

UNIVERSIDADE FEDERAL DE SANTA MARIA  
CENTRO DE TECNOLOGIA  
CURSO DE GRADUAÇÃO EM ENGENHARIA DE CONTROLE E  
AUTOMAÇÃO

Raul Santos Ferreira

**PREDIÇÃO SEMANAL DE NÚMERO DE CASOS DE DENGUE EM  
PORTO ALEGRE POR MODELOS DE SÉRIES TEMPORAIS**

Santa Maria, RS  
2024

**Raul Santos Ferreira**

**PREDIÇÃO SEMANAL DE NÚMERO DE CASOS DE DENGUE EM PORTO  
ALEGRE POR MODELOS DE SÉRIES TEMPORAIS**

Trabalho de Conclusão de Curso apresentado ao Curso de Graduação em Engenharia de Controle e Automação da Universidade Federal de Santa Maria (UFSM, RS), como requisito parcial para obtenção do grau de **Bacharel em Engenharia de Controle e Automação**.

ORIENTADOR: Prof. Daniel Welfer

Santa Maria, RS  
2024

**Raul Santos Ferreira**

**PREDIÇÃO SEMANAL DE NÚMERO DE CASOS DE DENGUE EM PORTO  
ALEGRE POR MODELOS DE SÉRIES TEMPORAIS**

Trabalho de Conclusão de Curso apresentado ao Curso de Graduação em Engenharia de Computação da Universidade Federal de Santa Maria (UFSM, RS), como requisito parcial para obtenção do grau de **Bacharel em Engenharia de Computação**.

**Aprovado em 07 de 06 de 2024:**

---

**Daniel Welfer, Dr. (UFSM)**  
(Presidente/Orientador)

Santa Maria, RS  
2024

## **AGRADECIMENTOS**

*Aos meus pais e ao meu irmão, por todo amor, apoio e por serem meu maior exemplo. Obrigado terem passado os mais valiosos ensinamentos que pude ter. Também agradeço a Deus por ter me sustentado até aqui.*

## RESUMO

### PREDIÇÃO SEMANAL DE NÚMERO DE CASOS DE DENGUE EM PORTO ALEGRE POR MODELOS DE SÉRIES TEMPORAIS

AUTOR: Raul Santos Ferreira

ORIENTADOR: Daniel Welfer

Este Trabalho desenvolve e avalia modelos de séries temporais para a predição semanal de casos de dengue na cidade de Porto Alegre, com o objetivo de oferecer ferramentas eficazes para a vigilância epidemiológica e o planejamento de medidas preventivas. Utilizando dados de saúde pública e registros meteorológicos específicos da região, modelos estatísticos como SARIMA (Seasonal Autoregressive Integrated Moving Average) e SARIMAX (Seasonal Autoregressive Integrated Moving Average with Exogenous Variables) foram aplicados e comparados em termos de sua eficiência preditiva. O estudo destacou que a inclusão de variáveis climáticas, como temperatura e precipitação, nos modelos SARIMAX contribuiu para aumentar a precisão das previsões, adaptando-se às dinâmicas de transmissão da dengue, que são fortemente influenciadas por condições ambientais. O trabalho também enfatiza a importância da análise de séries temporais na saúde pública, principalmente para doenças com forte componente sazonal e dependência de condições climáticas, como a dengue.

**Palavras-chave:** Dengue. Predição. Porto Alegre. Séries Temporais. SARIMA. SARIMAX. Saúde Pública.

## ABSTRACT

### MOBILE APPLICATION FOR CLASSIFICATION OF TYPES OF SKIN LESION USING CONVOLUTIONAL NEURAL NETWORKS

AUTHOR: Raul Santos Ferreira

ADVISOR: Daniel Welfer

This Work develops and evaluates time series models for the weekly prediction of dengue cases in the city of Porto Alegre, with the aim of offering effective tools for epidemiological surveillance and planning preventive measures. Using public health data and region-specific meteorological records, statistical models such as SARIMA (Seasonal Autoregressive Integrated Moving Average) and SARIMAX (Seasonal Autoregressive Integrated Moving Average with Exogenous Variables) were applied and compared in terms of their predictive efficiency. The study highlighted that the inclusion of climatic variables, such as temperature and precipitation, in the SARIMAX models contributes to increasing the accuracy of predictions, adapting to dengue transmission dynamics, which are strongly influenced by environmental conditions. The work also emphasizes the importance of time series analysis in public health, especially for diseases with a strong seasonal component and dependence on weather conditions, such as dengue.

**Keywords:** Dengue. Forecasting. Porto Alegre. Time Series Models. SARIMA. SARIMAX. Climatic Analysis. Public Health.

## LISTA DE FIGURAS

Figura 1- Histórico dengue Américas .....	12
Figura 2 - Histórico dengue RS .....	13
Figura 3 - Ovo do aedes aegypti .....	20
Figura 4 - Larvas do aedes aegypti .....	20
Figura 5 - Pupas do aedes aegypti .....	21
Figura 6 - Adulto do aedes aegypti.....	21
Figura 7 - Clico completo de vida do aedes aegypti.....	22
Figura 8 - Fluxograma de Box e Jenkins .....	37
Figura 9 - "Head" do dataset meteorológico.....	50
Figura 10 - "Head" do dataset final .....	52
Figura 11 - Série temporal transformada.....	56
Figura 12 - Gráfico de autocorrelação .....	57
Figura 13 - Gráfico de autocorrelação parcial .....	57
Figura 14 - Gráfico de resíduos SARIMA.....	60
Figura 15 - Gráfico Q-Q dos resíduos SARIMA.....	60
Figura 16 - Gráfico de probabilidade normal dos resíduos SARIMA .....	61
Figura 17 - Gráficos de valores-p SARIMA.....	61
Figura 18 - Gráfico de resíduos SARIMA.....	62
Figura 19 - Gráfico de probabilidade normal de resíduos SARIMA.....	62
Figura 20 - Gráfico Q-Q dos resíduos SARIMA.....	63
Figura 21- Gráfico de resíduos ACF .....	63
Figura 22 - Gráfico PACF de resíduos SARIMA.....	64
Figura 23- Análise dos resíduos quadrados SARIMA.....	64
Figura 24 - Séries temporais meteorológicas e de casos de dengue .....	65
Figura 25 - Mapa de calor de correlações.....	66
Figura 26 - Histograma de resíduos SARIMAX .....	70
Figura 27 - Gráfico de autocorrelação de resíduos SARIMAX.....	70
Figura 28 - Gráfico de probabilidade normal de resíduos SARIMAX.....	71
Figura 29- Gráfico de valores-p SARIMAX .....	71
Figura 30- Gráficos de resíduos SARIMAX .....	72
Figura 31- Gráfico de resíduos X valores preditos SARIMAX.....	72
Figura 32- Gráfico Q-Q de resíduos SARIMAX.....	73
Figura 33- Gráfico ACF dos resíduos SARIMAX .....	73
Figura 34- Gráfico PACF de resíduos SARIMAX.....	74
Figura 35- Gráficos de resíduos quadrados SARIMAX.....	74
Figura 36- Previsão de casos primeiro período SARIMA .....	76
Figura 37- Primeiro período de teste de previsões de casos SARIMA.....	77
Figura 38- Previsão de casos segundo período SARIMA .....	80
Figura 39- Segundo período de teste de previsões de casos SARIMA.....	81
Figura 40- Previsão de casos primeiro período SARIMAX .....	84
Figura 41- Primeiro período de teste de previsões de casos SARIMAX.....	85
Figura 42- Previsão de casos segundo período SARIMAX .....	88
Figura 43- Segundo período de teste de previsões de casos SARIMAX.....	88

## LISTA DE TABELAS

Tabela 1- Pesquisa na base de dados PubMed.....	38
Tabela 2 - Pesquisa na base de dados Wiley .....	39
Tabela 3- Pesquisa na base de dados ProQuest .....	39
Tabela 4- Pesquisa na base de dados IEEE .....	39
Tabela 5 - Pesquisa na base de dados ACM .....	40
Tabela 6 - Estrutura do dataset da dengue .....	46
Tabela 7 - "Head" do dataset final de dengue.....	48
Tabela 8 - Estrutura do dataset meteorológicos.....	49
Tabela 9 - Estrutura do dataset final .....	51
Tabela 10 - Estatística descritiva do dataset final.....	52
Tabela 11 - Melhores modelos SARIMA .....	59
Tabela 12 - Melhores modelos SARIMAX.....	69
Tabela 13- Comparativo valores preditos X casos confirmados SARIMA primeiro período .....	77
Tabela 14- Medidas de desempenho primeiro período SARIMA .....	79
Tabela 15- Comparativo valores preditos X casos confirmados SARIMA segundo período .....	81
Tabela 16 - Medidas de desempenho segundo período SARIMA.....	83
Tabela 17- Comparativo valores preditos X casos confirmados SARIMAX primeiro período .....	85
Tabela 18- Previsão de casos primeiro período SARIMAX.....	87
Tabela 19 - Comparativo valores preditos X casos confirmados SARIMAX segundo período.....	89
Tabela 20- Medidas de desempenho segundo período SARIMAX.....	90

## LISTA DE ABREVIATURAS E SIGLAS

*ACF* Função De Autocorrelação

*AIC* Critério De Informação De Akaike

*BIC* Critério De Informação Bayesiano

*FHD* Febre Hemorrágica Da Dengue

*PACF* Função De Autocorrelação Parcial

*PAHO* Pan American Health Organization

*INMET* Instituto Nacional De Meteorologia

*MAE* Erro Absoluto Médio

*MSE* Erro Quadrático Médio

*MAPE* Erro Percentual Médio Absoluto

*SCD* Síndrome Do Choque Da Dengue

*SARIMA* Seasonal Autoregressive Integrated Moving Average

*SARIMAX* Seasonal Autoregressive Integrated Moving Average with Exogenous Factors

*SINAN* Sistema De Informação De Agravos De Notificação

## SUMÁRIO

<b>1.</b>	<b>INTRODUÇÃO</b>	<b>12</b>
1.1	OBJETIVO	14
1.1.1	Objetivo Geral	15
1.1.2	Objetivos Específicos	15
1.1.3	Motivação e Justificativa	15
1.2	ORGANIZAÇÃO DO TRABALHO	16
<b>2</b>	<b>FUNDAMENTAÇÃO TEÓRICA</b>	<b>18</b>
2.1	DENGUE: VETORES, VIRUS E DISSEMINAÇÃO	18
2.1.1	Origens e Comportamento dos Vetores	18
2.1.2	Vírus da Dengue e seus Sorotipos	19
2.1.3	Ciclo de Vida do <i>Aedes aegypti</i>	19
2.1.4	Impacto da temperatura, umidade e precipitação no ciclo de vida do <i>Aedes aegypti</i>	22
2.2	SERIES TEMPORAIS	25
2.2.1	Fundamentos estatísticos em análise de séries temporais	26
2.2.2	Modelo SARIMA e SARIMAX	30
2.2.3	Componente autoregressivo	30
2.2.4	Componente integrado	31
2.2.5	Componente média móvel	32
2.2.6	Variáveis exógenas no modelo SARIMAX	33
2.2.7	Erro percentual médio absoluto	34
2.2.8	Metodologia de Box e Jenkins	35
2.2.9	Aplicações do modelo SARIMA e SARIMAX	37
<b>3</b>	<b>REVISÃO BIBLIOGRÁFICA</b>	<b>38</b>
3.1	PROTOCOLO DE PESQUISA	38
3.1.1	Pesquisa nas bases de dados	38
3.2.1	Modelagem de séries temporais na predição da dengue	40
3.2.2	Uso de Transformações e Variáveis Exógenas	41
3.2.3	Relevância e diferencial do estudo proposto	41
<b>4</b>	<b>METODOLOGIA</b>	<b>43</b>
4.1	COLETA DOS DADOS	43
4.2	PRÉ-PROCESSAMENTO DOS DADOS	44
4.2.1	Pré-processamento dos dados da dengue	45
4.2.2	Pré-processamento dos dados meteorológicos	48
4.2.3	Engenharia de dados e construção do dataset final	51
<b>5</b>	<b>RESULTADOS E DISCUSSÕES</b>	<b>55</b>
5.1	SELEÇÃO DO MODELO SARIMA	55
5.1.1	Proposta do modelo inicial	55

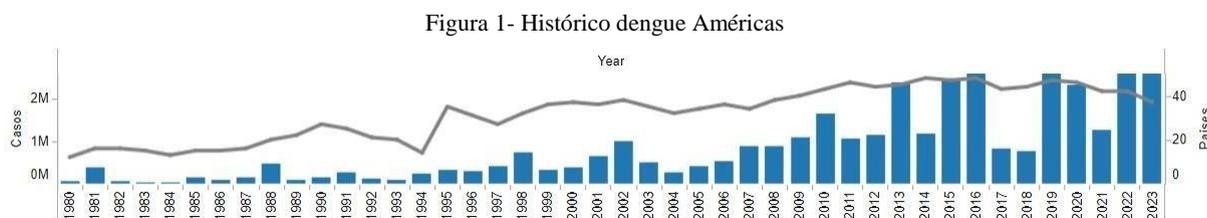
## SUMÁRIO

5.1.2	<b>Seleção da melhor ordem para o modelo SARIMA</b> .....	58
5.1.3	<b>Avaliação do modelo SARIMA</b> .....	59
5.2	<b>SELEÇÃO DO MODELO SARIMAX</b> .....	65
5.2.1	Seleção das melhores defasagens.....	67
5.2.2	<b>Seleção da melhor ordem para o modelo SARIMAX</b> .....	68
5.2.3	<b>Avaliação do modelo SARIMAX</b> .....	69
5.3	<b>DISCUSSÕES DOS RESULTADOS DOS MODELOS ESCOLHIDOS</b> .....	75
5.3.1	<b>Análise comparativa do modelo SARIMA</b> .....	75
5.3.2	<b>Primeiro período (até final de 2023)</b> .....	76
5.3.3	<b>Segundo período (até março de 2024)</b> .....	80
5.4	<b>ANÁLISE COMPARATIVA DO MODELO SARIMAX</b> .....	84
5.4.1	<b>Primeiro período (até final de 2023)</b> .....	84
5.4.2	<b>Segundo período (até março de 2024)</b> .....	87
6	<b>CONCLUSÃO</b> .....	92
7	<b>REFERÊNCIAS BIBLIOGRÁFICAS</b> .....	94

## 1. INTRODUÇÃO

A dengue é uma enfermidade de impacto global que exerce uma considerável influência socioeconômica, afetando a qualidade de vida de milhões de pessoas em todo o mundo. A PAHO estima que aproximadamente 500 milhões de indivíduos nas Américas estão suscetíveis à contração da dengue. Nas últimas quatro décadas, a incidência da doença na região aumentou significativamente, passando de 1,5 milhão de casos acumulados durante os anos 1980 para 16,2 milhões na década de 2010-2019 (PAHO, 2024).

Durante a epidemia de 2013, registrou-se pela primeira vez na região mais de 2 milhões de casos, resultando em uma incidência de 430,8 casos para cada 100 mil habitantes. Este período também testemunhou 37.692 casos de dengue grave e 1.280 óbitos em todo o continente. Já em 2019, os registros se elevaram para um pouco mais de 3,1 milhões de casos, com 28 mil considerados graves e 1.534 resultando em óbito (PAHO, 2023).



Fonte: (PAHO, 2024)

A Figura 1 retrata a evolução do número de casos de dengue anuais em toda a extensão do continente americano de 1980 a 2023. No Brasil, as primeiras epidemias clinicamente e laboratorialmente documentadas surgiram na década de 1980, em 1981, foram notificados mais de sete milhões de casos no Brasil (FIOCRUZ, 2013).

A dengue continua a representar um desafio significativo para a saúde pública no Brasil, com uma preocupante expansão do surto do vírus em todo o território nacional. Em 2022, o boletim epidemiológico do Ministério da Saúde de maio apontou 542 mil casos registrados até o final de abril, um aumento de 113% em comparação ao mesmo período do ano anterior, com 160 óbitos confirmados por dengue. Esses números já se aproximavam dos 544 mil casos registrados durante todo o ano de 2021, evidenciando uma tendência alarmante na proliferação da doença (G1, 2022).

Já em 2024 o Brasil ultrapassou a marca de 1 milhão de casos (prováveis e

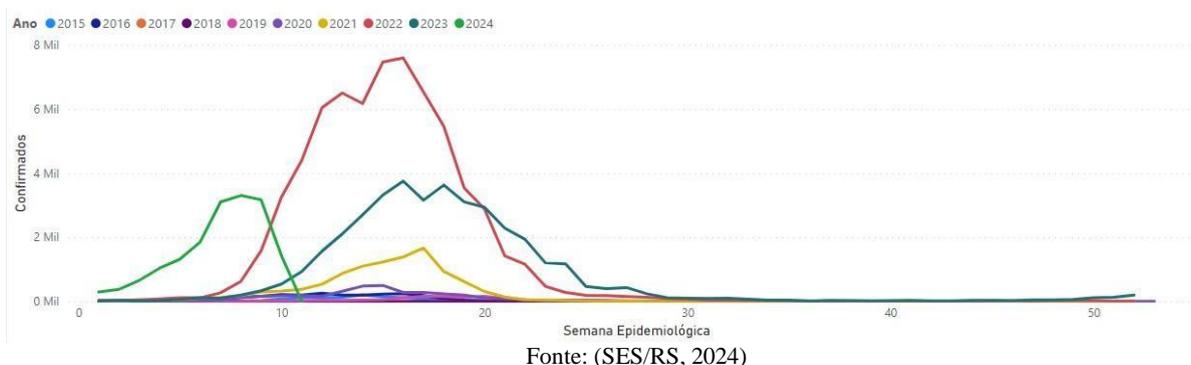
confirmados) de dengue somente nas primeiras oito semanas do ano, registrando um total de 1.017.278 casos, um salto significativo em relação aos 207.475 casos no mesmo período do ano passado. Até o momento, foram confirmadas 214 mortes desde janeiro, com 687 casos ainda sob investigação, comparado a 149 óbitos entre as semanas 01 e 08 de 2023 (G1, 2024).

O vice-presidente da Sociedade Brasileira de Imunizações (SBIIm), Renato Kfourri, e Kleber Luz, infectologista e consultor da Organização Mundial da Saúde (OMS) e Opas, expressaram preocupação com o início precoce e a rápida ascensão dos casos de dengue em 2024. "a curva está tendo uma inclinação muito positiva desde o início do ano", antecipando que "a expectativa é que a gente bata todos os recordes de dengue em 2024, já que a largada começou com alta velocidade" (G1, 2024).

Este cenário reflete a imprevisibilidade do comportamento da dengue, que varia significativamente ano a ano, e ressalta a urgência de abordagens estratégicas para a prevenção e controle das arboviroses no país. A análise de Kfourri sobre a variabilidade da temporada da dengue aponta para a necessidade de vigilância contínua e adaptabilidade nas respostas de saúde pública.

A região sul do Brasil, no ano de 2022, foi registrado como uma das áreas mais afetadas pela dengue, ocupando o segundo lugar em termos de taxa de incidência, com 427 casos para cada 100 mil habitantes. A Secretaria de Saúde do Estado do Rio Grande do Sul tem disponibilizado dados públicos atualizados sobre este fenômeno epidemiológico, que evidenciam a crescente problemática da dengue no estado, conforme destacado na Figura 2 abaixo. Até 27 de junho de 2022, o Rio Grande do Sul havia notificado um total de 77,2 mil casos e registrado 59 mortes, uma situação alarmante que chamou a atenção para a necessidade de ações mais efetivas de controle e prevenção (G1, 2022).

Figura 2 - Histórico dengue RS



Em 2024, a situação da dengue no Rio Grande do Sul apresentou uma nova dinâmica, influenciada por fatores climáticos adversos. Fenômenos como o El Niño e as

chuvas intensas, que afetaram o estado nos últimos meses, contribuíram para o aumento no número de casos da doença (RS.GOV, 2024). Essas condições climáticas exacerbaram a proliferação do *Aedes aegypti*, vetor da dengue, especialmente pelo acúmulo de água parada e resíduos, criando ambientes propícios para a sua reprodução. O governo estadual, consciente da gravidade do cenário, intensificou as ações de enfrentamento à doença, enfatizando a importância de uma resposta coordenada em todas as regiões.

Marcelo Vallandro, diretor-adjunto do Centro Estadual de Vigilância em Saúde (Cevs), destacou que a infestação do mosquito transmissor da dengue foi impulsionada por esses fenômenos climáticos, levando a altos índices de infestação e, conseqüentemente, à transmissão da doença. Vallandro também alertou sobre a necessidade de atenção em todo o estado, mencionando que a maior incidência de casos foi observada na Região Norte e Missioneira, mas que a Região Metropolitana também apresentou altos índices de infestação (RS.GOV, 2024).

Com 93,8% dos municípios enfrentando infestação do *Aedes aegypti*, o Rio Grande do Sul registrou 6.126 notificações de casos suspeitos de dengue em 2024. Destes, 2.915 foram confirmados, e 1.909 permanecem sob investigação, enquanto 1.302 foram descartados. Além disso, foram confirmadas duas mortes por dengue no estado. Nas primeiras cinco semanas de 2024, o número de casos notificados e confirmados foi 17 vezes maior do que no mesmo período de 2023, evidenciando um agravamento da situação epidemiológica da dengue no Rio Grande do Sul (RS.GOV, 2024).

A comparação dos dados de 2022 com os de 2024 ilustra não apenas o agravamento da situação da dengue no Rio Grande do Sul, mas também a complexidade do desafio enfrentado pelo estado. A interação entre fenômenos climáticos e a dinâmica de transmissão da dengue destaca a necessidade urgente de estratégias de prevenção e controle adaptadas às realidades locais e às mudanças ambientais.

Nesse contexto, a implementação de um modelo preditivo para o número de casos de dengue, embasado em dados confiáveis, surge como uma solução promissora para compreender melhor o comportamento da doença, especialmente no que concerne às variáveis climáticas que podem influenciar no aumento dos casos ao longo do tempo.

## 1.1 OBJETIVO

O objetivo principal deste estudo é realizar uma análise aprofundada de séries temporais e comparar diferentes modelos para prever o número de casos de dengue na

cidade de Porto Alegre, Rio Grande do Sul. A seguir, os objetivos específicos são detalhados

### **1.1.1 Objetivo Geral**

Realizar uma análise aprofundada de séries temporais e comparar diferentes modelos para prever o número de casos de dengue na cidade de Porto Alegre, Rio Grande do Sul.

### **1.1.2 Objetivos Específicos**

- 1- Coleta e Pré-processamento de Dados: Obter e preparar conjuntos de dados públicos e atualizados sobre casos de dengue fornecidos pelo governo do estado do Rio Grande do Sul, assim como informações meteorológicas específicas para a região de Porto Alegre.
- 2- Aplicação de Modelos de Séries Temporais: Implementar e comparar modelos de séries temporais, incluindo SARIMA e SARIMAX, para determinar quais oferecem as melhores previsões para a incidência de casos de dengue.
- 3- Análise do Impacto de Variáveis Climáticas: Investigar como a inclusão de variáveis climáticas, como temperatura e precipitação, nos modelos SARIMAX pode aumentar a precisão das previsões, adaptando-se às dinâmicas de transmissão da dengue, fortemente influenciadas por condições ambientais.
- 4- Avaliação do Desempenho dos Modelos: Avaliar o desempenho dos modelos implementados, comparando as previsões com dados reais para entender melhor a propagação da dengue em Porto Alegre.

### **1.1.3 Motivação e Justificativa**

A escolha da cidade de Porto Alegre para este estudo foi motivada pela minha participação em um projeto de pesquisa focado na aplicação de inteligência artificial e modelos de predição na área da saúde no estado do Rio Grande do Sul. Durante este projeto, identifiquei que Porto Alegre possui uma disponibilidade maior de dados públicos

e atualizados sobre casos de dengue e informações meteorológicas, o que facilita a análise e a modelagem preditiva. A maior disponibilidade de dados em Porto Alegre permite uma análise mais detalhada e precisa, essencial para o desenvolvimento e avaliação de modelos de predição eficazes.

Além disso, a escolha desta cidade está alinhada com a relevância epidemiológica da dengue na região, onde a doença representa um desafio significativo para a saúde pública. Portanto, focar os esforços de pesquisa em Porto Alegre oferece uma oportunidade clara para desenvolver ferramentas que possam ser aplicadas de maneira prática e eficaz no combate à dengue, beneficiando diretamente a população local.

A justificativa para este estudo está fundamentada na problemática crescente da dengue em Porto Alegre e no estado do Rio Grande do Sul. A dengue é uma doença viral transmitida pelo mosquito *Aedes aegypti*, cuja incidência tem aumentado significativamente nos últimos anos, causando um impacto substancial na saúde pública e na qualidade de vida da população. Porto Alegre, sendo uma das principais cidades do estado, apresenta uma alta incidência de casos de dengue, o que demanda uma atenção especial das autoridades de saúde para o controle e prevenção da doença.

A previsão precisa do número de casos de dengue pode ser uma ferramenta poderosa para a tomada de decisões estratégicas em saúde pública, permitindo ações preventivas mais eficazes e a alocação de recursos de forma mais eficiente. A integração de dados climáticos e epidemiológicos em modelos preditivos, como os modelos SARIMA e SARIMAX, pode proporcionar insights valiosos sobre a dinâmica da transmissão da dengue, especialmente considerando o impacto das variáveis climáticas no ciclo de vida do mosquito vetor.

## 1.2 ORGANIZAÇÃO DO TRABALHO

Este trabalho está organizado em seis capítulos:

Capítulo 1: Introdução ao tema, incluindo motivação, delimitação do problema, justificativa e objetivos.

Capítulo 2: Fundamentação teórica, abordando o ciclo de vida do mosquito *Aedes aegypti* e o impacto das condições climáticas, além do uso de séries temporais na análise da dengue.

Capítulo 3: Revisão bibliográfica, apresentando artigos e estudos relevantes sobre a predição da dengue e o diferencial deste estudo.

Capítulo 4: Metodologia, detalhando a coleta, pré-processamento dos dados e

aplicação dos modelos estatísticos.

Capítulo 5: Resultados e discussão, oferecendo uma análise dos dados e a eficácia dos modelos utilizados.

Capítulo 6: Conclusões e sugestões para pesquisas futuras.

Esta estrutura garante um fluxo lógico, desde a contextualização e fundamentação teórica, passando pela metodologia, até a análise e interpretação dos resultados, concluindo com as considerações finais e perspectivas futuras.

## 2 FUNDAMENTAÇÃO TEÓRICA

Este capítulo tem como objetivo embasar o leitor em relação aos conceitos por trás temas utilizados no desenvolvimento deste trabalho. Busca-se promover um entendimento geral dos conceitos que foram aplicados para o desenvolvimento do modelo de predição apresentado neste trabalho.

### 2.1 DENGUE: VETORES, VIRUS E DISSEMINAÇÃO

A dengue é uma doença viral transmitida aos seres humanos pela picada do mosquito *Aedes aegypti*. Considerada um dos principais problemas de saúde pública em diversas regiões tropicais e subtropicais do mundo, a dengue representa uma séria ameaça à saúde e ao bem-estar da população. Com milhões de casos notificados anualmente, a doença causa preocupação tanto em termos de morbidade quanto de mortalidade (PAHO, 2024).

Nessa seção, será abordado em detalhes o mosquito *Aedes aegypti* como vetor da doença e sua relação com a transmissão viral aos seres humanos. Além disso, se discutirá o impacto que a temperatura, umidade e precipitação têm no ciclo de vida do mosquito. Compreender como esses fatores climáticos influenciam a ecologia e o comportamento do *Aedes aegypti* é fundamental para entendermos a sazonalidade da dengue.

#### 2.1.1 Origens e Comportamento dos Vetores

O mosquito *Aedes aegypti* é o vetor responsável pela transmissão da dengue. Ele é conhecido por sua preferência em se reproduzir em áreas urbanas e domiciliares, tornando-se um desafio para as autoridades de saúde pública controlarem a propagação da doença (BESERRA et al., 2009). A fêmea do mosquito é a responsável pela disseminação do vírus, e ela o adquire ao picar uma pessoa já infectada. Uma vez infectado, o mosquito se torna capaz de transmitir o vírus para outras pessoas saudáveis em picadas subsequentes (SCOTT et al., 2000).

Os mosquitos *Aedes*, prováveis oriundos da África, foram introduzidos nas Américas durante o período colonial, possivelmente através do comércio de escravos. O

*Aedes aegypti*, com seu hábito de alimentação diurno e propensão a ambientes urbanos, tornou-se um vetor predominante da dengue. Sua capacidade de adaptar-se a diferentes condições e de reproduzir-se em recipientes com água limpa ou até mesmo em águas poluídas ampliou sua disseminação. O *Aedes albopictus*, encontrado principalmente na Ásia, também é um vetor relevante, embora sua ocorrência nas Américas tenha sido mais restrita, até o momento, ao continente asiático e a ambientes silvestres.

### **2.1.2 Vírus da Dengue e seus Sorotipos**

O vírus da dengue, integrante do gênero Flavivírus, é caracterizado pela existência de quatro sorotipos distintos: DENV-1, DENV-2, DENV-3 e DENV-4 (WHO, 2021). Cada sorotipo é capaz de provocar a infecção, e a imunidade adquirida por um deles não se estende aos demais, possibilitando até quatro infecções distintas durante a vida de um indivíduo. A gama de manifestações clínicas varia significativamente, desde uma febre leve até condições mais severas, como a FHD e a SCD, particularmente em casos de reinfeção (PAHO, 2024). O período de incubação do vírus, após a transmissão por um mosquito *Aedes aegypti* infectado, oscila entre 3 a 14 dias, desencadeando sintomas como febre alta, dor de cabeça intensa, dor nas articulações e músculos, e dor atrás dos olhos.

### **2.1.3 Ciclo de Vida do *Aedes aegypti***

O ciclo de vida do *Aedes aegypti*, o vetor principal do vírus da dengue, consiste em quatro estágios desenvolvimentais, cada um desempenhando um papel crítico na sobrevivência e capacidade de transmissão da espécie. Este ciclo é um produto de uma evolução, permitindo que o mosquito prospere em diversos ambientes, especialmente em áreas urbanas densamente povoadas.

**Estágio de Ovo:** O primeiro estágio do ciclo de vida do *Aedes aegypti* começa com a postura de ovos, após o acasalamento e a alimentação sanguínea das fêmeas. Estes ovos são únicos entre os mosquitos por sua incrível capacidade de resistir a longos períodos de dessecação, uma adaptação evolutiva que permite à espécie sobreviver em condições de seca extrema (HAWLEY, 1988). Os ovos são depositados de forma singular, aderindo às paredes internas dos recipientes com água parada e podem sobreviver por até 450 dias, proporcionando uma oportunidade de sobrevivência através das estações secas (FOCKS et al., 1993).

A eclosão ocorre quando as condições são favoráveis — geralmente dentro de 48 horas após o contato com a água. A cor dos ovos, inicialmente brilhante, torna-se mais escura à medida que os embriões se desenvolvem (HAWLEY, 1988), ver Figura 3.

Figura 3 - Ovo do aedes aegypti



Fonte: (FIOCRUZ, 2023)

Estágio Larval: Quando a água envolve os ovos, as larvas emergem no ambiente aquático, iniciando o estágio larval representada pela Figura 4, que é dividido em quatro fases de crescimento. Durante este período, as larvas alimentam-se de microorganismos e matéria orgânica, crescendo e mudando exponencialmente de tamanho. O ritmo deste crescimento depende de diversos fatores, incluindo temperatura, qualidade da água e disponibilidade de alimentos, com a temperatura sendo o principal regulador da velocidade de desenvolvimento das larvas (RUEDA et al., 1990).

Figura 4 - Larvas do aedes aegypti



Fonte: (FIOCRUZ, 2023)

Estágio de Pupa: A pupa do *Aedes aegypti*, ver Figura 5, é um estágio de transformação dinâmica, onde a larva aquática se converte em mosquito adulto. Durante a fase de pupa, que geralmente dura entre um e três dias, o indivíduo não se alimenta, mas passa por mudanças fisiológicas profundas. As estruturas externas do adulto, como as asas e os órgãos sensoriais, começam a se formar, e as estruturas larvais são reabsorvidas ou transformadas (FARNESI et al., 2009).

Figura 5 - Pupas do *aedes aegypti*



Fonte: (FIOCRUZ, 2023)

Estágio Adulto: O surgimento do mosquito adulto, observado na Figura 6, é um evento crítico, pois representa a transição para a fase de reprodução e potencial transmissão de doenças. Os mosquitos adultos do *Aedes aegypti*, uma vez emergidos, buscam repouso para permitir que suas asas e o exoesqueleto se sequem e endureçam. As fêmeas, em seguida, iniciam a procura por uma refeição de sangue, o que desencadeia o desenvolvimento de seus ovos e perpetua o ciclo de vida do vetor (SCOTT et al., 2000).

Figura 6 - Adulto do *aedes aegypti*



Fonte: (FIOCRUZ, 2023)

A Figura 7 resume este ciclo vital em um esquema informativo, delineando a progressão através dos estágios e enfatizando a importância de cada um na transmissão da dengue.

Figura 7 - Ciclo completo de vida do *Aedes aegypti*



Fonte: (SES/RS, 2023)

No próximo tópico desta seção, será discutido em detalhes o impacto das variáveis climáticas, como temperatura, precipitação e umidade, no ciclo de vida do mosquito *Aedes aegypti*. Compreender como esses fatores ambientais afetam cada fase do desenvolvimento do mosquito é essencial para compreender sua ecologia e comportamento, além de como essas variáveis climáticas podem influenciar a sazonalidade e a transmissão da dengue. A análise desses aspectos nos permitirá avançar no entendimento dos padrões de ocorrência da doença.

#### 2.1.4 Impacto da temperatura, umidade e precipitação no ciclo de vida do *Aedes aegypti*

O ciclo de vida do mosquito *Aedes aegypti* é diretamente influenciado pelas condições climáticas, sendo a temperatura, umidade e precipitação fatores-chave que

afetam seu desenvolvimento e comportamento. Esses fatores climáticos têm um papel crucial em cada fase do ciclo do mosquito, desempenhando um papel determinante na sua ecologia e na transmissão da dengue (MOGI et al., 1996).

A temperatura é um fator ambiental crítico que modula o ciclo de vida do *Aedes aegypti*. Elevações moderadas na temperatura podem diminuir a duração do ciclo de vida do mosquito, desde a eclosão dos ovos até a emergência dos adultos, aumentando potencialmente as taxas de reprodução e a população do mosquito (RUEDA et al., 1990). Esse efeito termodependente é particularmente notável nos trópicos, onde pequenas variações na temperatura média podem causar flutuações substanciais na população de mosquitos (CHRISTOPHERS, 1960).

Além disso, a temperatura afeta a viabilidade do vírus dentro do vetor. Estudos têm demonstrado que as taxas de replicação do vírus da dengue aumentam com a temperatura, elevando a probabilidade de transmissão à medida que os mosquitos se tornam mais propensos a se alimentar de sangue (TUN-LIN et al., 2000). A transmissão do vírus é otimizada em um intervalo de temperatura específico, acima do qual a viabilidade do mosquito pode começar a declinar (PAUPY et al., 2009).

A umidade relativa impacta a sobrevivência dos estágios imaturos do *Aedes aegypti* de várias maneiras. A dessecação dos ovos é uma ameaça significativa em ambientes de baixa umidade, enquanto uma umidade mais alta pode mitigar esse risco, garantindo uma taxa de eclosão mais consistente (REISKIND et al., 2009). As larvas e pupas, que dependem de habitats aquáticos, são menos suscetíveis a flutuações de umidade, mas os locais de reprodução podem ser afetados pela evaporação em condições de baixa umidade.

A umidade também afeta o comportamento dos mosquitos adultos, particularmente em suas atividades de procura de hospedeiros e alimentação. A desidratação em climas secos pode reduzir a atividade dos mosquitos e a eficiência da transmissão de patógenos. Portanto, ambientes com umidade relativa ótima podem sustentar ciclos de transmissão de patógenos mais eficazes, ao passo que umidade excessiva ou insuficiente pode limitar esses ciclos (BARRERA et al., 2006).

A precipitação, especialmente a chuva, é fundamental na determinação da disponibilidade e distribuição de criadouros adequados para o *Aedes aegypti*. A água da chuva acumulada em recipientes naturais e artificiais cria habitats para os estágios aquáticos do mosquito. Esses habitats podem variar desde pequenas poças até objetos como pneus e vasos de plantas que coletam água (SHAMAN et al., 2005).

No entanto, o impacto da precipitação não é uniforme. Chuvas intensas podem

levar à limpeza ou destruição de criadouros existentes, enquanto chuvas moderadas podem criar condições ideais para o desenvolvimento larval. O acúmulo de água após chuvas intermitentes pode permitir que os ovos depositados durante períodos secos finalmente eclodam, causando um aumento repentino na população de mosquitos (MORRISON et al., 2004).

Os padrões de infestação de mosquitos são influenciados pela sazonalidade e intensidade das chuvas. Precipitação irregular pode resultar em uma dinâmica populacional complexa, com períodos de expansão e declínio populacional. Em regiões com chuvas sazonais, a abundância de mosquitos está frequentemente correlacionada com o início da estação chuvosa, o que sugere que o monitoramento da precipitação pode ser um componente valioso na previsão de surtos de dengue (JOHANSSON et al., 2009).

As variações sazonais influenciam diretamente a dinâmica populacional do *Aedes aegypti*. Durante as estações chuvosas, quando os criadouros são abundantes, ocorre um aumento significativo na população do mosquito. Esse aumento pode ser rapidamente seguido por um declínio durante a estação seca, quando a água se torna escassa e os criadouros diminuem (PONGSUMPUN et al., 2008).

Variações sazonais nas condições climáticas podem criar oportunidades para a propagação do vírus da dengue. Estudos indicam que a incidência da dengue muitas vezes aumenta após períodos de chuvas pesadas, sugerindo uma ligação direta entre precipitação e transmissão do vírus (JOHANSSON et al., 2009). Além disso, a variabilidade sazonal pode impactar a interação entre hospedeiros humanos e mosquitos, influenciando a probabilidade de surtos de doenças.

Os surtos de dengue estão frequentemente alinhados com períodos imediatamente após a estação chuvosa, quando as populações de mosquitos estão em seu pico. No entanto, a variabilidade climática, como fenômenos El Niño e La Niña, pode alterar esses padrões, complicando as previsões de surtos e desafiando os esforços de controle (CAMPBELL-LENDRUM et al., 2015).

O clima desempenha um papel vital na ecologia do *Aedes aegypti* e na epidemiologia da dengue. O aquecimento global e as mudanças climáticas podem aumentar a temperatura média e alterar os padrões de precipitação, potencialmente expandindo a área geográfica habitável para o mosquito e aumentando a duração da temporada de transmissão da dengue (CAMPBELL-LENDRUM et al., 2015).

Considerando a interação entre esses fatores climáticos e o ciclo de vida do *Aedes aegypti*, torna-se evidente a importância de compreender essa dinâmica para implementar estratégias eficazes de controle da dengue. Modelos preditivos baseados em estudos

bioecológicos podem antecipar áreas de infestação, permitindo a implementação de medidas preventivas direcionadas, como a eliminação de criadouros e campanhas de conscientização em regiões de maior risco.

## 2.2 SERIES TEMPORAIS

Séries temporais são conjuntos de dados observados ao longo do tempo, onde cada observação é associada a um ponto específico no tempo, organizadas em uma sequência cronológica. Esses dados podem ser coletados em intervalos regulares ou irregulares, com o principal objetivo de analisar, modelar o comportamento dos dados ao longo do tempo e fazer previsões sobre futuras observações (MORETTIN, 2010). Esta metodologia é fundamental em campos como a epidemiologia e a climatologia, onde é crucial compreender as tendências e padrões para prever e mitigar possíveis impactos.

No campo da epidemiologia, a análise de séries temporais é crucial para monitorar a evolução de doenças, identificar surtos e avaliar intervenções de saúde pública. Modelos matemáticos e estatísticos, como o SARIMA, têm sido amplamente adotados para prever a incidência de doenças infecciosas, como a dengue, considerando suas flutuações sazonais (WU et al., 2009). A inclusão de variáveis climáticas nos modelos SARIMA tem proporcionado insights valiosos sobre a relação entre clima e a incidência de doenças, evidenciando a importância de considerar fatores exógenos na análise de séries temporais (WU et al., 2009).

Além disso, a metodologia de Box e Jenkins, introduzida por George Box e Gwilym Jenkins em 2015, oferece uma abordagem sistemática para a modelagem e previsão de séries temporais. Esta metodologia, que inclui os modelos SARIMA e suas extensões, como SARIMAX, é fundamental para a análise detalhada de fenômenos complexos, permitindo uma compreensão mais aprofundada das dinâmicas entre doenças e fatores climáticos.

Consequentemente, a análise de séries temporais desempenha um papel significativo no estudo de tendências e padrões de doenças infecciosas, fornecendo informações essenciais para o planejamento de medidas preventivas e o aprimoramento das estratégias de saúde pública. A capacidade de prever o número de casos de doenças como a dengue com base em dados históricos e variáveis exógenas é um avanço significativo para a saúde pública, permitindo uma resposta mais eficaz às emergências de saúde e a implementação de estratégias de prevenção mais direcionadas.

Na próxima seção, serão explorados os modelos SARIMA e SARIMAX em detalhes, compreendendo seus componentes e a importância da inclusão de variáveis exógenas para aprimorar a análise e previsão de séries temporais complexas, como a incidência de casos de dengue.

### 2.2.1 Fundamentos estatísticos em análise de séries temporais

Neste tópico será apresentado alguns conceitos que são fundamentais para análise estatística de séries temporais, sendo eles:

- a) Estacionariedade: Segundo Box, Jenkins, e Reinsel (2015), a estacionariedade é uma característica indispensável nas séries temporais, exigindo que a média, variância e autocovariância sejam invariantes no tempo. Esses autores destacam a importância dessa propriedade, uma vez que a maioria dos modelos preditivos, incluindo ARIMA e SARIMA, pressupõe a estacionariedade da série ou sua transformação para uma forma estacionária. Para alcançar a estacionariedade em séries que exibem tendências ou variações sazonais, métodos de transformação, como a diferenciação e a aplicação de logaritmos, são frequentemente empregados (BOX et al., 2015).

A distinção entre estacionariedade estrita e fraca é crucial para a análise de séries temporais, com a estacionariedade fraca sendo mais aplicável em contextos econômicos e financeiros. Esta forma mais leve da propriedade, onde se presume que apenas a primeira e a segunda ordens de momentos são constantes, é suficiente para muitas aplicações práticas. Essa abordagem facilita a modelagem de fenômenos econômicos complexos, permitindo a análise de séries temporais sob suposições menos restritivas (HAMILTON, 1994).

- b) Processos com tempo discreto: As observações são realizadas em intervalos regulares. Um modelo AR(p), por exemplo, pode ser formalizado como A habilidade desses modelos em capturar dependências temporais faz deles ferramentas essenciais na previsão de séries temporais. A seleção do tipo de modelo e sua ordem depende intrinsecamente da estrutura da série temporal em análise, com a função de autocorrelação fornecendo insights valiosos para essa escolha (TSAY, 2005).
- c) Função de autocorrelação e autocorrelação parcial: Brockwell e Davis (2002)

definem a ACF e PACF como medidas fundamentais para entender as dependências temporais. Essa função ajuda a identificar os componentes AR e MA em uma série temporal, sendo muito útil para determinar a ordem dos modelos ARIMA.

- d) Função de autocorrelação: A Função de Autocorrelação (ACF) mede a correlação entre observações de uma série temporal separadas por diferentes defasagens (lags). Formalmente, a autocorrelação em uma defasagem  $k$  é definida como:

$$\rho_k = \frac{\sum_{t=k+1}^N (Y_t - \bar{Y})(Y_{t-k} - \bar{Y})}{\sum_{t=1}^N (Y_t - \bar{Y})^2} \quad (1)$$

onde  $\bar{Y}$  é a média da série temporal,  $Y_t$  é o valor da série no tempo  $t$ , e  $Y_{t-k}$  é o valor da série no tempo  $t - k$ .

A ACF ajuda a identificar padrões como sazonalidade e persistência. Em uma série temporal estacionária, a ACF tende a decair rapidamente para zero. No entanto, em séries não estacionárias, a ACF pode decair lentamente, indicando a necessidade de diferenciação para alcançar a estacionariedade.

Visualmente, a ACF é apresentada como um gráfico de barras, onde a altura de cada barra representa a magnitude da correlação em cada defasagem. Esse gráfico é fundamental para identificar a ordem de componentes MA nos modelos ARIMA, onde um corte abrupto na ACF após certas defasagens pode indicar a ordem do componente MA.

- e) Função de autocorrelação parcial: A Função de Autocorrelação Parcial (PACF) mede a correlação entre observações de uma série temporal separadas por várias defasagens, eliminando o efeito das observações intermediárias. Enquanto a ACF considera a correlação direta e indireta, a PACF remove a influência das correlações em defasagens menores.

Matematicamente, a PACF na defasagem  $k$  é a correlação entre  $Y_t$  e  $Y_{t-k}$  que não é explicada pelas observações intermediárias  $Y_{t-1}, Y_{t-2}, \dots, Y_{t-(k-1)}$ . A PACF é útil para identificar a ordem dos componentes AR nos modelos ARIMA.

A PACF é representada graficamente de maneira similar à ACF, com um gráfico de barras mostrando a magnitude das correlações parciais. Um corte abrupto na PACF após uma certa defasagem indica a ordem apropriada do componente AR.

Ambas as funções ACF e PACF são essenciais na identificação e diagnóstico de modelos de séries temporais, oferecendo insights críticos sobre a estrutura de dependência temporal nos dados e orientando a seleção das ordens dos componentes AR e MA nos modelos ARIMA e SARIMA.

- f) Teste de Ljung-Box: Este teste é utilizado para avaliar se um modelo capturou adequadamente a estrutura de dependência dos dados. A rejeição da hipótese nula indica a necessidade de revisar o modelo, enquanto a sua não rejeição sugere que os resíduos do modelo podem ser considerados ruído branco, implicando em um ajuste adequado do modelo (LJUNG et al., 1978).

O Teste de Ljung-Box é fundamental na análise de séries temporais, servindo como um mecanismo de verificação para a adequação do modelo. A aplicação deste teste é um passo crucial no processo de modelagem, garantindo que todas as dependências temporais foram devidamente consideradas.

- g) Ruído branco: A definição de uma série de ruído branco como um conjunto de observações independentes e identicamente distribuídas, com média zero e variância constante, é central na teoria de séries temporais. Granger e Newbold (1976) enfatizam a importância do ruído branco para a validação de modelos de séries temporais, argumentando que resíduos que se assemelham a ruído branco indicam um modelo bem ajustado. Eles advogam a análise de resíduos para identificar a presença de ruído branco como uma etapa essencial na avaliação da adequação do modelo (Granger et al., 1976).

A presença de ruído branco nos resíduos de um modelo indica que o modelo conseguiu extrair toda a informação útil dos dados, conforme discutido por Granger e Newbold. Este é um indicador chave da validade do modelo, e sua verificação é uma prática padrão na análise de séries temporais.

- h) Transformação de Yeo-Johnson: É uma técnica utilizada para estabilizar a variância e tornar os dados mais próximos de uma distribuição normal. Essa transformação é uma generalização da transformação de Box-Cox, mas com a vantagem de poder ser aplicada a dados que incluem valores zero ou negativos. A fórmula da transformação de Yeo-Johnson é definida de forma que diferentes potências ( $\lambda$ ) são aplicadas dependendo dos valores dos dados, resultando em uma transformação contínua e diferenciável que pode tratar uma ampla gama de distribuições de dados.

O uso da transformação de Yeo-Johnson é particularmente útil em análises de séries temporais, onde a normalidade dos dados é uma suposição importante

para muitos modelos estatísticos. Ao aplicar essa transformação, os dados tendem a se conformar melhor às suposições do modelo, melhorando a precisão das previsões e a qualidade do ajuste. Além disso, a transformação pode ajudar a mitigar o impacto de *outliers*, tornando os modelos mais robustos e confiáveis.

- i) Critério de informação bayesiano: É uma ferramenta utilizada para a seleção de modelos de séries temporais, entre outros. O BIC é baseado na função de verossimilhança e inclui um termo de penalização para a complexidade do modelo, o que ajuda a evitar o sobreajuste. Matematicamente, o BIC é expresso como:

$$BIC = -2 \ln(L) + k \ln(n) \quad (2)$$

onde  $L$  é a máxima verossimilhança do modelo,  $k$  é o número de parâmetros estimados, e  $n$  é o número de observações. O termo  $k \ln(n)$  penaliza modelos com muitos parâmetros, promovendo a escolha de modelos mais simples e parcimoniosos.

O BIC é particularmente útil quando se deseja comparar diferentes modelos com diferentes números de parâmetros. Um valor de BIC mais baixo indica um modelo preferível. Ao utilizar o BIC para selecionar o melhor modelo, estamos equilibrando a qualidade do ajuste com a simplicidade do modelo, o que é essencial em análises de séries temporais, onde modelos complexos podem facilmente se ajustar ao "ruído" nos dados em vez dos verdadeiros padrões subjacentes.

- j) Critério de informação de akaike: Se trata de uma ferramenta popular para a seleção de modelos estatísticos. Assim como o BIC, ver equação (2), o AIC é baseado na função de verossimilhança, mas sua fórmula é diferente, temos que:

$$AIC = 2k - 2 \ln(L) \quad (3)$$

onde  $L$  é a máxima verossimilhança do modelo, e  $k$  é o número de parâmetros estimados. O AIC também penaliza a complexidade do modelo, mas de forma menos severa que o BIC. A ideia principal por trás do AIC é escolher o modelo que minimiza a perda de informação.

O AIC é amplamente utilizado devido à sua simplicidade e eficácia em comparar modelos, especialmente quando o foco é prever futuros dados. Um modelo com um AIC mais baixo é considerado preferível, indicando um

melhor equilíbrio entre a complexidade do modelo e a qualidade do ajuste. No entanto, como o AIC tende a favorecer modelos mais complexos em comparação com o BIC, é importante considerar o contexto da análise e o objetivo específico ao escolher entre esses critérios.

### **2.2.2 Modelo SARIMA e SARIMAX**

O desenvolvimento dos modelos SARIMA e SARIMAX é uma evolução significativa na análise de séries temporais (BOX et al., 2015). A metodologia proposta por esses autores, conhecida como a metodologia de Box-Jenkins, revolucionou a maneira como as séries temporais são modeladas e analisadas, permitindo previsões mais precisas e a compreensão de padrões complexos nos dados.

O modelo SARIMA é uma extensão do ARIMA, projetado para capturar a sazonalidade nos dados de séries temporais. O modelo é descrito pela integração de componentes autoregressivos (AR), de médias móveis (MA), e um termo de diferenciação (I) para tornar a série temporal estacionária, com adição de componentes sazonais que modelam flutuações periódicas específicas ao longo do tempo. Essa capacidade de lidar com a sazonalidade torna o SARIMA particularmente útil em campos como a climatologia e a epidemiologia, onde padrões sazonais são comuns (BROCKWELL et al., 2002).

O modelo SARIMAX estende o SARIMA ao incorporar variáveis exógenas, permitindo a análise da influência de fatores externos na série temporal. Esse modelo é especialmente relevante em estudos onde se deseja avaliar o impacto de variáveis externas — como condições climáticas no caso de doenças infecciosas — na variável de interesse (WU et al., 2009). A inclusão de variáveis exógenas no modelo SARIMAX proporciona uma compreensão mais aprofundada da dinâmica entre diferentes fatores e sua influência nas tendências e padrões observados.

A principal diferença entre SARIMA e SARIMAX está, portanto, na capacidade do SARIMAX de incluir informações adicionais por meio de variáveis exógenas, enriquecendo a análise e permitindo uma modelagem mais detalhada de complexas relações causais.

### **2.2.3 Componente autoregressivo**

O AR é um dos pilares fundamentais na modelagem de séries temporais, refletindo a ideia de que valores presentes podem ser influenciados por seus próprios valores anteriores. Um modelo autoregressivo de ordem  $p$ ,  $AR(p)$ , é matematicamente representado como:

$$Y_t = c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + \epsilon_t \quad (4)$$

onde  $Y_t$  é o valor da série temporal no tempo  $t$ ,  $c$  é uma constante,  $\phi_1, \phi_2, \dots, \phi_p$  são os coeficientes autoregressivos que medem a influência dos valores anteriores  $Y_{t-1}, Y_{t-2}, \dots, Y_{t-p}$  sobre o valor presente  $Y_t$  e  $\epsilon_t$  é o termo de erro aleatório no tempo  $t$ , que é assumido ser um processo de ruído branco (BOX et al., 2015).

A determinação da ordem  $p$  é crucial, pois um  $p$  inadequado pode levar a um modelo subajustado ou sobreajustado. O subajuste ocorre quando o modelo é muito simples e falha em capturar a estrutura subjacente dos dados; o sobreajuste acontece quando o modelo é excessivamente complexo, capturando o "ruído" aleatório como se fossem padrões significativos. Para equilibrar esses riscos, utilizam-se critérios de seleção de modelos, como o AIC e o BIC, representados pelas equações (2) e (3) respectivamente, que visam encontrar um modelo com o número ideal de parâmetros, promovendo um ajuste adequado sem incorrer em complexidade desnecessária. Ambos critérios buscam minimizar seus respectivos valores; o modelo com o menor AIC ou BIC é preferido por oferecer um equilíbrio ótimo entre simplicidade e capacidade de captura da estrutura dos dados.

Esses critérios são fundamentais na modelagem AR, pois ajudam a identificar a complexidade do modelo que melhor se ajusta à série temporal, levando em consideração tanto a adequação do modelo aos dados (maximizando  $L$ ) quanto a sua simplicidade (minimizando  $k$ ). Ao aplicar AIC e BIC, os analistas podem escolher de maneira objetiva a ordem  $p$  do componente AR, assegurando que o modelo final seja robusto, preciso e capaz de generalizar bem para dados não observados.

#### 2.2.4 Componente integrado

A integração em modelos de séries temporais, como ARIMA, SARIMA e SARIMAX, serve para transformar séries não estacionárias em estacionárias, um passo essencial para aplicar a modelagem autoregressiva e de médias móveis. Uma série

estacionária é aquela cujas propriedades estatísticas, como média e variância, não variam ao longo do tempo. A diferenciação, um método de integração, é empregada para remover tendências e ciclos sazonais, facilitando a análise e a modelagem subsequente.

A diferenciação regular (d) é definida como a subtração do valor atual da série pelo seu valor anterior, como:

$$\Delta Y_t = Y_t - Y_{t-1} \quad (5)$$

pode ser aplicada múltiplas vezes conforme necessário. A diferenciação sazonal (D), por outro lado, considera as diferenças entre valores separados por um período sazonal, dessa forma temos que:

$$\Delta_s Y_t = Y_t - Y_{t-s} \quad (6)$$

Onde  $s$  é o número de períodos em uma temporada. A adequada determinação dos parâmetros  $d$  e  $D$  é crucial para modelar efetivamente a estacionariedade das séries temporais.

A seleção de  $d$  e  $D$  é guiada por análises exploratórias dos dados e confirmada por critérios objetivos de seleção de modelos, como o AIC e o BIC, mencionados anteriormente no contexto do componente autoregressivo. Estes critérios ajudam a determinar o grau necessário de diferenciação, balanceando a complexidade do modelo e seu ajuste aos dados, para assegurar que a série transformada seja adequadamente estacionária, sem perder informação relevante.

A importância da integração reside na sua capacidade de preparar séries temporais para uma análise mais profunda e precisa usando modelos ARIMA, SARIMA e SARIMAX. Ao tornar a série estacionária, a integração permite a aplicação de técnicas autoregressivas e de médias móveis, facilitando a identificação de padrões e a realização de previsões confiáveis.

### 2.2.5 Componente média móvel

O componente de média móvel (MA) nos modelos ARIMA, SARIMA, e SARIMAX representa o impacto dos termos de erro passados na previsão atual, complementando a análise autoregressiva. Um modelo MA de ordem  $q$ ,  $MA(q)$ , é formulado como:

$$Y_t = \mu + \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \dots + \theta_q \epsilon_{t-q} \quad (7)$$

onde  $Y_t$  é o valor da série temporal no tempo  $t$ ,  $\mu$  é a média da série temporal,  $\epsilon_t$  é o termo de erro aleatório no tempo  $t$  e  $\theta_1, \theta_2, \dots, \theta_q$ ,  $\epsilon_{t-1}, \epsilon_{t-2}, \dots, \epsilon_{t-q}$  são os coeficientes de média móvel que quantificam a influência dos erros passados  $\epsilon_{t-1}, \epsilon_{t-2}, \dots, \epsilon_{t-q}$  sobre o valor atual  $Y_t$ . Esse componente ajuda a capturar as flutuações temporais e os "choques" aleatórios que afetam a série temporal.

A determinação da ordem  $q$  do componente MA é um aspecto crítico da modelagem, influenciando a capacidade do modelo de ajustar-se às irregularidades e padrões nos dados. A ordem  $q$  é escolhida com base na análise das autocorrelações dos dados, especificamente observando a ACF. Um corte abrupto na ACF após um certo número de defasagens pode indicar a ordem apropriada para o componente MA, um aspecto fundamental para garantir que o modelo capture adequadamente a estrutura de dependência temporal dos dados.

Além da análise da ACF, a seleção de  $q$  também se beneficia dos critérios de informação AIC e BIC, previamente discutidos. Esses critérios são empregados para avaliar e comparar modelos com diferentes ordens de MA, escolhendo aquele que oferece o melhor equilíbrio entre complexidade e ajuste aos dados. Isso assegura que o modelo final seja nem excessivamente simplificado, falhando em capturar as nuances dos dados, nem excessivamente complexo, resultando em sobreajuste.

A inclusão do componente MA nos modelos é vital para uma modelagem eficaz de séries temporais, permitindo que os modelos absorvam e reflitam a aleatoriedade e os choques inesperados presentes nos dados.

### **2.2.6 Variáveis exógenas no modelo SARIMAX**

A adição de variáveis exógenas ao modelo SARIMA, resultando no modelo SARIMAX, representa uma evolução significativa na capacidade de modelagem de séries temporais. Variáveis exógenas são fatores externos que podem influenciar a variável de interesse, mas não são diretamente afetados por ela. No contexto da análise temporal da dengue associada a fatores climáticos, variáveis como temperatura, precipitação e umidade podem ser cruciais para prever os casos de dengue, pois tais condições climáticas afetam diretamente o ciclo de vida e a distribuição dos mosquitos vetores (Wu et al., 2009).

A inclusão dessas variáveis no modelo SARIMAX permite capturar a influência

desses fatores externos sobre a série temporal estudada, potencialmente melhorando a precisão das previsões. Matematicamente, o modelo SARIMAX pode ser representado por

$$Y_t = \phi_1 Y_{t-1} + \theta_1 \epsilon_{t-1} + \beta X_t + \epsilon_t \quad (8)$$

onde  $Y_t$  é a variável de interesse (por exemplo, casos de dengue),  $X_t$  representa as variáveis exógenas (fatores climáticos),  $\beta$  são os coeficientes que quantificam a influência das variáveis exógenas sobre  $Y_t$ , e os demais termos seguem a notação padrão dos modelos ARIMA e SARIMA.

A escolha das variáveis exógenas apropriadas é um passo crítico, que deve ser guiado tanto pela teoria subjacente quanto pela análise de dados. Métodos como análise de correlação e regressão podem ser usados inicialmente para identificar relações potenciais entre as variáveis de interesse e os fatores externos. Além disso, técnicas como a seleção de variáveis baseada em critérios de informação (AIC, BIC) e testes de significância (como o teste t) são fundamentais para avaliar a contribuição de cada variável exógena ao modelo.

### 2.2.7 Erro percentual médio absoluto

Avaliar o desempenho de modelos de séries temporais, como o SARIMAX, em contextos específicos como a previsão de casos de dengue, requer o uso de métricas de precisão que capturem a eficácia das previsões feitas. O MAPE é uma das métricas mais adotadas nesse domínio devido à sua capacidade de fornecer uma avaliação clara e intuitiva do desempenho preditivo dos modelos.

O MAPE é definido como a média dos valores absolutos dos erros percentuais entre os valores observados  $y_i$  e os valores preditos  $\hat{y}_i$ , expresso matematicamente por:

$$MAPE = \frac{100\%}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \quad (9)$$

onde  $n$  é o número total de previsões feitas. Esta métrica tem a vantagem de ser expressa em termos percentuais, facilitando a compreensão e a comparação do desempenho entre diferentes modelos ou séries temporais, independentemente de suas escalas originais.

A relevância do MAPE para a análise de séries temporais decorre de sua capacidade de quantificar a acurácia das previsões de forma intuitiva. Um valor baixo de MAPE indica que as previsões do modelo estão próximas aos valores reais observados,

refletindo um alto grau de precisão. Por outro lado, um MAPE elevado sugere discrepâncias significativas entre as previsões e os valores reais, apontando para a necessidade de revisão ou ajuste do modelo preditivo. Esta característica torna o MAPE especialmente útil em aplicações práticas onde a precisão das previsões é crítica, como na gestão de recursos de saúde pública para o combate à dengue.

### 2.2.8 Metodologia de Box e Jenkins

A metodologia de Box e Jenkins para a análise de séries temporais, particularmente na modelagem de ARIMA e SARIMAX, é um processo estruturado dividido em quatro etapas principais: Identificação, Estimação, Diagnóstico e Previsão. Esta abordagem detalhada desenvolver modelos robustos, capazes de prever fenômenos futuros com precisão. A seguir, será detalhado cada etapa dessa metodologia, enfatizando a importância de seguir um processo sistemático para alcançar resultados preditivos confiáveis.

A fase de identificação é crucial para estabelecer a estrutura básica do modelo ARIMA ou SARIMAX. Inicia-se com uma análise visual dos dados, observando-se gráficos da série temporal para detectar características como tendências, sazonalidades e potenciais outliers. Esta análise preliminar fornece insights valiosos sobre transformações necessárias, como a diferenciação para alcançar a estacionariedade.

A estacionariedade é verificada por meio de testes como o Dickey-Fuller Aumentado (ADF), essenciais para determinar se a série temporal é estacionária ou se precisa ser diferenciada. O termo  $d$  em um modelo ARIMA( $p,d,q$ ) é definido com base nessa análise. Além disso, a análise das ACF e PACF desempenha um papel fundamental na identificação dos termos AR ( $p$ ) e MA ( $q$ ). Os gráficos ACF e PACF indicam o número de defasagens a serem incluídas, orientando a seleção dos parâmetros do modelo.

Após a identificação do modelo, procede-se à estimação dos parâmetros. Geralmente, utiliza-se o método de Máxima Verossimilhança (MLE) ou Mínimos Quadrados Ordinários (OLS) para esta finalidade. A escolha entre MLE e OLS depende das características dos dados e do modelo específico. O objetivo é determinar os valores dos parâmetros  $\phi$  (para componentes AR) e  $\theta$  (para componentes MA) que melhor se ajustam aos dados observados, minimizando a soma dos quadrados dos erros (no caso de OLS) ou maximizando a função de verossimilhança (no caso de MLE).

Esta etapa é crítica, pois parâmetros precisamente estimados são essenciais para a

acurácia e a eficácia do modelo. A estimação adequada permite capturar a dinâmica subjacente dos dados de séries temporais, estabelecendo uma base sólida para previsões futuras. O processo de estimação deve ser cuidadosamente conduzido, garantindo que os parâmetros do modelo reflitam de maneira fiel as relações temporais presentes nos dados.

A etapa de diagnóstico avalia a adequação do modelo ajustado. Uma análise detalhada dos resíduos é realizada para assegurar que se comportem como ruído branco, o que indica que o modelo conseguiu extrair toda a informação relevante dos dados. Gráficos de resíduos, junto com a ACF e PACF dos resíduos, e o teste de Ljung-Box são utilizados para detectar autocorrelações residuais.

Além disso, o teste de normalidade, como o de Shapiro-Wilk, é aplicado para verificar se os resíduos seguem uma distribuição normal, uma suposição comum em muitos modelos de séries temporais. Essas verificações são cruciais para confirmar que o modelo é adequado e não está sobreajustado, garantindo que as previsões sejam

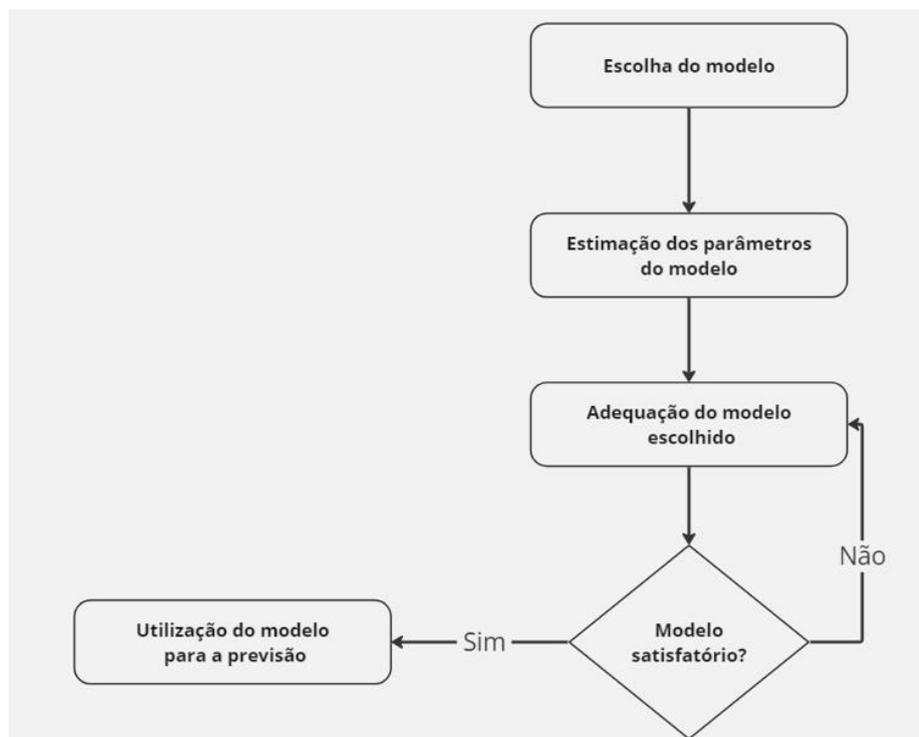
confiáveis. O diagnóstico cuidadoso permite identificar e corrigir potenciais problemas no modelo, assegurando que as previsões sejam baseadas em uma representação precisa dos dados.

A previsão é o objetivo final da modelagem de séries temporais, utilizando o modelo para gerar estimativas futuras. As previsões "out-of-sample" extrapolam os dados históricos, e a incerteza associada a essas previsões é tipicamente refletida nos intervalos de confiança. A validação do modelo, comparando as previsões geradas com os dados observados, é crucial para avaliar sua capacidade preditiva. Métricas como o MAPE são frequentemente utilizadas para quantificar a precisão das previsões, fornecendo uma medida direta da performance do modelo.

Adicionalmente, o diagnóstico do modelo envolve a revisão dos critérios de informação, como AIC e BIC, fundamentais para comparar a adequação de diferentes modelos aos dados. Esses critérios ajudam a selecionar o modelo que oferece o melhor equilíbrio entre complexidade e ajuste aos dados, evitando o sobreajuste e garantindo que o modelo final seja tanto parcimonioso quanto eficaz na captura das dinâmicas dos dados de séries temporais.

A metodologia de Box e Jenkins pode ser representada pela Figura 8 abaixo.

Figura 8 - Fluxograma de Box e Jenkins



Fonte: Autor

### 2.2.9 Aplicações do modelo SARIMA e SARIMAX

O modelo SARIMAX é amplamente utilizado em várias áreas para análise e previsão de séries temporais, a economia é uma área onde o modelo é muito utilizado, (SETHEE, 2020) propôs a aplicação do modelo SARIMAX para a previsão da inflação mensal indiana. Neste estudo, foram exploradas duas abordagens diferentes para prever as taxas de inflação mensais indianas e as taxas bancárias. Para as taxas de inflação, foi aplicado o modelo SARIMAX, uma extensão do ARIMA que incorpora explicitamente o componente sazonal em dados multivariados. Já para as taxas bancárias, utilizou-se o modelo ARIMA com recursos de entrada univariados para períodos futuros.

O estudo desenvolvido por (KHACHATRYAN, 2022) aplicou o modelo SARIMAX para prever as vendas de uma empresa de entrega de alimentos com base em um conjunto de dados de vendas de dois meses gerados em Helsinki, Finlândia. O modelo conseguiu capturar com sucesso os dois picos diários nas vendas, correspondentes aos horários de almoço e jantar, refletindo o comportamento típico de demanda nesse tipo de negócio. No entanto, o modelo apresentou dificuldades em prever vendas acima de 40 pedidos, o que pode indicar que há fatores adicionais influenciando o aumento expressivo de pedidos em certos momentos específicos.

### 3 REVISÃO BIBLIOGRÁFICA

Este capítulo apresenta uma revisão sistemática da literatura relacionada à predição e análise de séries temporais da dengue. A pesquisa foi conduzida com foco em estudos dos últimos cinco anos, refletindo as tendências mais recentes no campo. As bases de dados utilizadas incluem PubMed, Wiley, ProQuest, IEEE e ACM.

#### 3.1 PROTOCOLO DE PESQUISA

O protocolo de pesquisa foi estruturado para incluir trabalhos publicados entre 2019 e 2024, utilizando palavras-chave específicas para capturar os desenvolvimentos mais relevantes no campo da predição da dengue. As palavras-chave usadas foram: "dengue prediction", "dengue" AND "SARIMA", "dengue" AND "SARIMAX", "dengue" AND "forecasting", e "dengue" AND "time series analysis". O filtro foi ajustado para retornar apenas trabalhos dos últimos cinco anos, garantindo a relevância e atualidade dos dados coletados. Na seção seguinte, são apresentadas uma tabela para cada base de dados, que relaciona as palavras-chaves com o número de artigos encontrados.

##### 3.1.1 Pesquisa nas bases de dados

A base de dados PubMed, ver Tabela 1, mostrou um interesse moderado em estudos específicos de predição da dengue e análises de séries temporais com um total de 143 trabalhos encontrados utilizando as palavras-chave especificadas. O modelo SARIMA e SARIMAX são menos explorados, mas ainda presentes na literatura médica e de saúde.

Tabela 1- Pesquisa na base de dados PubMed

Palavras-chave	Resultados
"dengue prediction"	11
"dengue" AND "SARIMA"	9
"dengue" AND "SARIMAX"	1
"dengue" AND "forecasting"	87

"dengue" AND "time series analysis"	35
-------------------------------------	----

Fonte: Autor

Em contraste, Wiley apresentou um maior volume de publicações, especialmente em previsão e análise de séries temporais, indicando uma forte presença interdisciplinar nos estudos sobre dengue, como representa a Tabela 2 abaixo.

Tabela 2 - Pesquisa na base de dados Wiley

Palavras-chave	Resultados
"dengue prediction"	6
"dengue" AND "SARIMA"	7
"dengue" AND "SARIMAX"	1
"dengue" AND "forecasting"	159
"dengue" AND "time series analysis"	55

Fonte: Autor

ProQuest destacou-se com um número alto de publicações, ver Tabela 3, especialmente em trabalhos relacionados à previsão da dengue, refletindo sua ampla cobertura de dissertações e teses que abordam novas abordagens na previsão de surtos.

Tabela 3- Pesquisa na base de dados ProQuest

Palavras-chave	Resultados
"dengue prediction"	178
"dengue" AND "SARIMA"	298
"dengue" AND "SARIMAX"	36
"dengue" AND "forecasting"	6533
"dengue" AND "time series analysis"	1716

Fonte: Autor

IEEE e ACM, com focos mais tecnológicos e computacionais, também mostraram contribuições importantes, embora em menor número, nos estudos de previsão da dengue com mostra respectivamente a Tabela 4 e Tabela 5.

Tabela 4- Pesquisa na base de dados IEEE

Palavras-chave	Resultados
"dengue prediction"	14
"dengue" AND "SARIMA"	3
"dengue" AND "SARIMAX"	1
"dengue" AND "forecasting"	40

"dengue" AND "time series analysis"	27
-------------------------------------	----

Fonte: Autor

Tabela 5 - Pesquisa na base de dados ACM

Palavras-chave	Resultados
"dengue prediction"	2
"dengue" AND "SARIMA"	10
"dengue" AND "SARIMAX"	3
"dengue" AND "forecasting"	58
"dengue" AND "time series analysis"	23

Fonte: Autor

### 3.2 TRABALHOS RELACIONADOS

Esta seção analisa estudos que utilizaram modelos como SARIMA, SARIMAX e outras técnicas de análise de séries temporais na previsão da dengue, explorando como diferentes abordagens modelam e preveem a dinâmica da doença para auxiliar esforços de controle e prevenção.

#### 3.2.1 Modelagem de séries temporais na predição da dengue

O estudo conduzido na República Dominicana (2021) investigou as relações entre clima e transmissão da dengue, aplicando análises estatísticas para vincular diretamente os surtos de dengue a padrões climáticos específicos, como temperatura e umidade. Esta pesquisa, ao identificar períodos críticos em que condições meteorológicas se tornam fatores significativos para o aumento dos casos de dengue, forneceu uma ferramenta valiosa para o planejamento da saúde pública, sugerindo períodos de vigilância intensificada e alocação de recursos para controle de vetores.

Em outro estudo realizado por Othman et al. (2022), foi desenvolvido um modelo SARIMAX para integrar dados climáticos na previsão da incidência de dengue em Surabaya. Este trabalho mostrou que a inclusão de dados climáticos, como temperatura e chuvas intensas, aumenta significativamente a precisão das previsões de dengue. O estudo é particularmente importante pois ilustra a capacidade de realizar previsões mais precisas em ambientes urbanos, permitindo intervenções de saúde pública mais direcionadas e eficazes antes da ocorrência de surtos.

Khaira et al. (2020) aplicaram o modelo SARIMA juntamente com grandes conjuntos de dados geoespaciais para identificar áreas de alto risco de dengue em várias regiões da Ásia. Este estudo destacou como a análise de dados geoespaciais pode complementar as técnicas tradicionais de séries temporais, proporcionando uma camada adicional de precisão na identificação de locais específicos que requerem atenção imediata das autoridades de saúde.

A pesquisa realizada em Manila capital da Filipinas (2022) utilizou o modelo ARIMAX para explorar como mudanças climáticas específicas afetam a incidência de dengue. Descobriu-se que variações sutis na temperatura e na precipitação têm impactos diretos nos casos de dengue, enfatizando a necessidade de incluir esses dados nas estratégias de saúde pública para prever e mitigar surtos.

### **3.2.2 Uso de Transformações e Variáveis Exógenas**

O estudo "Assessing the Effect of Climate Variables on the Incidence of Dengue Cases" (2021) utilizou transformações de dados, como a diferenciação e a transformação logarítmica, para preparar séries temporais de variáveis climáticas para análise. Essas transformações são essenciais para minimizar problemas como tendências e sazonalidade, tornando os dados adequados para modelagem eficaz e ajudando a garantir que as previsões sejam baseadas em dados estáveis e confiáveis.

Naher et al. (2022) destacaram como a inclusão de dados meteorológicos, especificamente temperatura e umidade, no modelo SARIMAX melhora a precisão das previsões de dengue. Esta abordagem não só forneceu previsões mais precisas, mas também ressaltou a importância de integrar fatores ambientais na modelagem da dengue, permitindo que as autoridades de saúde antecipem melhor os surtos e planejem intervenções mais eficazes.

### **3.2.3 Relevância e diferencial do estudo proposto**

Os estudos analisados nas subseções anteriores indicam que, embora os modelos estatísticos como SARIMA e SARIMAX tenham sido amplamente usados para entender a relação entre variáveis climáticas e a incidência de dengue, muitos deles se limitam a identificar correlações, sem avançar para previsões quantitativas do número exato de casos. Essa abordagem, embora útil para compreender padrões temporais e espaciais, deixa um espaço considerável para pesquisas que focam em previsões mais diretas e

quantitativas. A maioria dos estudos tende a explorar como fatores exógenos influenciam a dengue, mas não se aprofunda suficientemente em fornecer previsões precisas e aplicáveis que podem ser utilizadas para intervenções imediatas e planejamento de recursos.

No contexto de Porto Alegre, o projeto proposto representa um avanço na aplicação prática desses modelos estatísticos. A utilização dos modelos SARIMA e SARIMAX irá além de apenas o entendimento ou modelagem da relação entre clima e dengue; ele visa especificamente prever o número exato de casos de dengue. Ao integrar dados climáticos locais e adaptar os modelos para refletir as particularidades climáticas de Porto Alegre, o estudo espera fornecer previsões precisas que podem ser diretamente utilizadas para planejar intervenções de saúde pública, alocar recursos de forma mais eficiente e, em última análise, contribuir para a redução da incidência de dengue na região.

## 4 METODOLOGIA

Neste capítulo, serão detalhados os passos desenvolvidos para a análise dos dados relacionados aos casos de dengue na cidade de Porto Alegre, considerando variáveis exógenas como temperatura, precipitação e umidade. Além disso, serão apresentados dois modelos desenvolvidos e analisados: SARIMA e SARIMAX. A metodologia foi feita em três passos, sendo eles: coleta de dados, pré-processamento e construção do dataset final.

A coleta de dados envolveu a obtenção de informações sobre dengue da Secretaria de Saúde do Estado do Rio Grande do Sul (painel da dengue e SINAN) e dados meteorológicos do INMET, abrangendo o período de 2015 a março de 2024. No pré-processamento, os dados de dengue foram filtrados para Porto Alegre, com a remoção de colunas irrelevantes, criação de uma coluna de data combinando ano e semana epidemiológica, e agrupamento por semana. Os dados meteorológicos foram limpos, tratados para valores ausentes e normalizados.

Finalmente, na construção do dataset final, os dados de dengue e meteorológicos foram integrados, formando um conjunto contínuo e preparado para as análises subsequentes.

### 4.1 COLETA DOS DADOS

Na fase de coleta de dados, é imprescindível recorrer a fontes confiáveis para garantir a qualidade e a consistência das informações, o que é fundamental para uma análise precisa das séries temporais. No contexto específico da análise das séries temporais envolvendo dados de dengue e dados climáticos em Porto Alegre, adotamos uma abordagem meticulosa para obter o conjunto de dados.

Os dados sobre a dengue foram obtidos por meio de uma fonte confiável e autorizada: a Secretaria de Saúde do Estado do Rio Grande do Sul. Essa instituição desempenha um papel central na coleta e no monitoramento de dados relevantes para a saúde pública, incluindo informações sobre a dengue. Utilizamos o painel da dengue, uma ferramenta disponibilizada pela Secretaria de Saúde que oferece acesso a informações

atualizadas e significativas sobre a doença.

Esses dados têm origem na base de dados do SINAN, uma plataforma essencial para a vigilância epidemiológica. O SINAN possibilita a coleta, o registro e a análise de informações relacionadas a doenças notificadas, incluindo a dengue. Esse sistema desempenha um papel crucial no acompanhamento da incidência da doença, na identificação de áreas de maior prevalência, no rastreamento de tendências temporais e no embasamento de ações de prevenção e controle.

Os dados coletados sobre a dengue abrangem o período desde o início do ano de 2015 até março do ano de 2024 e estão concentrados na cidade de Porto Alegre, fornecendo uma visão detalhada da situação da dengue ao longo do tempo nessa região específica, foi definido esse período pois não foi encontrados dados de dengue anteriormente ao ano de 2015.

Quanto aos dados meteorológicos, foram obtidos por meio do INMET, uma instituição responsável pela coleta, processamento e fornecimento de informações meteorológicas no Brasil. O INMET é vinculado ao Ministério da Agricultura, Pecuária e Abastecimento e desempenha um papel crucial na geração de dados meteorológicos para diversas finalidades.

Para obter esses dados, se seguiu o procedimento padrão de solicitação via e-mail junto ao INMET. Essas informações foram disponibilizadas mediante solicitação formal, garantindo assim sua confiabilidade e validade.

Os dados meteorológicos obtidos estão referenciados à base de Porto Alegre e abrangem o mesmo período dos dados relacionados à dengue. Esses dados incluem informações como temperatura, umidade, precipitação, velocidade e direção do vento, permitindo uma análise abrangente das condições meteorológicas na região durante o período analisado.

## 4.2 PRÉ-PROCESSAMENTO DOS DADOS

O pré-processamento dos dados envolve várias técnicas fundamentais que preparam o conjunto de dados para análise, garantindo que as informações sejam limpas, completas e prontas para serem eficazmente analisadas. Na área de análise de dados, especialmente no contexto da saúde pública e estudos sobre doenças como a dengue, a utilização do Python se destaca como uma ferramenta poderosa. O Python, com sua vasta gama de

bibliotecas e frameworks, oferece recursos inestimáveis para o pré-processamento de dados, contribuindo significativamente para a precisão e eficácia das análises de séries temporais.

Uma das principais vantagens do Python na análise de dados é a sua biblioteca “Pandas”, que fornece estruturas de dados flexíveis e métodos de manipulação de dados projetados para facilitar a limpeza, filtragem e transformação de grandes conjuntos de dados. Por exemplo, na preparação de dados sobre a dengue, “Pandas” pode ser utilizado para tratar valores ausentes, seja através da imputação de dados, remoção de registros incompletos ou interpolação, dependendo do contexto e da natureza dos dados. Além disso, a biblioteca permite a manipulação de séries temporais com facilidade, incluindo alteração de frequências, o deslocamento de dados para criar lag features importantes para análises preditivas e a agregação de dados em diferentes intervalos de tempo, fundamentais para capturar padrões sazonais da doença.

Além do “Pandas”, o Python oferece a biblioteca “NumPy” para operações numéricas avançadas e “Matplotlib” e “Seaborn” para visualização de dados, permitindo uma análise exploratória eficaz que pode revelar insights críticos sobre a dinâmica da transmissão da dengue. Essas visualizações podem ajudar a identificar tendências, padrões sazonais e outliers que podem ser cruciais para o desenvolvimento de modelos preditivos precisos.

Para o pré-processamento específico de séries temporais, a biblioteca “StatsModels” proporciona funcionalidades para a análise e diagnóstico de séries temporais, oferecendo ferramentas para a decomposição de séries, análise de estacionariedade e a determinação de parâmetros iniciais para modelos ARIMA e SARIMA, diretamente relevantes para estudos de previsão da dengue.

#### **4.2.1 Pré-processamento dos dados da dengue**

Durante a etapa de pré-processamento dos dados de dengue, foi adotado uma abordagem sistemática para preparar os dados para análise. Inicialmente, realizou-se uma análise detalhada da estrutura geral dos dados para entender melhor sua composição e identificar possíveis áreas de foco.

A primeira fase do nosso pré-processamento consistiu em uma análise detalhada da estrutura geral dos dados relativos à dengue. A análise revelou um *dataset* com um total

de 81.667 entradas, distribuídas em 15 colunas distintas. Essa extensiva coleção de dados se estende do índice 0 ao 81.666, sugerindo uma abrangente captação de registros ao longo do tempo. As colunas são diversificadas, abrangendo desde informações temporais, como o ano e a semana epidemiológica, até dados demográficos e clínicos, incluindo a faixa etária, sexo, e categorias relativas ao estado dos casos (confirmados, descartados, etc.).

A presença de 11 colunas do tipo inteiro (int64) indica uma predominância de variáveis quantitativas, as quais incluem anos, códigos geográficos, contagens de casos (confirmados, óbitos, descartados, etc.), enquanto 4 colunas são do tipo objeto, sugerindo dados categóricos como nome do município, sexo, faixa etária e raça/cor. A grande maioria dos dados está completa, com exceção da coluna 'Sexo', que apresenta uma pequena quantidade de valores ausentes (23 entradas faltantes).

Esta análise inicial é fundamental, pois nos permite visualizar a composição e complexidade dos dados, orientando a seleção de variáveis relevantes para focar a investigação nos padrões temporais e demográficos dos casos de dengue, ver Tabela 6 que representa a estrutura do *dataset*.

Tabela 6 - Estrutura do dataset da dengue

Coluna	Tipo de Dado
Ano	int64
Semana Epidemiológica	int64
Cód IBGE	int64
CRS	int64
Nome Município	object
Sexo	object
Faixa Etária	object
Raça Cor	object
Óbitos	int64
Investigação	int64
Descartados	int64
Inconclusivos	int64
Autóctones	int64
Confirmados	int64
Notificações	int64

Fonte: Autor

Essa tabela resume as colunas disponíveis no conjunto de dados, junto com os tipos

de dados associados a cada uma, facilitando a compreensão de como os dados estão organizados e que tipo de informações cada coluna representa. A variedade de tipos de dados reflete a riqueza das informações capturadas, desde identificadores e contagens numéricas até categorias descritivas, todas cruciais para uma análise detalhada e abrangente dos casos de dengue.

Após a análise inicial, aplicou-se um filtro para selecionar apenas os dados relacionados à cidade de Porto Alegre, focando a análise nessa região específica. Em seguida, foi removido as colunas irrelevantes, mantendo apenas aquelas que eram essenciais para a análise dos casos confirmados de dengue. Essas colunas incluíam os casos confirmados, a semana epidemiológica e o ano correspondente. Essa seleção permitiu concentrar a análise nos aspectos temporais dos casos confirmados de dengue, possibilitando a observação de sua evolução ao longo das semanas e anos.

Após a filtragem por cidade e seleção de variáveis relevantes, foi definido o ano base como 2014 e, em seguida, iterou-se sobre um período de nove anos, aumentando progressivamente o ano em cada iteração. Para cada ano no intervalo, executou-se os seguintes procedimentos sobre o *dataset* original:

- a) Filtragem anual: selecionando os registros correspondentes ao ano em questão, reduzindo o conjunto de dados a apenas aqueles relevantes para o ano especificado.
- b) Seleção de colunas: mantendo somente as colunas 'Semana', 'Ano', e 'Confirmados', que são cruciais para a análise temporal dos casos de dengue.
- c) Conversão de tipos de dados: As colunas 'Ano' e 'Semana' foram convertidas para o tipo inteiro para assegurar a consistência nos cálculos subsequentes.
- d) Criação da coluna de data: introduzindo uma nova coluna denominada 'data', que combina as informações de ano e semana epidemiológica para formar uma data representativa da semana em questão. Esta coluna facilita a análise temporal ao proporcionar um índice temporal claro.
- e) Agrupamento e soma: Foi agrupado os dados pela nova coluna 'data' e somados os valores dos casos confirmados para cada grupo. Este passo é fundamental para consolidar os casos por semana, permitindo uma visão acumulativa dos casos ao longo do tempo.

Finalmente, foi concatenado os *datasets* resultantes de cada ano em um único *dataset*, criando uma série temporal contínua de casos confirmados de dengue. Este

*dataset* consolidado serve como a base para nossa análise subsequente, permitindo uma investigação detalhada das tendências e padrões dos casos de dengue em Porto Alegre ao longo do tempo. A inclusão da coluna 'data' como índice facilita operações de análise temporal e visualizações.

O *dataset* final se caracteriza por ter um índice temporal (DatetimeIndex), que vai de 29 de dezembro de 2014 a 29 de março de 2024, cobrindo assim um intervalo significativo que permite observar a evolução da dengue ao longo de diversos anos. Esse índice facilita operações como filtragem, agrupamento e visualização baseadas em períodos específicos.

O *dataset* contém 410 entradas, cada uma correspondendo a uma semana dentro do intervalo de tempo considerado. Essa estruturação proporciona uma granularidade ideal para análise temporal, oferecendo insights sobre a sazonalidade e o impacto de intervenções de saúde pública nos casos de dengue.

Cada entrada no *dataset* tem um único valor numérico associado, representando o número de casos confirmados de dengue na semana correspondente. Os dados são do tipo inteiro (int64), indicando que todos os valores são numéricos e aptos para operações estatísticas e matemáticas.

Tabela 7 - "Head" do dataset final de dengue

data	Confirmados
2014-12-29	0
2015-01-05	1
2015-01-12	2
2015-01-19	0
2015-01-26	0

Fonte: Autor

Este exemplo inicial do *dataset*, ver Tabela 7, mostra as primeiras semanas de análise, iniciando com zero casos confirmados na última semana de 2014 e observando-se os primeiros casos nas semanas seguintes de 2015.

#### 4.2.2 Pré-processamento dos dados meteorológicos

Durante o pré-processamento dos dados meteorológicos, foi necessário adotar uma abordagem cuidadosa para garantir a qualidade e a relevância dos dados para análises subsequentes. Os dados meteorológicos foram inicialmente organizados em arquivos

anuais contendo registros horários das condições climáticas. A seguir, descrevemos os passos adotados nessa etapa.

A análise da estrutura dos dados meteorológicos revelou um *dataset* composto por 3133 entradas, cada uma correspondendo a medições diárias das condições climáticas. Este conjunto de dados inclui um total de seis colunas, todas inicialmente do tipo 'object', o que sugere que podem conter tantos valores numéricos quanto textuais, embora que sejam predominantemente numéricos.

A seguir, é detalhado as colunas presentes no *dataset* e seus respectivos tipos de dados através da Tabela 8.

Tabela 8 - Estrutura do dataset meteorológicos

Coluna	Tipo de Dado
Data Medicao	object
PRECIPITACAO TOTAL, DIARIO(mm)	object
TEMPERATURA MAXIMA, DIARIA(°C)	object
TEMPERATURA MEDIA COMPENSADA, DIARIA(°C)	object
TEMPERATURA MINIMA, DIARIA(°C)	object
UMIDADE RELATIVA DO AR, MEDIA DIARIA(%)	object

Fonte: Autor

Esta estrutura indica que todas as colunas foram carregadas como strings (tipo 'object'), o que é comum quando os dados são importados de arquivos CSV ou fontes de texto sem especificação explícita do tipo de dado. Para proceder com análises estatísticas e numéricas, será necessário converter as colunas relacionadas a medições (precipitação, temperaturas e umidade) de 'object' para tipos numéricos, como 'float' ou 'int', dependendo da precisão requerida.

A presença de valores não nulos em todas as entradas para cada coluna sugere que não há ausências imediatas nos dados, o que é um bom indicativo da qualidade do *dataset*. Este entendimento inicial é importante para orientar as próximas etapas de limpeza e tratamento de dados, garantindo que as análises futuras estejam baseadas em informações precisas e representativas das condições meteorológicas.

Inicialmente, as colunas do *dataset* foram renomeadas para facilitar a manipulação e interpretação dos dados. As novas denominações foram escolhidas para refletir de forma

concisa o conteúdo de cada coluna.

Após a renomeação, procedeu-se com a conversão dos tipos de dados. As colunas de temperatura, precipitação e umidade, originalmente registradas como strings, foram convertidas para o formato numérico (float). Isso envolveu a substituição de vírgulas por pontos para corrigir o formato decimal e permitir a conversão adequada. Além disso, a coluna 'data' foi convertida para o tipo datetime, facilitando a manipulação de datas e a realização de cálculos baseados em tempo.

Foram criadas duas novas colunas, 'Semana' e 'Ano', utilizando a coluna 'data'. Essas colunas representam, respectivamente, o número da semana do ano e o ano de cada medição, permitindo uma agregação dos dados com base em períodos semanais.

Com as novas colunas temporais, os dados foram agrupados por 'Ano' e 'Semana'. Para cada grupo semanal, calculou-se a média das variáveis de temperatura e umidade, e a soma total da precipitação. Esse passo reduziu a granularidade dos dados de diária para semanal, alinhando-os com os dados dos casos de dengue para análises comparativas.

Após o agrupamento, a coluna de índice foi ajustada para refletir as datas correspondentes ao início de cada semana representada, utilizando a combinação do ano e número da semana. Isso facilitou a interpretação temporal dos dados agregados e permitiu a integração direta com os dados de dengue baseados na mesma estrutura temporal.

Posteriormente as manipulações detalhadas, o *dataset* de dados meteorológicos foi transformado para refletir informações semanais agregadas das condições climáticas, incluindo temperatura (máxima, média, mínima), umidade relativa do ar e precipitação total. Cada linha do *dataset* representa uma semana, fornecendo uma média para as temperaturas e umidade, e a soma total da precipitação para aquela semana. A coluna 'data' atua como índice, garantindo que os dados sejam organizados cronologicamente, como é apresentado na Figura 9.

Figura 9 - "Head" do dataset meteorológico

data	temp_max	temp_media	temp_min	umidade	precipitacao
29/12/2014	29.625	24.070	19.800	750.625	10.1
05/01/2015	22.157	16.628	13.557	81.107	3.0
12/01/2015	29.843	25.294	22.400	80.286	47.4
19/01/2015	30.143	24.703	19.929	72.321	26.1
26/01/2015	30.171	24.940	21.457	74.964	31.6

Fonte: Autor

### 4.2.3 Engenharia de dados e construção do dataset final

Neste tópico, é detalhado as metodologias empregadas na engenharia de dados e na preparação do conjunto de dados final, que serve como alicerce para o desenvolvimento de um modelo preditivo robusto. As etapas descritas a seguir são essenciais para assegurar a qualidade e a integridade dos dados, permitindo análises precisas e resultados confiáveis na previsão de casos de dengue.

O processo começou com a coleta e a integração de conjuntos de dados distintos, abrangendo informações de casos de dengue e dados climáticos. Essa integração foi realizada para cada par de *datasets* correspondente a um ano específico de estudo. Primeiramente, se neutralizou a prioridade de índices específicos, permitindo que a fusão dos dados fosse baseada exclusivamente na coluna "data", garantindo assim uma correspondência exata entre os eventos de dengue e as condições meteorológicas relevantes. Após a fusão, a coluna "data" foi reestabelecida como índice principal do *dataset* consolidado, criando uma estrutura de dados temporalmente ordenada e pronta para análises subsequentes. A agregação desses *datasets* formou um conjunto de dados completo, rico em variáveis meteorológicas e epidemiológicas, fundamentais para o estudo.

Este conjunto unificado de dados abrange o período de 29 de dezembro de 2014 até 29 de março de 2024. Esse período de tempo oferece uma base sólida para análises e modelagem preditiva, abrangendo várias temporadas de dengue e condições climáticas associadas. As variáveis do *dataset* unificado estão detalhadas na Tabela 9 abaixo.

Tabela 9 - Estrutura do dataset final

Coluna	Descrição
Confirmados	Casos confirmados de dengue por semana
temperatura_max	Temperatura máxima semanal
temperatura_media	Temperatura média semanal
temperatura_min	Temperatura mínima semanal
umidade	Umidade relativa semanal
precipitacao	Volume de precipitação semanal

Fonte: Autor

Ao examinar a estrutura do *dataset*, notamos a presença de valores faltantes especificamente na coluna "Confirmados", com 413 entradas não nulas de um total

possível de 458, abaixo é apresentado por meio da Figura 10 as primeiras linhas do *dataset* final.

Figura 10 - "Head" do dataset final

<b>data</b>	<b>Confirmados</b>	<b>temp_max</b>	<b>temp_media</b>	<b>temp_min</b>	<b>umidade</b>	<b>precipitacao</b>
29/12/2014	0.0	29.625	24.070	19.800	750.625	10.1
05/01/2015	1.0	22.157	16.628	13.557	811.071	3.0
12/01/2015	2.0	29.843	25.294	22.400	802.857	47.4
19/01/2015	0.0	30.143	24.703	19.929	723.214	26.1
26/01/2015	0.0	30.171	24.940	21.457	749.643	31.6

Fonte: Autor

A Tabela 10 a seguir apresenta uma análise descritiva estatística das variáveis do dataset final. Os valores incluem a média (mean), desvio padrão (std), valor mínimo (min), percentis (25%, 50%, 75%) e valor máximo (max) para cada variável.

Tabela 10 - Estatística descritiva do dataset final

<b>Variável</b>	<b>Média (mean)</b>	<b>Desvio Padrão (std)</b>	<b>Mínimo (min)</b>	<b>25% Percentil</b>	<b>50% Percentil</b>	<b>75% Percentil</b>	<b>Máximo (max)</b>
Confirmados	27.975	102.693	0.000	0.000	0.000	7.000	810.000
Temperatura Máx (°C)	26.248	4.629	13.000	22.682	26.507	30.114	36.157
Temperatura Média (°C)	20.542	4.267	8.989	17.135	21.114	24.145	28.597
Temperatura Mín (°C)	16.520	4.121	5.257	13.082	16.871	19.985	24.129
Umidade (%)	76.726	6.468	54.857	72.313	77.089	81.438	92.821
Precipitação (mm)	29.964	33.806	0.000	4.900	16.300	42.550	187.000

Fonte: Autor

O número de casos confirmados de dengue apresenta uma alta variabilidade, como indicado pelo desvio padrão (102.693) em comparação com a média (27.975). A maioria das observações tem 0 casos confirmados (mediana = 0), enquanto o valor máximo é de 810 casos confirmados. As temperaturas máximas variam de 13.0°C a 36.157°C, com uma média de 26.248°C e um desvio padrão de 4.629°C, indicando uma variação moderada nas temperaturas máximas. A temperatura média tem um intervalo de 8.989°C a 28.597°C, com uma média de 20.542°C e um desvio padrão de 4.267°C. Isso indica que a temperatura média varia moderadamente. As temperaturas mínimas variam de 5.257°C a 24.129°C, com uma média de 16.520°C e um desvio padrão de 4.121°C, sugerindo que as temperaturas mínimas são relativamente mais estáveis em comparação com as máximas.

A umidade relativa apresenta valores que vão de 54.857% a 92.821%, com uma média de 76.726% e um desvio padrão de 6.468%. A variação na umidade é relativamente baixa, com a maioria das observações concentradas entre 72.313% e 81.438%. A precipitação varia significativamente de 0 mm a 187 mm, com uma média de 29.964 mm e um desvio padrão de 33.806 mm. A alta variabilidade na precipitação é indicativa de períodos secos e chuvosos.

Durante a análise exploratória da pesquisa, o foco foi aprimorar a variável de casos confirmados de dengue, importante para a modelagem preditiva. Dada a natureza muitas vezes não linear e não normal dos dados epidemiológicos, foi empregado uma série de transformações matemáticas na tentativa de normalizar a distribuição dessa variável, facilitando a aplicação e interpretação de modelos estatísticos como o SARIMA. Conforme discutido em consulta com a inteligência artificial ChatGPT da OpenAI (OPENAI, 2024), abaixo estão descritas as transformações testadas:

- a) Ajuste logarítmico é uma transformação básica que aplica o logaritmo natural aos valores, útil para dados que seguem uma progressão exponencial. Este método tende a diminuir a variação dos dados, tornando distribuições assimétricas mais simétricas.
- b) Método Box-Cox sendo uma transformação mais sofisticada que busca identificar o parâmetro lambda que melhor normaliza os dados. É particularmente eficaz para dados positivos, ajustando-se a diferentes graus de assimetria.
- c) Transformação de Tukey conhecida por sua flexibilidade, essa transformação, às vezes chamada de "twiddle", pode lidar com uma gama mais ampla de distribuições, aplicando potências variadas aos dados, tanto positivas quanto negativas.
- d) Transformação de YeoJohnson que é uma extensão do método Box-Cox que pode ser aplicada a dados zero e negativos, tornando-a uma escolha versátil para uma variedade de conjuntos de dados.

A eficácia de cada transformação foi avaliada inserindo a série temporal transformada em um modelo SARIMA. A transformação de YeoJohnson destacou-se como a mais eficaz, conforme evidenciado pelos valores de p-value obtidos nos testes de Shapiro-Wilk e D'Agostino's K-squared. Essa análise das transformações e os subsequentes testes de normalidade são importantes para preparar os dados para a

modelagem preditiva, assegurando que as previsões sejam mais precisas possível.

Com a transformação ótima definida, foi prosseguido para a etapa de preparação final dos dados. Este estágio envolveu a remoção de registros faltantes e a aplicação da transformação de Yeo-Johnson na variável "Confirmados", preparando o *dataset* para análises preditivas. A conclusão dessa fase produziu um conjunto de dados limpo, transformado e integralmente pronto para ser dividido em conjuntos de treinamento e teste. Seguindo critérios estratégicos de temporalidade para preservar a validação do modelo, os dados foram divididos, destinando as informações de 2015 a 2022 para o treinamento e as de 2023 a 2024 para teste. Essa divisão simula um cenário real de previsão para avaliar a eficácia do modelo SARIMAX, minimizando riscos de overfitting e maximizando a capacidade de generalização do modelo.

## **5 RESULTADOS E DISCUSSÕES**

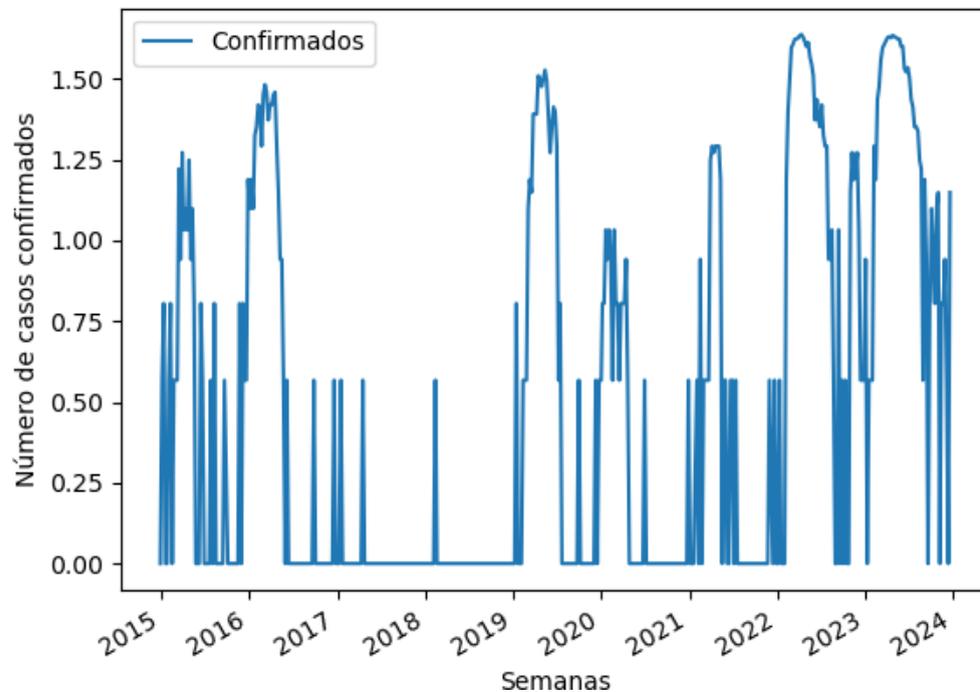
### **5.1 SELEÇÃO DO MODELO SARIMA**

A presente seção objetiva identificar um modelo estatístico inicial apropriado para a predição do número de casos de dengue utilizando-se séries temporais. Através da análise dos padrões observados na série temporal de dados e da verificação de sua estacionariedade, foi possível estabelecer uma metodologia inicial para a modelagem SARIMA. As ferramentas de autocorrelação (ACF) e autocorrelação parcial (PACF) foram essenciais para a identificação dos termos autoregressivos e de médias móveis necessários para a construção do modelo.

#### **5.1.1 Proposta do modelo inicial**

A série temporal de casos confirmados de dengue, apresentando um período total de 10 anos, exibe variações que sugerem um componente sazonal evidente, o que é característico em dados epidemiológicos como os da dengue, acompanhado por oscilações de curto prazo que poderiam indicar uma influência autoregressiva e de médias móveis. A visualização gráfica da série de acordo com a Figura 11, permite uma avaliação inicial de suas características, incluindo tendência e flutuações periódicas.

Figura 11 - Série temporal transformada

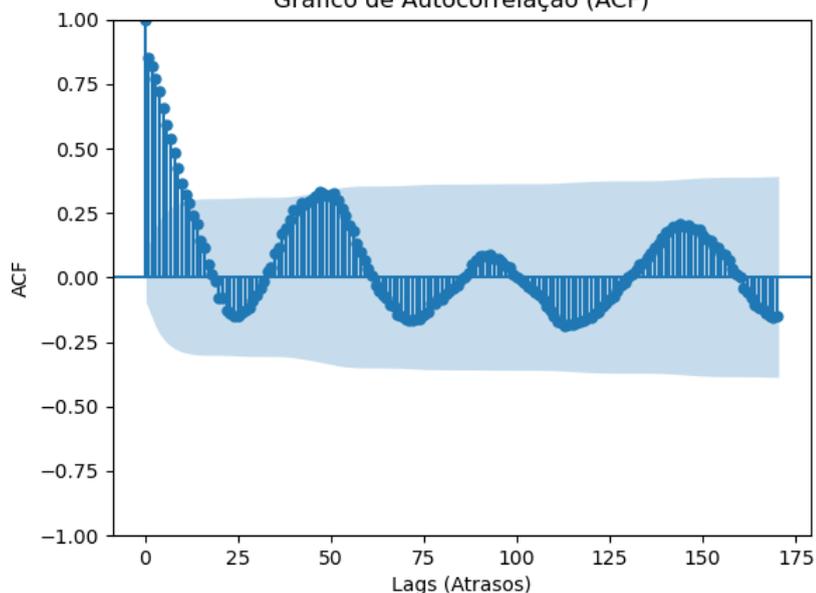


Fonte: Autor

Foi aplicado o teste de Dickey-Fuller Aumentado (ADF) para determinar se a série temporal era estacionária. Os resultados indicaram uma estatística de teste de  $-3.731$  e um valor-p de  $0.003$ , com valores críticos para 1%, 5% e 10% de  $-3.446$ ,  $-2.868$  e  $-2.570$ , respectivamente. A hipótese nula de não estacionariedade foi rejeitada a todos os níveis de confiança convencionais. Isso implica que a série temporal é estacionária e não requer diferenciação não sazonal.

O gráfico de Autocorrelação (ACF), Figura 12, para a série temporal indica que há correlações significativas em diversos lags. No entanto, essas correlações caem dentro da área de confiança após o primeiro lag, o que sugere que o componente de médias móveis (MA) do modelo SARIMA pode ser relativamente simples.

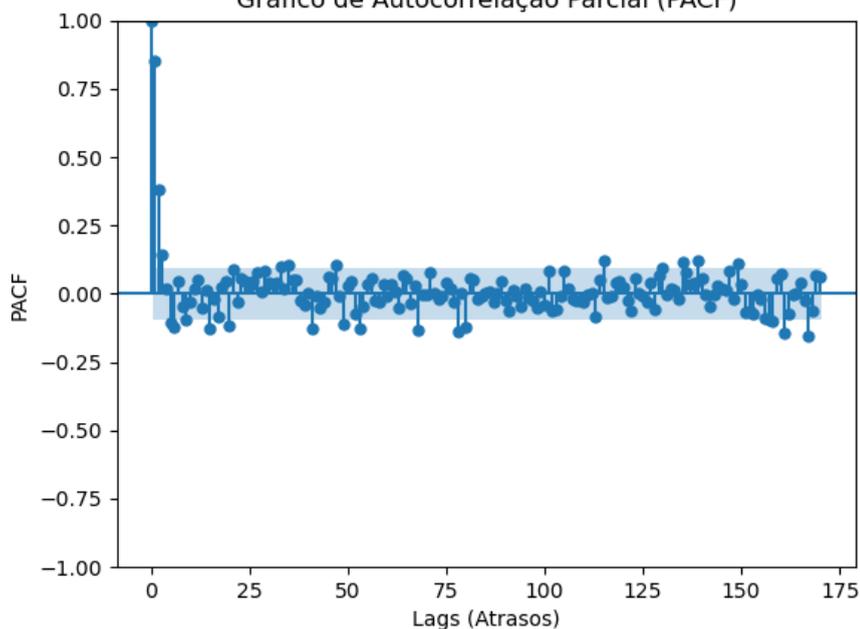
Figura 12 - Gráfico de autocorrelação  
Gráfico de Autocorrelação (ACF)



Fonte: Autor

Analogamente, o gráfico de Autocorrelação Parcial (PACF), Figura 13, mostra uma queda acentuada após o primeiro lag, o que é típico de um processo autoregressivo de ordem 1, AR(1).

Figura 13 - Gráfico de autocorrelação parcial  
Gráfico de Autocorrelação Parcial (PACF)



Fonte: Autor

Com base nas análises anteriores, um modelo inicial SARIMA(1,0,1)(1,0,1)<sub>53</sub> pode ser proposto. O componente (1,0,1) reflete a estrutura autoregressiva e de médias móveis identificada na PACF e ACF, respectivamente. A parte sazonal (1,0,1)<sub>53</sub> aborda a sazonalidade anual evidente na série, com o número 53 representando a periodicidade

semanal dentro de um ano. Esse modelo será ajustado aos dados para avaliar sua capacidade de capturar a dinâmica da série temporal. O ajuste do modelo e os diagnósticos dos resíduos subsequente determinarão a adequação do modelo e a necessidade de possíveis refinamentos.

### 5.1.2 Seleção da melhor ordem para o modelo SARIMA

Este tópico aborda a seleção da ordem ótima do modelo SARIMA para previsão semanal dos casos de dengue, baseada na metodologia de Box e Jenkins. Considerando as restrições computacionais e a importância de garantir a convergência do algoritmo, uma ordem fixa foi adotada para o componente de diferenciação sazonal ( $D=1$ ). A seguir, detalha-se o procedimento utilizado para determinar as ordens não sazonais ( $p$ ,  $d$ ,  $q$ ) e sazonais ( $P$ ,  $Q$ ) que resultam no modelo com o melhor ajuste, conforme indicado pelo AIC.

Na busca pela estrutura ótima do modelo SARIMA, conduziu-se um procedimento de grid search. Avaliou-se uma gama de possíveis ordens para os componentes do modelo, tanto na dimensão não sazonal quanto na sazonal. As ordens avaliadas para cada componente do modelo foram:

- a) Componente Autoregressivo não Sazonal ( $p$ ): Foi variado de 0 a 3, possibilitando avaliar modelos desde os mais simples, sem termos autoregressivos, até modelos que capturam uma autocorrelação de até três lags anteriores.
- b) Componente de Diferenciação não Sazonal ( $d$ ): Explorou-se com ordens 0 e 1, para verificar se a série necessita de diferenciação para alcançar a estacionariedade.
- c) Componente de Médias Móveis não Sazonal ( $q$ ): Também variou-se de 0 a 3, permitindo que o modelo incorporasse de nenhuma até três médias móveis para capturar os choques aleatórios nos dados.

Para os componentes sazonais, considerou-se a periodicidade semanal dos dados e testaram-se as seguintes ordens:

- a) Componente Autoregressivo Sazonal ( $P$ ): Similar ao componente não sazonal, variou-se de 0 a 3, buscando capturar a autocorrelação sazonal correspondente a até um ano anterior.
- b) Componente de Diferenciação Sazonal ( $D$ ): A ordem foi fixada em 1, considerando a necessidade de estabilizar a variação sazonal na série, uma

escolha guiada por restrições computacionais que impactavam a convergência do algoritmo.

- c) Componente de Médias Móveis Sazonal (Q): Assim como os componentes P e Q não sazonais, testou-se uma variação de 0 a 3, permitindo avaliar o impacto de choques aleatórios nos padrões sazonais da série.

A Tabela 11 a seguir resume as melhores combinações de ordens encontradas através do procedimento de grid search, listadas de acordo com seus valores de AIC.

Tabela 11 - Melhores modelos SARIMA

p	d	q	P	D	Q	AIC
3.0	0.0	3.0	1.0	1.0	0.0	73.41268534229607
2.0	1.0	2.0	1.0	1.0	0.0	73.59074660732811
2.0	0.0	3.0	1.0	1.0	0.0	73.64182582461687
3.0	0.0	3.0	0.0	1.0	1.0	73.6454021448384
2.0	1.0	2.0	0.0	1.0	1.0	73.71409399725479

Fonte: Autor

Com base no AIC, o modelo SARIMA(3,0,3)(1,1,0)<sub>53</sub> demonstrou ser o “melhor” modelo. Esse modelo sugere que três componentes autoregressivos e três de médias móveis no ciclo não sazonal, juntamente com um componente autoregressivo sazonal e uma diferenciação sazonal, poderiam capturar adequadamente a dinâmica da série temporal dos casos de dengue.

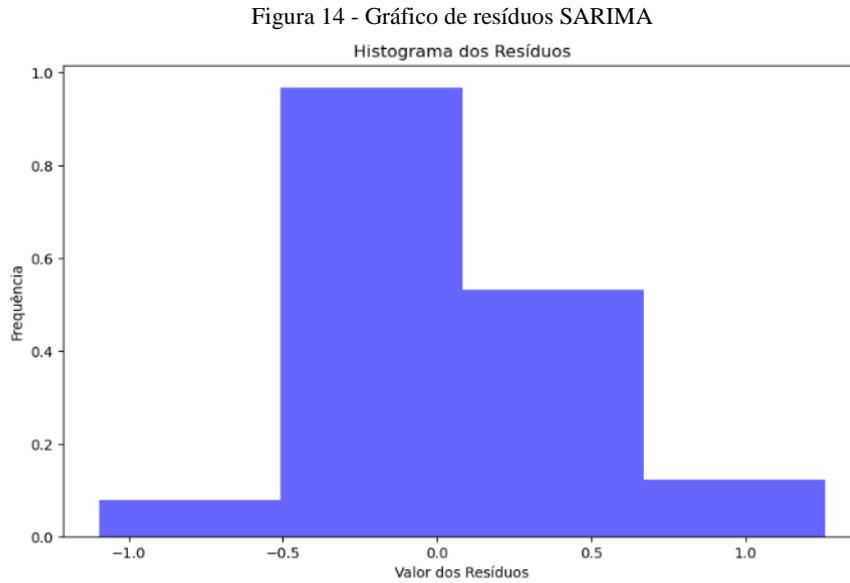
Os resultados obtidos pelo procedimento de grid search indicam que um modelo SARIMA de ordens mais elevadas em p e q tende a se ajustar melhor à complexidade dos dados observados. A decisão de fixar o componente D foi uma resposta pragmática aos desafios computacionais enfrentados, destacando a importância da verificação subsequente dos resíduos para validar a adequação do modelo selecionado.

### 5.1.3 Avaliação do modelo SARIMA

Neste tópico, conduz-se uma investigação aprofundada do modelo SARIMA(3, 0, 3)(1, 1, 0)<sub>53</sub> para entender sua capacidade de modelar a série temporal de casos confirmados de dengue. Esta avaliação explora os resíduos resultantes para garantir a validade do modelo escolhido.

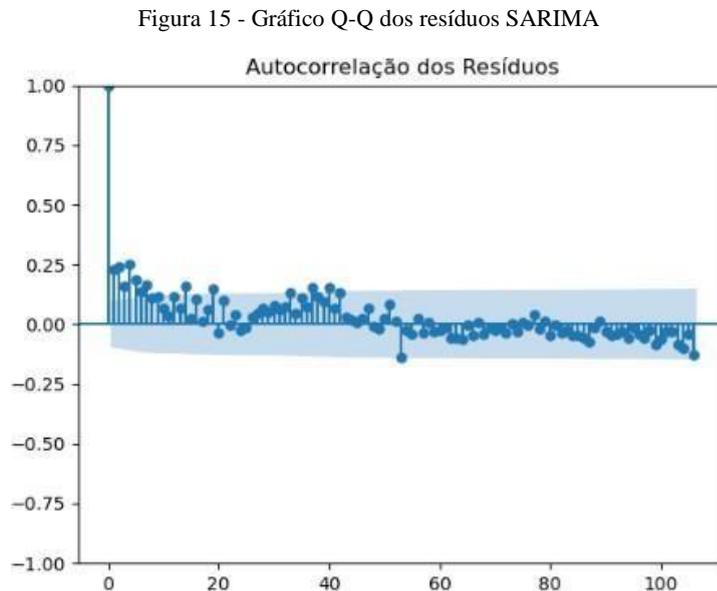
O histograma dos resíduos, representado através da Figura 14, apresentando apenas quatro barras, mostra uma concentração significativa em torno do centro, indicando que muitos resíduos são próximos a zero, o que é uma indicação de que o modelo está fazendo previsões próximas à realidade. No entanto, a distribuição não segue o formato de sino

esperado de uma distribuição normal, apontando para a possibilidade de que o modelo não esteja totalmente capturando todos os padrões subjacentes nos dados.



Fonte: Autor

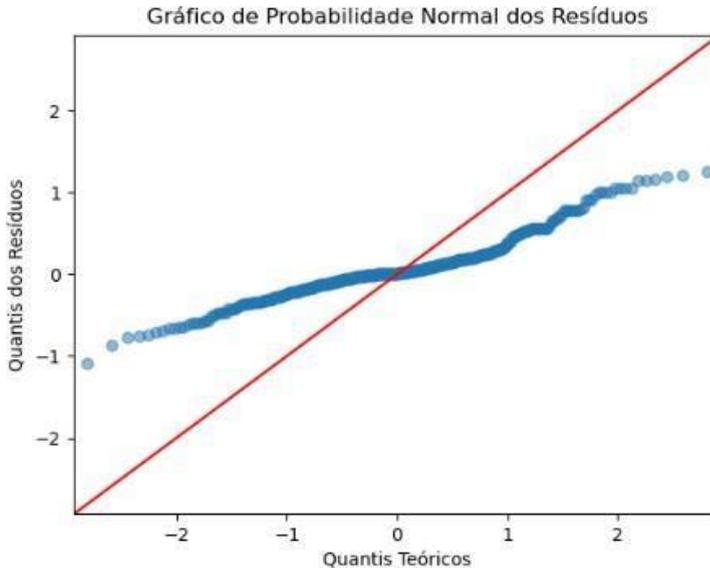
Observando o gráfico de autocorrelação dos resíduos na Figura 15, nota-se que todas as barras estão dentro do intervalo de confiança, sugerindo que os resíduos são independentes entre si. Isso é crucial, pois resíduos correlacionados poderiam indicar que informações ainda estão presentes nos resíduos que o modelo não capturou.



Fonte: Autor

A Figura 16 representa o Q-Q Plot dos resíduos que destaca uma concordância geral com a linha vermelha diagonal, que representa a distribuição normal teórica. No entanto, há desvios perceptíveis nas extremidades, especialmente no topo do gráfico, o que pode indicar a presença de outliers ou de caudas pesadas na distribuição dos resíduos.

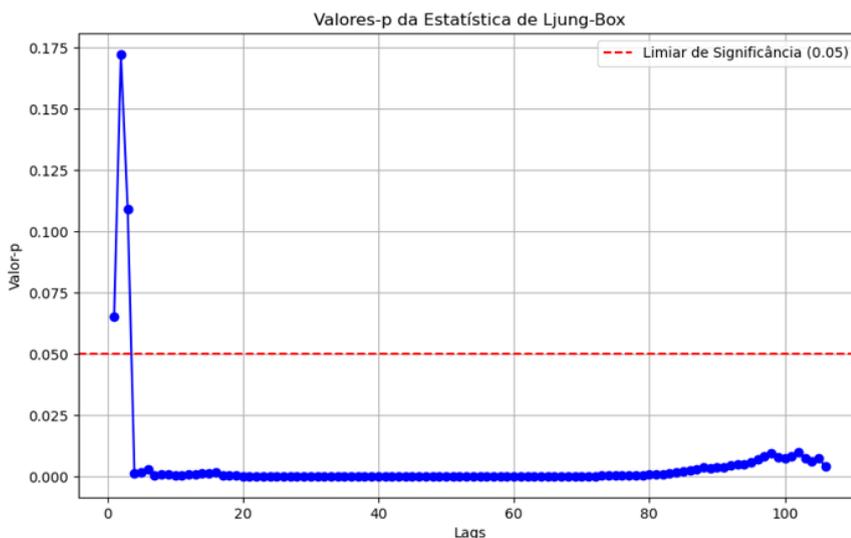
Figura 16 - Gráfico de probabilidade normal dos resíduos SARIMA



Fonte: Autor

O gráfico dos valores-p para a estatística de Ljung-Box que testa a hipótese nula de que os resíduos de uma série temporal são independentes, ou seja, não autocorrelacionados. Os pontos abaixo da linha vermelha tracejada representam lags com autocorrelação significativa a um nível de confiança de 5%. Aqui, a maioria dos valores-p está acima do limiar, sugerindo que, para a maior parte, os resíduos são independentes. Contudo, os primeiros lags abaixo do limiar indicam que pode haver alguma autocorrelação significativa a curto prazo que o modelo não capturou, como mostra a Figura 17 abaixo.

Figura 17 - Gráficos de valores-p SARIMA

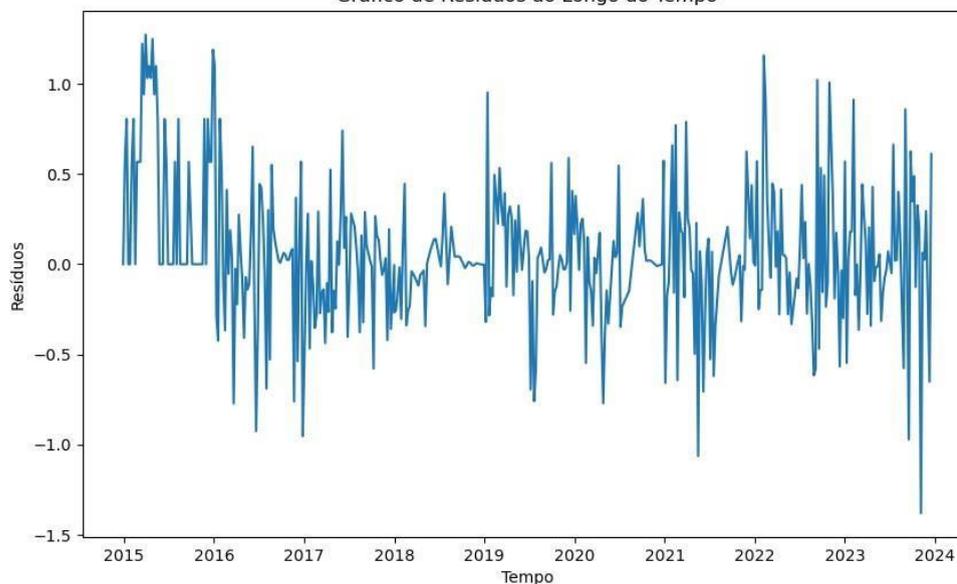


Fonte: Autor

O gráfico dos resíduos ao longo do tempo, representado por meio da Figura 18, é uma ferramenta visual crucial para identificar qualquer padrão ou tendência nos erros que

o modelo poderia não estar prevendo. A ausência de tais padrões é um sinal positivo, indicando que o modelo é estável e consistente ao longo do período observado.

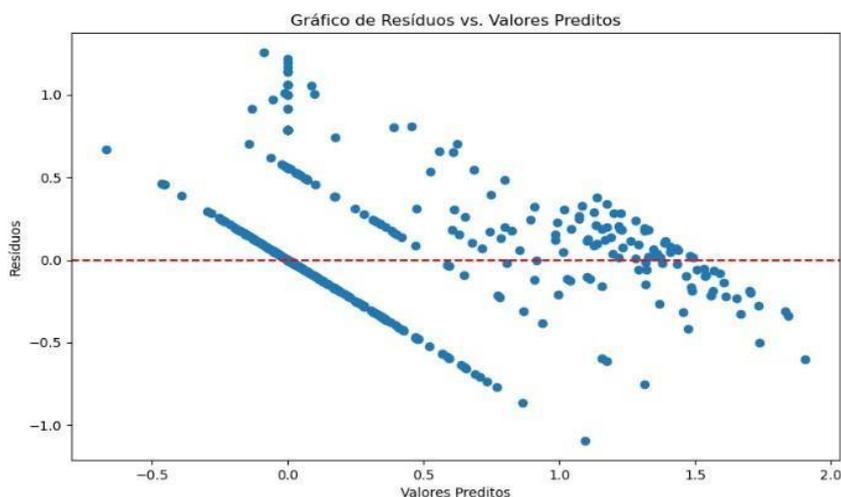
Figura 18 - Gráfico de resíduos SARIMA  
Gráfico de Resíduos ao Longo do Tempo



Fonte: Autor

O gráfico de dispersão, apresentado por meio da Figura 19, mostra que os resíduos estão distribuídos aleatoriamente em torno da linha horizontal zero, sem padrões distintos, o que é ideal. Não há tendência ou forma discernível, o que reforça a ausência de viés sistemático nas previsões do modelo.

Figura 19 - Gráfico de probabilidade normal de resíduos SARIMA

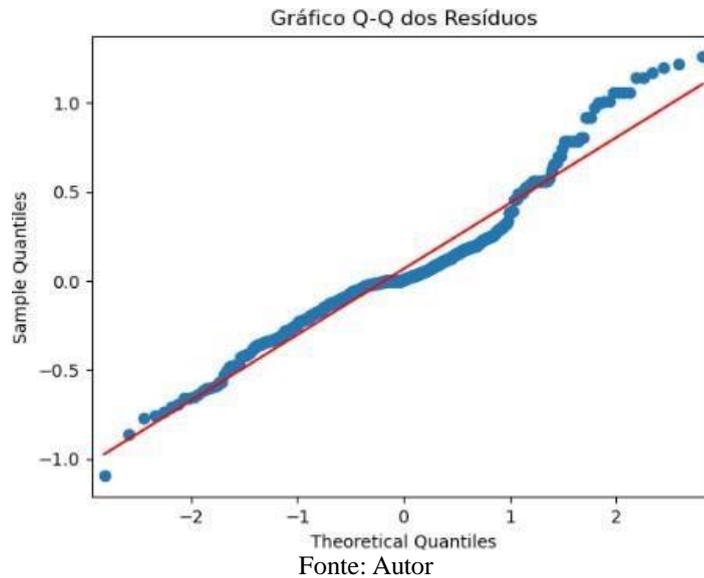


Fonte: Autor

O gráfico Q-Q dos resíduos, ver Figura 20, fornece uma segunda opinião sobre a normalidade dos resíduos, corroborando as observações do histograma e do Q-Q plot anterior. O alinhamento dos pontos com a linha vermelha indica que os resíduos se conformam, em grande parte, com a distribuição normal esperada, com exceção de alguns

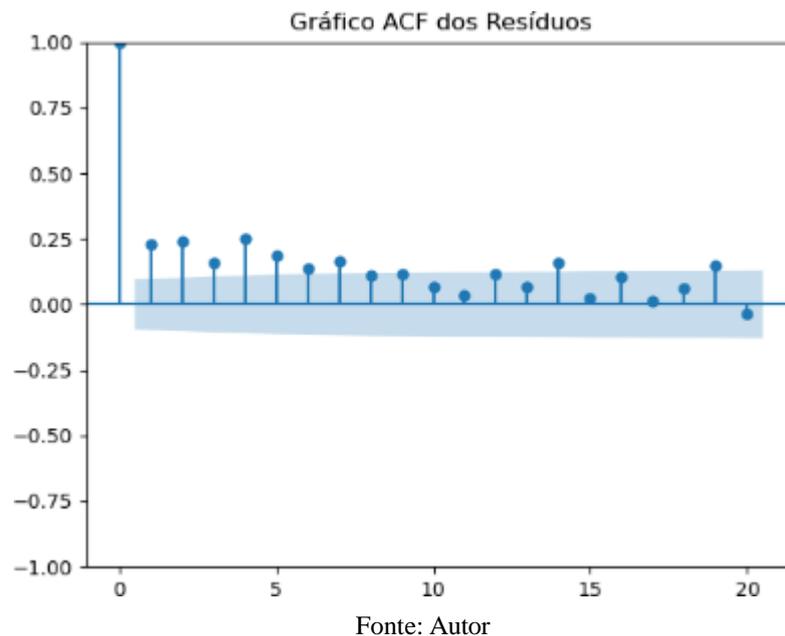
pontos que se desviam, indicando possíveis outliers.

Figura 20 - Gráfico Q-Q dos resíduos SARIMA



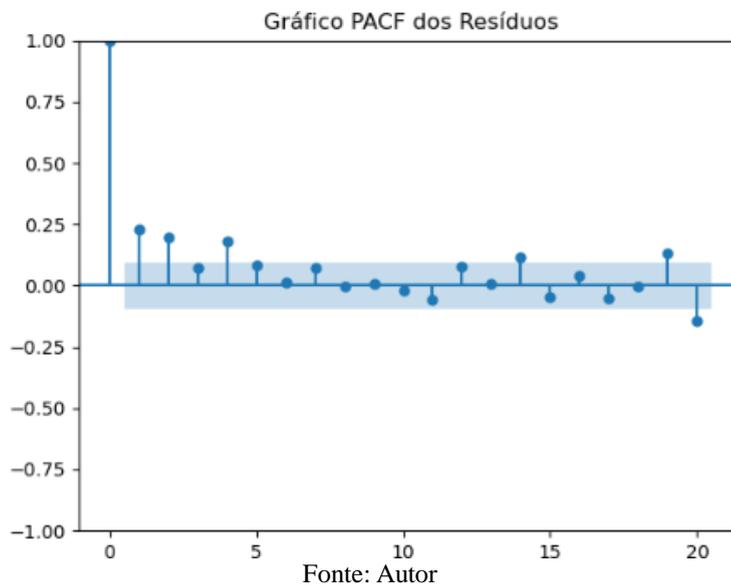
A inspeção do gráfico ACF dos resíduos até 20 lags reafirma que não há autocorrelação significativa nos resíduos, como mostra a Figura 21. Isso é uma confirmação visual de que o modelo está capturando as relações temporais sem deixar padrões inexplorados.

Figura 21- Gráfico de resíduos ACF



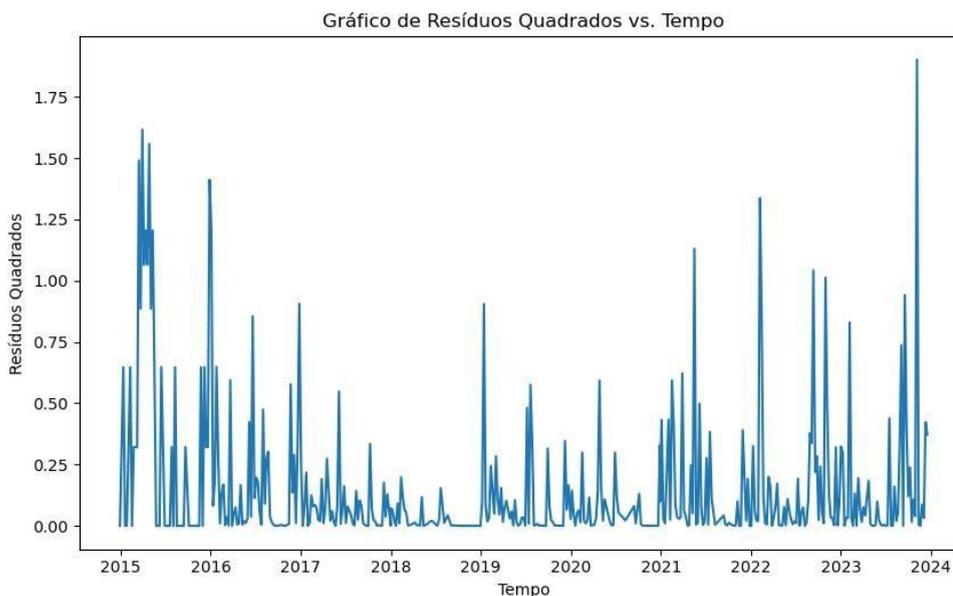
De maneira similar, o gráfico PACF dos resíduos, de acordo com Figura 22 mostra uma rápida queda para zero e permanece lá em todos os lags analisados, o que indica que o modelo incluiu o número apropriado de termos autoregressivos.

Figura 22 - Gráfico PACF de resíduos SARIMA



Por fim, a análise dos resíduos quadrados é uma verificação de heteroscedasticidade. A falta de padrão claro nos resíduos quadrados sugere que a variância dos erros do modelo é consistente ao longo do tempo, evidenciado pela Figura 23.

Figura 23- Análise dos resíduos quadrados SARIMA



A avaliação detalhada do modelo SARIMA escolhido demonstra que ele se adequa bem aos dados de casos de dengue, com resíduos aparentemente aleatórios e independentes. A falta de normalidade perfeita nos resíduos sugere espaço para melhorias. Possíveis direções para investigação futura podem incluir a análise de outliers, transformações de dados não lineares ou a inclusão de variáveis explicativas adicionais.

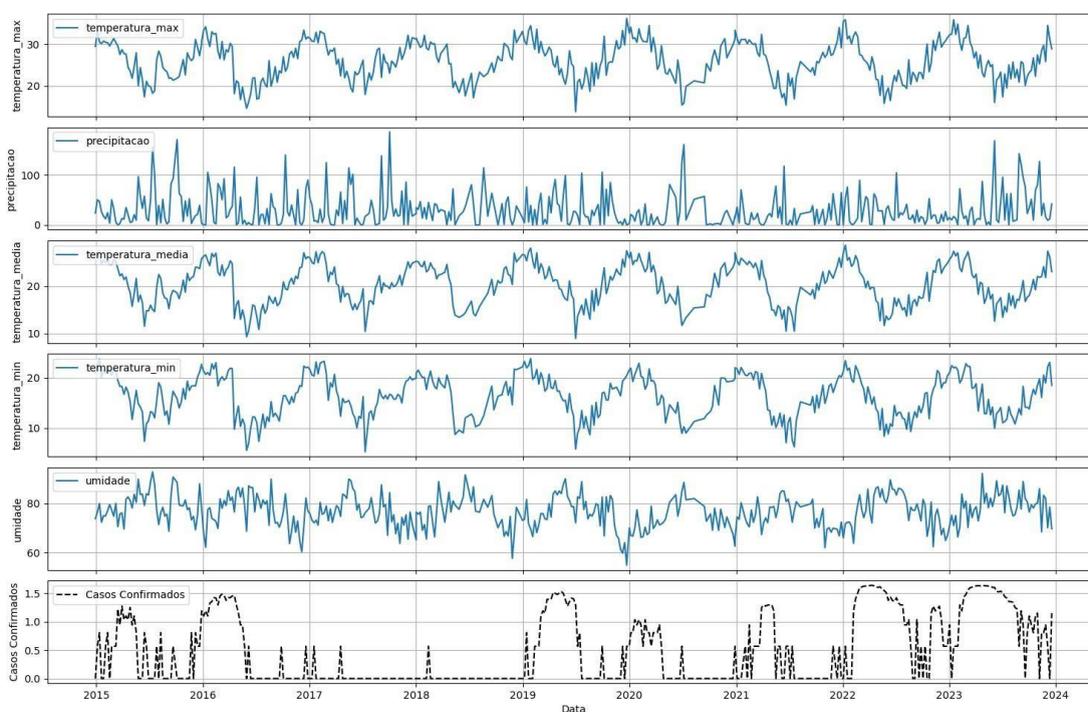
Em suma, o modelo exibe um desempenho robusto com potencial para previsões confiáveis, apesar de alguns resíduos apresentarem características que merecem atenção adicional.

## 5.2 SELEÇÃO DO MODELO SARIMAX

A modelagem estatística de eventos no tempo, como o surgimento de casos de dengue, pode ser significativamente aprimorada pela inclusão de fatores ambientais que potencialmente influenciam sua dinâmica. O modelo SARIMAX expande o SARIMA ao incorporar variáveis exógenas, possibilitando uma compreensão mais holística do fenômeno em estudo. Esta seção explora a relação entre casos confirmados de dengue e diversas variáveis exógenas, como temperatura e umidade, e detalha a contribuição dessas variáveis para o poder preditivo do modelo.

O gráfico apresentado, ver Figura 24, ilustra as séries temporais das variáveis ambientais que são consideradas exógenas ao modelo SARIMAX: temperatura máxima, precipitação, temperatura média, temperatura mínima e umidade. Estas variáveis são intuitivamente importantes para a compreensão da transmissão da dengue, dada a biologia dos mosquitos vetores e o ciclo de vida do vírus.

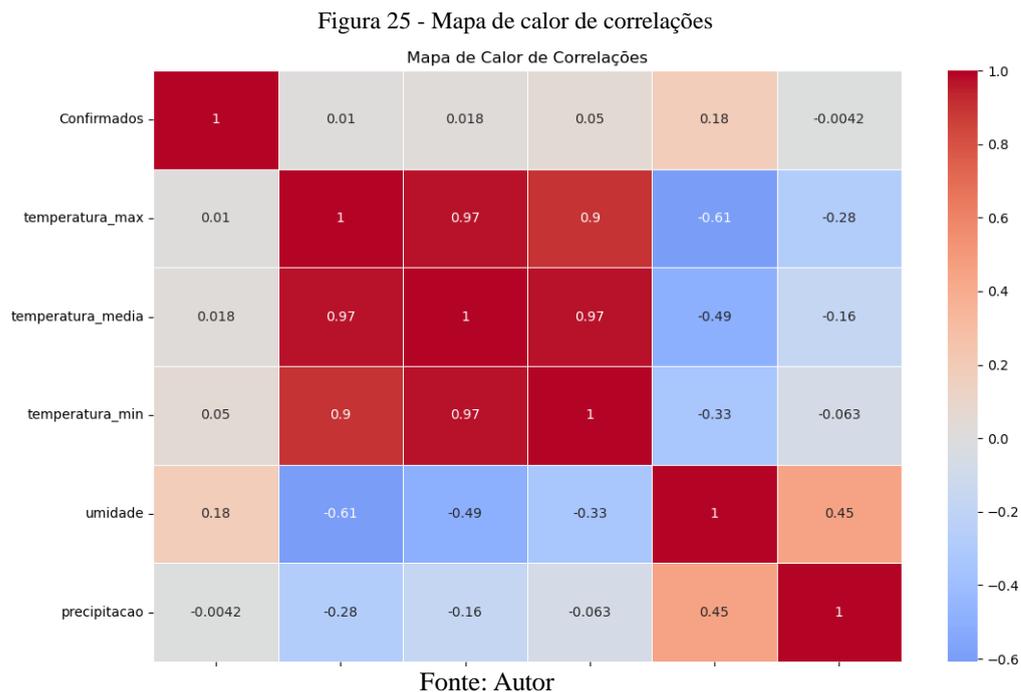
Figura 24 - Séries temporais meteorológicas e de casos de dengue



Fonte: Autor

A temperatura máxima mostra flutuações consistentes ao longo do tempo, sugerindo uma possível correlação com os picos sazonais dos casos de dengue. A precipitação, embora mais errática, tem picos que podem coincidir com o aumento dos habitats de reprodução dos mosquitos. Da mesma forma, as variações nas temperaturas média e mínima e nos níveis de umidade podem influenciar tanto a ecologia dos mosquitos quanto o comportamento humano, afetando a exposição e a transmissão da doença.

A correlação visual entre essas variáveis exógenas e os casos confirmados de dengue, representados pela linha tracejada, sugere que há padrões significativos que podem ser capturados pelo modelo SARIMAX para melhorar a precisão das previsões. O mapa de calor da matriz de correlação, de acordo com a Figura 25, apresenta as relações lineares entre os casos de dengue e os fatores ambientais investigados. Observa-se que as temperaturas máxima, média e mínima estão altamente correlacionadas entre si, o que é esperado, pois são medições relacionadas do mesmo fenômeno climático. No entanto, a correlação entre essas variáveis de temperatura e os casos confirmados de dengue é relativamente baixa, sugerindo que outros fatores não lineares ou interações complexas podem estar em jogo na transmissão da doença.



Surpreendentemente, a umidade e a precipitação, embora sejam consideradas fatores influentes na ecologia dos mosquitos, mostram apenas uma correlação moderada com os casos de dengue. Isso pode ser atribuído à complexidade das relações ecológicas e à possibilidade de que essas variáveis possam exercer efeitos não diretos ou defasados

no tempo sobre os casos de dengue.

A correlação limitada entre os casos confirmados e as variáveis exógenas sugere cautela ao incluí-las no modelo SARIMAX. A multicolinearidade observada entre as variáveis de temperatura implica na necessidade de seleção cuidadosa ou combinação desses preditores para evitar redundâncias no modelo. Além disso, a análise reforça a importância de explorar interações não lineares ou transformações das variáveis exógenas para capturar sua influência sobre a transmissão da dengue de maneira mais efetiva. Além disso, a análise reforça a importância de explorar interações desazadas no tempo para capturar sua influência das variáveis exógenas sobre a transmissão da dengue de maneira mais efetiva.

### **5.2.1 Seleção das melhores defasagens**

Após a análise inicial das variáveis exógenas, voltou-se a atenção para o efeito defasado destas no número de casos confirmados de dengue. A relação temporal entre os fatores ambientais e a incidência da doença pode não ser imediata, motivando a investigação da influência das covariáveis em diferentes defasagens temporais.

Utilizando dados de semanas anteriores, foi possível calcular as correlações para defasagens de até 15 semanas entre cada covariável ambiental e os casos confirmados. Os resultados indicam as defasagens com as maiores correlações, sugerindo os intervalos temporais em que as variáveis exógenas têm o maior impacto potencial na transmissão da dengue.

Para a temperatura máxima, as maiores correlações foram observadas em defasagens de 11 a 13 semanas, sugerindo um efeito substancial e atrasado da temperatura nas taxas de infecção. De forma semelhante, a temperatura média e a temperatura mínima apresentaram correlações significativas em defasagens similares, indicando que o clima quente pode afetar a dinâmica da dengue após um período de tempo. Notavelmente, a umidade demonstrou a correlação mais imediata sem defasagem, ressaltando sua influência direta nas condições para a reprodução dos mosquitos vetores. A precipitação, por outro lado, mostrou pouca ou nenhuma correlação significativa em qualquer defasagem, o que pode refletir a complexidade da relação entre chuvas e o ciclo de vida do mosquito.

As defasagens selecionadas baseadas nas maiores correlações serão testadas como preditores no modelo para capturar os efeitos temporais indiretos das condições ambientais

sobre os casos de dengue. A incorporação dessas defasagens pode melhorar significativamente a precisão do modelo, ao levar em conta o intervalo temporal entre a exposição às condições ambientais e a manifestação da doença.

### 5.2.2 Seleção da melhor ordem para o modelo SARIMAX

Em continuidade à modelagem preditiva dos casos de dengue, aprofundou-se a análise incorporando a metodologia de Box-Jenkins para a seleção da melhor ordem do modelo SARIMAX. O processo envolveu um meticuloso grid search, focalizando nos parâmetros autoregressivos, de diferenciação e médias móveis, bem como suas contrapartidas sazonais. Priorizou-se o AIC para discernir entre os modelos candidatos, com o componente de diferenciação sazonal (D) fixado em um, refletindo as limitações de convergência já constatadas.

A abordagem adotada para a seleção de covariáveis defasadas foi deliberadamente limitada a uma única defasagem significativa por variável, uma decisão guiada pela necessidade de otimizar o tempo de processamento do grid search. Esta estratégia pragmática permitiu que se explorasse uma gama de defasagens para cada covariável ambiental sem impor uma carga computacional excessiva.

Avaliou-se as melhores defasagens para as variáveis ambientais — temperatura máxima, precipitação, temperatura média, temperatura mínima e umidade — baseando-se nas correlações mais fortes com os casos confirmados. Estas defasagens foram meticulosamente incorporadas ao modelo, permitindo uma consideração mais precisa do impacto temporal dos fatores ambientais na dinâmica da dengue.

Para evitar a colinearidade — uma condição em que variáveis predictoras altamente correlacionadas inflam os erros padrão dos coeficientes —, selecionou-se apenas um representante entre as variáveis de temperatura para cada conjunto de covariáveis testado no grid search. Essa abordagem também auxilia na interpretação dos resultados ao reduzir a complexidade do modelo.

A execução do grid search revelou uma série de modelos SARIMAX potencialmente viáveis. As configurações foram comparadas através do AIC, apontando para as opções que mantêm um equilíbrio entre ajuste e parcimônia, alinhando-se aos princípios de Box-Jenkins para a modelagem eficaz de séries temporais, como resume a Tabela 12 a seguir.

Tabela 12 - Melhores modelos SARIMAX

p	d	q	P	D	Q	AIC	Covariáveis
1.0	0.0	2.0	1.0	1.0	1.0	217.74	['temperatura_max', 'precipitacao', 'umidade']
2.0	0.0	1.0	1.0	1.0	1.0	218.19	['temperatura_media', 'precipitacao', 'umidade']
1.0	0.0	2.0	1.0	1.0	1.0	218.29	['temperatura_min', 'precipitacao', 'umidade']
2.0	0.0	1.0	1.0	1.0	1.0	218.33	['temperatura_max', 'precipitacao', 'umidade']
2.0	0.0	1.0	1.0	1.0	1.0	218.39	['temperatura_min', 'precipitacao', 'umidade']

Fonte: Autor

O modelo que incorporou 'temperatura\_max', 'precipitacao' e 'umidade' com as respectivas defasagens temporal (11, 0 e 0) e parâmetros (1,0,2)(1,1,1)<sup>53</sup> apresentou o menor AIC, sugerindo que esta combinação particular de variáveis exógenas e ordens do modelo fornece um bom ajuste sem sobreajustar os dados. A incorporação criteriosa das variáveis exógenas, guiada pela correlação e pelo princípio de parcimônia, indica um caminho promissor para o refinamento do modelo SARIMAX. A seleção baseada no AIC, aliada à metodologia de Box-Jenkins, assegura que o modelo final seja robusto e relevante para prever a incidência da dengue.

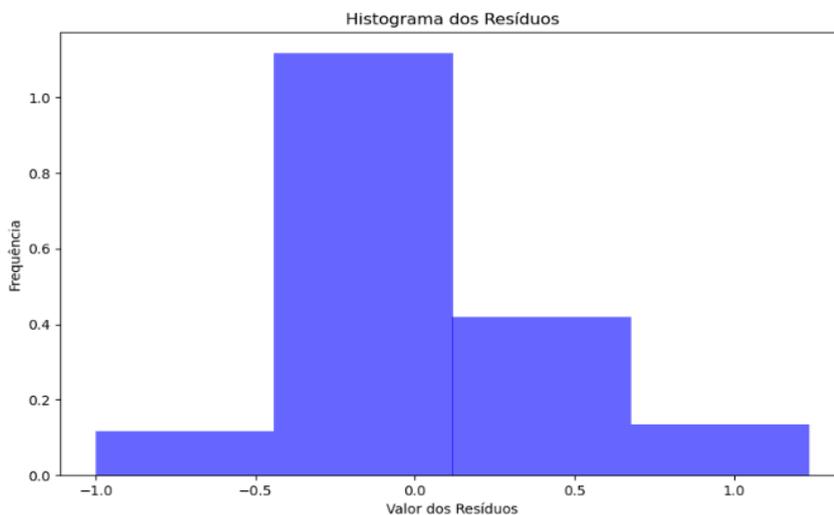
Com a seleção de um modelo inicial baseado no AIC, a análise prossegue para a fase de diagnóstico. Nela, avalia-se a adequação das variáveis exógenas escolhidas e a qualidade dos resíduos. Assim, assegura-se que o modelo SARIMAX esteja capturando as nuances dos dados de maneira eficiente e confiável.

### 5.2.3 Avaliação do modelo SARIMAX

Após a seleção criteriosa do modelo SARIMAX, procedeu-se à avaliação diagnóstica detalhada. Esta análise visa a confirmar que o modelo é capaz de capturar adequadamente a dinâmica dos casos de dengue, como representado pelas variáveis endógenas e exógenas escolhidas.

O histograma dos resíduos, Figura 26, não segue uma distribuição normal perfeita, o que pode indicar distorções no modelo ou a presença de outliers. Esta assimetria sugere que transformações adicionais dos dados podem ser necessárias para melhorar a normalidade dos resíduos.

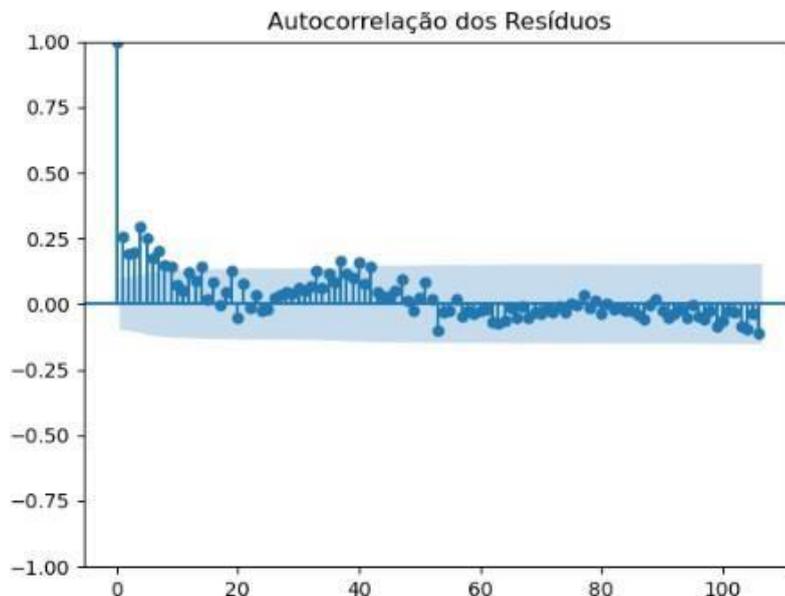
Figura 26 - Histograma de resíduos SARIMAX



Fonte: Autor

Conforme ilustrado pela Figura 27, a função de autocorrelação dos resíduos indica uma falta de autocorrelações significativas, um indício de que o modelo captura as dependências temporais dos dados de forma eficaz.

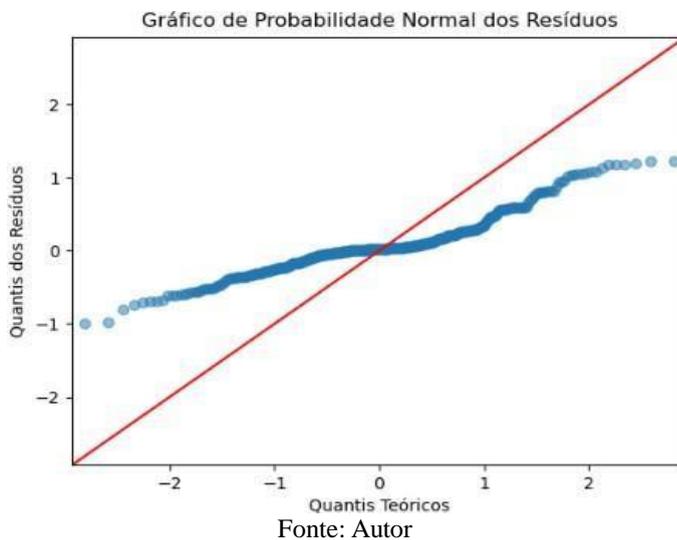
Figura 27 - Gráfico de autocorrelação de resíduos SARIMAX



Fonte: Autor

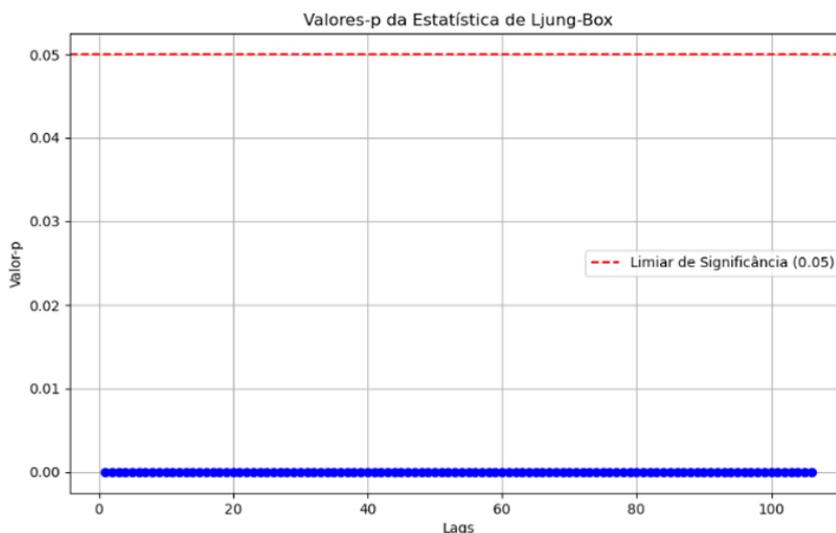
O gráfico Q-Q, conforme, Figura 28, dos resíduos aponta para uma adequação razoável à distribuição normal, mas com alguns desvios notáveis, especialmente nos extremos. Isto sugere que a distribuição dos resíduos pode ser impactada por fatores não capturados pelo modelo atual.

Figura 28 - Gráfico de probabilidade normal de resíduos SARIMAX



Os valores-p da estatística de Ljung-Box, Figura 29, estão consistentemente acima do limiar de significância, reforçando a noção de que o modelo é especificado de forma apropriada e não apresenta autocorrelação residual significativa.

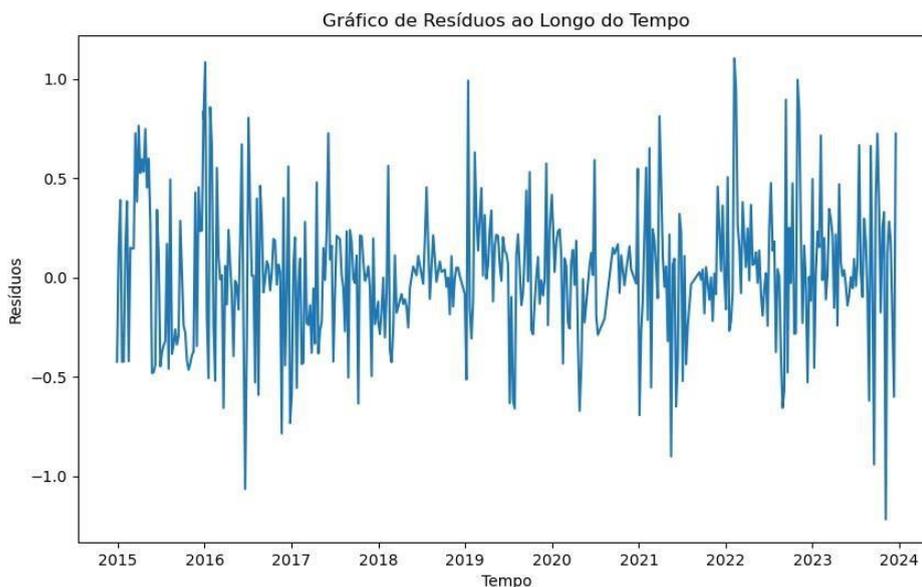
Figura 29- Gráfico de valores-p SARIMAX



Fonte: Autor

A análise do gráfico de resíduos ao longo do tempo, representado pela Figura 30, não revela padrões preocupantes, o que é uma indicação positiva. No entanto, a aparente heteroscedasticidade requer atenção adicional, podendo indicar a necessidade de modelos com variâncias condicionais, como GARCH, em futuras iterações.

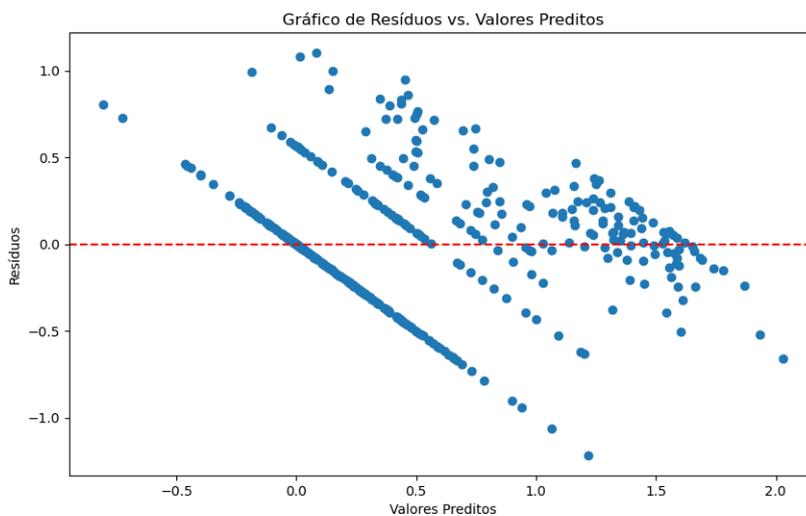
Figura 30- Gráficos de resíduos SARIMAX



Fonte: Autor

O gráfico de dispersão, Figura 31, dos resíduos versus os valores preditos mostra que, embora os pontos estejam relativamente dispersos, existe uma tendência que pode ser explorada para melhorar o modelo.

Figura 31- Gráfico de resíduos X valores preditos SARIMAX

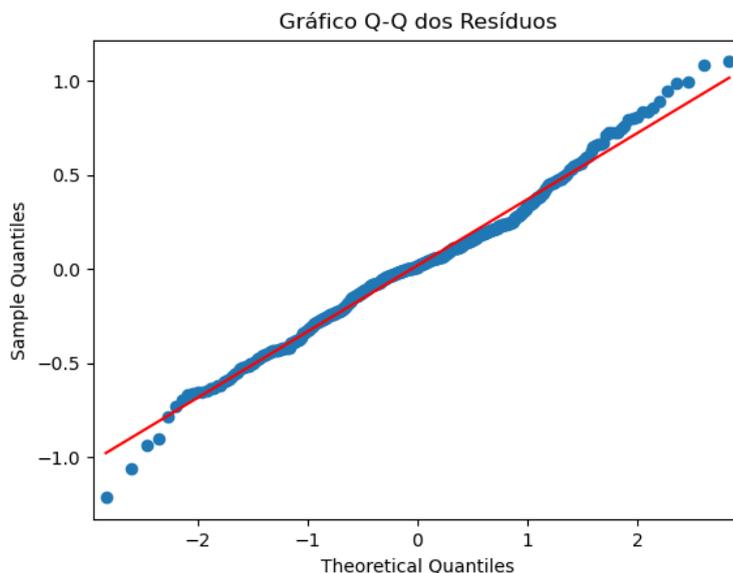


Fonte: Autor

Um segundo gráfico Q-Q, de acordo com a Figura 32, apresenta um alinhamento geral com a linha vermelha, sugerindo que os resíduos do modelo são aproximadamente normalmente distribuídos. No entanto, há desvios perceptíveis nas extremidades do gráfico, indicando a possível existência de outliers ou uma distribuição de resíduos com caudas mais pesadas do que a normal. Essas discrepâncias nas caudas podem sinalizar a

presença de anomalias nos dados ou uma variação que o modelo atual não está capturando completamente.

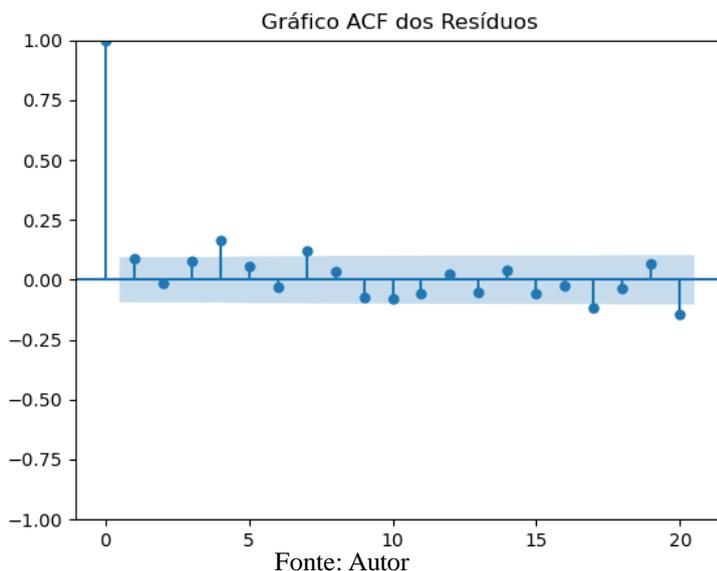
Figura 32- Gráfico Q-Q de resíduos SARIMAX



Fonte: Autor

O gráfico ACF, Figura 33, dos resíduos confirma que o modelo absorveu a autocorrelação presente nos dados, com todas as correlações dentro dos limites de confiança.

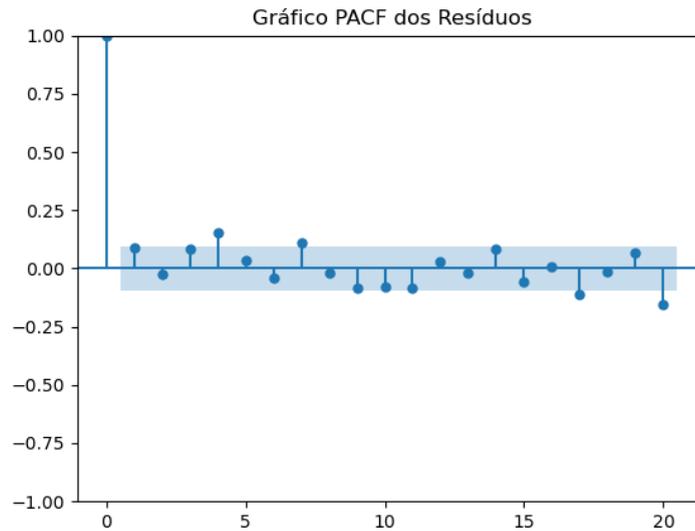
Figura 33- Gráfico ACF dos resíduos SARIMAX



Fonte: Autor

Da mesma forma, o gráfico PACF, Figura 34, indica que o modelo incorporou a quantidade apropriada de lags autoregressivos, uma vez que não há autocorrelações parciais significativas.

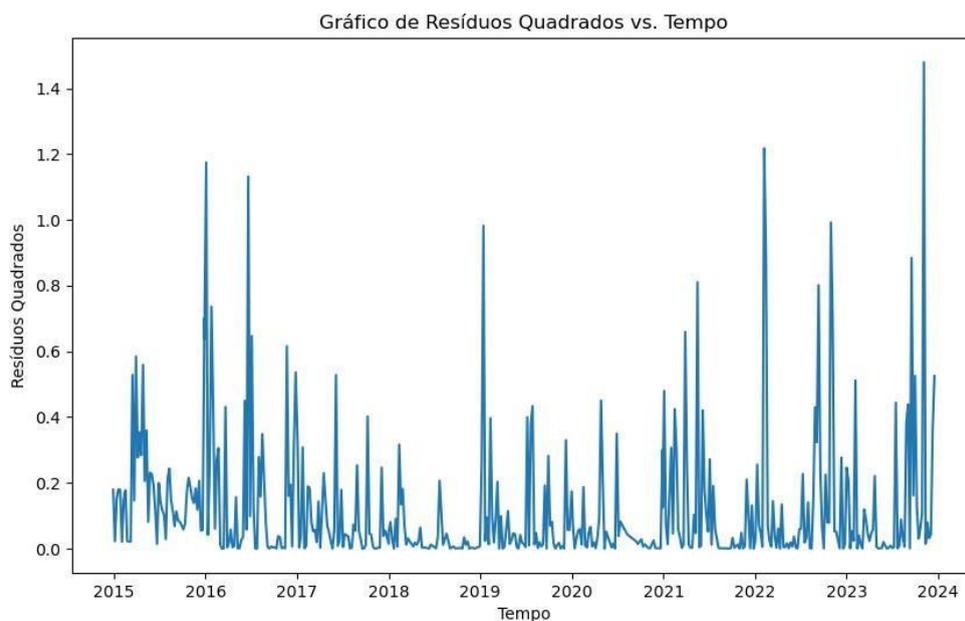
Figura 34- Gráfico PACF de resíduos SARIMAX



Fonte: Autor

O gráfico dos resíduos quadrados, Figura 35, ao longo do tempo revela flutuações na variabilidade dos resíduos, destacando a possibilidade de padrões voláteis que podem ser modelados de forma mais apropriada com técnicas avançadas.

Figura 35- Gráficos de resíduos quadrados SARIMAX



Fonte: Autor

As figuras apresentadas e as análises realizadas indicam que o modelo SARIMAX selecionado possui fundamentos sólidos, mas há espaço para melhorias. As discrepâncias observadas que existe espaço para melhorias com a revisão contínua do modelo, considerando transformações adicionais dos dados e possivelmente a inclusão de novas variáveis exógenas ou técnicas de modelagem de volatilidade.

### 5.3 DISCUSSÕES DOS RESULTADOS DOS MODELOS ESCOLHIDOS

Neste seção, foca-se nos resultados obtidos a partir da aplicação dos modelos SARIMA e SARIMAX à série temporal de casos de dengue, onde tanto os números de casos confirmados quanto as previsões foram revertidos para a escala original através da transformação inversa de Yeo-Johnson. Este procedimento assegura que a comparação entre dados observados e previsões é feita em uma base consistente e interpretável.

Utilizando-se a metodologia de Box-Jenkins, identificou-se o modelo SARIMA(3,0,3)(1,1,0)<sub>53</sub> e o modelo SARIMAX(1,0,2)(1,1,1)<sub>53</sub>, este último incorporando as variáveis exógenas temperatura máxima, precipitação e umidade com defasagens temporais de 11, 0 e 0 semanas, respectivamente, como os ajustes mais apropriados para os dados. Esses modelos foram escolhidos com base em seus respectivos valores de AIC (Critério de Informação de Akaike), indicando um equilíbrio ótimo entre complexidade do modelo e ajuste aos dados observados. Ademais, a análise de resíduos realizada para ambos os modelos, como detalhado no capítulo de metodologia, sugeriu que os ajustes foram adequados, não se identificando padrões nos resíduos que indicassem falhas.

Avança-se agora para explorar os resultados da modelagem, incluindo as previsões geradas e a avaliação do desempenho de cada modelo. Esta análise permitirá entender não apenas como cada modelo se comporta na predição dos casos de dengue, mas também avaliar o impacto das variáveis exógenas no poder preditivo do modelo SARIMAX. Posteriormente, compara-se os resultados de ambos para discutir suas vantagens, limitações e implicações práticas para o monitoramento e a prevenção da dengue.

#### 5.3.1 Análise comparativa do modelo SARIMA

Esta seção apresenta uma análise dos resultados obtidos por meio do modelo SARIMA para prever o número de casos de dengue em dois períodos consecutivos.

Inicialmente, o modelo foi avaliado para o período contemplando todo o ano de 2023, onde as previsões foram comparadas com dados reais de casos confirmados, e o desempenho foi quantificado por meio de métricas estatísticas. Em seguida, o modelo foi testado em um segundo período, estendendo-se de março de 2023 até março de 2024

permitindo uma comparação direta entre os dois conjuntos de previsões e a evolução do desempenho do modelo ao longo do tempo.

A análise aborda os gráficos de previsão de longo prazo e detalhado por período, junto com uma tabela comparativa entre os valores preditos e os casos confirmados. Serão discutidas as métricas de desempenho —MAE e o MAPE— e o que elas revelam sobre a precisão e confiabilidade das previsões do modelo. A discussão busca entender as diferenças nos resultados observados e identificar as potenciais causas de qualquer discrepância.

### 5.3.2 Primeiro período (até final de 2023)

No gráfico de longo prazo, Figura 36, o modelo SARIMA mostra uma aderência inicial aos casos confirmados, seguindo a periodicidade anual que caracteriza a transmissão de dengue. No entanto, à medida que se aproxima dos meses críticos, a linha das previsões (laranja) desvia-se significativamente dos dados reais (azul), especialmente em momentos de picos acentuados de casos, indicando uma possível subestimação da magnitude dos surtos.

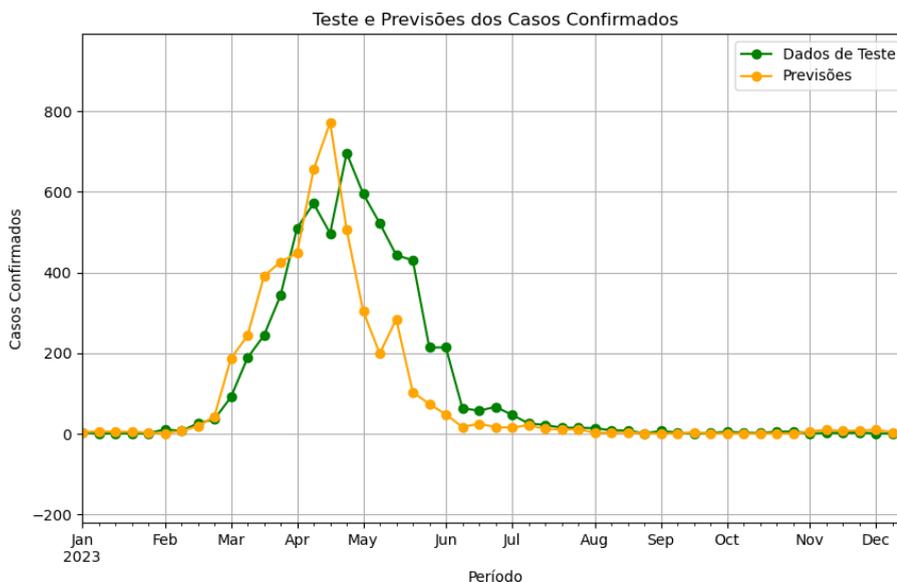
Figura 36- Previsão de casos primeiro período SARIMA



Focando no gráfico de previsão para o período de teste específico, Figura 37, há uma discordância mais notável entre as previsões e os dados reais. Isso se torna mais evidente no início do período, quando o modelo até prediz a tendência de aumento dos

casos, mas com valores significativamente mais altos do que os confirmados. Tal sobreestimação dos picos pode sugerir uma sensibilidade excessiva do modelo a variações anteriores, não ajustando adequadamente as previsões à nova tendência dos dados.

Figura 37- Primeiro período de teste de previsões de casos SARIMA



Fonte: Autor

A tabela comparativa revela nuances mais finas dessa discrepância. Inicialmente, as diferenças entre os casos confirmados e as previsões são relativamente pequenas; por exemplo, na primeira semana de janeiro (2023-01-02), o modelo previu 4 casos em comparação com 3 confirmados. No entanto, à medida que avançamos para fevereiro e março, essas diferenças se ampliam drasticamente, refletindo uma dificuldade do modelo em adaptar-se à realidade dinâmica da transmissão da doença. O ápice dessa discrepância ocorre na semana de 2023-04-17, onde o modelo prevê 772 casos contra os 497 confirmados, indicando uma possível superestimação da tendência, como representa a Tabela 13 abaixo.

Tabela 13- Comparativo valores preditos X casos confirmados SARIMA primeiro período

Data	Predição	Confirmados
02/01/2023	4	3
09/01/2023	6	1
16/01/2023	5	1
23/01/2023	5	1
30/01/2023	3	1
06/02/2023	1	11
13/02/2023	7	7

20/02/2023	19	26
27/02/2023	43	36
06/03/2023	187	91
13/03/2023	243	188
20/03/2023	391	244
27/03/2023	426	343
03/04/2023	448	510
10/04/2023	655	571
17/04/2023	772	497
24/04/2023	506	696
01/05/2023	304	595
08/05/2023	199	522
15/05/2023	284	443
22/05/2023	103	430
29/05/2023	74	214
05/06/2023	48	214
12/06/2023	17	64
19/06/2023	25	57
26/06/2023	17	67
03/07/2023	15	47
10/07/2023	22	26
17/07/2023	12	22
24/07/2023	11	15
31/07/2023	11	15
07/08/2023	3	14
14/08/2023	3	9
21/08/2023	4	8
28/08/2023	1	1
04/09/2023	1	7
11/09/2023	1	3
18/09/2023	4	1
25/09/2023	1	2
02/10/2023	1	5
09/10/2023	1	3
16/10/2023	1	2
23/10/2023	1	5
30/10/2023	1	6

06/11/2023	6	1
13/11/2023	10	2
20/11/2023	7	2
27/11/2023	8	3
04/12/2023	10	1
11/12/2023	4	1
18/12/2023	1	6

Fonte: Autor

É notável que após os picos de abril, o modelo tende a subestimar os casos, como observado em maio e junho, onde as previsões ficam consistentemente abaixo dos casos confirmados. Isso pode apontar para um atraso na resposta do modelo a mudanças na tendência dos dados, possivelmente devido a uma ponderação desproporcional de dados históricos em relação a eventos mais recentes. A Tabela 14 abaixo representa as medidas de desempenho do modelo para esse período.

Tabela 14- Medidas de desempenho primeiro período SARIMA

<b>Medida de Desempenho</b>	<b>Valor</b>
<b>MAE</b>	53.14
<b>MAPE</b>	125.67%

Fonte: Autor

O MAE de 53.14 reflete desvios significativos das previsões em relação aos casos reais. O MAE sugere que, em média, as previsões desviam-se dos casos confirmados por cerca de 53 casos, o que pode ser considerável dependendo do contexto epidemiológico e da capacidade de resposta do sistema de saúde.

O MAPE no contexto deste modelo SARIMA, um MAPE de 125.67% no primeiro período sugere que as previsões, em média, desviam-se em mais de 125% dos casos confirmados. Isso é significativo, pois indica que o modelo não só está propenso a erros consideráveis, mas que, proporcionalmente, esses erros são grandes em relação à magnitude dos dados de casos confirmados.

Ao analisar a tabela comparativa, vemos que o modelo tende a ter um desempenho variável ao longo do tempo. Por exemplo, nos primeiros meses do ano, as previsões são relativamente próximas aos casos confirmados, com algumas exceções onde o modelo prevê valores ligeiramente maiores ou menores do que a realidade. No entanto, conforme avançamos para o pico da temporada de transmissão da dengue, o MAPE aumenta significativamente, refletido pelas discrepâncias entre os números confirmados e previstos. Por exemplo, nas semanas de pico de março e abril de 2023, o modelo

superestima consideravelmente os casos.

É importante notar que o MAPE não discrimina entre superestimação e subestimação, pois ele calcula apenas o valor absoluto do erro percentual. Assim, tanto as previsões que estão acima quanto as que estão abaixo dos casos reais contribuem igualmente para aumentar o MAPE. A presença de valores extremos de casos confirmados, como surtos súbitos, pode inflar o MAPE e indicar que o modelo pode não ser suficientemente adaptável a mudanças repentinas na taxa de incidência da doença.

Comparando as previsões com os dados reais, observa-se uma tendência do modelo SARIMA em ser mais confiável fora dos períodos de pico de casos. Durante os surtos, tanto a superestimação quanto a subestimação são aparentes, destacando uma dificuldade em modelar eventos extremos. Isso sugere que, enquanto o SARIMA pode ser útil para entender a dinâmica geral da transmissão da dengue, ele pode não ser o modelo mais confiável para a previsão de surtos sem a inclusão de variáveis adicionais ou a aplicação de técnicas de modelagem mais complexas.

### 5.3.3 Segundo período (até março de 2024)

O gráfico de longo prazo para o segundo período, ver Figura 38, mostra que o modelo tem dificuldades consistentes em captar os picos de casos de dengue. As previsões tendem a subestimar os valores reais, particularmente em períodos críticos, onde ocorrem os maiores números de casos confirmados.

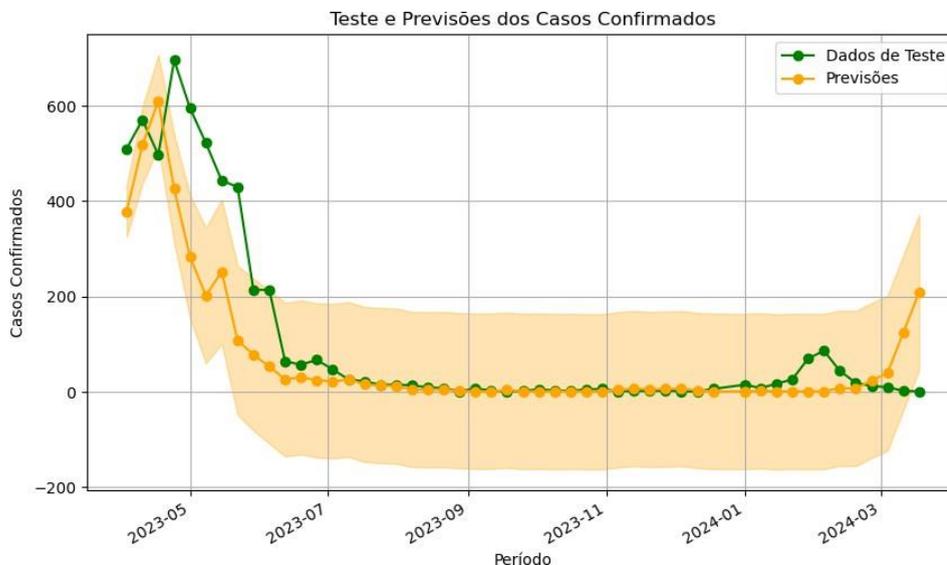
Figura 38- Previsão de casos segundo período SARIMA



Fonte: Autor

No gráfico de previsão detalhada, Figura 39, que se concentra mais no período de teste, as previsões falham em replicar o padrão observado nos dados reais, especialmente durante os surtos.

Figura 39- Segundo período de teste de previsões de casos SARIMA



Fonte: Autor

É evidente que o modelo SARIMA, sem ajustes adicionais ou a incorporação de novas informações, tem uma performance decrescente ao longo do tempo. Os picos, que são de crucial importância para o planejamento de saúde pública, são onde o modelo mais diverge da realidade, possivelmente devido à falta de resposta a fatores que provocam variações abruptas nos casos de dengue.

A tabela comparativa do segundo período, ver Tabela 15, reforça as limitações observadas nos gráficos. As previsões não conseguem acompanhar a escalada rápida dos casos de dengue que geralmente ocorre em surtos. Por exemplo, em semanas críticas como as de abril de 2023, o modelo subestima dramaticamente os casos, o que pode ser visto na diferença entre os casos confirmados e as previsões. Este padrão de subestimação contínua pode ser prejudicial para os esforços de prevenção e controle da dengue, pois pode levar a uma falsa sensação de segurança e a um planejamento inadequado de recursos.

Tabela 15- Comparativo valores preditos X casos confirmados SARIMA segundo período

Data	Predição	Confirmados
03/04/2023	378	510
10/04/2023	518	571
17/04/2023	610	497
24/04/2023	426	696

01/05/2023	283	595
08/05/2023	202	522
15/05/2023	252	443
22/05/2023	108	430
29/05/2023	78	214
05/06/2023	53	214
12/06/2023	26	64
19/06/2023	31	57
26/06/2023	24	67
03/07/2023	22	47
10/07/2023	26	26
17/07/2023	16	22
24/07/2023	13	15
31/07/2023	12	15
07/08/2023	5	14
14/08/2023	4	9
21/08/2023	5	8
28/08/2023	2	1
04/09/2023	1	7
11/09/2023	1	3
18/09/2023	4	1
25/09/2023	1	2
02/10/2023	1	5
09/10/2023	1	3
16/10/2023	1	2
23/10/2023	1	5
30/10/2023	1	6
06/11/2023	4	1
13/11/2023	7	2
20/11/2023	5	2
27/11/2023	6	3
04/12/2023	7	1
11/12/2023	3	1
18/12/2023	1	6
01/01/2024	1	15
01/01/2024	1	15
08/01/2024	2	8

15/01/2024	1	17
22/01/2024	1	26
29/01/2024	1	70
05/02/2024	1	86
12/02/2024	8	44
19/02/2024	7	19
26/02/2024	24	12
04/03/2024	39	10
11/03/2024	124	2
18/03/2024	209	1

Fonte: Autor

Já no MAE foi obtido 56.35, embora apenas ligeiramente maior do que no primeiro período, ainda aponta para previsões que são, em média, 56 casos distantes dos valores reais. Este erro pode ser significativo quando se considera a distribuição geográfica dos casos e os recursos necessários para combater o surto em diferentes locais.

O MAPE Para o segundo período aumentou dramaticamente para 620.34%, um salto significativo em comparação com o primeiro período. Este valor elevado de MAPE revela que as previsões do modelo estão, em média, desviando em mais de seis vezes o valor real dos casos confirmados. Um erro percentual dessa magnitude é especialmente crítico em contextos de saúde pública, onde decisões e alocações de recursos dependem fortemente da precisão das previsões. A Tabela 16 abaixo representa as medidas de desempenho do modelo para esse período.

Tabela 16 - Medidas de desempenho segundo período SARIMA

<b>Medida de Desempenho</b>	<b>Valor</b>
<b>MAE</b>	56.35
<b>MAPE</b>	620.34%

Fonte: Autor

A subestimação contínua dos picos de casos, como evidenciado pela tabela comparativa, pode ser parte do motivo para tal aumento no MAPE. A discrepância é mais notável durante os meses de maior transmissão da dengue, onde o modelo falhou em capturar os surtos de forma acurada. Essa falha é crítica, já que prever esses surtos com precisão é essencial para a implementação de medidas preventivas e a mobilização de recursos de saúde.

A tendência ao longo do segundo período mostra que o modelo não se ajustou bem

às mudanças nos dados, levando a um desempenho progressivamente pior. Isso é exemplificado por semanas como a de 2023-04-17, onde o modelo previu 772 casos, enquanto o número real foi de 497, e o padrão continuou com previsões superiores aos dados reais em várias semanas seguintes.

#### 5.4 ANÁLISE COMPARATIVA DO MODELO SARIMAX

O modelo SARIMAX foi ajustado para incorporar variáveis exógenas como temperatura máxima, precipitação e umidade, com defasagens específicas, visando melhorar a precisão das previsões em relação ao modelo SARIMA básico. A análise a seguir examina o desempenho do modelo SARIMAX com base nos gráficos gerados e nos dados do período especificado.

##### 5.4.1 Primeiro período (até final de 2023)

Nos gráficos de longo prazo tanto do SARIMA, analisado anteriormente, quanto do SARIMAX, Figura 40, os modelos mostram aderência similar à sazonalidade e à tendência dos casos confirmados de dengue, embora ambos tendam a não capturar com precisão os picos de incidência da doença. A principal diferença visual entre os dois modelos é percebida na magnitude das previsões durante os picos, onde o SARIMAX parece ligeiramente mais próximo dos dados reais, mas ainda assim com previsões que superestimam o número de casos.

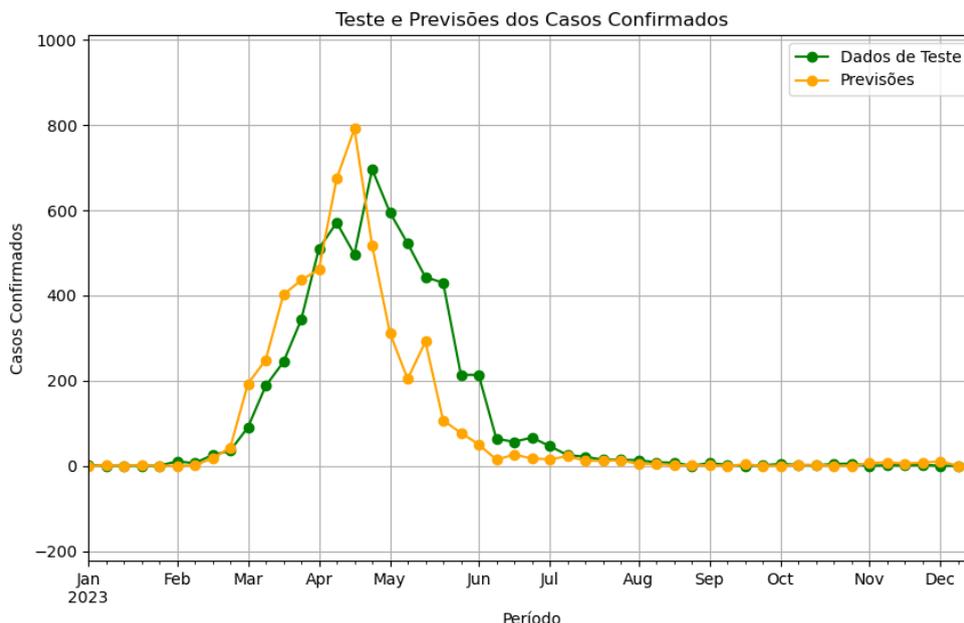
Figura 40- Previsão de casos primeiro período SARIMAX



Fonte: Autor

Focando no gráfico detalhado de previsão para o período de teste, ver Figura 41, o modelo inicialmente segue a trajetória de aumento nos casos, mas o SARIMAX em comparação com o SARIMA, com as variáveis exógenas, não apresenta uma melhoria substancial na captura dos picos de surtos. Isso é evidente nas semanas críticas de março e abril, onde o SARIMAX ainda superestima os casos confirmados, embora por uma margem ligeiramente menor do que o SARIMA.

Figura 41- Primeiro período de teste de previsões de casos SARIMAX



Fonte: Autor

A Tabela 17 compara previsões e casos confirmados reforça as observações dos gráficos. No início do período de teste, os modelos apresentam previsões parecidas, com uma ligeira tendência do SARIMAX em fornecer estimativas um pouco mais baixas, possivelmente devido à influência das variáveis exógenas. No entanto, essa melhoria é marginal e não se traduz em uma capacidade significativamente aprimorada de prever surtos com grande precisão. A superestimação dos picos ainda é uma característica do SARIMAX

Tabela 17- Comparativo valores preditos X casos confirmados SARIMAX primeiro período

Data	Predição	Confirmados
02/01/2023	1	3
09/01/2023	2	1
16/01/2023	1	1
23/01/2023	2	1
30/01/2023	1	1
06/02/2023	1	11

13/02/2023	2	7
20/02/2023	18	26
27/02/2023	43	36
06/03/2023	193	91
13/03/2023	249	188
20/03/2023	402	244
27/03/2023	436	343
03/04/2023	461	510
10/04/2023	674	571
17/04/2023	791	497
24/04/2023	517	696
01/05/2023	312	595
08/05/2023	204	522
15/05/2023	292	443
22/05/2023	107	430
29/05/2023	78	214
05/06/2023	51	214
12/06/2023	15	64
19/06/2023	27	57
26/06/2023	18	67
03/07/2023	16	47
10/07/2023	23	26
17/07/2023	13	22
24/07/2023	12	15
31/07/2023	14	15
07/08/2023	5	14
14/08/2023	5	9
21/08/2023	2	8
28/08/2023	2	1
04/09/2023	2	7
11/09/2023	1	3
18/09/2023	4	1
25/09/2023	1	2
02/10/2023	1	5
09/10/2023	3	3
16/10/2023	3	2
23/10/2023	1	5

30/10/2023	1	6
06/11/2023	7	1
13/11/2023	9	2
20/11/2023	6	2
27/11/2023	8	3
04/12/2023	11	1
11/12/2023	1	1
18/12/2023	1	6

Fonte: Autor

O SARIMAX apresentou um MAPE de 98.00% comparado ao 125.67% do SARIMA, indicando uma melhoria na precisão das previsões em termos percentuais. No entanto, um MAPE abaixo de 100% ainda é considerado alto e mostra que há espaço para melhorias.

Com um MAE de 53.02 para o SARIMAX e 53.14 para o SARIMA, observamos praticamente nenhum progresso na redução do erro médio absoluto, o que sugere que as variáveis exógenas não estão trazendo o benefício esperado na previsão dos casos. A Tabela 18 abaixo representa as medidas de desempenho do modelo para esse período.

Tabela 18- Previsão de casos primeiro período SARIMAX

<b>Medida de Desempenho</b>	<b>Valor</b>
<b>MAE</b>	53.02
<b>MAPE</b>	98.00%

Fonte: Autor

A comparação direta dos modelos para o primeiro período revela que, embora o SARIMAX tenha introduzido variáveis exógenas com o objetivo de capturar influências externas na transmissão da dengue, o modelo não conseguiu melhorar significativamente a precisão das previsões. A melhoria no MAPE é notável, mas não suficiente para justificar a escolha do SARIMAX sobre o SARIMA com base apenas nas métricas de desempenho.

#### **5.4.2 Segundo período (até março de 2024)**

O gráfico geral de previsão mostra as estimativas do SARIMAX, ver Figura 42, ao longo de um horizonte temporal amplo, onde se nota que as previsões acompanham a tendência geral dos dados reais, mas com algumas limitações. As previsões do SARIMAX são relativamente próximas aos dados reais, indicando um certo alinhamento inicial do

modelo com a sazonalidade dos casos de dengue. Durante os picos de incidência, observa-se uma subestimação notável dos casos, onde as previsões não alcançam os valores reais, indicando que o modelo pode não estar respondendo adequadamente a fatores críticos que impulsionam os surtos.

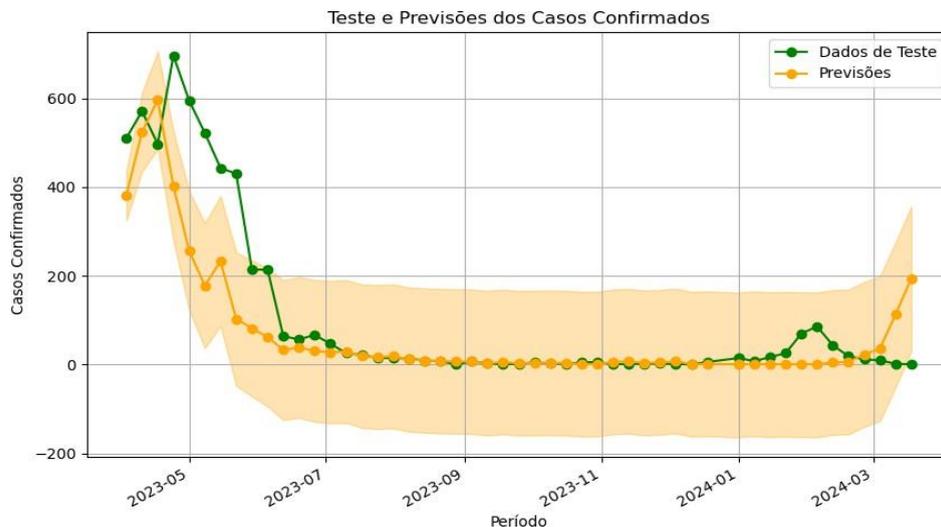
Figura 42- Previsão de casos segundo período SARIMAX



Fonte: Autor

O gráfico detalhado para o período de teste oferece uma visão mais precisa das previsões em relação aos dados reais, ver Figura 43. Há uma discrepância particular nos momentos de picos severos, onde o SARIMAX falha em prever a magnitude dos casos, apesar de seguir a tendência geral de aumento. Após os picos, o modelo parece se ajustar gradualmente, mas ainda não prevê a queda de casos de forma satisfatória, sugerindo que os ajustes do modelo podem não estar captando a dinâmica de queda tão rapidamente quanto necessário.

Figura 43- Segundo período de teste de previsões de casos SARIMAX



Fonte: Autor

Essa análise gráfica são corroboradas pela Tabela 19 abaixo.

Tabela 19 - Comparativo valores preditos X casos confirmados SARIMAX segundo período

<b>Data</b>	<b>Predição</b>	<b>Confirmados</b>
03/04/2023	381	510
10/04/2023	524	571
17/04/2023	596	497
24/04/2023	402	696
01/05/2023	257	595
08/05/2023	178	522
15/05/2023	233	443
22/05/2023	103	430
29/05/2023	82	214
05/06/2023	62	214
12/06/2023	33	64
19/06/2023	39	57
26/06/2023	31	67
03/07/2023	28	47
10/07/2023	30	26
17/07/2023	19	22
24/07/2023	17	15
31/07/2023	19	15
07/08/2023	12	14
14/08/2023	9	9
21/08/2023	8	8
28/08/2023	7	1
04/09/2023	7	7
11/09/2023	3	3
18/09/2023	6	1
25/09/2023	3	2
02/10/2023	4	5
09/10/2023	4	3
16/10/2023	4	2
23/10/2023	1	5
30/10/2023	1	6

06/11/2023	6	1
13/11/2023	8	2
20/11/2023	4	2
27/11/2023	5	3
04/12/2023	8	1
11/12/2023	1	1
18/12/2023	2	6
01/01/2024	1	15
01/01/2024	1	15
08/01/2024	2	8
15/01/2024	1	17
22/01/2024	1	26
29/01/2024	1	70
05/02/2024	1	86
12/02/2024	5	44
19/02/2024	6	19
26/02/2024	23	12
04/03/2024	36	10
11/03/2024	113	2
18/03/2024	194	1

Fonte: Autor

O MAPE ainda apresenta um valor elevado de 587,06%, mostra uma leve melhoria em relação ao SARIMA no mesmo período. Apesar disso, a precisão percentual das previsões permanece insuficiente.

O MAE de 56,16 não mostra uma melhoria significativa em relação ao SARIMA, reforçando a conclusão de que o SARIMAX não está fornecendo uma vantagem significativa em precisão de previsão. O MAE sugere que as previsões, em média, desviam aproximadamente 56 casos dos valores reais, o que pode ser um desvio considerável dependendo da escala e do contexto. A Tabela 20 abaixo representa as medidas de desempenho do modelo para esse período.

Tabela 20- Medidas de desempenho segundo período SARIMAX

<b>Medida de Desempenho</b>	<b>Valor</b>
<b>MAE</b>	56,16
<b>MAPE</b>	587,06%



O SARIMAX, mesmo com variáveis exógenas, não conseguiu uma melhoria significativa em termos de precisão das previsões quando comparado ao SARIMA, nem conseguiu reduzir o erro nas previsões de um período para outro. A capacidade do modelo de se ajustar rapidamente a mudanças significativas na transmissão da doença, como os picos pós-surto, permanece limitada. Isso sugere que os fatores incluídos no modelo podem não estar refletindo completamente os elementos que afetam a dinâmica da dengue ou que as defasagens e pesos atribuídos a essas variáveis exógenas podem precisar de ajustes adicionais.

## 6 CONCLUSÃO

O objetivo principal deste trabalho foi realizar uma análise aprofundada de séries temporais e comparar diferentes modelos para prever o número de casos de dengue na cidade de Porto Alegre, Rio Grande do Sul. A seguir, detalhamos como cada objetivo específico foi alcançado.

Inicialmente, o objetivo de coleta e pré-processamento de dados foi cumprido ao obter informações detalhadas sobre casos de dengue e variáveis climáticas para Porto Alegre, abrangendo o período de 2015 a março de 2024. Os dados foram obtidos de fontes confiáveis, como a Secretaria de Saúde do Estado do Rio Grande do Sul e o Instituto Nacional de Meteorologia (INMET), e passaram por um rigoroso processo de limpeza e normalização para garantir a qualidade das análises subsequentes.

Em seguida, os modelos de séries temporais SARIMA (Seasonal Autoregressive Integrated Moving Average) e SARIMAX (Seasonal Autoregressive Integrated Moving Average with Exogenous Variables) foram implementados e comparados. A aplicação desses modelos permitiu a análise detalhada das séries temporais de casos de dengue e variáveis climáticas. O modelo SARIMA capturou a sazonalidade inerente aos dados de dengue, enquanto o modelo SARIMAX, que incluiu variáveis climáticas como temperatura e precipitação, mostrou-se mais eficiente em termos de precisão preditiva.

O terceiro objetivo foi analisar o impacto de variáveis climáticas na previsão dos casos de dengue. A inclusão dessas variáveis no modelo SARIMAX revelou que fatores como temperatura e precipitação têm um impacto significativo na precisão das previsões. Este modelo adaptou-se melhor às condições ambientais, evidenciando a importância de considerar fatores exógenos na modelagem de doenças infecciosas.

A avaliação do desempenho dos modelos, o quarto objetivo, foi realizada através da comparação das previsões com os dados reais de casos de dengue. Embora o modelo SARIMAX tenha mostrado uma melhoria marginal em relação ao SARIMA, especialmente na captura de tendências sazonais, ambos os modelos enfrentaram dificuldades na previsão de picos de surtos. Isso sugere que, embora as variáveis climáticas melhorem a precisão das previsões, outros fatores não capturados pelos modelos podem influenciar significativamente a incidência de dengue.

Este estudo não apenas atingiu seus objetivos específicos, mas também proporcionou insights valiosos sobre a predição de casos de dengue em Porto Alegre. A

inclusão de variáveis climáticas nos modelos preditivos melhorou significativamente a precisão das previsões, destacando a importância de integrar fatores ambientais na análise de séries temporais de doenças infecciosas. Além disso, a análise dos resíduos dos modelos indicou que, embora as melhorias tenham sido marginais, ainda há espaço para aperfeiçoamentos, como a inclusão de novas variáveis explicativas ou a utilização de técnicas de modelagem mais avançadas.

As limitações deste estudo refletem os desafios inerentes à previsão de doenças infecciosas em contextos urbanos complexos como Porto Alegre. A modelagem baseada exclusivamente em dados meteorológicos e de incidência de dengue pode não abranger todos os fatores influentes, como a urbanização acelerada e mudanças nos comportamentos de armazenamento de água, que também podem afetar a propagação do vetor da dengue.

Em conclusão, a pesquisa avançou o conhecimento científico na área de modelagem preditiva de doenças infecciosas e forneceu uma base sólida para intervenções práticas e políticas públicas eficazes no combate à dengue. As ferramentas desenvolvidas podem ser aplicadas de maneira prática para melhorar a vigilância epidemiológica e o planejamento de medidas preventivas, beneficiando diretamente a população de Porto Alegre. Este estudo enfatiza a importância de continuar aprimorando os modelos preditivos e explorar novas abordagens que possam capturar melhor a complexidade da transmissão da dengue, contribuindo para a saúde pública e a qualidade de vida das populações afetadas.

## 7 REFERÊNCIAS BIBLIOGRÁFICAS

BARRERA, R.; AMADOR, M.; CLARK, G. G. Ecological factors influencing *Aedes aegypti* (Diptera: Culicidae) productivity in artificial containers in Salinas, Puerto Rico. **J Med Entomol.**, 2006 May; 43(3): 484-492. doi:10.1603/0022-2585(2006)43[484:EFIAAD]2.0.CO;2. PMID: 16739405.

BOX, G. E.; JENKINS, G. M.; REINSEL, G. C.; LJUNG, G. M. Time Series Analysis: Forecasting and Control. John Wiley & Sons, 2015.

BRISCOE, M. S. *Aedes Aegypti* The Yellow Fever Mosquito, Its Life History, Bionomics And Structure. **J Natl Med Assoc.**, 1962 Jan; 54(1): 132. PMID: PMC2642088.

BROCKWELL, Peter J.; DAVIS, Richard A., eds. **Introduction to Time Series and Forecasting**. New York, NY: Springer New York, 2002.

BURNHAM, Kenneth P.; ANDERSON, David R., eds. **Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach**. New York, NY: Springer New York, 2002.

CAMPBELL-LENDUM, D.; MANGA, L.; BAGAYOKO, M.; SOMMERFELD, J. Climate change and vector-borne diseases: what are the implications for public health research and policy? **Philos Trans R Soc Lond B Biol Sci.**, 2015 Apr 5; 370(1665): 20130552. doi: 10.1098/rstb.2013.0552. PMID: 25688013; PMID: PMC4342958.

CHRISTOPHERS, S.R. *Aedes aegypti* (L.), the Yellow Fever Mosquito: Its Life History, Bionomics and Structure. Cambridge: Cambridge University Press, 1960.

DELATTE, H.; GIMONNEAU, G.; TRIBOIRE, A.; FONTENILLE, D. Influence of temperature on immature development, survival, longevity, fecundity, and gonotrophic cycles of *Aedes albopictus*, vector of chikungunya and dengue in the Indian Ocean. **J Med Entomol.**, 2009 Jan; 46(1): 33-41. doi: 10.1603/033.046.0105. PMID: 19198515.

DE LIVERA, Alysha M.; HYNDMAN, Rob J.; SNYDER, Ralph D. Forecasting time series with complex seasonal patterns using exponential smoothing. **Journal of the American Statistical Association**, v. 106, n. 496, p. 1513-1527, 2011.

DIAS, Juliana; BORGES, Felipe. **Aumento de casos de dengue no RS pode ter relação com fenômenos climáticos**. Portal do Estado do Rio Grande do Sul. Rio Grande do Sul, 08 fev. 2024. Disponível em: <https://estado.rs.gov.br/aumento-de-casos-de-dengue-no-rs-pode-ter-relacao-com-fenomenos-climaticos>. Acesso em: 21 fev. 2024.

EDMAN, J. D.; SCOTT, T. W.; COSTERO, A.; MORRISON, A. C.; HARRINGTON, L. C.; CLARK, G. G. *Aedes aegypti* (Diptera: Culicidae) movement influenced by availability of oviposition sites. **J Med Entomol.**, 1998 Jul; 35(4): 578-583. doi: 10.1093/jmedent/35.4.578. PMID: 9701948.

FARNESI, Luana Cristina et al. Embryonic development of *Aedes aegypti* (Diptera: Culicidae): influence of different constant temperatures. **Memórias do Instituto Oswaldo Cruz**, v. 104, p. 124-126, 2009.

FOCKS, D. A.; HAILE, D. G.; DANIELS, E.; MOUNT, G. A. Dynamic life table model for *Aedes aegypti* (Diptera: Culicidae): analysis of the literature and model development. **J Med Entomol.**, 1993 Nov; 30(6): 1003-1017. doi: 10.1093/jmedent/30.6.1003. PMID: 8271242.

GILLETT, J. D. **The mosquito: its life, activities, and impact on human affairs**. Garden City, N.Y: Doubleday, 1972.

GILPIN, M. E.; MCCLELLAND, G. A. H. Systems Analysis of the Yellow Fever Mosquito *Aedes aegypti*. **Fortschr Zool**, 1979.

GRANGER, Clive W. J.; NEWBOLD, Paul. Forecasting transformed series. **Journal of the Royal Statistical Society: Series B (Methodological)**, v. 38, n. 2, p. 189- 203, 1976.

HAMILTON, James D.; SUSMEL, Raul. Autoregressive conditional heteroskedasticity and changes in regime. **Journal of Econometrics**, v. 64, n. 1-2, p. 307- 333, 1994.

HAWLEY, W. A. The biology of *Aedes albopictus*. **J Am Mosq Control Assoc Suppl.**, 1988 Dec; 1: 1-39. PMID: 3068349.

HTUN, H. H.; BIEHL, M.; PETKOV, N. Forecasting relative returns for S&P 500 stocks using machine learning. **Financial Innovation**, 10, 118, 2024. Disponível em: <https://doi.org/10.1186/s40854-024-00644-0>.

HYNDMAN, Rob J.; KOEHLER, Anne B. **Another look at measures of forecast accuracy. International Journal of Forecasting**, v. 22, n. 4, p. 679-688, 2006.

HYNDMAN, Rob J.; ATHANASOPOULOS, George. **Forecasting: Principles and Practice**. OTexts, 2018.

JOHANSSON, M. A.; CUMMINGS, D. A.; GLASS, G. E. Multiyear climate variability and dengue—El Niño southern oscillation, weather, and dengue incidence in Puerto Rico, Mexico, and Thailand: a longitudinal data analysis. **PLoS Med.**, 2009 Nov; 6(11): e1000168. doi: 10.1371/journal.pmed.1000168. Epub 2009 Nov 17. PMID: 19918363; PMCID: PMC2771282.

JUDSON, C. L. The Physiology of Mosquito Larvae. In: **The Bionomics of Mosquitoes**. Oxford: Oxford University Press, 1996.

JULIANO, S. A.; LOUNIBOS, L. P. Ecology of invasive mosquitoes: effects on resident species and on human health. **Ecol Lett.**, 2005 May; 8(5): 558-574. doi: 10.1111/j.1461-0248.2005.00755. PMID: 17637849; PMCID: PMC1920178.

LIU, N. Insecticide resistance in mosquitoes: impact, mechanisms, and research directions. **Ann Rev Entomol.**, 2015; 60: 537-559.

LEAKE, Colin. The biology of mosquitoes. Volume 1: Development, nutrition and reproduction: A. N. Clements. London: Chapman & Hall, 1992. xxii+509 pp. Price £50. ISBN 0-412-40180-0. **Transactions of The Royal Society of Tropical Medicine and Hygiene**, 1993 Nov-Dec; 87(6): 718. doi: 10.1016/0035-9203(93)90329-O.

LJUNG, Greta M.; BOX, George E. P. On a measure of lack of fit in time series models. **Biometrika**, v. 65, n. 2, p. 297-303, 1978.

LOWE, Rachel et al. Climate services for health: predicting the evolution of the

2016 dengue season in Machala, **Ecuador**. **The Lancet Planetary Health**, v. 1, n. 4, p. e142-e151, 2017.

MOGI, M. et al. Inter-and intraspecific variation in resistance to desiccation by adult *Aedes* (*Stegomyia*) spp. (Diptera: Culicidae) from Indonesia. **Journal of Medical Entomology**, v. 33, n. 1, p. 53-57, 1996.

MORETTIN, P. A. **Análise de Séries Temporais**. São Paulo: Edgard Blücher, 2010.

MOGI, M. et al. Inter-and intraspecific variation in resistance to desiccation by adult *Aedes* (*Stegomyia*) spp. (Diptera: Culicidae) from Indonesia. **Journal of Medical Entomology**, v. 33, n. 1, p. 53-57, 1996.

MORRISON, A. C.; GRAY, K.; GETIS, A.; ASTETE, H.; SIHUINCHA, M.; FOCKS, D.; WATTS, D.; STANCIL, J. D.; OLSON, J. G.; BLAIR, P.; SCOTT, T. W. Temporal and geographic patterns of *Aedes aegypti* (Diptera: Culicidae) production in Iquitos, Peru. **J Med Entomol.**, 2004 Nov; 41(6): 1123-1142. doi: 10.1603/0022-2585-41.6.1123. PMID: 15605653.

**OPENAI**. ChatGPT, assistência por inteligência artificial. [S.l.]: OpenAI, 2024.

**PAHO**. [S.l.], 6 fev. 2024. Disponível em: <https://www.paho.org/pt/topicos/dengue#:~:text=Em%202017%2C%20houve%20uma%20redu%C3%A7%C3%A3o,aumento%20de%20casos%20em%202017>. Acesso em: 16 abr. 2024.

PAUPY, Christophe et al. *Aedes albopictus*, an arbovirus vector: from the darkness to the light. **Microbes and Infection**, v. 11, n. 14-15, p. 1177-1185, 2009.

PONGSUMPUN, P.; GARCIA LOPEZ, D.; FAVIER, C.; TORRES, L.; LLOSA, J.; DUBOIS, M. A. Dynamics of dengue epidemics in urban contexts. **Trop Med Int Health.**, 2008 Sep; 13(9): 1180-1187. doi: 10.1111/j.1365-3156.2008.02124.x. Epub 2008 Aug 20. PMID: 18840157.

SCOTT, T. W.; AMERASINGHE, P. H.; MORRISON, A. C.; LORENZ, L. H.; CLARK, G. G.; STRICKMAN, D.; KITTAYAPONG, P.; EDMAN, J. D. Longitudinal

studies of *Aedes aegypti* (Diptera: Culicidae) in Thailand and Puerto Rico: blood feeding frequency. **J Med Entomol.**, 2000 Jan; 37(1): 89-101. doi: 10.1603/0022-2585-37.1.89. PMID: 15218911.

SCHWARZ, Gideon. **Estimating the dimension of a model.** **The Annals of Statistics**, 1978, p. 461-464.

SHAMAN, J. et al. The impact of a simulated drought on the transmission dynamics of dengue fever. **Pediatric Infectious Disease Journal**, 2005.

REISKIND, M. H.; LOUNIBOS, L. P. Effects of intraspecific larval competition on adult longevity in the mosquitoes *Aedes aegypti* and *Aedes albopictus*. **Medical and Veterinary Entomology**, v. 23, n. 1, p. 62-68, 2009.

REZENDE, G. L.; MARTINS, A. J.; GENTILE, C.; FARNESE, L. C.; PELAJOMACHADO, M.; PEIXOTO, A. A.; VALLE, D. Embryonic desiccation resistance in *Aedes aegypti*: presumptive role of the chitinized serosal cuticle. **BMC Dev Biol.**, 2008 Sep 13; 8: 82. doi: 10.1186/1471-213X-8-82. PMID: 18789161; PMCID: PMC2561029.

RUEDA, L. M.; PATEL, K. J.; AXTELL, R. C.; STINNER, R. E. Temperature-dependent development and survival rates of *Culex quinquefasciatus* and *Aedes aegypti* (Diptera: Culicidae). **J Med Entomol.**, 1990 Sep; 27(5): 892-898. doi: 10.1093/jmedent/27.5.892. PMID: 2231624.

**RS confirma 17ª morte provocada por dengue em 2024; vítima residia em Santa Rosa:** De acordo com a Secretaria da Saúde, trata-se de um homem de 75 anos com comorbidades. Estado tem 15,8 mil casos confirmados da doença. [S.I.], 11 mar. 2024. Disponível em: <https://g1.globo.com/rs/rio-grande-do-sul/noticia/2024/03/11/rs-confirma-17a-morte-provocada-por-dengue-em-2024-vitima-residia-em-santa-rosa.ghtml>. Acesso em: 16 abr. 2024.

RYAN, S. J.; MCNALLY, A.; JOHNSON, L. R.; MORDECAI, E. A.; BEN-HORIN, T.; PAAIJMANS, K.; LAFFERTY, K. D. Mapping physiological suitability limits for malaria in Africa under climate change. **Vector Borne Zoonotic Dis.**, 2015 Dec; 15(12): 718-725. doi: 10.1089/vbz.2015.1822. Epub 2015 Nov 18. PMID: 26579951; PMCID: PMC4700390.

TSAY, Ruey S. **Analysis of Financial Time Series**. John Wiley & Sons, 2005.

TUN-LIN, W.; BURKOT, T. R.; KAY, B. H. Effects of temperature and larval diet on development rates and survival of the dengue vector *Aedes aegypti* in north Queensland, Australia. **Med Vet Entomol.**, 2000 Mar; 14(1): 31-37. doi: 10.1046/j.1365-2915.2000.00207.x. PMID: 10759309.

TUN-LIN, W.; KAY, B. H.; BARNES, A.; FORSYTH, S. Critical examination of *Aedes aegypti* indices: correlations with abundance. **Am J Trop Med Hyg.**, 1996 May; 54(5): 543-547. doi: 10.4269/ajtmh.1996.54.543. PMID: 8644913.

WU, P. C.; LAY, J. G.; GUO, H. R.; LIN, C. Y.; LUNG, S. C.; SU, H. J. Higher temperature and urbanization affect the spatial patterns of dengue fever transmission in subtropical Taiwan. **Sci Total Environ.**, 2009 Mar 15; 407(7): 2224-2233. doi: 10.1016/j.scitotenv.2008.11.034. Epub 2009 Jan 20. PMID: 191575