



UEPB

**UNIVERSIDADE ESTADUAL DA PARAÍBA
PRÓ-REITORIA DE PÓS-GRADUAÇÃO E PESQUISA
PROGRAMA DE PÓS-GRADUAÇÃO EM CIENCIA E TECNOLOGIA EM SAÚDE**

FÁBIO JÚNIOR FRANCISCO DA SILVA

**PREVISÃO DE INTERNAÇÕES HOSPITALARES DE DENGUE POR MEIO DE
SÉRIES TEMPORAIS**

**CAMPINA GRANDE/PB
2019**

FÁBIO JÚNIOR FRANCISCO DA SILVA

**PREVISÃO DE INTERNAÇÕES HOSPITALARES DE DENGUE POR MEIO DE
SÉRIES TEMPORAIS**

Dissertação apresentada ao Programa de Pós-graduação em Ciência e Tecnologia em Saúde (PPGCTS), da Universidade Estadual da Paraíba (UEPB), como requisito à obtenção do título de Mestre em Ciência e Tecnologia em Saúde.

Orientador: Dr. Wellington Candeia de Araújo

**CAMPINA GRANDE/PB
2019**

É expressamente proibido a comercialização deste documento, tanto na forma impressa como eletrônica. Sua reprodução total ou parcial é permitida exclusivamente para fins acadêmicos e científicos, desde que na reprodução figure a identificação do autor, título, instituição e ano do trabalho.

S586p Silva, Fábio Júnior Francisco da.
Previsão de internações hospitalares de dengue por meio de séries temporais [manuscrito] / Fábio Júnior Francisco da Silva. - 2019.
119 p. : il. colorido.
Digitado.
Dissertação (Mestrado em Profissional em Ciência e Tecnologia em Saúde) - Universidade Estadual da Paraíba, Pró-Reitoria de Pós-Graduação e Pesquisa, 2020.
"Orientação : Prof. Dr. Wellington Candeia de Araújo, Coordenação do Curso de Computação - CCT."
1. Aedes aegypti. 2. Previsão de séries temporais. 3. Inteligência artificial. 4. Modelos estatísticos. I. Título
21. ed. CDD 600


FABIO JUNIOR FRANCISCO DA SILVA

**PREVISÃO DE INTERNAÇÕES HOSPITALARES DE DENGUE POR MEIO
DE SÉRIES TEMPORAIS**

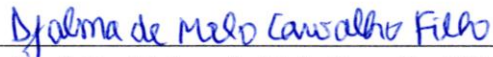
Dissertação apresentada ao Curso de Mestrado em Ciência e Tecnologia em Saúde da Universidade Estadual da Paraíba, como requisito à obtenção do título de mestre.

Aprovada em: 20/12/2019.


BANCA EXAMINADORA



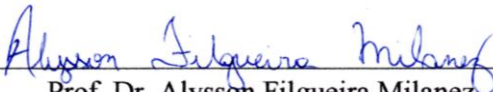
Prof. Dr. Wellington Candeia de Araújo
Universidade Estadual da Paraíba (UEPB)



Prof. Dr. Djalma de Melo Carvalho Filho
Universidade Estadual da Paraíba (UEPB)



Prof. Dr. Rodrigo Alves Costa
Universidade Estadual da Paraíba (UEPB)



Prof. Dr. Alysso Filgueira Milanez
Externo (IFPB)

À minha amada esposa Ianne Raquel por ser
companheira;

Ao nosso filho Theo Arthur;

Aos meus irmãos Misias e Juliana;

Aos meus pais Sebastião e Irene;

À minha sogra Iracema e cunhadas Rafaela e
Samara;

E aos meus sobrinhos Henrique e Maya.

Dedico

AGRADECIMENTOS

A “caminhada” percorrida para a conclusão deste trabalho de pesquisa foi tão desafiadora quanto realizadora. Agradeço o apoio de pessoas muito especiais em minha vida por conseguir finalizá-lo

A DEUS, primeiramente, pela dádiva da vida, esperança, perseverança e sabedoria a mim atribuída, o que permitiu alcançar mais um objetivo.

Ao Grande Professor Dr. Wellington Candeia de Araujo, meu orientador, que acompanhou, não só a produção deste trabalho, mas também deu o maior apoio para a conclusão da pós-graduação.

Ao amigo Francisco Anderson pelo companheirismo e compartilhamento das experiências vivenciadas durante a realização de todo o mestrado.

Aos meus familiares que me encorajaram com incentivos e compreensão e outros que de alguma forma contribuíram com esta realização.

Aos profissionais da UEPB: colaboradores, supervisores, coordenadores, secretários e outros pela dedicação à realização desta pós-graduação.

A Ianne Raquel, esposa amada e companheira pela dedicação, compreensão, motivação, sempre ao meu lado, incentivando e ajudando em todos os momentos.

Ao nosso filho Theo Arthur que mesmo sendo um bebê foi paciente e dormiu a noite toda, o que permitiu terminar a pesquisa e a escrita deste trabalho.

E também a todos os amigos que contribuíram de alguma forma com a realização desta pós-graduação.

“Contem o que vai acontecer no futuro, e saberemos que vocês são mesmo deuses. Façam alguma coisa, boa ou má, para que a vejamos e os respeitemos”

Isaías 41:23

RESUMO

O *Aedes aegypti* é o vetor das comorbidades febre amarela, dengue, *chikungunya* e *zika*. Dentre elas, a dengue é a mais comum e a que causa mais mortes, se tornando preocupação de saúde pública no Brasil e no mundo. O Brasil apresenta condições ideais para a proliferação do mosquito e concentra a maior parte dos casos de dengue nas Américas. Controlar o vetor é desafiador e precisa de ações estratégicas envolvendo governo e a sociedade civil. A infecção pelo vírus da dengue pode ser assintomática, branda ou ocasionar doença grave que coloca em risco a vida do paciente, o qual precisará ser internado em hospital para tratamento. Sabe-se que os leitos hospitalares são limitados, a demanda de internações por dengue é representativa e a internação gera custo. A incidência de dengue apresenta tendência e sazonalidade, principais componentes de uma série temporal. Para contribuir com a gestão estadual e municipal de saúde no combate ao *Aedes aegypti*, bem como a gestão hospitalar das internações, esta pesquisa definiu um algoritmo para a previsão de internações hospitalares por dengue utilizando modelos estatísticos computacionais para análise de séries temporais com foco no planejamento e gestão do combate à doença. Foram utilizadas as técnicas estatísticas (Suavização Exponencial, Modelo de Média Móvel Integrado Autorregressivo (ARIMA), Redes Neurais Artificiais Autorregressivas, Combinação de modelos, Regressão Linear e método inocentes/ingênuo) para realizar a previsão das séries temporais com horizonte de 8 semanas e estimativa real de 4 semanas, pois o DATASUS libera os dados com 4 semanas de atraso. O ajuste dos parâmetros de cada modelo foi executado de forma automática no ambiente de desenvolvimento integrado (IDE) RStudio, por funções do pacote *forecast* que contém implementações dos modelos estatísticos de séries temporais para a linguagem de programação R. A utilização de diversos métodos de previsão aplicados concomitantemente sobre a mesma série temporal melhorou a precisão da previsão. O algoritmo realizou cerca de 8 previsões a cada 10 com erro médio percentual absoluto (MAPE) inferior a 26%. Esta é mais uma estratégia com potencial para ser utilizada no controle da dengue no Brasil, pois o algoritmo criado pode ser a base para o desenvolvimento, em trabalhos futuros, de um serviço *web* que forneça aos gestores da saúde (Ministros, Secretários, Diretores e Coordenadores) a realização de previsão de internações por dengue de acordo com a realidade local. A partir deste estudo surgiram oportunidades de pesquisa para exploração de métodos estatísticos que lidam com séries temporais contáveis para a previsão de internações por dengue com foco no estabelecimento de saúde onde os dados ocorrem,

bem como a realização de Revisão Sistemática da Literatura, considerando o métodos clássicos e técnicas de aprendizagem de máquina, para responder as perguntas: Quais métodos estatísticos para a previsão de séries temporais (discretas/contáveis ou contínuas) são mais utilizados? Os métodos clássicos são mais eficientes que as abordagens com aprendizagem de máquina? Quais métricas de erro são mais seguras para a escolha do modelo mais preciso entre os candidatos a determinada previsão de série temporal?

Palavras-chave: Controle do *Aedes aegypti*. Internação por dengue. Previsão de séries temporais. Inteligência artificial. Modelos estatísticos.

ABSTRACT

Aedes aegypti is the vector of comorbidities yellow fever, dengue fever, chikungunya and zika. Among them, dengue is the most common disease and the one that causes the most deaths, becoming a public health concern in Brazil and worldwide. Brazil presents ideal conditions for mosquito proliferation and concentrates most dengue cases in the Americas. Controlling the vector is challenging and needs strategic action involving government and civil society. Infection by dengue virus can be asymptomatic, mild or cause a serious disease that puts the patient's life at risk, who will need to be admitted to a hospital for treatment. It is known that hospital beds are limited, the demand for dengue is representative and hospitalization generates costs. The incidence of dengue has tendency and seasonality, the main components of a time series. To contribute to state and municipal health management in combating *Aedes aegypti*, as well as hospital management of dengue hospitalizations, this research created an algorithm for the prediction of dengue hospitalizations using computational statistical models for time series analysis focusing on planning and management of disease control. The main statistical techniques used (Exponential Smoothing, Autoregressive Integrated Moving Average Model (ARIMA), Autoregressive Artificial Neural Networks, Model combination, Linear Regression and Naive Methods) to forecast 8-week time series and a real estimate of 4-weeks, because DATASUS releases data is 4-weeks late. The adjustment of the parameters of each model was performed automatically in the Integrated Development Environment (IDE) RStudio, by functions of the forecast package that contains implementations of the statistical models of time series for the programming language R. Using multiple forecasting methods applied concomitantly over the same time series has improved forecast accuracy. The algorithm made about 8 predictions every 10 with Mean Absolute Percentage Error (MAPE) below 26%. This is another strategy with potential to be used in dengue control in Brazil, because the algorithm created may be the basis for the development, in the future works, of a web service that provides health managers (Ministers, Secretaries, Directors and Coordinators) the possibility of predicting hospitalizations for dengue according to reality of each one. From this study, research opportunities arose to explore statistical methods that deal with count time series for predicting dengue hospitalizations with a focus on the health establishment where the data occur, as well as conducting a Systematic Literature Review, considering the classical methods and machine learning techniques, to answer the questions: Which statistical methods for forecasting time

series (discrete/count or continuous) are most used? Are classical methods more efficient than machine learning approaches? Which error metrics are safer for choosing the most accurate model among candidates for a given time series forecast?

Key words: Aedes aegypti Control. Dengue hospitalization. Forecast of time series. Artificial intelligence. Statistical models.

LISTAS DE SIGLAS E ABREVIATURAS

AB	Atenção Básica
ABNT	Associação Brasileira de Normas Técnicas
ACE	Agente de Combate às Endemias
ACF	Função de Autocorrelação
ACS	Agente Comunitário de Saúde
AIC	Critério de Informação de Akaike
AICc	Critério de Informação de Akaike Corrigido
AIH	Autorizações de Internação Hospitalar
ANN	Rede Neural Artificial
AR	Autorregressivo
ARIMA	Modelo de Média Móvel Integrado Autorregressivo
BIC	Critério de Informação Bayesiano de Schwarz
CAPES	Coordenação de Aperfeiçoamento de Pessoal de Nível Superior
CBCD	Centro Colaborador da OMS para a Classificação de Doenças em Português
CE	Critérios de Exclusão
CEP	Código de Endereçamento Postal
CI	Critérios de Inclusão
CIB	Comitê Intergestores Bipartite
CID	Classificação Estatística Internacional de Doenças
CHS	Ciências Humanas e Sociais
CONEP	Comissão Nacional de Ética em Pesquisa
CEP	Comitês de Ética em Pesquisa
DATASUS	Departamento de Informática do SUS
DeCS	Descritores em Ciências da Saúde
DIVE	Diretoria de Vigilância Epidemiológica
DICIO	Dicionário de Português Online
DRIFT	Modelo de previsão baseado no método inocente (<i>Naive</i>)
EI	Excluídos por Irrelevância
ESB	Estratégia de Saúde Bucal
ESF	Estratégia Saúde da Família
ETS	Erro, Tendência e Sazonalidade

G1	Portal de notícias da Globo
HOLT	Modelo de previsão por tendência linear de Holt
IBGE	Instituto Brasileiro de Geografia e Estatística
IEEE	Instituto de Engenheiros Eletricistas e Eletrônicos
INMET	Instituto Nacional de Meteorologia
INDA	Infraestrutura Nacional de Dados Abertos
KPSS	Teste de Estacionariedade de Kwiatkowski, Phillips, Schmidt e Shin
LOS	Tempo em Internação Hospitalar
MA	Média Móvel
MAE	Erro Médio Absoluto
MAPE	Erro Médio Percentual Absoluto
MASE	Erro Médio Absoluto Escalado
MC1	Modelo Combinado 1
MC2	Modelo Combinado 2
ME	Erro Médio
MeSH	<i>Medical Subject Headings</i>
mmHg	milímetros de mercúrio
MS	Ministério da Saúde
NNAR	Rede Neural Autorregressiva
OMS	Organização Mundial da Saúde
PA	Pressão Arterial
PACF	Função Parcial de Autocorrelação
PNCD	Programa Nacional de Controle da Dengue
PSF	Programa Saúde da Família
R ²	R quadrado ajustado
RMSE	Raiz do Erro Médio Quadrado
RS	Revisão Sistemática
RSL	Revisão Sistemática de Literatura
SARIMA	Modelo de Média Móvel Integrado Autorregressivo Sazonal
SARIMAX	Modelo de Média Móvel Integrado Autorregressivo Sazonal Multivariado
SCNES	Sistema de Cadastro Nacional de Estabelecimentos de Saúde
SIH	Sistema de Informações Hospitalares
SIHD	Sistema de Informação Hospitalar Descentralizado

SINAN	Sistema de Informação de Agravos de Notificação
SNAIVE	Técnica de previsão inocente para série com sazonalidade
SIS	Sistema de Informação em Saúde
STL	Modelo de previsão por decomposição da série temporal
TBATS	Suavização exponencial combinada com Fourier e transformação de Box-cox
TDIC	Tecnologia Digital da Informação e Comunicação
TMLS	Modelo de previsão utilizando regressão linear
UBS	Unidade Básica de Saúde
UF	Unidade da Federação
WHO	Organização Mundial da Saúde

SUMÁRIO

1	INTRODUÇÃO	16
2	REVISÃO DA LITERATURA	23
2.1	O MOSQUITO Aedes Aegypti	23
2.1.1	Controle do <i>Aedes aegypti</i> no Brasil	25
2.1.2	Internações hospitalares por dengue no Brasil	27
2.2	SÉRIES TEMPORAIS	30
2.2.1	Métricas de avaliação de modelo e precisão	32
2.2.2	Modelos estatísticos para previsão de séries temporais	34
2.2.2.1	<i>Suavização Exponencial</i>	34
2.2.2.2	ARIMA	35
2.2.2.3	Redes neurais artificiais	36
2.2.2.4	Modelos combinados	37
2.2.2.5	Ajuste automático de parâmetros para a escolha do melhor modelo	37
2.3	SÉRIES TEMPORAIS APLICADAS AO ESTUDO DA DENGUE	46
2.4	PESQUISAS RELACIONADAS	48
3	MATERIAL E MÉTODO	52
3.1	CARACTERIZAÇÃO DA PESQUISA	52
3.1.1	Levantamento bibliográfico	52
3.1.2	Fonte e seleção dos dados	55
3.2	PREVISÃO COM MELHOR PRECISÃO	56
3.2.1	ARIMA	58
3.2.2	Suavização exponencial	58
3.2.3	Rede Neural Autorregressiva	59
3.2.4	Outros métodos estatísticos	59
3.2.5	Modelo combinado	59
3.2.6	Seleção do modelo com melhor precisão	60
3.3	ASPECTOS ÉTICOS	62
4	RESULTADOS E DISCUSSÕES	65
4.1	ANALISE EXPLORATÓRIA DOS DADOS	65
4.2	ESCOLHA DO MODELO COM MELHOR PRECISÃO	73
4.3	ALGORITMO PARA PREVISÃO DE INTERNAÇÕES HOSPITALARES POR DENGUE	75

5	CONSIDERAÇÕES FINAIS	79
5.1	LIMITAÇÕES E TRABALHOS FUTUROS	81
	REFERÊNCIAS.....	84
	APÊNDICE A – Decomposição das séries temporais por Unidade da Federação	90
	APÊNDICE B – Protocolo de revisão sistemática da literatura sobre séries temporais para a previsão de internações hospitalares por dengue	101

1 INTRODUÇÃO

1 INTRODUÇÃO

O *Aedes aegypti* é um mosquito oriundo da África que se desenvolve na água e se espalhou pelo mundo transmitindo febre amarela, dengue, *chikungunya*, *zika* e, recentemente, microcefalia. Com ciclo de vida de quatro fases (ovo, larva, pupa e adulto), em condições ambientais favoráveis, depois que o ovo eclode, leva-se cerca de 7 a 10 dias para que a fêmea se torne adulta. Ao contrair o vírus, o mosquito fêmea começa a transmiti-lo ao ser humano entre 8 e 12 dias (DIVE, 2019).

A dengue é a doença mais transmitida pelo *Aedes aegypti* e se tornou um dos principais desafios de Saúde Pública no Brasil e no mundo, exigindo medidas de combate ao mosquito, implementadas com a maior participação possível de toda a população (ROCHA; DANTAS; CÂNDIDO, 2014).

A Organização Mundial da Saúde (*World Health Organization* - WHO) alerta para o aumento mundial da dengue (de 2,2 milhões em 2010 para mais de 3,34 milhões em 2016). América, Sudeste da Ásia e Pacífico Ocidental são as regiões mais afetadas passando de 1,2 milhões de casos em 2008 para mais de 3,34 milhões em 2016. Em 2015 ocorreram 2,35 milhões de casos só na América, dos quais 10.200 casos se manifestaram como dengue severa e levaram 1.181 pessoas à morte. Estima-se que meio milhão de pessoas com dengue severa precisam de hospitalização a cada ano, dos quais 12.500 não resistem e morrem (WHO, 2019).

Em 2019 no Brasil, até o mês de março, ocorreu aumento de 303% nos casos suspeitos de dengue em comparação com o mesmo período no ano de 2018 (G1, 2019). Até o mês de maio de 2019, já haviam ocorrido 10.391 internações por dengue a mais do que no ano de 2018 inteiro (SIH/SUS, 2019). As internações hospitalares acarretam custos, bem como os leitos disponíveis para novas internações são limitados.

O custo da internação hospitalar está associado ao tempo de internação (*Length of Stay* - LOS). Em pesquisa realizada na cidade de Zhongshan, Província de Guangdong, China, Zhang, Yuan e Wang (2017) estimaram uma média de 7,2 dias por paciente internado por dengue e custo médio de US\$ 499,64 dólares, aproximadamente 3,71% do produto interno bruto per capita da cidade no período.

No Brasil, a média do custo hospitalar por dengue foi de US\$ 1.622,40 dólares para pacientes que necessitaram de hemoderivados no tratamento e US\$ 550,20 dólares para os demais casos. O tempo de permanência médio de pacientes em internação hospitalar em

virtude de dengue é de 3 dias e existe uma concentração maior de casos próxima a costa marítima brasileira (VIEIRA MACHADO *et al.*, 2014; SILVA *et al.*, 2019). Comparando o preço da internação entre Brasil e China, naquele o valor é mais que o dobro deste.

O combate ao *Aedes aegypti* é importante para que diversas doenças sejam evitadas, mas a ação deve ser coletiva (governo e sociedade civil). Neste sentido, faz-se necessário compreender holisticamente¹ o comportamento desse vetor e desenvolver estratégias para previsão de novos casos das doenças transmitidas por ele. “Saber o que vai acontecer antes mesmo dos primeiros sinais se manifestarem pode propiciar melhor aproveitamento dos efeitos benéficos de eventos futuros ou uma preparação antecipada de eventuais efeitos adversos” (ANTUNES; CARDOSO, 2015).

Técnicas estatísticas de análise e previsão de séries temporais têm sido utilizadas para realizar a previsão de incidência de dengue e novas internações hospitalares por doenças diversas, por exemplo, as pesquisas de Chakraborty, Chattopadhyay e Ghosh (2019), Zhou *et al.* (2018), Villani *et al.* (2017) e Lizzi (2012). Diante disso seria possível realizar a previsão das novas internações hospitalares por dengue?

A incidência e conseqüentemente as internações hospitalares por dengue apresentam tendência e sazonalidade, os principais componentes de uma série temporal. Além disso, os dados sobre internações hospitalares estão disponíveis, se enquadraram no estudo de séries temporais e os principais métodos estatísticos ganharam recursos computacionais de apoio com o pacote *forecast* para a linguagem de programação R.

R é um ambiente e linguagem de programação livre e colaborativa que permite o desenvolvimento de software que envolve computação por métodos e técnicas estatísticas. Está disponível para os principais Sistemas Operacionais. O pacote *forecast* para R fornece métodos e ferramentas para visualizar, analisar e realizar previsão em séries temporais univariadas, incluindo suavização exponencial, modelagem automática da técnica ARIMA e outros (R CORE TEAM, 2019; HYNDMAN *et al.*, 2019).

Os dados em conjunto com os métodos estatísticos disponíveis têm potencial para a **“Criação de um algoritmo de previsão de internações hospitalares de dengue por meio de séries temporais”**. De acordo com Cormen *et al.* (2012), pode-se definir “algoritmo”

¹ Segundo Dicio (2019), **holística** é adjetivo que significa busca para entender os fenômenos ou a realidade por completo, e não somente como resultado da união de suas partes; analisar e entender algo por inteiro.

como procedimento computacional (sequência de etapas) bem definido que **recebe** um valor ou conjunto de valores como **entrada** e **produz** algum valor ou conjunto de valores como **saída**, ou seja, uma ferramenta para resolver um problema computacional bem especificado.

Frente aos desafios existentes para o controle do *Aedes aegypti*, bem como a necessidade de implementação de estratégias que subsidiem a gestão do combate a esse vetor transmissor da dengue e consequentemente as internações hospitalares, esta pesquisa **criou um algoritmo para a previsão de internações hospitalares por dengue utilizando modelos estatísticos computacionais para análise de séries temporais**.

Para chegar-se ao objetivo geral do presente estudo, definiram-se os seguintes objetivos específicos:

- 1. Gerar série temporal por semana epidemiológica e UF.**
- 2. Realizar análise exploratória das séries temporais geradas.**
- 3. Criar algoritmo para a previsão de hospitalizações por dengue em conformidade com a série temporal analisada.**
- 4. Avaliar o algoritmo de previsão utilizando a métrica MAPE para ajuste de parâmetros e verificação de precisão.**

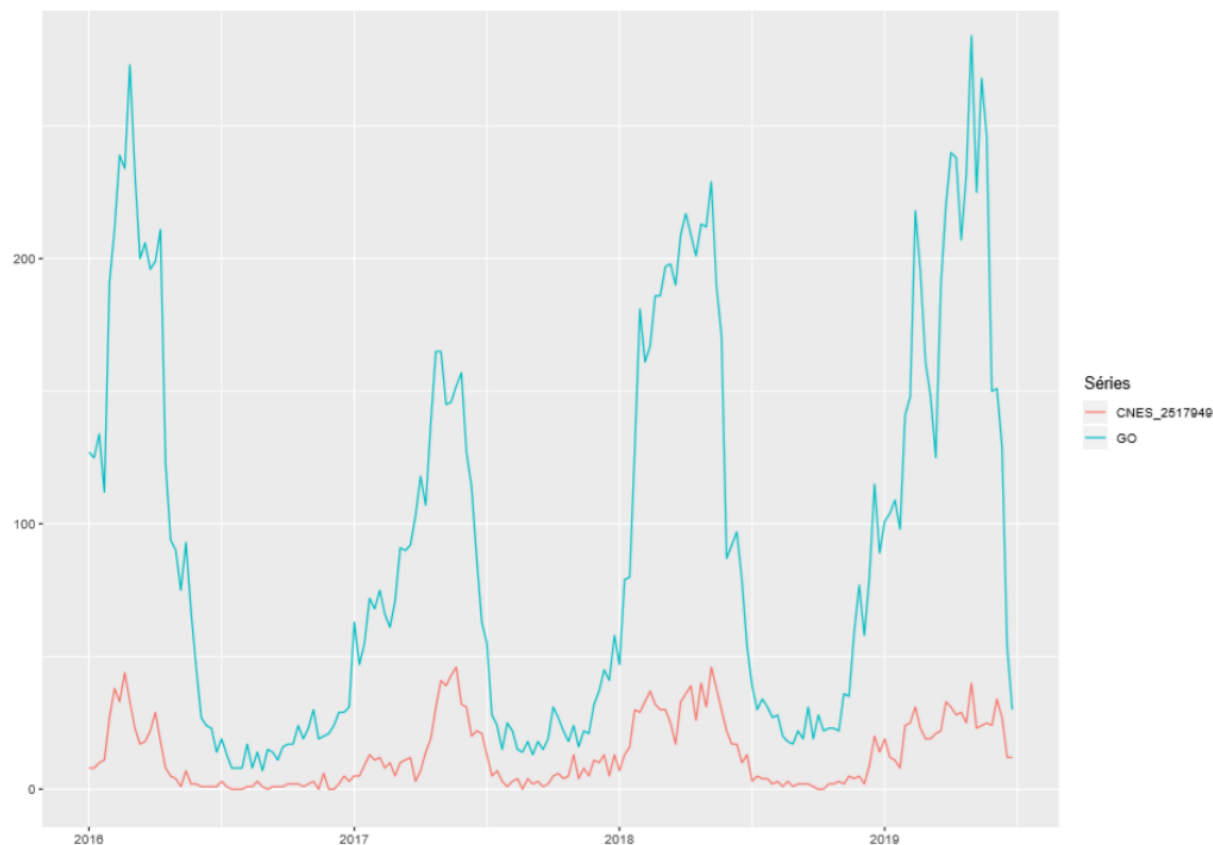
A previsão tem precisão passível de erro, o importante é reduzir esse erro o máximo possível. As pesquisas relacionadas trazem como resultado o uso de determinada técnica de previsão supondo que a mesma vai lidar com a série temporal em diferentes épocas. Esse fato não se verificou na previsão de internações hospitalares por dengue, pois ao variar a época de previsão, o método mais preciso foi diferente. Neste sentido, este trabalho utilizou diversos métodos estatísticos concomitantemente e sobre os mesmos dados com o objetivo de reduzir o erro da previsão.

Como a disponibilidade de leitos hospitalares para internação de novos pacientes é limitada, a previsão de demanda se torna ferramenta fundamental para a gestão, pois, contribui para o aumento da eficiência na alocação de recursos físicos e humanos necessários ao cuidado do paciente, diminuindo as lotações e aumentando a qualidade do cuidado dos internados, já que a demanda seria conhecida com antecedência (ZHOU *et al.*, 2018). Ainda segundo os autores, poucas publicações exploram a previsão de novas internações.

A Figura 1 mostra a série histórica do estado de Goiás e da Maternidade e Hospital São Judas Tadeu (Cadastro Nacional de Estabelecimentos de Saúde – CNES nº 2517949), Goiânia-GO, organizada por semana epidemiológica. É possível perceber que a tendência das

internações por dengue no estabelecimento de saúde acompanha a tendência das ocorrências a nível estadual.

FIGURA 1 - SÉRIES TEMPORAIS DE INTERNAÇÃO POR DENGUE NO ESTADO DE GOIÁS E NO ESTABELECIMENTO DE SAÚDE CNES Nº 2517949



Fonte: Autoria própria, 2019.

A série histórica de internações por dengue (por semana epidemiológica) da maioria dos estabelecimentos de saúde conveniados ao Sistema Único de Saúde (SUS) avaliados se caracterizam como discretas e/ou contáveis e não apresentaram quantidade de dados e padrões suficientes para aplicação direta de previsão por meio de técnicas estatísticas que lidam com séries não contáveis ou discretas.

Por outro lado, a série histórica por UF conta com a quantidade de dados necessária para a realização da previsão. Como a tendência das internações hospitalares por dengue dos estabelecimentos de saúde acompanha a tendência das internações ocorridas na UF onde o estabelecimento está localizado (Figura 1), então é possível realizar previsão para a UF e observar quais hospitais estão sendo mais procurados.

A previsão das internações por dengue pode servir de base para o monitoramento constante pelas gestões estaduais e municipais de saúde, bem como a gestão hospitalar. Em suma, realizando-se a previsão de internações por dengue a nível estadual, é possível observar quais hospitais (na abrangência da UF) estão recebendo mais internações por dengue e alertá-los sobre a tendência de crescimento, além de verificar quais são os municípios de residência dos pacientes.

Segundo Brasil (2009) é responsabilidade da esfera estadual, entre outros:

- Elaborar o plano estadual de prevenção e controle de epidemias de dengue, coordenar a elaboração dos planos regionais e apoiar a elaboração dos planos municipais;
- Realizar supervisão nos municípios, com reuniões periódicas de monitoramento;
- Definir e regular, no âmbito da CIB, fluxos regionais para garantir a atenção integral aos pacientes com dengue;
- Garantir o acesso dos pacientes aos serviços sob gestão estadual, conforme pactuação, incluindo suporte laboratorial e regulação de leitos;
- Mobilizar e instrumentalizar entidades da sociedade organizada e da iniciativa privada, de âmbito estadual, para atuarem no enfrentamento da dengue.

Os dados de internações hospitalares são encaminhados ao DATASUS que disponibiliza cronograma anual de envio pelos estabelecimentos de saúde conveniados ao SUS, por exemplo, para dados produzidos na competência 11 (onze) de 2019, isto é, internações ocorridas no mês de novembro, o prazo para envio foi definido do dia 5 (cinco) ao dia 30 (trinta) de dezembro de 2019 (SIHD, 2019).

Os dados das internações hospitalares não estão disponíveis em tempo real, pois existe atraso de um mês (aproximadamente 4 semanas), ou seja, no início de janeiro de 2020 tem-se acesso aos dados até o final de novembro de 2019. Logo definiu-se horizonte de previsão de internações hospitalares por dengue de 8 semanas para uma previsão real de 4 semanas, pois o crescimento do erro é diretamente proporcional ao aumento do horizonte de previsão a partir dos dados reais.

O principal risco desta pesquisa esteve relacionado à possibilidade de inexistência dos dados para os estados brasileiros incluídos no estudo ou a impossibilidade de previsão pela utilização dos métodos estatísticos selecionados. Mas, das 27 (vinte sete) séries temporais de internações hospitalares por dengue extraídas do DATASUS, uma para cada Unidade da

Federação, apenas 5 (cinco) apresentaram característica de série temporal discreta e/ou contável, as demais se enquadraram nos métodos estatísticos utilizados.

O algoritmo pode falhar quanto ao erro na previsão das internações hospitalares por dengue. O erro é a diferença entre o valor previsto e o valor real observado. As internações ocorridas e não registradas levam a dados faltantes que podem interferir negativamente na previsão, ou seja, aumentar o erro. Por isso, a estratégia utilizada para minimizar o erro ao máximo possível foi fundamental para o sucesso da pesquisa.

Na pesquisa de Zhou *et al.* (2018), a média dos erros entre os três modelos de previsão utilizado foi de 27%. Em Lizzi (2012), o menor erro foi de 28%. Para esta pesquisa estabeleceu-se como maior valor para o erro, a porcentagem de 26%. O erro fora do intervalo de confiança estabelecido (26%), pode levar os gestores de saúde a alocar recursos financeiros físicos e/ou humanos: (1) desnecessários nos casos de previsão muito acima da realidade; ou (2) deixar de alocar recursos necessários em caso de previsão muito abaixo da realidade.

Esta pesquisa foi organizada em 5 (cinco) capítulos: 1 - Introdução que apresenta o contexto, problema e justificativa; 2 - Revisão da literatura abordando o controle *do Aedes aegypti* e as internações hospitalares por dengue no Brasil, as definições pertinentes sobre séries temporais (modelos estatísticos para previsão, métricas de avaliação, precisão de modelos e aplicação no estudo da dengue) e pesquisas relacionadas; 3 - Material e método que discorre sobre a caracterização da pesquisa, levantamento bibliográfico com auxílio de Revisão Sistemática da Literatura, fonte, seleção e tratamento dos dados abertos sobre internações hospitalares, modelos estatísticos de previsão de séries temporais utilizados, aferição de precisão com base no erro e aspectos éticos; 4 - Resultados e discussões com análise exploratória dos dados, escolha do modelo de previsão de séries temporais com melhor precisão e definição do algoritmo para previsão de internações hospitalares por dengue; e 5 - Considerações finais, o qual mostra as conclusões com base na pesquisa, bem como as limitações encontradas e oportunidade para trabalhos futuros.

2 REVISÃO DA LITERATURA

2 REVISÃO DA LITERATURA

Este capítulo apresenta os conceitos necessários para a compreensão dos objetivos da pesquisa e os trabalhos relacionados, ou seja, o estado da arte (WAZLAWICK, 2017). Para tanto, a pesquisa por referências bibliográficas foi realizada em duas abordagens:

- 1) Levantamento das referências por método tradicional, utilizando as palavras chave em buscas na *web* por bases de publicações científicas com potencial para os objetivos estabelecidos neste trabalho;
- 2) Execução de protocolo de Revisão Sistemática da Literatura (RSL) até o passo da seleção dos estudos encontrados, utilizando a *String* de busca ("*time series*" AND (*predict** OR *forecast**) AND *hospital** AND dengue), que foi aplicada nas bases de busca: PubMed, Portal de Periódicos da CAPES, IEEE e Scopus e retornou um total de 1144 publicações, das quais 11 foram selecionadas pelos critérios de inclusão e exclusão.

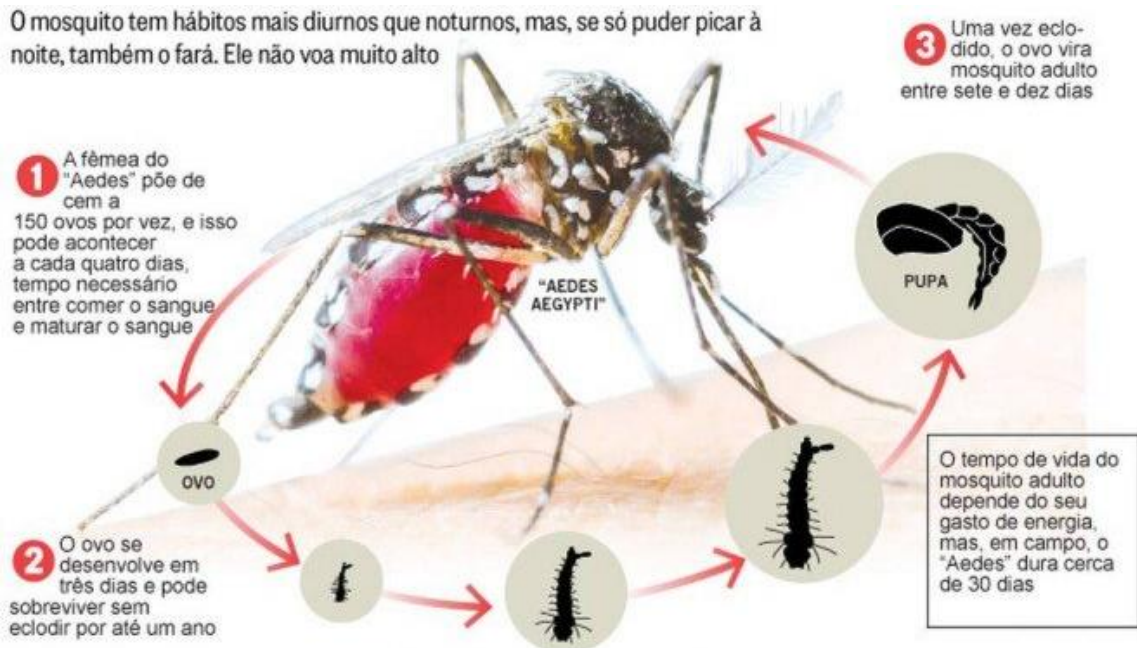
2.1 O MOSQUITO AEADES AEGYPTI

O *Aedes aegypti* é um inseto que apresenta desenvolvimento indireto com metamorfose completa e ciclo de vida de quatro fases: ovo, larva (composto por quatro estágios larvais), pupa e adulto. As fases envolvendo ovo, larva e pupa ocorrem na água. Proveniente da África, em 2017 encontrava-se em muitos lugares do mundo, principalmente nas regiões tropicais e subtropicais. Esse mosquito depende de concentração humana para se estabelecer já que a fêmea necessita de sangue, fonte de proteína para a maturação dos ovos (DIVE, 2019; TSCHIEDEL; BERVIAN, 2017).

A Figura 2 apresenta o ciclo de vida do mosquito transmissor da dengue. Os ovos eclodem quando entram em contato com a água e o desenvolvimento até a fase adulta do mosquito leva um período de 7 a 10 dias (se houver condições ambientais favoráveis). Na vida adulta, o mosquito precisa se reproduzir para manter a sobrevivência da espécie. No acasalamento, a fêmea do mosquito precisa de sangue para nutrir seus ovos e quando pica uma pessoa infectada adquire o vírus e num período de 8 a 12 dias está pronta para transmiti-lo para outras pessoas durante o resto de sua vida (DIVE, 2019).

O mosquito é transmissor de diversas doenças, a dengue está entre elas, sendo uma das principais ameaças à saúde pública ao redor do mundo. O Brasil apresenta cerca de 70% dos casos ocorridos nas Américas (COELHO *et al.*, 2016).

FIGURA 2 - CICLO DE VIDA DO Aedes Aegypti



Fonte: Denise Valle, pesquisadora do Laboratório de Biologia Molecular de Flavivirus do Instituto Oswaldo Cruz, da Fiocruz Editora de Arte

Fonte: CANDIDA; OUCHANA, 2016.

Febre amarela, *chikungunya* e *zika* são outras doenças infecciosas causadas por um vírus da família *Flaviviridae* e transmitidas, na maioria das vezes, pelo mosquito *Aedes aegypti*". Em 2016 confirmou-se, com reconhecimento da Organização Mundial da Saúde (OMS), a associação do Zika vírus com a microcefalia, mais uma doença causada pelo vetor, a qual eclodiu no Nordeste brasileiro no final do ano de 2015 (ELPIDIO DE SÁ *et al.*, 2017; REIS *et al.*, 2017).

O combate e/ou controle ao mosquito precisa ser realizado em conjunto com órgãos governamentais e a participação efetiva dos diversos segmentos da sociedade, contemplando linha de financiamento para ações integradas, de saneamento básico, controle do vetor, vigilância em saúde, educação em saúde, comunicação e mobilização social (BRASIL, 2009; MESQUITA; PARENTE; COELHO, 2017).

O principal método para evitar a transmissão do vírus da dengue é combater o vetor transmissor com o apoio de toda a população (ROCHA; DANTAS; CÂNDIDO, 2014; WHO,

2019). Não é difícil, porém é desafiador, principalmente no Brasil. Segundo Bandeira e Arruda (2013, p. 2):

O Brasil por ser um país tropical, oferece condições ambientais de proliferação do mosquito *Aedes aegypti* causador da dengue, porém não é só o aspecto ambiental que deve ser considerado de forma isolada, também são relevantes as questões culturais, as condições de saneamento básico da população e a rápida adaptação do mosquito vetor às condições ambientais em que se encontra. É uma preocupação não apenas para órgãos governamentais, mas também para os educadores, pois as formas de manter o meio ambiente livre desse mosquito é adotar práticas diárias de combate e trabalho coletivo.

O Brasil é lugar com condições ideais para a proliferação desse vetor e a definição de Políticas Públicas eficientes no seu controle se torna um desafio, pois o modelo de crescimento econômico implementado na região causou crescimento urbano desordenado e não foi possível preencher lacunas (dificuldades para garantir o abastecimento regular e contínuo de água, a coleta e o destino adequado dos resíduos sólidos) no setor de infraestrutura, fatores que contribuem diretamente com o desenvolvimento do mosquito (BRASIL, 2009).

2.1.1 Controle do *Aedes aegypti* no Brasil

O Brasil evoluiu a política pública de combate ao vetor transmissor da dengue. Antes de 1996, as ações eram centralizadas e acreditava-se que era possível erradicar o *Aedes aegypti*, entretanto, percebeu-se a impossibilidade para a erradicação perante as condições ideais para a rápida proliferação do mosquito. Então, o Ministério da Saúde (MS) elaborou plano para o controle da doença que incluiu ações do governo e da sociedade civil (RIBEIRO; MOURA, 2013).

O controle do vetor transmissor da dengue foi inserido no âmbito do Programa de Saúde da Família (PSF). O PSF foi criado em 1994 com equipe multiprofissional de, no mínimo: 1 médico, 1 enfermeiro, 1 técnico ou auxiliar de enfermagem e entre 4 e 6 Agentes Comunitários de Saúde (ACS). Com o passar do tempo, a equipe mínima foi ampliada com a inclusão de 1 odontologista e 1 técnico ou auxiliar de saúde bucal (ROCHA; DANTAS; CÂNDIDO, 2014). O PSF se transformou em Estratégia de Saúde da Família (ESF) e Estratégia de Saúde Bucal (ESB). Alocadas nas Unidades Básicas de Saúde (UBS), cada

equipe cuida de no máximo 4 mil pessoas com média recomendada de 3 mil pessoas. Para cada ACS, a quantidade de pessoas não pode ultrapassar 750 (BRASIL, 2012).

Com base em uma população de aproximadamente 208,5 milhões de pessoas, em maio de 2019, o Brasil tinha a ESF atendendo 63,45% a ESB atendendo 52,66% e ACS atendendo 63,47% das pessoas no território brasileiro. 73,82% da população brasileira tinha acesso aos serviços de Atenção Básica (AB) (E-GESTOR, 2019). Essa capacidade instalada pode ser eficientemente utilizada para o controle do vetor transmissor da dengue.

Além dos profissionais da ESF, o MS incorporou na AB o profissional Agente de Combate as Endemias (ACE), o qual deve atuar nas ações de prevenção e controle de doenças e agravos à saúde, em interação com os ACS e a equipe de AB (BRASIL, 2018). Em 2002, o MS criou o Programa Nacional de Controle da Dengue (PNCD) com o objetivo de integrar as ações de combate do vetor na AB, mobilizando a ESF como coordenadora das atividades dos ACS e ACE que devem agir de forma integrada e complementar a participação ativa da população local (ROCHA; DANTAS; CÂNDIDO, 2014).

A partir das ações da política pública de combate à dengue, em 2009 foram formuladas as Diretrizes Nacionais para Prevenção e Controle de Epidemias de Dengue, com o objetivo de fornecer parâmetros para que os planos de prevenção sigam a mesma lógica de organização em todo o País. As diretrizes definem as responsabilidades de cada esfera do SUS (União, estados e municípios) no controle da dengue, e ainda, auxiliam estados e municípios na organização de suas atividades de prevenção e controle, em períodos de baixa transmissão ou em situações epidêmicas, contribuindo para evitar a ocorrência de óbitos e para reduzir o impacto das epidemias (BRASIL, 2009; RIBEIRO; MOURA, 2013).

Santos *et al.* (2014), realizou estudo comparativo sobre a dengue entre Brasil e Colômbia para avaliar o conhecimento, atitudes e práticas. A conclusão é que o Brasil tem bom controle relativo aos índices de infestação do *Aedes aegypti*, enquanto a Colômbia apresenta alto risco de transmissão. O estudo confirma o estado da arte, o qual evidencia que a estratégia de controle vertical da dengue é inadequada, pois não inclui métodos contextualizados, soluções alternativas e capacitação local.

O Brasil conseguiu estabelecer rede de controle do vetor transmissor da dengue com recursos humanos e conhecimentos, mas ainda não consegue aproveitar com eficiência tais recursos, pois carece de educação permanente para os profissionais de saúde sobre o *Aedes aegypti* e registro eletrônico com geoprocessamento. A capacitação dos recursos humanos e o registro confiável dos dados se tornaram ferramentas primordiais no controle do vetor que

encontra resistência na capacidade de gerência de saúde local e conscientização da população em geral.

É importante destacar que “Quase tudo que acontece, acontece em algum lugar. Saber o local onde algo acontece pode ser fundamental” (LONGLEY *et al.*, 2012, p. 4). Faz sentido afirmar que os acontecimentos estão ancorados no espaço e no tempo. Considerando os dados como acontecimentos é justificável que análises com o objetivo de responder a algum questionamento precisam levar em consideração o lugar e tempo em que tais dados foram gerados.

As análises realizadas com base em dados espaciais possibilitam ao técnico avaliar não só quantitativamente os dados, como também relacionar as informações de saúde com dados ambientais, socioeconômicos e com a posição que o evento ocupa na superfície terrestre, a fim de acompanhar as permanentes mudanças do espaço geográfico e detectar áreas e populações sujeitas a agravos de saúde (FIGUEIREDO, 2009, p. 30).

Em se tratando da dengue, para ter uma visão holística dos locais onde os casos estão surgindo, seria necessário coletar as coordenadas geográficas de onde o paciente, provavelmente, foi infectado (o Brasil ainda não faz). “O reconhecimento geográfico é o primeiro passo para o planejamento das atividades de controle vetorial e consiste na identificação e numeração de quarteirões, bem como na localização e especificação do tipo de imóvel dentro de cada quarteirão. Sua atualização deve ser realizada após o encerramento das atividades de cada ciclo” (BRASIL, 2009).

2.1.2 Internações hospitalares por dengue no Brasil

A infecção pelo vírus da dengue pode passar de assintomática até ocasionar doença grave que coloque em risco a vida do paciente, tudo depende da relação vírus-hospedeiro. Os países endêmicos de dengue da Ásia e da América Latina relatam que a Febre por Dengue Hemorrágica e a Síndrome do Choque por Dengue estão entre as principais causas de internações hospitalares (CHAKRABORTY; CHATTOPADHYAY; GHOSH, 2019; DIAS *et al.*, 2010).

Pessoas podem ter dengue e não perceberem, uma vez que os sintomas não se manifestem visivelmente. Quando os sintomas começam a se manifestar, a pessoa precisa

procurar o serviço de saúde mais próximo para receber os cuidados de acordo com seu caso clínico.

Em caso de suspeita de dengue, a OMS recomenda critérios para a internação hospitalar do paciente: I) grávidas, crianças, idosos, obesos, diabéticos, pessoas com insuficiência renal ou doença hemolítica crônica podem tornar a dengue ou seu manejo mais complicado; II) pacientes que moram sozinhos ou distante de estabelecimentos de saúde sem garantia de transporte; e III) pacientes que precisam de tratamento de emergência e encaminhamento urgente devido a dengue severa (WHO, 2009).

De acordo com Dias *et al.* (2010), considerando o paciente com suspeita de dengue, a internação hospitalar é indicada na presença de **sinais de alarme** ou aumento do hematócrito na vigência de **hidratação adequada**. Se o paciente estiver com plaquetopenia menor que 20.000/mm³, mesmo sem repercussão clínica, ele deve ser internado e reavaliado clínica e laboratorialmente a cada 12 horas.

Os sinais de alarme são: dor abdominal intensa e contínua; vômitos persistentes; hipotensão postural; hipotensão arterial; pressão diferencial < 20 mmHg (PA convergente); hepatomegalia dolorosa; hemorragias importantes (hematêmese e/ou melena); extremidades frias, cianose; pulso rápido e fino; agitação e/ou letargia; diminuição da diurese; diminuição repentina da temperatura corpórea ou hipotermia; aumento repentino do hematócrito e desconforto respiratório.

Sobre o registro eletrônico do atendimento em saúde, o Brasil ampliou o interesse em coletar os registros dos atendimentos realizados aos usuários em estabelecimentos de saúde vinculados ao SUS e armazená-los, eletronicamente, em bancos de dados, os quais são utilizados como base para elaboração de políticas de saúde, planejamento e gestão. Entre os sistemas mantidos pelo DATASUS está o SIH/SUS que recebe, mensalmente, o registro das internações hospitalares ocorridas nos estabelecimentos conveniados ao SUS. Esses dados têm origem nas Autorizações de Internação Hospitalar (AIH), destinadas ao pagamento das internações em hospitais públicos e privados conveniados ao SUS, o que lhe confere confiabilidade para produção de conhecimento científico sobre saúde coletiva (BITTENCOURT; CAMACHO; LEAL, 2006).

No SIH/SUS, as doenças que causaram a internação hospitalar são identificadas pela Classificação Estatística Internacional de Doenças (CID), na versão em português brasileiro mantida pelo MS, CID-10. As doenças transmitidas pelo *Aedes aegypti* são descritas pelos

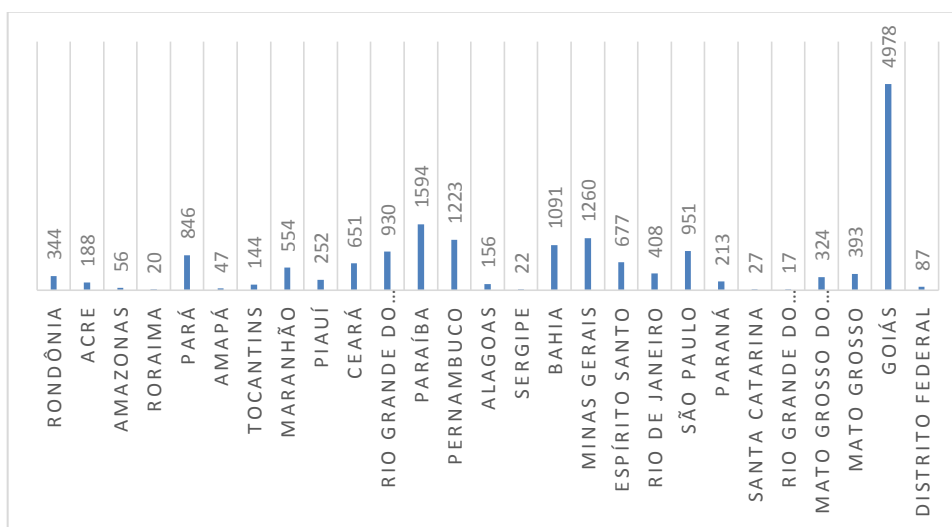
códigos: A95 Febre amarela; A95.0 Febre amarela silvestre; A95.1 Febre amarela urbana; A95.9 Febre amarela não especificada; A90 Dengue clássica; A91 Febre hemorrágica devida ao vírus do dengue; A92.0 Febre de Chikungunya (SIH/SUS, 2019).

O Zika vírus ainda não tinha uma classificação CID-10 própria, mas estava enquadrado como A92.8, outras febres virais especificadas transmitidas por mosquito. No final de 2015, a OMS definiu que o Zika deveria ser codificado como: U06 Doença pelo Zika vírus e U06.9 Doença pelo Zika vírus, não especificada, excluindo-se o até então CID-10 A92.8 (CBCD, 2018).

Os dados das internações hospitalares do mês atual devem ser enviados pelo estabelecimento de saúde de ocorrência para o DATASUS até o final do mês seguinte. É possível enviar dados de internações que ocorreram no prazo de até 90 dias até o mês atual. Como o envio dos dados é condição para que o estabelecimento receba do SUS pela internação hospitalar realizada, a flexibilização do envio permite que o estabelecimento não perca recursos perante a burocracia para a apresentação da internação hospitalar de seus pacientes.

O DATASUS disponibiliza publicamente os dados que servem a pesquisadores para o desenvolvimento de pesquisas dos mais variados tipos e com propósitos diversificados. Conforme Gráfico 1, foram registradas no SIH/SUS 17.453 internações por dengue em 2018, das quais 136 pessoas morreram. Os estados de Goiás, Paraíba e Minas Gerais apresentaram os maiores números de internações.

GRÁFICO 1 - INTERNAÇÕES HOSPITALARES POR DENGUE EM 2018 NO BRASIL



Fonte: (SIH/SUS, 2019).

Em 2019, até o mês de março, teve-se aumento de 303% nos casos notificados da doença no país em comparação com o mesmo período no ano anterior (G1, 2019). Até maio de 2019, as internações hospitalares por dengue registradas no SIH/SUS já somavam-se 27.844, das quais 205 pessoas faleceram (SIH/SUS, 2019). Os dados do primeiro semestre de 2019 já superam os do ano de 2018. As internações por dengue, bem como os óbitos, aumentam de acordo com o crescimento dos casos suspeitos de dengue por local de residência do paciente. “A quase totalidade dos óbitos por dengue é evitável e depende, na maioria das vezes, da qualidade da assistência prestada e da organização da rede de serviços de saúde” (BRASIL, 2009).

Os leitos hospitalares e os recursos (financeiros, físicos e humanos) são limitados. Por isso, é fundamental controlar o vetor transmissor da dengue para que os pacientes cheguem a necessitar o mínimo possível de serviços hospitalares. Na impossibilidade de evitar a internação hospitalar, a gestão precisa conhecer a demanda para que sejam tomadas decisões orientadas ao planejamento.

2.2 SÉRIES TEMPORAIS

Série temporal ou histórica é um conjunto de dados coletados, ordenadamente, em intervalos regulares de tempo, o qual apresenta observações vizinhas dependentes (CHATFIELD, 1996; LATORRE; CARDOSO, 2001). Ela pode ser composta por tendência (T_t), ciclo (C_t), sazonalidade (S_t) e termo aleatório (a_t), onde t representa, na escala temporal, o instante em que o dado ou observação foi registrada. Conforme Hyndman e Athanasopoulos (2018):

- A **tendência** é o aumento ou diminuição da ocorrência dos dados ao longo do tempo por período constante ou quando ocorre mudança de direção, isto é, passagem de tendência crescente para decrescente.
- O fator **sazonalidade** apresenta frequência fixa e conhecida ancorado no tempo, podendo ser identificado a partir dos dias da semana, semanas do mês, meses do ano ou época.
- O comportamento **cíclico** pode ser confundido com o sazonal, mas se diferencia pelo seu caráter eventual e sem frequência fixa.
- O **termo aleatório** ou erro diz respeito as influências na série que não estão relacionadas à tendência, sazonalidade ou ciclo.

A decomposição de uma série temporal e a análise de seus elementos individualmente permite a compreensão dos fenômenos que a originou, ou seja:

Passamos a conhecer como determinado processo se comportará quando conseguirmos caracterizar sua **tendência**, reconhecer sua **variação sazonal e cíclica**, dimensionar sua **variação aleatória** e identificar os fatores associados, responsáveis pelo impacto significativo sobre suas medidas (ANTUNES; CARDOSO, 2015).

Matematicamente, a série temporal (y_t) pode ser decomposta em fatores aditivos $y_t = T_t + S_t + C_t + a_t$. Alternativamente pode se decompor a série temporal em fatores multiplicativos $y_t = T_t * S_t * C_t * a_t$. A decomposição por adição ou multiplicação depende da natureza dos dados (HYNDMAN; ATHANASOPOULOS, 2018). Um exemplo de série temporal seria a coleta da **quantidade diária de pessoas internadas** em unidade hospitalar por ter contraído dengue. Neste caso, tem-se a medida (quantidade de pessoas), o fato (internações hospitalares por dengue) e a unidade de tempo (por dia).

Uma série temporal pode ser classificada em: discreta ou contínua, determinística ou estocástica, estacionária ou não-estacionária e linear ou não-linear com aplicação nas diversas áreas do conhecimento: economia, física, marketing, demografia, controle de processos, etc. (CHAKRABORTY; CHATTOPADHYAY; GHOSH, 2019; CHATFIELD, 1996). De acordo com a classificação, pode-se adotar um ou outro método estatístico mais adequado à análise e previsão de séries temporais.

A estacionariedade, por exemplo, é fator preponderante para a utilização do método (*AutoRegressive Integrated Moving Average* - ARIMA). Para ser **estacionária**, a série temporal não deve apresentar mudança sistemática na média e na variância, bem como não tem variações periódicas, nem tendência ou sazonalidade. Uma série temporal não-estacionária pode ser transformada em estacionária utilizando-se processos de diferenciação para estabilizar a média e a variância, fazendo-se com que sejam diminuídas ou removidas a tendência e a sazonalidade (HYNDMAN; ATHANASOPOULOS, 2018).

Para realizar previsão de séries temporais, faz-se necessário ajustar parâmetros em função matemática de acordo com os padrões encontrados nos dados existentes. Os melhores ajustes podem ser avaliados por métricas que indicam qual o melhor modelo para cada situação.

2.2.1 Métricas de avaliação de modelo e precisão

Segundo Hyndman e Athanasopoulos (2018), quando existem muitos preditores candidatos, ou seja, parâmetros a serem ajustados para encontrar um modelo de previsão de série temporal dada determinada abordagem estatística, a solução é usar uma ou mais métricas de precisão. Seguem as mais populares:

- 1) R quadrado ajustado (R^2);
- 2) Critério de Informação de Akaike (*Akaike Information Criterion - AIC*);
- 3) Critério de Informação de Akaike Corrigido (*Corrected Akaike Information Criterion - AICc*);
- 4) Critério de Informação Bayesiano de Schwarz (*Schwarz's Bayesian Information Criterion - BIC*).

Para AIC, AICc e BIC deseja-se encontrar os menores valores possíveis e para R^2 , procura-se o modelo com maior valor. Considerando que o R^2 se torna menos adequado para a previsão, uma vez que tende a selecionar muitas variáveis preditoras, bem como na realidade raramente ou nunca existem requisitos suficientemente adequados para a perfeição do BIC, recomenda-se utilizar uma das estatísticas AICc ou AIC (HYNDMAN; ATHANASOPOULOS, 2018).

O AIC é amplamente utilizado, pois inclui em sua fórmula de cálculo termo de penalização para o número de parâmetros incluídos no modelo, o que preserva o princípio de parcimônia ao qualificar o ajuste aos dados. Esse critério possibilita a comparação de modelos com diferentes números de parâmetros em sua fórmula (CHATFIELD, 2003).

As métricas supramencionadas são úteis no ajuste dos parâmetros para modelos concorrentes e da mesma classe, mas para comparar o desempenho de previsão entre modelos de classes diferentes, bem como a precisão da previsão deve-se utilizar outra abordagem.

É recomendado dividir a série temporal em dados de treino (80%) e dados de teste (20%), selecionar o modelo com base nos dados de treino e realizar comparação das previsões com os dados de teste, pois são dados novos e permitirão verificar o erro entre os valores previstos e os valores reais; quanto menor for o erro, melhor é o modelo de previsão. Essa medida de desempenho ou precisão permite realizar comparações entre modelos derivados de abordagens estatísticas diferentes (HYNDMAN; ATHANASOPOULOS, 2018).

A precisão de modelos de previsão de séries temporais pode ser calculada com os métodos que Villani *et al.* (2017) descrevem:

- A Equação (1) representa a Raiz do Erro Médio Quadrado (*Root Mean Square Error - RMSE*):

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n (y_t - \hat{y}_t)^2} \quad (1)$$

- Pela Equação (2) é possível calcular o Erro Médio Absoluto (*Mean Absolute Error - MAE*):

$$MAE = \frac{1}{n} \sum_{t=1}^n |y_t - \hat{y}_t| \quad (2)$$

- A Equação (3) possibilita o cálculo do Erro Médio Percentual Absoluto (*Mean Absolute Percentage Error - MAPE*):

$$MAPE = \frac{1}{n} \sum_{t=1}^n \frac{|y_t - \hat{y}_t|}{y_t} \quad (3)$$

Considerando o número de observações e/ou dados n da série temporal, o erro é calculado fazendo-se a diferença entre o valor real (y_t) e o valor previsto (\hat{y}_t). O parâmetro t representa determinada observação ao longo das n iterações a serem feitas no somatório.

O erro aumenta com o tamanho do horizonte de previsão e a instabilidade da série causando desempenho ruim no modelo, pois a incerteza sobre estimativas futuras cresce à medida que se distancia do presente, e intercorrências inesperadas são mais ou menos frequentes nas atividades humanas, o que implica risco adicional de erro não previsto. Quando o comportamento da série é muito instável, melhores projeções devem ser feitas apenas a partir de um passado recente da mesma (ANTUNES; CARDOSO, 2015; LATORRE; CARDOSO, 2001).

Um ótimo modelo de previsão apresenta residuais (diferença entre os dados reais e os valores ajustados pelo modelo) sem autocorrelação, média aproximadamente igual a zero, variância constante e distribuição normal. Se existir autocorrelação nos resíduos, o modelo pode ser melhorado. Espera-se que os resíduos sejam um ruído branco (*White noise*) (HYNDMAN; ATHANASOPOULOS, 2018). As métricas para avaliação do desempenho de previsão de séries temporais norteiam a escolha dos métodos mais adequados aos dados analisados.

2.2.2 Modelos estatísticos para previsão de séries temporais

A análise de séries temporais pode apresentar um ou mais dos objetivos (descrição, explicação, previsão e controle) (CHATFIELD, 1996):

- A **descrição** consegue identificar os componentes da série temporal, bem como a natureza dos dados e seu comportamento. Geralmente utiliza-se a plotagem dos dados em gráficos.
- Quando se deseja saber como uma variável interfere em outra, onde duas ou mais observações estão no mesmo conjunto de dados, usa-se a **explicação** como forma de criar outras séries temporais e entender os relacionamentos existentes.
- A **previsão** de séries temporais é uma estimativa de prováveis valores no futuro. São muito importantes nas vendas, indústria, economia, etc.
- O **controle** tem relação direta com a avaliação da qualidade de processos na manufatura de produtos.

Como o objetivo desta pesquisa é realizar a **previsão** de internações por dengue, serão exploradas as técnicas estatísticas: Suavização Exponencial, ARIMA, Redes Neurais Artificiais Regressivas e a combinação dos modelos anteriores.

2.2.2.1 Suavização Exponencial

Suavização exponencial é um caso especial do algoritmo *Holt-Winters* que faz uso de médias móveis exponencialmente ponderadas, em que os dados mais recentes da série têm maior peso. Pode ser aplicado quando não há evidência de efeitos sazonais ou tendência sistemáticos (COWPERTWAIT; METCALFE, 2009). Para previsão de séries temporais utilizando-se Suavização Exponencial, Ferreira *et al.* (2016), observa:

- Quando não há tendência e nem sazonalidade, usa-se Suavização Exponencial Simples;
- Para séries com tendência e sem sazonalidade, é recomendada a Suavização Exponencial de Holt; e
- Quando apresenta sazonalidade, aplica-se a Suavização de Holt-Winters.

Com variações nas combinações dos componentes de Tendência e Sazonalidade, são possíveis diversos métodos de suavização exponencial. Cada método é resultado da

combinação (T, S), onde T representa a tendência e S representa a sazonalidade da série (HYNDMAN; ATHANASOPOULOS, 2018).

2.2.2.2 ARIMA

Os modelos de Suavização Exponencial e ARIMA são complementares na previsão de séries temporais e com isso se tornaram os mais usados. Enquanto o primeiro se baseia na tendência e na sazonalidade dos dados, o segundo descreve a autocorrelação (HYNDMAN; ATHANASOPOULOS, 2018).

Entre os modelos clássicos para a previsão de séries temporais, o ARIMA é o mais popular, mas a precisão de sua previsão é restrita devido à sua incapacidade para capturar as relações de **não-linearidade** das séries temporais do mundo real. Ele funciona bem com dados lineares e estacionários (CHAKRABORTY; CHATTOPADHYAY; GHOSH, 2019; ZHOU *et al.*, 2018).

A função de autocorrelação (*Autocorrelation function* - ACF) e a função parcial de autocorrelação (*Partial Autocorrelation Function* - PACF) são utilizadas para examinar a estacionariedade e período entre as observações. Elas permitem identificar a ordem dos Termos MA e AR do modelo ARIMA. Também pode ser utilizado o Teste de Dickey-Fuller aumentado (*Augmented Dickey-Fuller* - ADF) (VILLANI *et al.*, 2017; ZHOU *et al.*, 2018).

O ARIMA foi elaborado por Box e Jenkin em 1970. O nome do modelo vem de uma combinação: AR (*Autoregressive Models*), que é modelo autorregressivo, ou seja, projeta a variável de interesse usando valores passados da mesma variável e pode ser representado por $AR(p)$, onde p representa a ordem de defasagem a partir do valor atual; MA (*Moving Average Models*), enquanto Modelo de Média Móvel, usa os erros de previsões anteriores para projetar, regressivamente, o valor da variável resposta e pode ser representado por $MA(q)$, onde q é a ordem de defasagem do erro a partir do valor atual (HYNDMAN; ATHANASOPOULOS, 2018; VILLANI *et al.*, 2017).

Ainda segundo os autores, quando se combina diferenciação com o modelo de autorregressão e o de médias móveis, tem-se o Modelo de Média Móvel Integrado Autorregressivo (ARIMA) para lidar com séries temporais não-sazonais, o qual pode ser representado por $ARIMA(p, q, d)$, onde d representa o grau de derivação. O modelo ARIMA também pode lidar com séries temporais sazonais acrescentando-se parâmetros para a parte sazonal $ARIMA(p, q, d)(P, Q, D)m$, onde m é o número de observações. As letras minúsculas

referem-se aos parâmetros da parte não-sazonal e as letras maiúsculas aos parâmetros da parte sazonal.

Segundo Chakraborty, Chattopadhyay e Ghosh (2019), o método de previsão consiste em 3 etapas interativas: identificação do modelo, ajustes dos parâmetros do modelo e análise dos resíduos:

- Na **identificação ou seleção do modelo**, processos de diferenciação são aplicados em casos de dados não-estacionários. Com a condição de estacionariedade garantida, plotam-se em gráficos as funções ACF e PACF para analisar e selecionar os tipos de modelos AR e MA.
- O **ajuste de parâmetros** é a etapa que se preocupa com a otimização do modelo utilizando-se de métricas como AIC e/ou BIC.
- A **análise residual** é executada para a verificação do desempenho do modelo selecionado nas etapas anteriores.

A dificuldade das técnicas que envolvem AR é a identificação do modelo, pois é possível que pessoas diferentes identifiquem modelos de ordens diferentes para a mesma série temporal (LATORRE; CARDOSO, 2001).

2.2.2.3 *Redes neurais artificiais*

Modelos de análise e previsão de séries temporais com base em Redes Neurais, são modelos com representação matemática das conexões cerebrais, uma rede de neurônios organizada em camadas de entrada e camadas de saída. Tal abordagem consegue identificar relações não-lineares entre as variáveis de resposta e seus preditores. Na série temporal, o modelo de Rede Neural Autorregressiva (*Neural Network Autoregression Model - NNAR*), *feed-forward*, usa valores defasados da série como dados da camada de entrada e pode ser representada por $NNAR(p, k)$, onde p representa os dados da camada de entrada e k a quantidade de nós da camada oculta, considerando que a rede tem apenas uma camada oculta (HYNDMAN; ATHANASOPOULOS, 2018).

De acordo com Chakraborty, Chattopadhyay e Ghosh (2019), para realizar a previsão, o modelo NNAR usa interativamente como função de ativação dentro nos nós da rede, a função logística. Na previsão de um valor futuro, faz-se com os dados disponíveis, enquanto na previsão de dois ou mais valores futuros, utiliza-se os dados disponíveis juntamente com as

previsões anteriormente realizadas. O processo acontece até que o intervalo de previsões definido seja realizado.

2.2.2.4 Modelos combinados

Clemen (1989), concluiu que a precisão nos resultados previstos aumentava com a combinação de múltiplas previsões e em muitos casos a simples média das previsões individuais melhoraram o desempenho do resultado da previsão. As combinações de modelos podem utilizar outras abordagens mais elaboradas além da utilização da média dos resultados das previsões individuais.

A razão para confiar em modelo combinado em vez dos resultados individuais está no fato de que no mundo real a ocorrência dos dados é dinâmica e sofre influência de diversos fatores. Os modelos estatísticos disponíveis não têm flexibilidade para se adequarem perfeitamente a todas as realidades. Chakraborty, Chattopadhyay e Ghosh (2019), por exemplo, percebendo que os resíduos gerados pelo modelo ARIMA ainda contavam com características da série temporal analisada, utilizaram rede neural para realizar previsão com os resíduos e melhoraram a precisão da previsão pela combinação dos dois resultados.

Foram realizadas pesquisas com o uso de médias ponderadas e outras abordagens com elaboração mais complexas, mas o uso da média simples não foi superado. Vale destacar que a combinação é quase sempre próxima ou melhor que os resultados das previsões individuais (HYNDMAN; ATHANASOPOULOS, 2018).

2.2.2.5 Ajuste automático de parâmetros para a escolha do melhor modelo

Os métodos estatísticos de análise e previsão de séries temporais: $ARIMA(p, d, q)(P, D, Q)s$, $NNAR(p, P, k)m$ e Suavização Exponencial (N, A, A_a) combinados dois a dois com (N, A, M) ; requerem ajustes de parâmetros para minimizarem o erro na previsão de resultados futuros (modelo aceitável). Dependendo da natureza dos dados, cada método, individualmente, pode originar diversos modelos a depender do intervalo de tempo e componentes da série temporal analisada (HYNDMAN; ATHANASOPOULOS, 2018; LIZZI, 2012; SILVA, 2018; ZHOU *et al.*, 2018).

Com base na quantidade de parâmetros requeridos pelos métodos supramencionados, é possível inferir dificuldade ou até impossibilidade de utilização e comparação de desempenho

entre todos os possíveis modelos numa mesma análise de séries temporais visando o melhor ajuste.

A computação automática para a escolha de determinado modelo antes de sua utilização na previsão de observações futuras em séries temporais facilita a aplicabilidade dos métodos. Hyndman e Khandakar (2008), analisaram os métodos ARIMA e suas derivações, bem como os de Suavização Exponencial e automatizaram-nos, criando no pacote *forecast* para R as funções *auto.arima()* e *ETS()*, as quais têm a capacidade de receber qualquer série temporal, identificar seus componentes e ajustar os parâmetros para escolher o modelo mais adequado.

O Pseudocódigo 1 apresenta a chamada da função *auto.arima* com os principais argumentos necessários ao seu funcionamento. Esta função foi desenvolvida com o objetivo de retornar o melhor modelo de previsão para uma série temporal univariada com base no método ARIMA. Ela faz uma pesquisa a partir dos parâmetros informados, usando o AIC, AICc ou BIC e retorna como resultado o modelo com melhor ajuste aos dados analisados. Os comentários que não estão incluídos no Pseudocódigo 1, estão apresentados no Quadro 1 por causa da quantidade de palavras envolvidas na explicação.

PSEUDOCÓDIGO 1 – CHAMADA DA FUNÇÃO AUTO.ARIMA COM SEUS ARGUMENTOS

```

auto.arima(
  y, // Série temporal
  d = NA, // Ordem da primeira derivada
  D = NA, // Ordem da primeira derivada para a parte sazonal
  max.p = 5, // Valor máximo de p
  max.q = 5, // Valor máximo de q
  max.P = 2, // Valor máximo P
  max.Q = 2, // Valor máximo de Q
  max.order = 5,
  max.d = 2, // Máximo valor para a diferenciação da parte não sazonal
  max.D = 1, // Valo máximo para a diferenciação da parte sazonal
  start.p = 2, // Valor inicial de p para o procedimento passo a passo
  start.q = 2, // Valor inicial de q para o procedimento passo a passo
  start.P = 1, // Valor inicial de P para o procedimento passo a passo
  start.Q = 1, // Valor inicial de Q para o procedimento passo a passo
  stationary = FALSE,
  seasonal = TRUE,
  ic = c("aicc", "aic", "bic"),
  stepwise = TRUE,
  nmodels = 94, // Máximo número de modelos na busca passo a passo
  trace = FALSE,
  approximation = (length(x) > 150 | frequency(x) > 12),
  method = NULL,
  truncate = NULL,
  xreg = NULL,
  test = c("kpss", "adf", "pp"), // Teste de raiz unitária a ser usado
  test.args = list(), // Argumentos adicionais para sazonalidade
  seasonal.test = c("seas", "ocsb", "hegy", "ch"),

```



```

seasonal.test.args = list(),
allowdrift = TRUE,
allowmean = TRUE,
lambda = NULL,
biasadj = FALSE,
parallel = FALSE, // Para uso em programação paralela
num.cores = 2, // Quantidade de processos paralelos
x = y, // Depreciação da série
... // Outros parâmetros do método ARIMA
)

```

Fonte: Hyndman *et al.* (2019)

O Quadro 1 apresenta a descrição por argumento e/ou variável da função *auto.arima* apresentada no Pseudocódigo 1. As explicações mais diretas, ou seja, mais curtas, estão presentes no próprio Pseudocódigo 1, as demais explicações para compreensão da citada função estão no quadro que segue.

QUADRO 1 - DESCRIÇÃO DOS PRINCIPAIS PARÂMETROS UTILIZADOS NA FUNÇÃO AUTO.ARIMA

Argumento	Descrição
<i>d</i>	Se o valor não for informado, então é escolhido pelo argumento <i>test</i> .
<i>D</i>	Se o valor não for informado, então é escolhido pelo argumento <i>season.test</i> .
<i>max.order</i>	Máximo valor de $p + q + P + Q$ nos casos em que a seleção do modelo não é passo a passo, ou seja, testa todas as possibilidades.
<i>stationary</i>	Se receber o valor <i>TRUE</i> , então restringe a busca por modelos estacionários.
<i>seasonal</i>	Se receber o valor <i>FALSE</i> , então restringe a busca por modelos não sazonais.
<i>ic</i>	Critério de informação a ser usado na seleção do modelo.
<i>stepwise</i>	Se receber o valor <i>TRUE</i> , então a seleção do modelo é mais rápida (busca passo a passo). Caso contrário, a seleção considera todas as possibilidades de combinações, o que pode ser muito lento, principalmente para modelos sazonais.
<i>trace</i>	Se receber o valor <i>TRUE</i> , então reporta a lista de modelos ARIMA considerados.
<i>approximation</i>	Se receber o valor <i>TRUE</i> , então o cálculo é feito através de somas condicionais de quadrados e o critério de informação utilizado para a seleção do modelo é aproximado. O modelo final ainda é calculado usando a estimativa de máxima verossimilhança. A aproximação deve ser usada para séries temporais longas ou um período sazonal alto para evitar tempos de computação excessivos.
<i>method</i>	Método de ajuste, uma probabilidade máxima ou minimizador da soma dos quadrados condicionais. O padrão, a menos que haja valores ausentes, é usar a

Argumento	Descrição
	soma condicional dos quadrados para encontrar os valores iniciais e, em seguida, a probabilidade máxima.
<i>truncate</i>	Um valor inteiro que indica quantas observações devem ser usadas na seleção de modelo. Os últimos valores da variável <i>truncate</i> da série são usados para selecionar um modelo quando <i>truncate</i> não é <i>NULL</i> e <i>approximation = TRUE</i> . Todas as observações são usadas se <i>truncate = NULL</i> ou <i>approximation = FALSE</i> .
<i>xreg</i>	Opcionalmente, um vetor numérico ou matriz de regressores externos que deve ter o mesmo número de linhas que <i>y</i> .
<i>test.args</i>	Argumentos adicionais passados para o teste de raiz unitária.
<i>seasonal.test</i>	Determina qual método deve ser usado para selecionar o número de diferenças sazonais. O método padrão é uma medida da sazonalidade calculada a partir da decomposição da série temporal.
<i>seasonal.test.args</i>	Argumentos adicionais a serem passados para o teste de raiz unitária quando a série temporal é sazonal.
<i>allowdrift</i>	Se <i>TRUE</i> , então modelos com termos <i>drift</i> são considerados.
<i>allowmean</i>	Se <i>TRUE</i> , então modelos com média diferente de zero são considerados.
<i>lambda</i>	Parâmetro para a aplicação da transformação <i>de Box-Cox</i> . Se <i>lambda="auto"</i> então a transformação é automaticamente calculada utilizando a função <i>BoxCox.lambda</i> . Se o parâmetro for <i>NULL</i> , então a transformação é ignorada. De outra forma, é necessário estimar o valor do parâmetro <i>lambda</i> .
<i>biasadj</i>	Se <i>biasadj = TRUE</i> , então é feito um ajuste para produzir previsões médias e valores ajustados. Se os dados transformados forem usados para produzir previsões e valores ajustados, uma transformação inversa regular resulta em previsões medianas.
<i>num.cores</i>	Permite que o usuário especifique a quantidade de processos paralelos que podem ser utilizados nos casos em que <i>parallel = TRUE</i> e <i>stepwise = FALSE</i> .

Fonte: Hyndman *et al.* (2019)

O Pseudocódigo 2 apresenta os principais argumentos da função ETS. As explicações sobre as variáveis declaradas estão como comentário no próprio pseudocódigo e no Quadro 2. Se os valores das variáveis *alpha*, *beta*, *gamma* ou *phi* for *NULL*, então eles são estimados e passados por parâmetro.

PSEUDOCÓDIGO 2 – CHAMADA DA FUNÇÃO ETS COM SEUS ARGUMENTOS

```

ets(
  y, // Série temporal
  model = "ZZZ",
  damped = NULL,
  alpha = NULL, // Valor de alpha
  beta = NULL, // Valor de beta
  gamma = NULL, // valor de gama
  phi = NULL, // Valor de phi
  additive.only = FALSE,
  lambda = NULL,
  biasadj = FALSE,
  lower = c(rep(1e-04, 3), 0.8),
  upper = c(rep(0.9999, 3), 0.98),
  opt.crit = c("lik", "amse", "mse", "sigma", "mae"),
  nmse = 3,
  bounds = c("both", "usual", "admissible"),
  ic = c("aicc", "aic", "bic"),
  restrict = TRUE,
  allow.multiplicative.trend = FALSE,
  use.initial.values = FALSE,
  na.action = c("na.contiguous", "na.interp", "na.fail"),
  ...
)

```

Fonte: Hyndman *et al.* (2019)

O Quadro 2 apresenta uma explicação detalhada sobre cada variável envolvida na chamada da função ETS. A complexidade da função exige a compreensão dos seus parâmetros para sua utilização.

QUADRO 2 - DESCRIÇÃO DOS PRINCIPAIS PARÂMETROS UTILIZADOS NA FUNÇÃO ETS

Argumento	Descrição
<i>model</i>	Três caracteres são utilizados. A primeira letra indica o tipo do cálculo de Erro ("A", "M" ou "Z"); a segunda letra indica o tipo do cálculo da tendência ("N", "A", "M" ou "Z"); e a terceira letra indica o tipo do cálculo da sazonalidade ("N", "A", "M" ou "Z"). Nos casos supramencionados, "N" = nenhum, "A" = aditivo, "M" = multiplicativo e "Z" = selecionado automaticamente.
<i>damped</i>	Se assumir o valor <i>TRUE</i> , usa uma tendência amortecida (aditiva ou multiplicativa). Se assumir o valor <i>NULL</i> , são testadas tendências amortecidas e não amortecidas e o melhor modelo (de acordo com o critério de informação <i>ic</i>) é retornado.
<i>additive.only</i>	Se assumir o valor <i>TRUE</i> , considera apenas modelos aditivos. O padrão é <i>FALSE</i> .
<i>lambda</i>	Parâmetro para a transformação de <i>Box-Cox</i> . Se <i>lambda</i> = "auto", a

Argumento	Descrição
	transformação é selecionada automaticamente usando a função <i>BoxCox.lambda</i> . A transformação é ignorada se <i>lambda = NULL</i> . Caso contrário, os dados para a transformação precisam ser estimados. Quando <i>lambda</i> é especificado, <i>aditivo.only</i> é definido como <i>TRUE</i> .
<i>biasadj</i>	Usa uma média retro transformada ajustada para a transformação de <i>Box-Cox</i> . Se os dados transformados forem usados para produzir previsões e valores ajustados, uma transformação regular inversa resulta em previsões medianas. Quando <i>biasadj = TRUE</i> , é feito um ajuste para produzir previsões médias e valores ajustados.
<i>lower</i>	Menor valor admitido para os parâmetros (alfa, beta, gama e phi).
<i>upper</i>	Maior valor admitido para os parâmetros (alfa, beta, gama e phi).
<i>opt.crit</i>	Critério de otimização: "mse" (erro médio quadrático), "amse" (MSE médio nos primeiros horizontes de previsão nmse), "sigma" (desvio padrão dos resíduos), "mae" (média de resíduos absolutos) ou "lik" (histórico de probabilidade, o padrão).
<i>nmse</i>	Número de etapas para MSE médio de várias etapas ($1 \leq nmse \leq 30$).
<i>bounds</i>	Esse parâmetro pode ser: "usual" indica que todos os parâmetros devem estar entre os valores <i>lower</i> e <i>upper</i> especificados; "admissible" indica que os parâmetros devem estar num intervalo admissível; "both" (parâmetro padrão) faz a interseção entre os anteriores.
<i>ic</i>	Critério de informação a ser usado na seleção do modelo.
<i>restrict</i>	Se <i>restrict = TRUE</i> (padrão), então os modelos com variação infinita não são permitidos.
<i>allow.multiplicative.trend</i>	Se <i>TRUE</i> , modelos com tendência multiplicativa são permitidos ao procurar um modelo. Caso contrário, modelos multiplicativos não são permitidos. Esse argumento é ignorado se um modelo de tendência multiplicativo for declarado explicitamente.
<i>use.initial.values</i>	Se <i>TRUE</i> e a variável <i>model</i> for da classe "ets", então os valores iniciais no modelo não serão reestimados.
<i>na.action</i>	Indica o que deve acontecer quando os dados contêm valores faltantes. Por padrão, a maior parte contínua da série temporal é considerada.

Fonte: Hyndman *et al.* (2019)

Para a Rede Neural Autorregressiva (NNAR) do tipo “*feed-forward networks*” com uma camada oculta, o pacote *forecast* do R conta com a função *nnetar()* que ajusta os

parâmetros $NNAR(p, P, k)m$ automaticamente, caso eles não sejam definidos (HYNDMAN; ATHANASOPOULOS, 2018). O Pseudocódigo 3 contém os principais argumentos da função *nnetar*, os quais estão comentados no próprio pseudocódigo e no Quadro 3.

PSEUDOCÓDIGO 3 – CHAMADA DA FUNÇÃO NNETAR COM SEUS ARGUMENTOS

```
nnetar(
  y, // Série temporal
  p,
  P = 1, // Número de atrasos sazonais utilizados como entrada
  size,
  repeats = 20,
  xreg = NULL,
  lambda = NULL,
  model = NULL,
  subset = NULL,
  scale.inputs = TRUE,
  x = y, // Indica a depreciação da série
  ...
)
```

Fonte: Hyndman *et al.* (2019)

O Quadro 3 apresenta uma explicação mais detalhada sobre as variáveis utilizadas na chamada da função *nnetar*. Tal explanação ajuda na compreensão da utilização da função para a previsão de séries temporais.

QUADRO 3 - DESCRIÇÃO DOS PRINCIPAIS PARÂMETROS UTILIZADOS NA FUNÇÃO NNETAR

Argumento	Descrição
<i>p</i>	Número de atrasos não sazonais usados como entrada. Para séries temporais não sazonais o padrão é o número de defasagens (de acordo com a AIC) para um modelo linear de AR (<i>p</i>). Para séries temporais sazonais, o mesmo método é usado, mas com dados ajustados sazonalmente (de uma decomposição STL).
<i>size</i>	Número de nós da camada oculta. O padrão é metade do número de nós de entrada (incluindo regressores externos, se fornecido) mais 1.
<i>repeats</i>	Número de redes para lidar com diferentes pesos iniciais aleatórios. Elas são calculadas como médias ao produzir previsões.
<i>xreg</i>	Opcionalmente, um vetor ou matriz de regressores externos, que deve ter o mesmo número de linhas da série temporal. Deve ser numérico.
<i>lambda</i>	Parâmetro para a transformação de <i>Box-Cox</i> . Se <i>lambda</i> = "auto", a transformação é calculada automaticamente usando a função <i>BoxCox.lambda</i> . A transformação é ignorada se <i>lambda</i> = NULL. Caso contrário, os dados para a transformação precisam serem estimados. Quando <i>lambda</i> é especificado, <i>aditivo.only</i> é definido como TRUE.

Argumento	Descrição
<i>model</i>	Saída de uma chamada anterior para <i>nnetar</i> . Se o modelo for passado por parâmetro, esse mesmo modelo será ajustado em <i>y</i> sem reavaliar nenhum outro parâmetro.
<i>subset</i>	Vetor opcional que especifica um subconjunto de observações a serem usados no ajuste. Pode ser um vetor de índice inteiro ou um vetor lógico desde que tenha o mesmo tamanho de <i>y</i> . Todas as observações são usadas por padrão.
<i>cale.inputs</i>	Se <i>TRUE</i> , as entradas são dimensionadas subtraindo a média da coluna e dividindo por seus respectivos desvios padrão. Se <i>lambda</i> não for <i>NULL</i> , a escala é aplicada após a transformação <i>Box-Cox</i> .

Fonte: Hyndman *et al.* (2019)

As funções *auto.arima*, *ets* e *nnetar* retornam o modelo com o melhor ajuste aos dados da série temporal analisada de acordo com as especificações de cada método, o qual é utilizado pela função *forecast* para realizar a previsão da série temporal. Os principais argumentos da função *forecast* estão apresentados no Pseudocódigo 4 e comentados, detalhadamente no Quadro 4.

PSEUDOCÓDIGO 4 – CHAMADA DA FUNÇÃO FORECAST COM SEUS ARGUMENTOS

```
forecast(object, ...)
# S3 method for default
forecast(object, ...)
# S3 method for ts
forecast(
  object,
  h = ifelse(frequency(object) > 1, 2 * frequency(object), 10),
  level = c(80, 95),
  fan = FALSE,
  robust = FALSE,
  lambda = NULL,
  biasadj = FALSE,
  find.frequency = FALSE,
  allow.multiplicative.trend = FALSE,
  model = NULL,
  ...
)
```

Fonte: Hyndman *et al.* (2019)

O Quadro 4 apresenta os principais comentários sobre as variáveis declaradas na chamada da função *forecast*. Está é a função do pacote *forecast* disponível para a linguagem R que realiza a previsão de séries temporais. Ela aceita como parâmetro (*object*) os modelos citados anteriormente e outros desde que estejam de acordo com as especificações.

QUADRO 4 - DESCRIÇÃO DOS PRINCIPAIS PARÂMETROS UTILIZADOS NA FUNÇÃO FORECAST

Argumento	Descrição
<i>object</i>	Série temporal ou modelo de série temporal para o qual são necessárias previsões.
<i>h</i>	Número de períodos para previsão.
<i>level</i>	Nível de confiança para intervalos de previsão.
<i>robust</i>	Se <i>TRUE</i> , a função consegue lidar com valores ausentes e <i>outliers</i> no objeto. Este argumento é válido apenas quando o objeto é da classe <i>ts</i> .
<i>find.frequency</i>	Se <i>TRUE</i> , a função determina o período apropriado, se os dados forem de período desconhecido.
<i>allow.multiplicative.trend</i>	Se <i>TRUE</i> , modelos da classe <i>ETS</i> com tendências multiplicativas são permitidos. Caso contrário, apenas os modelos aditivos ou modelos <i>ETS</i> sem tendência são permitidos.
<i>model</i>	Um objeto que descreve um modelo de série temporal; por exemplo, <i>ets</i> , <i>arima</i> , <i>bats</i> , <i>tbats</i> , <i>nnetar</i> , etc.

Fonte: Hyndman *et al.* (2019)

Dessa forma, pode-se utilizar a função *forecast()* para a realização da previsão, passando como parâmetros, individualmente, o modelo *auto.arima()*, *ETS()*, *nnetar()* ou outros, executados previamente, e o horizonte de previsão.

A automatização na escolha dos modelos entre as possibilidades disponíveis por cada método viabiliza a utilização de diversos modelos para encontrar o que mais se ajusta à série temporal de acordo com sua natureza e intervalo de tempo. A previsão dos casos futuros de dengue sempre estará sujeita a erros devido à dinâmica de mudanças dos fatores que influenciam a doença, o que implica em cuidado quanto às decisões de planejamento e tomadas com base nas previsões (CHAKRABORTY; CHATTOPADHYAY; GHOSH, 2019). Minimizar, ao máximo, o erro é o principal fator na abordagem de previsão.

A escolha automática de modelo de previsão de séries temporais foi aplicada, por exemplo, em Ruchiraset e Tantrakarnapa (2018), que utilizaram o *auto.arima()* com a função *stepwise* (seleciona a melhor ordem do modelo com base no valor AIC) do pacote *forecast* do R para obter o melhor ajuste dos parâmetros ARIMA $(p, d, q)(P, D, Q)s$ com o objetivo de modelar a previsão de internações por pneumonia.

2.3 SÉRIES TEMPORAIS APLICADAS AO ESTUDO DA DENGUE

Na epidemiologia, a necessidade de prever o futuro para intervir nas ações do presente é assunto de vida ou morte, pois a redução de doenças na população depende da efetividade deste esforço. A análise de séries temporais contempla tal preocupação e pode derivar conhecimentos sobre a movimentação recente das medidas de interesse em saúde, prever resultados e reconhecer fatores que interferem sobre eles (ANTUNES; CARDOSO, 2015).

Na análise de uma série temporal, primeiramente deseja-se modelar o fenômeno estudado para, a partir daí, descrever o comportamento da série, fazer estimativas e, por último, avaliar quais os fatores que influenciaram o comportamento da série, buscando definir relações de causa e efeito entre duas ou mais séries (LATORRE; CARDOSO, 2001).

A dengue apresenta características sazonais e no Brasil, esse período coincide com o verão, devido ao aumento da temperatura e a maior ocorrência de chuvas, quando o mosquito encontra as condições ideais de reprodução. Cerca de 70% dos casos de dengue ocorrem nos períodos de janeiro a maio, mas apesar disso o Brasil tem condições climáticas para o desenvolvimento do mosquito transmissor da dengue o ano inteiro (RIBEIRO; MOURA, 2013). A vinculação da ocorrência dos casos de dengue ao aumento da temperatura e presença de chuvas, faz com que a série histórica desta doença tenha potencial para se aplicar com sucesso técnicas de previsão de séries temporais.

Modelos estatísticos de análise de séries temporais têm sido utilizados para monitorar e prever a **incidência** de dengue e outras doenças infecciosas (LIZZI, 2012). Ainda de acordo com a autora, os modelos ARIMA (útil quando os dados da série temporal exibem flutuações sazonais periódicas) são os mais utilizados na compreensão da dinâmica da doença, permitindo predições do número de casos em períodos subsequentes à serie estudada.

O Quadro 5 mostra o resumo das técnicas estatísticas de previsão de séries temporais utilizadas no estado da arte e os modelos selecionados em estudos de incidência e internações hospitalares por dengue e outras doenças infecciosas. Encontraram-se aplicações com Redes Neurais Artificiais e o modelo ARIMA, bem como o uso de modelos combinados. É consenso entre os pesquisadores dos trabalhos relacionados que o modelo ARIMA é o mais poderoso para lidar com a previsão de séries temporais, mas é importante destacar a possibilidade de utilização de outras técnicas diante da realidade encontrada nos dados analisados.

QUADRO 5 - TÉCNICAS DE ANÁLISE E PREVISÃO DE SÉRIES TEMPORAIS MAIS UTILIZADAS NA PREVISÃO DE INCIDÊNCIA DE DENGUE E INTERNAÇÕES HOSPITALARES

Autor(es)	Técnicas de séries temporais usadas	Modelo selecionado
Chakraborty, Chattopadhyay e Ghosh (2019)	ARIMA, NNAR e ARIMA-NNAR	ARIMA-NNAR apresentou melhor precisão para dados com linearidade e não-linearidade suficientes
Zhou <i>et al.</i> (2018)	SARIMA, NARNN e SARIMA-NARNN	O método híbrido SARIMA-NARNN não melhorou os resultados encontrados
Silva (2018)	ANN	ANN apresentam sucesso na previsão de incidência de dengue, desde que tenham seus parâmetros ajustados de acordo com os dados analisados
Ruchiraset e Tantrakarnapa (2018)	ARIMA	ARIMA (1,0,2)(2,0,0)[12] com poluição do ar (PM ₁₀)
Villani <i>et al.</i> (2017)	ARIMA	SARIMA (0,1,0,12) apresentou o melhor ajuste
Lizzi (2012)	SARIMA e SARIMAX	Os modelos SARIMAX (1,1,1) (1,1,1) ₁₂ para Campinas e SARIMAX (1,1,1) (0,1,1) ₁₂ Para Ribeirão Preto apresentaram melhor precisão

Fonte: Autoria própria, 2019.

Realizar previsões de doenças infecciosas em tempo real, na prática é um desafio (Logístico, Computacional e Estatístico). Pesquisadores trabalham para entender e estabelecer conjunto de melhores práticas para a implementação de algoritmos de previsão em tempo real (REICH *et al.*, 2016):

- **Logisticamente** os dados históricos devem estar disponíveis de forma padronizada para ser processados, analisados e as previsões devem ser formalmente registradas e arquivadas em banco de dados disponíveis publicamente.

- São necessários recursos **computacionais** para processar e analisar os dados com foco na criação do modelo de previsão.
- **Estatisticamente**, os desafios estão relacionados a análises para treinamento, seleção e validação dos modelos de previsão apropriados considerando a natureza dos dados.

O motivo apresentado pode explicar o fato de que as pesquisas encontradas, todas, abordam a previsão de séries temporais a curto prazo e a longo prazo, não se preocupando com a previsão em tempo real. A estimativa de demanda pode ser ferramenta indispensável no apoio à tomada de decisões na gestão dos serviços de saúde e a previsão de séries temporais possui potencial para gerar tais estimativas.

2.4 PESQUISAS RELACIONADAS

Em Chongqing, China, Zhou *et al.* (2018), utilizaram análise de séries temporais para prever o número diário e mensal de novas internações de pacientes em unidade hospitalar. A pesquisa utilizou o ARIMA em combinação com NARNN.

Três abordagens foram aplicadas para a obtenção dos resultados: ARIMA para série temporal com sazonalidade (SARIMA), NARNN e SARIMA-NARNN. Concluiu-se que o modelo SARIMA-NARNN não superou o desempenho dos modelos que o constituíram, pois a precisão foi menor.

Estudo ecológico com uso de análise de série temporal, conduzido por Lizzi (2012), com dados das cidades de Ribeirão Preto e Campinas, ambas no estado de São Paulo, mostrou que é possível realizar “predição do número mensal de casos de dengue por modelos de séries temporais”.

O estudo utilizou os registros mensais de casos confirmados de dengue, entre os anos de 1998 e 2008, obtidos do Sistema de Informação de Agravos de Notificação (SINAN). A série temporal foi analisada utilizando-se os modelos de Box & Jenkins, popularmente conhecidos como ARIMA ou SARIMA com extensões para o modelo SARIMAX que permitem incorporar covariáveis (chuvas e temperatura mínima ocorridas em meses anteriores ao caso notificado de dengue) como as temperaturas máxima, média e mínima e a precipitação média a cada mês.

Os resultados demonstram que os ajustes dos modelos SARIMA variaram de uma população a outra, sugerindo diferentes padrões temporais da doença de acordo com características locais. Conclui-se que os modelos SARIMA apresentam sucesso na previsão

de novos casos de dengue e podem ser utilizados com benefícios em programas de prevenção e planejamentos de serviços públicos de Vigilância em Saúde.

A pesquisa de Silva (2018), intitulada de “Metodologia de monitoramento de epidemias: uma abordagem baseada em redes neurais artificiais”, utilizou análise de séries temporais com dados dos casos de dengue obtidos a partir do SINAN e as variáveis climáticas obtidas do Instituto Nacional de Meteorologia (INMET). Os dados foram filtrados para os municípios de Belém, Parauapebas, Altamira e Santarém, estado do Pará, e serviram como base para a criação de sistema de alerta de surtos de dengue e estudo de caso aplicado nos supramencionados municípios.

Os resultados indicaram que as Redes Neurais Artificiais (RNA) conseguem aprender o comportamento de cada série individual e criar previsões com precisão satisfatória, mas são necessários diversos ajustes para encontrar o melhor modelo que represente a série em análise dependendo do intervalo de tempo analisado.

A proposta de pesquisa de Chakraborty, Chattopadhyay e Ghosh (2019), desenvolvida na cidade de Kolkata, estado indiano de West Bengal, foi criar uma abordagem híbrida com a utilização do modelo ARIMA para lidar com os aspectos lineares dos dados e modelo de NNAR, para lidar com as características não-lineares. A pesquisa levou em consideração, apenas, a série histórica dos casos de dengue (série temporal univariada), desprezando os conhecimentos sobre a epidemiologia da dengue e as variáveis climáticas.

O Modelo híbrido ARIMA-NNAR foi formado com ARIMA identificando os padrões lineares da série temporal e NNAR utilizando os valores residuais do ARIMA para lidar com os padrões não-lineares dos dados. O resultado foi observado pela melhoria na capacidade para a previsão de epidemias de dengue, com a metodologia híbrida superando as deficiências dos métodos individuais.

Na província de Chiang Mai na Tailândia, Ruchiraset e Tantrakarnapa (2018), desenvolveram pesquisa com o objetivo de utilizar séries temporais para prever, mensalmente, internações por pneumonia associadas à poluição do ar que é medida em 10 microgramas por metro cúbico (PM_{10}) de concentração, monóxido de carbono (CO), dióxido de nitrogênio (NO_2), dióxido de enxofre (SO_2) e ozônio (O_3) e variáveis climáticas: chuva, umidade e temperatura. Eles utilizaram o modelo ARIMA sazonal em doze cenários planejados e conseguiram sucesso na abordagem que apontou a poluição do ar como fator preponderante na ocorrência dos casos e, conseqüentemente, no modelo de série temporal selecionado.

O modelo ARIMA também foi utilizado por Villani *et al.* (2017), para estimar, mensalmente, a demanda de diabéticos que necessitariam do Serviço Médico de Emergência pré-hospitalar. A pesquisa foi conduzida no estado de Victoria na Austrália e demonstrou que é possível conhecer a demanda antes que ela chegue ao Serviço de Emergência.

3 MATERIAL E MÉTODO

3 MATERIAL E MÉTODO

Neste capítulo são apresentados: o local onde foi realizada a pesquisa, as principais características, delineamento, a coleta e seleção dos dados, bem como as ferramentas computacionais utilizadas para execução.

3.1 CARACTERIZAÇÃO DA PESQUISA

Uma pesquisa científica pode ser classificada seguindo critérios. Wazlawick (2017), afirma que é possível diferenciar os tipos de pesquisa de acordo sua natureza, objetivos e procedimentos técnicos. Neste sentido, esta pesquisa se caracteriza:

- Quanto à natureza como *trabalho original*, pois busca apresentar conhecimento novo a partir de observações e teorias construídas para explicá-las (pesquisa exploratória). As observações são o registro das internações hospitalares por dengue no Brasil.
- Do ponto de vista dos objetivos, trata-se de *pesquisa explicativa*, já que busca analisar dados observados para encontrar seus fatores determinantes, ou seja, os padrões existentes para a previsão de séries temporais (pesquisa quantitativa).
- No que diz respeito aos procedimentos técnicos, utiliza-se de *pesquisa bibliográfica* para a fundamentação teórica, bem como *pesquisa experimental* para experimentar diferentes técnicas estatísticas visando analisar a relação de dependência (autocorrelação) na série temporal analisada e encontrar o melhor modelo de previsão com base nos dados em análise (estudo de caso).

A pesquisa desenvolveu-se no CAMPUS I, da Universidade Estadual da Paraíba (UEPB), na cidade de Campina Grande/PB, mas tem abrangência nacional, uma vez que lida com dados de internações hospitalares por dengue de todas as Unidades da Federação do Brasil (APÊNDICE A) para previsão de série temporal.

3.1.1 Levantamento bibliográfico

A análise e compreensão do estado da arte seguiu duas abordagens: Protocolo de Revisão Sistemática da Literatura (APÊNDICE B); e pesquisa bibliográfica tradicional em bases eletrônicas.

A questão principal de pesquisa da RSL foi: Quais técnicas de séries temporais têm sido pesquisadas para criar modelos de previsão de internações hospitalares? As questões secundárias foram:

1. Dentre as técnicas de previsão de séries temporais encontradas, quais estão relacionadas com a previsão de incidência de dengue?
2. Dentre as técnicas de previsão de séries temporais encontradas, quais estão relacionadas com previsão de internações por dengue?
3. Dentre as técnicas de previsão de séries temporais encontradas, quais estão relacionadas com previsão de internações por doenças diversas?
4. Dentre as técnicas de previsão de séries temporais encontradas, quais são mais indicadas para a previsão de internações hospitalares e/ou incidência de dengue?

Os critérios de inclusão (CI) utilizados foram:

- **CI-01** Publicações com séries temporais relacionadas à previsão de internações hospitalares por dengue;
- **CI-02** Publicações com séries temporais relacionadas à previsão de internações hospitalares diversas;
- **CI-02** Publicações que apresentaram séries temporais na previsão de incidência de dengue.

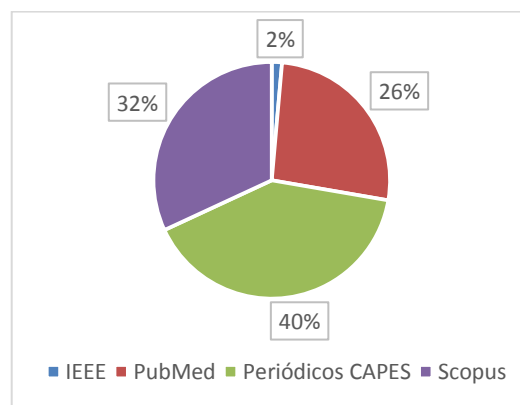
Os critérios de exclusão (CE) estabelecidos foram:

- **CE-01** Publicações com séries temporais não relacionadas à previsão de internações hospitalares ou incidência de dengue;
- **CE-02** Pesquisas sem relação com as palavras-chave (podem ser considerados sinônimos);
- **CE-03** Trabalhos não escritos em inglês e/ou português;
- **CE-04** Publicações em que as palavras-chave da expressão de busca não constaram no título, resumo e/ou texto da publicação (considerar apenas os elementos textuais);
- **CE-05** Trabalhos publicados como artigos curtos ou pôsteres;
- **CE-06** Pesquisas que apresentem análises preditivas sem métodos e/ou técnicas de séries temporais;
- **CE-07** Publicações duplicadas;
- **CE-08** Publicações como revisão de literatura;
- **CE-09** Trabalho relacionado a inteligência artificial, *big data* ou *data mining*;

- **CE-10** Pesquisas com previsão de internação ou incidência de dengue sem a aplicação de séries temporais.

A expressão de busca ("*time series*" AND (*predict** OR *forecast**) AND *hospital** AND dengue) foi aplicada nas bases de busca: PubMed, Portal de Periódicos da CAPES, IEEE e Scopus e retornou um total de 1144 publicações distribuídas por base como se apresenta no Gráfico 2. A seleção das publicações foi realizada a partir da leitura do título e resumo de cada publicação, classificando-as de acordo com os critérios de inclusão e exclusão estabelecidos.

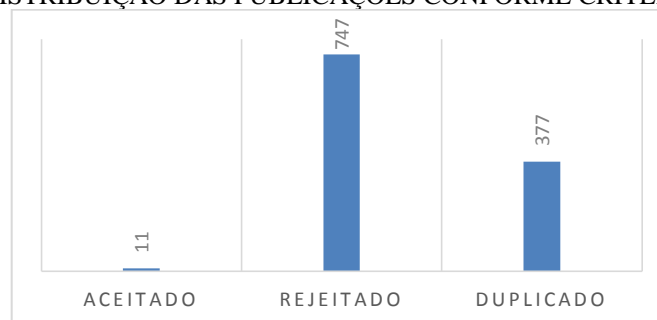
GRÁFICO 2 - DISTRIBUIÇÃO DAS PUBLICAÇÕES POR BASE DE BUSCA



Fonte: Autoria própria, 2019.

Realizada a etapa de seleção das publicações conforme os critérios de seleção estabelecidos, restaram-se 11 publicações. O Gráfico 3 mostra o número de publicações duplicadas, rejeitadas e aceitadas. A partir do estado da arte, não se encontrou muitas pesquisas relacionadas à previsão de internações hospitalares ou incidência por dengue com análise de séries temporais, o que explica a pequena quantidade de estudos selecionados frente aos pesquisados.

GRÁFICO 3 - DISTRIBUIÇÃO DAS PUBLICAÇÕES CONFORME CRITÉRIOS DE SELEÇÃO



Fonte: Autoria própria, 2019.

A partir da seleção dos estudos pela execução do Protocolo de Revisão Sistemática, buscou-se ampliar o número de referências com pesquisas via *web* por meio do método tradicional. O resultado deste levantamento embasou o capítulo de revisão da literatura.

3.1.2 Fonte e seleção dos dados

Hyndman e Athanasopoulos (2018), afirmam que a quantidade de dados que formam a série temporal é preponderante para a realização de análise e previsão. Não existe consenso sobre a quantidade mínima ou máxima de registros, tudo depende do número de parâmetros requerido por cada modelo e da aleatoriedade dos dados. O tamanho necessário da amostra aumenta com o número de parâmetros a ser estimados e a quantidade de ruídos existentes nos dados. Séries muito pequenas ou muito grandes prejudicam o ajuste do modelo. Logo, sugere-se não mais do que 200 observações. Neste sentido utilizaram-se os dados de 182 semanas epidemiológicas compreendidas entre a semana 1 de 2016 e a semana 26 de 2019 (APÊNDICE A) obtidos do portal do DATASUS.

Os dados foram acessados no site do DATASUS com os critérios: **dados; RD – AIH Reduzida; 2016, 2017, 2018 e 2019; Todos estados brasileiros; Todos os meses entre 2016 e o primeiro semestre de 2019** (DATASUS, 2019). Os arquivos estão consolidados por mês e mantêm nomenclatura que indica arquivo do tipo reduzido (RD), o estado de produção do arquivo (UF), o ano de realização dos atendimentos com os dois últimos dígitos (AA) e o mês de atendimento (MM).

Os arquivos estão disponíveis no formato *.dbc*, os quais são visualizados utilizando a ferramenta TabWin que apresenta limitações sobre a forma como se deseja agrupá-los. As limitações restringem o planejamento e a ação do pesquisador aos recursos de seleção e agrupamento de dados disponíveis no TabWin. Como alternativa, tem-se o pacote *read.dbc* para a linguagem de programação R, o qual é capaz de ler os arquivos disponibilizados pelo DATASUS de maneira flexível e independente de Sistema Operacional (PETRUZALEK, 2016).

O dicionário de dados da estrutura dos arquivos AIH “RD” (arquivo reduzido liberado pelo DATASUS) mostra que a base de dados tem 113 colunas (variáveis) e inúmeras linhas. Foi preciso realizar seleção das internações provenientes de dengue a partir dos códigos CID-10 (A90 Dengue clássico e A91 Febre hemorrágica devida ao vírus do dengue) para restringir o número de registros, linhas da base de dados, bem como o número de colunas.

Selecionadas as internações por dengue, criou-se uma base de dados (stInterDengueUF.csv) para cada estado brasileiro. A base de dados tem a quantidade de internações hospitalares por dengue agrupadas pela semana epidemiológica de ocorrência e foi salva em arquivo de texto com Valores Separados por Vírgula (*Comma Separated Values - CSV*). O arquivo stInterDengueUF.csv contém a série temporal de determinado estado e foi utilizada para a previsão das internações futuras.

3.2 PREVISÃO COM MELHOR PRECISÃO

Para a previsão das séries temporais, foi utilizado o pacote *forecast* e os gráficos necessários foram gerados com o pacote *ggplot2*, ambos disponíveis para a linguagem de programação R. Os algoritmos foram criados com o auxílio do Ambiente Integrado de Desenvolvimento (*Integrated Development Environment - IDE*) RStudio. O computador utilizado foi um notebook com configuração básica (Memória RAM de 4 GB *Single Channel* DDR3L 1600 MHz, Processador Intel(R) Core(TM) i3-5010U de 2,10 GHz e Disco Rígido com espaço para 500 GB de armazenamento).

Cada série temporal foi dividida em dados de treinamento (da 1ª semana de 2016 até a 10ª semana de 2019) e dados de teste (da semana 11 a semana 18 de 2019). Como precisava-se de 8 semanas para a calcular o erro na precisão da previsão, então considerou-se o restante da série temporal para a realização do treinamento do modelo.

Villani *et al.* (2017), afirmam que tal estratégia permite ao modelo utilizar os dados de treinamento para identificar os padrões e criar o melhor ajuste e os dados de teste para validar o modelo, uma vez que são dados novos que o modelo não conhece e permitem averiguar a precisão. A divisão dos dados em conjunto de treinamento e teste observou a variação sazonal que indica aumento do número de casos de dengue com picos de máximo no meio do primeiro semestre de cada ano.

A análise de séries temporais considerou a tendência, sazonalidade, estacionariedade, os *outliers* e os residuais para que fossem construídos modelos de previsão baseados na natureza dos dados (HYNDMAN; ATHANASOPOULOS, 2018). Neste sentido, o tratamento dos *outliers* foi obtido com a função *tsclean()* que usa interpolação linear com possível ajuste sazonal para substituir os valores faltantes, encontrar e eliminar os *outliers* na série temporal. Os modelos foram testados com e sem a presença dos *outliers* para verificar a melhoria ou

não da precisão, logo existiram duas séries, uma original e outra com tratamento dos valores discrepantes.

Por causa do atraso de 4 semanas que o DATASUS tem para liberar os dados de internações publicamente, a previsão foi definida para horizonte de 8 semanas e com isso tem-se previsão real de 4 semanas. Como a presença dos *outliers* e a realização de transformações nos dados interferem diretamente na precisão da previsão, bem como o modelo ARIMA é o mais utilizados nas pesquisas, foram planejados 10 cenários de experimentos:

1. Série temporal original;
2. Série temporal original com aplicação da transformação de Box-Cox;
3. Série temporal sem *outliers*;
4. Série temporal sem *outliers* e aplicação da transformação de Box-Cox;
5. Série temporal original e os parâmetros (*approximation=FALSE*, *stepwise=FALSE*) e sem garantia de estacionariedade da série (sem diferenciação explícita). Neste caso, o modelo se ocupa de escolher a melhor diferenciação no método ARIMA;
6. Série temporal sem *outliers* e os parâmetros (*approximation=FALSE*, *stepwise=FALSE*) e sem garantia de estacionariedade da série (sem diferenciação explícita) no método ARIMA;
7. Série temporal original e os parâmetros (*approximation=FALSE*, *stepwise=FALSE*, *stationary=TRUE*). Neste caso, tem-se a garantia de estacionariedade da série (com diferenciação explícita) no método ARIMA;
8. Série temporal sem *outliers* e os parâmetros (*approximation=false*, *stepwise=FALSE*). Neste caso, tem-se a garantia de estacionariedade da série (com diferenciação explícita) no método ARIMA;
9. Utilização da série temporal original, ou seja, como foi obtida do DATASUS;
10. Série temporal sem *outliers*, isto é, foram eliminados os valores discrepantes em relação à maioria dos dados presentes na série histórica analisada.

Sobre os quais foi experimentada a previsão de internações hospitalares por dengue utilizando-se todos os métodos estatísticos apontados nesta pesquisa. Cada método estatístico de previsão de séries temporais tem suas particularidades (número de parâmetros a ser ajustados); as seções seguintes explicam como cada método aqui utilizado foi abordado.

3.2.1 ARIMA

O modelo ARIMA $(p, d, q)(P, D, Q)$ s é o que apresenta maior número de parâmetros a serem ajustados nos dados da série temporal de treinamento e fazer isso analisando gráficos e utilizando abordagem não automática tornaria os objetivos da pesquisa difíceis de serem alcançados. Neste sentido, foi utilizada a função *auto.arima(treino)*, que foi criada por Hyndman e Khandakar (2008) e tem a capacidade de ajustar os parâmetros e selecionar o modelo mais adequado, dentre todos os candidatos. O retorno da função *auto.arima(treino)* foi armazenado na variável **ARIMA**.

A próxima etapa é utilizar a função *forecast(ARIMA, 8)*, a qual recebe como parâmetros o modelo ajustado e o horizonte de previsão e realiza a estimativa das internações hospitalares por dengue para as 8 semanas seguintes. Depois da previsão, realizou-se a aferição do erro do modelo com base nos dados de teste.

3.2.2 Suavização exponencial

Seguindo o exemplo do modelo ARIMA, Hyndman e Khandakar (2008), também automatizaram o processo de ajuste dos parâmetros para o modelo de suavização exponencial e criaram a função *ETS(treino)*. O nome da função é abreviação para seus componentes: Erro, Tendência e Sazonalidade. Esta função consegue automatizar o processo de ajuste dos parâmetros e retorna o melhor modelo de acordo com a série temporal analisada, o qual foi armazenado na variável **ETS**.

Após a escolha do modelo de melhor ajuste, utilizou-se a função *forecast(ETS, 8)* com os devidos parâmetros, que realizou a previsão para as 8 semanas seguintes. Terminou-se com o cálculo do erro MAPE do modelo sobre os dados de teste.

Utilizou-se também nesta categoria o método TBATS, uma técnica que aplica suavização exponencial agregada com a transformação de Box-Cox, erros ARMA e componentes de tendência e sazonalidade na previsão das séries temporais. O ajuste automático dos parâmetros foi realizado com a função *tbats(treino)* e a previsão com a função *forecast(TBATS, 8)*.

O método da tendência linear de Holt (HOLT) retorna como previsão a tendência da série temporal analisada. Esta técnica foi utilizada com a função *holt(treino)* passando como parâmetros o conjunto de dados de treinamento e a quantidade de semanas a ser previstas.

3.2.3 Rede Neural Autorregressiva

Para a utilização da abordagem baseada em Rede Neural Autorregressiva, foi utilizada a função *nnetar(treino)*, a qual realizou o ajuste dos parâmetros com base na série temporal analisada e retornou o modelo de melhor ajuste. O retorno da função foi armazenado na variável **NNAR**.

A previsão foi realizada com a função *forecast(NNAR, 8)* considerando o horizonte de 8 semanas para estimativa de internações hospitalares por dengue. A validação do modelo foi feita utilizando os dados de teste.

3.2.4 Outros métodos estatísticos

Também foram utilizados outros métodos estatísticos de previsão de séries temporais:

- O modelo de previsão por decomposição da série temporal (STL) que segue a mesma rotina dos modelos apresentados anteriormente com o ajuste automático de parâmetros. Foi utilizada a função *stlf(treino)* para o ajuste dos parâmetros e a previsão foi realizada com a função *forecast(STL, 8)*;
- O método inocente que lida com sazonalidade (SNAIVE) repete como previsão o último intervalo sazonal do ano. Foi utilizada a função *snaive(treino, 8)* passando como parâmetros a série temporal e a quantidade de semanas a serem previstas;
- O método *DRIFT* é uma derivação do método inocente e tem por finalidade encontrar a sazonalidade da série temporal analisada. Ele foi utilizado a partir da função *rwf(treino, drift=TRUE, 8)*.
- A técnica de regressão linear também foi utilizada a partir da função *tslm(treino ~ trend + season)* que utiliza a decomposição da série temporal em tendência e sazonalidade para a realização da previsão. A previsão foi feita a partir da função *forecast(TSLM, 8)*.

3.2.5 Modelo combinado

O modelo combinado foi formado pela multiplicação por um fator de ajuste das previsões obtidas a partir dos métodos: HOLT, SNAIVE e a projeção da série temporal utilizada na previsão. A previsão do modelo combinado também levou em consideração o horizonte de 8 semanas de previsão. A estimativa do erro para o modelo foi realizada com a

utilização dos dados de teste tal qual as demais abordagens. O fator de ajuste foi empiricamente escolhido utilizando-se as razões: $\frac{1}{2}$, $\frac{1}{3}$, $\frac{1}{4}$ e $\frac{1}{5}$. Entre os quais, $\frac{1}{2}$ (0,5) foi o que apresentou melhor resultado nos experimentos.

3.2.6 Seleção do modelo com melhor precisão

Após a previsão de cada série temporal por cenário definido, foram calculadas as principais métricas de avaliação de desempenho e precisão com a função *accuracy()*, para verificar o erro entre a previsão e os valores reais. O resultado foi registrado com o nome da série temporal, o modelo de melhor ajuste por método, o cenário e o valor da métrica de erro MAPE com base nos dados de teste.

Os dez métodos de previsão de séries temporais comentados nesta pesquisa foram utilizados para realizar a previsão de cada série temporal com base nos cenários definidos e de acordo com o algoritmo do Pseudocódigo 5.

PSEUDOCÓDIGO 5 - ALGORITMO PARA A ESCOLHA DO MODELO DE PREVISÃO DE SÉRIES TEMPORAIS MAIS PRECISO

```
MELHOR-MODELO(T) // recebe a lista de séries temporais
1  M //lista com o melhor modelo por UF
2
3  for i = 1 to T.comprimento
4      treino = window(T[i], start=c(2016, 1), end=c(2019, 10))
5      trainholt = window(T[i], start=c(2019, 1), end=c(2019, 10))
6      trainaive = window(T[i], start=c(2018, 1), end=c(2019, 10))
7      trainets = window(T[i], start=c(2018, 38), end=c(2019, 10))
8      projecao = window(T[i], start=c(2019, 3), end=c(2019, 10))
9      teste = window(T[i], start=c(2019, 11), end=c(2019, 18))
10
11     ARIMA = forecast(auto.arima(treino), 8)
12     ETS = forecast(ets(treinoets), 8)
13     STL = forecast(stlf(treino), 8)
14     NNAR = forecast(nnetar(treino), 8)
15     TBATS = forecast(tbats(treino), 8)
16     TSLM = forecast(tslm(treino ~ trend + season) , 8)
17     HOLT = holt(trainholt, 8)
18     SNAIVE = snaive(trainaive, 8)
19     DRIFT = rwf(trainholt, drift=TRUE, 8)
20     MC1 = (projecao + HOLT[["mean"]])*0,5
21     MC2 = (SNAIVE[["mean"]] + DRIFT[["mean"]] + projecao)*0,5
22
```

```
23     D = {accuracy(ARIMA, teste), accuracy(ETS, teste),
          accuracy(STL, teste), accuracy(NNAR, teste), accuracy(TBATS,
          teste), accuracy(TSLM, teste), accuracy(HOLT, teste),
          accuracy(SNAIVE, teste), accuracy(DRIFT, teste),
          accuracy(MC1, teste), accuracy(MC2, teste)}
24
25     mape = 100 // para pegar o menor erro percentual
26     for n = 1 to 11
27         if (D[n] < mape)
28             mape = D[n]
29
30     M[i] = {T[i], D[mape].method, mape}
31
32 return M // retorna a lista de melhores modelos por UF
```

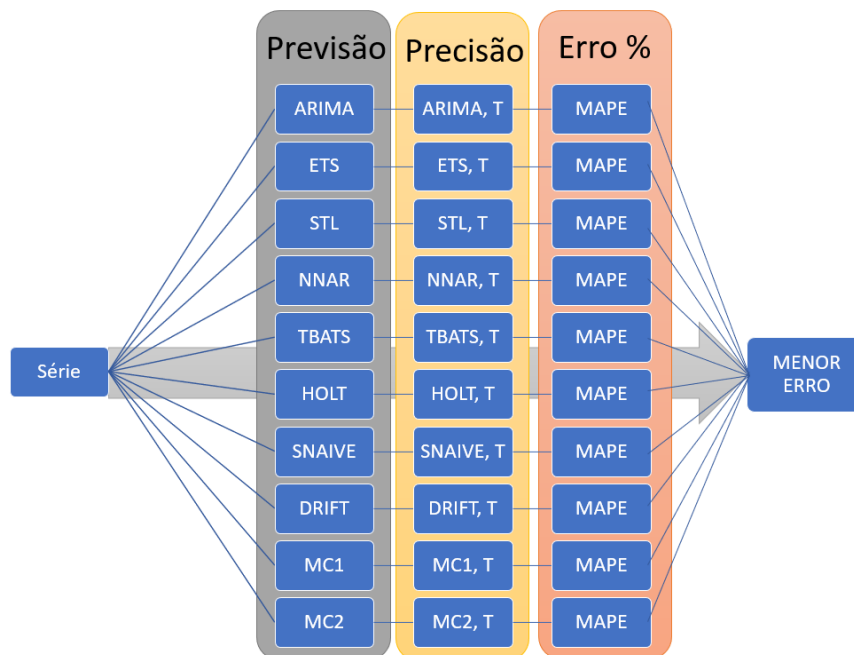
Fonte: Autoria própria, 2019

A Figura 3 ilustra o algoritmo do Pseudocódigo 5 mostrando o processo de seleção do melhor modelo de previsão das internações hospitalares por dengue para um período de 8 semanas ($h=8$):

1. Na previsão, a série temporal foi submetida aos métodos estatísticos;
2. Os parâmetros foram ajustados para encontrar o modelo de melhor ajuste;
3. Na precisão, utilizando os dados de teste da mesma série calculou-se o erro percentual absoluto médio (MAPE) para verificar a porcentagem média de erros;
4. Comparou-se o valor do MAPE entre os modelos candidatos para encontrar o menor.

O menor MAPE indicou o modelo mais preciso para a precisão de determinada série temporal dentre os 10 candidatos.

FIGURA 3 - FLUXOGRAMA ILUSTRATIVO DA PREVISÃO DE INTERNAÇÕES HOSPITALARES POR DENGUE CONSIDERANDO TODOS OS MODELOS ESTATÍSTICOS DE PREVISÃO UTILIZADOS



Fonte: Autoria própria, 2019.

Esperava-se que em 80% ou mais das séries temporais analisadas e previstas, os métodos estatísticos conseguissem realizar previsões de internações por dengue com Erro Percentual Médio Absoluto (MAPE) de, no máximo, 26%.

Comprovada a possibilidade de previsão de internações hospitalares por dengue a partir da série histórica dos dados disponíveis, precisou-se pesar na escolha do modelo de previsão com melhor precisão entre os 10 disponíveis para o cenário onde os dados de teste não estão disponíveis, ou seja, de fato prever o futuro. Os modelos redundantes foram desconsiderados na abordagem.

3.3 ASPECTOS ÉTICOS

Órgãos governamentais produzem e colocam dados abertos à disposição das pessoas, possibilitando sua leitura, acompanhamento, reutilização em novos projetos, sites e aplicativos. No Brasil, a Lei de Acesso à Informação (Lei nº 12.527 de 18 de novembro de 2011), que regulamenta o acesso a dados e informações mantidas pelo governo é ordenamento jurídico de democratização para o acesso à informação nacional (BRASIL, 2011; CASALI *et al.*, 2018).

A política dos dados abertos no Brasil foi instituída pela Instrução Normativa nº 4 de 12 de abril de 2012 e embasou a criação da Infraestrutura Nacional de Dados Abertos (INDA) que disponibiliza informações publicamente (DATASUS, 2012):

Segundo a norma, um dado público é todo aquele gerado ou sob a guarda do governo que não tenha o seu acesso restrito por alguma legislação, já o **dado aberto** é o dado público digitalizado e acessível na web, processável por máquina, exposto sob uma **licença aberta** que permita a livre utilização, reutilização, distribuição, consumo, cruzamento etc, exigindo no máximo o crédito da autoria e o compartilhamento pela mesma licença, e estruturado em **formato aberto**, que é um formato livre de patentes ou qualquer restrição de utilização e cuja especificação esteja documentada publicamente.

No Brasil, a Comissão Nacional de Ética em Pesquisa (CONEP) e os Comitês de Ética em Pesquisa (CEP) constituem o sistema CEP/CONEP que avalia, no âmbito da pesquisa científica em Ciências Humanas e Sociais (CHS), se aquilo que o pesquisador pretende fazer implica em risco para as pessoas participantes e qual o nível do risco.

Para a realização de pesquisa com seres humanos é necessário submeter o projeto ao CEP e esperar sua aprovação. No entanto, entre outros, nos casos em que a pesquisa utilize informações de acesso público, nos termos da Lei nº 12.527/2011 ou informações de domínio público, a mesma não será registrada nem avaliada pelo sistema CEP/CONEP (BRASIL, 2016; GUERRIERO, 2016).

Os dados das internações hospitalares utilizados nesta pesquisa são abertos, estão disponíveis publicamente e anonimizados o que dispensou a submissão da pesquisa ao Comitê de Ética em Pesquisa.

4 RESULTADOS E DISCUSSÕES

4 RESULTADOS E DISCUSSÕES

Os resultados e discussões desta pesquisa foram produzidos a partir de análise exploratória das séries temporais utilizadas, bem como experimentação de técnicas de previsão para decidir sobre o modelo estatístico com melhor precisão entre os candidatos a previsão de determinada série temporal. O algoritmo de previsão de internações hospitalares por dengue foi definido com base nos resultados encontrados.

4.1 ANÁLISE EXPLORATÓRIA DOS DADOS

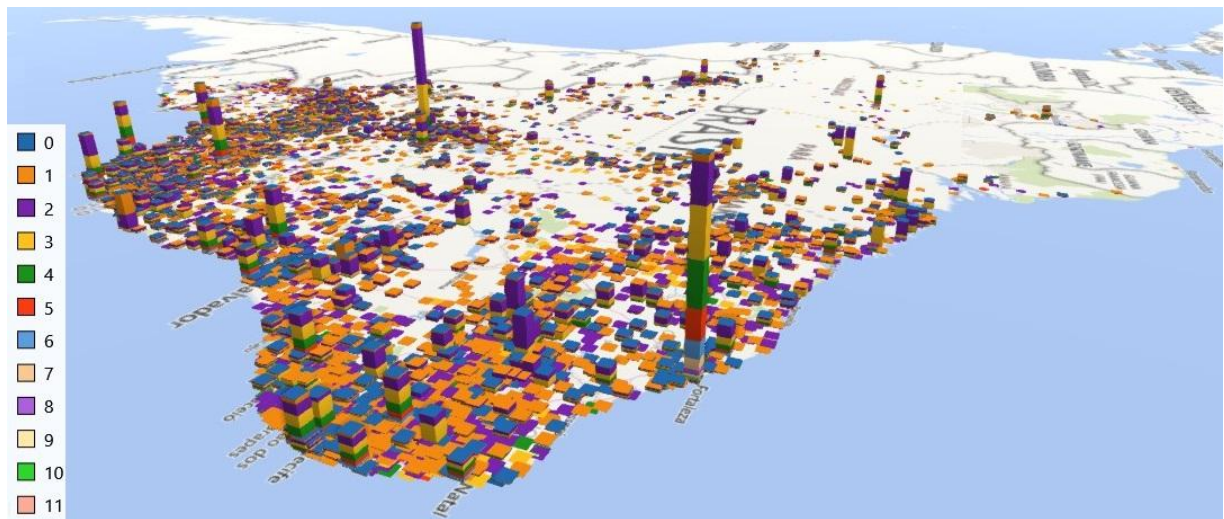
As séries temporais obtidas a partir dos estados brasileiros foram decompostas (APÊNDICE A) em seus componentes: tendência/ciclo, sazonalidade e restante. O restante diz respeito às influências contidas nos dados por elementos que não são sazonalidade, ciclo ou tendência. A tendência e o ciclo foram decompostos juntos.

A maioria das séries temporais analisadas apresentou tendência que variam entre decrescente, constante e crescente. Todas as séries temporais apresentaram variação constante na sua sazonalidade com concentração de máximos (picos) que ocorreram sempre no primeiro semestre e mínimos que aconteceram no segundo semestre de cada ano do período analisado. O teste de hipótese, KPSS, realizado para a verificação da estacionariedade nas séries temporais apresentou valor p maior que 0,05 para todos os casos, ou seja, todas são não estacionárias.

Alguns estados apresentaram quantidades máximas de internações hospitalares por dengue abaixo de 25 ocorrências por semana epidemiológica. Tais séries temporais: Amazonas (AM), Roraima (RR), Amapá (AP) e Rio Grande do Sul (RS) foram excluídas por não se adequarem aos métodos estatísticos utilizados durante os experimentos, pois pela quantidade de internações por semana epidemiológica, se enquadram na categoria de séries temporais discretas e/ou contáveis. Os dados do estado do Acre (AC), também foram desconsiderados, pois não estavam disponíveis para os meses de maio e junho de 2019.

A Figura 4 mostra que as internações por dengue no período de 2009 a 2018 estão concentradas mais na região próxima à costa marítima do Brasil. Isso se deve a presença de uma maior densidade populacional nessa área em comparação com a Região Norte do País onde estão localizados os estados: AM, RR, AP e AC, fator que pode explicar a baixa quantidade de internações por dengue nesses estados.

FIGURA 4 - INTERNAÇÕES POR DENGUE DISTRIBUÍDAS POR CIDADE DE RESIDÊNCIA DO PACIENTE E CATEGORIZADA POR QUANTIDADE DE DIAS DE PERMANÊNCIA NO HOSPITAL



Fonte: Silva *et al.* (2019).

Antes da realização dos experimentos, os dados foram divididos em conjunto de treinamento (1ª semana de 2016 até a 10ª semana de 2019) e conjunto de teste (11ª semana de 2019 até a 18ª semana de 2019). Como critério de divisão considerou-se toda a série temporal para treinamento dos modelos, exceto as 8 últimas semanas da série em análise, as quais foram utilizadas para a realização do teste.

Como o modelo de suavização exponencial (*ETS*) não lida com mais de 24 observações (semanas epidemiológicas), então o conjunto de dados de treinamento dele foi definido da 38ª semana de 2018 à 10ª semana de 2019. Por se tratar da mesma série temporal, bem como foi utilizada o último período da mesma, então não se tem prejuízo com essa abordagem.

Os experimentos sobre a capacidade de previsão dos modelos estatísticos (*ARIMA*, *ETS*, *TBATS*, *STL*, *NNAR* e *REGRESSÃO LINEAR*) foram realizados em 4 cenários:

- Cenário 1 – série temporal original;
- Cenário 2 – série temporal original com aplicação da transformação de *Box-Cox*;
- Cenário 3 – série temporal sem *outliers*;
- Cenário 4 – série temporal sem *outliers* e aplicação da transformação de *Box-Cox*.

O Quadro 6 apresenta o resultado dos quatro cenários supramencionados. A tonalidade de verde indica aproximação do valor *MAPE* a zero (condição perfeita). A tonalidade de

vermelho indica o afastamento do valor MAPE de zero. Quanto maior o valor MAPE, maior é o erro de previsão e pior é a precisão do modelo.

Nas séries temporais analisadas, os modelos estatísticos utilizados com o ajuste automático de parâmetros conseguiram realizar a previsão com horizonte de 8 semanas em cerca de 63,63% dos casos (menor MAPE entre os modelos candidatos com valor inferior a 32%). Em 37,47% dos casos, os modelos não conseguiram previsão com MAPE de no máximo de 32%.

Da análise do cenário 1 (dados originais) em comparação com o cenário 2 (dados originais transformados). Em ambos os casos não foi realizado o tratamento para a eliminação dos *outliers*. Observa-se que a aplicação de transformação melhorou a precisão da previsão em cerca de 18,18%, mas o cenário 3 que apresenta eliminação de *outliers* conseguiu superar a melhoria da aplicação das transformações.

O resultado do cenário 4, que considera a eliminação dos *outliers* e a aplicação de transformação nos dados antes da previsão, em comparação com o cenário 3 (os dados sem *outliers* e sem a aplicação de transformação) mostra que houve melhoria na precisão, apenas, em um caso (MG) dos 22 casos analisados. Sugerindo que não é necessário aplicar transformação nos dados para a realização da previsão de séries temporais de internações por dengue.

Quando se compara o cenário 1 (dados originais) com o cenário 3 (dados sem *outliers*), tem-se uma melhora na precisão da previsão das internações por dengue em cerca de 36,36% dos casos analisados. Logo, sugere-se utilizar a série original e em seguida a série com a eliminação dos *outliers*.

Com relação aos modelos escolhidos pelo ajuste automático dos parâmetros, observa-se que o mesmo método estatístico gerou modelos diferentes de acordo com cada série temporal e cenário do experimento. Por conta dos recursos computacionais, a classe ARIMA (Quadro 6) foi utilizada com limitante de busca no ajuste automático de seus parâmetros, bem como não foi garantida a estacionariedade da série temporal antes de submetê-la ao método.

A regressão linear teve sua precisão superada na única vez que conseguiu sucesso (série temporal do estado de Goiás - GO) e por esse motivo não foi considerada na continuidade dos experimentos.

QUADRO 6 - MELHOR PRECISÃO NA PREVISÃO DE INTERNAÇÕES HOSPITALARES POR DENGUE CONSIDERANDO (CENÁRIO, SÉRIE TEMPORAL E MÉTODOS DE PREVISÃO)

UF	CENÁRIO 1		CENÁRIO 2		CENÁRIO 3		CENÁRIO 4	
	MODELO	MAPE	MODELO	MAPE	MODELO	MAPE	MODELO	MAPE
MG	ETS(M,A,N)	35,19	ETS(A,A,N)	36,69	ETS(M,A,N)	47,25	ARIMA(2,1,0)(0,1,0)[52]	39,72
SP	TBATS(0,347, {0,0}, 0,859, {<52,2>})	39,14	TBATS(0,347, {0,0}, 0,859, {<52,2>})	39,14	ARIMA(5,2,0)(1,0,1)[52]	30,82	BATS(0,556, {0,0}, 0,91, -)	31,36
BA	ETS(A,N,N)	26,69	NNAR(2,1,2)[52]	24,69	NNAR(4,1,3)[52]	13,22	NNAR(2,1,2)[52]	25,37
MA	TBATS(0,248, {0,0}, -, {<52,1>})	43,28	ARIMA(4,0,0)(0,1,0)[52]	38,65	STL + ETS(A,Ad,N)	22,34	ARIMA(4,0,0)(0,1,0)[52]	38,05
GO	Regressão Linear	22,75	Regressão Linear	19,78	ARIMA(1,1,1)(1,1,0)[52]	13,88	ARIMA(1,1,1)(1,1,0)[52]	13,83
PE	ETS(M,N,N)	54,68	ARIMA(2,1,1)(1,0,0)[52]	52,55	ARIMA(0,2,2)(1,0,0)[52]	47,87	ARIMA(2,1,1)(1,0,0)[52]	52,57
PR	ARIMA(2,1,2)(1,0,0)[52]	45,18	ETS(A,N,N)	49,78	ETS(A,N,N)	49,78	ETS(A,N,N)	49,78
RN	ETS(A,N,N)	49,30	ETS(A,N,N)	49,30	ETS(A,N,N)	49,3	ETS(A,N,N)	49,30
CE	STL + ETS(A,Ad,N)	17,96	ARIMA(0,1,1)(0,1,0)[52]	35,04	STL + ETS(A,N,N)	27,65	ARIMA(0,1,1)(0,1,0)[52]	35,01
MS	ARIMA(3,2,1)(1,0,0)[52]	25,89	BATS(1, {5,2}, -, -)	28,26	ARIMA(1,2,1)	74,73	TBATS(1, {0,0}, 0,966, {<52,7>})	77,03
PA	TBATS(0,041, {0,0}, -, {<52,6>})	18,90	TBATS(0,041, {0,0}, -, {<52,6>})	18,90	TBATS(0,436, {0,0}, -, {<52,1>})	19,96	TBATS(0,436, {0,0}, -, {<52,1>})	19,96
ES	STL + ETS(A,N,N)	26,31	NNAR(4,1,3)[52]	27,5	STL + ETS(A,N,N)	27,51	NNAR(4,1,3)[52]	27,6
RJ	STL + ETS(A,Ad,N)	35,82	BATS(1, {0,0}, -, -)	60,91	TBATS(1, {0,0}, 0,895, {<52,5>})	43,88	TBATS(1, {0,0}, 0,895, {<52,5>})	43,88
PB	STL + ETS(A,N,N)	15,48	BATS(0,359, {0,0}, 0,867, -)	43,00	Regressão Linear	23,61	BATS(0,407, {0,0}, 0,868, -)	43,85
RO	ARIMA(1,1,3)(1,0,1)[52]	28,42	ETS(A,N,N)	33,99	BATS(1, {0,0}, -, -)	25,19	BATS(1, {0,0}, -, -)	25,19
PI	STL + ETS(A,N,N)	33,88	TBATS(1, {0,0}, -, {<52,6>})	34,91	TBATS(1, {0,0}, -, {<52,5>})	31,56	TBATS(1, {0,0}, -, {<52,5>})	31,56
MT	ETS(A,N,N)	17,79	ETS(A,N,N)	17,79	ETS(A,N,N)	17,79	ETS(A,N,N)	17,79
AL	ETS(M,N,N)	49,80	ETS(A,N,N)	49,77	ETS(M,N,N)	49,8	ETS(A,N,N)	49,77
DF	ETS(A,A,N)	21,31	ETS(A,A,N)	21,29	ETS(A,A,N)	21,76	ETS(A,A,N)	21,79
TO	TBATS(1, {0,0}, -, {<52,1>})	20,68	TBATS(1, {0,0}, -, {<52,1>})	20,68	TBATS(1, {0,0}, -, {<52,7>})	21,78	TBATS(1, {0,0}, -, {<52,7>})	21,78

UF	CENÁRIO 1		CENÁRIO 2		CENÁRIO 3		CENÁRIO 4	
	MODELO	MAPE	MODELO	MAPE	MODELO	MAPE	MODELO	MAPE
SC	BATS(1, {1,3}, -, -)	67,28	BATS(1, {1,3}, -, -)	67,28	ETS(A,N,N)	69,63	ETS(A,N,N)	69,63
SE	ETS(A,A,N)	44,72	ETS(A,A,N)	44,16	NNAR(9,1,6) [52]	38,05	BATS(1, {0,0}, 1, -)	40,53

Fonte: Autoria própria, 2019.

A técnica ARIMA está entre as preferidas dos pesquisadores. Ela é utilizada em diversos trabalhos com previsão de séries temporais (VILLANI *et al.*, 2017, CHAKRABORTY; CHATTOPADHYAY; GHOSH, 2019, RUCHIRASET; TANTRAKARNAPA, 2018). Assim, com base no mesmo período utilizado nos cenários que originaram os resultados do Quadro 6 e considerando toda a capacidade do método ARIMA, ou seja, sem limitações, realizaram-se experimentos de previsão de séries temporais em mais 4 cenários:

- Cenário 5 – série temporal original e os parâmetros (*approximation=FALSE, stepwise=FALSE*) e sem garantia de estacionariedade da série (sem diferenciação explícita). Neste caso, o modelo se ocupou de escolher a ordem de diferenciação, se necessário;
- Cenário 6 – série temporal sem *outliers* e os parâmetros (*approximation=FALSE, stepwise=FALSE*) e sem garantia de estacionariedade da série (sem diferenciação explícita);
- Cenário 7 – série temporal original e os parâmetros (*approximation=FALSE, stepwise=FALSE, stationary=TRUE*). Neste caso, tem-se a garantia de estacionariedade da série (com diferenciação explícita);
- Cenário 8 – série temporal sem *outliers* e os parâmetros (*approximation=FALSE, stepwise=FALSE*). Neste caso, tem-se a garantia de estacionariedade da série (com diferenciação explícita).

No Quadro 7, ao comparar os cenários 7 e 8 que apresentam os resultados da previsão com ARIMA utilizando a diferenciação da série original (cenário 7) ou a diferenciação da série sem *outliers* (cenário 8), evidencia-se que garantir a estacionariedade da série antes de submetê-la ao método ARIMA não melhora o desempenho da previsão, pelo contrário, o piora na maioria dos casos.

O cenário 6, que apresenta a eliminação de *outliers* na série sem diferenciação explícita, melhorou a precisão e o desempenho dos modelos em 45,45% dos casos analisados

em comparação com o cenário 5, que trata da série temporal original, também sem diferenciação explícita.

Comparando-se os resultados do Quadro 7 com os apresentados no Quadro 6, observa-se que nos estados de Goiás (GO) e Paraná (PR), em ambos os casos, os resultados foram semelhantes. O método ARIMA somente conseguiu melhorar o desempenho na previsão da série temporal do estado do Alagoas (AL) (aproximadamente 5 pontos percentuais de diferença). Nos demais casos, foi superado pelos concorrentes.

Portanto, pode-se utilizar a função *auto.arima()* passando como parâmetro apenas a série temporal, deixando o ajuste de parâmetros a cargo da função, que consegue se adequar aos dados analisados. Isso é importante, pois diminui-se o tempo para treinamento do modelo, uma vez que a busca pelo modelo mais preciso não será estendida. A transformação aplicada à série temporal pode ser dispensada, uma vez que os parâmetros dos modelos se ajustam automaticamente aos dados analisados.

QUADRO 7 - MELHOR PRECISÃO NA PREVISÃO DE INTERNAÇÕES HOSPITALARES POR DENGUE CONSIDERANDO (CENÁRIO, SÉRIE TEMPORAL E MÉTODO DE PREVISÃO ARIMA)

SÉRIE	CENÁRIO 5		CENÁRIO 6		CENÁRIO 7		CENÁRIO 8	
	MODELO	MAPE	MODELO	MAPE	MODELO	MAPE	MODELO	MAPE
MG	ARIMA(2,1,3)	52,12	ARIMA(0,1,4)	55,74	ARIMA(2,0,3)	100,94	ARIMA(0,0,4)	99,55
SP	ARIMA(0,2,4) (0,0,1)[52]	51,94	ARIMA(0,2,3)	36,74	ARIMA(0,0,4)	101,93	ARIMA(0,0,3)	99,8
BA	ARIMA(3,1,2)	33,41	ARIMA(1,1,3) (0,0,1)[52]	25,54	ARIMA(3,0,2)	101,82	ARIMA(1,0,3) (0,0,1)[52]	98,16
MA	ARIMA(0,1,4) (1,0,0)[52]	53,08	ARIMA(1,1,2)	49,76	ARIMA(0,0,4) (1,0,0)[52]	100,59	ARIMA(1,0,2)	99,52
GO	ARIMA(0,1,0) (1,1,0)[52]	25,01	ARIMA(0,1,1) (1,1,0)[52]	13,93	ARIMA(2,0,1) (1,0,0)[52]	46,53	ARIMA(0,1,1) (1,1,0)[52]	13,93
PE	ARIMA(0,2,4)	64,37	ARIMA(0,2,2)	48,21	ARIMA(0,0,3)	101,34	ARIMA(0,0,2)	96,39
PR	ARIMA(4,1,0)	42,44	ARIMA(1,1,2)	56,19	ARIMA(4,0,0)	99,31	ARIMA(1,0,2)	101,98
RN	ARIMA(4,1,0)	82,77	ARIMA(2,1,2)	49,78	ARIMA(4,0,0)	106,19	ARIMA(2,0,2)	91,35
CE	ARIMA(2,1,2) (0,1,1)[52]	27,58	ARIMA(1,1,1) (0,1,1)[52]	31,02	ARIMA(0,0,5)	100,12	ARIMA(4,0,0)	100,8
MS	ARIMA(2,2,3)	27,85	ARIMA(0,2,2)	74,74	ARIMA(2,0,2)	99,85	ARIMA(0,0,2)	99,61
PA	ARIMA(3,1,2)	31,06	ARIMA(1,1,2)	30,63	ARIMA(3,0,2)	101,34	ARIMA(1,0,2)	100,91
ES	ARIMA(2,1,2)	27,2	ARIMA(1,1,3)	28,22	ARIMA(2,0,2)	100,86	ARIMA(1,0,3)	100,7

SÉRIE	CENÁRIO 5		CENÁRIO 6		CENÁRIO 7		CENÁRIO 8	
	MODELO	MAPE	MODELO	MAPE	MODELO	MAPE	MODELO	MAPE
RJ	ARIMA(0,1,5)	64,86	ARIMA(5,2,0)	65,69	ARIMA(0,0,5)	100,64	ARIMA(5,1,0)	100,74
PB	ARIMA(0,1,5)	49,16	ARIMA(1,1,4)	49,48	ARIMA(0,0,5)	99,54	ARIMA(1,0,4)	99,69
RO	ARIMA(3,1,2)	69,80	ARIMA(0,1,5)	32,56	ARIMA(3,0,2)	114,85	ARIMA(0,0,5)	101,76
PI	ARIMA(2,0,1) (1,1,0)[52]	53,32	ARIMA(2,0,0) (1,1,0)[52]	70,03	ARIMA(2,0,1)	64,45	ARIMA(1,0,0)	98,85
MT	ARIMA(1,1,0)	17,69	ARIMA(0,1,1) (1,1,0)[52]	28,01	ARIMA(1,0,0)	100	ARIMA(1,0,4)	98,16
AL	ARIMA(0,1,3)	44,73	ARIMA(0,1,3)	44,73	ARIMA(0,0,3)	94,66	ARIMA(4,0,1)	99,32
DF	ARIMA(2,1,2) (1,0,0)[52]	20,16	ARIMA(2,1,2) (1,0,0)[52]	20,97	ARIMA(2,0,2) (1,0,0)[52]	98,38	ARIMA(2,0,2) (1,0,0)[52]	99,23
TO	ARIMA(0,1,1)	25,72	ARIMA(0,1,1) (1,1,0)[52]	26,24	ARIMA(0,0,1)	97,39	ARIMA(0,0,1) (1,0,0)[52]	96,5
SC	ARIMA(0,1,3)	75,13	ARIMA(1,1,2)	73,87	ARIMA(0,0,3)	97,98	ARIMA(1,0,2)	98,8
SE	ARIMA(1,1,2)	56,90	ARIMA(2,1,3)	44,53	ARIMA(1,0,2)	104,27	ARIMA(1,0,3)	101,7

Fonte: Autoria própria, 2019.

Na previsão de séries temporais, existem métodos estatísticos mais simples do que os supramencionados, mas que podem surpreender na precisão a depender dos dados analisados, assim como a combinação entre modelos (HYNDMAN; ATHANASOPOULOS, 2018). Pensando nisso, mais 2 cenários foram criados:

- Cenário 9 – utilização da série temporal original, ou seja, como foi obtida do DATASUS.
- Cenário 10 – série temporal com eliminação de *outliers*, isto é, foram substituídos os valores discrepantes em relação a maioria dos dados presentes na série histórica.

Os métodos estatísticos adicionais utilizados foram: HOLT, SNAIVE e DRIFT. Mais uma vez, os dados foram divididos em base de treinamento (da 1ª semana de 2016 a 10ª semana de 2019) e base de testes (da 11ª a 18ª semana epidemiológica do ano de 2019). Para o modelo HOLT, a base de treinamento ficou definida da 1ª semana a 10ª semana de 2019. No modelo inocente sazonal (SNAIVE), a base de treinamento tinha dados da semana epidemiológica 1 de 2018 a semana 10 de 2019. O modelo ETS foi executado sobre o período da semana 38 de 2018 até a semana 10 de 2019.

Para a criação dos modelos combinados (MC1 e MC2) utilizou-se uma projeção de 8 semanas (3ª a 10ª semana epidemiológica de 2019), a qual foi utilizada para gerar a previsão de 8 semanas futuras em combinação com outros modelos:

- MC1 – modelo combinado 1 (projeção + HOLT) * 0,5
- MC2 – modelo combinado 2 (SNAIVE + DRIFT + projeção) * 0,5

No Quadro 8, observa-se que 22,72% das previsões realizadas tiveram o valor MAPE acima de 26%. Em 77,28% dos casos analisados no experimento, os métodos utilizados para a previsão de internações hospitalares por dengue conseguiram precisão com MAPE abaixo de 26%. Confirma-se também que a eliminação de *outliers* é importante, em alguns casos (18,18%), ou seja, aproximadamente 2 a cada 10 previsões, para melhorar a precisão da previsão, mas sem diferença significativa, pois apenas no estado de SP, a diferença foi de 11 pontos percentuais.

QUADRO 8 - MELHOR PRECISÃO NA PREVISÃO DE INTERNAÇÕES HOSPITALARES POR DENGUE CONSIDERANDO (CENÁRIO, SÉRIE TEMPORAL E MÉTODOS DE PREVISÃO SIMPLES E COMPLEXOS)

SÉRIE	CENÁRIO 9		CENÁRIO 10	
	MODELO	MAPE	MODELO	MAPE
MG	HOLT	30,35	HOLT	34,90
SP	DRIFT	29,67	HOLT	17,43
BA	MC1	15,65	MC1	15,65
MA	HOLT	24,36	STL	22,34
GO	HOLT	13,69	MC1	13,44
PE	SNAIVE	21,57	MC2	23,45
PR	HOLT	25,54	HOLT	25,54
RN	HOLT	47,51	HOLT	47,51
CE	STL	17,96	STL	27,65
MS	DRIFT	21,19	HOLT	71,59
PA	MC1	15,66	MC1	15,66
ES	NNAR	23,83	STL	27,50
RJ	MC2	23,93	MC2	23,93
PB	STL	15,48	MC2	16,88
RO	ARIMA	28,42	TBATS	25,19
PI	STL	33,88	TBATS	31,56
MT	ETS	17,79	ETS	17,79
AL	MC2	36,29	MC2	36,29
DF	DRIFT	15,36	DRIFT	15,36
TO	TBATS	20,68	TBATS	21,78
SC	SNAIVE	64,35	SNAIVE	64,34

SÉRIE	CENÁRIO 9		CENÁRIO 10	
	MODELO	MAPE	MODELO	MAPE
SE	HOLT	39,56	HOLT	39,56

Fonte: Autoria própria, 2019.

Os experimentos realizados nos 10 cenários supramencionados mostram que é possível realizar a previsão de internações hospitalares por dengue para horizonte de 8 semanas utilizando-se diversas técnicas de previsão de séries temporais concomitantemente sobre o mesmo conjunto de dados. A utilização de métodos simples e combinados de previsão de séries temporais conseguiram melhorar a precisão.

Mas, entre os modelos candidatos qual tem melhor desempenho e precisão na previsão de determinada série temporal? Essa não é uma resposta fácil, pois é preciso esperar o fato acontecer (obtenção das internações reais) para realizar o cálculo do valor MAPE, pois estes dados só estarão disponíveis no futuro.

A escolha do modelo mais preciso entre os concorrentes precisa ser feita com base na série temporal analisada. Neste caso, o modelo que melhor se ajusta aos dados, nem sempre será o modelo mais preciso na previsão, pois tem-se o risco de *overfitting*, isto é, o modelo se ajusta demais aos padrões existentes nos dados gerando um valor de erro MAPE bem próximo de zero, mas quando se realiza previsão em dados reais, o valor do erro MAPE tende a crescer demais e a precisão do modelo acaba sendo não tão boa quanto seus concorrentes.

A decisão sobre o modelo mais preciso considerando apenas os dados originais das internações hospitalares fica sobre 10 possibilidades. Quando se inclui a previsão sobre os dados originais e os dados com eliminação de *outliers*, o espaço de escolha aumenta para 20 possibilidades.

Considerando que em apenas 2 casos a cada 10, a previsão de séries temporais sem *outliers* teve uma pequena melhoria e também a necessidade de reduzir os espaço de possibilidades de previsões de séries temporais de internações por dengue, então é suficiente para os objetivos desta pesquisa realizar a previsão de internações por dengue utilizando somente os dados originais.

4.2 ESCOLHA DO MODELO COM MELHOR PRECISÃO

Cada série temporal foi submetida aos 10 modelos de previsão (NNAR, STL, ARIMA, TBATS, ETS, SNAIVE, HOLT, DRIFT, MC1 e MC2) e os resultados avaliados com base no MAPE do conjunto de dados da previsão em comparação com o conjunto de dados de teste

foram registrados. O modelo que apresentou o menor MAPE foi o escolhido para ser utilizado na previsão de determinada série histórica. Para a pesquisa foi possível dividir a base de dados em treinamento e teste e mensurar o MAPE verificando o desempenho de cada modelo.

Em um caso real, os dados de teste não estão disponíveis, pois o conjunto de dados inteiro até a atualidade é utilizado para a previsão das 8 semanas epidemiológicas seguintes. É indicado utilizar o MAPE com base no conjunto de dados de treinamento para verificar o ajuste do modelo, mas os experimentos mostraram que o melhor ajuste do modelo aos dados não é garantia de que ele seja o melhor em precisão da previsão perante seus concorrentes para determinada série temporal.

As séries temporais analisadas apresentam tendência bem definida, bem como sazonalidade (APÊNDICE A). Os métodos estatísticos (ETS, DRIFT e TBATS) foram desconsiderados pela redundância perante a precisão dos outros, ou seja, o resultado dos demais métodos de previsão eram muito próximos aos deles.

As previsões que se iniciam mais próximas da série temporal original tem mais probabilidade de melhorar sua precisão perante as demais. Este fato ficou confirmado observando-se que 86,36% das previsões com melhor MAPE representam os quatro métodos com previsão iniciada mais próxima da série temporal original e 77,28% das previsões mais precisas foram dos 3 métodos com previsões mais próximas da série original. Os modelos que apresentaram previsão que não acompanharam a tendência e a sazonalidade da série original tiveram as piores precisões em relação a seus concorrentes.

Entre os 3 modelos com previsão iniciada o mais próximo possível da série temporal original, em 83,33% dos casos analisados, o que apresentou melhor precisão foi aquele que acompanhou a tendência e variação do período anterior à previsão das internações por dengue.

Portanto, para a escolha do modelo mais preciso entre os candidatos à previsão das internações hospitalares por dengue, aplicar-se-á a série temporal aos métodos estatísticos: NNAR, STL, ARIMA, SNAIVE, HOLT, MC1 e MC2. Como a técnica DRIFT foi desconsiderada, então definiu-se o modelo MC2 como sendo: (SNAIVE + PROJECAO + HOLT) * 0,5.

Após a realização da previsão, deve-se escolher os três modelos que apresentarem maior proximidade do início da previsão com a série temporal original. A previsão vai mostrar três casos (aumento nas internações por dengue, estabilidade no número de internações ou diminuição das internações por dengue). Para decidir sobre qual caso tem mais

probabilidade de ocorrer observa-se, entre as três previsões, àquela que acompanha a tendência e a variação do período anterior à previsão, pois ela tem maior probabilidade de acontecer.

4.3 ALGORITMO PARA PREVISÃO DE INTERNAÇÕES HOSPITALARES POR DENGUE

Os dados de internações por dengue dos estados brasileiros que se deseja realizar a previsão, precisam ser obtidos no site do DATASUS. É necessário considerar 167 semanas epidemiológicas a contar da semana atual para o passado, ou seja, aproximadamente 3 anos e 5 meses de dados. Ao extrair a série temporal dos dados obtidos, submete-se a mesma ao algoritmo de previsão do Pseudocódigo 6 e tem-se o resultado. Pelo cronograma de trabalho da pesquisa e domínio dos pesquisadores sobre a linguagem de programação R, o algoritmo foi definido com base no pacote *forecast*.

PSEUDOCÓDIGO 6 - ALGORITMO DE PREVISÃO DE INTERNAÇÕES HOSPITALARES POR DENGUE

```
PREVISAO-INTERNACOES-DENGUE(s) // recebe a série temporal
1   Sat = ? // recebe semana epidemiológica atual
2
3   treino = window(s, start=c(sat - 166))
4   trainholt = window(s, start=c(sat - 9))
5   trainaive = window(s, start=c(sat - 62))
6   projecao = window(s, start=c(sat - 7))
7
8   G = {window(s, start=c(sat - 51)),
9   ARIMA = forecast(auto.arima(treino, approximation=F,
10  stepwise=F), 8),
11  STL = forecast(stlf(treino), 8),
12  NNAR = forecast(nnetar(treino), 8),
13  HOLT = holt(trainholt, 8),
14  SNAIVE = snaive(trainaive,8),
15  MC1 = (projecao + HOLT[["mean"]])*0,5,
16  MC2 = (SNAIVE[["mean"]] + HOLT[["mean"]] + projecao)*0.5}
17
18  plot G
19  T = (os 3 métodos com previsão mais próxima de s em plot G)
20  X = (O modelo em T que acompanha a tendência e variação de s)
21 return X
```

Fonte: Autoria própria, 2019.

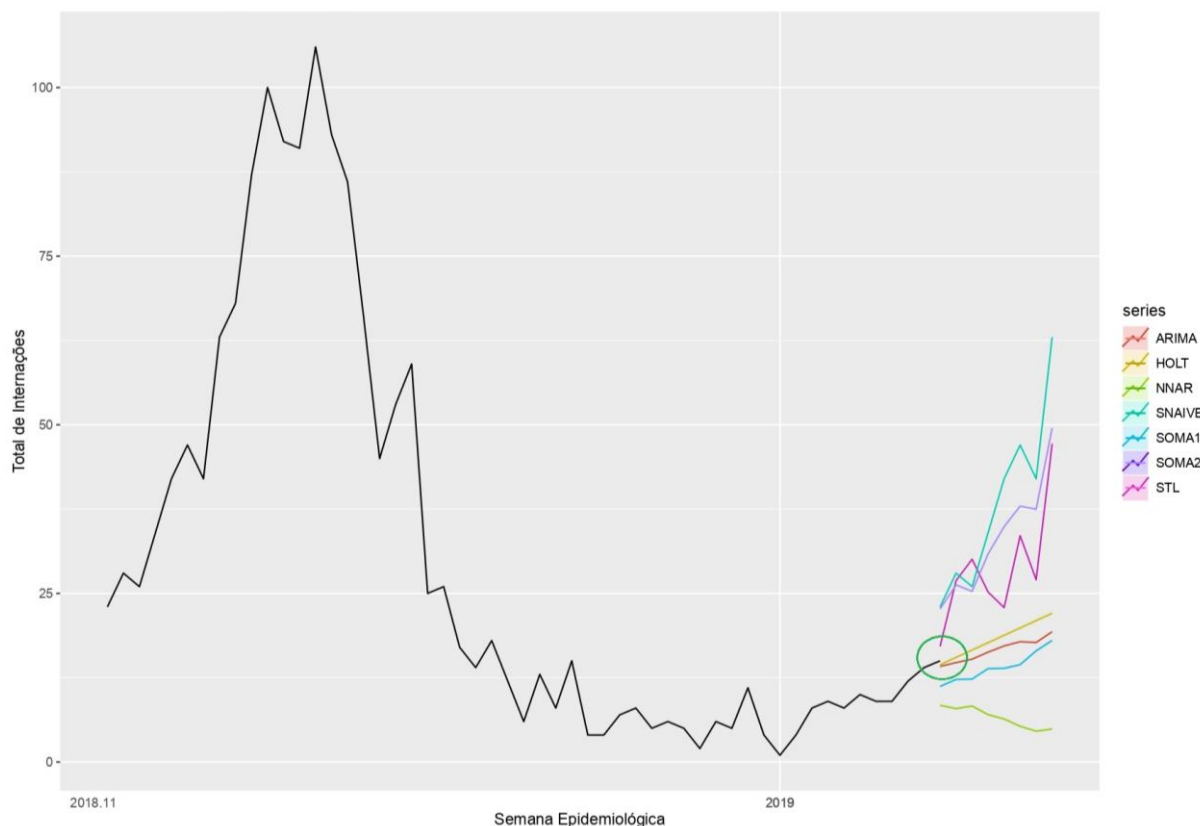
Ao receber a série temporal, o algoritmo precisa que seja definida a semana epidemiológica atual (*sat*) que será utilizada como última semana em cada um dos conjuntos de dados de treinamento (*treino*, *trainholt*, *trainaive*, e *projecao*), bem como precisa ser definida a semana inicial do período de previsão ($start=c(sat - x)$). A variável $x + 1$ indica a quantidade de semanas que será considerada em cada conjunto de dados (linhas 3 a 6 do algoritmo) utilizados no treinamento dos modelos estatísticos que irão prever as internações hospitalares por dengue.

Empiricamente, foi escolhido o valor de ajuste (0,5) para os modelos MC1 e MC2, os quais se mostraram com melhor precisão enquanto modelos combinados a partir das técnicas HOLT e SNAIVE, as quais capturam a tendência e a sazonalidade do último período da série a ser prevista.

A Figura 5 ilustra a utilização do algoritmo de previsão das internações por dengue. Nela é possível observar que a série temporal foi submetida aos sete métodos de previsão de séries temporais estabelecidos. O círculo verde indica quais os três modelos tiveram o início de sua previsão mais próxima da série original (HOLT, ARIMA e STL). A plotagem da série temporal considerou o período da semana 11 de 2018 até a semana 11 de 2019, mas a previsão foi realizada com base no período citado nos experimentos. Nela, pode-se observar que as 11 semanas, último período da série, apresentam tendência de crescimento junto de uma variação. Entre os três modelos, o que mais se aproxima da série original no último período é o STL, logo, ele tem mais probabilidade de acontecer na realidade.

A previsão com STL indica aumento considerável dos casos de dengue. Já com ARIMA e HOLT, o aumento é mais leve. No caso da PB, não se teve uma previsão que indique a diminuição dos casos de dengue entre os três modelos selecionados. Confirmando os dados da Figura 5, o STL foi o modelo que apresentou menor MAPE entre os candidatos (15,48%).

FIGURA 5 - EXEMPLIFICAÇÃO DA PREVISÃO DE INTERNAÇÕES HOSPITALARES POR DENGUE PARA A SÉRIE TEMPORAL PB



Fonte: Aatoria própria, 2019.

Logo, a combinação da previsão mais próxima da série original e depois a observação de qual modelo, entre os três restantes, acompanham a tendência e variação da série original, pode ser utilizado para a escolha do modelo de previsão de internações hospitalares mais preciso entre os candidatos.

Nos testes, obtiveram-se cerca de 2 erros para cada 10 previsões realizadas, tanto na escolha dos três modelos mais próximos da série histórica original perante seus concorrentes, como para a observação sobre a tendência e sazonalidade quanto à escolha do modelo mais preciso entre os três restantes.

5 CONSIDERAÇÕES FINAIS

5 CONSIDERAÇÕES FINAIS

A missão do pesquisador na previsão de series temporais é minimizar o erro de forma que o mesmo se aproxime de zero (ideal). Nos trabalhos relacionados a esta pesquisa, os autores realizaram previsões de séries temporais indicando que o método utilizado para a previsão em determinada época é o mesmo que se adequa às demais épocas a serem previstas. Isso não aconteceu na previsão das internações por dengue, pois considerando a mesma série temporal, na Bahia por exemplo, à medida que se variava a época de previsão, variava-se também o método com melhor precisão.

A principal contribuição desta pesquisa é utilizar uma abordagem que faz uso de diversas técnicas de previsão de séries temporais ao mesmo tempo, para os mesmos dados e na mesma época com o objetivo de minimizar o erro produzido. O resultado foi que em 77,28% dos casos analisados nos experimentos, os métodos estatísticos utilizados para a previsão de internações hospitalares por dengue conseguiram precisão com erro abaixo de 26%.

Para o desenvolvimento da pesquisa foi fundamental a decomposição das séries temporais analisadas, mostrando que todas têm variação constante na sua sazonalidade com pontos de máximo concentrados sempre no primeiro semestre e pontos de mínimos acontecendo no segundo semestre de cada ano. Com base no teste de hipótese KPSS, elas são não estacionárias.

O modelo estatístico de previsão por regressão linear teve sua precisão superada pelos concorrentes e por isso não foi considerado na continuidade dos experimentos. É suficiente utilizar os dados originais para conseguir os resultados desejados. O tratamento dos *outliers*, seja, sua eliminação, não fez muita diferença na maioria das previsões realizadas, mas quando utilizadas as séries temporais sem *outliers*, dobra-se a quantidade de métodos candidatos à previsão de determinada série temporal. A transformação aplicada aos dados antes da realização da previsão, também, pode ser desconsiderada.

A realização de diferenciação explícita para garantir a estacionariedade da série antes da previsão com o método ARIMA não melhorou a precisão. A função *auto.arima(treino)* consegue se adequar aos dados analisados e realizar os devidos ajustes nos parâmetros necessários para este método, que sendo considerado um dos mais utilizados para a realização da previsão de séries temporais, não conseguiu se sair tão bem quanto o esperado na previsão

de internações por dengue. Uma causa provável é que a predominância da quantidade de dados por semana epidemiológica tenha interferido no desempenho desta técnica estatística.

A pesquisa confirma que o tratamento de *outliers* é importante, em alguns casos (18,18%), para melhorar a precisão da previsão, mas sem diferença significativa. Como o tratamento de *outliers* aumenta a quantidade de modelos no conjunto das possíveis escolhas do mais preciso, pode-se desconsiderar esta abordagem na previsão de séries temporais de internações por dengue.

Os experimentos realizados nos 10 cenários apresentados nesta pesquisa trouxeram evidências de que é possível realizar a previsão de internações hospitalares por dengue para horizonte de 8 semanas utilizando-se diversas técnicas de previsão de séries temporais sobre o mesmo conjunto de dados e ao mesmo tempo seguindo algoritmo projetado para tal fim. A utilização de métodos simples e combinados de previsão de séries temporais conseguem melhorar a precisão da previsão diante dos métodos mais complexos.

Os experimentos mostraram que o melhor ajuste do modelo candidato aos dados da série temporal não é garantia de que ele seja o melhor em precisão diante dos demais concorrentes, mesmo com a indicação do menor valor MAPE com base no conjunto de dados de treinamento que indica o quanto determinado modelo de previsão se ajusta aos dados analisados para previsões futuras.

As previsões que se iniciam mais próximas da série temporal original tem mais probabilidade de melhorar sua precisão perante as demais. Este fato ficou confirmado em 86,36% das previsões com melhor MAPE representa os quatro métodos com previsão iniciada mais próxima possível da série temporal original.

Percebeu-se também que entre os três modelos com início de previsão mais próximo da série temporal original, aquele que seguiu a tendência e a variação da série no período anterior à previsão teve maior probabilidade de ser o mais preciso. Isso foi confirmado em 83,33% dos casos analisados.

Portanto, para se escolher o modelo mais preciso entre os sete concorrentes, primeiro observam-se quais os três que mais se aproximam da série temporal original e entre eles, escolhe-se aquele que acompanha a tendência e a variação da série no período anterior a previsão.

Para minimizar o erro, foi adotada a seguinte abordagem de mitigação diante dos objetivos da pesquisa: Escolher o modelo (Rede Neural Autorregressiva, ARIMA, Suavização Exponencial, Modelo Combinado, Previsão por Decomposição ou método ingênuo/inocente),

o qual apresentou o melhor desempenho (erro de previsão dentro do intervalo de confiança estabelecido) com base nos dados da série temporal analisada.

O resultado foi que em 77,28% dos casos analisados, os métodos estatísticos utilizados para a previsão de internações hospitalares por dengue conseguiram precisão com erro abaixo de 26%.

Assim, o algoritmo programado para a previsão de internações hospitalares por dengue apresenta erro dentro do intervalo de confiança estabelecido e pode ser utilizado para a implementação de aplicação que possibilite ao gestor conhecer, antecipadamente, a provável demanda.

5.1 LIMITAÇÕES E TRABALHOS FUTUROS

Esta pesquisa lidou, exclusivamente, com dados de internações hospitalares por dengue do território brasileiro e utilizou as técnicas de previsão de séries temporais implementadas no pacote *forecast* disponível para a linguagem de programação R.

A confiabilidade dos dados utilizados não foi verificada diretamente em virtude da sua quantidade e abrangência, mas como se trata de condição para o recebimento dos recursos financeiros do SUS pelas internações hospitalares realizadas nos estabelecimentos de saúde, acredita-se que os mesmos são confiáveis. Ressalta-se que não fica excluída a possibilidade de os estabelecimentos de saúde não enviarem todos os dados produzidos sobre internações e/ou realizar adulterações nos mesmos.

Na evidência de Unidades da Federação com baixas ocorrências de internações por dengue considerando as semanas epidemiológicas, o que classificaria os dados como série temporal contável, isto é, discreta em vez de contínua, os métodos estatísticos utilizados podem não ser os mais adequados, pois os mesmos tentam encontrar padrões nos dados, o que implica na disponibilidade em quantidade e variedade.

Neste sentido, como trabalhos futuros, pretende-se explorar outros métodos estatísticos que realizam previsão em séries temporais contáveis, isto é, aquelas que apresentam baixa quantidade de dados por período de tempo. A possibilidade de previsão de séries temporais contáveis abre espaço para a previsão de internações por dengue com foco no estabelecimento de saúde onde os dados ocorrem.

O algoritmo desenvolvido nesta pesquisa pode ser a base para a realização da previsão de internações por dengue, mas não chega a ser uma aplicação com elementos que facilitem o

seu uso no dia a dia da gestão da saúde pública. Assim, pretende-se desenvolver, com base nele, um serviço *web* que conte com facilidade de uso, o qual possa ser utilizado no âmbito da gestão em saúde no que diz respeito ao controle do vetor transmissor da dengue no Brasil.

O erro na previsão das internações hospitalares por dengue é fator determinante para o sucesso do algoritmo e sua precisão depende da aproximação deste erro a zero. Logo, é importante revisar, no futuro, o quanto o estado da arte avançou nos estudos sobre a redução do erro em previsão de séries temporais, ou seja, o cálculo do erro previsto.

Técnicas de aprendizagem de máquina estão avançando rapidamente. Logo, pretende-se realizar Revisão Sistemática da Literatura com o objetivo de responder às perguntas: Quais métodos estatísticos para a previsão de séries temporais (discretas/contáveis ou contínuas) são mais utilizados? Os métodos clássicos são mais eficientes que as novas abordagens com aprendizagem de máquina? Quais métricas de erro são mais seguras para a escolha do modelo mais preciso entre os candidatos a determinada previsão de série temporal?

A definição do algoritmo de previsão de internações hospitalares por dengue considerou pacotes disponíveis para a linguagem de programação R. Pretende-se também utilizar a linguagem de programação Python e os recursos disponíveis para a realização de previsão de séries temporais e realizar uma comparação entre as duas tecnologias. Previsão de séries temporais em R versus previsão de séries temporais em Python.

O horizonte de previsão foi de 8 semanas epidemiológicas e se estendesse o horizonte de previsão para 12, 16 ou 20 semanas, será que o erro na precisão da previsão aumentaria a ponto de tornar inviável tal abordagem?

REFERÊNCIAS

RERERÊNCIAS

ANTUNES, J. L. F.; CARDOSO, M. R. A. Uso da análise de séries temporais em estudos epidemiológicos. **Epidemiologia e Serviços de Saúde**, v. 24, n. 3, p. 565–576, set. 2015.

BANDEIRA, A. B.; ARRUDA, G. **CONTROLE DA DENGUE: UM DESAFIO PARA A ESCOLA E A SOCIEDADE**. In: PARANÁ. Secretaria de Estado da Educação. Superintendência de Educação. Os Desafios da Escola Pública Paranaense na Perspectiva do Professor PDE, 2013. Curitiba: SEED/PR., 2016. v. 1, (Cadernos PDE). Disponível em: <http://www.gestaoescolar.diaadia.pr.gov.br/modules/conteudo/conteudo.php?conteudo=20>. Acesso em: 01 jan. 2020.

BITTENCOURT, S. A; CAMACHO, L. A. B.; LEAL, M. C. **O Sistema de Informação Hospitalar e sua aplicação na saúde coletiva**. Cad. Saúde Pública, v. 22, n. 1, p. 19-30, 2006.

BRASIL. Ministério da Saúde. Secretaria de Vigilância em Saúde. Departamento de Vigilância Epidemiológica. **Diretrizes nacionais para prevenção e controle de epidemias de dengue**. Brasília: Ministério da Saúde, 2009. 160 p. – (Série A. Normas e Manuais Técnicos).

BRASIL. Ministério da Saúde. **Política Nacional de Atenção Básica**. Brasília: Ministério da Saúde, 2012. (Série E. Legislação em Saúde).

BRASIL. Lei nº 13.595, de 5 de janeiro de 2018. Altera a Lei nº 11.350, de 5 de outubro de 2006, para dispor sobre a reformulação das atribuições, a jornada e as condições de trabalho, o grau de formação profissional, os cursos de formação técnica e continuada e a indenização de transporte dos profissionais Agentes Comunitários de Saúde e Agentes de Combate às Endemias. **Diário Oficial da União**. 74. ed. Brasília, DF, 18 abr. 2018. Seção 1, p. 3.

BRASIL. Lei nº 12.527, de 18 de novembro de 2011 que regula o acesso a informações previsto no inciso XXXIII do art. 5º, no inciso II do § 3º do art. 37 e no § 2º do art. 216 da Constituição Federal; altera a Lei nº 8.112, de 11 de dezembro de 1990; revoga a Lei nº 11.111, de 5 de maio de 2005, e dispositivos da Lei nº 8.159, de 8 de janeiro de 1991; e dá outras providências. **Diário Oficial da União**. Edição Extra. Brasília, DF, 18 nov. 2011. Seção 1, p. 1.

BRASIL. Ministério da Saúde. Conselho Nacional de Saúde. Resolução nº 510 de 7 de abril de 2016 que dispõe sobre as normas aplicáveis a pesquisas em Ciências Humanas e Sociais cujos procedimentos metodológicos envolvam a utilização de dados diretamente obtidos com os participantes ou de informações identificáveis ou que possam acarretar riscos maiores do que os existentes na vida cotidiana. **Diário Oficial da União**. 98. ed. Brasília, DF, 25 mai. 2016. Seção 1, p. 44.

CANDIDA, S.; OUCHANA, G. Chafarizes, praças e lagos preocupam vizinhos, que temem focos de ‘Aedes’. **O Globo**, Rio de Janeiro, 13 fev. 2016, 12:26. Disponível em: <https://oglobo.globo.com/rio/chafarizes-pracas-lagos-preocupam-vizinhos-que-temem-focos-de-aedes-18665449>. Acesso em: 01 jan. 2020.

CASALI, M. H. *et al.* Desenvolvimento de uma Ferramenta Web para Migração e Visualização de Dados Abertos do DATASUS. **Revista Eletrônica de Sistemas de Informação e de Gestão Tecnológica**, v. 9, n. 3, 2018.

CBCD. **CID-10**. Disponível em: <http://www.datasus.gov.br/cid10/V2008/cid10.htm>. Acesso em: 01 jan. 2020.

CHAKRABORTY, T.; CHATTOPADHYAY, S.; GHOSH, I. Forecasting dengue epidemics using a hybrid methodology. **Physica A: Statistical Mechanics and its Applications**, v. 527, p. 121266, ago. 2019.

CHATFIELD, C. **The Analysis of Time Series: An Introduction**. 5. ed. New York: Chapman and Hall/CRC, 1996.

CHATFIELD, C. **The Analysis of Time Series: An Introduction**. 6. ed. New York: Chapman and Hall/CRC, 2003.

CLEMEN, R. T. Combining forecasts: A review and annotated bibliography. **International Journal of Forecasting**, v. 5, n. 4, p. 559–583, jan. 1989.

COELHO, G. E. *et al.* Sensitivity of the Dengue Surveillance System in Brazil for Detecting Hospitalized Cases. **PLOS Neglected Tropical Diseases**, v. 10, n. 5, p. e0004705, 18 maio 2016.

CORMEN, T. H. *et al.* **Algoritmos: Teoria e Prática**. Tradução: Arlete Simille Marques. 3. ed. Rio de Janeiro: Elsevier, 2012.

COWPERTWAIT, P. S. P.; METCALFE, A. V. **Introductory Time Series with R**. New York: Springer-Verlag New York, 2009.

DATASUS. **Início, Serviços, Transferência/Download de Arquivos, Arquivos de Dados**. Disponível em: <http://www2.datasus.gov.br/DATASUS/index.php?area=0901&item=1&acao=25>. Acesso em: 01 jan. 2020.

DATASUS. **Infraestrutura Nacional de Dados Abertos**. 2012. Disponível em: <http://www2.datasus.gov.br/DATASUS/index.php?acao=11&id=33150>. Acesso em: 01 jan. 2020.

DIAS, L. B. A. *et al.* Dengue: transmissão, aspectos clínicos, diagnóstico e tratamento. **Medicina (Ribeirão Preto. Online)**, v. 43, n. 2, p. 143, 30 jun. 2010.

DIVE. Diretoria de Vigilância Epidemiológica. **Vigilância e controle aedes aegypti: orientações técnicas para pessoal decampo**. 2. ed. Florianópolis: Secretaria de Estado da Saúde de Santa Catarina, 2019.

DICIO. **Holística**. Disponível em: <https://www.dicio.com.br/holistica/>. Acesso em: 01 jan. 2020.

E-GESTOR. **Relatório: histórico de cobertura**. Disponível em: <https://egestorab.saude.gov.br/paginas/acesoPublico/relatorios/relHistoricoCobertura.xhtml>. Acesso em: 01 jan. 2020.

ELPIDIO DE SÁ, F. *et al.* Produção de sentidos parentais no cuidado de crianças com microcefalia por vírus zika. **Revista Brasileira em Promoção da Saúde**, v. 30, n. 4, p. 1–10, 6 dez. 2017.

FERREIRA, P. C. *et al.* **Análise de Séries Temporais em R: um curso introdutório**. 1. ed. Rio de Janeiro: Elsevier, 2017.

FIGUEREDO, Jessé Miranda de. **Análise espaço-temporal dos casos de dengue no município de Ribeirão Preto (SP) pela técnica de geoprocessamento**. 2009. 124 f. Dissertação (Mestrado em Tecnologia Ambiental) - Universidade de Ribeirão Preto, Ribeirão Preto, 2009.

G1. Casos de dengue no Brasil crescem 303% em 2019. **g1.globo.com**, 16 abr. 2019, 10:48. Disponível em: <https://g1.globo.com/bemestar/noticia/2019/04/16/casos-de-dengue-no-brasil-crescem-303percent-em-2019.ghtml>. Acesso em: 01 jan. 2020.

GUERRIERO, I. C. Z. Resolução nº 510 de 7 de abril de 2016 que trata das especificidades éticas das pesquisas nas ciências humanas e sociais e de outras que utilizam metodologias próprias dessas áreas. **Ciência & Saúde Coletiva**, v. 21, p. 2619-2629, 2016.

HYNDMAN, R. J. *et al.* forecast: Forecasting functions for time series and linear models. **R package version 8,9**. Disponível em: <http://pkg.robjhyndman.com/forecast>. Acesso em: 01 jan. 2020.

HYNDMAN, R. J.; ATHANASOPOULOS, G. **Forecasting: Principles and Practice**. 2. ed. Melbourne, Australia: OTexts, 2018.

HYNDMAN, R. J.; KHANDAKAR, Y. Automatic Time Series Forecasting: The forecast Package for R. **Journal of Statistical Software**, v. 27, n. 3, p. 1–22, 2008.

LATORRE, M. DO R. D. DE O.; CARDOSO, M. R. A. Análise de séries temporais em epidemiologia: uma introdução sobre os aspectos metodológicos. **Revista Brasileira de Epidemiologia**, v. 4, n. 3, p. 145–152, nov. 2001.

LIZZI, E. A. D. S. **Predição do número mensal de casos de dengue por modelos de séries temporais**. 2012. 131 páginas. Dissertação (Mestrado em Saúde na Comunidade) – Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto, 2012.

LONGLEY, P. A. *et al.* **Sistemas e ciência da informação geográfica**. 3. ed. Porto Alegre: Bookman, 2012.

MESQUITA, F. O. DE S.; PARENTE, A. S.; COELHO, G. M. P. Agentes Comunitários de Saúde e Sgntes de Combate a Endemias: Desafios para controle do Aedes aegypti. **Id on Line REVISTA DE PSICOLOGIA**, v. 11, n. 36, p. 64, 30 jul. 2017.

PETRUZALEK, D. **read.dbc**. 2016. Disponível em: <https://cran.r-project.org/web/packages/read.dbc/README.html>. Acesso em: 01 jan. 2020.

R CORE TEAM. R: A language and environment for statistical computing. **R Foundation for Statistical Computing**, Vienna, Austria. 2019. Disponível em: <https://www.R-project.org/>. Acesso em: 01 jan. 2020.

REICH, N. G. *et al.* Challenges in Real-Time Prediction of Infectious Disease: A Case Study of Dengue in Thailand. **PLOS Neglected Tropical Diseases**, v. 10, n. 6, p. e0004761, 15 jun. 2016.

REIS, C. A. DOS *et al.* **Análise de uma Dinâmica Populacional não Linear Simplificada do Aedes aegypti via Dinâmica Zero**. In: CONFERÊNCIA BRASILEIRA DE DINÂMICA, CONTROLE E APLICAÇÕES, 2017, São José do Rio Preto: DINCON, 2017.

RIBEIRO, A. L. N.; MOURA, G. L. DE. Análise das políticas públicas de combate à dengue. **O Manancial - Repositório Digital da UFSM**, p. 1–22, 2013.

ROCHA, D. C.; DANTAS, R. T.; CÂNDIDO, G. A. POLÍTICAS PÚBLICAS PARA A SAÚDE E O PAPEL DA ATENÇÃO BÁSICA DE SAÚDE NO CONTROLE E PREVENÇÃO DA DENGUE NO PAÍS. **revista da universidade vale do rio verde**, p. 754–764, 2014.

RUCHIRASET, A.; TANTRAKARNAPA, K. Time series modeling of pneumonia admissions and its association with air pollution and climate variables in Chiang Mai Province, Thailand. **Environmental Science and Pollution Research**, v. 25, n. 33, p. 33277–33285, 26 nov. 2018.

SANTOS, S. L. DOS *et al.* Dengue in Brazil and Colombia: a study of knowledge, attitudes, and practices. **Revista da Sociedade Brasileira de Medicina Tropical**, v. 47, n. 6, p. 783–787, dez. 2014.

SIH/SUS. **MORBIDADE HOSPITALAR DO SUS - POR LOCAL DE INTERNAÇÃO - BRASIL**. Disponível em: <http://tabnet.datasus.gov.br/cgi/tabcgi.exe?sih/cnv/niuf.def>. Acesso em: 01 jan. 2020.

SIHD. **Datas para entrega de remessas do SIHD pelo gestor ao DATASUS**. Disponível em: <http://www2.datasus.gov.br/SIHD/data-envio-remessas-SIHD>. Acesso em: 01 jan. 2020.

SILVA, W. R. S. **Metodologia de monitoramento de epidemias: uma abordagem baseada em redes neurais artificiais**. 2018. 75 f. Dissertação (Mestrado em Engenharia Elétrica) - Universidade Federal do Pará, Instituto de Tecnologia, Belém, 2018. Programa de Pós-Graduação em Engenharia Elétrica. Disponível em: <http://repositorio.ufpa.br/jspui/handle/2011/10056>. Acesso em: 01 jan. 2020.

SILVA, F. J. F. *et al.* **Visualização de grande quantidade de dados: uma análise do tempo de internação por dengue no Brasil**. In: CONGRESSO SERTANEJO DE COMPUTAÇÃO. 2., 2019, Patos/PB. *Anais eletrônicos...* Patos/PB: UEPB, 2019.

TSCHIEDEL, M.; BERVIAN, P. V. **A educação ambiental como mecanismo de controle da proliferação do mosquito *Aedes aegypti* na educação básica**. In: CONGRESSO INTERNACIONAL DE EDUCAÇÃO CIENTÍFICA E TECNOLÓGICA, 4, 2017, Santo Ângelo. Disponível em: http://www.santoangelo.uri.br/anais/ciecitec/2017/resumos/poster/trabalho_2881.docx. Acesso em: 01 jan. 2020.

VIEIRA MACHADO, A. A. *et al.* Direct Costs of Dengue Hospitalization in Brazil: Public and Private Health Care Systems and Use of WHO Guidelines. **PLoS Neglected Tropical Diseases**, v. 8, n. 9, p. e3104, 4 set. 2014.

VILLANI, M. *et al.* Time series modelling to forecast prehospital EMS demand for diabetic emergencies. **BMC Health Services Research**, v. 17, n. 1, p. 332, 5 dez. 2017.

WAZLAWICK, R. S. **Metodologia de pesquisa em ciência da computação**. 2. ed. São Paulo: Elsevier, 2017.

WHO. **Dengue: guidelines for diagnosis, treatment, prevention and control**. 3. ed. Geneva: World Health Organization, 2009.

WHO. Dengue and Severe Dengue. **World Health Organization**, 15 abr. 2019. Disponível em: <https://www.who.int/news-room/fact-sheets/detail/dengue-and-severe-dengue>. Acesso em: 01 jan. 2020.

ZHANG, J. H.; YUAN, J.; WANG, T. Direct cost of dengue hospitalization in Zhongshan, China: Associations with demographics, virus types and hospital accreditation. **PLoS Neglected Tropical Diseases**, v. 11, n. 8, p. 1–15, 2017.

ZHOU, L. *et al.* Time series model for forecasting the number of new admission inpatients. **BMC Medical Informatics and Decision Making**, v. 18, n. 1, p. 39, 15 dez. 2018.

APÊNDICES

APÊNDICE A – Decomposição das séries temporais por Unidade da Federação

Este apêndice apresenta breve caracterização territorial administrativa do Brasil, bem como a decomposição, com o uso da função *mstl()* do pacote *forecast* para linguagem R, das séries temporais de internações hospitalares por dengue no período de janeiro de 2016 a junho de 2019.

Cada Unidade da Federação brasileira teve uma série histórica nomeada com sua sigla de UF, a qual foi decomposta em seus elementos: Tendência/Ciclo (o ciclo foi decomposto junto com a tendência), Sazonalidade e Restante (dados que sofrem influência de elementos que não são tendência e/ou sazonalidade).

Também foi feito o teste de hipótese (*Kwiatkowski-Phillips-Schmidt-Shin* – KPSS) para verificação da condição de estacionariedade. A estacionariedade foi testada utilizando-se a função *ur.kpss()* do pacote *urca* disponível para a linguagem R. O KPSS admite como hipótese nula que a série temporal é estacionária (se o valor p for menor que 0,05).

O Brasil está dividido em 5 (cinco) regiões administrativas: Norte, Centro-Oeste, Sul, Sudeste e Nordeste. A Figura A apresenta a concentração populacional por cidade brasileira de acordo com o censo populacional realizado pelo IBGE no ano de 2010 e estimativa populacional para o ano de 2012. O agrupamento de Unidades da Federação por região é o seguinte:

- I. Norte (Tocantins, Pará, Amapá, Roraima, Amazonas, Acre e Rondônia);
- II. Centro-Oeste (Goiás, Mato Grosso do Sul, Mato Grosso e Distrito Federal);
- III. Sul (Paraná, Santa Catarina e Rio Grande do Sul);
- IV. Sudeste (São Paulo, Rio de Janeiro, Espírito Santo e Minas Gerais); e
- V. Nordeste (Piauí, Maranhão, Ceará, Rio Grande do Norte, Paraíba, Pernambuco, Sergipe, Alagoas e Bahia).

As regiões Sudeste, Nordeste, Sul e Centro-Oeste são as que apresentam a maior densidade populacional no território brasileiro. A região Norte, mesmo sendo a maior em território, apresenta menor densidade populacional. Pelo que se observa no mapa, as pessoas estão concentradas mais perto da costa marítima.

Ao analisar os gráficos de cada série temporal, é importante observar a concentração populacional e o período de chuvas em determinado estado, pois esses dois fatores influenciam no desenvolvimento do mosquito *Aedes aegypti*. Para o desenvolvimento desta pesquisa, foi suficiente realizar a análise apenas dos componentes da série temporal sem se preocupar com os fatores que influenciam na ocorrência dos dados, uma vez que se trata de um estudo para a realização de previsão de séries temporais com o objetivo de criar algoritmo para estimar as internações por dengue em cada estado do Brasil.

FIGURA A - DISTRIBUIÇÃO POPULACIONAL NO BRASIL



Fonte: Silva et al. (2019)

As séries temporais decompostas foram agrupadas por região do Brasil onde a Unidade da Federação se encontra. Cada Figura que segue apresenta quatro gráficos: Dados (a série temporal original), Tendência (os ciclos e a tendência da série temporal decomposta), Sazonalidade (os ciclos sazonais da série) e Restante (aquilo que não se enquadra em tendência, ciclo ou sazonalidade). Não foram feitas considerações sobre o Restante, pois nada pode-se afirmar a partir da decomposição da série temporal.

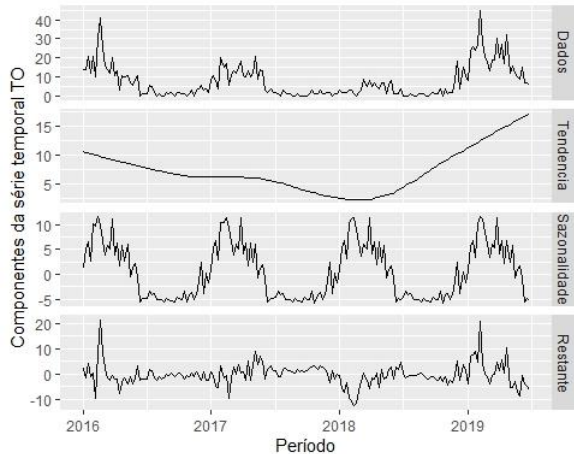
I) Região Norte

É formada pelos estados: Tocantins, Pará, Amapá, Roraima, Amazonas, Acre e Rondônia.

As Figuras A.1 e A.2 apresentam os elementos constituintes das séries temporais dos estados do Tocantins e Pará, respectivamente. Em Tocantins, os Dados tem quantidades variando entre 0 e 45, a tendência é decrescente, constante e crescente, a sazonalidade tem variação constante com pontos de máximo ocorrendo entre meados e final do primeiro semestre de cada ano e a série temporal não é estacionária (valor $p = 0,38$). No estado do Pará, os Dados tem

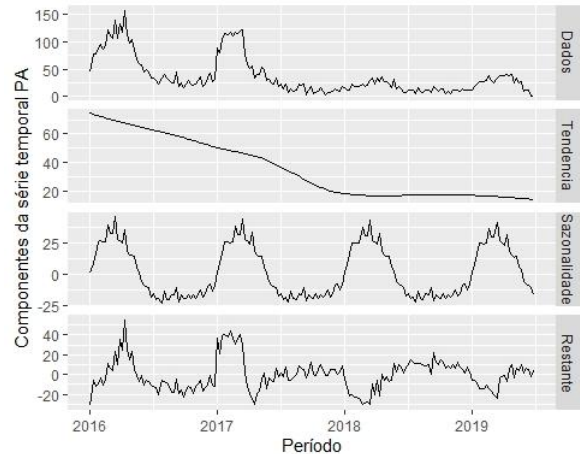
valores entre 1 e 158, a tendência é decrescente, constante e depois levemente crescente, a sazonalidade tem variação constante com picos no meio do primeiro semestre de cada ano e a série temporal não é estacionária (valor $p = 1,53$).

FIGURA A.1 - DECOMPOSIÇÃO DA SÉRIE TEMPORAL DO TOCANTINS



Fonte: Autoria própria, 2019

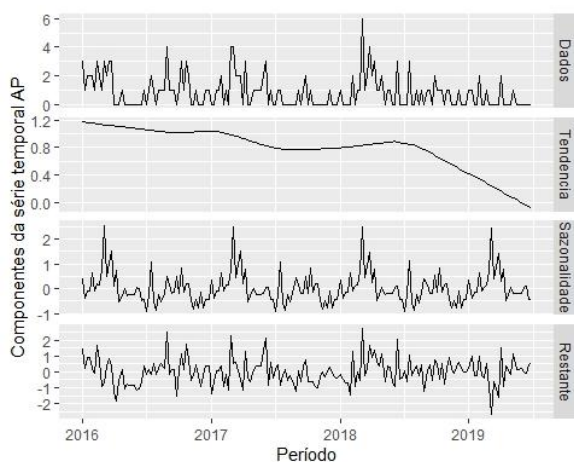
FIGURA A.2 - DECOMPOSIÇÃO DA SÉRIE TEMPORAL DO PARÁ



Fonte: Autoria própria, 2019

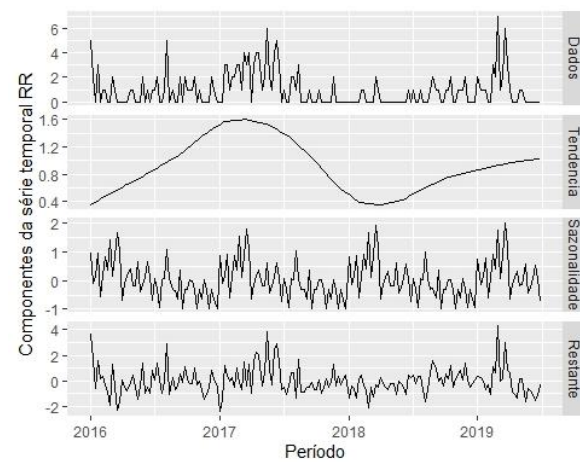
As Figuras A.3 e A.4 apresentam a decomposição das séries temporais dos estados do Amapá e Roraima. No estado de Amapá, os Dados tem valores entre 0 e 6, a tendência é decrescente, a sazonalidade tem variação constante com picos no início do primeiro semestre de cada ano e a série temporal não é estacionária (valor $p = 0,54$). Em Roraima, a série temporal tem Dados com quantidades entre 0 e 7, a tendência é crescente, decrescente e depois crescente, a sazonalidade tem variação constante com pontos de máximo ocorrendo entre início e meados do primeiro semestre de cada ano e a série temporal não é estacionária (valor $p = 0,26$).

FIGURA A.3 - DECOMPOSIÇÃO DA SÉRIE TEMPORAL DO AMAPÁ



Fonte: Autoria própria, 2019.

FIGURA A.4 - DECOMPOSIÇÃO DA SÉRIE TEMPORAL DE RORAIMA

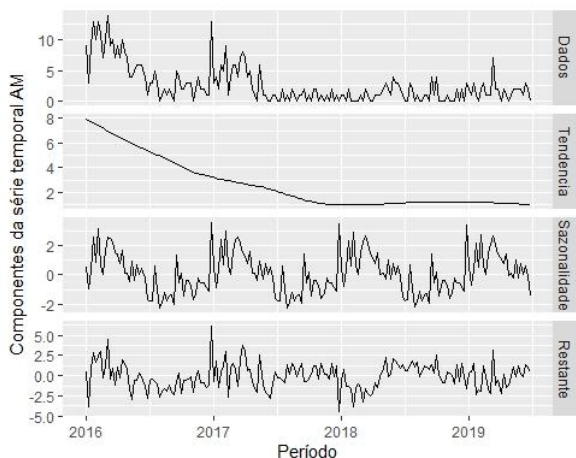


Fonte: Autoria própria, 2019.

As Figuras A.5 e A.6 apresentam os componentes das séries temporais dos estados do Amazonas e do Acre. No Amazonas, os Dados tem valores compreendidos entre 0 e 14, a

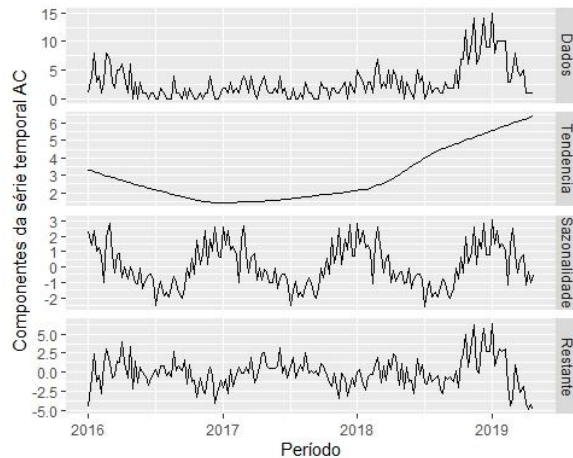
tendência é decrescente e depois constante, a sazonalidade tem variação constante com picos no início do primeiro semestre de cada ano e a série temporal não é estacionária (valor $p = 1,76$). No Acre, os Dados tem valores compreendidos entre 0 e 15, a tendência é decrescente, constante e depois crescente, a sazonalidade tem variação constante com picos no final do último semestre e início do primeiro semestre de cada ano e a série temporal não é estacionária (valor $p = 1,04$).

FIGURA A.5 - DECOMPOSIÇÃO DA SÉRIE TEMPORAL DO AMAZONAS



Fonte: Autoria própria, 2019.

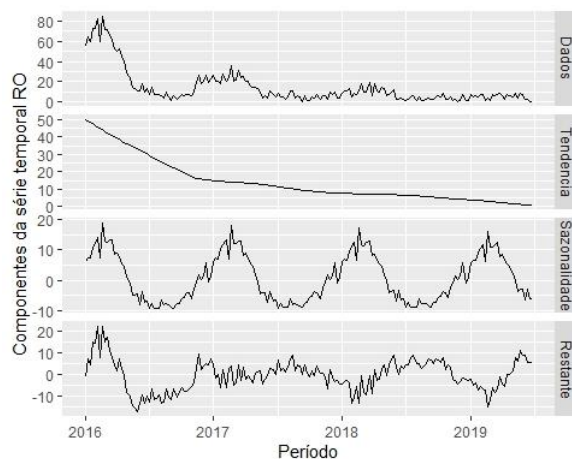
FIGURA A.6 - DECOMPOSIÇÃO DA SÉRIE TEMPORAL DO ACRE



Fonte: Autoria própria, 2019.

A Figura A.7 mostra os componentes da série temporal de Rondônia com quantidades compreendidas entre 0 e 85, a tendência é decrescente, a sazonalidade tem variação constante com ponto de máximo no início do primeiro semestre de cada ano e a série temporal não é estacionária (valor $p = 1,55$).

FIGURA A.7 - DECOMPOSIÇÃO DA SÉRIE TEMPORAL DE RONDÔNIA



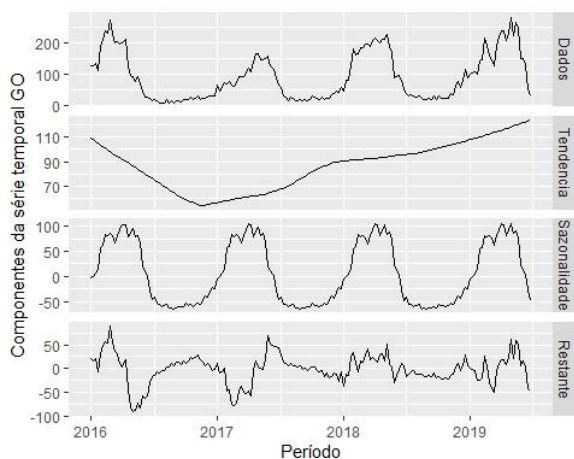
Fonte: Autoria própria, 2019.

II) Região Centro-Oeste

É formada pelas Unidades da Federação: Goiás, Mato Grosso do Sul, Mato Grosso e Distrito Federal.

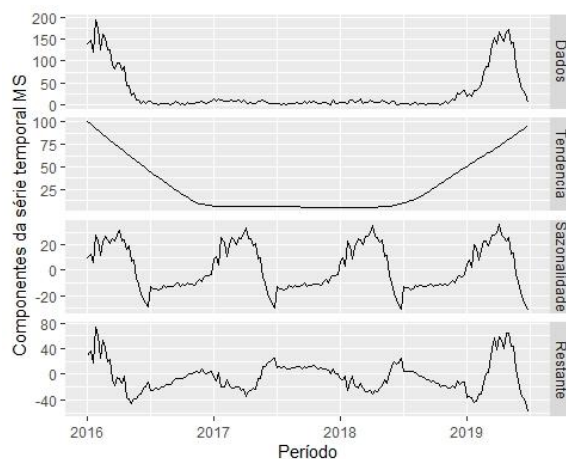
As Figuras A.8 e A.9 apresentam os componentes das séries temporais dos estados de Goiás e estado e Mato Grosso do Sul. Em Goiás, os Dados tem quantidades variando entre 7 e 284, a tendência é decrescente e crescente, a sazonalidade tem variação constante com picos que ocorrem em meados do primeiro semestre de cada ano e a série temporal não é estacionária (valor $p = 0,32$). Já no MS, os Dados tem quantidades entre 0 e 195, a tendência é decrescente, constante e depois crescente, a sazonalidade tem variação constante com picos no meio do primeiro semestre de cada ano e a série temporal não é estacionária (valor $p = 0,47$).

FIGURA A.8 - DECOMPOSIÇÃO DA SÉRIE TEMPORAL DO ESTADO DE GOIÁS



Fonte: Autoria própria, 2019.

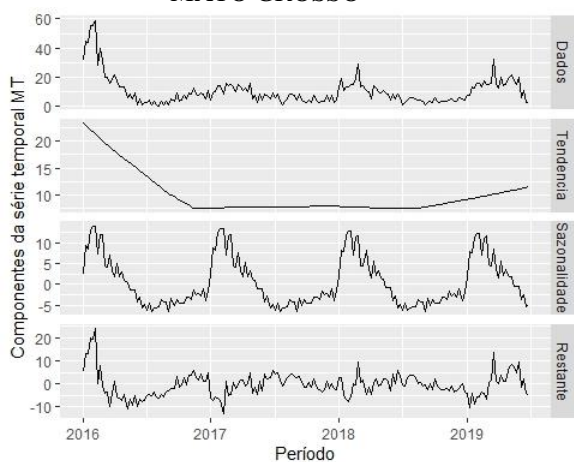
FIGURA A.9 - DECOMPOSIÇÃO DA SÉRIE TEMPORAL DO MATO GROSSO DO SUL



Fonte: Autoria própria, 2019.

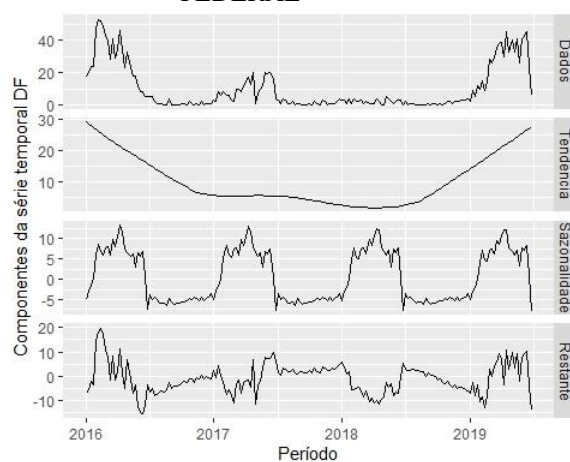
As Figuras A.10 e A.11 apresentam os componentes das séries temporais do Mato Grosso e do Distrito Federal. No Mato Grosso, os Dados apresentaram quantidades compreendidas entre 0 e 59, a tendência é decrescente e depois constante, a sazonalidade tem variação constante com picos no início do primeiro semestre de cada ano e a série temporal não é estacionária (valor $p = 0,40$). No DF, que faz parte do estado de Goiás, os Dados tem quantidades compreendidas entre 0 e 53, a tendência é decrescente, constante e crescente, a sazonalidade tem variação constante com picos no meio do primeiro semestre de cada ano e a série temporal não é estacionária (valor $p = 0,49$).

FIGURA A.10 - DECOMPOSIÇÃO DA SÉRIE TEMPORAL DO ESTADO DE MATO GROSSO



Fonte: Autoria própria, 2019.

FIGURA A.11 - DECOMPOSIÇÃO DA SÉRIE TEMPORAL DO DISTRITO FEDERAL



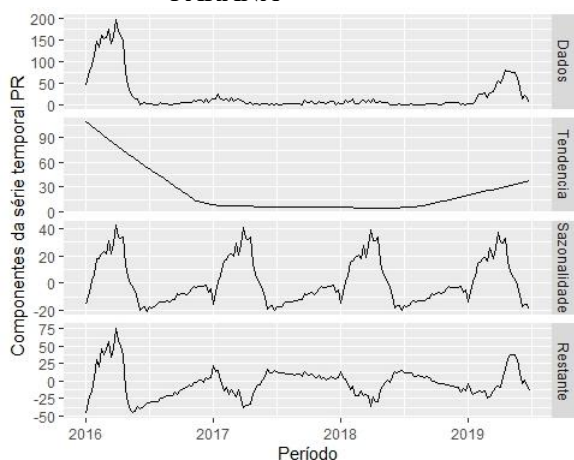
Fonte: Autoria própria, 2019.

III) Região Sul

É composta pelos estados: Paraná, Santa Catarina e Rio Grande do Sul.

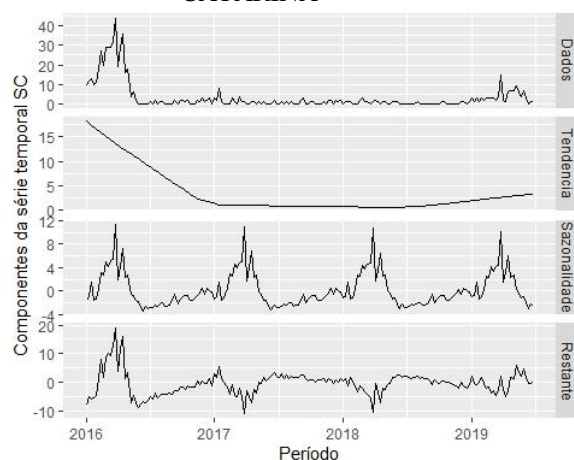
As Figuras A.12 e A.13 apresentam os componentes das séries temporais dos estados do Paraná e Santa Catarina, respectivamente. O Paraná tem Dados com quantidades entre 0 e 199, a tendência é decrescente, constante e depois levemente crescente, a sazonalidade tem variação constante com máximos no meio do primeiro semestre de cada ano e a série temporal não é estacionária (valor $p = 0,71$). No estado de Santa Catarina, a série temporal tem Dados com quantidades entre 0 e 44, a tendência é decrescente, constante e depois levemente crescente, a sazonalidade tem variação constante com picos no meio do primeiro semestre de cada ano e a série temporal não é estacionária (valor $p = 0,84$).

FIGURA A.12 - DECOMPOSIÇÃO DA SÉRIE TEMPORAL DO ESTADO DO PARANÁ



Fonte: Autoria própria, 2019.

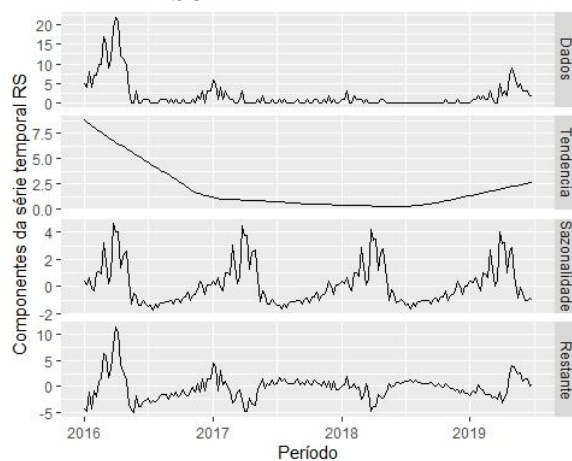
FIGURA A.13 - DECOMPOSIÇÃO DA SÉRIE TEMPORAL DE SANTA CATARINA



Fonte: Autoria própria, 2019.

A Figura A.14 apresenta os componentes da série temporal do Rio Grande do Sul, onde os Dados tem quantidades compreendidos entre 0 e 22, a tendência é decrescente, constante e levemente crescente a sazonalidade tem variação constante com picos que ocorrem sempre no meio do primeiro semestre de cada ano e a série temporal não é estacionária (valor $p = 0,78$).

FIGURA A.14 - DECOMPOSIÇÃO DA SÉRIE TEMPORAL DO RIO GRANDE DO SUL



Fonte: Autoria própria, 2019.

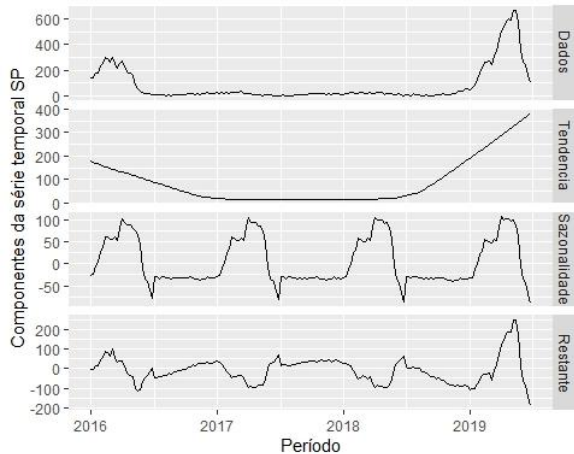
IV) Região Sudeste

Composta pelos estados: São Paulo, Rio de Janeiro, Espírito Santo e Minas Gerais.

As Figuras A.15 e A.16 apresentam os componentes das séries temporais de São Paulo e Rio de Janeiro, respectivamente. Em SP, a série temporal tem Dados com quantidades entre 1 e 670, a tendência é decrescente, constante e depois crescente, a sazonalidade tem variação constante com pontos de máximo ocorrendo no final do primeiro semestre de cada ano e a série

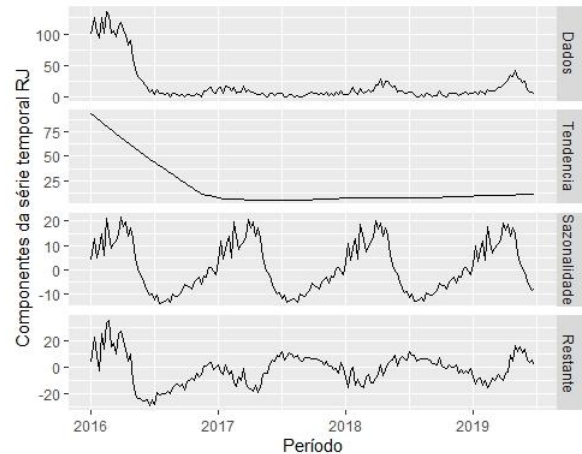
temporal não é estacionária (valor $p = 0,68$). A série temporal do RJ apresenta Dados com quantidades compreendidas entre 0 e 136, a tendência é decrescente e constante, a sazonalidade tem variação constante com pontos de máximo no início e no meio do primeiro semestre de cada ano e a série temporal não é estacionária (valor $p = 1,05$).

FIGURA A.15 - DECOMPOSIÇÃO DA SÉRIE TEMPORAL DE SÃO PAULO



Fonte: Autoria própria, 2019.

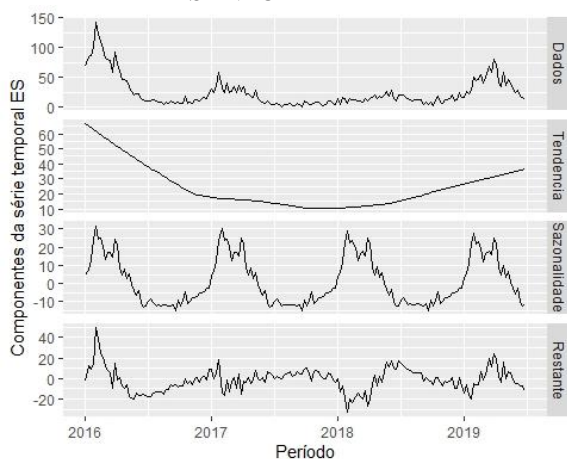
FIGURA A.16 - DECOMPOSIÇÃO DA SÉRIE TEMPORAL DO RIO DE JANEIRO



Fonte: Autoria própria, 2019.

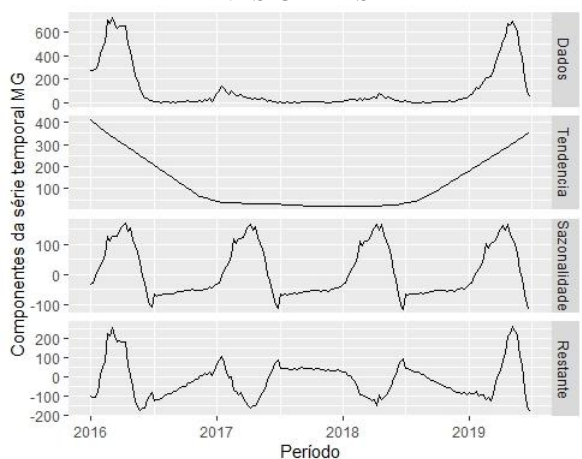
As Figuras A.17 e A.18 apresentam os componentes das séries temporais dos estados do Espírito Santo e Minas Gerais. No ES, os Dados tem quantidades entre 1 e 143, a tendência é decrescente, constante e crescente, a sazonalidade tem variação constante com picos que ocorrem no início do primeiro semestre de cada ano e a série temporal não é estacionária (valor $p = 0,59$). Em Minas Gerais, os Dados tem quantidades variando entre 2 e 724, a tendência é decrescente, constante e depois crescente, a sazonalidade tem variação constante com picos ocorrendo em meados do primeiro semestre de cada ano e a série temporal não é estacionária (valor $p = 0,47$).

FIGURA A.17 - DECOMPOSIÇÃO DA SÉRIE TEMPORAL DO ESPÍRITO SANTO



Fonte: Autoria própria, 2019.

FIGURA A.18 - DECOMPOSIÇÃO DA SÉRIE TEMPORAL DO ESTADO DE MINAS GERAIS



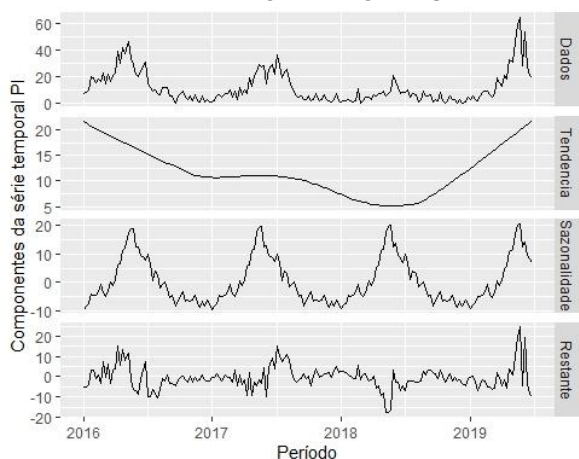
Fonte: Autoria própria, 2019.

V) Região Nordeste

É composta pelas Unidades da Federação: Piauí, Maranhão, Ceará, Rio Grande do Norte, Paraíba, Pernambuco, Sergipe, Alagoas e Bahia.

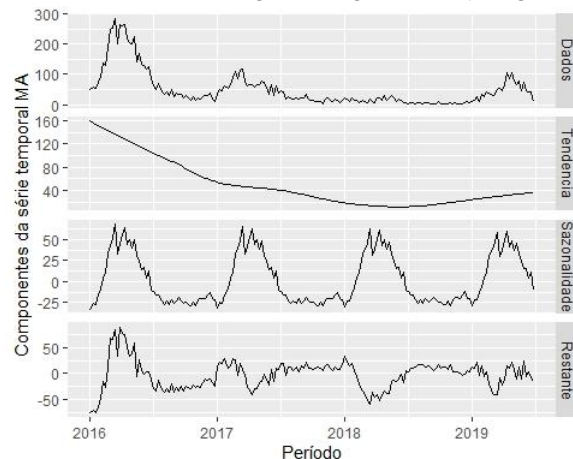
As Figuras A.19 e A.20 mostram os componentes das séries temporais do Piauí e do Maranhão. O Piauí tem Dados com quantidades compreendidas entre 0 e 65, a tendência é decrescente, constante, decrescente, constante e depois crescente, a sazonalidade tem variação constante com pontos de máximo no final do primeiro semestre de cada ano e a série temporal não é estacionária (valor $p = 0,30$). Já no Maranhão, os Dados tem quantidades entre 2 e 287, a tendência é decrescente e depois levemente crescente, a sazonalidade tem variação constante com picos que ocorrem em meados do primeiro semestre de cada ano e a série temporal não é estacionária (valor $p = 1,37$).

FIGURA A.19 - DECOMPOSIÇÃO DA SÉRIE TEMPORAL DO PIAUÍ



Fonte: Autoria própria, 2019.

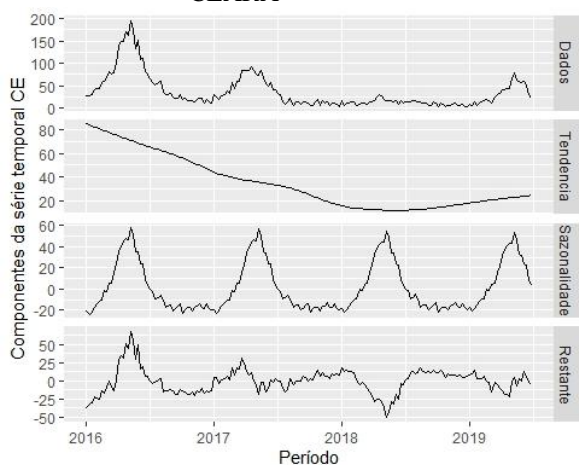
FIGURA A.20 - DECOMPOSIÇÃO DA SÉRIE TEMPORAL DO MARANHÃO



Fonte: Autoria própria, 2019.

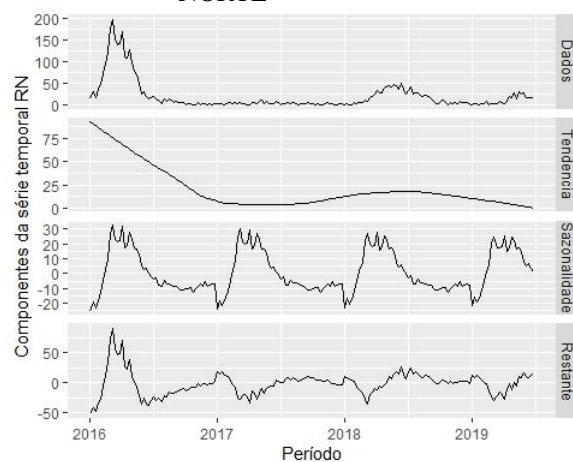
As Figuras A.21 e A.22 mostram os componentes das séries temporais dos estados do Ceará e Rio Grande do Norte. No Ceará, os Dados tem quantidade entre 3 e 196, a tendência é decrescente, constante e levemente crescente, a sazonalidade tem variação constante com picos no final do primeiro semestre de cada ano e a série temporal não é estacionária (valor $p = 1,17$). Já no Rio Grande do Norte, a série temporal tem Dados com quantidades entre 0 e 198, a tendência é decrescente e levemente constante, a sazonalidade tem variação constante com picos que ocorrem no meio do primeiro semestre de cada ano e a série temporal não é estacionária (valor $p = 0,70$).

FIGURA A.21 - DECOMPOSIÇÃO DA SÉRIE TEMPORAL DO ESTADO DO CEARÁ



Fonte: Autoria própria, 2019.

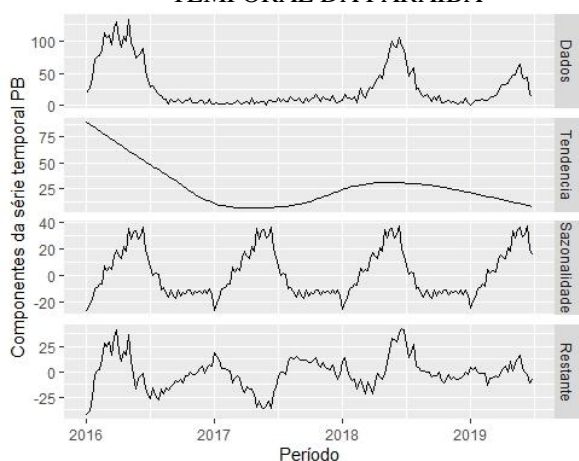
FIGURA A.22 - DECOMPOSIÇÃO DA SÉRIE TEMPORAL DE RIO GRANDE DO NORTE



Fonte: Autoria própria, 2019.

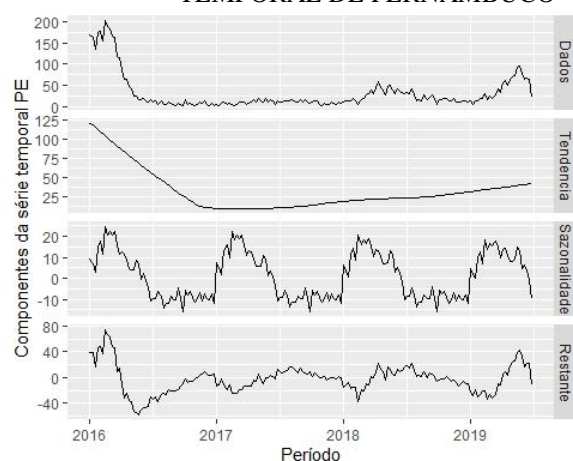
As Figuras A.23 e A.24 evidenciam os componentes das séries temporais do estado da Paraíba e do Pernambuco. A Paraíba apresenta Dados com quantidades que estão entre 1 e 134, a tendência é decrescente, constante, crescente, constante e depois decrescente a sazonalidade tem variação constante com picos no final do primeiro semestre de cada ano e a série temporal não é estacionária (valor $p = 0,48$). O Pernambuco tem Dados com quantidades entre 1 e 204, a tendência é decrescente, constante e depois levemente crescente, a sazonalidade tem variação constante com máximos no início do primeiro semestre de cada ano e a série temporal não é estacionária (valor $p = 0,53$).

FIGURA A.23 - DECOMPOSIÇÃO DA SÉRIE TEMPORAL DA PARAÍBA



Fonte: Autoria própria, 2019.

FIGURA A.24 - DECOMPOSIÇÃO DA SÉRIE TEMPORAL DE PERNAMBUCO

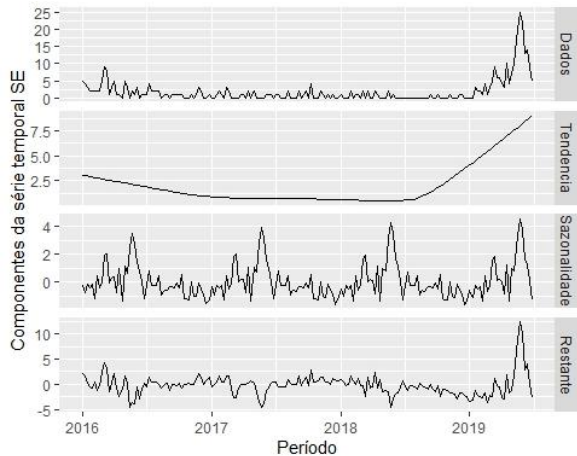


Fonte: Autoria própria, 2019.

As Figuras A.25 e A.26 apresentam os componentes das séries temporais de Sergipe e Alagoas. No estado de SE, os Dados tem quantidades compreendidas entre 0 e 25, a tendência é levemente decrescente, constante e depois crescente, a sazonalidade tem variação constante com pontos de máximo que ocorrem no final do primeiro semestre de cada ano e a série temporal não

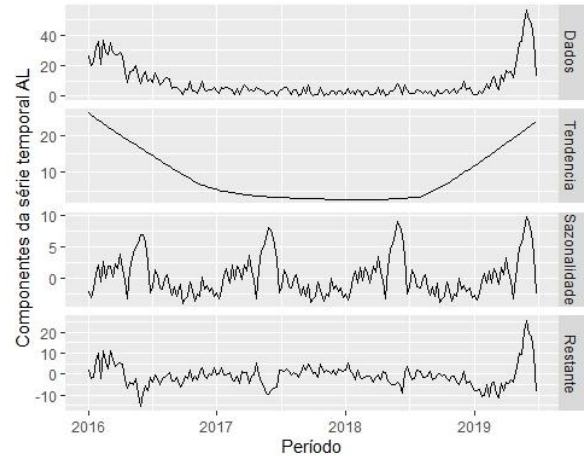
é estacionária (valor $p = 0,57$). No estado de Alagoas, os Dados tem valores entre 0 e 57, a tendência é decrescente, constante e depois crescente, a sazonalidade tem variação constante com picos no final do primeiro semestre de cada ano e a série temporal não é estacionária (valor $p = 0,57$).

FIGURA A.25 - DECOMPOSIÇÃO DA SÉRIE TEMPORAL DE SERGIPE



Fonte: Autoria própria, 2019.

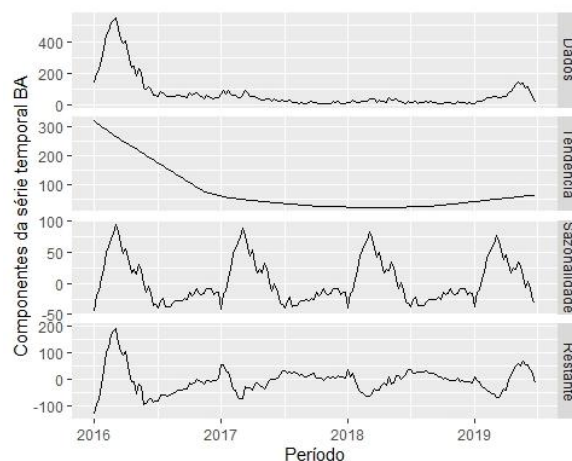
FIGURA A.26 - DECOMPOSIÇÃO DA SÉRIE TEMPORAL DE ALAGOAS



Fonte: Autoria própria, 2019.

A Figura A.27 apresenta os componentes da série temporal da Bahia, em que os Dados tem quantidades compreendidas entre 9 e 551, a tendência é decrescente e depois constante, a sazonalidade tem variação constante com picos que ocorrem no início do semestre de cada ano e a série temporal não é estacionária (valor $p = 1,27$).

FIGURA A.27 - DECOMPOSIÇÃO DA SÉRIE TEMPORAL DO ESTADO DA BAHIA



Fonte: Autoria própria, 2019.

APÊNDICE B – Protocolo de revisão sistemática da literatura sobre séries temporais para a previsão de internações hospitalares por dengue

1 INTRODUÇÃO

Febre amarela, dengue, *chikungunya*, *zika* e microcefalia são doenças causadas por um único vetor, o mosquito *Aedes aegypti*, considerando pelo Ministro da Saúde em 2017 como o maior desafio de saúde pública naquele ano. Dentre as doenças citadas, a mais recorrente é a dengue.

Discute-se na literatura que os esforços empregados no controle do vetor devem ir além do uso de simples comunicação vertical, abrangendo os diversos setores da sociedade (governo e sociedade civil) envolvendo políticas públicas desenvolvidas intersetorialmente de forma integrada e com financiamento, para eliminar os locais onde o mosquito possa se desenvolver e evitar custos com o tratamento de pacientes contagiados com dengue e/ou outras doenças.

Quando o combate ao vetor transmissor da dengue não é eficiente, as pessoas infectadas podem ser internadas em hospitais públicos e/ou privados. O Sistema Único de Saúde (SUS) brasileiro mantém, por meio de seu Departamento de Informática (DATASUS), o registro público de todas as internações ocorridas em Hospitais Públicos e privados conveniados ao SUS. Esses dados têm origem nas Autorizações de Internações de Hospitalares (AIH), destinadas ao pagamento das internações em hospitais públicos e privados conveniados ao SUS, o que lhe confere confiabilidade para produção de conhecimento científico sobre saúde coletiva (BITTENCOURT; CAMACHO; LEAL, 2006).

O desafio de combate ao mosquito *Aedes aegypti* tem instigado diversos estudos que faz uso da Tecnologia Digital da Informação e Comunicação (TDIC) no sentido de contribuir para seu controle.

Na cidade de Recife foi desenvolvido um projeto de análise de dados para implantação de *Data Mart* como ferramenta para tomada de decisão em combate aos vírus da dengue, *zika* e *chikungunya*. A solução proposta foi utilizar ferramentas e técnicas de *Business Intelligence* (BI) para análise e mapeamento de bases de dados em geral, com ênfase na saúde pública, para que através desta análise possa-se obter um sistema de apoio à tomada de decisão a favor do combate a doenças epidêmicas de Dengue, *Zika* e *Chikungunya* (CARVALHO *et al.*, 2017).

Outra solução, foi o desenvolvimento de ambiente *web* (DeuZikaChico) para o monitoramento e combate de epidemias de doenças causadas pelo mosquito em questão. DeuZikaChico faz uso dos comentários nas redes sociais, Informações Geográficas Ambientais e Geoprocessamento para cumprir o objetivo pretendido (OLIVEIRA *et al.*, 2016).

São diversas as soluções que usam TDIC e modelos estatísticos com o propósito de integrar o conjunto de esforços dedicados a controlar o desenvolvimento do mosquito transmissor da dengue e outras doenças, mas ainda existe uma lacuna no sentido da utilização de séries temporais para a previsão de internações hospitalares por dengue.

Para avaliar mais detalhadamente esta lacuna, o presente protocolo é o planejamento para a realização de Revisão Sistemática da Literatura para levantar o estado da arte identificando estudos sobre o uso de séries temporais aplicadas à predição e/ou previsão de internações hospitalares por dengue no contexto da saúde pública com foco no planejamento e gestão.

2 PROTOCOLO DE ESTUDO PARA REVISÃO SISTEMÁTICA

O protocolo faz parte do planejamento da Revisão Sistemática da Literatura (RSL). Sendo o primeiro de uma série de documentos a serem produzidos, ele visa orientar e documentar todos os procedimentos necessários à condução da RS que se consolida com a execução e a sumarização.

2.1 CONTEXTO

Compreender e ter conhecimentos da demanda e dos custos de internações hospitalares no SUS é relevante para os gestores da saúde, uma vez que os subsidia de informações para que sejam tomadas decisões mais coerentes com a realidade. É de fundamental importância que os hospitais utilizem técnicas para estimar custos com internações hospitalares (ROSA; CHANG; MATTOS, 2016).

A previsão de internações hospitalares por dengue utilizando séries temporais pode permitir a estimativa de custos e as necessidades de outros recursos essenciais à manutenção dos leitos hospitalares em quantidade e condições adequadas à internação dos pacientes por dengue. É a oportunidade para o gestor da unidade hospitalar conhecer com antecedência sua demanda futura.

2.2 OBJETIVO E QUESTÕES DE PESQUISA

2.2.1 Objetivo

Analisar o estado da arte, através de Revisão Sistemática da Literatura, **com o propósito** de identificar estudos sobre séries temporais **com relação a** previsão de internações hospitalares e/ou incidência de dengue **do ponto de vista** dos pesquisadores **no contexto** da saúde coletiva **com foco**, no planejamento e gestão.

2.2.2 Questões de pesquisa

2.2.2.1 Questões primárias

Quais técnicas de séries temporais têm sido pesquisadas para criar modelos de previsão de internações hospitalares?

2.2.2.2 Questões secundárias

Dentre as técnicas de previsão de séries temporais encontradas, quais estão relacionadas com a previsão de incidência de dengue?

Dentre as técnicas de previsão de séries temporais encontradas, quais estão relacionadas com previsão de internações por dengue?

Dentre as técnicas de previsão de séries temporais encontradas, quais estão relacionadas com previsão de internações por doenças diversas?

Dentre as técnicas de previsão de séries temporais encontradas, quais são mais indicadas para a previsão de internações hospitalares e/ou incidência de dengue?

2.3 ESCOPO DA PESQUISA

Para tornar a pesquisa viável no que diz respeito ao custo, esforço e tempo empregados na realização do acesso aos dados que abrangem o estudo, serão realizadas consultas em bibliotecas digitais através de seus mecanismos de buscas. Na indisponibilidade de estudos em sua forma digital, serão realizadas consultas manuais, quando possível.

2.3.1 Critérios adotados para seleção das fontes

Serão consideradas como fontes das publicações, as bibliotecas digitais disponíveis online, bases de indexação, periódicos, eventos da área e consulta a especialistas. As fontes precisam ter vinculação com o portal de periódicos da CAPES para tornar mais acessível o seu conteúdo.

As bibliotecas digitais e bases de indexação precisam atender aos seguintes requisitos:

- Permitir busca por expressões lógicas ou equivalentes;
- Incluir em sua base publicações da área de séries temporais que possuam relação direta com a temática pesquisada e/ou similar;
- Permitir consulta no texto completo das publicações indexadas.

É importante salientar que os resultados devem ser únicos quando a busca utilizar o mesmo conjunto de palavras-chaves. Se a unicidade dos resultados não for garantida, então será preciso registrar estratégia para minimizar os efeitos colaterais das limitações encontradas.

2.3.2 Restrições

A pesquisa está restrita ao período de 01 de janeiro de 2013 até dia 31 de dezembro de 2018, levando em consideração, exclusivamente, as fontes selecionadas segundo os critérios estabelecidos.

2.4 IDIOMAS

Os idiomas selecionados serão Português e Inglês. A escolha da língua inglesa se justifica pela sua adoção como forma de publicação em conferências e periódicos internacionais relevantes e correlacionados com o tema de pesquisa, como também, publicações nacionais. O idioma português será escolhido em virtude da sua adoção em periódicos e conferências nacionais na área de saúde.

2.5 MÉTODOS DE BUSCA DAS PUBLICAÇÕES

Serão planejadas e utilizadas expressões de busca na pesquisa por fontes bibliográficas via Internet (modo digital) para a obtenção de versões digitais das publicações e, quando disponíveis e relevantes, serão selecionadas fontes não-digitais (modo manual), as quais serão analisadas manualmente de acordo com os critérios estabelecidos pela expressão de busca.

2.5.1 Expressão geral de busca

Para a definição das palavras-chave que compõem a expressão de busca é importante escolher termos que sejam descritores nas bases de dados utilizados para indexar as publicações. Após a definição do tema, faz-se necessário realizar pesquisa sobre a relação das palavras-chave escolhidas com os termos de indexação utilizados pelas bases onde estão os estudos publicados. Na área da saúde existe uma ferramenta online chamada DeCS que apresenta descritores em Ciências da Saúde. Ela foi criada para servir como referência na indexação de artigos de revistas científicas, livros, anais de congressos, relatórios técnicos, e outros. Ainda na área da saúde existe o *Medical Subject Headings* (MeSH) que tem como propósito fornecer uma terminologia organizada para indexação e catalogação de informações biomédicas.

Para pesquisa por publicações em português deve-se utilizar a seguinte expressão de busca:

- “séries temporais” AND (predição OR previsão) AND (“internações hospitalares” OR hospitalizações) AND dengue.

Para pesquisa em inglês deve-se utilizar a expressão de busca que segue:

- “*time series*” AND (*prediction OR predict OR predicting OR forecast OR forecasting*) AND (“*hospital admissions*” OR *hospitalizations*) AND *dengue*.

2.5.2 Busca manual

Nas consultas de forma manual, deve-se usar como critério de seleção, as palavras-chave presentes na expressão geral de busca, primeiramente nos títulos, resumos (*abstracts*) e palavras-chave dos artigos. Para artigos escritos na língua portuguesa e que apresentem o resumo em inglês, *abstract*, deve-se primeiro pesquisar no *abstract*, em caso de dúvida sobre a seleção do artigo, deve-se pesquisar nas palavras-chave, resumo e na introdução.

2.6 PROCEDIMENTOS PARA SELEÇÃO E CRITÉRIOS

A definição da estratégia de busca é fundamental para o sucesso e imparcialidade na seleção das fontes necessárias ao desenvolvimento da Revisão Sistemática. Nesse sentido, será realizada uma pesquisa (com base na expressão de busca) por um pesquisador da área para identificar publicações com potencialidades para serem usadas.

A seleção das publicações será feita levando em consideração os critérios de inclusão, exclusão e qualidade que serão pré-estabelecidos. Deverá sempre acontecer um consenso entre os pesquisadores sobre as publicações selecionadas, mas diante de avaliações conflitantes sobre alguma publicação, a mesma será sempre incluída dentre os estudos selecionados.

Serão selecionadas publicações que apresentem métodos e técnicas de modelagem computacional com séries temporais votadas para a predição de internações hospitalares no âmbito da saúde com foco no planejamento e gestão.

2.6.1 Ferramentas computacionais de apoio

O processo de planejamento, execução e sumarização de uma Revisão Sistemática é sempre complexo ao passo que envolve o registro documental de todos os procedimentos executados. Nesse sentido, é de fundamental importância a adoção de ferramentas computacionais que auxiliem no gerenciamento da condução de uma Revisão Sistemática.

Para o presente estudo será adotada a ferramenta de gerenciamento de RS StArt que foi desenvolvida pelo Laboratório de Pesquisa em Engenharia de Software (LaPES) da Universidade Federal de São Carlos (UfSCar). O propósito da ferramenta é dar suporte ao pesquisador, apoiando a aplicação dos passos envolvidos na condução da RS. A StArt tem sido usada por alunos de pós-graduação que têm relatado o apoio positivo de seu uso e as vantagens em relação a outras ferramentas.

2.6.2 Procedimentos de seleção

Para selecionar as publicações relacionadas ao objetivo e as questões de estudo, serão executadas as etapas que seguem.

2.6.4.1 Seleção e catalogação preliminar das publicações

Será executada a expressão de busca na base de indexação (Portal de Periódicos da CAPES). A busca irá identificar e retornar um conjunto de publicações sobre estudos com relação ao tema que serão divididas em pastas com o nome da respectiva base de dados de origem de cada publicação. A partir dessa primeira consulta será definida quais bases de indexação serão consideradas para busca das publicações utilizadas na presente RS.

Após a catalogação das publicações por bases de indexação, todas serão agrupadas em um repositório denominado de "*Unclassified Papers*". Em seguida, será executado o processo de seleção levando em consideração os critérios de inclusão e exclusão que irá classificar e agrupar as publicações nos seguintes repositórios: *Accepted Papers* (publicações que serão analisadas semanticamente), *Rejected Papers* (estudos que não têm relação com o tema de pesquisa) e *Duplicated Papers* (trabalhos duplicados).

2.6.2.2 Primeiro Filtro – Seleção das publicações relevantes

Os estudos serão selecionados para ser incluídos, isto é, apresentam relevância para o tema de estudo ou excluídos, ou seja, não apresentam relação com o tema de estudo. A seleção dos estudos e classificação como incluídos ou excluídos será feita levando em consideração a definição dos critérios de inclusão e exclusão.

Serão definidos e identificados os seguintes Critérios de Inclusão (CI):

- **CI-01** Publicações com séries temporais relacionadas a previsão de internações hospitalares por dengue;

- **CI-02** Publicações com séries temporais relacionadas a previsão de internações hospitalares diversas;
- **CI-02** Publicações que apresentam séries temporais na previsão de incidência de dengue. Serão definidos e identificados os seguintes Critérios de Exclusão (CE):
- **CE-01** Excluir publicações com séries temporais não relacionadas a previsão de internações hospitalares ou incidência de dengue;
- **CE-02** Não serão selecionadas publicações sem relação com as palavras-chave (podem ser considerados sinônimos);
- **CE-03** Não serão selecionadas publicações que não estejam escritas em inglês e/ou português;
- **CE-04** Se as palavras-chave da expressão de busca não constarem no título, resumo e/ou texto da publicação, a mesma será excluída (considerar apenas os elementos textuais);
- **CE-05** Não serão selecionados trabalhos publicados como artigos curtos ou pôsteres;
- **CE-06** Serão excluídos trabalhos que apresentam análises preditivas sem apresentar os métodos e/ou técnicas de séries temporais;
- **CE-07** Não serão selecionadas as publicações duplicadas;
- **CE-08** Serão excluídos os trabalhos publicados como revisão de literatura;
- **CE-09** Trabalho relacionado a inteligência artificial, *big data* ou *data mining*;
- **CE-10** Trabalho com previsão de internação ou incidência de dengue sem a aplicação de séries temporais.

Nesta etapa, será realizada uma análise sintática que visa selecionar os estudos em incluídos ou excluídos. Isso acontecerá pela comparação das palavras da expressão geral de busca com sua presença no título, *abstract* e *Keywords* de cada publicação. Os critérios de inclusão e exclusão poderão sofrer alterações à medida que o protocolo é testado e validado.

2.6.4.3 Segundo Filtro – Seleção das publicações a partir de critérios de qualidade

A aplicação da expressão de busca de acordo com o 1^o Filtro (título, *abstract* e *Keywords*) para a seleção preliminar não é suficiente para tornar as publicações selecionadas adequadas ao contexto da pesquisa, uma vez que somente avalia aspectos sintáticos de cada estudo. É preciso analisar, também, os estudos encontrados nos aspectos semânticos através da leitura do resumo e/ou abstracts e classificá-los de acordo com os critérios de inclusão e exclusão.

Na análise semântica, a classificação por critérios de inclusão e exclusão é eficiente, pois a leitura do resumo e/ou *abstract* acontece apenas uma vez. Como os pesquisadores sempre devem

estar em consenso sobre a seleção dos estudos, em caso de divergência, deve-se criar um novo critério ou mesclar critérios semelhantes.

2.6.4.2 Terceiro Filtro – Seleção dos dados relevantes

Após a aplicação do 1º e 2º filtros, é preciso refinar ainda mais a seleção para especializar o universo de fontes pesquisadas. Neste sentido, as publicações que não respeitarem os critérios abaixo serão Excluídas por Irrelevância (EI):

- Apresentar técnicas de séries temporais para previsão de internações hospitalares (STPI);
- Descrever técnicas de séries temporais para previsão de internações hospitalares por dengue (STPID).

A exclusão por irrelevância será marcada como EI+STPI ou EI+STPID. Esses critérios estão sujeitos às alterações mediante o teste e validação do protocolo.

2.6.3 Outros critérios de inclusão

Considerar na inclusão as publicações que foram citadas nas referências bibliográficas dos estudos selecionados e, sendo relevantes, não foram identificadas pelas palavras-chave do estudo utilizadas na composição da expressão geral de busca.

2.7 PROCEDIMENTO PARA EXTRAÇÃO DE DADOS

Os dados das publicações precisam ser extraídos e para isso será necessário definir critérios e procedimentos com o intuito de sistematizar o processo.

2.7.1 Seleção e Catalogação preliminar dos dados coletados

Criados os repositórios: *Unclassified Papers*, *Accepted Papers*, *Rejected Papers* e *Duplicated Papers*, as referências completas selecionadas a partir da fonte consultada serão armazenadas no repositório de dados indicado de acordo com os critérios de seleção.

2.7.2 Seleção dos dados relevantes

As referências serão analisadas e submetidas aos critérios de seleção estabelecidos. Quando atender aos critérios de seleção deverão ser marcadas como “incluído pelo critério

[sigla/número do filtro]”, quando não atender, deverão ser marcadas como “excluída pelo critério [sigla/número do filtro]”.

2.7.3 Extração dos dados

Será criado um banco de dados, onde para cada publicação relevante à pesquisa serão extraídos os dados de acordo com a Tabela 1.

Quando não existir resposta para algum dos campos de extração de dados indicados na Tabela 1, será utilizada a sigla NA (Não se Aplica).

2.7.4 Sumarização dos resultados

Será produzido relatório da Revisão Sistemática com uma análise qualitativa sobre a utilização de séries temporais para a previsão de internações hospitalares. Constará no relatório:

Tabela 1 – Formulário de extração de dados

DADOS A SER EXTRAÍDOS	DESCRIÇÃO
Título	Título do estudo
Autor	Autor (es) da publicação
Ano	Ano de publicação
Veículo de publicação	Onde o estudo foi publicado
<i>Keywords</i>	Palavras utilizadas para sintetizar e/ou descrever determinada publicação
Siglas importantes	Abreviações e/ou siglas que são de relevância significativa dentro do estudo.
Objetivo do estudo	Qual o Objetivo Proposto?
Técnicas e/ou Abordagens e/ou métodos	Quais as técnicas, abordagens, ou métodos são empregados no estudo.
Descrição da técnica	Resumo detalhado da técnica, abordagem ou método proposto pelo estudo
Tipo de pesquisa	A pesquisa descrita no estudo pode ser teórica ou empírica
Contextualização	Qual o contexto abordado? (Dengue ou outros casos epidemiológicos)

DADOS A SER EXTRAÍDOS	DESCRIÇÃO
Implementação	Recursos computacionais utilizados para implementar a técnica, abordagem ou método proposto pelo estudo.
Variáveis	Quantas variáveis foram utilizadas e de qual natureza?
Fonte dos dados	Origem dos dados que foram utilizados no estudo
Preparação dos dados	Técnicas de preparação e/ou pré-processamento dos dados
Resultados	Conclusões do estudo
Limitações	As limitações encontradas na aplicação da técnica, abordagem ou método
Métricas de desempenho	Métricas utilizadas para avaliação do modelo e seu desempenho
Oportunidades de pesquisa	Indicação de trabalhos futuros

Fonte: Autoria própria, 2018.

Tabelas, gráficos e outras formas de representação, que permitam ilustrar quantitativamente a análise dos estudos avaliados. Os resultados serão tabulados, sem a realização de meta-análise.

2.8 PROCEDIMENTOS PARA ANÁLISE

A análise será quantitativa e qualitativa, identificando a quantidade de publicações relacionadas com o tema de pesquisa e o relacionamento entre elas de acordo com a abrangência de cada uma.

2.8.1 Análise quantitativa

Na análise quantitativa serão considerados dados extraídos das publicações segundo os critérios:

- Número de publicações selecionadas para fazer parte do estudo;
- Quantidade de métodos de séries temporais encontrados que se relacionam com a questão de pesquisa;
- Quantidade de métodos de séries temporais mais indicados para a previsão de internações hospitalares.

2.8.2 Análise qualitativa

A análise qualitativa será realizada para discutir os dados quantitativos, com o intuito de relacionar as publicações selecionadas com as questões de pesquisa declaradas.

3 PLANEJAMENTO E EXECUÇÃO DA REVISÃO SISTEMÁTICA

O capítulo anterior apresentou o protocolo onde constam os procedimentos e critérios utilizados para a condução da RS. Ele deve ser o guia para a execução e consolidação da Revisão Sistemática. Para o planejamento e elaboração do protocolo e condução do estudo seguiu-se o processo definido em Montoni (2007), que sugere as seguintes atividades:

- 1 Realizar Prospecção sobre o tema de interesse;
- 2 Definir protocolo;
- 3 Testar e avaliar o protocolo;
- 4 Executar a pesquisa;
- 5 Analisar os resultados da pesquisa;
- 6 Empacotar e publicar os resultados.

É de fundamental importância que esse processo seja seguido e que seus resultados sejam documentados com o objetivo de permitir auditoria.

3.1 DEFINIÇÃO DO ESCOPO E ESTUDOS PRELIMINARES

Definido o tema de estudo, o primeiro passo foi encontrar as palavras-chave adequadas para compor a expressão de busca a ser utilizada nas fontes das publicações. Foram selecionados os termos: séries temporais, previsão, predição, internação hospitalar, hospitalizações e dengue.

Para tornar mais efetiva a busca nas bases de publicações, realizou-se uma pesquisa na ferramenta online, DeCS, com os termos supramencionados. Os resultados para os descritores relacionados com as palavras-chave estão apresentados em Inglês, Português e Espanhol. Como para a RS em questão interessa apenas Inglês e Português, a Tabela 2 apresenta as correspondências para esses dois idiomas.

Tabela 2 – Associação entre palavras-chave e descritores

DESCRITOR(ES) PORTUGUÊS	DESCRITOR(ES) INGLÊS
Previsão e/ou predição	<i>Forecasting</i>
Séries temporais	<i>Time Series Studies</i>
Hospital	<i>Hospital</i>
Dengue	<i>Dengue</i>

Fonte: Autoria própria, 2018.

Da leitura de estudos, identificou-se também como palavras chave: *hospital admission* e *hospitalized* (para internações hospitalares) e *predict* (para previsão ou predição). Ainda sobre

as palavras-chave e sua relação com os descritores, foi utilizado o recurso "PubMed Advanced Search Builder" da base de indexação PubMed para identificar a quantidade de ocorrências por descritor. A Tabela 3 mostra o resultado.

Tabela 3 – Termo indexado e quantidade de ocorrências

TERMO INDEXADO	FREQUÊNCIA
Forecasting	84458
Time series	26620
Hospitalization	173041
Predict	298664
Hospital admission	20687
Hospitalized	90477
Dengue	19420

Fonte: Autoria própria, 2018.

A frequência expressiva dos termos de indexação ou descritores mostrados na Tabela 3 é importante pois mostra que os mesmos são relevantes para os estudos que são desenvolvidos.

Definidos os termos de busca, foi realizada uma pesquisa sobre o tema: modelos para previsão de internações hospitalares, como estavam sendo encontrados muitos resultados que mostravam análise preditivas utilizando diferentes modelos estatísticos, decidiu-se filtrar mais o tema e optar por levantar o estado da arte que tivesse relação com a previsão de internações hospitalares e incidência de dengue utilizando técnicas de séries temporais.

3.2 DEFINIÇÃO DAS PALAVRAS-CHAVE E CALIBRAGEM DA EXPRESSÃO DE BUSCA

Utilizando a expressão de busca: ***"time series"AND (prediction OR predict OR predicting OR forecast OR forecasting) AND ("hospital admissions"OR hospitalizations) AND dengue***, foi realizada consulta no Portal de Periódicos da CAPES considerando o período de 2013 a 2018. A consulta retornou 95 resultados.

Os resultados obtidos com a busca, mesmo considerando a relevância e abrangência das palavras-chave, apresentaram uma tendência de poucas publicações que teriam alguma relação ainda não comprovadas com o tema de pesquisa. A dificuldade poderia estar na definição das palavras-chave e/ou na definição do tema de estudo.

Fez-se necessário realizar processo para calibrar a expressão de busca envolvendo a definição da base de dados de busca para testes de protocolo, identificação de publicações que

deveriam compor o grupo de controle, identificação de expressão de busca inicial, testes da expressão de busca e análise dos resultados retornados pela expressão de busca.

Como resultado de expressão de busca calibrada obteve-se: "*time series*"**AND** (*predict****OR forecast***) **AND hospital*** **AND dengue**. A utilização de caracteres curingas (*) fazem com que sejam obtidas variações das palavras marcadas com os mesmos. É importante salientar que a base de indexação precisa suportar esse recurso. Com a expressão de busca calibrada, obtiveram-se 462 publicações como resultado de consulta no Portal de Periódicos da CAPES.

3.3 IDENTIFICAÇÃO DE PUBLICAÇÕES DE CONTROLE E PALAVRAS-CHAVE

Após ter-se definido o tema do estudo, foi feita uma nova rodada de testes da expressão de busca. Desta vez foi utilizada a base de busca da *PubMed*. A escolha das palavras-chave esteve condicionada às publicações de controle selecionadas. Uma lista inicial destas publicações foi identificada durante os testes dos estudos iniciais, outras foram adicionadas durante os testes com as expressões de busca à medida que novos estudos foram retornados.

O conjunto de publicações de controle surgiu por meio da calibragem da expressão de busca. Para os casos em que muitas publicações foram retornadas foi necessário:

- Verificar se a estratégia de busca estava retornando estudos essenciais ou se outros requisitos não inerentes ao estudo estão sendo considerados pela estratégia;
- Verificar se não há palavras-chave demais, desnecessárias;
- Verificar se a máquina de busca era eficiente;
- Não utilizar a inclusão de muitas cláusulas restritivas (AND, NOT, OR, etc.), para não limitar o escopo da busca desnecessariamente;
- Revisar a lista de publicações através da aplicação dos filtros presentes no protocolo.

Dessa forma, o processo para calibragem da expressão de busca envolveu:

- 1 Definição da base de busca para testes do protocolo;
- 2 Identificação de publicações que deveriam compor o grupo de controle;
- 3 Identificação da expressão de busca inicial;
- 4 Testes da expressão de busca;
- 5 Análise dos resultados retornados pela expressão de busca.

Esse processo foi feito de forma iterativa, com os passos 2 a 5 sendo executados continuamente até que o resultado fosse considerado satisfatório.

3.3.1 Primeira rodada

Foi executada pesquisa na base *PubMed* utilizando a expressão de busca: (((**time series**[Abstract]) AND **predict***[Abstract]) AND **hospital***[Abstract]) AND **dengue**[Abstract] AND (('2013/01/01"[PDat] : "2018/12/31"[PDat])), a qual retornou 10 resultados dos quais 1 faz parte do grupo de controle.

A expressão de busca: (((("Abstract":time series) AND "Abstract":predict*) AND "Abstract":hospital) AND "Abstract":dengue), foi executada na base de indexação *IEEE Xplore*, mas não retornou resultados mesmo sem informar período de tempo.

O portal de periódicos da CAPES não permite consulta apenas no abstract, mas tudo indica que não existem muitas publicações sobre o tema de pesquisa, considerando título e resumo. A estratégia pensada para contornar as limitações foi estender a expressão de busca para todos os elementos de cada publicação (*Title, Abstract, keywords, full text*) e aplicá-la novamente nas bases de indexação.

3.3.2 Segunda rodada

Como a estratégia de busca utilizada na Primeira rodada não obteve muito sucesso, decidiu-se buscar todos os estudos que estivessem relacionados com a expressão de busca: "**time series**"AND **predict*** AND **dengue** AND **hospital***, a qual leva em consideração o título, o resumo, as palavras-chave e o próprio conteúdo dos estudos.

Foi executada, no portal de periódicos da CAPES, a expressão de busca: "**time series**"AND **predict*** AND **dengue** AND **hospital***. A busca foi realizada no dia 27 de novembro de 2018, considerando o período de 2013 até 2018. A busca foi realizada utilizando a opção "Busca por assunto". Foram encontradas 465 publicações. Das publicações encontradas, identificaram-se 4 que integraram o grupo de controle.

3.3.3 Terceira rodada

É importante salientar que a expressão de busca foi executada na máquina de busca "*Compendex (Engineering Village)*", considerando o período de 2013 a 2018. A consulta foi realizada no dia 3 de dezembro de 2018 e retornou apenas 3 resultados, mas sem relação direta com o objetivo do estudo. Também foi considerada a base *ScienceDirect (Elsevier)*, a qual recebeu a expressão de busca e retornou 92 resultados, mas apenas 1 apresentou relação direta com o objetivo de pesquisa.

Como os termos da expressão de busca são fundamentais para o estudo em questão e a busca realizada na Segunda rodada retornou um conjunto satisfatório de publicações, decidiu-se

realizar um último ajuste da *String* de busca para: **"time series"AND (predict* OR forecast*) AND hospital* AND dengue** e aplicá-la nas bases de indexação. Depois disso, a *String* de busca não recebeu mais refinamentos.

3.4 DEFINIÇÃO DAS MÁQUINAS DE BUSCA

Na realização de testes com o protocolo, a base que retornou mais resultados foi o Portal de Periódicos da CAPES, uma vez que indexa várias bases de publicações. Para garantir maior cobertura dos resultados, decidiu-se utilizar também a base *PubMed* pela sua importância para a área da saúde, a *IEEE Xplore* e *Scopus*. Considerou-se que estas quatro bases de busca seriam suficientes, pois indexam a maioria do estado da arte relevante no contexto deste estudo.

3.4.1 Portal de periódicos da CAPES

A expressão de busca utilizada foi: **"time series"AND (predict* OR forecast*) AND hospital* AND dengue**. A busca foi realizada no dia 16 de janeiro de 2019, considerando o período de 2013 até 2018. A busca foi realizada utilizando a opção "Busca por assunto". Foram encontradas 462 publicações, as quais foram adicionadas em "Meu espaço", dentro do Portal de Periódicos da CAPES, para em seguida serem exportadas como BibTex. Foi preciso exportar em 16 arquivos, uma vez que cada exportação somente comporta 30 publicações.

3.4.2 IEEE Xplore

A expressão de busca utilizada foi: **"time series"AND (predict* OR forecast*) AND hospital* AND dengue**. A busca foi realizada aplicando a *String* de busca na aba "*Command Search*", marcando a opção "*Full Text & Metadata*". A busca foi realizada no dia 16 de janeiro de 2019, considerando o período de 2013 até 2018. Foram encontradas 16 publicações. As publicações foram exportadas em formato RIS.

3.4.3 PubMed/MEDLINE

A expressão de busca utilizada foi: **"time series"AND (predict* OR forecast*) AND hospital* AND dengue AND (("2013/01/01"[PDat]: "2018/12/31"[PDat]))**. A busca foi realizada no dia 15 de janeiro de 2019, considerando o período de 01/01/2013 até 31/12/2018. A

busca foi realizada utilizando no espaço principal de busca. Foram encontradas 310 publicações, as quais foram exportadas no formato MEDLINE.

3.4.4 Scopus

A expressão de busca utilizada foi: **"time series"AND (predict* OR forecast*) AND hospital* AND dengue**. A busca foi realizada no dia 17 de janeiro de 2019, considerando o período de 01/01/2013 até 31/12/2018. A busca foi realizada utilizando no espaço principal de busca. Foram encontradas 356 publicações, as quais foram exportadas no formato RIS.

3.4.5 Instrumento para consulta manual

Na consulta manual, deve-se procurar as palavras-chave constantes na expressão geral de busca (Português e Inglês). A busca deve levar em consideração títulos/titles e resumos/abstracts dos estudos e registrar os dados preenchendo o formulário conforme Tabela 4.

Tabela 4 – Formulário para busca manual

ARTIGO	ANO	LÍNGUA	Nº PALAVRAS CHAVE	RESULTADO

Fonte: Autoria própria, 2018.

Quando as palavras-chave não existirem na publicação a mesma será incluída no formulário de consulta manual com o resultado marcado como excluído e informar o critério de exclusão.

3.5 IDENTIFICAÇÃO DO PERÍODO DE BUSCA

Não se encontraram muitas publicações sobre o uso de séries temporais para a previsão de internações hospitalares e/ou incidência de dengue, então considerou-se como período de busca janeiro de 2013 até dezembro de 2018. Outra estratégia para a definição do período de busca seria identificar a publicação do grupo de controle mais antiga e então realizar as buscas a partir dessa data. Essa estratégia é eficaz à medida que poucas publicações são encontradas considerando períodos mais recentes.

REFERÊNCIAS

BITTENCOURT, S. A.; CAMACHO, L. A. B.; LEAL, M. d. C. O sistema de informação hospitalar e sua aplicação na saúde coletiva. *Cadernos de Saúde Pública*, SciELO Public Health, v. 22, p. 19–30, 2006. Citado na página 5.

CARVALHO, N. M. *et al.* Projeto de análise de dados para implantação de data mart como ferramenta para tomada de decisão em combate aos vírus da dengue, zika e chikungunya. *Revista InterScientia*, v. 5, n. 2, p. 106–123, 2017. Citado na página 5.

OLIVEIRA, A. L. F. A. *et al.* **Deuzikachico**: o poder da agi no monitoramento e combate de epidemias como a de dengue, zika e chikungunya. 2016. Citado na página 5.

ROSA, N. M.; CHANG, J. J.; MATTOS, C. A. Aplicação do método híbrido ARIMA-RNA para a predição dos custos de internação pelo sistema único de saúde na cidade de São Paulo.

XXXVI ENCONTRO NACIONAL DE ENGENHARIA DE PRODUÇÃO, 2016.

Disponível em:

www.abepro.org.br/biblioteca/TN_STO_228_331_29001.pdf. Acesso em: 26 jun. 2018.

Citado na página 7.