non-tuberculous mycobacterial diseases in the Czech Republic. **International Journal of Environmental Research and Public Health**, v. 16, n.20, p. 3969, 2019.

MONTEIRO, A. M. V. *et al.* **Análise espacial de dados geográficos**. Planaltina: Embrapa, 2004.

MORETTIN, P. A.; TOLOI, C. **Análise de séries temporais**. In: *Análise de séries temporais*. 2006.

NASCIMENTO, L. A. *et al.* Geração automatizada de séries temporais de dados posicionais utilizando o IBGE-PPP/Automated generation of time series of positional data using the IBGE-PPP. **Boletim de Ciências Geodésicas**, v. 23, n. 1, p. 150, 2017.

NOBRE, A. A.; SCHMIDT, A. M.; LOPES, H. F. Spatio-temporal models for mapping the incidence of malaria in Pará. **Environmetrics: The official journal of the International Environmetrics Society**, v. 16, n. 3, p. 291-304, 2005.

NURHAYATI, N.; PASARIBU, U. S.; NESWAN, O. Application of generalized space-time autoregressive model on GDP data in West European countries. **Journal of Probability and Statistics**, v. 2012, 2012.

PACE, R. K.; BARRY, R.; W., G. O.; SIRMANS, F. C. A method for spatial temporal forecasting with application to real estate prices. **International Journal of Forecasting**. Louisiana, n.16, p. 229-246, 2000

- PELLISSARI, D. M.; DIAZ-QUIJANO, F. A. Household crowding as a potential mediator of socioeconomic determinants of tuberculosis incidence in Brazil. **PloSone**, v. 12, n. 4, p. e0176116, 2017.
- PFEIFER, P. E.; DEUTRCH, S. J. A three-stage iterative procedure for space-time modeling phillip. **Technometrics**, v. 22, n. 1, p. 35-47, 1980a.
- PFEIFER, P. E.; DEUTSCH, S. J. Identification and interpretation of first order space-time ARMA models. **Technometrics**, v. 22, n. 3, p. 397-408, 1980b.
- PFEIFER, P. E.; DEUTSCH, S. J. A STARIMA model-building procedure with application to description and regional forecasting. **Transactions of the Institute of British Geographers**, p. 330-349, 1980c.
- PLOUBIDIS, G. B. *et al.* Social determinants of tuberculosis in Europe: a prospective ecological study. **European Respiratory Journal**, v. 40, n. 4, p. 925-930, 2012.
- PRADO, R. R. do *et al.* A epidemia de aids no Estado de São Paulo: uma aplicação do modelo espaço-temporal bayesiano completo. **Revista da Sociedade Brasileira de Medicina Tropical**, v. 42, n. 5, p. 537-542, 2009.
- R Core Team (2023). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- RAO, T. S.; ANTUNES, A. M. C. Spatio-temporal model ling of temperature time series: a comparativestudy. In: **Time series analysis and applications to geophysical systems**. Springer, New York, NY, 2004.
- RIBEIRO S.N. *et al.* **Pneumologia Newton Bethlem**. Ed.: Atheneu, 4^a edição, pag 379-448, 1995
- RODRIGUES-JUNIOR, A. L.; CASTILHO, E. A. de. A epidemia de AIDS no Brasil, 1991-2000: descrição espaço-temporal. **Rev Soc Bras Med Trop**, v. 37, n. 4, p. 312-7, 2004.
- RODRIGUES-JUNIOR, A. L.; CASTILHO, E. A. de. A AIDS nas regiões de fronteira no Brasil de 1990 a 2003. **Revista Panamericana de Salud Pública**, v. 25, p. 31-38, 2009.
- SÁFADI, T. Uso de séries temporais na análise de vazão de água na represa de Furnas. **Ciência e Agrotecnologia**, v. 28, n. 1, p. 142-148, 2004.
- SANTOS NETO, M. *et al.* Pulmonary tuberculosis in São Luis, State of Maranhão, Brazil: space and space-time risk clusters for death (2008-2012). **Revista da Sociedade Brasileira de Medicina Tropical**, v. 48, n. 1, p. 69-76, 2015.
- SHOESMITH, G. L. Space–time autoregressive models and forecasting national, regional and state crime rates. **International journal of forecasting**, v. 29, n. 1, p. 191-201, 2013.
- SHUMWAY, R. H.; STOFFER, D. S. **Time Series Analysis and Its Applications With R Examples**. New York: Springer, 2006.
- SMITH, L. H. Plot arrangement for variety experiment with corn. **Proceedings of the American Society of Agronomy**, New York, v. 107, n. 9, p. 84-89, 1910

SMITH, T. E. **Spatial weight matrices**. 2013. Disponível em: <http://www.seas.upenn.edu/~ese502/lab-content/extra_materials/>. Acesso 31 de março de 2020.

SZWARCWALD, C. L. *et al.* Dinâmica da epidemia de AIDS no Município do Rio de Janeiro, no período de 1988-1996: uma aplicação de análise estatística espaço-temporal. **Cadernos de Saúde Pública**, v. 17, p. 1123-1140, 2001.

TAQUETTE, S. R. *et al.* The AIDS epidemic in adolescents between the ages of 13 and 19 in the City of Rio de Janeiro, Brazil: space-time description. **Revista da Sociedade Brasileira de Medicina Tropical**, v. 44, n. 4, p. 467-470, 2011.

TYSZLER, M. **Econometria espacial: discutindo medidas para a matriz de ponderação espacial**. Tese de Doutorado, 2006.

WHO: World Health Organization- Global Tuberculosis Report 2022
Disponível em : <[https://www.thelancet.com/journals/lanmic/article/PIIS2666-5247\(22\)00359-7/fulltext#articleInformation](https://www.thelancet.com/journals/lanmic/article/PIIS2666-5247(22)00359-7/fulltext#articleInformation)>

ZHENG *et al.* **Short-term forecasting of water logging a turbanstorm-water logging monitoring sites based starma model**. *Progress in geography*, 33(7). 949-957. 2014

ZHOU, M.; BUONGIORNO, J. Space-time modeling of timber prices. **Journal of agricultural and resource economics**, p. 40-56, 2006.

ZWARCOWALD, C. L.; BASTOS, F. I.; CASTILHO, E. A. The dynamics of the AIDS epidemic in Brazil: a space-time analysis in the period 1987-1995. **The Brazilian Journal of Infectious Diseases**, São Paulo, v. 2, n. 4, p. 175-186, 1998.

APÊNDICE A- Código computacional desse trabalho desenvolvido em R.

```
#Lendo od dados
```

```
d1 = Dados_TB_PREVISÃO      # Dados originais sem tendencia
```

```
library(starma)
```

```
d= stcenter(d1)
```

```
result.bic = NULL
```

```
result.mape = NULL
```

```
#####Criando as matrizes
```

```
W#####
```

```
#####
```

```
#####
```

```
o0=diag(09)
```

```
#Contiguidade
```

```
o1 = t(matrix(c(
```

```
0.000000, 0.000000, 0.142857, 0.142857, 0.142857, 0.142857,
0.142857, 0.142857, 0.142857,
```

```
0.000000, 0.000000, 0.500000, 0.500000, 0.000000, 0.000000,
0.000000, 0.000000, 0.000000,
```

```
0.250000, 0.250000, 0.000000, 0.250000, 0.000000, 0.250000,
0.000000, 0.000000, 0.000000,
```

```
0.333333, 0.333333, 0.333333, 0.000000, 0.000000, 0.000000,
0.000000, 0.000000, 0.000000,
```

```
0.500000, 0.000000, 0.000000, 0.000000, 0.000000, 0.000000,
0.500000, 0.000000, 0.000000,
```

```
0.333333, 0.000000, 0.333333, 0.000000, 0.000000, 0.000000,
0.000000, 0.000000, 0.333333,
```

```
0.333333, 0.000000, 0.000000, 0.000000, 0.333333, 0.000000,
0.000000, 0.333333, 0.000000,
```

```
0.333333, 0.000000, 0.000000, 0.000000, 0.000000, 0.000000,
0.333333, 0.000000, 0.333333,
```

```

0.333333, 0.000000, 0.000000, 0.000000, 0.000000, 0.333333,
  0.000000, 0.333333, 0.000000)
,9,9))
w1 = list(order0 = o0,order1 = o1) #Contiguidade ordem 1

#0 PB- 1 IDH

o2 = t(matrix(c(
0.00000, 0.03931, 0.04441, 0.02262, 0.79937, 0.01903,
  0.03036, 0.02524, 0.01966,
0.05223, 0.00000, 0.45519, 0.07081, 0.04979, 0.04902,
  0.17702, 0.09371, 0.05223,
0.06361, 0.49069, 0.00000, 0.06605, 0.06026, 0.04771,
  0.13739, 0.08378, 0.05051,
0.03139, 0.07395, 0.06399, 0.00000, 0.03053, 0.16639,
  0.12325, 0.30252, 0.20798,
0.80549, 0.03776, 0.04239, 0.02217, 0.00000, 0.01873,
  0.02947, 0.02466, 0.01933,
0.01992, 0.03861, 0.03485, 0.12547, 0.01945, 0.00000,
  0.05339, 0.08095, 0.62736,
0.04786, 0.21006, 0.15124, 0.14004, 0.04611, 0.08045,
  0.00000, 0.23631, 0.08793,
0.03541, 0.09895, 0.08206, 0.30585, 0.03433, 0.10853,
  0.21027, 0.00000, 0.12460,
0.01951, 0.03902, 0.03501, 0.14878, 0.01904, 0.59511,
  0.05536, 0.08816, 0.00000)
,9,9))
w2 = list(order0 = o0,order1 = o2) # 0 Prev + 1IDH

```

```

#0,25 PB - 0,75 IDH

```

```

o3 =

```

```

t(matrix(c(

```

```

0.00000, 0.16878, 0.08933, 0.08136, 0.40067, 0.03920,
  0.11883, 0.06452, 0.03731,
0.13619, 0.00000, 0.15312, 0.12676, 0.09582, 0.04120,
  0.32398, 0.08428, 0.03865,
0.04487, 0.09533, 0.00000, 0.45839, 0.03669, 0.03509,
  0.18076, 0.11669, 0.03218,
0.04209, 0.08128, 0.47211, 0.00000, 0.03499, 0.03914,
  0.13352, 0.16123, 0.03565,
0.45678, 0.13539, 0.08327, 0.07710, 0.00000, 0.04071,
  0.10448, 0.06335, 0.03891,
0.03261, 0.04248, 0.05812, 0.06294, 0.02971, 0.00000,
  0.04867, 0.08311, 0.64236,
0.08007, 0.27057, 0.24249, 0.17391, 0.06176, 0.03942,
  0.00000, 0.09513, 0.03665,
0.05876, 0.09512, 0.21157, 0.28382, 0.05061, 0.09099,
  0.12857, 0.00000, 0.08056,
0.03196, 0.04103, 0.05488, 0.05903, 0.02924, 0.66149,
  0.04659, 0.07578, 0.00000)

```

,9,9))

w3 = list(order0 = o0,order1 = o3) #0,25 Prev + 0,75 IDH

#0,5 PB - 0,5 IDH

o4 = t(matrix(c(

```

0.00000, 0.18313, 0.01562, 0.07358, 0.03962, 0.00703,
  0.65925, 0.01561, 0.00617,
0.38880, 0.00000, 0.03056, 0.11144, 0.10734, 0.01437,
  0.30428, 0.03053, 0.01267,
0.00094, 0.00087, 0.00000, 0.00120, 0.00068, 0.00077,
  0.00097, 0.99397, 0.00061,
0.25478, 0.18176, 0.06868, 0.00000, 0.08917, 0.02691,
  0.28679, 0.06860, 0.02331,
0.21186, 0.27034, 0.05991, 0.13771, 0.00000, 0.03192,
  0.19985, 0.05987, 0.02854,
0.06354, 0.06120, 0.11551, 0.07026, 0.05397, 0.00000,
  0.06423, 0.11560, 0.45569,

```

```

0.68099, 0.14804, 0.01653, 0.08555, 0.03860, 0.00734,
0.00000, 0.01651, 0.00643,
0.00094, 0.00087, 0.99397, 0.00119, 0.00068, 0.00077,
0.00096, 0.00000, 0.00062,
0.06094, 0.05895, 0.10069, 0.06651, 0.05273, 0.49792,
0.06151, 0.10075, 0.00000)
,9,9))

```

```
w4 = list(order0 = o0,order1 = o4) #0,5 Prev + 0,5 IDH
```

```
#0,75 PB- 0,25 IDH
```

```

o5 = t(matrix(c(
0.00000, 0.14793, 0.10571, 0.15222, 0.18553, 0.04882,
0.14645, 0.17382, 0.03953,
0.00703, 0.00000, 0.00293, 0.24955, 0.03471, 0.00175,
0.69875, 0.00380, 0.00148,
0.13516, 0.07883, 0.00000, 0.07977, 0.08610, 0.11597,
0.07850, 0.34494, 0.08073,
0.01474, 0.50819, 0.00604, 0.00000, 0.08209, 0.00358,
0.37446, 0.00786, 0.00304,
0.06905, 0.27165, 0.02506, 0.31553, 0.00000, 0.01438,
0.25879, 0.03340, 0.01213,
0.08654, 0.06507, 0.16080, 0.06553, 0.06852, 0.00000,
0.06491, 0.12034, 0.36829,
0.00747, 0.74927, 0.00313, 0.19718, 0.03545, 0.00187,
0.00000, 0.00405, 0.00159,
0.19466, 0.08950, 0.30212, 0.09088, 0.10050, 0.07602,
0.08901, 0.00000, 0.05730,
0.08103, 0.06394, 0.12942, 0.06432, 0.06679, 0.42582,
0.06380, 0.10487, 0.00000)
,9,9))

```

```
w5 = list(order0 = o0,order1 = o5) ##0,75 Prev + 0,25 IDH
```

#1 PB- 0 IDH

```
o6 = t(matrix(c(
  0.00000, 0.07323, 0.09953, 0.06131, 0.13280, 0.04724,
    0.06701, 0.48358, 0.03530,
  0.04693, 0.00000, 0.02704, 0.24156, 0.10462, 0.01840,
    0.50542, 0.04076, 0.01527,
  0.18211, 0.07719, 0.00000, 0.06942, 0.10409, 0.16453,
    0.07327, 0.22930, 0.10009,
  0.04332, 0.26626, 0.02680, 0.00000, 0.08047, 0.01885,
    0.51003, 0.03844, 0.01583,
  0.17268, 0.21228, 0.07398, 0.14812, 0.00000, 0.04531,
    0.17587, 0.13548, 0.03627,
  0.10565, 0.06422, 0.20111, 0.05967, 0.07793, 0.00000,
    0.06196, 0.11709, 0.31235,
  0.03598, 0.42340, 0.02150, 0.38762, 0.07261, 0.01488,
    0.00000, 0.03160, 0.01241,
  0.49325, 0.06487, 0.12783, 0.05550, 0.10627, 0.05341,
    0.06003, 0.00000, 0.03884,
  0.09672, 0.06526, 0.14988, 0.06138, 0.07641, 0.38267,
    0.06335, 0.10434, 0.00000)
,9,9))
```

w6 = list(order0 = o0,order1 = o6) ##1 Prev + 0 IDH

#####

#Modelo 1- Contiguidade

stcor.test(d,w1,tlag=NULL)

stacf(d,w1)

stpacf(d,w1)

ar1 = matrix(0, 25, 2)

ar1[1,1] = 1

```
ar1[2,1] = 1
```

```
ar1[3,1] = 1
```

```
ar1[4,1] = 1
```

```
ar1[5,1] = 1
```

```
#ar1[6,1] = 1
```

```
ar1[7,1] = 1
```

```
ar1[13,1] = 1
```

```
ar1[6,2] = 1
```

```
ma1 = matrix(0,25,2)
```

```
ma1[12,1] = 1
```

```
ma1[6,1] = 1
```

```
#ma1[14,1] = 1
```

```
#ma1[15,1] = 1
```

```
#ma1[18,1] = 1
```

```
ma1[12,2] = 1
```

```
m1=starma(d,w1,ar1,ma1)
```

```
summary(m1)
```

```
stcor.test(m1$residuals,w1)
```

```
stacf(m1$residuals,w1)
```

```
stpacf(m1$residuals,w1)
```

```
print(capture.output(m1))
```

```
m1$bic
```

```
m1$loglik
result.bic = rbind(result.bic,c(m1$bic,m1$loglik))

res1 = m1$residuals

#####
#####

#Modelo 2- 0 Prev + 1 IDH
stcor.test(d,w2,tlag=NULL)
stacf(d,w2)
stpacf(d,w2)

ar2 = matrix(0, 25, 2)
ar2[1,1] = 1
ar2[2,1] = 1
ar2[17,1] = 1
ar2[19,1] = 1
#ar2[10,2] = 1
ar2[6,2] = 1

ma2 = matrix(0,25,2)
ma2[1,1] = 1
ma2[3,1] = 1
ma2[5,1] = 1
ma2[8,1] = 1
ma2[14,1] = 1
ma2[21,1] = 1
#ma2[2,2] = 1
#ma2[12,2] = 1
```

```

m2=starma(d,w2,ar2,ma2)
summary(m2)
stcor.test(m2$residuals,w2)

stacf(m2$residuals,w2)
stpacf(m2$residuals,w2)

print(capture.output(m2))
m2$bic
m2$loglik
result.bic = rbind(result.bic,c(m2$bic,m2$loglik))
#####
#####
#Modelo 3- 0,25 Prev + 0,75 IDH
stcor.test(d,w3,tlag=NULL)
stacf(d,w3)
stpacf(d,w3)

ar3 = matrix(0, 25, 2)
ar3[1,1] = 1
ar3[2,1] = 1
ar3[17,1] = 1
ar3[19,1] = 1
ar3[6,2] = 1
ar3[3,2] = 1

ma3 = matrix(0,25,2)

```

```
ma3[1,1] = 1
ma3[3,1] = 1
ma3[5,1] = 1
ma3[8,1] = 1
ma3[14,1] = 1
ma3[21,1] = 1
ma3[4,2] = 1

m3=starma(d,w3,ar3,ma3)
summary(m3)
stcor.test(m3$residuals,w3)

stacf(m3$residuals,w3)
stpacf(m3$residuals,w3)

print(capture.output(m3))
m3$bic
m3$loglik
result.bic = rbind(result.bic,c(m3$bic,m3$loglik))
#####
#####
#Modelo 4- 0,5 Prev + 0,5 IDH
stcor.test(d,w4,tlag=NULL)
stacf(d,w4)
stpacf(d,w4)

ar4 = matrix(0, 25, 2)
ar4[1,1] = 1
```

```
ar4[2,1] = 1
```

```
ar4[17,1] = 1
```

```
ar4[19,1] = 1
```

```
#ar2[10,2] = 1
```

```
ar4[12,2] = 1
```

```
ma4 = matrix(0,25,2)
```

```
ma4[1,1] = 1
```

```
ma4[3,1] = 1
```

```
ma4[5,1] = 1
```

```
ma4[8,1] = 1
```

```
ma4[14,1] = 1
```

```
ma4[21,1] = 1
```

```
m4=starma(d,w4,ar4,ma4)
```

```
summary(m4)
```

```
stcor.test(m4$residuals,w4)
```

```
stacf(m4$residuals,w4)
```

```
stpacf(m4$residuals,w4)
```

```
print(capture.output(m4))
```

```
m4$bic
```

```
m4$loglik
```

```
result.bic = rbind(result.bic,c(m4$bic,m4$loglik))
```

```
#####
```

```
#####
```

```
#Modelo 5- 0,75 Prev + 0,25 IDH
```

```
stcor.test(d,w5,tlag=NULL)
```

```
stacf(d,w5)
```

```
stpacf(d,w5)
```

```
ar5 = matrix(0, 25, 2)
```

```
ar5[1,1] = 1
```

```
ar5[2,1] = 1
```

```
ar5[17,1] = 1
```

```
ar5[19,1] = 1
```

```
ar5[10,2] = 1
```

```
#ar5[19,2] = 1
```

```
ma5 = matrix(0,25,2)
```

```
ma5[1,1] = 1
```

```
ma5[3,1] = 1
```

```
ma5[5,1] = 1
```

```
ma5[8,1] = 1
```

```
ma5[14,1] = 1
```

```
ma5[21,1] = 1
```

```
#ma5[2,2] = 1
```

```
#ma5[12,2] = 1
```

```
m5=starma(d,w5,ar5,ma5)
```

```
summary(m5)
```

```
stcor.test(m5$residuals,w5)
```

```
stacf(m5$residuals,w5)
```

```
stpacf(m5$residuals,w5)
```

```

print(capture.output(m5))
m5$bic
m5$loglik
result.bic = rbind(result.bic,c(m5$bic,m5$loglik))
#####
#####
#Modelo 6- 1 Prev + 0 IDH
stcor.test(d,w6,tlag=NULL)
stacf(d,w6)
stpacf(d,w6)

ar6 = matrix(0, 25, 2)
ar6[1,1] = 1
ar6[2,1] = 1
ar6[17,1] = 1
ar6[19,1] = 1
#ar2[10,2] = 1
ar6[6,2] = 1

ma6 = matrix(0,25,2)
ma6[1,1] = 1
ma6[3,1] = 1
ma6[5,1] = 1
ma6[8,1] = 1
ma6[14,1] = 1
ma6[21,1] = 1

m6=starma(d,w6,ar6,ma6)

```

```

summary(m6)
stcor.test(m6$residuals,w6)

stacf(m6$residuals,w6)
stpacf(m6$residuals,w6)

print(capture.output(m6))

m6$bic
m6$loglik
result.bic = rbind(result.bic,c(m6$bic,m6$loglik))

result.bic

#####

#ESTIMAÇÃO

z246 = t(matrix(c(-0.073446682,    -0.4079631,   -0.360384654, -0.328919963, -
1.50139899,
                1.015206549,    1.294273044, -0.511671554, -0.865481698),1,9))
z247 = t(matrix(c(0.312537356,    0.2701843,    1.256510251, 0.326770613,
0.499033430,
                1.523347162,    -0.001074675, 0.509522204, 0.863332348),1,9))
z248 = t(matrix(c(-0.290562704,    -0.8149633,   -1.079004612, 0.326770613, -
0.001074675,
                -1.694876716,    1.294273044, 0.509522204, -2.162092232),1,9))
z249 = t(matrix(c(0.216041347,    0.4058138,    0.268407809, -0.001074675, -
0.501182780,
                0.337685733,    -0.864639821, -0.001074675, -0.001074675),1,9))
z250 = t(matrix(c(-0.338810709,    -0.1367042,   -0.180729664, -1.312455825, -
0.501182780,

```

```

-1.525496512,    0.430707898, -1.532865311, 0.863332348),1,9))
z251 = t(matrix(c(0.336661359,    -0.2723336,   0.358235304, 1.965997051,
0.499033430,
0.507065937,    -1.296422394, 1.020119082, 172.773937),1,9))
z.obs = cbind(z246, z247, z248, z249, z250, z251)

```

```
z245 = matrix(as.numeric(d[245,1:9]))
```

```
z244 = matrix(as.numeric(d[244,1:9]))
```

```
z243 = matrix(as.numeric(d[243,1:9]))
```

```
z242 = matrix(as.numeric(d[242,1:9]))
```

```
z241 = matrix(as.numeric(d[241,1:9]))
```

```
z240 = matrix(as.numeric(d[240,1:9]))
```

```
z239 = matrix(as.numeric(d[239,1:9]))
```

```
z238 = matrix(as.numeric(d[238,1:9]))
```

```
z237 = matrix(as.numeric(d[237,1:9]))
```

```
z236 = matrix(as.numeric(d[236,1:9]))
```

```
z235 = matrix(as.numeric(d[235,1:9]))
```

```
z234 = matrix(as.numeric(d[234,1:9]))
```

```
z233 = matrix(as.numeric(d[233,1:9]))
```

```
z232 = matrix(as.numeric(d[232,1:9]))
```

```
z231 = matrix(as.numeric(d[231,1:9]))
```

```
z230 = matrix(as.numeric(d[230,1:9]))
```

```
z229 = matrix(as.numeric(d[229,1:9]))
```

```
z228 = matrix(as.numeric(d[228,1:9]))
```

```
z227 = matrix(as.numeric(d[227,1:9]))
```

```
z226 = matrix(as.numeric(d[226,1:9]))
```

```
z225 = matrix(as.numeric(d[225,1:9]))
```

```
z224 = matrix(as.numeric(d[224,1:9]))
```

```
z223 = matrix(as.numeric(d[223,1:9]))
```

```
z222 = matrix(as.numeric(d[222,1:9]))
```

z221 = matrix(as.numeric(d[221,1:9]))

z220 = matrix(as.numeric(d[220,1:9]))

z219 = matrix(as.numeric(d[219,1:9]))

#####

###PREVISÃO MODELO 1

res1 =m1\$residuals

r1.234 = matrix(as.numeric(res1[234,1:9]))

r1.240 = matrix(as.numeric(res1[240,1:9]))

r1.235 = matrix(as.numeric(res1[235,1:9]))

r1.241 = matrix(as.numeric(res1[241,1:9]))

r1.236 = matrix(as.numeric(res1[236,1:9]))

r1.242 = matrix(as.numeric(res1[242,1:9]))

r1.237 = matrix(as.numeric(res1[237,1:9]))

r1.243 = matrix(as.numeric(res1[243,1:9]))

r1.238 = matrix(as.numeric(res1[238,1:9]))

r1.244 = matrix(as.numeric(res1[244,1:9]))

r1.239 = matrix(as.numeric(res1[239,1:9]))

r1.245 = matrix(as.numeric(res1[245,1:9]))

p1.246= -0.882892*z245-0.773638*z244-0.713891*z243-0.617834*z242-0.594380*z241-
0.102638*(w1\$order1%*z240)-0.050737*z239+0.112902*z233-
0.531313*r1.240+0.085113*r1.234+0.139749*(w1\$order1%*r1.234)

p1.247= -0.882892*p1.246-0.773638*z245-0.713891*z244 -0.617834*z243-0.594380*z242-
0.102638*(w1\$order1%*z241)-0.050737*z240+0.112902*z234-
0.531313*r1.241+0.085113*r1.235+0.139749*(w1\$order1%*r1.235)

p1.248= -0.882892*p1.247-0.773638*p1.246-0.713891*z245 -0.617834*z244-
0.594380*z243-0.102638*(w1\$order1%*z242)-0.050737*z241+0.112902*z235-
0.531313*r1.242+0.085113*r1.236+0.139749*(w1\$order1%*r1.236)

```
p1.249= -0.882892*p1.248-0.773638*p1.247-0.713891*p1.246 -0.617834*z245-
0.594380*z244-0.102638*(w1$order1%*z243)-0.050737*z242+0.112902*z236-
0.531313*r1.243+0.085113*r1.237+0.139749*(w1$order1%*r1.237)
```

```
p1.250= -0.882892*p1.249-0.773638*p1.248-0.713891*p1.247 -0.617834*p1.246-
0.594380*z245-0.102638*(w1$order1%*z244)-0.050737*z243+0.112902*z237-
0.531313*r1.244+0.085113*r1.238+0.139749*(w1$order1%*r1.238)
```

```
p1.251= -0.882892*p1.250-0.773638*p1.249-0.713891*p1.248 -0.617834*p1.247-
0.594380*p1.246-0.102638*(w1$order1%*z245)-0.050737*z244+0.112902*z238-
0.531313*r1.245+0.085113*r1.239+0.139749*(w1$order1%*r1.239)
```

```
p.1 = t(cbind(p1.246, p1.247, p1.248, p1.249, p1.250, p1.251))
```

```
library(Metrics)
```

```
# Calculando MAPE
```

```
mape1 <- mape(t(z.obs),p.1)
```

```
print(paste0("MAPE: ", round(mape1, 2), "%"))
```

```
#####
```

```
###PREVISÃO MODELO 2
```

```
res2 =m2$residuals
```

```
r2.225 = matrix(as.numeric(res2[225,1:9]))
```

```
r2.226 = matrix(as.numeric(res2[226,1:9]))
```

```
r2.227 = matrix(as.numeric(res2[227,1:9]))
```

```
r2.228 = matrix(as.numeric(res2[228,1:9]))
```

```
r2.229 = matrix(as.numeric(res2[229,1:9]))
```

```
r2.230 = matrix(as.numeric(res2[230,1:9]))
```

```
r2.231 = matrix(as.numeric(res2[231,1:9]))
```

$$r2.232 = \text{matrix}(\text{as.numeric}(\text{res2}[232,1:9]))$$

$$r2.233 = \text{matrix}(\text{as.numeric}(\text{res2}[233,1:9]))$$

$$r2.234 = \text{matrix}(\text{as.numeric}(\text{res2}[234,1:9]))$$

$$r2.235 = \text{matrix}(\text{as.numeric}(\text{res2}[235,1:9]))$$

$$r2.236 = \text{matrix}(\text{as.numeric}(\text{res2}[236,1:9]))$$

$$r2.237 = \text{matrix}(\text{as.numeric}(\text{res2}[237,1:9]))$$

$$r2.238 = \text{matrix}(\text{as.numeric}(\text{res2}[238,1:9]))$$

$$r2.239 = \text{matrix}(\text{as.numeric}(\text{res2}[239,1:9]))$$

$$r2.240 = \text{matrix}(\text{as.numeric}(\text{res2}[240,1:9]))$$

$$r2.241 = \text{matrix}(\text{as.numeric}(\text{res2}[241,1:9]))$$

$$r2.243 = \text{matrix}(\text{as.numeric}(\text{res2}[243,1:9]))$$

$$r2.242 = \text{matrix}(\text{as.numeric}(\text{res2}[242,1:9]))$$

$$r2.244 = \text{matrix}(\text{as.numeric}(\text{res2}[244,1:9]))$$

$$r2.245 = \text{matrix}(\text{as.numeric}(\text{res2}[245,1:9]))$$

$$\begin{aligned} p2.246 = & -0.474527*z^{245}-0.398836*z^{244}- \\ & 0.055454*(w2\$order1\%*\%z^{240})+0.060673*z^{229}+0.053959*z^{227}-0.411553*r2.245- \\ & 0.382648*r2.243 -0.085927*r2.241+0.111602*r2.238-0.084277*r2.234+0.072946*r2.225 \end{aligned}$$

$$r2.246 = z^{246} - p2.246$$

$$\begin{aligned} p2.247 = & -0.474527*p2.246-0.398836*z^{245}- \\ & 0.055454*(w2\$order1\%*\%z^{241})+0.060673*z^{230}+0.053959*z^{222}-0.411553*r2.246- \\ & 0.382648*r2.244 -0.085927*r2.242+0.111602*r2.239-0.084277*r2.235+0.072946*r2.226 \end{aligned}$$

$$r2.247 = z^{247} - p2.247$$

$$\begin{aligned} p2.248 = & -0.474527*p2.247-0.398836*p2.246- \\ & 0.055454*(w2\$order1\%*\%z^{242})+0.060673*z^{231}+0.053959*z^{223}-0.411553*r2.247- \\ & 0.382648*r2.245 -0.085927*r2.243+0.111602*r2.240-0.084277*r2.236+0.072946*r2.227 \end{aligned}$$

$$r2.248 = z^{248} - p2.248$$

```
p2.249 = -0.474527*p2.248-0.398836*p2.247-
0.055454*(w2$order1%*z243)+0.060673*z232+0.053959*z224-0.411553*r2.248-
0.382648*r2.246 -0.085927*r2.244+0.111602*r2.241-0.084277*r2.237+0.072946*r2.228
```

```
r2.249 = z249 - p2.249
```

```
p2.250 = -0.474527*p2.249-0.398836*p2.248-
0.055454*(w2$order1%*z244)+0.060673*z233+0.053959*z225-0.411553*r2.249-
0.382648*r2.247 -0.085927*r2.245+0.111602*r2.242-0.084277*r2.238+0.072946*r2.229
```

```
r2.250 = z250 - p2.250
```

```
p2.251 = -0.474527*p2.250-0.398836*p2.249-
0.055454*(w2$order1%*z245)+0.060673*z234+0.053959*z226-0.411553*r2.250-
0.382648*r2.248 -0.085927*r2.246+0.111602*r2.243-0.084277*r2.239+0.072946*r2.230
```

```
r2.251 = z251 - p2.251
```

```
p.2 = t(cbind(p2.246, p2.247, p2.248, p2.249, p2.250, p2.251))
```

```
mape2 <- mape(t(z.obs),p.2)
```

```
print(paste0("MAPE: ", round(mape2, 2), "%"))
```

```
#####
```

```
###PREVISÃO MODELO 3
```

```
res3 =m3$residuals
```

```
r3.225 = matrix(as.numeric(res3[225,1:9]))
```

```
r3.226 = matrix(as.numeric(res3[226,1:9]))
```

```
r3.227 = matrix(as.numeric(res3[227,1:9]))
```

```
r3.228 = matrix(as.numeric(res3[228,1:9]))
```

```
r3.229 = matrix(as.numeric(res3[229,1:9]))
```

```
r3.230 = matrix(as.numeric(res3[230,1:9]))
```

```
r3.231 = matrix(as.numeric(res3[231,1:9]))
```

```
r3.232 = matrix(as.numeric(res3[232,1:9]))
```

```
r3.233 = matrix(as.numeric(res3[233,1:9]))
```

$$r3.234 = \text{matrix}(\text{as.numeric}(\text{res3}[234,1:9]))$$

$$r3.235 = \text{matrix}(\text{as.numeric}(\text{res3}[235,1:9]))$$

$$r3.236 = \text{matrix}(\text{as.numeric}(\text{res3}[236,1:9]))$$

$$r3.237 = \text{matrix}(\text{as.numeric}(\text{res3}[237,1:9]))$$

$$r3.238 = \text{matrix}(\text{as.numeric}(\text{res3}[238,1:9]))$$

$$r3.239 = \text{matrix}(\text{as.numeric}(\text{res3}[239,1:9]))$$

$$r3.240 = \text{matrix}(\text{as.numeric}(\text{res3}[240,1:9]))$$

$$r3.241 = \text{matrix}(\text{as.numeric}(\text{res3}[241,1:9]))$$

$$r3.243 = \text{matrix}(\text{as.numeric}(\text{res3}[243,1:9]))$$

$$r3.242 = \text{matrix}(\text{as.numeric}(\text{res3}[242,1:9]))$$

$$r3.244 = \text{matrix}(\text{as.numeric}(\text{res3}[244,1:9]))$$

$$r3.245 = \text{matrix}(\text{as.numeric}(\text{res3}[245,1:9]))$$

$$\begin{aligned} p3.246 = & -0.5002350*z245-0.419225*z244+0.095248*(w3\$order1\%*\%z243)- \\ & 0.100074*(w3\$order1\%*\%z240)+0.059229*z229 +0.050781*z227-0.384451*r3.245- \\ & 0.406950*r3.243 +0.155979*(w3\$order1\%*\%r3.242) -083533*r3.241 +0.107680*r3.238- \\ & 0.0079597*r3.232 +0.072888*r3.225 \end{aligned}$$

$$r3.246 = z246 - p3.246$$

$$\begin{aligned} p3.247 = & -0.5002350*p3.246-0.419225*z245+0.095248*(w3\$order1\%*\%z244)- \\ & 0.100074*(w3\$order1\%*\%z241)+0.059229*z230 +0.050781*z228-0.384451*r3.246- \\ & 0.406950*r3.244 +0.155979*(w3\$order1\%*\%r3.243) -083533*r3.242 +0.107680*r3.239- \\ & 0.0079597*r3.233 +0.072888*r3.226 \end{aligned}$$

$$r3.247 = z247 - p3.247$$

$$\begin{aligned} p3.248 = & -0.5002350*p3.247-0.419225*p3.246 +0.095248*(w3\$order1\%*\%z245)- \\ & 0.100074*(w3\$order1\%*\%z242) +0.059229*z231 +0.050781*z229-0.384451*r3.247- \\ & 0.406950*r3.245 +0.155979*(w3\$order1\%*\%r3.244) -083533*r3.243 +0.107680*r3.240- \\ & 0.0079597*r3.234 +0.072888*r3.227 \end{aligned}$$

$$r3.248 = z248 - p3.248$$

$$\begin{aligned} p3.249 = & -0.5002350*p3.248-0.419225*p3.247 +0.095248*(w3\$order1\%*\%p3.246)- \\ & 0.100074*(w3\$order1\%*\%z243) +0.059229*z232 +0.050781*z230-0.384451*r3.248- \end{aligned}$$

```
0.406950*r3.246 +0.155979*(w3$order1%*%r3.245) -083533*r3.243 +0.107680*r3.240-
0.0079597*r3.234 +0.072888*r3.227
```

```
r3.249 = z249 - p3.249
```

```
p3.250 = -0.5002350*p3.249-0.419225*p3.248 +0.095248*(w3$order1%*%p3.247)-
0.100074*(w3$order1%*%z244) +0.059229*z233 +0.050781*z231-0.384451*r3.249-
0.406950*r3.247 +0.155979*(w3$order1%*%r3.246) -083533*r3.244 +0.107680*r3.241-
0.0079597*r3.235 +0.072888*r3.228
```

```
r3.250 = z250 - p3.250
```

```
p3.251 = -0.5002350*p3.250-0.419225*p3.249 +0.095248*(w3$order1%*%p3.248)-
0.100074*(w3$order1%*%z245) +0.059229*z234 +0.050781*z232-0.384451*r3.250-
0.406950*r3.248 +0.155979*(w3$order1%*%r3.247) -083533*r3.245 +0.107680*r3.242-
0.0079597*r3.236 +0.072888*r3.229
```

```
r3.251 = z251 - p3.251
```

```
p.3 = t(cbind(p3.246, p3.247, p3.248, p3.249, p3.250, p3.251))
```

```
mape3 <- mape(t(z.obs),p.3)
```

```
print(paste0("MAPE: ", round(mape3/1000, 2), "%"))
```

```
#####
```

```
###PREVISÃO MODELO 4
```

```
res4 =m4$residuals
```

```
r4.225 = matrix(as.numeric(res4[225,1:9]))
```

```
r4.226 = matrix(as.numeric(res4[226,1:9]))
```

```
r4.227 = matrix(as.numeric(res4[227,1:9]))
```

```
r4.228 = matrix(as.numeric(res4[228,1:9]))
```

```
r4.229 = matrix(as.numeric(res4[229,1:9]))
```

```
r4.230 = matrix(as.numeric(res4[230,1:9]))
```

```
r4.231 = matrix(as.numeric(res4[231,1:9]))
```

$$r4.232 = \text{matrix}(\text{as.numeric}(\text{res4}[232,1:9]))$$

$$r4.233 = \text{matrix}(\text{as.numeric}(\text{res4}[233,1:9]))$$

$$r4.234 = \text{matrix}(\text{as.numeric}(\text{res4}[234,1:9]))$$

$$r4.235 = \text{matrix}(\text{as.numeric}(\text{res4}[235,1:9]))$$

$$r4.236 = \text{matrix}(\text{as.numeric}(\text{res4}[236,1:9]))$$

$$r4.237 = \text{matrix}(\text{as.numeric}(\text{res4}[237,1:9]))$$

$$r4.238 = \text{matrix}(\text{as.numeric}(\text{res4}[238,1:9]))$$

$$r4.239 = \text{matrix}(\text{as.numeric}(\text{res4}[239,1:9]))$$

$$r4.240 = \text{matrix}(\text{as.numeric}(\text{res4}[240,1:9]))$$

$$r4.241 = \text{matrix}(\text{as.numeric}(\text{res4}[241,1:9]))$$

$$r4.243 = \text{matrix}(\text{as.numeric}(\text{res4}[243,1:9]))$$

$$r4.242 = \text{matrix}(\text{as.numeric}(\text{res4}[242,1:9]))$$

$$r4.244 = \text{matrix}(\text{as.numeric}(\text{res4}[244,1:9]))$$

$$r4.245 = \text{matrix}(\text{as.numeric}(\text{res4}[245,1:9]))$$

$$\begin{aligned} p4.246 = & -0.472258 * z^{245} - 0.397105 * z^{244} + 0.067630 * (w4\$order1 \% \% z^{234}) \\ & + 0.063593 * z^{229} + 0.053298 * z^{227} - 0.412886 * r4.245 - 0.379336 * r4.243 - 0.081325 * r4.241 \\ & + 0.114197 * r4.238 - 0.084099 * r4.232 + 0.072200 * r4.225 \end{aligned}$$

$$r4.246 = z^{246} - p4.246$$

$$\begin{aligned} p4.247 = & -0.472258 * p4.246 - 0.397105 * z^{245} + 0.067630 * (w4\$order1 \% \% z^{235}) \\ & + 0.063593 * z^{230} + 0.053298 * z^{228} - 0.412886 * r4.246 - 0.379336 * r4.244 - 0.081325 * r4.242 \\ & + 0.114197 * r4.239 - 0.084099 * r4.233 + 0.072200 * r4.226 \end{aligned}$$

$$r4.247 = z^{247} - p4.247$$

$$\begin{aligned} p4.248 = & -0.472258 * p4.247 - 0.397105 * p4.246 + 0.067630 * (w4\$order1 \% \% z^{236}) \\ & + 0.063593 * z^{231} + 0.053298 * z^{229} - 0.412886 * r4.247 - 0.379336 * r4.245 - 0.081325 * r4.243 \\ & + 0.114197 * r4.240 - 0.084099 * r4.234 + 0.072200 * r4.227 \end{aligned}$$

$$r4.248 = z^{248} - p4.248$$

$$\begin{aligned} p4.249 = & -0.472258 * p4.248 - 0.397105 * p4.247 + 0.067630 * (w4\$order1 \% \% z^{237}) \\ & + 0.063593 * z^{232} + 0.053298 * z^{230} - 0.412886 * r4.248 - 0.379336 * r4.246 - 0.081325 * r4.244 \\ & + 0.114197 * r4.241 - 0.084099 * r4.235 + 0.072200 * r4.228 \end{aligned}$$

$r_{4.249} = z_{249} - p_{4.249}$

$p_{4.250} = -0.472258 * p_{4.249} - 0.397105 * p_{4.248} + 0.067630 * (w_4 \text{order1} \% \% z_{238})$
 $+ 0.063593 * z_{233} + 0.053298 * z_{231} - 0.412886 * r_{4.249} - 0.379336 * r_{4.247} - 0.081325 * r_{4.245}$
 $+ 0.114197 * r_{4.242} - 0.084099 * r_{4.236} + 0.072200 * r_{4.229}$

$r_{4.250} = z_{250} - p_{4.250}$

$p_{4.251} = -0.472258 * p_{4.250} - 0.397105 * p_{4.249} + 0.067630 * (w_4 \text{order1} \% \% z_{239})$
 $+ 0.063593 * z_{234} + 0.053298 * z_{232} - 0.412886 * r_{4.250} - 0.379336 * r_{4.248} - 0.081325 * r_{4.246}$
 $+ 0.114197 * r_{4.243} - 0.084099 * r_{4.237} + 0.072200 * r_{4.230}$

$r_{4.251} = z_{251} - p_{4.251}$

$p_4 = t(\text{cbind}(p_{4.246}, p_{4.247}, p_{4.248}, p_{4.249}, p_{4.250}, p_{4.251}))$

`mape4 <- mape(t(z.obs), p.4)`

`print(paste0("MAPE: ", round(mape4, 2), "%"))`

#####

###PREVISÃO MODELO 5

`res5 = m5$residuals`

`r5.225 = matrix(as.numeric(res5[225,1:9]))`

`r5.226 = matrix(as.numeric(res5[226,1:9]))`

`r5.227 = matrix(as.numeric(res5[227,1:9]))`

`r5.228 = matrix(as.numeric(res5[228,1:9]))`

`r5.229 = matrix(as.numeric(res5[229,1:9]))`

`r5.230 = matrix(as.numeric(res5[230,1:9]))`

`r5.231 = matrix(as.numeric(res5[231,1:9]))`

`r5.232 = matrix(as.numeric(res5[232,1:9]))`

`r5.233 = matrix(as.numeric(res5[233,1:9]))`

`r5.234 = matrix(as.numeric(res5[234,1:9]))`

`r5.235 = matrix(as.numeric(res5[235,1:9]))`

$$r5.236 = \text{matrix}(\text{as.numeric}(\text{res5}[236,1:9]))$$

$$r5.237 = \text{matrix}(\text{as.numeric}(\text{res5}[237,1:9]))$$

$$r5.238 = \text{matrix}(\text{as.numeric}(\text{res5}[238,1:9]))$$

$$r5.239 = \text{matrix}(\text{as.numeric}(\text{res5}[239,1:9]))$$

$$r5.240 = \text{matrix}(\text{as.numeric}(\text{res5}[240,1:9]))$$

$$r5.241 = \text{matrix}(\text{as.numeric}(\text{res5}[241,1:9]))$$

$$r5.243 = \text{matrix}(\text{as.numeric}(\text{res5}[243,1:9]))$$

$$r5.242 = \text{matrix}(\text{as.numeric}(\text{res5}[242,1:9]))$$

$$r5.244 = \text{matrix}(\text{as.numeric}(\text{res5}[244,1:9]))$$

$$r5.245 = \text{matrix}(\text{as.numeric}(\text{res5}[245,1:9]))$$

$$\begin{aligned} p5.246 = & -0.484493*z245 - 0.405465*z244 + 0.0744258*(w5\$order1\%*\%z236) \\ & + 0.061747*z229 + 0.053297*z227 - 0.402303*r5.245 - 0.394470*r5.243 - 0.092133*r5.241 \\ & + 0.107737*r5.238 - 0.086845*r5.232 + 0.075687*r5.225 \end{aligned}$$

$$r5.246 = z246 - p5.246$$

$$\begin{aligned} p5.247 = & -0.484493*p5.246 - 0.405465*z245 + 0.0744258*(w5\$order1\%*\%z237) \\ & + 0.061747*z230 + 0.053297*z228 - 0.402303*r5.246 - 0.394470*r5.244 - 0.092133*r5.242 \\ & + 0.107737*r5.239 - 0.086845*r5.233 + 0.075687*r5.226 \end{aligned}$$

$$r5.247 = z247 - p5.247$$

$$\begin{aligned} p5.248 = & -0.484493*p5.247 - 0.405465*p5.246 + 0.0744258*(w5\$order1\%*\%z238) \\ & + 0.061747*z231 + 0.053297*z229 - 0.402303*r5.247 - 0.394470*r5.245 - 0.092133*r5.243 \\ & + 0.107737*r5.240 - 0.086845*r5.234 + 0.075687*r5.227 \end{aligned}$$

$$r5.248 = z248 - p5.248$$

$$\begin{aligned} p5.249 = & -0.484493*p5.248 - 0.405465*p5.247 + 0.0744258*(w5\$order1\%*\%z239) \\ & + 0.061747*z232 + 0.053297*z230 - 0.402303*r5.248 - 0.394470*r5.246 - 0.092133*r5.244 \\ & + 0.107737*r5.241 - 0.086845*r5.235 + 0.075687*r5.228 \end{aligned}$$

$$r5.249 = z249 - p5.249$$

$$\begin{aligned} p5.250 = & -0.484493*p5.249 - 0.405465*p5.248 + 0.0744258*(w5\$order1\%*\%z240) \\ & + 0.061747*z233 + 0.053297*z231 - 0.402303*r5.249 - 0.394470*r5.247 - 0.092133*r5.245 \\ & + 0.107737*r5.242 - 0.086845*r5.236 + 0.075687*r5.229 \end{aligned}$$

```
r5.250 = z250 - p5.250
```

```
p5.251 = -0.484493*p5.250 - 0.405465*p5.249 + 0.0744258*(w5$order1%*z241)
+ 0.061747*z234 + 0.053297*z232 - 0.402303*r5.250 - 0.394470*r5.248 - 0.092133*r5.246
+ 0.107737*r5.243 - 0.086845*r5.237 + 0.075687*r5.230
```

```
r5.251 = z251 - p5.251
```

```
p.5 = t(cbind(p5.246, p5.247, p5.248, p5.249, p5.250, p5.251))
```

```
mape5 <- mape(t(z.obs), p.5)
```

```
print(paste0("MAPE: ", round(mape5, 2), "%"))
```

```
#####
```

```
###PREVISÃO MODELO 6
```

```
res6 = m6$residuals
```

```
r6.225 = matrix(as.numeric(res5[225,1:9]))
```

```
r6.226 = matrix(as.numeric(res5[226,1:9]))
```

```
r6.227 = matrix(as.numeric(res5[227,1:9]))
```

```
r6.228 = matrix(as.numeric(res5[228,1:9]))
```

```
r6.229 = matrix(as.numeric(res5[229,1:9]))
```

```
r6.230 = matrix(as.numeric(res5[230,1:9]))
```

```
r6.231 = matrix(as.numeric(res5[231,1:9]))
```

```
r6.232 = matrix(as.numeric(res5[232,1:9]))
```

```
r6.233 = matrix(as.numeric(res5[233,1:9]))
```

```
r6.234 = matrix(as.numeric(res5[234,1:9]))
```

```
r6.235 = matrix(as.numeric(res5[235,1:9]))
```

```
r6.236 = matrix(as.numeric(res5[236,1:9]))
```

```
r6.237 = matrix(as.numeric(res5[237,1:9]))
```

```
r6.238 = matrix(as.numeric(res5[238,1:9]))
```

```
r6.239 = matrix(as.numeric(res5[239,1:9]))
```

$$r6.240 = \text{matrix}(\text{as.numeric}(\text{res5}[240,1:9]))$$

$$r6.241 = \text{matrix}(\text{as.numeric}(\text{res5}[241,1:9]))$$

$$r6.243 = \text{matrix}(\text{as.numeric}(\text{res5}[243,1:9]))$$

$$r6.242 = \text{matrix}(\text{as.numeric}(\text{res5}[242,1:9]))$$

$$r6.244 = \text{matrix}(\text{as.numeric}(\text{res5}[244,1:9]))$$

$$r6.245 = \text{matrix}(\text{as.numeric}(\text{res5}[245,1:9]))$$

$$p6.246 = -0.466437 * z245 - 0.393419 * z244 - 0.081542 * (w5\$order1 \% \% z240) + 0.060969 * z229 + 0.052716 * z227 - 0.417912 * r6.245 - 0.376180 * r6.243 - 0.083955 * r6.241 + 0.112249 * r6.238 - 0.084057 * r6.232 + 0.070676 * r6.225$$

$$r6.246 = z246 - p6.246$$

$$p6.247 = -0.466437 * p6.246 - 0.393419 * z245 - 0.081542 * (w5\$order1 \% \% z241) + 0.060969 * z230 + 0.052716 * z228 - 0.417912 * r6.246 - 0.376180 * r6.244 - 0.083955 * r6.242 + 0.112249 * r6.239 - 0.084057 * r6.233 + 0.070676 * r6.226$$

$$r6.247 = z247 - p6.247$$

$$p6.248 = -0.466437 * p6.247 - 0.393419 * p6.246 - 0.081542 * (w5\$order1 \% \% z242) + 0.060969 * z231 + 0.052716 * z229 - 0.417912 * r6.247 - 0.376180 * r6.245 - 0.083955 * r6.243 + 0.112249 * r6.240 - 0.084057 * r6.234 + 0.070676 * r6.227$$

$$r6.248 = z248 - p6.248$$

$$p6.249 = -0.466437 * p6.248 - 0.393419 * p6.247 - 0.081542 * (w5\$order1 \% \% z243) + 0.060969 * z232 + 0.052716 * z230 - 0.417912 * r6.248 - 0.376180 * r6.246 - 0.083955 * r6.244 + 0.112249 * r6.241 - 0.084057 * r6.235 + 0.070676 * r6.228$$

$$r6.249 = z249 - p6.249$$

$$p6.250 = -0.466437 * p6.249 - 0.393419 * p6.248 - 0.081542 * (w5\$order1 \% \% z244) + 0.060969 * z233 + 0.052716 * z231 - 0.417912 * r6.249 - 0.376180 * r6.247 - 0.083955 * r6.245 + 0.112249 * r6.242 - 0.084057 * r6.236 + 0.070676 * r6.229$$

$$r6.250 = z250 - p6.250$$

```
p6.251= -0.466437*p6.250-0.393419*p6.249 -0.081542*(w5$order1%*% z245)
+0.060969*z234 +0.052716*z232 -0.417912*r6.250 -0.376180*r6.248 -0.083955*r6.246
+0.112249*r6.243 -0.084057*r6.237 +0.070676*r6.230
```

```
r6.251 = z251 - p6.251
```

```
p.6 = t(cbind(p6.246, p6.247, p6.248, p6.249, p6.250, p6.251))
```

```
mape6 <- mape(t(z.obs),p.6)
```

```
print(paste0("MAPE: ", round(mape6, 2), "%"))
```

```
#####
```

```
#####
```