



MATHEUS FERES FREITAS

**MATRIZES SOCIOECONÔMICAS NO AJUSTE DE
MODELOS STARMA APLICADOS A DADOS
EPIDEMIOLÓGICOS.**

LAVRAS- MG

2019

MATHEUS FERES FREITAS

**MATRIZES SOCIOECONÔMICAS NO AJUSTE DE MODELOS STARMA
APLICADOS A DADOS EPIDEMIOLÓGICOS.**

Dissertação apresentada à Universidade Federal de Lavras, como parte das exigências do curso do programa de Pós-Graduação em Estatística e Experimentação Agropecuária, área de concentração em Estatística Aplicada, para a obtenção do título de mestre.

Prof. Dra. Thelma Sáfadi

Orientadora

LAVRAS- MG

2019

**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 socioeconômicas no ajuste de modelos STARMA
aplicados a dados epidemiológicos. / Matheus Feres Freitas. - 2019.
82 p. : il.

Orientador(a): Thelma Sáfadi.

Dissertação (mestrado acadêmico) - Universidade Federal de
Lavras, 2019.

Bibliografia.

1. Matriz de vizinhança. 2. STARMA. 3. Tuberculose. I.
Sáfadi, Thelma. II. Título.

MATHEUS FERES FREITAS

**MATRIZES SOCIOECONÔMICAS NO AJUSTE DE MODELOS STARMA
APLICADOS A DADOS EPIDEMIOLÓGICOS.**

**SOCIOECONOMIC MATRICES IN ADJUSTMENT OF STARMA MODELS
APPLIED TO EPIDEMIOLOGICAL DATA.**

Dissertação 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 mestre.

APROVADA em 08 de fevereiro de 2019.

Dr. Thelma Sáfadi- UFLA

Dr. Renato Ribeiro de Lima – UFLA

Dr. Paulo Henrique Sales Quimarães- UFLA

Dr. Alessandra Querino da Silva- UFGD

Prof. Dr. Thelma Sáfadi

Orientadora

LAVRAS

2019

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.

Agradeço meus pais que sempre foram muito presentes em todos os sentidos. Foram fundamentais durante minha graduação e igualmente importantes durante meu mestrado. Esta é a segunda vez que os agradeço por apoiarem e acreditarem em um sonho que não é apenas meu. Nossos esforços têm sido muito válidos, a cada dia que passa o doutorado tem se tornado uma realidade mais próxima.

Mel, 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. Juntamente com você pude contar com outros parceiros de república, que tornaram tudo mais leve e divertido. Por estes momentos alegres, que ficarão para sempre, também sou grato à minha irmã Dani e ao Cristhian.

Minha gratidão pela paciência, pelos ensinamentos e orientação da professora Thelma Sáfi. Agradeço também a todos os professores e técnicos do Departamento de Estatística.

Muito obrigado a todos os meus colegas de classe, foi muito bom estar dentro de uma turma tão colaborativa. De maneira especial, agradeço à Denize e Haiany pela amizade e ao Lucas pelo auxílio na maior dificuldade deste trabalho.

Sou grato à equipe Number One Lavras pelo apoio e a compreensão. Em especial agradeço a Giovana pelo apoio e pela generosidade. Foi bom contar com vocês!

Por fim, direciono meus agradecimentos a FAPEMIG e a CAPES pelo apoio financeiro por meio da concessão da bolsa de mestrado. O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES)- Código de financiamento 00. O presente trabalho também foi realizado com o apoio da Fundação de Amparo à Pesquisa de Minas Gerais (FAPEMIG).

SUMÁRIO

1	INTRODUÇÃO.....	09
2	OBJETIVOS.....	11
3	REVISÃO BIBLIOGRÁFICA	12
4	REFERENCIAL TEÓRICO	15
4.1	Tuberculose e sua propagação	15
4.2	Séries temporais: Principais conceitos.	18
4.3	Modelos de Box e Jenkins e suas variações	21
4.4	Modelos espaço- temporais: Principais conceitos	23
4.5	Matriz de ponderação espacial (W)	24
4.5.1	Matriz de ponderação espacial segundo critérios geográficos.....	26
4.5.2	Matriz de ponderação espacial segundo critérios socioeconômicos	27
4.6	Modelo STARMA e suas variantes	28
5	MATERIAIS E MÉTODOS.....	32
5.1	Materiais	32
5.1.1	Dos dados	32
5.1.2	Análise descritiva dos dados	32
5.2	Metodologia	36
5.2.1	Ajuste dos modelos de séries temporais.....	36
5.2.2	Obtenção das matrizes de vizinhança	37
5.2.3	Obtenção dos modelos STARMA.....	39
6	RESULTADOS E DISCUSSÕES	42
6.1	Ajuste dos modelos de séries temporais	42
6.2	Matrizes de vizinhança socioeconômicas	45
6.3	Identificação, estimação e diagnóstico dos modelos STARMA.....	46
7	CONSIDERAÇÕES FINAIS.....	54
	REFERÊNCIAS	56
	APÊNDICES.	62

RESUMO

Neste trabalho estudou-se a utilização de matrizes de vizinhança socioeconômicas em modelos espaço temporais da classe autorregressivo e de médias móveis (STARMA). O conjunto de dados escolhido é composto por nove séries temporais que quantificam a taxa de incidência de Tuberculose, observadas entre 2002 e 2017, nas seguintes cidades mineiras: Belo Horizonte, Betim, Contagem, Governador Valadares, Juiz de Fora, Lavras, Montes Claros, Pouso Alegre e Uberlândia. Uma vez que a maior parte das cidades encontram-se geograficamente distantes, foi necessária a utilização de matrizes de vizinhança socioeconômicas. As matrizes foram obtidas por meio de duas variáveis socioeconômicas: o IDH municipal e o investimento anual médio na saúde básica. A obtenção do modelo foi feita computacionalmente e consistiu de três etapas: Identificação, estimação e diagnóstico do modelo. Concluiu-se que, ao contrário do imaginado, é possível observar a existência de autocorrelação espaço-temporal na taxa de incidência de tuberculose, mesmo em cidades que se encontram geograficamente distantes. A distância entre as áreas observadas neste trabalho, fez com que as matrizes de vizinhança socioeconômicas se tornem a opção mais apropriada no ajuste de modelos STARMA para o conjunto de dados utilizado nesse trabalho.

Palavras-chave: Matriz de vizinhança socioeconômica. STARMA. Tuberculose.

ABSTRACT

In this work the use of socioeconomic neighborhood matrices was studied in time-space models of autorregressive and moving averages (STARMA) class. The selected data set is composed of nine time series that quantify the incidence rate of Tuberculosis observed between 2002 and 2017 in the following cities: Belo Horizonte, Betim, Contagem, Governador Valadares, Juiz de Fora, Lavras, Montes Claros, Pouso Alegre and Uberlândia. Since most cities are geographically distant, the use of socioeconomic neighborhood matrices was necessary. The matrices were obtained through two socioeconomic variables: the municipal IDH and the average annual investment in basic health. The model was obtained computationally and consisted of three stages: Identification, estimation and diagnosis of the model. It was concluded that, contrary to the imagined, it is possible to observe the existence of space-time autocorrelation in the incidence rate of tuberculosis, even in cities that are geographically distant. The distance between the areas observed in this work has made the socio-economic neighborhood matrices become the most appropriate option in the adjustment of STARMA models to the data used in this work.

Keywords: Socioeconomic neighborhood matrix. STARMA. Tuberculosis.

1 INTRODUÇÃO

A análise de séries temporais é capaz de fornecer informações e previsões que auxiliam em importantes tomadas de decisões. A análise e o ajuste de modelos para séries temporais, tem sido aplicado nas mais diversas áreas, tais como: economia, epidemiologia, sismologia, psicologia e também em estudos agrícolas, sociais, hídricos e pluviométricos.

Os modelos espaço-temporais, são mais recentes e um pouco mais complexos do que os modelos para séries temporais. Os modelos espaço-temporais também apresentam comportamento linear e podem ser empregados para ajustar dados econômicos, agrícolas, climáticos, ambientais e outros.

A principal diferença entre os modelos temporais e espaço-temporais é o número de variáveis correlacionadas. Enquanto para as séries temporais é observada apenas a correlação temporal, nos modelos espaço-temporais o conjunto de dados apresentará correlação espacial e temporal.

De maneira sintetizada, entende-se que um modelo espaço-temporal se adequa a dados de uma determinada variável, observada ao longo do tempo, em diferentes localidades. Por este motivo é necessário conhecer e modelar a interação entre cada uma das áreas nas quais a variável de interesse será mensurada.

A relação existente entre os locais de interesse é explicitada por meio da matriz de ponderação espacial (**W**), sendo também denominada matriz de vizinhança. Esta matriz, que busca indicar o grau de similaridade entre as regiões de interesse, pode ser construída de diferentes maneiras. Para os modelos espaço-temporais da classe autorregressiva e de médias móveis, classe de interesse deste trabalho, a obtenção de uma matriz **W** apropriada é de extrema importância e impactará diretamente no processo de ajuste do modelo.

Usualmente as matrizes de vizinhança são construídas em função de critérios geográficos como, por exemplo, a existência ou inexistência de fronteiras, o tamanho de fronteira compartilhada ou ainda em distâncias observadas entre duas áreas da região de interesse. Embora seja menos comum, é possível também obter matrizes **W** baseando-se em critérios socioeconômicos. Neste caso, o grau de similaridade entre duas áreas da região de interesse é dado em função de uma mesma variável socioeconômica mensurada nessas áreas.

Neste trabalho investigou-se a aplicabilidade de matrizes de vizinhança socioeconômicas em modelos espaço-temporais autorregressivos e de médias móveis (STARMA) aplicados a dados epidemiológicos. Também foi verificado se diferentes modelos espaço-temporais se ajustam à taxa de incidência de tuberculose observada entre janeiro de 2002 e dezembro de 2017 em nove cidades mineiras.

2 OBJETIVOS

O principal objetivo deste trabalho foi avaliar a utilização de matrizes de vizinhança socioeconômicas no ajuste de modelos da classe STARMA que foram utilizados no ajuste de dados epidemiológicos.

No entanto, para que o objetivo final fosse alcançado satisfatoriamente foi necessário o cumprimento de alguns objetivos específicos, dentre os quais se destacam:

- Ajustar modelos temporais para as nove cidades que compõem o conjunto de dados;
- Construir diferentes matrizes de ponderação espacial baseadas em critérios socioeconômicos;
- Ajustar modelos espaço-temporais da classe STARMA, para as matrizes de ponderação espacial construídas;
- Validar o ajuste dos modelos por meio de critérios apropriados;
- Analisar qual dentre as matrizes de vizinhança considerada foi a mais apropriada para o ajuste do modelo STARMA aplicado ao conjunto de dados proposto.

3 REVISÃO BIBLIOGRÁFICA

Os modelos para séries temporais lineares, univariados ou multivariados, são amplamente aplicados nos mais diversos temas de pesquisa. Usualmente denominados séries temporais, estes modelos são especialmente interessantes por sua relativa simplicidade associada à sua capacidade preditiva.

Apesar de maior simplicidade, os modelos univariados são de grande aplicabilidade. Apresentam resultados satisfatórios na explicação e/ou previsão de muitos fenômenos que despertam o interesse dos pesquisadores e assim contribuem com muitos estudos. Dentre os modelos univariados destacam-se os das classes Autorregressivo Integrado de Médias Móveis (ARIMA) e Sazonal Autorregressivo Integrado de Médias Móveis (SARIMA).

No que se refere a modelos estatísticos aplicados à área de saúde pública, os modelos de séries temporais são de grande relevância. Este fato é confirmado não somente pela expressiva quantidade de publicação da área nas quais são utilizados estes modelos, mas especialmente, pelo interesse de periódicos tais como, Revista Brasileira de Epidemiologia e Epidemiologia e Serviço de Saúde em publicarem trabalhos puramente conceituais, que visam elucidar conceitos e a aplicação de séries temporais. Como exemplos deste tipo de trabalho tem-se Antunes e Cardoso (2015) e Latorre et al. (2001).

Além dos modelos temporais, existem também os modelos espaço-temporais. Os modelos espaço-temporais da classe STARMA podem, de maneira simplista, serem entendidos como um caso específico de modelos Vetoriais Autorregressivos e de Médias Móveis (VARMA) (JIN, 2017). Dentre os modelos desta classe destacam-se os modelos Espaço-Temporal Autorregressivo (STAR) e Espaço-Temporal Autorregressivo de Médias Móveis (STARMA).

Os modelos espaço-temporais, especialmente da classe STARMA, são bem mais contemporâneos quando comparados aos de séries temporais. Os primeiros trabalhos com modelos da classe STARMA foram realizados na década de 80 por Pfeifer e Deutch (1980a). No entanto, considerando-se tal contemporaneidade percebe-se que análises espaço-temporais têm se mostrado úteis em estudos das mais diversas áreas. Muitas pesquisas são realizadas

buscando-se conhecer tendências e o comportamento espaço-temporal de fenômenos de interesse das mais diferentes áreas do conhecimento.

Os padrões e as tendências espaço-temporais começaram a despertar o interesse dos pesquisadores da saúde apenas na segunda metade da década de 90. Os primeiros trabalhos nesta área foram publicados por Jacquez (1996) e Zwartwald, Bastos e Castilho (1998). O primeiro trabalho estudou padrões de agrupamento e disseminação de patologias no espaço-tempo. No segundo trabalho foi realizada uma análise espaço-temporal da epidemia de AIDS no Brasil, ocorrida entre os anos de 1987 e 1995.

O estudo dos aspectos espaço-temporais de propagação de epidemias tem especial interesse pelas epidemias causadas por HIV e tuberculose, mas existem também trabalhos sobre outras enfermidades, tais como malária e obesidade. Como exemplo citam-se Nobre, Schmidt e Lopes (2005) que estudaram os padrões espaço-temporais de incidência de malária no Pará e Chaves et al. (2010) que analisaram a evolução do sobrepeso e da obesidade na população masculina adolescente entre os anos de 1980 a 2005.

Os padrões e a tendência espaço-temporal da tuberculose foram tratados por Hino, Santos e Villa (2005), por Melo et al. (2012) e por Santos Neto et al. (2015). O primeiro trabalho restringiu-se ao número de casos de tuberculose na cidade de Ribeirão Preto- SP relativos ao período compreendido entre 1998 e 2002. No segundo trabalho, foram considerados os casos de tuberculose ocorridos entre 1997 e 2006 em Rondônia e na Amazônia Ocidental. Santos Neto e seus colaboradores consideraram em sua pesquisa os casos de tuberculose confirmados em São Luiz- MA entre 2008 e 2012.

Guimarães et al. (2012), em um estudo mais abrangente também se interessou por padrões espaço-temporais de tuberculose, com maior ênfase na tendência temporal da doença em diversas regiões. O autor e seus colaboradores trataram a tendência espaço-temporal da tuberculose, da HIV e da pobreza, buscando com isso estudar a relação entre as patologias e o nível econômico do Brasil, das Américas e do mundo.

A disseminação do vírus HIV foi o assunto de interesse de Rodrigues-Junior e Castilho (2004), que abordaram a descrição espaço-temporal, considerando os diagnósticos realizados em todo território nacional entre os anos 1991 e 2000. Taquette et al. (2001) pesquisaram sobre a mesma epidemia, no entanto se restringiram aos casos de adolescentes entre 13 e 19 anos, na cidade do Rio de Janeiro. Rodrigues-Junior e Castilho (2009)

estudaram de maneira um pouco mais aprofundada os padrões espaço-temporais da epidemia de AIDS observadas nas fronteiras do Brasil, no entanto o trabalho restringiu-se a apresentar apenas os modelos de séries temporais ajustados.

Prado et al. (2009) apresentou um modelo espaço-temporal que se ajustava ao número de casos diagnosticados de AIDS ocorridos em São Paulo entre os anos de 1990 e 2005, contudo o modelo espaço-temporal autorregressivo (STAR) foi fundamentado em abordagem bayesiana. Szwarcwald et al. (2001) foi quem aprofundou-se mais na aplicação de modelos espaço-temporais aplicados a dados epidemiológicos de HIV, em seu trabalho a autora e seus colaboradores apresentaram um modelo da classe STAR que se ajustou ao número de casos confirmados na cidade do Rio de Janeiro entre 1988 e 1996.

É importante salientar que a maior parte dos trabalhos de grande relevância científica na área de saúde pública, não efetuaram de fato o ajuste de modelos espaço-temporais, mas se restringem a abordar a tendência e/ou a distribuição espaço-temporal de epidemias. Isto porque tais estudos destinam-se, em sua grande maioria, a analisar a taxa de incidência de determinada patologia e relaciona-la com indicadores econômicos, culturais, e ambientais da região estudada.

4 REFERENCIAL TEÓRICO

Nesta seção serão apresentados os conceitos e as definições fundamentais ao desenvolvimento deste trabalho.

Na primeira subseção será definida a tuberculose e apresentadas algumas estatísticas brasileiras e mundiais sobre a doença. Esta parte será finalizada destacando-se alguns aspectos sobre a patologia e sua disseminação.

Nas duas subseções seguintes serão tratados os conceitos fundamentais aos modelos temporais e apresentados os principais modelos de séries temporais, comumente chamados de modelos de Box e Jenkins.

Nas subseções restantes serão definidos conceitos espaço-temporais importantes, tais como: estacionariedade e matriz de ponderação espacial. Por último serão apresentados os principais modelos espaço-temporais.

4.1 Tuberculose e sua propagação

A tuberculose (TB), é uma doença de caráter transmissível, causada por uma micobactéria aeróbica, que se divide em vários gêneros, dos quais se destacam: *M. tuberculosis*, que é o agente etiológico da tuberculose em humanos; *Mycobacterium africanum*, causador da tuberculose em humanos, apenas em certas regiões da África; *Mycobacterium bovis*, *Mycobacterium caprae* e *Mycobacterium pinnipedii*, que causam TB em mamíferos silvestres e domesticados e *Mycobacterium microti*, que causa TB apenas em ratos (GAGNEUX, 2018).

De acordo com Misrha e Surolia (2018), a grande devastação, gerada pela doença, decorre do fato de o patógeno utilizar as reservas nutricionais do hospedeiro, levando-o a déficits funcionais consideráveis, desencadeando até mesmo, o óbito. Na grande maioria dos casos, a progressão da doença é sanada e a infecção inicial produz uma lesão granulomatosa compacta, auto-limitada, produzindo um meio eficaz, de conter a disseminação da bactéria (LEEMANS et al., 2001).

Anteriormente designada como "Consumo", a tuberculose possui uma alta capacidade destrutiva sobre o infectado, devido à manifestação de ciclos persistentes de febre, tosse sangrenta, dores no peito e perda significativa de massa muscular. A doença atingia tamanha devastação que, em meados do século XVII, foi apelidada como "capitão de todos esses

homens de morte". Driblando as tentativas de controle, tomadas pelas autoridades de saúde na época, a tuberculose foi a infecção crônica dominante, durante a primeira metade do século 20, incluindo a Segunda Guerra Mundial (MISRHA E SUROLIA, 2018).

Ocasionalmente principalmente pelo *Mycobacterium tuberculosis*, a TB pode apresentar algumas predisposições para seu desencadeamento, como fatores genéticos e condições ambientais (FERREIRA et al., 2005). Guimarães et al. (2012), destacam que dentre os fatores predisponentes para a tuberculose, destaca-se a coinfeção por HIV, o ambiente de pobreza e o crescimento populacional desordenado. Além disso, o autor ressalta que a gravidade e o impacto negativo da doença variam na razão inversa do índice de desenvolvimento humano (IDH).

De acordo com Coelho et al. (2010), no ano de 2005 a tuberculose consistia na maior causa de morbi-mortalidade mundial, entre as doenças infecto-contagiosas, totalizando um quarto das mortes evitáveis em adultos. Hijjar (2001), em concordância com Guimarães et al. (2012), justifica que tamanho aumento da taxa de infecção por *Mycobacterium tuberculosis*, se deu em detrimento ao aumento da pobreza, à má distribuição de renda e à urbanização acelerada.

Em um panorama mundial, quando são observados os países ou as regiões do mundo, a incidência da TB se dissemina de maneiras distintas. No entanto, observa-se que 21% dos infectados são habitantes de países desenvolvidos, enquanto 79% das infecções são observadas em países pobres e em desenvolvimento (COELHO et al., 2010).

Embora os índices de infecção tenham caído consideravelmente ao longo dos anos, a doença ainda é considerada uma questão de saúde pública e classificada, pela Organização Mundial da Saúde (OMS), como sendo uma doença negligenciada (BRASIL, 2010). Dados epidemiológicos de 2016, relatam o número de 10,4 milhões de pessoas infectadas, dentre as quais 1,8 milhões foram a óbitos (PELLISSARI; DIAZ-QUIJANO, 2017).

A transmissão da TB acontece exclusivamente por meio das vias respiratórias. A doença é transmitida quando gotículas que contenham o bacilo são expelidas, sob a forma de aerossóis, por pessoas infectadas, e inaladas por pessoas não portadoras da TB. O portador da TB pode expelir partículas contaminadas através da fala, do espirro e da tosse e segundo ‘ as gotículas mais leves podem permanecer em suspensão, no ar, por diversas horas. Uma vez depositados em roupas e utensílios, dificilmente, os bacilos se dispersarão novamente em

aerossóis. Portanto, nesse caso, não são relevantes na transmissão da doença.

Embora o pulmão seja o órgão mais acometido pela TB, a infecção pode, de diferentes formas, se disseminar gerando o comprometimento de outros órgãos. A tuberculose extrapulmonar, contribui com aproximadamente 10-20% dos casos diagnosticados, podendo atingir taxas bem maiores, de até 60%, naqueles com imunodeficiência acentuada, como, por exemplo, na coinfeção tuberculose/HIV. Os acometimentos extrapulmonares mais frequentes em nosso meio são: o pleural, o linfonodal e o urogenital, além do acometimento do sistema nervoso central (SNC). A manifestação clínica da doença, dependerá do local comprometido, podendo causar os mais diversos sinais e sintomas (BETHLEM, 2012).

Entre as manifestações de tuberculose extrapulmonar, a tuberculose pleural, é a mais frequente, sendo em alguns casos indetectável por meio de radiografia (BETHLEM, 2012). A segunda manifestação extrapulmonar mais frequente é a tuberculose ganglionar (TG). A TG pode ser dividida em: mediastinal, intra-abdominal e periférica, sendo este último tipo de infecção uma manifestação frequente nos pacientes com HIV, atingindo taxas de até 60% (FANLO e TIBERIO, 2007).

Em menor frequência, a TB também pode acometer o trato geniturinário. Neste caso a infecção inicia-se no córtex-renal e pode se disseminar pelo sistema urinário (ureter, bexiga e uretra), de maneira contaminar o aparelho genital, sendo o sexo masculino, mas susceptível a contaminação genital. (RIBEIRO et al., 1995).

O caso mais raro de acometimento extrapulmonar se dá no Sistema Nervoso Central (SNC). Devido a ampla vacinação de recém-nascidos com a vacina BCG, a incidência de meningoencefalite tuberculosa é mais baixa. Estima-se que este tipo de infecção correspondeu a apenas 1% das formas extrapulmonares observadas no Brasil no fim dos anos 90. Embora este tipo de infecção possa acontecer em qualquer faixa etária, especialmente nos pacientes soro positivos, a infecção se instala, preferencialmente, em crianças menores de 6 anos de idade (BETHLEM, 2012).

Segundo Watts (2012), Ducati et al. (2006) e Ramam et al. (2011), além da alta taxa de propagação da doença, um grande desafio vivenciado há décadas, era a deficiência nos métodos diagnósticos. O teste cutâneo de Mantoux, por exemplo, tratava de um método seguro, barato e simples, mas que não gerava uma boa representação da real infectividade da população, uma vez que indicava apenas a exposição prévia à micobactéria.

Na tentativa de evoluir, a radiografia do tórax, passou a ser utilizada como metodologia para o diagnóstico da tuberculose. Porém, se mostrava não segura para o controle da doença, uma vez que acusava a infecção apenas em seus estágios mais evoluídos. Somente com o decorrer do tempo e a evolução tecnológica, é que foi possível elevar o campo de diagnóstico da doença, contando então, com o método de tomografia computadorizada.

Atualmente, além dos exames por imagem, citados anteriormente, o diagnóstico da tuberculose conta com exames mais específicos e seguros, como: baciloscopia, teste rápido molecular para tuberculose e cultura para micobactéria (MINISTÉRIO DA SAÚDE, 2018).

No Brasil, a atuação do Programa Nacional de Controle da Tuberculose (PNCT) é de fundamental importância para a redução da morbidade e mortalidade ocasionada pela TB. Os sucessos do PNCT, têm sido notáveis, pois, apesar do país ainda figurar entre os 22 países no mundo com maior incidência da doença, houve no período entre 1990 e 2010 a redução de 38,4% na taxa da incidência e 35,8% na taxa de mortalidade (BRASIL, 2011).

Entre as estratégias adotadas pelo PNCT, está a busca passiva de casos novos, centrada no exame baciloscópico do escarro, ou testes rápidos automatizados, como o GeneXpert. Nas últimas décadas, o PNCT recomendou que também se fizesse a busca ativa dos contatos de pessoas que convivessem regularmente, no mesmo ambiente do paciente diagnosticado com TB, principalmente no domicílio (BRASIL, 2011).

4.2 Séries temporais: Principais conceitos

Uma dentre as diversas maneiras de se obter previsões é através da análise e do ajuste de séries temporais. Define-se como série temporal qualquer conjunto de dados ordenado no tempo, isto implica dizer que este conjunto de dados apresentará autocorrelação. (MORETTIN E TOLOI, 2006)

A análise de séries temporais é consideravelmente presente em pesquisas (BOX; JENKINS; REINSEL, 2008) das mais diversas áreas do conhecimento, tais como: economia, epidemiologia, medicina, sismologia, geofísica, meteorologia, geodesia, (NASCIMENTO et al., 2017; SHUMWAY; STOFFER 2006), ciências sociais e políticas, sociologia, história, psicologia (SÁFADI, 2004).

Uma das características mais importantes de uma série temporal é a dependência que cada observação apresenta com suas observações vizinhas. Se, por exemplo, entre duas ou mais variáveis de um conjunto de dados existe uma relação de dependência, de maneira que uma pode ser descrita em função da outra, define-se que estas variáveis apresentam correlação. De maneira análoga, para o caso de séries temporais, tem-se que a relação de dependência é observada em uma única variável em relação a ela mesma, e por isto, esta relação de dependência é denominada autocorrelação (MORETTIN; TOLOI, 2006).

Quando deseja-se analisar o comportamento de determinado conjunto de dados e realizar previsões, por meio do ajuste de séries temporais, é fundamental que se observe a existência de autocorrelação.

Apesar de ser amplamente utilizado na busca de previsões, é preciso destacar que a análise de uma série temporal não se limita apenas à sua capacidade preditiva, uma vez que pode ser de interesse do pesquisador conhecer o comportamento desta série, estudar a tendência temporal de determinada variável de interesse ou ainda verificar a existência de sazonalidade na série.

Uma série temporal será estacionária sempre que não apresentar tendência e sazonalidade. Graficamente a estacionariedade é observada quando uma série temporal se desenvolve aleatoriamente em torno de um valor constante.

Uma série temporal apresentará tendência sempre que evidenciar comportamento crescente ou decrescente. Dessa forma, a observação do gráfico de uma série temporal pode auxiliar na conclusão de existência ou inexistência de tendência naquela série. No entanto, ressalta-se que apenas a análise gráfica de uma série temporal não é suficiente para concluir se tal série apresenta ou não tendência. Para isto existem testes apropriados, como por exemplo, o teste de Cox-Stuart (BROCKWELL, DAVIS E CALDER, 2002; MORETTIN; TOLOI, 2006).

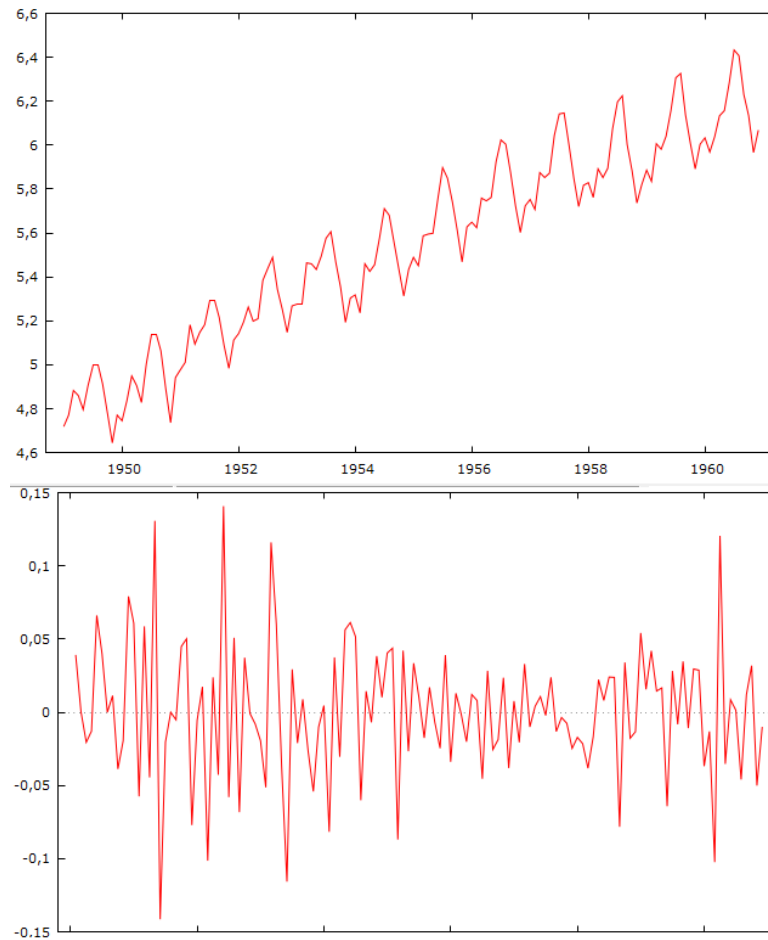
A componente sazonal, quando presente em uma série temporal, faz com que a série tenha comportamento periódico. A sazonalidade pode ser semanal, mensal, anual ou de qualquer outra periodicidade, dependendo da natureza da série. Para verificar a existência de sazonalidade em uma série temporal podem ser usados testes paramétricos e não paramétricos. Morettin e Tolo (2006) ressaltam o Teste de Kruskal-Wallis e o teste F.

Na prática, é muito incomum encontrar séries temporais que sejam estacionárias. A grande maioria das séries temporais apresentam tendência e/ou sazonalidade. No entanto, alguns modelos como, por exemplo os modelos de Box e Jenkins que serão apresentados na seção seguinte, são apropriados apenas para séries temporais estacionárias. Por este motivo, para algumas séries temporais, é necessário estimar e eliminar a tendência e a sazonalidade antes de realizar o ajuste da série.

A tendência pode ser estimada, por exemplo, por meio de um polinômio ou de uma função exponencial, enquanto a sazonalidade pode ser estimada através de uma soma de senos e cossenos. No entanto, é mais comum que a tendência e a sazonalidade sejam eliminadas por meio da tomada de diferenças da série temporal. A tendência pode ser eliminada de uma série temporal por meio de diferenças simples, enquanto a sazonalidade é eliminada por meio de diferenças sazonais.

A diferenciação de uma série temporal é realizada por meio do operador diferença, denotado por Δ . Para o caso da diferença simples, tem-se Δ^d , sendo 'd' o número de diferenças simples aplicadas à série. De maneira semelhante, a diferenciação sazonal é denotada por Δ^D sendo 'D' o número de diferenças sazonais tomadas da série. Não é recomendável que se tomem muitas diferenças de uma série temporal, uma vez que esta operação ocasiona a perda de observações do conjunto de dados original. No gráfico superior da Figura 1 é exibida uma série temporal não estacionária, onde é possível observar a existência de tendência e sazonalidade. No gráfico inferior da mesma figura, é exibida a mesma série temporal após aplicação de uma diferenciação simples e uma diferença sazonal, de modo a tornar-se estacionária.

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



Fonte: Do autor (2018)

4.3 Modelos de Box e Jenkins e suas variações

Existem muitos modelos que podem ser utilizados no ajuste de uma série temporal, de modo que a escolha do modelo mais adequado dependerá, dentre outros fatores, das características da série a ser modelada.

Também denominados modelos de Box e Jenkins, os modelos Autorregressivos e de Médias Móveis (ARMA) e suas variações são apropriados ao ajuste de séries estacionárias. Portanto, as componentes de tendência e sazonalidade devem ser eliminadas da série original, caso elas existam (MORETTIN E 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. O modelo ARIMA (p,d,q) é dado pela equação 1, note que ao excluir-se o operador de diferenciação simples da equação 1 (para o caso de d= 0), tem-se um modelo ARMA (p,q)

$$\Phi(B)\Delta^d Z_t = \theta(B)\epsilon_t \quad (1)$$

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; ϵ_t a componente aleatória não correlacionada do modelo, ou simplesmente um ruído branco, com média nula e variância constante (BROCKWELL, DAVIS E CALDER, 2002; MORETTIN; TOLOI, 2006).

Ainda considerando-se a equação 1, um modelo AR (p) é obtido excluindo-se o operador de diferença simples e fazendo-se $\theta(B) = 1$. Da maneira semelhante, um modelo MA (q) é obtido excluindo-se o operador de diferença simples e fazendo-se $\Phi(B) = 1$.

Os modelos da classe Sazonal Autorregressivo Integrado de Médias Móveis (SARIMA(p,d,q)(P,D,Q)) devem ser ajustados para os casos em que a série evidenciar um comportamento periódico menor que 12 meses, denominado sazonalidade. De maneira análoga ao modelo ARIMA, a componente sazonal deve ser eliminada, isto pode ser feito por meio de ‘D’ diferenciações sazonais. Um modelo SARIMA (p,d,q)(P,D,Q)_s pode ser representado por:

$$\Phi(B^s)\Phi(B)\Delta^d\Delta^D Z_t = \theta(B^s)\theta(B)\epsilon_t \quad (2)$$

sendo $\Phi(B^s)$ 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 $(1 - \theta_1 B^s - \theta_2 B^{2s} - \dots - \theta_p B^{qs})$; s é o período sazonal; Δ^d é o operador de diferença simples; Δ^D o operador de diferença sazonais; ϵ_t é a componente aleatória não correlacionada do modelo, ou simplesmente um ruído branco, com média nula e variância constante (BROCKWELL, DAVIS E CALDER, 2002; MORETTIN; TOLOI, 2006).

É importante que após o ajuste, tanto de modelos da classe ARIMA quanto de modelos da classe SARIMA, seja verificado se o modelo obtido é adequado. Um modelo

estará bem ajustado quando seus resíduos se comportarem como um ruído branco, isto é, o resíduo do modelo ajustado deve ser uma variável aleatória independente e identicamente distribuída com média nula e variância constante (MORETTIN; TOLOI, 2006). Outra forma de se fazer esta verificação é através do teste de Box e Pierce (1970), também conhecido como teste da estatística Q.

4.4 Modelos espaço-temporais: Principais conceitos

Os modelos espaço-temporais se caracterizam por ajustar conjuntos de dados que apresentem além da correlação temporal a correlação espacial. Entende-se que um conjunto de dados apresenta correlação quando existe dependência entre seus elementos. De maneira que, um conjunto de dados apresenta correlação temporal quando os elementos deste conjunto de dados se distribuírem ao longo do tempo de maneira dependente. De forma análoga, dizer que um conjunto de dados apresenta correlação espacial significa dizer que existe entre os elementos deste conjunto de dados um grau de associação espacial, de forma que suas disposições espaciais e suas interações com seus vizinhos devem ser consideradas durante o ajuste do modelo (JIN, 2017; PFEIFER; DEUTRCH, 1980a).

Afim de elucidar o conceito de correlação espaço-temporal, propõem-se a seguinte situação: Considere que determinada variável de interesse será monitorada simultaneamente em N diferentes locais durante T instantes de tempo, de forma que para todo instante $t \in T$, tem-se N observações desta mesma variável. Entende-se que este conjunto de dados apresentará correlação espaço-temporal se, em um dado instante de tempo tomar-se aleatoriamente uma dentre as N observações, e tal observação relacionar-se simultaneamente com observações obtidas em sua vizinhança e com as observações obtidas anteriormente e/ou posteriormente naquele local.

Outro conceito espaço-temporal relevante é de estacionariedade. Pfeifer e Deutch (1980a) afirmam que ao ajustar modelos da classe STARMA supõem-se não somente a estacionariedade temporal, mas também a estacionariedade espacial. Os autores explicam que isto implica dizer que a dependência verificada entre uma observação e suas observações vizinhas deve-se manter a mesma para todas as N regiões de interesse.

4.5 Matriz de ponderação espacial (W)

Para a definição e o ajuste dos modelos da classe STARMA é necessário que primeiramente seja definida a matriz de ponderação espacial. A matriz de ponderação espacial denotada por **W**, trata-se de uma matriz quadrada com dimensão $N \times N$ e $N^2 - N$ elementos de covariância (ALMEIDA, 2012).

Também denominada por matriz de vizinhança, matriz de proximidade espacial e matriz de pesos espaciais, trata-se de uma matriz que busca descrever a variabilidade espacial dos dados (MONTEIRO et al., 2004). De maneira sumária, pode-se dizer que a matriz de vizinhança tem por objetivo explicitar o grau de proximidade entre duas sub-regiões contidas na região de interesse, dado um critério apropriado. Almeida define o objetivo da matriz **W** como:

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).

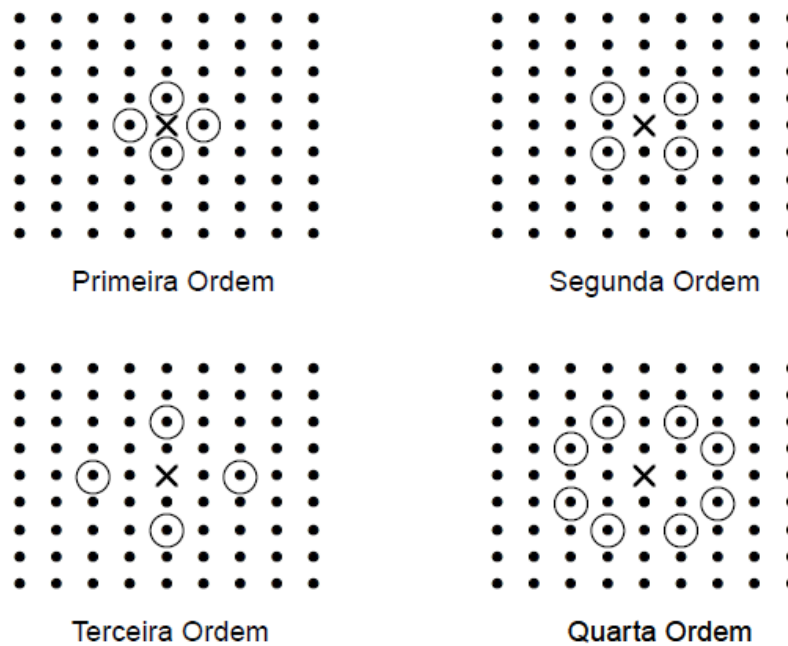
Desta forma, dada uma região de interesse constituída por N diferentes áreas: A_1, A_2, \dots, A_N , a matriz de ponderação espacial de ordem 'l' será representada de maneira semelhante a exibida abaixo, onde cada peso espacial $w_{i,j}$ expressa numericamente o grau de proximidade entre a região i e a região j para todo $i \neq j$.

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

É necessário que, após 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, respeitando-se então $\sum_j^N w_{ij} = 1$.

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'. A ordem da matriz de vizinhança indica uma hierarquia de proximidade entre uma região 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. Na Figura 2 é apresentada uma imagem desenvolvida por Pfeifer e Deutsch(1980a) que ilustra uma área e seus vizinhos de primeira, segunda, terceira e quarta ordem.

Figura 2- Vizinhança de primeira, segunda, terceira e quarta ordem.



Fonte: Pfeifer e Deutsch (1980a), Biz (2014).

A construção de uma matriz de vizinhança não é trivial e nem única. Martins, Biz e Ozaki (2014), em concordância com Pfeifer e Deutch (1980a), salientam que a definição dos pesos $w_{ij}^{(l)}$ é uma decisão subjetiva do construtor do modelo, sendo que este deve escolher pesos que reflitam a distribuição espacial do sistema em estudo. Rao e Antunes (2004), Biz (2014) e Smith (2013), são citados em muitos trabalhos como importantes estudos dedicados à definição da matriz de pesos \mathbf{W} .

Nas duas subseções seguintes serão exibidas diferentes maneiras de se obter a matriz de ponderação espacial. Na subseção 4.5.1 serão apresentadas matrizes de vizinhança criadas segundo critérios geográficos, isto é, considerando-se distâncias ou fronteiras. Na subseção 4.5.2 será apresentado o conceito de matrizes de proximidade espacial considerando-se parâmetros socioeconômicos.

Na seção 4.6 será apresentado formalmente o modelo STARMA e sua equação matemática. Nesta seção, se tornará mais evidente a importância de uma escolha apropriada da matriz \mathbf{W} e seu impacto sobre o ajuste de modelos da classe STARMA.

4.5.1 Matriz de ponderação espacial segundo 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 regiões em estudo.

As matrizes de vizinhança baseadas em fronteiras podem ser, antes de sua normalização, de natureza binária ou não. Esse tipo de matriz de vizinhança supõe que regiões que compartilham maior extensão de fronteira apresentam maior grau de proximidade (ALMEIDA, 2012).

A atribuição de pesos espaciais tendo em vista as fronteiras das áreas de interesse pode ser feita de duas maneiras distintas: A primeira consiste em atribuir pesos espaciais apenas verificando se duas áreas compartilham, ou não, fronteiras em comum. Enquanto na segunda os pesos são atribuídos considerando-se a extensão da fronteira compartilhada entre as áreas.

Dentre as matrizes de vizinhança obtidas considerando-se fronteiras destacam-se a Matriz de Contiguidade. Essa é uma das matrizes de vizinhança mais utilizada e se destaca por sua simplicidade.

Como principal desvantagem da atribuição de pesos espaciais considerando-se fronteiras, cita-se a dificuldade de mensurar o comprimento das fronteiras e dos perímetros de cada área. Outra desvantagem a ser considerada é que, em casos com áreas de tamanhos muito heterogêneos, será obtida uma matriz W desbalanceada, no sentido de que as áreas com fronteiras maiores tenderão a apresentar um número maior de vizinhos (ALMEIDA, 2012).

A construção de matrizes de vizinhança baseada em distâncias, fundamenta-se na ideia de que duas regiões geograficamente próximas terão maior interação espacial (ALMEIDA, 2012). Diversas maneiras são propostas para a construção da matriz W considerando-se a distância geográfica, dentre elas destacam-se: A matriz exponencial da distância, a matriz da distância inversa, a matriz de k-vizinhos e sua modificação proposta por Tyszler (2006).

Possivelmente, a matriz de vizinhança mais adequada, dentre as obtidas considerando-se distâncias é a matriz exponencial da distância (ALMEIDA, 2012). No entanto, de acordo como mesmo autor, a ponderação dos pesos espaciais para esta matriz dependerá de um parâmetro de amortecimento que muitas vezes é definido a priori, de maneira arbitrária e independente do restante das análises.

Outra desvantagem de se atribuir pesos espaciais considerando-se distâncias é a definição da métrica a ser usada. O mais comum é a opção por utilizar-se a distância Euclidiana, porém a utilização da Distância do Grande Circulo é mais apropriada para os casos em que as regiões de interesse estiverem geograficamente distantes (ALMEIDA, 2012; JIN, 2017).

4.5.2 Matriz de ponderação espacial segundo critérios socioeconômicos

Por vezes, pode ser mais pertinente medir o grau de similaridade entre duas áreas de interesse através 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 através de qualidade dos serviços básicos a população tais como: saúde, segurança, lazer e etc.

Nesses casos é recomendável a construção de matrizes de vizinhança baseadas em critérios socioeconômicos. Este conceito fundamenta-se no pensamento de que áreas com características iguais têm maior proximidade, ainda que geograficamente distantes (ALMEIDA, 2012), de maneira que a disposição e as iterações espaciais podem ser melhores representadas por meio de critérios socioeconômicos que a partir de critérios geográficos.

Não existe na literatura muitas instruções para a construção de matrizes de ponderação espacial socioeconômicas, menos ainda equações que norteiem a ponderação dos pesos espaciais para esse tipo de matriz de vizinhança. Uma das maneiras de se construir uma matriz de pesos espaciais socioeconômica é considerando-se a similaridade entre as regiões de interesse.

Em uma matriz de vizinhança socioeconômica de similaridade, os pesos espaciais ($w_{i,j}$) são atribuídos a partir do grau de semelhança existente entre duas áreas, dada uma variável socioeconômica de interesse. No entanto a ponderação dos pesos espaciais pode ser feita de muitas maneiras distintas, como ilustra o seguinte exemplo.

Imagine um estudo espaço-temporal cuja região de interesse é o sudeste brasileiro. 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, das seguintes formas:

$$i. \quad w_{i,j} = \begin{cases} 1, & \text{se } \frac{IDH_i + IDH_j}{2} \geq IDH_m \\ 0, & \text{caso contrário} \end{cases} \quad (3)$$

$$ii. \quad w_{i,j} = e^{-|IDH_i - IDH_j|} \quad (4)$$

$$iii. \quad w_{i,j} = \frac{1}{|IDH_i - IDH_j|^b} \quad (5)$$

sendo, respectivamente, IDH_i e IDH_j os Índice de Desenvolvimento Humano (IDH) dos estados i e j da região em análise, IDH_m o IDH médio do Sudeste brasileiro e b um fator de amortecimento.

Existem ainda matrizes de vizinhança socioeconômicas baseadas nos conceitos de dissimilaridade e de fluxo. Maiores esclarecimentos podem ser encontrados em Almeida (2012).

4.6 Modelo STARMA e suas variantes

Dentre os modelos espaço-temporais existe uma classe denominada autorregressivos de médias móveis espaço-temporais (STARMA). Jin (2017) classifica o modelo STARMA como um caso especial da classe de modelos autorregressivos e de médias móveis vetoriais (VARMA).

Os estudos envolvendo modelos da classe STARMA tiveram início no ano de 1975 com os trabalhos de Cliff e Ord (1975) e Martin e Oeppen (1975). Mas foi apenas após os trabalhos de Pfeifer e Deutsch (1980a, 1980b, 1980c) que esta classe de modelos teve notoriedade. Devido às restrições computacionais da década de 80, a implementação computacional completa do modelo era de extrema dificuldade o que fez com que esta classe de modelos caísse em esquecimento. Com os avanços da tecnologia, especialmente a evolução computacional, aproximadamente duas décadas mais tarde a classe de modelos STARMA tornou-se novamente objeto de interesse da ciência.

Esta classe de modelos é utilizada para descrever dados de séries temporais espacialmente localizados e se caracteriza pela dependência linear defasada no tempo e no espaço (PFEIFER ; DEUTRCH, 1980a, 1980b; MARTINS; BIZ; OZAKI, 2014).

É importante destacar que modelos da classe STARMA, assim como os modelos temporais da classe ARMA, são aplicáveis apenas a processos estacionários. Isto implica que a estrutura de covariância de $Z(t)$ não se alterará com o tempo (PFEIFER E DEUTRCH,

1980a; MARTINS; BIZ; OZAKI, 2014). Também de maneira semelhante aos modelos de Box e Jenkins, os modelos da classe STARMA são obtidos por meio do procedimento iterativos de três estágios, sendo estes estágios: Identificação do modelo, estimação dos parâmetros e diagnóstico do ajuste (CHEYSSON, 2016). Maiores esclarecimentos dos três estágios do processo iterativo propostos por Box e Jenkins, e estendidos para os modelos espaço-temporais, podem ser obtidas em Box e Jenkins (2008) e Pfeifer e Deutsch (1980a).

Para a construção de um modelo STARMA é necessário o uso do operador de atraso espacial $L^{(l)}$. O operador de defasagem espacial de l -ésima ordem, é definido pela equação 6. Note que este operador é diretamente dependente matriz de ponderação espacial (\mathbf{W}), anteriormente apresentada, uma vez que depende dos pesos espaciais (w_{ij}).

$$\begin{cases} L^{(0)}Z_j(t) = Z_j(t) \\ L^{(l)}Z_j(t) = \sum_{j=1}^N w_{ij}^{(l)} Z_j(t), \quad l > 0 \end{cases} \quad (6)$$

sendo $Z_j(t)$ a j -ésima série temporal ajustada.

O mesmo operador de atraso espacial, foi definido matricialmente por Pfeifer e Deutsch (1980a), como segue na Equação 7

$$L^{(l)}Z(t) = \begin{cases} \mathbf{W}^{(0)}Z(t) = \mathbf{W}^{(0)}Z(t) = \mathbf{I}_N Z(t), & l = 0 \\ \mathbf{W}^{(l)}Z(t), & l > 0 \end{cases} \quad (7)$$

sendo \mathbf{I}_N uma matriz identidade de dimensão $N \times N$ e $Z(t)$ a matriz de séries temporais que serão ajustadas pelo modelo.

O modelo STARMA pode ser escrito como apresentado na Equação 8

$$z_i(t) = \sum_{k=1}^p \sum_{l=0}^{\lambda_k} \phi_{kl} L^{(l)} z_i(t-k) - \sum_{k=1}^q \sum_{l=0}^{m_k} \theta_{kl} L^{(l)} \epsilon_i(t-k) + \epsilon_i(t) \quad (8)$$

em que $z_i(t)$ representa uma observação no local i realizada no instante t ; ‘ p ’ é a ordem auto regressiva e ‘ q ’ a ordem de médias móveis do modelo, λ_k é a ordem máxima de vizinhança do k -ésimo termo autorregressivo e m_k a ordem máxima espacial do k -ésimo termo de médias móveis, ϕ_{kl} e θ_{kl} são parâmetros do modelo, L^l é o operador de defasagem espacial e $\epsilon_i(t)$ são erros normais e aleatórios associados ao modelo, com $E[\epsilon_i(t)] = 0$ e

$$E[\epsilon_i(t)\epsilon_j(t+s)'] = \begin{cases} \sigma^2, & i = j \text{ e } s = 0 \\ 0, & \text{caso contrário} \end{cases}$$

O modelo STARMA também pode ser reescrito matricialmente explicitando-se a matriz de vizinhança \mathbf{W} , como mostra a equação 9

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

Nos modelos para séries temporais, foi destacado que o modelo ARMA trata-se da junção de outros dois modelos, o modelo AR(p) e o modelo MA(q). De modo análogo, tem-se que o modelo STARMA trata-se da união do modelo espaço-temporal autorregressivo (STAR) com o modelo espaço-temporal de médias móveis (STMA). Nas Equações 10 e 11, são apresentados, respectivamente, os modelos STAR e STMA

$$\mathbf{Z}(t) = \sum_{k=1}^p \sum_{l=0}^{\lambda_k} \phi_{kl} \mathbf{W}^l \mathbf{Z}(t-k) + \boldsymbol{\epsilon}(t), \quad (10)$$

$$\mathbf{Z}(t) = \boldsymbol{\epsilon}(t) - \sum_{k=1}^q \sum_{l=0}^{m_k} \theta_{kl} \mathbf{W}^l \boldsymbol{\epsilon}(t-k). \quad (11)$$

Caso o conjunto de dados não se comporte como um processo estacionário, deve-se de maneira semelhante aos modelos temporais, tomar diferenças simples e/ou sazonais do conjunto de séries temporais considerados. Para o caso de diferenciação simples serão obtidos modelos espaço-temporais autorregressivo integrado de médias móveis (STARIMA).

O modelo STARIMA $(p_\lambda, d, q_m)(P_\lambda, D, Q_M)$ é apresentado nas equações 12 e 13. Note que o primeiro modelo, apresentado na equação 12, trata-se de um STARMA do qual foram tomados d diferenças simples. O segundo modelo trata-se de um modelo STARMA do qual foram tomadas d diferenciações simples e D diferenças sazonais.

$$\nabla^d \mathbf{Y}(t) = \sum_{k=1}^p \sum_{l=0}^{\lambda_k} \phi_{kl} \mathbf{W}^l \nabla^d \mathbf{Y}(t-k) - \sum_{k=1}^q \sum_{l=0}^{m_k} \theta_{kl} \mathbf{W}^l \boldsymbol{\epsilon}(t-k) + \boldsymbol{\epsilon}(t) \quad (12)$$

$$\begin{aligned}
(I - \sum_{k=1}^P \sum_{l=0}^{\Lambda_k} \Phi_{lk} \mathbf{W}_l B^{kS}) \Phi_{p,\lambda}(B) \nabla_S^D \nabla^d \mathbf{Y}(t) \\
= (I - \sum_{k=1}^Q \sum_{l=0}^{M_k} \theta_{kL} \mathbf{W}_l B^{kS}) \theta_{q,m}(B) + \epsilon(t) \quad (13)
\end{aligned}$$

onde $\mathbf{Y}(t)$ representa a matriz de séries temporais não estacionárias a ser ajustada.

O último modelo, apresentado na Equação 13, é representado de maneira simplificado por Martins, Biz e Okazi (2014) através da Equação 14

$$\Phi_{P,\Lambda}(B^S) \Phi_{p,\lambda}(B) \nabla_S^D \nabla^d \mathbf{Y}(t) = \theta_{Q,M}(B^S) \theta_{q,m}(B) + \epsilon(t), \quad (14)$$

sendo $\Phi_{P,\Lambda}(B^S) = I - \sum_{k=1}^P \sum_{l=0}^{\Lambda_k} \Phi_{lk} \mathbf{W}_l B^{kS}$; $\nabla_S^D = (I - B^S)^D$;

$\theta_{Q,M}(B^S) = I - \sum_{k=1}^Q \sum_{l=0}^{M_k} \theta_{kL} \mathbf{W}_l B^{kS}$; P a ordem autorregressiva sazonal; D o número de diferenças sazonais; Q a ordem de médias móveis sazonal; S o período sazonal; Λ a ordem espacial autorregressivo e M a ordem espacial de médias móveis.

5 MATERIAIS E MÉTODOS

Esta seção iniciará expondo o conjunto de dados utilizado e suas estatísticas descritivas. Em sequência será apresentada a metodologia aplicada no decorrer do trabalho. Primeiramente será explicitada a metodologia aplicada no ajuste das nove séries temporais de interesse. Em seguida, será apresentada a metodologia empregada 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 Materiais

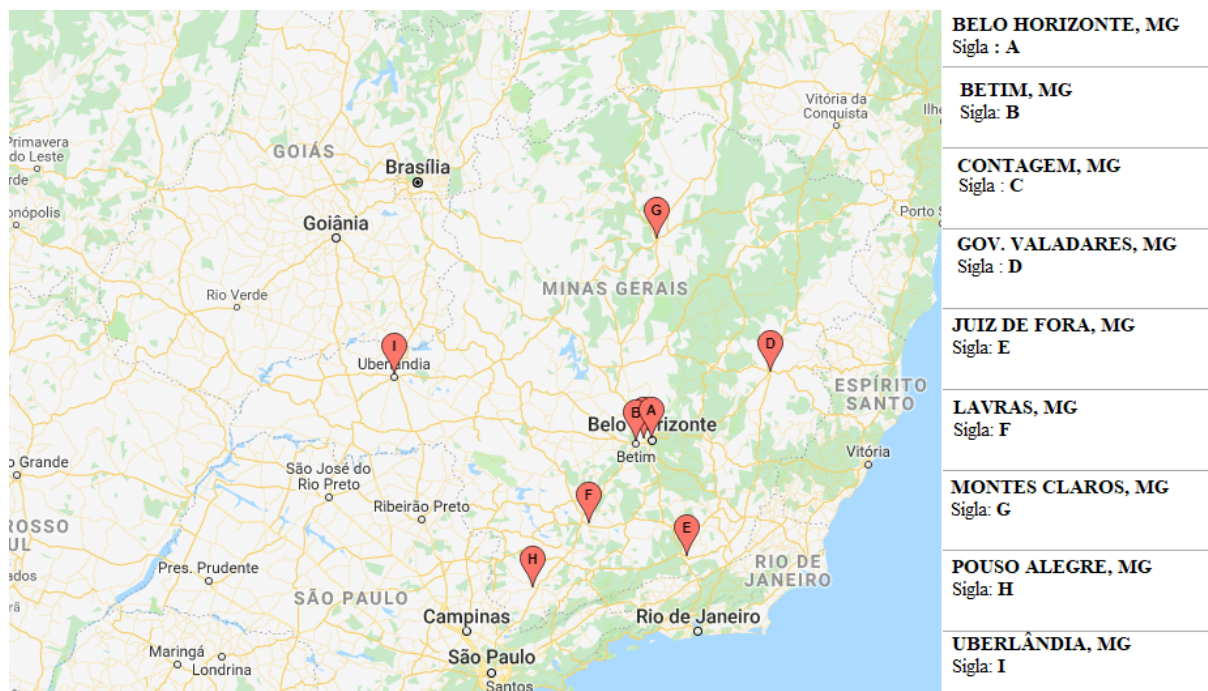
5.1.1 Dos materiais

Os dados foram obtidos do site do Departamento de Informática do Sistema Único de Saúde (DATASUS), e quantificam casos confirmados de tuberculose em nove cidades mineiras, sendo elas: Belo Horizonte, Betim, Contagem, Divinópolis, Juiz de Fora, Montes Claros, Pouso Alegre, Uberlândia e Lavras. O conjunto de dados trata-se de observações mensais coletadas entre janeiro de 2002 e dezembro de 2017.

5.1.2 Análise descritiva dos dados

Na Figura 3, são exibidas as nove cidades utilizadas no trabalho. Note que dentre as nove cidades apenas três (Belo Horizonte, Betim e Contagem) se encontram geograficamente próximas.

Figura 3- Indicação das cidades utilizadas no trabalho.

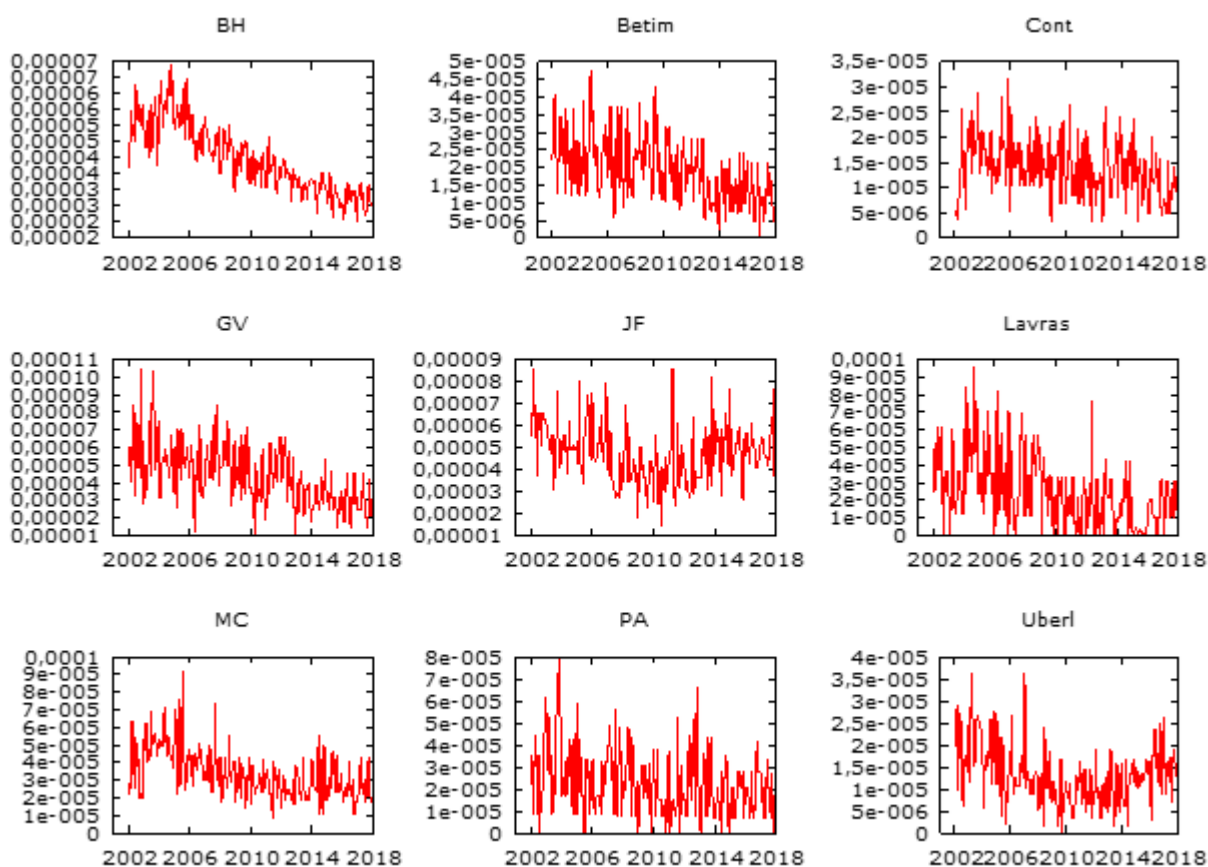


Fonte: Do autor (2018)

O mapa exibido na Figura 3 foi obtido através de uma ferramenta online chamada BatchGeo. Esta ferramenta se encontra disponível no site: < <https://pt.batchgeo.com/>> e trata-se de uma plataforma que, por meio do Google Maps e dados tabulados, gera novos gráficos personalizados.

Na Figura 4 é exibida a representação gráfica das nove séries temporais. Estes gráficos ilustram a taxa de incidência de TB observada entre 2002 e 2017 em cada uma das nove cidades de interesse. Note o comportamento decrescente no gráfico das séries temporais de Belo Horizonte, Betim, Contagem, Governador Valadares e Lavras. Enquanto que, a série de Pouso Alegre apresenta comportamento constante e as séries de Juiz de Fora, Montes Claros e Uberlândia apresentaram um aumento na incidência de tuberculose a partir do ano de 2010.

Figura 4- Gráfico das séries temporais das nove cidades de interesse: Taxa de incidência de TB entre janeiro de 2002 e dezembro de 2017.



Fonte: Do autor (2018)

Salienta-se que no site do DATASUS não são fornecidas as taxas de incidência mensal da tuberculose. Os valores da taxa de incidência mensal de cada uma das cidades de interesse foram obtidos por meio da divisão do número de casos de TB confirmados (obtidos no DATASUS) pela população estimada da cidade em questão.

A estimativa da população para as nove cidades foi realizada por modelos de regressão linear simples, baseando-se nos dados censitários disponibilizados pelo IBGE. Na Tabela 1 são exibidos os modelos de regressão obtidos. Os parâmetros dos modelos foram obtidos computacionalmente por meio do software R (R CORE TEAM, 2017) e adotou-se um nível de significância de 0,05. As populações estimadas que foram obtidas através dos modelos apresentados na Tabela 1 podem ser verificadas no Apêndice A.

Tabela 1- Modelos de regressão utilizados na estimação das populações para cidade de interesse.

CIDADE	MODELO DE REGRESSÃO
Belo Horizonte	População= $-6,688 \times 10^7 + 3,455 \times 10^4 \cdot \text{ano}$
Betim	População= $-1,398 \times 10^4 + 7,141 \cdot \text{ano}$
Contagem	População= $-1,255 \times 10^4 + 6,542 \cdot \text{ano}$
Governador Valadares	População= $-4,859 \times 10^6 + 2,552 \times 10^3 \cdot \text{ano}$
Juiz de Fora	População= $-1,195 \times 10^7 + 6,197 \times 10^3 \cdot \text{ano}$
Lavras	População= $-2,050 \times 10^6 + 1,060 \times 10^3 \cdot \text{ano}$
Montes Claros	População= $-1,077 \times 10^7 + 5,536 \times 10^3 \cdot \text{ano}$
Pouso Alegre	População= $-4,661 \times 10^6 + 2,39 \times 10^3 \cdot \text{ano}$
Uberlândia	População = $-2,367 \times 10^7 + 1,208 \times 10^4 \cdot \text{ano}$

Fonte: Do autor (2018)

Na Tabela 2, são exibidas estatísticas descritivas para a série temporal de cada uma das nove cidades. A maior incidência de TB foi observada em Juiz de Fora enquanto a menor taxa de incidência ocorreu em Uberlândia.

Tabela 2- Análise descritiva das séries temporais

CIDADE	MÉDIA	DESVIO PADRÃO	COEFICIENTE DE VARIABILIDADE (%)
Belo Horizonte	$3,980 \times 10^{-5}$	$1,072 \times 10^{-5}$	26,9
Betim	$1,915 \times 10^{-5}$	$9,359 \times 10^{-6}$	48,9
Contagem	$1,379 \times 10^{-5}$	$5,722 \times 10^{-6}$	41,5
Governador Valadares	$4,371 \times 10^{-5}$	$1,739 \times 10^{-5}$	39,8
Juiz de Fora	$4,731 \times 10^{-5}$	$1,340 \times 10^{-5}$	28,3
Lavras	$2,497 \times 10^{-5}$	$2,001 \times 10^{-5}$	80,3
Montes Claros	$3,403 \times 10^{-5}$	$1,436 \times 10^{-5}$	42,2
Pouso Alegre	$2,259 \times 10^{-5}$	$1,528 \times 10^{-5}$	67,6
Uberlândia	$1,322 \times 10^{-5}$	$6,734 \times 10^{-6}$	50,9

Fonte: Do autor (2018)

5.2 Metodologia

5.2.1 Ajuste dos modelos de séries temporais

Toda a análise e ajuste de séries temporais foi executado pelo software Gretl (COTTRELL; LUCCHETTI, 2017). Iniciou-se a análise do conjunto das nove séries temporais verificando-se a necessidade de transformação dos dados. Para isto foi aplicado o teste de Amplitude X Média. Dada a necessidade de transformação de algumas séries temporais, utilizou-se o teste de Box-Cox para determinar qual transformação seria a mais apropriada. Para maiores esclarecimentos sobre o teste de Box-Cox recomenda-se a leitura de Box e Cox(1964) e Silva (2013).

Foi utilizado o teste de Cox-Stuart, popularmente conhecido como Teste do Sinal, para verificar a existência de tendência nas séries. A presença de sazonalidade foi analisada por meio do gráfico da função de autocorrelação e do teste F para as variáveis Dummies Sazonais.

Desta forma, as nove séries temporais puderam ser ajustadas satisfatoriamente por meio de modelos da classe ARIMA. A validação do ajuste foi realizada por meio do teste de Box-Pierce.

5.2.2 Obtenção das matrizes de vizinhança

Foram construídas duas matrizes de vizinhança (**W**) de primeira ordem, ambas baseadas em critérios socioeconômicos. As matrizes de vizinhança socioeconômicas foram construídas baseando-se apenas no conceito de similaridade. Foram utilizadas duas variáveis socioeconômicas distintas, sendo elas o investimento direto na saúde básica municipal, obtido no site do DATASUS e o IDH municipal.

Na primeira matriz de vizinhança considerou-se a média de investimento anual em saúde básica das nove cidades de interesse. Na segunda matriz de ponderação socioeconômica considerou-se o último IDH das cidades mineiras, uma vez que se sabe que a tuberculose é fortemente relacionada a questões econômicas, sociais e ambientais (GUIMARÃES et al., 2012). Para a obtenção dos pesos espaciais de tais matrizes de vizinhança utilizaram-se as Equações 15 e 16. Note que a Equação 15 trata-se de uma variante da Equação 5, anteriormente definida, para a qual tem-se o parâmetro $b=1$.

$$w_{i,j} = \frac{1}{|IDH_i - IDH_j|} \quad (15)$$

$$w_{i,j} = \frac{1}{|In_i - In_j|} \quad (16)$$

em que IDH_i é o IDH da i -ésima cidade; IDH_j é o IDH da j -ésima cidade; In_i é investimento médio da i -ésima cidade em saúde básica e In_j é o investimento médio da j -ésima cidade em saúde básica.

Para o cálculo do investimento anual médio em saúde básica considerou-se o investimento dos anos de 2015, 2016 e 2017. Os valores de investimento na saúde básica das nove cidades de interesse são exibidos na Tabela 3, juntamente com seus respectivos IDH.

Tabela 3- Cidades de interesse: Investimento médio em saúde básica e IDH

CIDADE	INVESTIMENTO MÉDIO	IDH
Belo Horizonte	R\$ 139242499,76	0,810
Betim	R\$ 23702072,82	0,749
Contagem	R\$ 31710230,38	0,756
Governador Valadares	R\$ 23634755,92	0,727
Juiz de Fora	R\$ 27877446,56	0,778
Lavras	R\$ 6728080,66	0,782
Montes Claros	R\$ 38182719,07	0,770
Pouso Alegre	R\$ 8502476,86	0,774
Uberlândia	R\$ 34677774,99	0,789

Fonte: Do autor (2018)

Ressalta-se que o IDH trata-se de uma medida de desenvolvimento humano, que utiliza como critério indicadores de educação, longevidade e de renda. Para o estado de Minas Gerais o IDH médio é de 0,731. Tal indicador varia entre 0 e 1 e é categorizado da seguinte maneira (Wikipédia, 2018):

$$IDH = \begin{cases} 0,000 \leq IDH \leq 0,499: \textit{Muito baixo} \\ 0,500 \leq IDH \leq 0,599: \textit{Baixo} \\ 0,600 \leq IDH \leq 0,699: \textit{Médio} \\ 0,700 \leq IDH \leq 0,799: \textit{Alto} \\ 0,800 \leq IDH \leq 1,000: \textit{Muito Alto} \end{cases}$$

Observe que dentre as cidades em estudo apenas a cidade de Governador Valadares está abaixo do IDH médio estadual. No entanto, todas as cidades apresentam IDH classificados como 'Alto', exceto Belo Horizonte que tem seu IDH na classe 'Muito Alto'.

5.2.3 Obtenção dos modelos STARMA

Para as duas matrizes de vizinhança construídas foi ajustado um modelo da classe STARMA. O ajuste dos modelos foi feito através do software estatístico livre R (R CORE TEAM, 2017), utilizando-se um pacote denominado ‘starma’ de autoria de Felix Cheysson (CHEYSSON, 2016).

O pacote ‘starma’ fornece todas as ferramentas necessárias para identificar, estimar e diagnosticar modelos da classe STARMA para séries espaço-temporais. O pacote utiliza o procedimento de construção do modelo iterativo de três estágios desenvolvido por Box e Jenkins (1970) e estendido para a modelagem espaço-temporal proposta por Pfeifer e Deutsch (1980a). Além disto a estimação dos parâmetros do modelo, que anteriormente apresentava alto custo computacional, foi otimizada por meio do uso de um filtro de Kalman (CHEYSSON, 2016).

É importante salientar que o pacote ‘starma’, assim como no ajuste de séries temporais, se aplica somente a dados estacionários. Por este motivo todas as funções que serão apresentadas em seguida, foram aplicadas após a eliminação da tendência das séries temporais que compõem a série espaço-temporal de interesse.

Os três estágios do procedimento de obtenção do modelo STARMA, realizados pelo pacote ‘starma’ são:

i-Identificação:

Nesta etapa do processo de obtenção do modelo é identificada a ordem do modelo por meio das funções ‘stacf()’ e ‘stpacf()’. Por meio da análise dos gráficos da função de autocorrelação espaço-temporal (stacf) e da função de autocorrelação parcial espaço-temporal (stpacf) são determinadas as ordens médias móveis e autorregressivas, respectivamente.

Jin (2017) esclarece em seu trabalho que 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 $STAR(p_{\lambda_0}, p_{\lambda_1}, \dots, p_{\lambda_p})$ 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*.

ii-Estimação:

O método de estimação proposto por Pfeifer e Deutsch (1980a) utiliza o método da máxima verossimilhança. Este método, além de ser necessário a suposição de normalidade dos resíduos, é de alto custo computacional, o que ocasionou dificuldades e o desuso do modelo STARMA na década de 80. As demonstrações matemáticas necessárias para a obtenção dos parâmetros por meio do método da Máxima Verossimilhança podem ser encontradas em Jin (2017).

A estimação dos parâmetros através do pacote ‘*starma*’ é feito utilizando-se o filtro de Kalman (CHEYSSON, 2016). O filtro de Kalman, por meio de equações matriciais, obtém de maneira recursiva a estimativa dos parâmetros. Tais estimativas são consideradas ótima no sentido do princípio dos mínimos quadrados (JIN, 2017).

A estimação dos parâmetros, por meio da utilização do filtro de Kalman, além de ocasionar uma considerável otimização computacional é extremamente eficiente quando aplicada a grandes conjuntos de dados (CHEYSSON, 2016). E, apesar de ser o mais comum, não é necessário supor-se normalidade residual ao utilizar-se o filtro de Kalman na obtenção de estimativas dos parâmetros de modelos STARMA (JIN, 2017).

No pacote ‘*starma*’ a estimação dos parâmetros é obtida a partir da função ‘*starma()*’. É importante salientar que antes da obtenção dos parâmetros é necessário que o conjunto de dados seja centralizado por meio da função ‘*stcenter()*’ de modo que a série espaço-temporal passe a ter média nula e erro padrão unitário (CHEYSSON, 2016).

iii-Diagnóstico

Assim como no ajuste de modelos para séries temporais, após a estimação dos parâmetros é necessário verificar se o modelo obtido é apropriado. Também de maneira semelhante às séries temporais, uma série espaço-temporal estará bem ajustada quando seu resíduo apresentar comportamento semelhante ao de um ruído branco.

No pacote ‘*starma*’ a maneira de verificar se modelo obtido foi devidamente ajustado é por meio da função ‘*stcot.test()*’. Esta função trata-se de uma extensão do teste de Box-Pierce,

na qual é testada a hipótese nula de não correlação residual. Para esse teste utiliza-se a Equação 17 (CHEYSSON, 2016).

$$N \sum (T - s)[\hat{\rho}_l(s)]^2 \sim \chi^2(slag; tlag) \quad (17)$$

onde 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 a defasagem temporal considerada; $\hat{\rho}_l(s)$ a função de autocorrelação espaço-temporal para l-ésima ordem de vizinhança e s-ésima defasagem temporal; *slag* maior ordem de vizinhança considerada; *tlag* número de defasagem temporal utilizado na construção do gráfico da função de autocorrelação espaço-temporal, que pode ser obtido pela função '*stacf()*'.

6 RESULTADOS E DISCUSSÕES

Nesta seção serão relatados os resultados obtidos. Primeiramente serão apresentados os resultados decorrentes da análise dos modelos de séries temporais. Em seguida serão apresentadas as duas matrizes de vizinhança socioeconômicas. Por último serão tratados os modelos STARMA.

6.1 Ajuste dos modelos de séries temporais

Inicialmente verificou-se a necessidade de transformação dos dados por meio do teste Amplitude X Média, para isto considerou-se o nível de significância de 0,05. Na Tabela 4 são exibidos os resultados dos testes e suas respectivas transformações, quando necessárias. O teste de Box-Cox foi utilizado para determinar a transformação adequada.

Nas séries de Belo Horizonte, Governador Valadares e Lavras foi aplicada a transformação logarítmica enquanto que nas séries de Betim e Pouso Alegre aplicou-se a transformação raiz quadrada.

Tabela 4- Resultados do teste de Box-Pierce e transformações aplicadas.

SÉRIE TEMPORAL	TESTE BOX-PIERCE (Estatística Q)	TRANSFORMAÇÃO APLICADA
Belo Horizonte	0,001	log(Belo Horizonte)
Betim	0,001	$\sqrt{\text{Betim}}$
Contagem	0,491	-
Governador Valadares	0,001	log(Governador Valadares)
Juiz de Fora	0,494	-
Lavras	0,004	log(Lavras)
Montes Claros	0,084	-
Pouso Alegre	0,002	$\sqrt{\text{Pouso Alegre}}$
Uberlândia	0,060	-

Fonte: Do autor (2018)

A não estacionariedade das nove séries temporais foi verificada inicialmente por meio da análise dos seus gráficos da função de autocorrelação, exibidos no Apêndice B. Em seguida constatou-se, por meio do teste de Box-Pierce, que as séries não se comportavam como um ruído branco. Para o teste considerou-se o nível de significância de 0,05 e $\chi^2_{5\%;35} = 49,82$, o resultado é exibido na Tabela 5.

Tabela 5- Resultado do teste de Box-Pierce para as nove séries temporais.

SÉRIE TEMPORAL	TESTE BOX-PIERCE (Estatística Q)
log(Belo Horizonte)	2845,82
$\sqrt{\text{Betim}}$	346,03
Contagem	62,99
log(Governador Valadares)	398,44
Juiz de Fora	124,48
log(Lavras)	70,82
Montes Claros	496,53
$\sqrt{\text{Pouso Alegre}}$	50,89
Uberlândia	341,74

Fonte: Do autor (2018)

A análise dos gráficos da função de autocorrelação, exibidos no Apêndice B, também sugeriu, para todas as séries temporais, a existência de tendência e a inexistência de sazonalidade, o que posteriormente foi verificado. Constatou-se a existência de tendência em todas as séries temporais por meio do teste do Cox-Stuart. A inexistência de sazonalidade foi verificada através do teste F para as variáveis Dummies Sazonais.

As nove séries temporais foram ajustadas por meio de um modelo da classe ARIMA(0,1,1). Na Tabela 6 são apresentadas as estimativas dos parâmetros dos modelos com seus respectivos valores de erro padrão, foi considerado o nível de significância de 0,05. Note

que apenas a série temporal de Betim apresentou valor estatisticamente significativo para a constante do modelo.

Tabela 6- Tabela de ajuste dos modelos: Parâmetros ajustados com seus respectivos erros padrão entre parênteses.

SÉRIE TEMPORAL	PARÂMETROS AJUSTADOS (ERRO PADRÃO)	
	Constante	θ_1
log (Belo Horizonte)	-	-0,820578 (0,039672)
$\sqrt{\text{Betim}}$	-1,04077x10 ⁻⁵ (-2,81002x10 ⁻⁶)	-0,920830 (0,025733)
Contagem	-	-0,921338 (0,045206)
log(Governador Valadares)	-	-0,898043 (0,038904)
Juiz de Fora	-	-0,885861 (0,066620)
log(Lavras)	-	-0,924733 (0,036265)
Montes Claros	-	-0,892134 (0,029108)
$\sqrt{\text{Pouso Alegre}}$	-	-0,960087 (0,021224)
Uberlândia	-	-0,891910 (0,028828)

Fonte: Do autor (2018)

O teste de Box- Pierce foi novamente aplicado para verificar se os modelos foram devidamente ajustados. Como o esperado, após o ajuste os resíduos de todos os modelos apresentaram comportamento semelhante ao de um ruído branco. Na Tabela 7, são exibidos os resultados do teste de Box-Pierce, novamente considerando 0,05 como nível de

significância e $\chi_{5\%;35}^2 = 49,82$. No Apêndice C são exibidos os gráficos das nove séries temporais que compõem o conjunto de dados e suas respectivas séries ajustadas.

Tabela 7- Resultado do teste de Box-Pierce após o ajuste das nove séries temporais.

SÉRIE TEMPORAL	TESTE BOX-PIERCE (Estatística Q)
log(Belo Horizonte)	29,04
$\sqrt{\text{Betim}}$	43,48
Contagem	39,61
log(Governador Valadares)	30,71
Juiz de Fora	28,31
log(Lavras)	27,83
Montes Claros	48,13
$\sqrt{\text{Pouso Alegre}}$	49,33
Uberlândia	27,72

Fonte: Do autor (2018)

6.2 Matrizes de vizinhança socioeconômicas

Para a obtenção das matrizes de vizinhança foram consideradas duas variáveis socioeconômicas, o IDH municipal e o investimento direto na saúde básica municipal.

A ponderação dos pesos espaciais para a matriz de vizinhança baseada no IDH municipal foi feita por meio da Equação 15, enquanto que para a matriz de vizinhança baseada no investimento médio anual em saúde básica foi feito pela Equação 16. A seguir são exibidas as matrizes obtidas após normalização:

$$W_{Inv} = \begin{matrix} & \text{B.H.} & \text{Betim} & \text{Contagem} & \text{Gov. Val.} & \text{J. de Fora} & \text{Lavras} & \text{M. Claros} & \text{P. Alegre} & \text{Uberl.} & \\ \begin{matrix} 0,0000 \\ 0,0005 \\ 0,1807 \\ 0,0547 \\ 0,0099 \\ 0,0088 \\ 0,0116 \\ 0,0077 \\ 0,0092 \end{matrix} & \begin{matrix} 0,1232 \\ 0,0000 \\ 0,0938 \\ 0,2425 \\ 0,0649 \\ 0,0809 \\ 0,0770 \\ 0,0915 \\ 0,0881 \end{matrix} & \begin{matrix} 0,1324 \\ 0,0080 \\ 0,0000 \\ 0,2193 \\ 0,2875 \\ 0,0468 \\ 0,1814 \\ 0,0433 \\ 0,3258 \end{matrix} & \begin{matrix} 0,1231 \\ 0,9576 \\ 0,0936 \\ 0,0000 \\ 0,2599 \\ 0,0692 \\ 0,0804 \\ 0,0663 \\ 0,0876 \end{matrix} & \begin{matrix} 0,1278 \\ 0,0154 \\ 0,1973 \\ 0,2190 \\ 0,0000 \\ 0,0554 \\ 0,1135 \\ 0,0518 \\ 0,1421 \end{matrix} & \begin{matrix} 0,1074 \\ 0,0038 \\ 0,0302 \\ 0,0550 \\ 0,0521 \\ 0,0000 \\ 0,0370 \\ 0,6671 \\ 0,0346 \end{matrix} & \begin{matrix} 0,1409 \\ 0,0044 \\ 0,1168 \\ 0,0639 \\ 0,1069 \\ 0,0371 \\ 0,0000 \\ 0,0338 \\ 0,2755 \end{matrix} & \begin{matrix} 0,1089 \\ 0,0042 \\ 0,0326 \\ 0,0614 \\ 0,0568 \\ 0,6598 \\ 0,0451 \\ 0,0000 \\ 0,0363 \end{matrix} & \begin{matrix} 0,1361 \\ 0,0059 \\ 0,2548 \\ 0,0844 \\ 0,1620 \\ 0,0419 \\ 0,3818 \\ 0,0384 \\ 0,0000 \end{matrix} & \begin{matrix} \text{B.H.} \\ \text{Betim} \\ \text{Cont.} \\ \text{G. V.} \\ \text{J.F.} \\ \text{Lav.} \\ \text{M.C.} \\ \text{P.A.} \\ \text{Uberl.} \end{matrix} \end{matrix} \quad (17)$$

$$W_{IDH} = \begin{matrix} & \text{B.H.} & \text{Betim} & \text{Contagem} & \text{Gov. Val.} & \text{J. de Fora} & \text{Lavras} & \text{M. Claros} & \text{P. Alegre} & \text{Uberl.} & \\ \begin{matrix} 0,0000 \\ 0,0429 \\ 0,0424 \\ 0,0633 \\ 0,0369 \\ 0,0493 \\ 0,0369 \\ 0,0332 \\ 0,1009 \end{matrix} & \begin{matrix} 0,0765 \\ 0,0000 \\ 0,3268 \\ 0,2387 \\ 0,0407 \\ 0,0419 \\ 0,0702 \\ 0,0478 \\ 0,0529 \end{matrix} & \begin{matrix} 0,0864 \\ 0,3739 \\ 0,0000 \\ 0,1811 \\ 0,0537 \\ 0,0531 \\ 0,1053 \\ 0,0664 \\ 0,0642 \end{matrix} & \begin{matrix} 0,0562 \\ 0,1189 \\ 0,0789 \\ 0,0000 \\ 0,0231 \\ 0,0251 \\ 0,0343 \\ 0,0375 \\ 0,0342 \end{matrix} & \begin{matrix} 0,1458 \\ 0,0902 \\ 0,1040 \\ 0,1030 \\ 0,0000 \\ 0,3454 \\ 0,1843 \\ 0,3613 \\ 0,1925 \end{matrix} & \begin{matrix} 0,1666 \\ 0,0793 \\ 0,0880 \\ 0,0955 \\ 0,2953 \\ 0,0000 \\ 0,1229 \\ 0,1807 \\ 0,3026 \end{matrix} & \begin{matrix} 0,1166 \\ 0,1246 \\ 0,1634 \\ 0,1221 \\ 0,1476 \\ 0,1151 \\ 0,0000 \\ 0,3613 \\ 0,1115 \end{matrix} & \begin{matrix} 0,1296 \\ 0,1047 \\ 0,1271 \\ 0,1117 \\ 0,2953 \\ 0,1727 \\ 0,3686 \\ 0,0000 \\ 0,1412 \end{matrix} & \begin{matrix} 0,2222 \\ 0,0654 \\ 0,0693 \\ 0,0847 \\ 0,1074 \\ 0,1973 \\ 0,0776 \\ 0,0964 \\ 0,0000 \end{matrix} & \begin{matrix} \text{B.H.} \\ \text{Betim} \\ \text{Cont.} \\ \text{G. V.} \\ \text{J.F.} \\ \text{Lav.} \\ \text{M.C.} \\ \text{P.A.} \\ \text{Uberl.} \end{matrix} \end{matrix} \quad (18)$$

Por meio da análise da primeira matriz percebe-se a discrepância existente entre os valores investidos em saúde básica em Belo Horizonte e as demais cidades, isto corrobora os valores exibidos na Tabela 3. Na matriz exibida na Equação 17, observa-se que o maior valor para peso espacial ocorre entre Governador Valadares e Betim, isto ocorre pois o valor investido em saúde básica nesses municípios são próximo.

Também mesma Tabela 3 também são exibidos o IDH dos nove municípios, para esta variável é possível perceber que não há grande discrepâncias por este motivo observa-se a inexistência de valores demasiadamente grandes ou pequenos na matriz W_{IDH} .

6.3 Identificação, estimação e diagnóstico dos modelos STARMA

Como visto anteriormente, a obtenção de modelos STARMA, por meio do pacote ‘starma’ é realizada através das seguintes etapas: Identificação, estimação e diagnóstico do modelo. Nesta subseção será apresentado o resultado de cada uma das três etapas, para os dois modelos STARMA de interesse.

i-Identificação:

Antes da identificação da ordem dos modelos foi aplicado a extensão do teste de Box-Pierce para séries espaço-temporais. Este teste está implementado por meio da função ‘*scor.teste()*’ e, dentre outras utilidades, serve para verificar a existência de autocorrelação espaço-temporal no conjunto de dados.

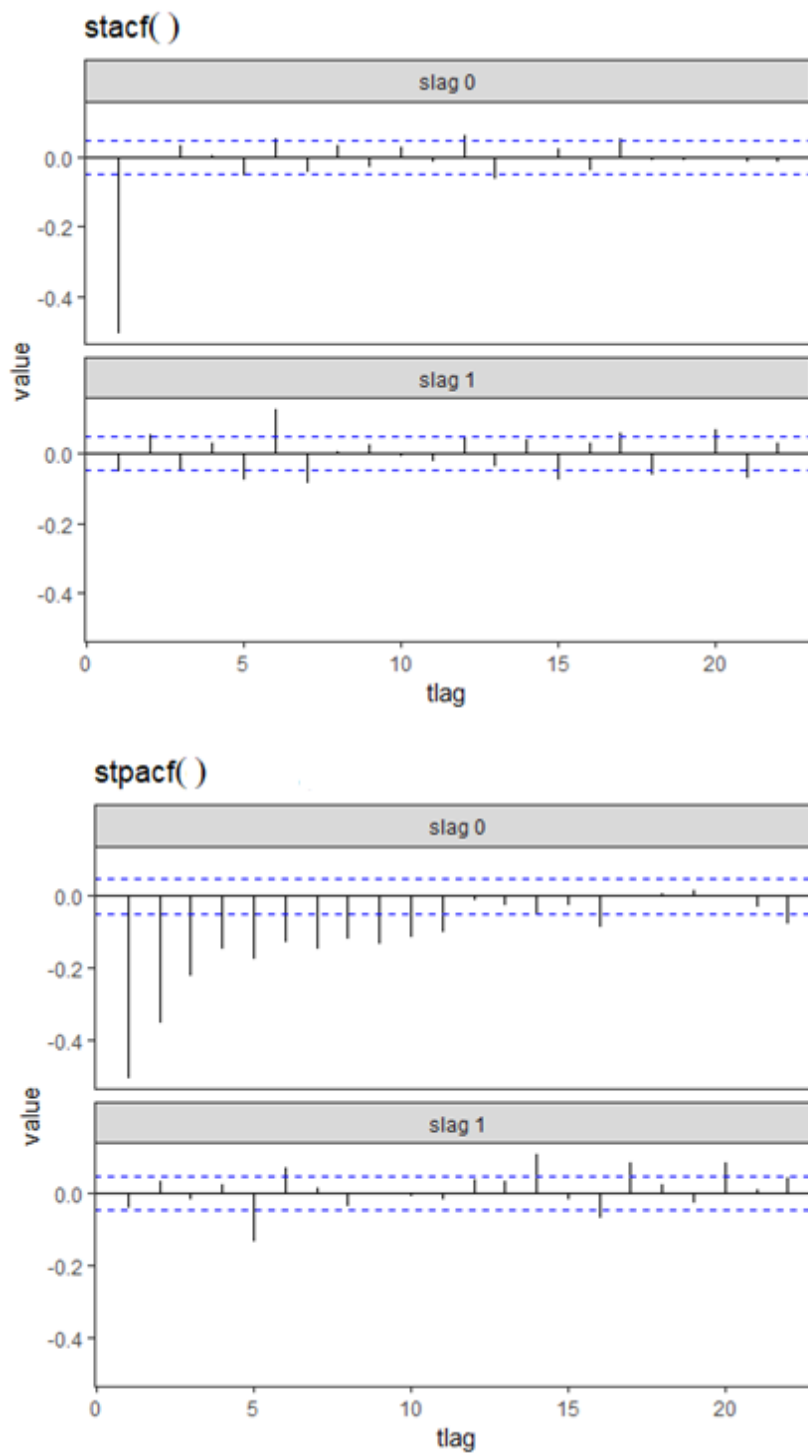
Para ambas as matrizes de vizinhança obtidas o teste indicou que a hipótese de não correlação deve ser rejeitada. Portanto conclui-se que os dados apresentam correlação espaço-temporal. É importante enfatizar que este teste trata-se de uma extensão empírica do teste de Box-Pierce (CHEYSSON, 2016).

Por meio das funções '*stacf()*' e '*stpacf()*' foram obtidos, respectivamente, o gráfico da função de auto correlação espaço-temporal (*stacf*) e o gráfico da função de auto correlação parcial espaço-temporal (*stpacf*). Estes gráficos servem como ferramenta de identificação da ordem do modelo e são exibidos nas figuras 5 e 6. Na Figura 5 são exibidos os gráficos para a matriz de vizinhança baseada no investimento anual médio na saúde básica, enquanto que na Figura 6 tem-se os gráficos da *stacf* e *stpacf* para a matriz **W** baseada no IDH. Note que são exibidos os gráficos das funções para vizinhança de ordem nula (*slag 0*) e de primeira ordem (*slag 1*).

Os gráficos das Figuras 5 e 6, em concordância com o resultado do teste espaço-temporal de Box-Pierce, indicam a existência de correlação espaço-temporal. Observe a existência de auto correlação temporal (*slag 0*) e a existência de correlação espaço-temporal considerando-se a primeira ordem de vizinhança espacial (*slag1*). Desta maneira, entende-se que existe alguma relação entre a taxa de incidência de TB das nove cidades. Este resultado indica que, diferente do que se imaginava, a taxa de incidência de tuberculose de uma região não influencia apenas suas proximidades.

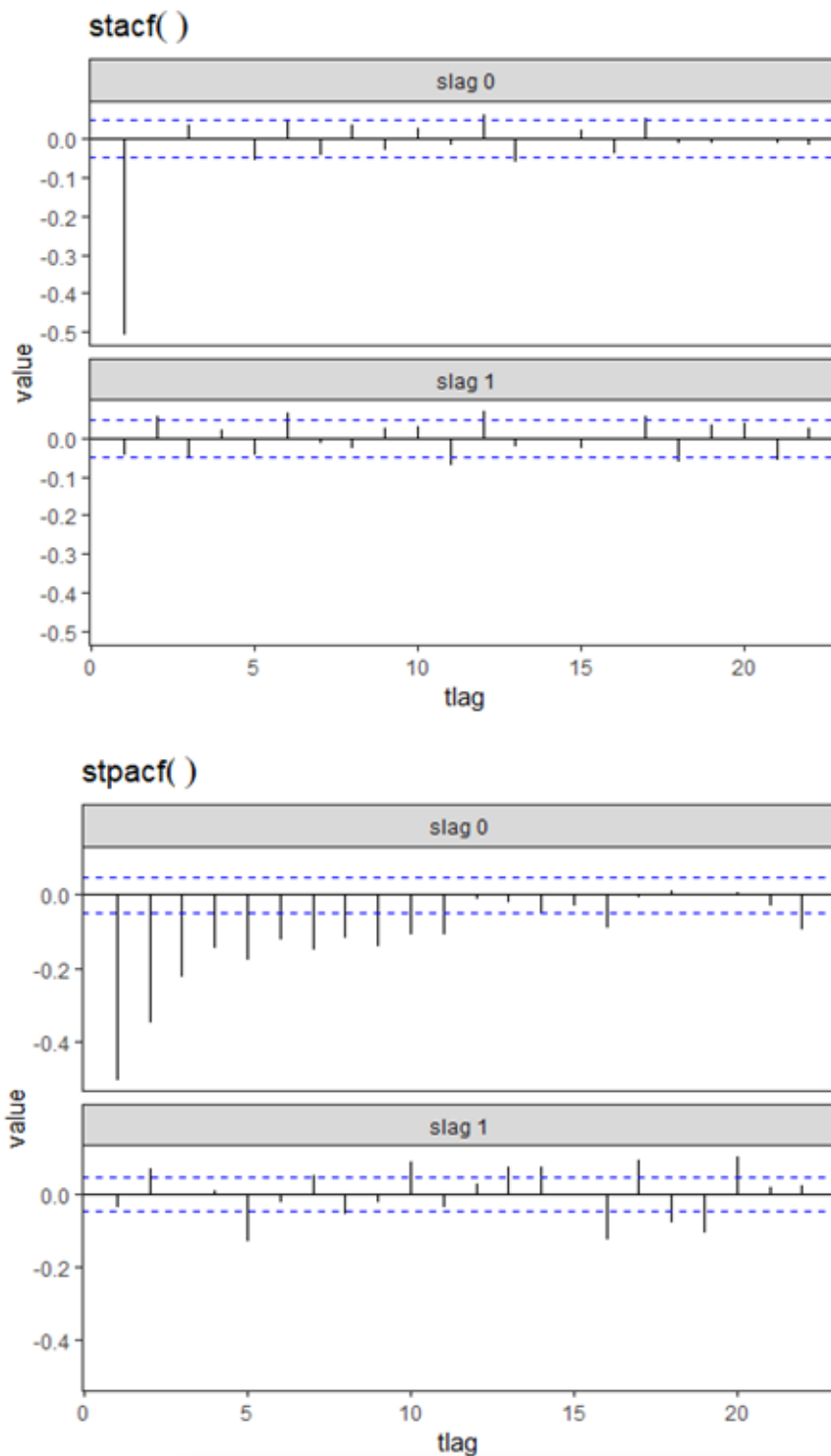
A existência de correlação espaço-temporal, observada no gráfico da função de autocorrelação espaço-temporal (*slag 1*), permite concluir que, por exemplo, a taxa de incidência de tuberculose observada em Belo Horizonte relaciona-se não somente com a taxa de cidades próximas como Betim e Contagem, mas também com a taxa de incidência de tuberculose das cidades mais distantes, tais como Montes Claros, Uberlândia.

Figura 5- Gráficos stacf e stpacf para matriz de vizinhança baseada no investimento anual médio em saúde básica.



Fonte: Do autor (2019)

Figura 6- Gráficos da *stacf* e *stpacf* para matriz de vizinhança baseada no IDH.



Fonte: Do autor (2019)

Para ambas as matrizes de vizinhança socioeconômicas é observado decaimento exponencial no gráfico da função *stpacf*, considerando-se a ordem de vizinhança igual a zero, isto é, considerando-se apenas a correlação temporal. Isto sugere o ajuste de modelos da

classe STMA. No entanto, como o decaimento exponencial, em ambos os casos, não é observado para vizinhança de ordem 1, pode ser que o ajuste de modelos STARMA seja necessário.

Inicialmente ajustou-se um modelo STMA (5_1), tal denotação indica que foram tomadas cinco defasagens no tempo e uma defasagem no espaço. Na tentativa de obter modelos mais adequados, foram ajustados modelos da classe STARMA. Por questões de parcimônia, optou-se por tentar inicialmente o ajuste de um modelo de ordem 2, isto é, um modelo STARMA ($2_1, 2_1$).

ii-Estimação e diagnóstico do modelo:

Na etapa de estimação são obtidas as estimativas dos parâmetros do modelo por meio do filtro de Kalman. Por este motivo é necessário definir o número de iterações que serão necessárias para que as estimativas dos parâmetros converjam. De acordo com Cheysson (2016), o filtro de Kalman tem se mostrado muito eficiente na estimação dos parâmetros de modelos STARMA, de maneira que não são necessárias muitas iterações para atingir a convergência da estimativa dos parâmetros.

A definição do número de iterações foi feita empiricamente. Foram testados os seguintes números de iterações: 2, 5, 10, 50. Observou-se que 2 iterações foram suficientes para que houvesse convergência até a quinta casa decimal. Para analisar se houve ou não convergência foram observados os valores estimados dos parâmetros e o valor do critério de informação bayesiano (BIC).

Por meio da função '*starma()*' ajustaram-se os modelos considerando-se duas iterações, as matrizes **W** obtidas, a ordem autorregressiva e ordem de médias móveis. Na Tabela 8 é exibido o resultado do ajuste para o caso da matriz **W** baseada no investimento anual médio em saúde básica. Na Tabela 9 é apresentado o resultado para o ajuste considerando a matriz **W** obtida através do IDH.

Tabela 8- Resultado do ajuste para o modelo STARMA ($6_1, 6_1$) considerando-se duas iterações e W baseada no investimento anual médio em saúde básica.

PARÂMETRO	ESTIMATIVA (ERRO PADRÃO)	valor-p
$\phi_{5,0}$	-0,068844 (0,018883)	0,000275
$\phi_{6,1}$	0,241055 (0,041939)	$1,07 \times 10^{-8}$
$\theta_{1,0}$	-0,830716 (0,025199)	$< 2,20 \times 10^{-16}$
$\theta_{6,1}$	-0,119665 (0,055938)	0,032558

Fonte: Do autor (2019)

Tabela 9- Resultado do ajuste para o modelo STARMA ($6_1, 1_1$) considerando-se duas iterações e W baseada no IDH municipal.

PARÂMETRO	ESTIMATIVA (ERRO PADRÃO)	valor-p
$\phi_{5,0}$	-0,067049 (0,018669)	0,000338
$\phi_{2,1}$	0,172787 (0,036590)	$2,52 \times 10^{-6}$
$\phi_{6,1}$	0,132327 (0,036384)	0,000284
$\theta_{1,0}$	-0,893149 (0,025600)	$< 2,20 \times 10^{-16}$

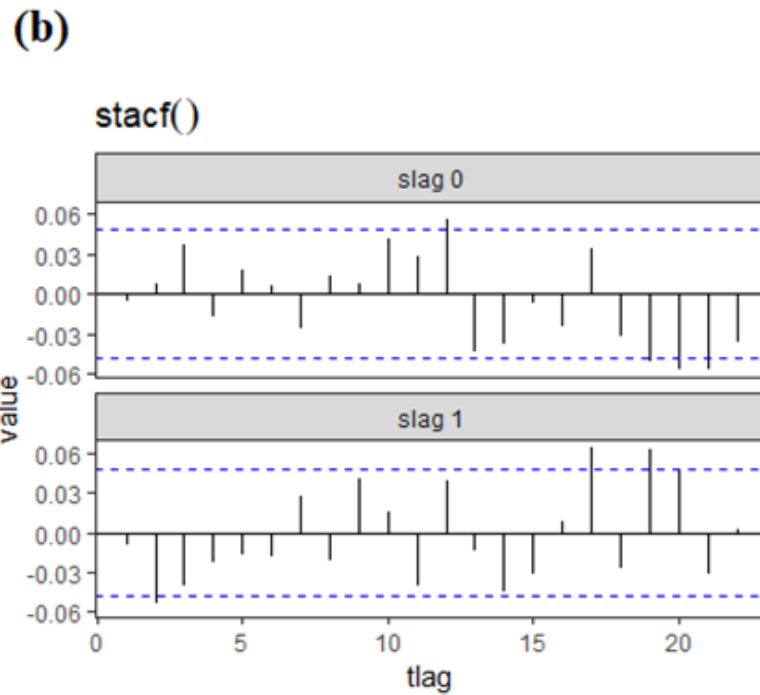
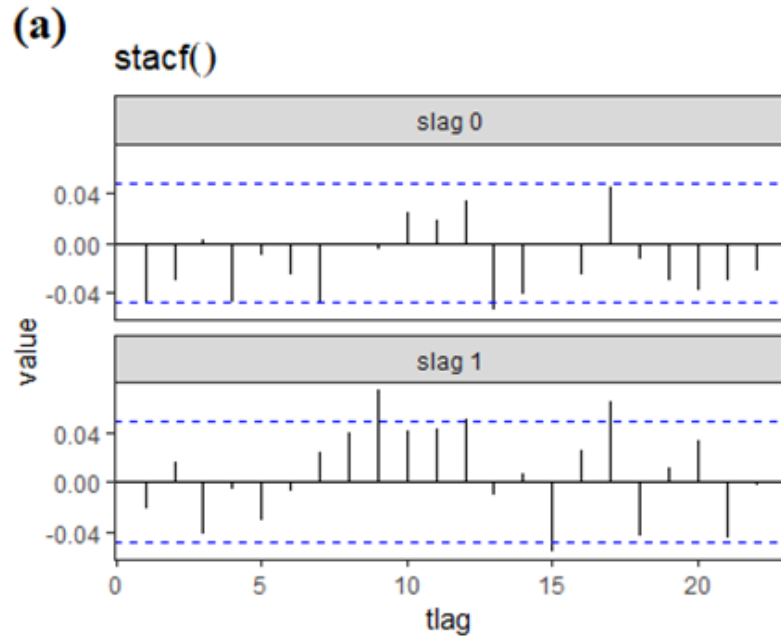
Fonte: Do autor (2019)

Uma vez que modelos autorregressivos são modelos que se descrevem em função de seus próprios valores antecedentes. A estimativa de coeficientes autorregressivos de sexta ordem, para ambos modelos, sugere que a taxa de incidência de TB em um dado mês é relacionada a taxa de incidência dos últimos seis meses.

Uma maneira de se diagnosticar a adequação do modelo ajustado, consiste em verificar se os resíduos do modelo apresentam comportamento semelhante a um ruído branco. Para isto é necessário verificar se os resíduos são independentes e tem distribuição gaussiana multivariada com média nula e variância constante. Outra maneira de se fazer este diagnóstico é através da análise do gráfico da *stacf*. Na Figura 7 são exibidos os gráficos da *stacf* para as duas matrizes de vizinhança socioeconômicas: No canto superior (Figura 7 (a)) é exibido o gráfico da *stacf* considerando-se a matriz **W** baseada no investimento anual médio em saúde básica. É possível ver que mesmo após o ajuste dos modelos os resíduos ainda apresentam auto correlação. O mesmo ocorre para **W** baseada no IDH (Figura 7 (b)).

Observe a existência de correlação estatisticamente significativas a medida que se torna maior a defasagem temporal. Isto sugere que as séries temporais que compõem a série espaço-temporal apresentam memória longa. Por desconhecer modelos espaço-temporais que se apliquem a processos com memória longa optou-se pelo ajuste dos modelos, já apresentados nas tabelas 8 e 9.

Figura 7- Gráficos da função de auto correlação residual após o ajuste dos modelos.



Fonte: Do autor (2019)

7 CONSIDERAÇÕES FINAIS

Durante o desenvolvimento deste trabalho foram depreendidas muitas conclusões, não apenas em relação às matrizes de vizinhança socioeconômicas ou a classe de modelos STARMA. Mas também sobre a propagação da Tuberculose e sua relação com variáveis econômicas, ambientais e sociais.

No que se refere às matrizes de vizinhança socioeconômicas e o seu impacto sobre modelos da classe STARMA, resultados inesperados ocasionaram divergências que poderão ser interessantes a estudos futuros. Diante do conjunto de dados original, composto pelas nove séries temporais, foi observada a existência de autocorrelação espaço-temporal. No entanto, não foi possível obter modelos STARMA capazes de modelar completamente a estrutura de autocorrelação espaço-temporal presente nos dados. Em síntese, tal divergência ocasionou os seguintes questionamentos: O conjunto de dados (a taxa de incidência de tuberculose) apresenta mesmo autocorrelação espaço-temporal? Seriam as matrizes de vizinhança socioeconômicas o motivo que impossibilitou o ajuste dos modelos STARMA?

Os padrões e a tendência espaço-temporal da incidência de tuberculose já foram tratados em muitos outros trabalhos, como foi visto na revisão bibliográfica, o que corrobora com a suposição de existência de autocorrelação espaço-temporal no conjunto de dados. Além disto, a autocorrelação espaço-temporal presente no conjunto de dados foi verificada de duas maneiras: através do gráfico da função de autocorrelação espaço-temporal e por meio da extensão espaço-temporal do teste de Box-Pierce. O resultado do teste evidencia de maneira clara e direta que a hipótese de não correlação deve ser rejeitada. Apesar de parecer inquestionável, são necessários estudos mais aprofundados sobre a extensão espaço-temporal do teste de Box- Pierce como destaca Cheysson.:

O teste pode ser aplicado a uma série espaço-temporal para testar a hipótese nula de não correlação. É útil verificar se os resíduos de um modelo STARMA são ruído branco multivariado ... Por favor, note que esta é uma extensão empírica e ainda não foi objeto de um trabalho. As especificações das matrizes de peso também não foram estudadas e podem levar a inconsistências (CHEYSSON, 2016, p.10).

Quanto às matrizes de vizinhança socioeconômicas aplicadas a modelos da classe STARMA, não é possível afirmar que são apropriadas ou inapropriadas. É possível, baseado neste trabalho, afirmar que tais matrizes de vizinhança são adequadas na construção de modelos STARMA que se ajustarão a taxa de incidência de tuberculose, uma vez que sabe-se

que esta patologia é fortemente ligada a variáveis socioeconômicas. É mais provável que a estrutura de autocorrelação observada no conjunto de dados, não tenha sido completamente ajustada devido a escolha das equações de ponderação dos pesos do que devido às variáveis socioeconômicas obtidos. No entanto, diante do conjunto de dados utilizado neste trabalho não havia a possibilidade de utilização de matrizes de vizinhança baseada em critérios geográficos, o que enaltece ainda mais a conveniência da utilização de matrizes de vizinhanças socioeconômicas.

Por enquanto os questionamentos anteriormente apresentados continuam em aberto, sem uma resposta clara e completamente consistente. Para que possam ser inteiramente respondidos são necessários mais estudos, não somente sobre matrizes de vizinhança socioeconômicas, mas também sobre modelos STARMA e sobre a extensão espaço-temporal do teste de Box-Pierce.

Ainda assim, por meio deste trabalho foi possível concluir que a autocorrelação espaço-temporal pode ser observada ao se considerar a taxa de incidência de tuberculose para cidades geograficamente distantes. Esta conclusão estimula novas pesquisas sobre a utilização de matrizes socioeconômicas em modelos STARMA, que buscam explicar a taxa de propagação diferentes epidemias. Além de explicitar a correlação existente entre variáveis socioeconômicas e a propagação da tuberculose.

REFERÊNCIAS

- ALMEIDA, E. **Econometria espacial aplicada**. Campinas–SP. Alínea, 2012.
- AMÂNCIO, C. T.; NASCIMENTO, L. F. C. Asma e poluentes ambientais: um estudo de séries temporais. **Revista da Associação Médica Brasileira**, v. 58, n. 3, p. 302-307, 2012.
- ANTUNES, J. L. F.; WALDMAN, E. A. A tuberculose através do século: séries temporais para a mortalidade em São Paulo, Brasil, 1900-1997. **Cadernos de Saúde Pública**, v. 15, n. 3, p. 463-476, 1999.
- 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.
- BAUMONT, C. **Spatial effects in housing price models: do housing prices capitalize urban development policies in the agglomeration of Dijon (1999)?**. Tese de Doutorado. Laboratoire d'économie et de gestion (LEG), 2004.
- BERTOLOZZI, M. R. et al. O controle da tuberculose: um desafio para a saúde pública. **Revista de Medicina**, v. 93, n. 2, p. 83-89, 2014.
- 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.
- BIAZEVIC, M. G. H. et al. Tendências de mortalidade por câncer de boca e orofaringe no Município de São Paulo, Brasil, 1980/2002. **Cadernos de Saúde Pública**, v. 22, p. 2105-2114, 2006.
- BIZ, G. **Simulações de pesos espaciais para o modelo STARMA e aplicações**. Tese de Doutorado. Universidade de São Paulo, 2014
- BOX, G. E. P; COX, D. R. An analysis of transformations. **Journal of the Royal Statistical Society. Series B (Methodological)**, p. 211-252, 1964.
- 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 Saude Publica**, 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.

- BRESSAN, A. A. Tomada de decisão em futuros agropecuários com modelos de previsão de séries temporais. **RAE-eletrônica**, v. 3, n. 1, 2004
- BROCKWELL, P. J.; DAVIS, R. A.; CALDER, M. V. **Introduction to time series and forecasting**. New York: Springer, 2002.
- 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.
- 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 JUNIOR, L. M. et al. **Análise longitudinal dos preços do carvão vegetal, no Estado de Minas Gerais**. **Revista Árvore**, Viçosa, v. 30, n. 3, p. 429-438, maio/jun, 2006.
- COELHO JUNIOR, L. M. et al. Análise temporal da Borracha Natural Brasileira. **Cerne**, Lavras, v. 15, n. 1, p. 19-26, jan./mar, 2009.
- 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.
- COTTRELL, Allin; LUCCHETTI, Riccardo. Gretl. Version 2017d. 2017. Disponível em: <http://gretl.sourceforge.net/>.
- DEPARTAMENTO DE INFORMÁTICA DO SISTEMA ÚNICO DE SAÚDE (DATASUS). Brasília, 2018. Disponível em: <http://tabnet.datasus.gov.br/cgi/tabcgi.exe?sinannet/cnv/tubercbr.def>.>. Acessado em: Agosto de 2018
- DUCATI, R. G. et al. The resumption of consumption: a review on tuberculosis. **Memórias do Instituto Oswaldo Cruz**, v. 101, n. 7, p. 697-714, 2006.
- FANLO, P.; TIBERIO, G. Tuberculosis extrapulmonar. In: **Anales del sistema sanitario de Navarra**. Gobierno de Navarra. Departamento de Salud, 2007. p. 143-162.
- FARIAS, H. P.; SÁFADI, T. Causalidade entre as principais bolsas de valores do mundo. **Revista de administração Mackenzie**. V. 11, n2. São Paulo. 2010
- FERREIRA, A. 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, C. et al. Internações e óbitos e sua relação com a poluição atmosférica em São Paulo, 1993 a 1997. **Revista de Saúde Pública**, v. 38, p. 751-757, 2004.

- GARRIDO, R.; MAHMASSANI, H. Forecasting short-term freight transportation demand: Poisson STARMA model. **Transportation Research Record: Journal of the Transportation Research Board**, n. 1645, p. 8-16, 1998.
- GAGNEUX, S. Ecology and evolution of Mycobacterium tuberculosis. **Nature Reviews Microbiology**, v. 16, n. 4, p. 202, 2018.
- GIACOMINI, R.; GRANGER, C. W. J. **Aggregation of space-time process. Journal of Econometrics**, Califórnia, n. 118, p. 7- 26, 2004
- GLASBEY, C. A.; ALLCROFT, D. J. A spatio temporal auto-regressive moving average model for solar radiation. **Biomathematics and Statistics Scotland**, n. 57, p. 143- 154, 2008
- GRUDNITSKI, G.; OSBURN, L. **Forecasting S&P and Gold Futures Prices: An Application of Neural Networks, in.: The Journal of Futures Markets**. Vol. 13, nº 6, p.631-643, 1993.
- 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.
- 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.
- JACQUEZ, G. M. Disease cluster statistics for imprecise space-time locations. **Statistics in Medicine**, v. 15, n. 7-9, p. 873-885, 1996.
- 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.
- JORGE, L. A. B.; AOKI, H.; PONTINHA, A. de A. S. **Relação entre a Precipitação Anual e o Incremento em Volume de Árvores Amostradas em Plantações de Pinus taeda**. Florestas e ambientes, 2012.
- KAMARIANAKIS, Y.; PRASTACOS, P. Space-time modeling of traffic flow. **Computers&Geosciences**, v. 31, n. 2, p. 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.
- LEEMANS, J. C. et al. Depletion of alveolar macrophages exerts protective effects in pulmonary tuberculosis in mice. **The Journal of Immunology**, v. 166, n. 7, p. 4604-4611, 2001.

MACIEL, Ethel Leonor Noia; SALES, Carolina Maia Martins. A vigilância epidemiológica da tuberculose no Brasil: como é possível avançar mais?. **Epidemiologia e Serviços de Saúde**, v. 25, p. 175-178, 2016.

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.

MINISTÉRIO DA SAÚDE. Brasília, 2018. Disponível em: <<http://portalms.saude.gov.br/saude-de-a-z/tuberculose>>. Acessado em 12 de Setembro de 2018.

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

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.

OLIVEIRA, R. A.; BUONGIORNO, J.; KMIOTEK, A. M. **Time series forecasting models of lumber cash, futures, and basis prices**. **Forest Science**, Lawrence, v. 23, n. 2, p. 269-279, 1977.

OTERO, U. B.; ROZENFELD, S.; GADELHA, A. J. Óbitos por desnutrição em idosos, São Paulo e Rio de Janeiro: análise de séries temporais. 1980-1996. **Revista Brasileira de Epidemiologia**, v. 4, p. 191-205, 2001.

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.

- PERES, M. F. T. et al. Evolução dos homicídios e indicadores de segurança pública no Município de São Paulo entre 1996 a 2008: um estudo ecológico de séries temporais. **Ciência & Saúde Coletiva**, v. 17, p. 3249-3257, 2012.
- 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.
- 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 (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- RAMAM, M. et al. How useful is the Mantoux test in the diagnosis of doubtful cases of cutaneous tuberculosis?. **International journal of dermatology**, v. 50, n. 11, p. 1379-1382, 2011.
- 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, 4ª edição, pag 379-448, 1995
- RODRIGUES-JÚNIOR, 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-JÚNIOR, 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.
- 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.
- 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.
- 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.

SILVA, M. L. da; SILVA, J. M. A. da. Análise do comportamento temporal dos preços do carvão vegetal: aplicação e avaliação da metodologia “Box e Jenkins”. **Revista Árvore**, Viçosa, v. 20, n. 1, p. 57-67, 1996.

SILVA, Ana Hermínia Andrade. **Testes escore para transformação de dados em regressões lineares**. Dissertação de mestrado. Universidade Federal do Pernambuco, 2013.

SILVA, S. B. **Análise de séries temporais: Um estudo de caso aplicado a casos de hanseníase no município de Ji-Paraná/RO**. Trabalho de conclusão de curso, Universidade Federal de Rondônia, 2015

SMITH, T. E. **Spatial weight matrices**. Disponível em: <http://www.seas.upenn.edu/~ese502/lab-content/extra_materials/>. Acessado 01 de março de 2018.

SOARES, N. S. et al. Análise de previsões do preço da borracha natural no Brasil. **Scientia Forestalis**, Piracicaba, v. 36, n. 80, p. 285-294, dez, 2008.

SONG, N. **Structural and forecastings of twoodlumber models with a time series approach**. 2003. Disponível em: <http://etd.lsu.edu/docs/available/etd-07122006-093251/unrestricted/Song_dis.pdf>. Acesso em: Janeiro de 2018.

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.

SZWARCOWALD, 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.

TOMÉ, E. A. et al. Tendências da mortalidade infantil no Município de Guarulhos: análise do período de 1971 a 1998. **Revista Brasileira de Epidemiologia**, v. 4, p. 153-167, 2001.

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

WATTS, G. Who annual report finds world at a crossroad on tuberculosis. **BMJ: British Medical Journal (Online)**, v. 345, 2012.

WIKIÉDIA. Disponível em: <https://pt.wikipedia.org/wiki/Lista_de_munic%C3%ADpios_de_Minhas_Gerais_por_IDH-M>. Acessado 04 de dezembro de 2018.

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.

APÊNDICE A- Estimativas populacionais obtidas.

1- Belo Horizonte

ANO	POPULAÇÃO ESTIMADA ATRAVÉS DO MODELO DE REGRESSÃO
2002	2292422
2003	2326972
2004	2361522
2005	2396072
2006	2430622
2007	2465172
2008	2499722
2009	2534272
2010	2568822
2011	2603372
2012	2637922
2013	2672472
2014	2707022
2015	2741572
2016	2776122
2017	2810672

2- Betim

ANO	POPULAÇÃO ESTIMADA ATRAVÉS DO MODELO DE REGRESSÃO
2002	320157
2003	327298
2004	334439
2005	341580
2006	348721
2007	355862
2008	363003
2009	370144
2010	377285
2011	384426
2012	391567
2013	398708
2014	405847
2015	412990
2016	420131
2017	427272

3- Contagem

ANO	POPULAÇÃO ESTIMADA ATRAVÉS DO MODELO DE REGRESSÃO
2002	550101
2003	556643
2004	563185
2005	569727
2006	567269
2007	582811
2008	589353
2009	595895
2010	602437
2011	608979
2012	615521
2013	622063
2014	628605
2015	635147
2016	641689
2017	648231

4- Governador Valadares

ANO	POPULAÇÃO ESTIMADA ATRAVÉS DO MODELO DE REGRESSÃO
2002	249999
2003	252551
2004	255103
2005	257655
2006	260207
2007	262759
2008	265311
2009	267864
2010	270416
2011	272967
2012	275520
2013	278072
2014	280624
2015	283176
2016	285728
2017	288280

5- Juiz de Fora

ANO	POPULAÇÃO ESTIMADA ATRAVÉS DO MODELO DE REGRESSÃO
2002	456394
2003	462591
2004	468788
2005	474985
2006	481182
2007	487379
2008	493576
2009	499773
2010	505970
2011	512167
2012	518364
2013	524561
2014	530758
2015	536955
2016	543152
2017	5493498

6- Lavras

ANO	POPULAÇÃO ESTIMADA ATRAVÉS DO MODELO DE REGRESSÃO
2002	81126
2003	82189
2004	83252
2005	84315
2006	85378
2007	86441
2008	87504
2009	88567
2010	90693
2011	91756
2012	92819
2013	93882
2014	94945
2015	95008
2016	97071
2017	99134

7- Montes Claros

ANO	POPULAÇÃO ESTIMADA ATRAVÉS DO MODELO DE REGRESSÃO
2002	313072
2003	318608
2004	324144
2005	329680
2006	335216
2007	340752
2008	346288
2009	351824
2010	357360
2011	362896
2012	368432
2013	373968
2014	379504
2015	385040
2016	390576
2017	396112

8- Pouso Alegre

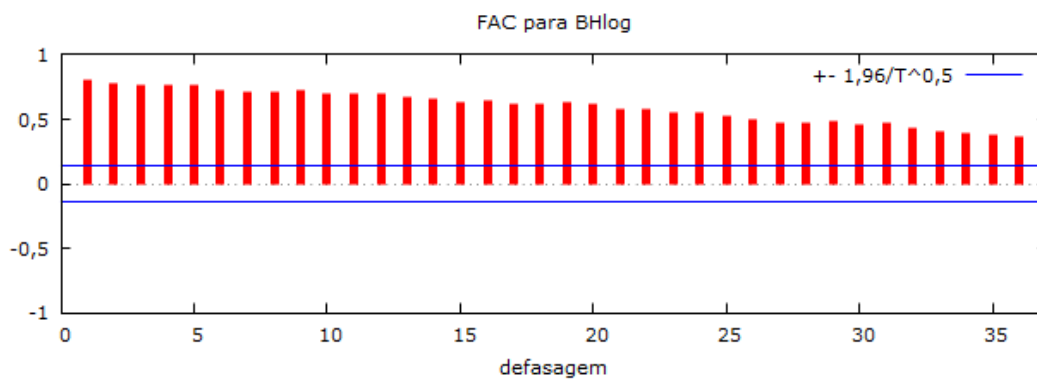
ANO	POPULAÇÃO ESTIMADA ATRAVÉS DO MODELO DE REGRESSÃO
2002	111744
2003	114128
2004	116512
2005	118896
2006	121280
2007	123664
2008	126048
2009	128432
2010	130816
2011	133200
2012	135584
2013	137968
2014	142736
2015	143953
2016	145120
2017	147504

9- Uberlândia

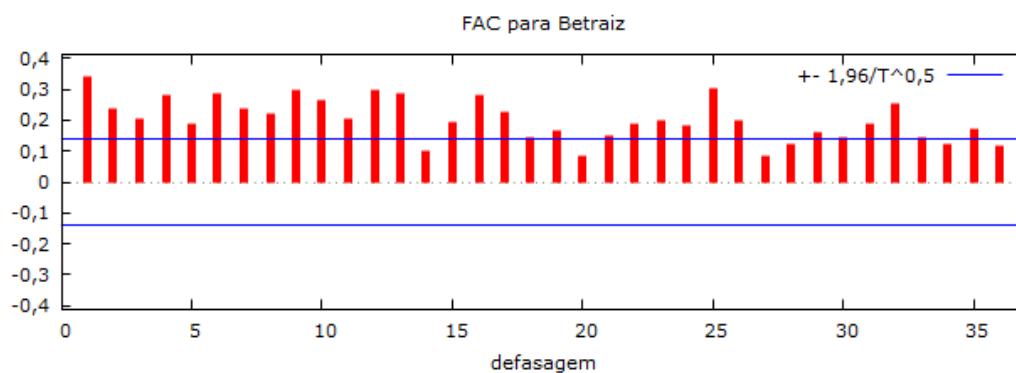
ANO	POPULAÇÃO ESTIMADA ATRAVÉS DO MODELO DE REGRESSÃO
2002	514160
2003	526240
2004	538320
2005	550400
2006	562480
2007	574560
2008	586640
2009	598720
2010	610800
2011	622880
2012	635960
2013	647040
2014	659120
2015	671200
2016	683280
2017	695360

APÊNDICE B- Gráfico da função de auto correlação das séries temporais após transformações apropriadas.

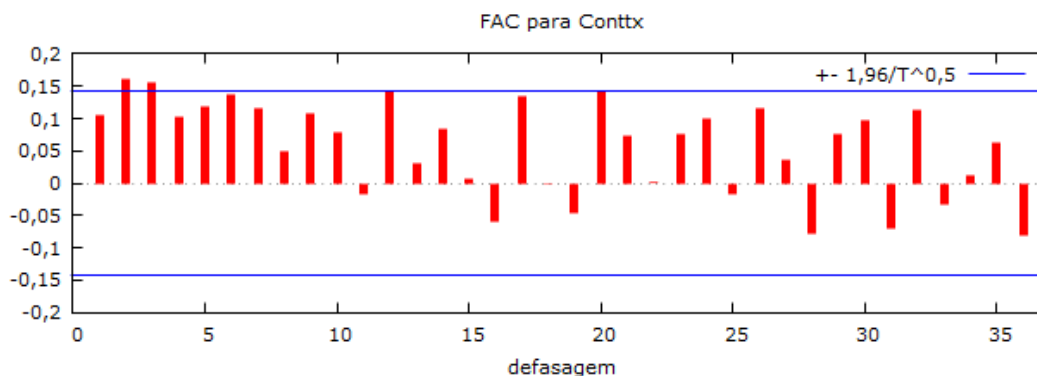
Belo Horizonte



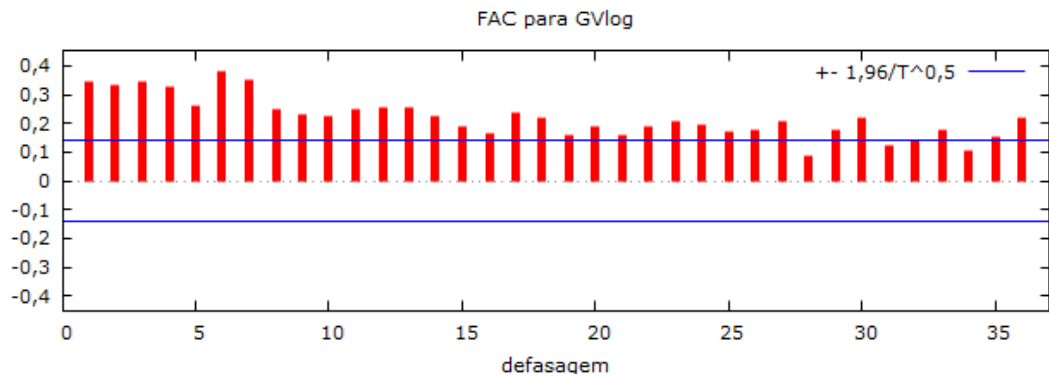
Betim



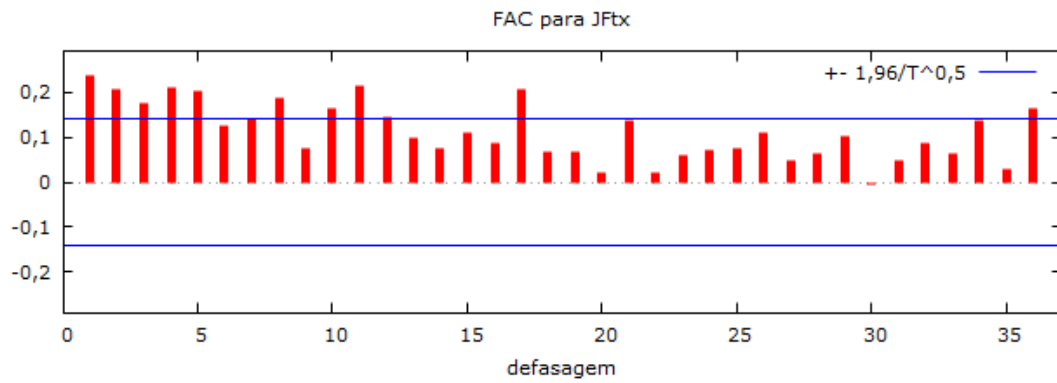
Contagem



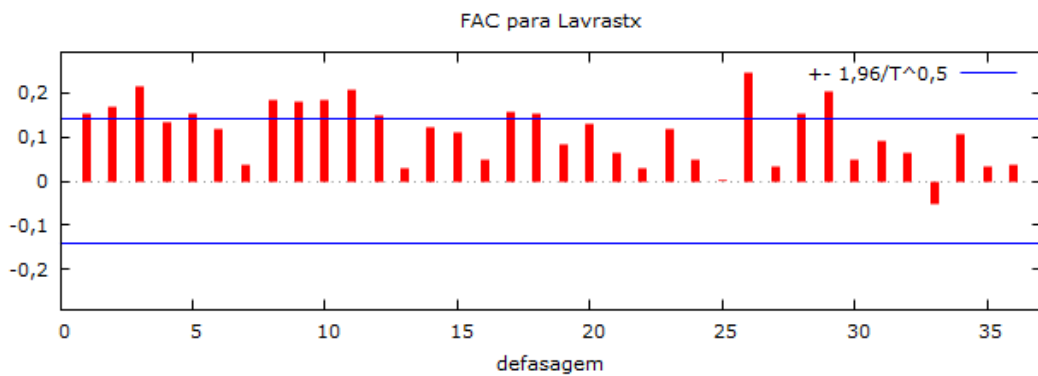
Governador Valadares



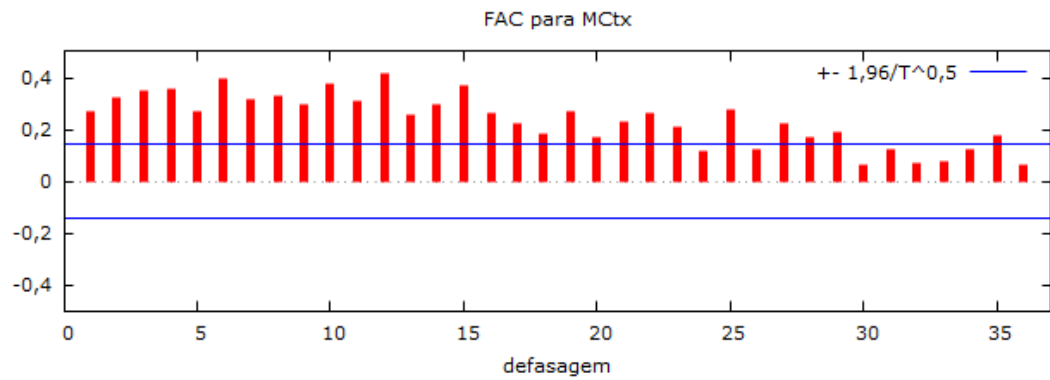
Juiz de Fora



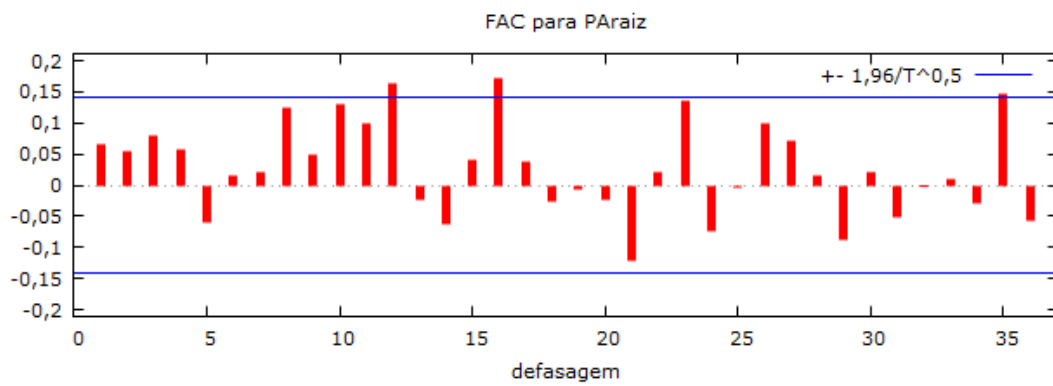
Lavras



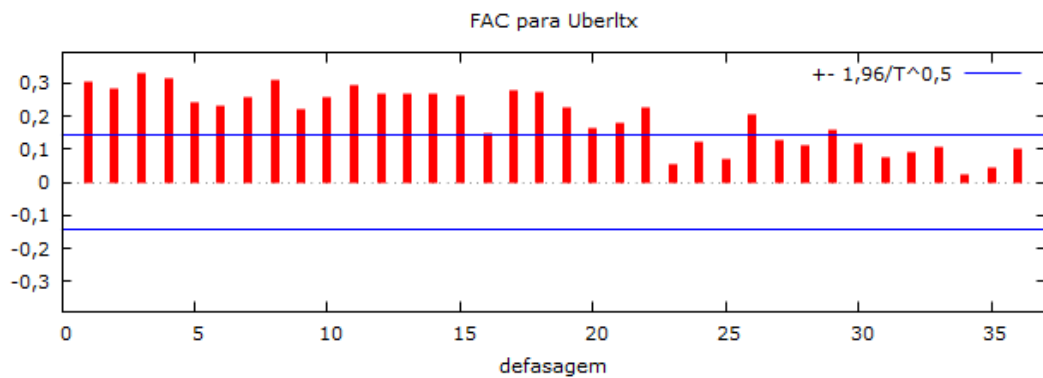
Montes Claros



Pouso Alegre

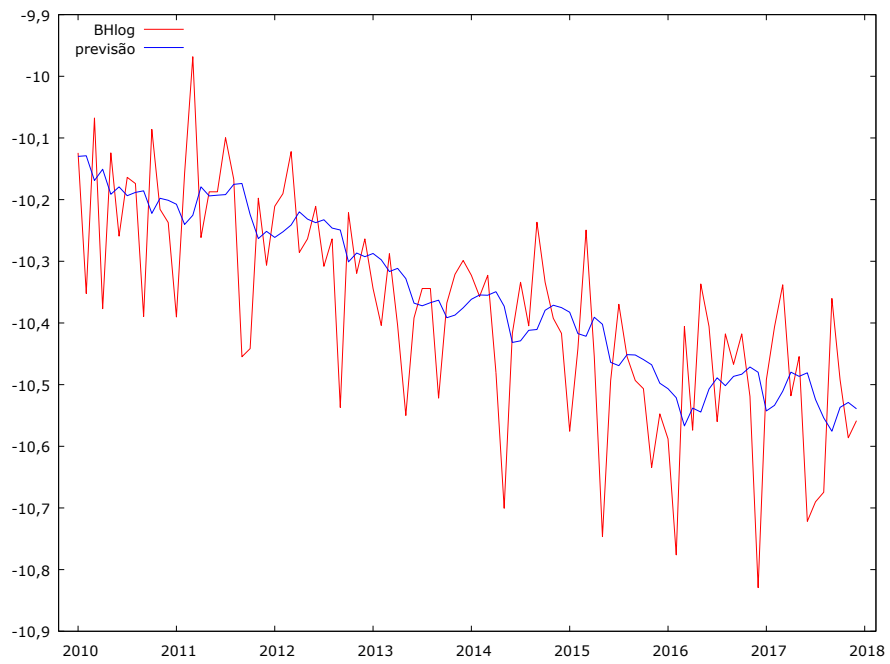


Uberlândia

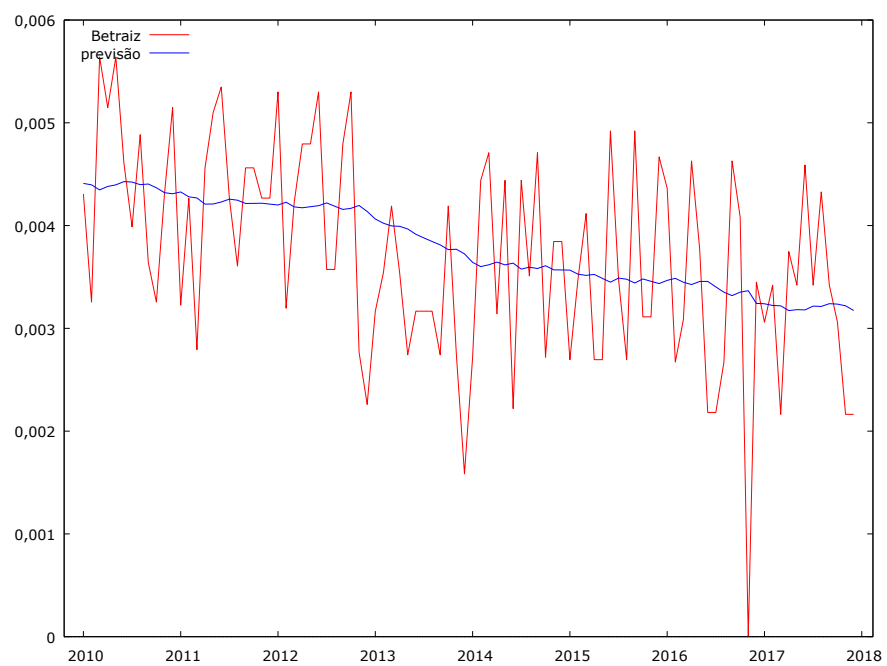


APÊNDICE C- Gráfico das séries temporais e das séries ajustadas

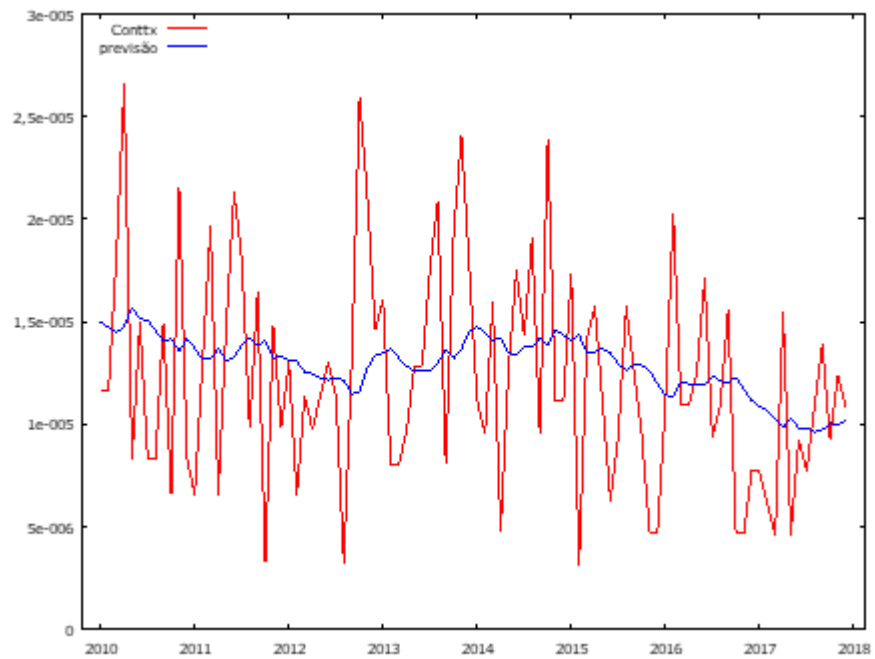
Belo Horizonte



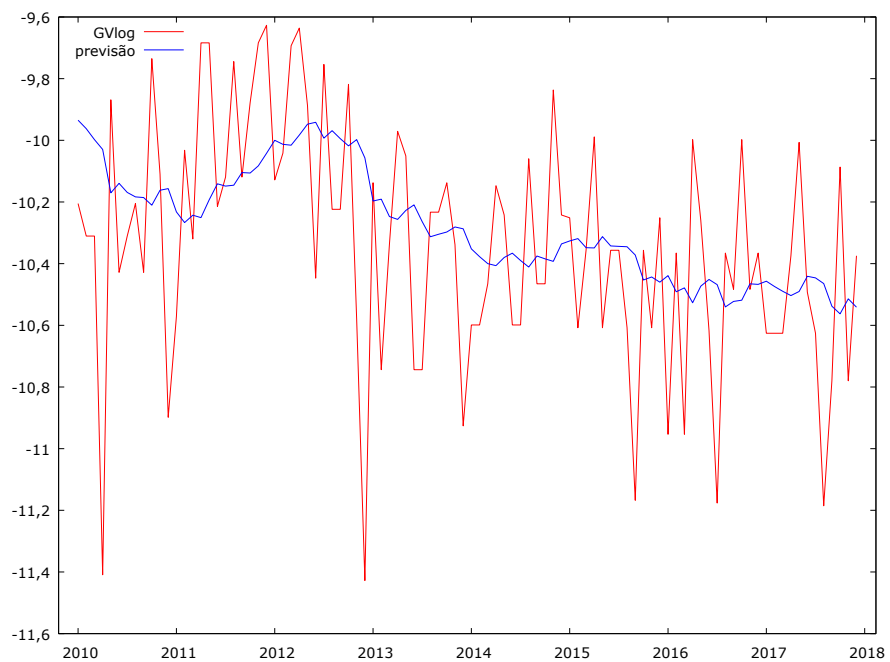
Betim



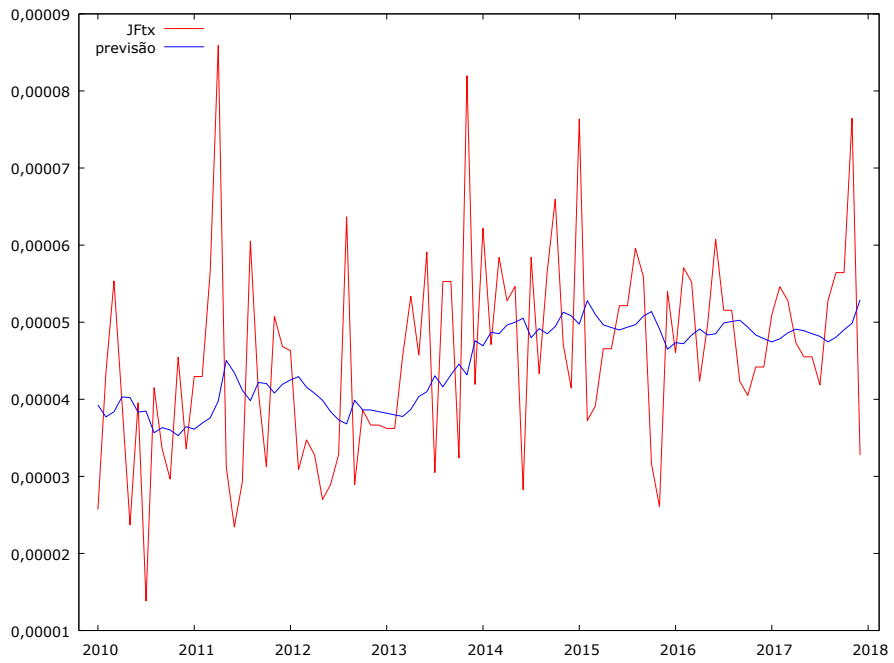
Contagem



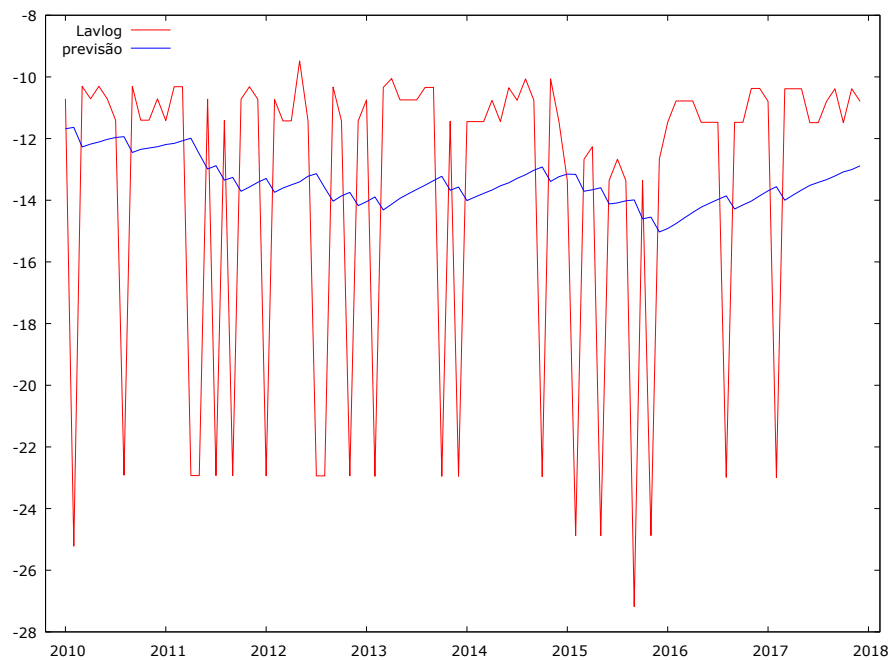
Governador Valadares



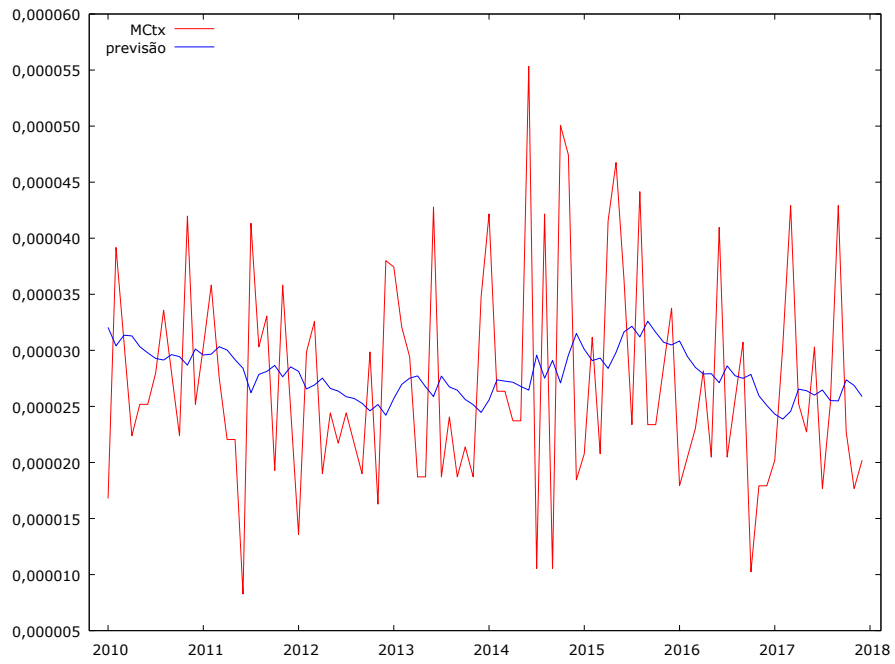
Juiz de Fora



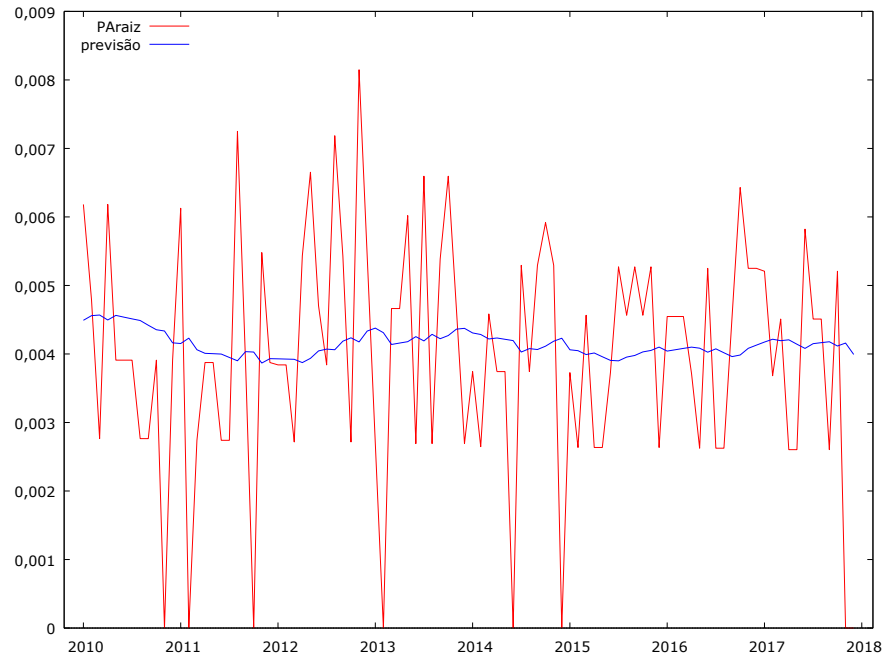
Lavras



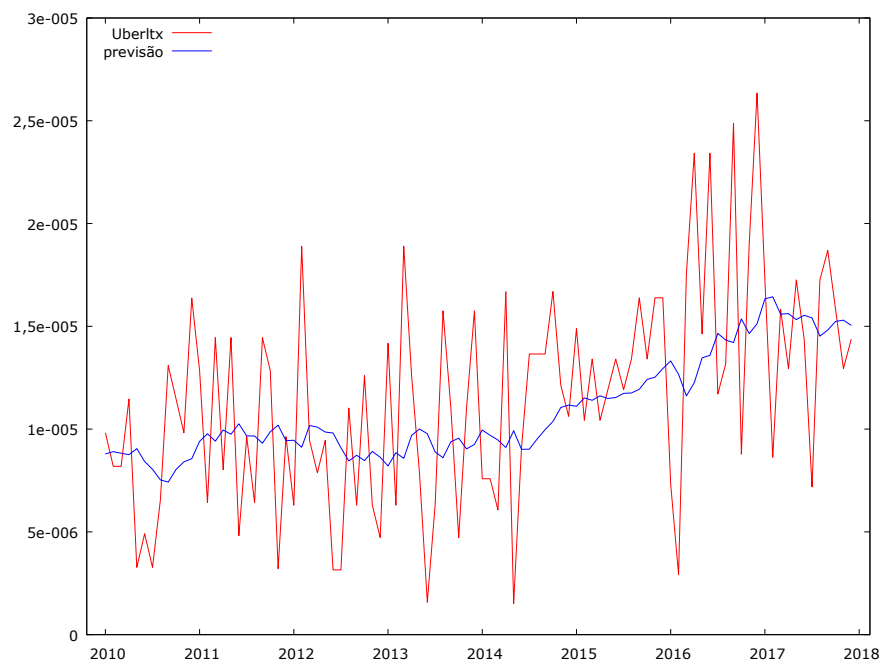
Montes Claros



Pouso Alegre



Uberlândia



APÊNDICE D- Código computacional utilizado na identificação, estimação e no diagnóstico do modelo.

```
#PRIMEIRO MODELO
#Modelo considerando-se W baseada no IDH
dados = read.table("dadosDifTx.csv",sep = ";",dec=".", h = T) #Entrada de dados
o0=diag(9)
o1 = matrix(c(0,0.0429,0.0424,0.0633,0.0369,0.0493,0.0369,0.0332,0.1009,
0.0765,0,0.3268,0.2387,0.0407,0.0419,0.0702,0.0478,0.0529,
0.0864,0.3739,0,0.1811,0.0537,0.0531,0.1053,0.0664,0.0642,
0.0562,0.1189,0.0789,0,0.0231,0.0251,0.0343,0.0375,0.0342,
0.1458,0.0902,0.1040,0.1030,0,0.3454,0.1843,0.3613,0.1925,
0.1666,0.0793,0.0880,0.0955,0.2953,0,0.1229,0.1807,0.3026,
0.1166,0.1246,0.1634,0.1221,0.1476,0.1151,0,0.3613,0.1115,
0.1296,0.1047,0.1271,0.1117,0.2953,0.1727,0.3686,0,0.1412,
0.2222,0.0654,0.0693,0.0847,0.1074,0.1973,0.0776,0.0964,0),9,9)
w1 = list(order0 = o0,order1 = o1) #Criando W baseada no IDH

library(starma)

d=stcenter(dados,center=TRUE, scale=TRUE) #Centralizando os dados (dados sem tendência)

stcor.test(d,w1,tlag=NULL) #Extensão do teste de Box-Pierce para coorelação

#Etapa de identificação do modelo
stacf(d,w1,tlag.max =NULL, use.ggplot = TRUE)
stpacf(d,w1,tlag.max = NULL, use.ggplot = TRUE)

#Etapa de ajuste do modelo
ar <- matrix(0, 6,2)
ar[2,2] = 1 # phi2,1
ar[5,1] = 1 # phi5,0
ar[6,2] = 1 #phi6,1

ma <- matrix(0,2,2)
ma[1,1] = 1 #theta1,0

#Ajuste do modelo
mod1=starma(d, w1, ar, ma)
summary(mod1)

#Diagnóstico do modelo
stcor.test(mod1$residuals,w1,fitdf =4)#Teste de correlação residual
stacf(mod1$residuals,w1) #Função de auto correlação residual
```

```

#Modelo considerando-se W baseada no Inv
dados = read.table("dadosDifTx.csv",sep = ";",dec=".", h = T) #Entrada de dados
o0=diag(9)
o2 = matrix(c(0,0.0005,0.1807,0.0547,0.0099,0.0088,0.0116,0.0077,0.0092,
0.1232,0,0.0938,0.2425,0.0649,0.0809,0.0770,0.0915,0.0881,
0.1324,0.0080,0,0.2193,0.2875,0.0468,0.1416,0.0433,0.3258,
0.1231,0.9576,0.0936,0,0.2599,0.0692,0.0804,0.0663,0.0876,
0.1278,0.0154,0.1973,0.2190,0,0.0554,0.1135,0.0518,0.1421,
0.1074,0.0038,0.0302,0.0550,0.0521,0,0.0370,0.6671,0.0346,
0.1409,0.0044,0.1168,0.0639,0.1069,0.0371,0,0.0338,0.2755,
0.1089,0.0042,0.0326,0.0614,0.0568,0.6598,0.0451,0,0.0364,
0.1361,0.0059,0.2548,0.0844,0.1620,0.0419,0.3818,0.0384,0),9,9)
w2 = list(order0 = o0,order1 = o2)
library(starma)

d2=stcenter(dados,center=TRUE, scale=TRUE) #Centralizando os dados (dados sem tendência)

stcor.test(d2,w2,tlag=NULL) #Extensão do teste de Box-Pierce para coorelação
#Etapa de identificação do modelo
stacf(d2,w2,tlag.max =NULL, use.ggplot = TRUE)
stpacf(d2,w2,tlag.max = NULL, use.ggplot = TRUE)
#Etapa de ajuste do modelo
ar <- matrix(0,6,2)
ar[6,2]=1
ar[5,1] = 1
ma <- matrix(0,6,2)
ma[1,1] = 1
ma[6,2]= 1

mod2=starma(d2, w2, ar, ma)
summary(mod2)

stcor.test(mod2$residuals,w2,fitdf =4)
stacf(mod2$residuals,w2)
stpacf(mod2$residuals,w2)

mod3=starma(dad, w3, ar, ma)
summary(mod3)

stcor.test(mod3$residuals,w3,fitdf =4)
stacf(mod3$residuals,w3)
stpacf(mod3$residuals,w3)

#Investimento para grande BH

o4 = matrix(c(0,0.0648,0.0693,0.4821,0,0.9307,0.5179,0.9352,0),3,3)
w4 = list(diag(3),o4)

stcor.test(dad,w4,tlag=NULL)

stacf(dad,w4,tlag.max =NULL, use.ggplot = TRUE)
stpacf(dad,w4,tlag.max = NULL, use.ggplot = TRUE)

ar <- matrix(0, 3,2)
#ar[2,2]=1
#ar[3,2] = 1

ma <- matrix(0,7,2)
ma[1,1] = 1
ma[4,2] = 1
ma[5,2]=1
ma[6,2] = 1

```



```

#APENAS BH, BETIM E CONTAGEM
dado = read.table("bh.csv",sep = ";",dec=".", h = T) #Apenas séries de BH, Betim e Contagem

library(starma) #e todas sem tendência
dad=stcenter(dado,center=TRUE, scale=TRUE)

#IDH para grande BH
o3 = matrix(c(0,0.1029,0.1148,0.4695,0,0.8852,0.5305,0.8971,0),3,3)
w3 = list(diag(3),o3)

stcor.test(dad,w3,tlag=NULL)

stacf(dad,w3,tlag.max =NULL, use.ggplot = TRUE)
stpacf(dad,w3,tlag.max = NULL, use.ggplot = TRUE)

ar <- matrix(0, 6,2)
#ar[3,2] = 1
#ar[4,2] = 1
#ar[5,2] = 1

ma <- matrix(0,8,2)
ma[1,1] = 1
#ma[4,2] = 1
ma[5,2] = 1
ma[6,2] = 1
ma[7,2] = 1

mod3=starma(dad, w3, ar, ma)
summary(mod3)

stcor.test(mod3$residuals,w3,fitdf =4)
stacf(mod3$residuals,w3)
stpacf(mod3$residuals,w3)

#Investimento para grande BH
o4 = matrix(c(0,0.0648,0.0693,0.4821,0,0.9307,0.5179,0.9352,0),3,3)
w4 = list(diag(3),o4)

stcor.test(dad,w4,tlag=NULL)

stacf(dad,w4,tlag.max =NULL, use.ggplot = TRUE)
stpacf(dad,w4,tlag.max = NULL, use.ggplot = TRUE)

ar <- matrix(0, 3,2)
#ar[2,2]=1
#ar[3,2] = 1

ma <- matrix(0,7,2)
ma[1,1] = 1
ma[4,2] = 1
ma[5,2]=1
ma[6,2] = 1
mod4=starma(dad, w4, ar, ma)
summary(mod4)

stcor.test(mod4$residuals,w4,fitdf =4)
stacf(mod4$residuals,w4)
stpacf(mod4$residuals,w4)

#CONTIGUIDADE
wcont = matrix(c(0,0,0.5,0,0,0.5,1,1,0),3,3)
w5=list(diag(3),wcont)

stcor.test(dad,w5,tlag=NULL)

stacf(dad,w5,tlag.max =NULL, use.ggplot = TRUE)
stpacf(dad,w5,tlag.max = NULL, use.ggplot = TRUE)

ar <- matrix(0, 6,2)
#ar[2,2]=1
#ar[3,2] = 1
#ar[4,2] = 1

ma <- matrix(0, 6,2)
ma[1,1] = 1
ma[2,2] = 1

```

```
ma[3,2] = 1
ma[4,2] = 1
ma[5,2] = 1
ma[6,2] = 1

mod5=starma(dad, w5, ar, ma)
summary(mod5)

stcor.test(mod5$residuals,w5,fitdf =6)
stacf(mod5$residuals,w5)
stpacf(mod5$residuals,w5)

res=mod5$residuals
```