



Universidade Federal de Uberlândia
Instituto de Matemática e Estatística

Bacharelado em Estatística

**INFLUÊNCIA DE VARIÁVEIS
CLIMÁTICAS DEFASADAS NA
INCIDÊNCIA DE DENGUE EM
UBERLÂNDIA-MG: UMA ANÁLISE DE
SÉRIES TEMPORAIS**

Giulia Ferreira Gonzaga de Oliveira

Uberlândia-MG

2026

Giulia Ferreira Gonzaga de Oliveira

**INFLUÊNCIA DE VARIÁVEIS
CLIMÁTICAS DEFASADAS NA
INCIDÊNCIA DE DENGUE EM
UBERLÂNDIA-MG: UMA ANÁLISE DE
SÉRIES TEMPORAIS**

Trabalho de conclusão de curso apresentado à Coordenação do Curso de Bacharelado em Estatística como requisito parcial para obtenção do grau de Bacharel em Estatística.

Orientador: Profa. Dra. Nadia Giaretta Biase

**Uberlândia-MG
2026**



**Universidade Federal de Uberlândia
Instituto de Matemática e Estatística**

Coordenação do Curso de Bacharelado em Estatística

A banca examinadora, conforme abaixo assinado, certifica a adequação deste trabalho de conclusão de curso para obtenção do grau de Bacharel em Estatística.

Uberlândia, ____ de _____ de 20____

BANCA EXAMINADORA

Profa. Dra. Nadia Giaretta Biase

Prof. Dr. Ednaldo Carvalho Guimarães

Prof. Dr. Rildo Aparecido Costa

**Uberlândia-MG
2026**

AGRADECIMENTOS

Em primeiro lugar, agradeço a Deus por ter me sustentado até aqui. Agradeço ao meu pai, Giovanni, e à minha mãe, Andressa, por sempre acreditarem em mim e se esforçarem para que este sonho se tornasse realidade. Agradeço também aos meus irmãos, Giovanna e Arthur, por todo o apoio.

Agradeço aos amigos que trilharam esse caminho junto a mim, em especial a Maria Cecília, que desde o primeiro período se fez presente em todos os momentos. Sem o seu apoio, incentivo e colaboração, todos esses anos teriam sido muito mais pesados. A Elisa e a Gabriela, por terem cruzado o meu caminho e trazerem consigo leveza, amor e momentos felizes. Vocês três deixaram tudo mais especial.

Agradeço ao meu namorado, Vitor, que, apesar de ter se juntado a mim na reta final desta caminhada, esteve ao meu lado todos os dias, incentivando, apoiando e dando força. Agradeço ainda aos familiares e amigos que, mesmo de longe, se fizeram presentes, sempre oferecendo apoio e motivação.

Agradeço à minha orientadora, Nádia Giaretta Biase, por toda a ajuda, disponibilidade, correções e orientação. Por fim, agradeço aos professores Ednaldo Carvalho Guimarães e Rildo Aparecido Costa por aceitarem o convite para compor a banca examinadora e contribuírem para este trabalho.

RESUMO

A dengue é uma doença presente de forma contínua no Brasil, e sua transmissão está associada a condições climáticas, especialmente temperatura e precipitação. O presente trabalho teve como objetivo analisar o comportamento histórico dos casos de dengue em Uberlândia-MG, no período de janeiro de 2014 a abril de 2026, por meio de modelos de séries temporais das classes SARIMA e SARIMAX, avaliando a influência de variáveis climáticas defasadas no tempo. Os dados de casos foram obtidos do Sistema de Informação de Agravos de Notificação (SINAN), e os dados climáticos de temperatura e precipitação foram obtidos da plataforma NASA POWER. A metodologia seguiu a abordagem de Box-Jenkins, incluindo teste de tendência (Cox-Stuart) e de sazonalidade (Fisher), transformação logarítmica para estabilização da variância, identificação e estimação de modelos SARIMA, e posteriormente a incorporação de variáveis exógenas climáticas defasadas, selecionadas por meio da função de correlação cruzada (CCF). O modelo SARIMA(1, 1, 0)(1, 1, 1)₁₂ foi selecionado como modelo de referência, apresentando os menores critérios de informação entre os modelos com coeficientes estatisticamente significativos. A inclusão da temperatura defasada em cinco meses como variável exógena resultou no modelo SARIMAX final, que apresentou menor AIC e coeficiente estatisticamente significativo, evidenciando uma associação inversa entre a temperatura e a incidência posterior da doença. O diagnóstico dos resíduos indicou ausência de autocorrelação nas primeiras defasagens, mas autocorrelação residual significativa nas defasagens sazonais mais longas, atribuída ao comportamento atípico do surto de 2025. A validação foi realizada por meio de previsão *rolling* (um passo à frente), tendo em vista o caráter atípico desse surto. Os modelos mostraram-se complementares, pois o SARIMA apresentou desempenho preditivo ligeiramente superior, enquanto o SARIMAX permitiu quantificar a influência da temperatura sobre os casos. Os resultados reforçam o potencial das variáveis climáticas como instrumento de apoio à vigilância epidemiológica e ao planejamento de ações de controle vetorial no município.

Palavras-chave: Box-Jenkins, variáveis exógenas, arbovirose, epidemia, saúde pública.

ABSTRACT

Dengue is an endemic disease in Brazil, and its transmission is associated with climatic conditions, particularly temperature and precipitation. This study aimed to analyze the historical behavior of dengue cases in Uberlândia, Minas Gerais, Brazil, from January 2014 to April 2026, through SARIMA and SARIMAX time series models, evaluating the influence of time-lagged climatic variables. Case data were obtained from the Notifiable Diseases Information System (SINAN), and climatic data on temperature and precipitation were obtained from the NASA POWER platform. The methodology followed the Box-Jenkins approach, including a trend test (Cox-Stuart) and a seasonality test (Fisher), logarithmic transformation to stabilize variance, identification and estimation of SARIMA models, and subsequent incorporation of lagged climatic exogenous variables, selected through cross-correlation function (CCF) analysis. The SARIMA(1, 1, 0)(1, 1, 1)₁₂ model was selected as the reference model, presenting the lowest information criterion values among the models with statistically significant coefficients. The inclusion of temperature lagged by five months as an exogenous variable resulted in the final SARIMAX model, which showed a lower AIC and a statistically significant coefficient, evidencing an inverse association between temperature and the subsequent incidence of the disease. Residual diagnostics indicated the absence of autocorrelation at the first lags, but significant residual autocorrelation at the longer seasonal lags, attributed to the atypical behavior of the 2025 outbreak. Model validation was performed through rolling (one-step-ahead) forecasting, given the atypical nature of that outbreak. The models proved to be complementary, as SARIMA showed slightly higher predictive performance, while SARIMAX made it possible to quantify the influence of temperature on dengue incidence. The results reinforce the potential of climatic variables as a tool to support epidemiological surveillance and vector control planning in the municipality.

Keywords: Box-Jenkins, exogenous variables, arbovirus, epidemic, public health.

SUMÁRIO

Lista de Figuras	I
Lista de Tabelas	II
1 Introdução	1
2 Fundamentação Teórica	3
2.1 A Dengue e sua Relação com Fatores Climáticos	3
2.2 Séries Temporais	4
2.2.1 Estacionariedade	5
2.2.2 Tendência e Sazonalidade	5
2.2.3 Funções de Autocorrelação e Autocorrelação Parcial	7
2.3 Modelos ARIMA, SARIMA e SARIMAX	8
2.3.1 Modelo Autorregressivo Integrado de Médias Móveis (ARIMA)	8
2.3.2 Modelo SARIMA	9
2.3.3 Modelo SARIMAX	9
2.4 Modelos SARIMAX Aplicados à Epidemiologia da Dengue	10
2.5 Metodologia de Box-Jenkins	10
2.6 Função de Correlação Cruzada (CCF)	11
2.7 Critérios de Seleção do Modelo	11
2.7.1 Critério de Informação de Akaike (AIC)	11
2.7.2 Critério de Informação Bayesiano (BIC)	11
2.8 Diagnóstico dos Resíduos	12
2.9 Métricas de Avaliação do Modelo	13
3 Metodologia	14
3.1 Caracterização da Pesquisa	14
3.2 Fonte e Período dos Dados	14
3.2.1 Dados de Casos de Dengue	14
3.2.2 Dados Climáticos	14
3.3 Divisão entre Treino e Validação	15
3.4 Pré-processamento dos Dados	15
3.4.1 Transformação da Série de Casos	15
3.4.2 Transformação das Variáveis Climáticas	15
3.5 Testes de Hipótese	15
3.6 Identificação, Estimação e Diagnóstico dos Modelos SARIMA	16
3.7 Incorporação de Variáveis Climáticas Exógenas	16
3.7.1 Identificação das Defasagens Temporais	16
3.7.2 Seleção do Lag e da Variável	16
3.7.3 Padronização das Variáveis Exógenas	17
3.8 Validação dos Modelos	17

3.9	Softwares Utilizados	18
4	Resultados	19
4.1	Análise Exploratória	19
4.2	Testes de Hipótese	23
4.3	Seleção do Modelo SARIMA	24
4.3.1	Diagnóstico de Resíduos – Modelo M4	24
4.3.2	Validação por Previsão Rolling – Modelo M4	26
4.4	Incorporação de Variáveis Climáticas – Modelo SARIMAX	27
4.4.1	Função de Correlação Cruzada	27
4.4.2	Seleção do Lag por Busca Sistemática	28
4.4.3	Modelo SARIMAX Final	29
4.4.4	Validação por Previsão Rolling – Modelo SARIMAX	31
4.5	Comparativo Final entre os Modelos	31
5	Considerações Finais	33
	Referências Bibliográficas	35
	Apêndice A Dados Utilizados	38

LISTA DE FIGURAS

4.1	Série mensal de casos de dengue em Uberlândia-MG (2014–2025).	19
4.2	Funções de autocorrelação (FAC) e de autocorrelação parcial (FACP) da série original de casos.	20
4.3	Relação entre média e desvio padrão em subperíodos da série original.	20
4.4	Série de casos de dengue após transformação logarítmica.	21
4.5	Relação entre média e desvio padrão em subperíodos da série logarítmica, indicando estabilização da variância.	21
4.6	FAC e FACP da série logarítmica de casos.	22
4.7	Série logarítmica após uma diferenciação simples ($d = 1$).	22
4.8	FAC e FACP da série logarítmica diferenciada, utilizadas na identificação dos ordens do modelo.	23
4.9	Resíduos do modelo M4 SARIMA(1, 1, 0)(1, 1, 1) ₁₂ ao longo do tempo.	25
4.10	Histograma dos resíduos do modelo M4.	25
4.11	FAC e FACP dos resíduos do modelo M4.	26
4.12	Periodograma acumulado dos resíduos do modelo M4.	26
4.13	Função de correlação cruzada (CCF) entre as variáveis climáticas e a série logarítmica de casos de dengue.	28
4.14	Resíduos do modelo SARIMAX (temperatura defasada em cinco meses) ao longo do tempo.	30
4.15	Periodograma acumulado dos resíduos do modelo SARIMAX.	30

LISTA DE TABELAS

4.1	Resultados dos testes de hipótese aplicados à série.	23
4.2	Comparativo de critérios de informação entre os modelos SARIMA candidatos. .	24
4.3	Coefficientes estimados do modelo M4.	24
4.4	Validação <i>rolling</i> do modelo M4 SARIMA.	27
4.5	Melhores configurações da busca sistemática de defasagens.	28
4.6	Coefficientes estimados do modelo SARIMAX com temperatura defasada em 5 meses.	29
4.7	Validação <i>rolling</i> do modelo SARIMAX.	31
4.8	Comparativo final entre os modelos M4 SARIMA e SARIMAX.	31
A.1	Casos mensais confirmados de dengue em Uberlândia-MG.	38
A.2	Temperatura média mensal (°C) em Uberlândia-MG.	38
A.3	Precipitação total mensal (mm) em Uberlândia-MG obtida da plataforma NASA POWER.	39

1. INTRODUÇÃO

A dengue é uma arbovirose transmitida pelo mosquito *Aedes aegypti* e representa um dos maiores desafios da saúde pública no Brasil. A doença está presente de forma contínua no país, mas, em determinados anos, enfrenta surtos epidêmicos de grande magnitude, que pressionam o sistema de saúde e geram consequências sociais e econômicas relevantes. Uberlândia-MG segue esse padrão: o município apresenta histórico recorrente de epidemias de dengue, com registros de surtos expressivos ao longo das últimas décadas, como o de 2019, apontado por Leite [12] como o maior da série por ele analisada.

A transmissão da dengue não depende apenas da presença do vetor, mas também das condições ambientais. Viana e Ignotti [23] destacam que, embora fatores urbanos e sociais influenciem a dinâmica da doença, as condições climáticas, especialmente temperatura e precipitação, desempenham papel importante na determinação do momento e da intensidade dos surtos. A precipitação favorece a formação de criadouros ao acumular água em recipientes, enquanto a temperatura acelera o ciclo reprodutivo do mosquito e a replicação viral.

Um ponto importante dessa relação é que seus efeitos não ocorrem de forma imediata. Entre a variação climática e o aumento nas notificações existe um intervalo de tempo, conhecido como defasagem temporal (*lag*). Esse intervalo envolve o desenvolvimento larval do mosquito, sua maturação, o processo de transmissão e o período de incubação do vírus até a manifestação clínica e o registro do caso. A identificação desse intervalo é um dos focos deste trabalho, que busca explorar o papel das defasagens temporais na relação entre clima e dengue. Evidências desse comportamento já foram observadas na literatura. Souza et al. [21], ao modelarem a incidência de dengue na Paraíba por meio de modelos de defasagem distribuída, identificaram um intervalo de até cinco meses entre as condições climáticas e alterações na tendência da incidência.

Uberlândia-MG apresenta características que a tornam um cenário relevante para esse tipo de análise. O município apresenta clima tropical bem definido, com estação chuvosa e quente entre outubro e março, além de um processo de urbanização acelerado. Essas características contribuem para a ocorrência recorrente de epidemias de dengue observadas no município ao longo das últimas décadas [12]. A análise do período de janeiro de 2014 a abril de 2026, com dados do Sistema de Informação de Agravos de Notificação (SINAN) e da plataforma NASA POWER, permite observar tanto episódios epidêmicos quanto um padrão sazonal mais regular da doença.

Compreender como a dengue responde às variações climáticas pode contribuir para o plane-

jamento em saúde pública. Estimar o intervalo entre mudanças climáticas e aumento de casos pode oferecer uma janela de tempo útil para a antecipação de ações de controle vetorial, antes do crescimento mais acentuado da curva epidêmica.

Nesse contexto, a análise de séries temporais se mostra uma abordagem adequada. Modelos da classe SARIMA (*Seasonal Autoregressive Integrated Moving Average*) e sua extensão SARIMAX (*Seasonal Autoregressive Integrated Moving Average with Exogenous Regressors*), que incorpora variáveis exógenas, permitem representar tendência, sazonalidade e a influência de covariáveis climáticas sobre a evolução da doença [15]. Neste trabalho, essas variáveis são consideradas com defasagens, cuja identificação é realizada por meio da função de correlação cruzada (CCF). As defasagens identificadas são então incorporadas aos modelos, com o objetivo de avaliar a influência da temperatura e da precipitação sobre a incidência futura de casos.

Os sucessivos surtos registrados no município ao longo dos últimos anos indicam que estratégias baseadas principalmente na reação após o aumento de casos têm se mostrado insuficientes para conter a expansão da doença. Isso reforça a necessidade de abordagens com maior capacidade de antecipação. Do ponto de vista metodológico, o uso de modelos de séries temporais como SARIMA e SARIMAX permite incorporar aspectos como sazonalidade, tendência e defasagens que não são capturados por abordagens mais simples. Do ponto de vista aplicado, a possibilidade de antecipar períodos de maior risco pode apoiar ações de controle vetorial, alocação de recursos e planejamento de campanhas preventivas.

Diante disso, este trabalho tem como objetivo modelar a série temporal mensal de casos de dengue em Uberlândia-MG por meio da metodologia de Box e Jenkins, buscando identificar um modelo capaz de descrever a dinâmica da doença e avaliar o efeito defasado das variáveis climáticas sobre sua incidência. Especificamente, pretende-se caracterizar a evolução dos casos entre 2014 e 2026, identificar padrões de tendência e sazonalidade, analisar a relação entre temperatura e precipitação com a incidência por meio da função de correlação cruzada, investigar a existência de defasagens temporais, ajustar e comparar modelos SARIMA e SARIMAX e avaliar a capacidade preditiva dos modelos por meio de previsões para períodos subsequentes ao intervalo analisado.

2. FUNDAMENTAÇÃO TEÓRICA

2.1 A DENGUE E SUA RELAÇÃO COM FATORES CLIMÁTICOS

A relação entre a dengue e o clima está diretamente ligada à biologia de seu vetor, o mosquito *Aedes aegypti*, cujo ciclo de vida e capacidade de transmissão dependem das condições ambientais locais. Nesse contexto, temperatura, precipitação e umidade relativa do ar são apontadas como determinantes ambientais chave para a proliferação do vetor [23].

Estudos conduzidos em diferentes regiões do Brasil corroboram essa relação, embora com particularidades locais importantes. Azevedo et al. [3], ao analisarem a dinâmica temporal da dengue no Distrito Federal entre 2013 e 2023, identificaram associação positiva entre o aumento da temperatura e da precipitação e a elevação dos casos da doença. Os autores destacam o ano de 2022 como o de maior incidência do período, com média anual próxima de 40 mil casos, coincidindo com os maiores valores dessas variáveis na série. Ressaltam, contudo, que essa relação não é estritamente linear, sendo também influenciada por fatores sociais e comportamentais, como sugerido pela queda abrupta de casos durante o período de isolamento social da pandemia de COVID-19 em 2021, mesmo sob condições climáticas favoráveis ao vetor.

Em pesquisa realizada em Rondonópolis-MT, Santos et al. [20] ajustaram modelos de regressão da classe dos Modelos Lineares Generalizados (Poisson e binomial negativo) para avaliar o efeito da temperatura média, da temperatura máxima e da umidade relativa do ar sobre os casos de dengue notificados entre 2001 e 2015. Os resultados indicaram uma relação *inversa* entre a temperatura média e o número de casos, estimando-se uma redução de aproximadamente 43% a cada grau Celsius de aumento. Por outro lado, observou-se relação *direta* com a temperatura máxima (aumento de aproximadamente 77% por grau) e com a umidade relativa do ar (aumento de aproximadamente 17% por ponto percentual). Nesse estudo, não foi identificada associação estatisticamente significativa entre a precipitação pluviométrica e a incidência de dengue no município, o que ilustra a heterogeneidade geográfica dessas relações e reforça a necessidade de análises específicas para cada localidade.

Essa relação inversa entre temperatura média e incidência, à primeira vista contraintuitiva, é explicada pelos próprios autores a partir do ciclo biológico do vetor. O desenvolvimento do *Aedes aegypti* ocorre de forma mais eficiente dentro de uma faixa específica de temperatura (entre aproximadamente 24°C e 28°C, segundo Fernandes et al., citados por Santos et al. [20]).

Assim, temperaturas médias muito elevadas podem, em determinados contextos, ultrapassar o intervalo ótimo para o desenvolvimento do mosquito. Esse tipo de evidência reforça que a relação entre temperatura e incidência de dengue não é uniforme entre diferentes localidades.

Especificamente para o município de Uberlândia-MG, objeto deste estudo, Leite [12], em tese de doutorado defendida no Programa de Pós-Graduação em Geografia da Universidade Federal de Uberlândia, analisou a relação entre variáveis climáticas e os casos de dengue no período de 2010 a 2019. O autor identificou que o aumento dos casos esteve associado à temperatura mínima acima de 18°C , à temperatura média principalmente na faixa de 26°C a 28°C e à umidade relativa do ar acima de 70%. Por outro lado, observou-se redução dos casos em anos com médias mensais de temperatura máxima acima de 30°C , sugerindo que temperaturas muito elevadas podem comprometer a sobrevivência do vetor, conforme descrito por Santos et al. [20].

Leite [12] destaca ainda que a maior concentração histórica dos casos de dengue em Uberlândia ocorre entre outubro e março, período que coincide com a estação chuvosa da região. O autor também aponta que o ano de 2019 representou o maior surto da série por ele analisada, com 31.328 casos notificados. Considerando a continuidade dos surtos observados no país nos anos mais recentes, reforça-se a importância de uma análise atualizada da série de Uberlândia, incorporando os anos posteriores ao período estudado.

2.2 SÉRIES TEMPORAIS

Denomina-se série temporal qualquer sequência de dados observados e registrados em sucessivos instantes de tempo, geralmente espaçados de forma uniforme [6]. O que diferencia esse tipo de dado de um conjunto de observações independentes é a presença de dependência entre valores ao longo do tempo: uma observação em determinado instante tende a estar relacionada com aquelas que a precederam. É essa estrutura de dependência que se busca descrever e modelar, com o objetivo de compreender o mecanismo gerador dos dados, caracterizar seu comportamento e gerar previsões para períodos futuros [16].

No estudo de séries temporais, é comum considerar que o comportamento da série resulta da combinação de diferentes componentes. A tendência está associada a movimentos de longo prazo, enquanto a sazonalidade está ligada a oscilações que se repetem em intervalos regulares. A componente cíclica refere-se a flutuações de duração variável, e a componente irregular ou aleatória reúne variações não explicadas pelas demais [10]. A identificação e o tratamento dessas componentes são etapas centrais no processo de modelagem. No caso da dengue, essa decomposição é especialmente relevante, já que a doença apresenta um padrão sazonal bem definido e pode sofrer alterações ao longo do tempo em função de fatores climáticos, ambientais e sociais. Essa estrutura será explorada na modelagem dos casos de dengue analisados neste trabalho.

2.2.1 ESTACIONARIEDADE

Diz-se que uma série temporal é estacionária quando suas características estatísticas permanecem estáveis ao longo do tempo. Em termos práticos, isso significa que a série oscila em torno de um nível médio aproximadamente constante, com dispersão que não se altera à medida que o tempo avança [6]. Essa propriedade é requisito para grande parte das técnicas de modelagem de séries temporais. Na prática, porém, é comum que as séries observadas não apresentem esse comportamento de forma espontânea, exibindo variações no nível ou na inclinação ao longo do período analisado [16].

Quando a série não satisfaz a condição de estacionariedade, recorre-se a transformações que a aproximem desse comportamento. O procedimento mais usual é a tomada de diferenças entre observações consecutivas, repetida tantas vezes quanto necessário. Seguindo a notação de Morettin e Toloi [16], a diferença de primeira ordem da série Z_t é definida como:

$$\Delta Z_t = Z_t - Z_{t-1}. \quad (2.1)$$

Se uma única diferenciação não for suficiente, pode-se aplicar a diferença de segunda ordem, obtida ao diferenciar novamente a série já transformada:

$$\Delta^2 Z_t = \Delta(\Delta Z_t) = Z_t - 2Z_{t-1} + Z_{t-2}. \quad (2.2)$$

De forma geral, a diferença de ordem n corresponde à aplicação sucessiva do operador de diferença:

$$\Delta^n Z_t = \Delta(\Delta^{n-1} Z_t). \quad (2.3)$$

Na maioria das aplicações, uma ou, no máximo, duas diferenciações são suficientes para que a série se torne estacionária [16].

A diferenciação atua principalmente sobre a média da série, mas não corrige instabilidades na variância. Quando a dispersão dos dados varia ao longo do tempo, caracterizando heterocedasticidade, é recomendável aplicar previamente uma transformação que estabilize a variância. Entre as alternativas disponíveis está a família de transformações de Box-Cox [4], controlada por um parâmetro λ estimado a partir dos dados. O caso particular $\lambda = 0$ corresponde à aplicação do logaritmo natural, transformação frequentemente utilizada quando a variância cresce proporcionalmente ao nível da série.

2.2.2 TENDÊNCIA E SAZONALIDADE

Uma forma comum de interpretar uma série temporal consiste em decompô-la em partes que representam comportamentos distintos. Considerando as observações Z_t , com $t = 1, \dots, n$, o modelo de decomposição aditiva expressa cada valor da série como a soma de três componentes:

$$Z_t = T_t + S_t + a_t, \quad (2.4)$$

em que T_t representa a tendência, S_t a componente sazonal e a_t o termo aleatório, assumido com média nula e variância constante σ_a^2 [16].

A tendência traduz a direção predominante da série no longo prazo, podendo indicar crescimento ou queda. A sazonalidade, por sua vez, corresponde a padrões que se repetem em intervalos regulares, como picos que ocorrem nos mesmos meses a cada ano. Esses padrões podem ser determinísticos, quando se mantêm relativamente estáveis ao longo do tempo, ou estocásticos, quando apresentam variações entre os ciclos. É importante distinguir a sazonalidade dos ciclos: enquanto a primeira ocorre em períodos fixos, geralmente inferiores a um ano, os ciclos correspondem a oscilações de duração mais longa e menos regular [16]. No caso da dengue, a sazonalidade assume papel central, dada a concentração recorrente de casos nos meses mais quentes e chuvosos do ano.

TESTE DE COX-STUART PARA TENDÊNCIA

A verificação da existência de tendência em uma série pode ser realizada por meio do teste de Cox-Stuart [8], um procedimento não paramétrico baseado na comparação entre observações separadas no tempo. A ideia consiste em dividir a série em duas metades e comparar cada observação da primeira metade com a correspondente na segunda.

Considerando uma série com n observações, formam-se pares do tipo (Z_i, Z_{i+c}) , em que $c = n/2$ quando n é par. Para cada par, analisa-se o sinal da diferença $D_i = Z_{i+c} - Z_i$, contabilizando-se o número de diferenças positivas, S^+ , e negativas, S^- . Na ausência de tendência, espera-se um equilíbrio entre esses sinais, de modo que a contagem segue uma distribuição binomial com probabilidade 0,5:

$$S \sim \text{Binomial}(n'; 0,5), \quad (2.5)$$

em que $n' = S^+ + S^-$ representa o número de pares com diferença não nula. Embora, para amostras grandes, a distribuição binomial possa ser aproximada pela normal, neste trabalho o p -valor foi calculado diretamente a partir da distribuição binomial exata, o que garante validade independentemente do tamanho da amostra. A hipótese nula do teste corresponde à ausência de tendência. Sua rejeição, ao nível de significância de 5% adotado neste estudo, sugere a presença de tendência na série e, conseqüentemente, a necessidade de aplicar diferenciação para alcançar a estacionariedade.

TESTE DE FISHER PARA SAZONALIDADE

A detecção de componentes sazonais determinísticas pode ser realizada por meio do teste de Fisher, baseado na análise do periodograma [17]. O periodograma decompõe a variabilidade da série em contribuições associadas a diferentes frequências, representando os dados como uma combinação de funções senoidais. Essa representação permite identificar a presença de ciclos dominantes na série.

Segundo Priestley [17], o periodograma na frequência f_i é dado por:

$$I_p(f_i) = \frac{2}{n} \left[\left(\sum_{t=1}^n a_t \cos \frac{2\pi i}{n} t \right)^2 + \left(\sum_{t=1}^n a_t \sin \frac{2\pi i}{n} t \right)^2 \right], \quad (2.6)$$

com $0 < f_i < \frac{1}{2}$ e $t = 1, 2, \dots, n$, sendo $I_p(f_i)$ a intensidade associada à frequência f_i . A estatística do teste compara a maior dessas intensidades com a soma total:

$$g = \frac{\max I_p}{\sum_{p=1}^{n/2} I_p}. \quad (2.7)$$

O valor crítico correspondente, z_α , para um nível de significância α , é dado por:

$$z_\alpha = 1 - \left(\frac{\alpha}{2} \right)^{\left(\frac{n}{2} - 1 \right)^{-1}}. \quad (2.8)$$

Sob a hipótese nula de ausência de sazonalidade, espera-se que nenhuma frequência se destaque de forma significativa. Quando g excede z_α , rejeita-se essa hipótese, indicando a presença de periodicidade relevante na série. Esse resultado sustenta a inclusão de uma componente sazonal no modelo, seja na formulação SARIMA ou em sua extensão SARIMAX.

2.2.3 FUNÇÕES DE AUTOCORRELAÇÃO E AUTOCORRELAÇÃO PARCIAL

Um dos aspectos centrais da análise de séries temporais é a caracterização da dependência entre observações separadas por diferentes intervalos de tempo. Essa dependência é descrita por duas funções fundamentais: a função de autocorrelação (FAC) e a função de autocorrelação parcial (FACP) [6].

A função de autocorrelação mede o grau de correlação linear entre observações da série separadas por uma defasagem k , isto é, entre Z_t e Z_{t-k} . Para uma série observada $z_1, z_2, \dots, z_N$, a autocorrelação amostral na defasagem k é estimada por:

$$r_k = \frac{c_k}{c_0}, \quad \text{em que} \quad c_k = \frac{1}{N} \sum_{t=1}^{N-k} (z_t - \bar{z})(z_{t+k} - \bar{z}), \quad (2.9)$$

em que c_k representa a autocovariância amostral na defasagem k , c_0 corresponde à variância amostral da série e $\bar{z}$ é a média das observações. Os valores de r_k , calculados para diferentes defasagens, permitem descrever a estrutura de dependência temporal da série e auxiliam na identificação de padrões como tendência e sazonalidade.

A função de autocorrelação parcial, por sua vez, mede a correlação entre Z_t e Z_{t-k} após a remoção do efeito das observações intermediárias. Enquanto a FAC capta a correlação total associada a cada defasagem, a FACP isola a contribuição direta de cada uma delas, descontando as correlações já explicadas por defasagens menores. Denotando por ϕ_{kj} o j -ésimo coeficiente de uma representação autorregressiva de ordem k , em que ϕ_{kk} é o último coeficiente, esses coeficientes satisfazem o sistema de equações de Yule-Walker:

$$\rho_j = \phi_{k1}\rho_{j-1} + \phi_{k2}\rho_{j-2} + \cdots + \phi_{kk}\rho_{j-k}, \quad j = 1, 2, \dots, k, \quad (2.10)$$

cuja resolução sucessiva para $k = 1, 2, 3, \dots$ fornece a sequência ϕ_{kk} , denominada função de autocorrelação parcial. Cada ϕ_{kk} corresponde à correlação entre Z_t e Z_{t-k} ajustada pelas variáveis intermediárias $Z_{t-1}, Z_{t-2}, \dots, Z_{t-k+1}$.

Na metodologia de Box-Jenkins, a análise conjunta da FAC e da FACP constitui a principal ferramenta para a identificação das ordens do modelo [6]. O comportamento de truncamento ou de decaimento gradual dessas funções orienta a escolha das ordens dos componentes autorregressivo (p) e de médias móveis (q), bem como de seus componentes sazonais correspondentes.

2.3 MODELOS ARIMA, SARIMA E SARIMAX

2.3.1 MODELO AUTORREGRESSIVO INTEGRADO DE MÉDIAS MÓVEIS (ARIMA)

Entre os métodos consagrados para a modelagem de séries temporais destaca-se a abordagem proposta por Box e Jenkins [5], que recorre aos modelos autorregressivos integrados de médias móveis, $ARIMA(p, d, q)$, apropriados para séries não estacionárias, desde que possam ser tornadas estacionárias por meio de diferenciação [16]. Segundo Fava [9], esses modelos combinam três componentes: a parte autorregressiva (AR), a etapa de integração (I) e a parte de médias móveis (MA).

No componente autorregressivo de ordem p , denotado $AR(p)$, supõe-se que o valor corrente da série dependa linearmente de seus próprios valores passados, conforme:

$$Z_t = \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \cdots + \phi_p Z_{t-p} + a_t, \quad (2.11)$$

cujo operador associado é

$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p, \quad (2.12)$$

sendo B o operador de retardo, que satisfaz $B^k Z_t = Z_{t-k}$, e a_t o termo de erro, suposto ruído branco, ou seja, uma sequência de variáveis aleatórias não correlacionadas, de média zero e variância constante.

O componente de médias móveis de ordem q , indicado por $MA(q)$, relaciona o valor corrente da série a uma combinação de erros aleatórios passados:

$$Z_t = \mu + a_t - \theta_1 a_{t-1} - \cdots - \theta_q a_{t-q}. \quad (2.13)$$

Tomando $\bar{Z}_t = Z_t - \mu$, essa expressão pode ser reescrita em termos do operador de médias móveis:

$$\bar{Z}_t = (1 - \theta_1 B - \cdots - \theta_q B^q) a_t = \theta(B) a_t, \quad (2.14)$$

com

$$\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q. \quad (2.15)$$

A reunião das partes autorregressiva e de médias móveis produz os modelos ARMA(p, q), aplicáveis a séries estacionárias:

$$\bar{Z}_t = \phi_1 \bar{Z}_{t-1} + \dots + \phi_p \bar{Z}_{t-p} + a_t - \theta_1 a_{t-1} - \dots - \theta_q a_{t-q}. \quad (2.16)$$

Quando a série de interesse não é estacionária, aplica-se a diferenciação até que essa condição seja satisfeita, e o modelo resultante passa a ser denominado ARIMA. Nesse caso, a série diferenciada é modelada como um processo ARMA, e a forma compacta do modelo ARIMA(p, d, q) é dada por:

$$\phi(B) \Delta^d Z_t = \theta(B) a_t, \quad (2.17)$$

em que d indica quantas diferenciações foram necessárias, e p e q são as ordens dos respectivos operadores.

2.3.2 MODELO SARIMA

Muitas séries temporais apresentam padrões que se repetem periodicamente ao longo do tempo. Quando essa repetição ocorre de forma sistemática, torna-se necessário incorporar componentes que representem a sazonalidade da série. Com esse objetivo, Box e Jenkins [5] estenderam o modelo ARIMA ao incluir termos sazonais, dando origem ao modelo SARIMA(p, d, q)(P, D, Q) $_s$, que pode ser escrito como:

$$\phi_p(B) \Phi_P(B^s) \Delta^d \Delta_s^D Z_t = \theta_q(B) \Theta_Q(B^s) a_t, \quad (2.18)$$

em que:

- $\Phi_P(B^s) = 1 - \Phi_1 B^s - \dots - \Phi_P B^{Ps}$ corresponde ao operador autorregressivo sazonal de ordem P ;
- $\Theta_Q(B^s) = 1 - \Theta_1 B^s - \dots - \Theta_Q B^{Qs}$ corresponde ao operador de médias móveis sazonal de ordem Q ;
- Δ^d representa o operador de diferenciação simples de ordem d ;
- $\Delta_s^D = (1 - B^s)^D$ representa o operador de diferenciação sazonal de ordem D , em que s indica o período sazonal, por exemplo, $s = 12$ para dados mensais com ciclo anual.

2.3.3 MODELO SARIMAX

O modelo SARIMAX (*Seasonal Autoregressive Integrated Moving Average with eXogenous variables*) amplia o SARIMA ao permitir a inclusão de variáveis externas, ou covariáveis, na

modelagem da série. Isso possibilita quantificar a influência de fatores externos à dinâmica interna da série sobre a variável de interesse [10]. Em termos gerais, o modelo combina uma estrutura de regressão para as covariáveis exógenas com um componente SARIMA, responsável por descrever a dependência temporal remanescente. Sua forma geral é:

$$\phi_p(B)\Phi_P(B^s)\Delta^d\Delta_s^D Z_t = \alpha + X_t\beta + \theta_q(B)\Theta_Q(B^s)a_t, \quad (2.19)$$

em que X_t é a matriz $n \times k$ que reúne as variáveis exógenas, β é o vetor $k \times 1$ de coeficientes associados a essas variáveis e α é o termo constante. Neste trabalho, as covariáveis consideradas são as variáveis climáticas, temperatura e precipitação, incorporadas ao modelo com defasagem temporal.

2.4 MODELOS SARIMAX APLICADOS À EPIDEMIOLOGIA DA DENGUE

A aplicação de modelos SARIMAX na previsão de casos de dengue com a inclusão de variáveis climáticas exógenas tem sido documentada em diferentes estudos brasileiros. Em geral, esses trabalhos apontam a temperatura como uma covariável relevante, embora a magnitude e a significância dos efeitos variem conforme a região e o período analisados.

Morais et al. [15], ao analisarem a série de casos de dengue no município de Garanhuns-PE entre 2010 e 2019, ajustaram inicialmente um modelo SARIMA(0, 1, 3)(0, 1, 1)₁₂, posteriormente estendido para um modelo SARIMAX com a inclusão das variáveis temperatura e umidade. Os autores observam que a incorporação dessas covariáveis resultou em aumento do critério de informação de Akaike (AIC), de 1114,92 para 1117,97, em comparação ao modelo SARIMA sem variáveis exógenas. Ainda assim, o modelo SARIMAX apresentou ajuste considerado adequado e manteve parcimônia, com a temperatura exercendo efeito de maior magnitude sobre o número de casos em relação à umidade.

Um aspecto metodológico relevante desse estudo é que as variáveis climáticas foram incorporadas no mesmo período de ocorrência dos casos, sem a investigação de possíveis defasagens temporais entre as condições climáticas e a resposta epidemiológica. Esse aspecto merece atenção porque, como discutido anteriormente, a relação entre clima e dengue tende a ocorrer com algum atraso. Assim, embora os resultados de Moraes et al. [15] evidenciem o potencial dos modelos SARIMAX para incorporar informações climáticas na previsão de casos, também reforçam a necessidade de explorar de forma mais direta o papel das defasagens temporais nessa relação.

2.5 METODOLOGIA DE BOX-JENKINS

O ajuste de modelos das classes ARIMA e SARIMA segue a metodologia proposta por Box e Jenkins [5], organizada em um ciclo iterativo que se repete até a obtenção de um modelo

considerado adequado [16]. Esse ciclo envolve quatro etapas principais. A especificação, na qual se define um conjunto inicial de modelos candidatos. A identificação, em que as funções de autocorrelação (FAC) e de autocorrelação parcial (FACP) auxiliam na escolha das ordens do modelo. A estimação dos parâmetros do modelo selecionado. E, por fim, a verificação, etapa em que a qualidade do ajuste é avaliada a partir da análise dos resíduos. Caso o modelo não apresente desempenho satisfatório, o processo é retomado com uma nova especificação, caso contrário, segue-se para a etapa de previsão.

2.6 FUNÇÃO DE CORRELAÇÃO CRUZADA (CCF)

A função de correlação cruzada (*Cross-Correlation Function* – CCF) mede a associação linear entre duas séries temporais quando uma delas é deslocada no tempo em relação à outra, para diferentes valores de defasagem [6]. Por meio dessa função, é possível avaliar se os movimentos de uma série antecedem, acompanham ou sucedem os da outra. Em estudos epidemiológicos, essa ferramenta é particularmente útil para estimar o intervalo de tempo entre a variação de um fator climático e a correspondente alteração no número de casos de uma doença. Essa informação orienta a escolha da defasagem com que as variáveis exógenas são incorporadas ao modelo SARIMAX.

2.7 CRITÉRIOS DE SELEÇÃO DO MODELO

Ao longo da modelagem, diferentes modelos candidatos são ajustados, o que exige a adoção de um critério objetivo para compará-los. A seleção baseia-se nos critérios de informação apresentados a seguir, que buscam equilibrar a qualidade do ajuste com a complexidade do modelo.

2.7.1 CRITÉRIO DE INFORMAÇÃO DE AKAIKE (AIC)

O critério proposto por Akaike [2] é definido por:

$$AIC = -2 \ln L(\hat{\theta}) + 2k, \quad (2.20)$$

em que $L(\hat{\theta})$ corresponde ao valor máximo da função de verossimilhança e k ao número de parâmetros estimados. O primeiro termo favorece modelos com melhor ajuste aos dados, enquanto o segundo penaliza o uso excessivo de parâmetros. Entre os modelos comparados, prefere-se aquele que apresenta menor valor de AIC.

2.7.2 CRITÉRIO DE INFORMAÇÃO BAYESIANO (BIC)

O critério de Schwarz [19] é expresso por:

$$BIC = -2 \ln L(\hat{\theta}) + k \ln(n), \quad (2.21)$$

em que n representa o número de observações utilizadas no ajuste. Sua lógica é semelhante à do AIC, mas a penalização da complexidade é mais severa, pois cresce com o logaritmo do tamanho da amostra. Assim como no AIC, o modelo preferido é aquele com menor valor de BIC.

2.8 DIAGNÓSTICO DOS RESÍDUOS

A validação de um modelo envolve a análise de seus resíduos, que, em um ajuste adequado, devem se comportar como ruído branco, isto é, não apresentar autocorrelação, ter média próxima de zero e variância constante.

Uma primeira avaliação pode ser feita por meio da inspeção gráfica da função de autocorrelação (FAC) dos resíduos, procedimento clássico da etapa de verificação da metodologia de Box-Jenkins [6]. Em um modelo bem especificado, espera-se que as autocorrelações residuais permaneçam, em sua maioria, dentro dos limites de confiança, sem evidenciar padrões sistemáticos. A presença de correlações significativas em determinadas defasagens, especialmente naquelas associadas ao período sazonal, sugere que parte da dependência temporal não foi capturada pelo modelo, indicando a necessidade de reespecificação. Neste trabalho, essa análise foi complementada pela inspeção da função de autocorrelação parcial (FACP) dos resíduos, do gráfico dos resíduos ao longo do tempo e de seu histograma. Utilizou-se também o periodograma acumulado [6], que permite verificar a presença de periodicidade remanescente nos resíduos.

Além da análise gráfica, a independência dos resíduos pode ser avaliada por meio de testes formais. O teste de Box-Pierce [7] tem como hipótese nula que os resíduos constituem ruído branco. A não rejeição dessa hipótese sugere ausência de autocorrelação relevante, o que sustenta a adequação do modelo.

De forma complementar, emprega-se o teste de Ljung-Box [13], uma versão aprimorada do teste de Box-Pierce, com melhor desempenho em amostras finitas. Diferentemente de uma avaliação restrita a poucas defasagens, o teste de Ljung-Box permite analisar a autocorrelação dos resíduos de forma conjunta até uma defasagem previamente definida. Sua estatística é dada por:

$$Q = n(n+2) \sum_{k=1}^m \frac{\hat{\rho}_k^2}{n-k}, \quad (2.22)$$

em que n é o número de observações, $\hat{\rho}_k$ é a autocorrelação amostral dos resíduos na defasagem k e m é o número de defasagens consideradas. Sob a hipótese nula de que os resíduos são não correlacionados, a estatística Q segue aproximadamente uma distribuição qui-quadrado com $m - p$ graus de liberdade, em que p corresponde ao número de parâmetros estimados no modelo. Esse teste é particularmente útil na análise de séries sazonais, pois permite avaliar a presença de autocorrelação residual em defasagens mais longas, incluindo aquelas associadas ao período sazonal.

2.9 MÉTRICAS DE AVALIAÇÃO DO MODELO

A qualidade preditiva dos modelos foi avaliada por meio de duas métricas de erro. A primeira, o Erro Percentual Absoluto Médio (*Mean Absolute Percentage Error* – MAPE), expressa o erro de previsão em termos relativos [10]:

$$MAPE = \frac{1}{n} \sum_{t=1}^n \left| \frac{Y_t - \hat{Y}_t}{Y_t} \right| \times 100, \quad (2.23)$$

em que Y_t e $\hat{Y}_t$ representam, respectivamente, o valor observado e o valor previsto no instante t , e n o número de observações consideradas. Por ser adimensional, o MAPE permite comparar a precisão de previsões entre séries com magnitudes distintas. Sua principal limitação ocorre na presença de valores nulos, caso em que o cálculo se torna inviável [10].

A segunda métrica é a Raiz do Erro Quadrático Médio (*Root Mean Square Error* – RMSE) [10]. Parte-se do Erro Quadrático Médio (EQM), que atribui maior peso a erros de maior magnitude:

$$EQM = \frac{1}{n} \sum_{t=1}^n a_t^2, \quad (2.24)$$

em que $a_t = Y_t - \hat{Y}_t$ representa o erro de previsão no instante t . O RMSE é definido como a raiz quadrada dessa quantidade:

$$RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n a_t^2}. \quad (2.25)$$

Por penalizar de forma mais acentuada os desvios maiores, o RMSE auxilia na identificação de modelos com erros de previsão elevados. Para ambas as métricas, valores menores indicam melhor desempenho preditivo [10].

3. METODOLOGIA

3.1 CARACTERIZAÇÃO DA PESQUISA

A presente pesquisa caracteriza-se como um estudo descritivo e exploratório, com abordagem quantitativa, voltado à análise estatística de séries temporais. O objetivo é compreender o comportamento histórico da dengue em Uberlândia-MG e avaliar como variáveis climáticas, consideradas com defasagem temporal, influenciam a incidência da doença.

3.2 FONTE E PERÍODO DOS DADOS

3.2.1 DADOS DE CASOS DE DENGUE

Os dados de casos de dengue em Uberlândia-MG foram obtidos a partir do Sistema de Informação de Agravos de Notificação (SINAN), disponibilizado pelo Departamento de Informática do Sistema Único de Saúde (DATASUS), referentes ao período de janeiro de 2014 a abril de 2026. Em razão do elevado volume de registros, a leitura e o processamento inicial dos arquivos anuais foram realizados com o pacote `data.table` do software R, que permite a importação seletiva de colunas e a filtragem eficiente de grandes bases de dados.

Do conjunto de variáveis disponibilizadas pelo SINAN, foram selecionadas as colunas referentes à data de início dos sintomas (`DT_SIN_PRI`), ao município de residência (`ID_MN_RESI`) e à classificação final do caso (`CLASSI_FIN`). Os registros foram filtrados de modo a reter apenas os casos cujo município de residência correspondia a Uberlândia-MG e cuja classificação final indicava confirmação de dengue, considerando as diferentes nomenclaturas adotadas ao longo dos anos da base (dengue, dengue com sinais de alarme, dengue grave, dengue clássica e dengue com complicações).

Após a filtragem dos arquivos anuais, os registros foram consolidados em uma única base. Em seguida, os casos foram agregados por mês de início dos sintomas, gerando a série temporal mensal de casos confirmados de dengue, com 148 observações mensais.

3.2.2 DADOS CLIMÁTICOS

Os dados climáticos referentes à temperatura média mensal e à precipitação total mensal foram obtidos da plataforma NASA POWER (*Prediction Of Worldwide Energy Resources*),

disponibilizada pela *National Aeronautics and Space Administration* (NASA), para o ponto geográfico correspondente a Uberlândia-MG (latitude $-18,92^\circ$; longitude $-48,26^\circ$). A temperatura foi obtida a partir da variável T2M (temperatura média do ar a 2 metros de altura, em $^\circ\text{C}$) e a precipitação a partir da variável PRECTOTCORR, originalmente em mm/dia, posteriormente convertida para mm/mês considerando o número de dias de cada mês.

A escolha dessa fonte se deve à ausência de valores faltantes em toda a série analisada, em contraste com a base de estações meteorológicas convencionais, que apresentou lacunas pontuais no período. Quanto à temperatura, optou-se pela utilização da temperatura média mensal por ser a métrica fornecida de forma contínua e padronizada pela plataforma, compatível com a escala de agregação mensal adotada neste estudo.

3.3 DIVISÃO ENTRE TREINO E VALIDAÇÃO

Para a construção e a validação dos modelos, a série foi dividida em dois períodos: um período de treino, utilizado para a identificação, estimação e diagnóstico dos modelos, correspondente a janeiro de 2014 a outubro de 2025 (142 observações), e um período de validação, utilizado para avaliar a capacidade preditiva dos modelos fora da amostra, correspondente a novembro de 2025 a abril de 2026 (6 observações).

3.4 PRÉ-PROCESSAMENTO DOS DADOS

3.4.1 TRANSFORMAÇÃO DA SÉRIE DE CASOS

Em razão da heterocedasticidade identificada na série original de casos, verificada por meio da análise gráfica da relação entre a média e o desvio padrão em subperíodos da série, foi aplicada a transformação logarítmica à série de casos de dengue. O objetivo foi estabilizar a variância e tornar a série mais adequada aos pressupostos dos modelos da classe ARIMA/SARIMA.

3.4.2 TRANSFORMAÇÃO DAS VARIÁVEIS CLIMÁTICAS

A variável de precipitação apresentou distribuição assimétrica, com a maior parte das observações concentrada em valores baixos e poucos meses com valores altos. Por esse motivo, foi aplicada a transformação logarítmica $\ln(\text{precipitação} + 1)$, com a finalidade de reduzir essa assimetria. A adição de uma unidade permite a aplicação da transformação mesmo em meses com precipitação igual a zero. A variável de temperatura, por apresentar distribuição aproximadamente simétrica, não necessitou de transformação.

3.5 TESTES DE HIPÓTESE

Após a transformação logarítmica, foram aplicados os seguintes testes estatísticos à série de casos:

- **Teste de Cox-Stuart:** utilizado para verificar a presença de tendência determinística na série, com o objetivo de determinar a ordem de diferenciação simples (d) necessária;
- **Teste de Fisher:** aplicado à série já diferenciada, para verificar a presença de sazonalidade significativa e corroborar a utilização de um componente sazonal anual ($s = 12$), em coerência com a natureza mensal dos dados.

3.6 IDENTIFICAÇÃO, ESTIMAÇÃO E DIAGNÓSTICO DOS MODELOS SARIMA

A identificação das ordens dos modelos candidatos foi realizada com base na análise gráfica das funções de autocorrelação (FAC) e autocorrelação parcial (FACP) da série diferenciada, complementada pelo uso da função `auto.arima` do pacote `forecast` do software R como referência inicial.

Os modelos candidatos foram estimados pelo método de máxima verossimilhança e comparados por meio dos critérios de informação de Akaike (AIC) e Bayesiano (BIC), priorizando-se modelos com menores valores nesses critérios e com todos os coeficientes estatisticamente significativos ao nível de 5%.

O diagnóstico dos modelos foi conduzido por meio da análise gráfica dos resíduos, do periodograma acumulado e dos testes de Box-Pierce e Ljung-Box, com a finalidade de verificar se os resíduos se comportam de forma compatível com ruído branco.

3.7 INCORPORAÇÃO DE VARIÁVEIS CLIMÁTICAS EXÓGENAS

3.7.1 IDENTIFICAÇÃO DAS DEFASAGENS TEMPORAIS

A identificação inicial das defasagens temporais (*lags*) entre as variáveis climáticas e os casos de dengue foi realizada por meio da função de correlação cruzada (CCF), aplicada entre cada variável climática e a série logarítmica de casos. Para a precipitação, utilizou-se a série transformada em escala logarítmica, por meio da expressão $\log(\text{precipitação} + 1)$, visando reduzir a assimetria. A CCF foi utilizada como ferramenta exploratória para indicar defasagens candidatas, posteriormente avaliadas nos modelos SARIMAX por meio dos critérios de informação e da significância estatística dos coeficientes.

3.7.2 SELEÇÃO DO LAG E DA VARIÁVEL

Com base nos lags indicados pela CCF, foi realizada uma busca sistemática (*grid search*), testando-se combinações de defasagens de 1 a 8 meses para temperatura e precipitação, de forma isolada e conjunta, totalizando 24 modelos candidatos. A definição do limite máximo

de oito meses fundamentou-se em dois aspectos complementares. Do ponto de vista biológico, esse intervalo é superior ao ciclo completo entre o estímulo climático e o início dos sintomas dos casos, que compreende o desenvolvimento larval, a maturação do vetor, a transmissão e a incubação viral, estimado em torno de quatro a cinco meses, garantindo margem suficiente para capturá-lo integralmente. Do ponto de vista empírico, a função de correlação cruzada apontou correlações relevantes nas defasagens de quatro e cinco meses, de modo que o limite de oito meses abrange uma faixa em torno desses valores, sem alcançar defasagens mais longas (próximas a doze meses), nas quais as correlações observadas refletem predominantemente o padrão sazonal anual invertido, e não uma relação de defasagem causal direta.

A seleção do modelo final considerou o menor valor de BIC entre as configurações testadas, associado à significância estatística dos coeficientes das variáveis exógenas ao nível de 5%. O critério BIC foi priorizado por impor penalização mais severa à complexidade do modelo, favorecendo a escolha de estruturas mais parcimoniosas. Cabe destacar que, na etapa de busca sistemática, todos os modelos candidatos foram ajustados sobre uma mesma janela de observações, correspondente à maior defasagem testada (oito meses), de modo a assegurar a comparabilidade dos critérios de informação entre as configurações. Uma vez selecionada a defasagem, o modelo final foi reajustado utilizando todas as observações válidas para essa defasagem específica. Para a comparação entre o modelo com e sem a variável exógena, ambos foram ajustados sobre essa mesma amostra efetiva, garantindo a validade da comparação dos critérios de informação.

3.7.3 PADRONIZAÇÃO DAS VARIÁVEIS EXÓGENAS

As variáveis climáticas exógenas selecionadas foram padronizadas (média zero e desvio padrão unitário) antes de sua inclusão no modelo, de forma a viabilizar a comparação entre os coeficientes estimados e garantir maior estabilidade numérica no processo de otimização.

3.8 VALIDAÇÃO DOS MODELOS

Em virtude do comportamento atípico observado na série durante o ano de 2025, caracterizado por um surto de magnitude muito superior à média histórica nos primeiros meses do ano, seguido de declínio abrupto, a validação dos modelos não foi realizada por meio de previsão direta de múltiplos passos à frente. Essa abordagem ampliaria o efeito da tendência de alta capturada durante o treinamento e poderia comprometer a interpretação dos resultados.

Optou-se, portanto, pela metodologia de previsão *rolling* (*one-step-ahead forecast*), na qual o modelo é reestimado a cada novo mês do período de validação, incorporando a observação real do mês imediatamente anterior à base de treinamento e gerando a previsão de apenas um passo à frente. Durante esse procedimento, manteve-se fixa a estrutura do modelo previamente selecionado, incluindo as ordens do modelo SARIMA e a defasagem da variável exógena, sendo reestimados apenas os parâmetros a cada atualização da base. Essa abordagem é mais adequada para a validação de modelos em séries com alta variabilidade epidêmica, pois reduz o acúmulo

de erro de previsão ao longo do horizonte de validação.

O desempenho preditivo dos modelos foi avaliado por meio do Erro Percentual Absoluto Médio (MAPE) e da Raiz do Erro Quadrático Médio (RMSE), calculados sobre o período de validação, bem como pela verificação da proporção de observações reais contidas dentro do intervalo de confiança de 95% gerado pelas previsões.

3.9 SOFTWARES UTILIZADOS

Todas as análises foram realizadas no software R [18] (versão 4.4), um ambiente estatístico de código aberto e gratuito. Foram utilizados principalmente os pacotes `forecast` [11], `lmtest` [26], `randtests` [14], `GeneCycle` [1] e `tseries` [22], além dos pacotes `dplyr` [24] e `zoo` [25] para manipulação dos dados.

4. RESULTADOS

4.1 ANÁLISE EXPLORATÓRIA

A série de casos de dengue em Uberlândia-MG, no período de janeiro de 2014 a outubro de 2025 (142 observações), apresentou picos epidêmicos sazonais com grande variação de amplitude, evidenciando o comportamento característico da doença na região. Entre esses eventos, destaca-se o surto de 2025, de magnitude muito superior à observada nos demais anos da série. A Figura 4.1 apresenta a série mensal de casos.

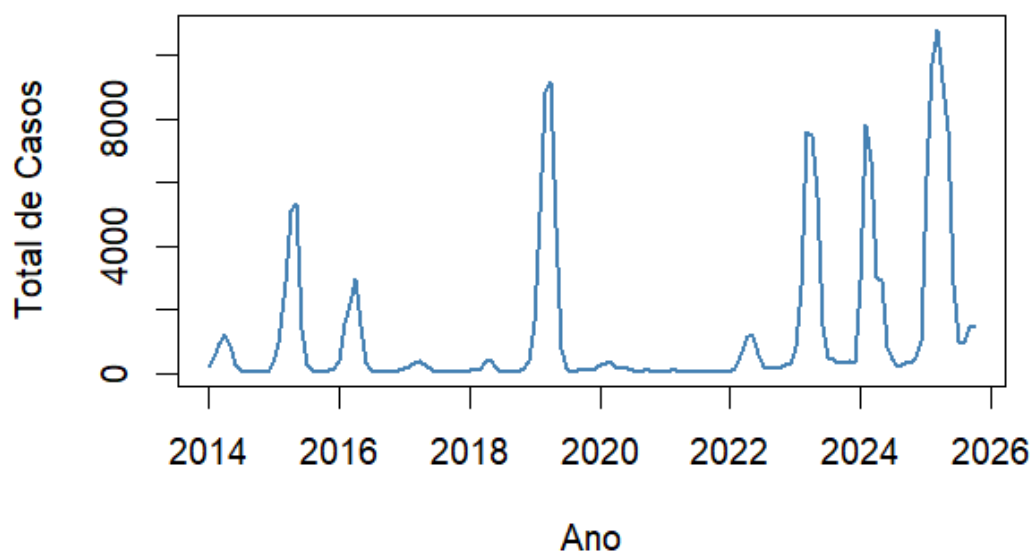


Figura 4.1: Série mensal de casos de dengue em Uberlândia-MG (2014–2025).

As funções de autocorrelação (FAC) e de autocorrelação parcial (FACP) da série original, apresentadas na Figura 4.2, evidenciam características típicas de uma série não estacionária. Na FAC, observa-se um decaimento gradual das autocorrelações ao longo das defasagens, indicando forte persistência temporal. Além disso, verificam-se picos significativos em múltiplos de 12 meses, o que sugere a presença de um padrão sazonal anual. A FACP, por sua vez, apresenta um pico expressivo e dominante na primeira defasagem, com as demais defasagens majoritariamente dentro dos limites de significância, à exceção de algumas defasagens na região sazonal. Esse padrão indica que grande parte da dependência temporal de curto prazo é explicada pela

observação imediatamente anterior. Em conjunto, esses resultados sustentam a necessidade de aplicar diferenciações regular e sazonal para tornar a série estacionária antes da modelagem.

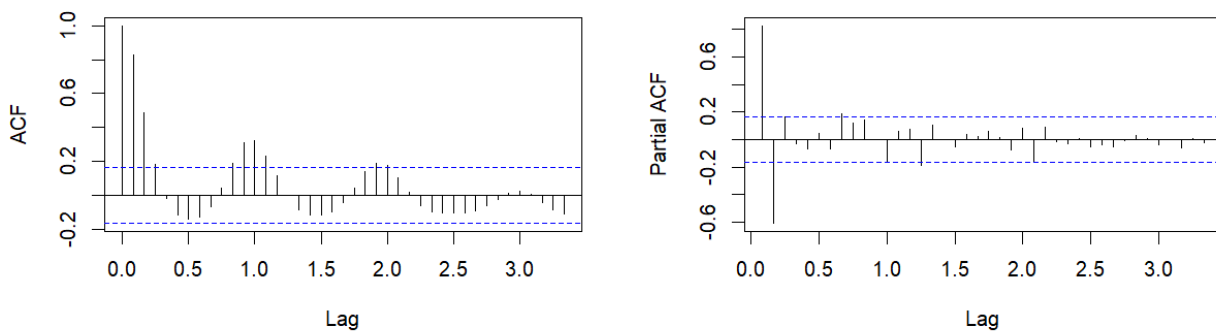


Figura 4.2: Funções de autocorrelação (FAC) e de autocorrelação parcial (FACP) da série original de casos.

A análise da relação entre a média e o desvio padrão dos casos em subperíodos da série indicou um comportamento multiplicativo, compatível com a presença de heterocedasticidade na série original, conforme ilustrado na Figura 4.3. Observa-se que, à medida que a média dos casos aumenta, o desvio padrão também tende a crescer, evidenciando dependência entre o nível e a variabilidade da série.

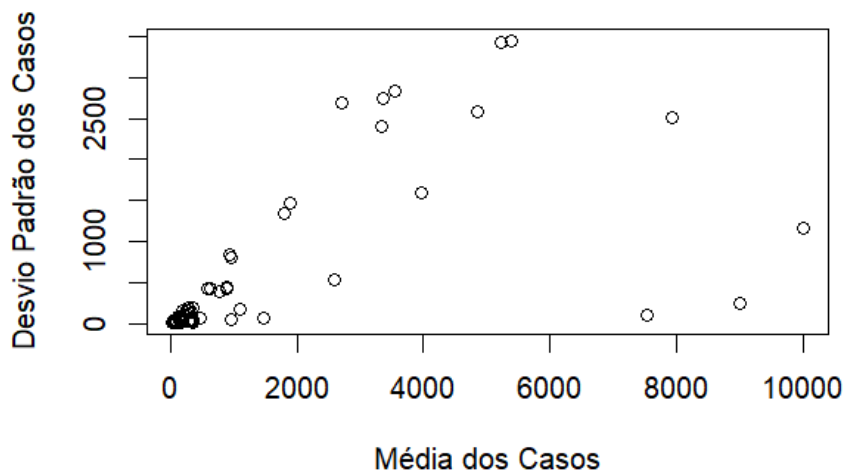


Figura 4.3: Relação entre média e desvio padrão em subperíodos da série original.

A adequação da transformação logarítmica foi corroborada pelo procedimento de Box-Cox, que indicou parâmetro ótimo $\lambda = -0,176$, valor suficientemente próximo de zero para sustentar a utilização da transformação em logaritmo natural. A série transformada é apresentada na Figura 4.4.

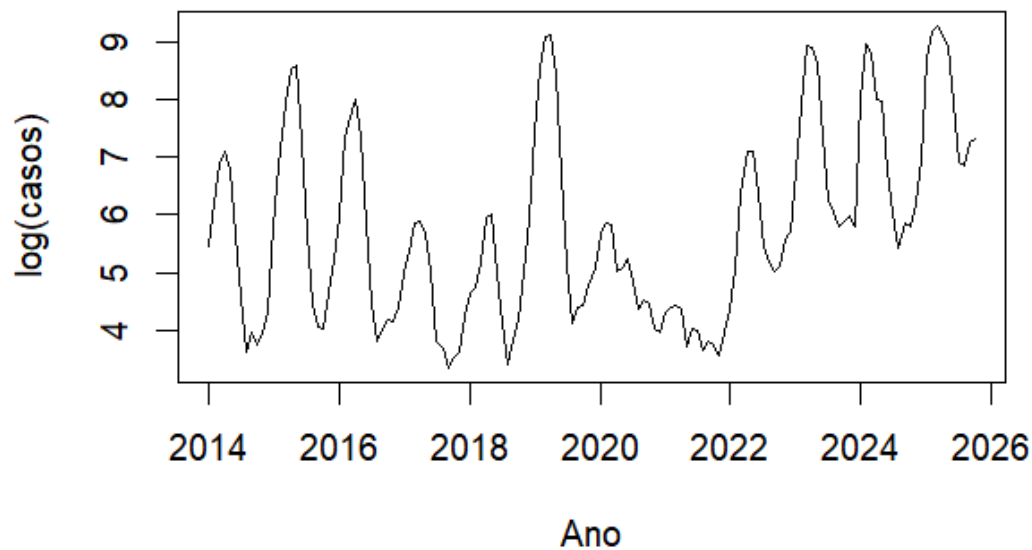


Figura 4.4: Série de casos de dengue após transformação logarítmica.

Após a transformação logarítmica, a relação entre a média e o desvio padrão tornou-se menos estruturada, com distribuição mais homogênea dos pontos, indicando estabilização da variância e redução da heterocedasticidade presentes na série original, conforme apresentado na Figura 4.5. Esse resultado torna a série mais adequada aos pressupostos dos modelos da classe SARIMA.

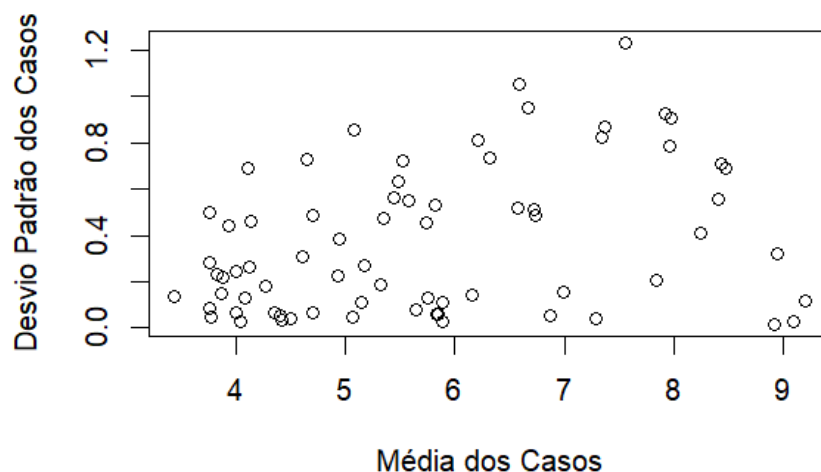


Figura 4.5: Relação entre média e desvio padrão em subperíodos da série logarítmica, indicando estabilização da variância.

As funções de autocorrelação (FAC) e de autocorrelação parcial (FACP) da série logarítmica são apresentadas na Figura 4.6. A FAC ainda exibe um decaimento gradual das autocorrelações, sugerindo que a série permanece não estacionária em média. Além disso, observam-se picos significativos em defasagens múltiplas de 12 meses, evidenciando a persistência do padrão

sazonal anual. Na FACP, destaca-se um coeficiente expressivo na primeira defasagem, seguido por poucos coeficientes significativos nas demais defasagens. Esses resultados indicam que, embora a transformação logarítmica tenha sido eficaz na estabilização da variância, ainda se fazem necessárias diferenciações regular e sazonal para remover a não estacionariedade e adequar a série à modelagem SARIMA.

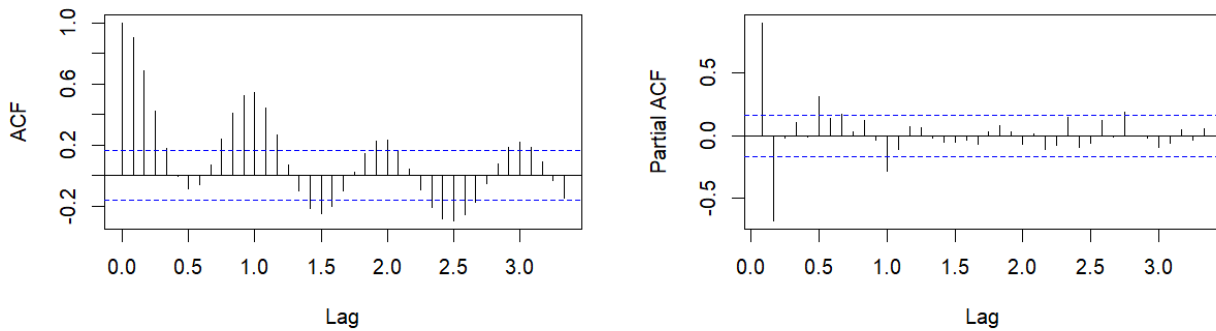


Figura 4.6: FAC e FACP da série logarítmica de casos.

Após a aplicação de uma diferenciação simples ($d = 1$), a série logarítmica passou a oscilar em torno de um nível aproximadamente constante, conforme apresentado na Figura 4.7. Esse comportamento indica que a diferenciação regular foi eficaz na remoção da tendência presente na série original, contribuindo para a obtenção de estacionariedade em média.

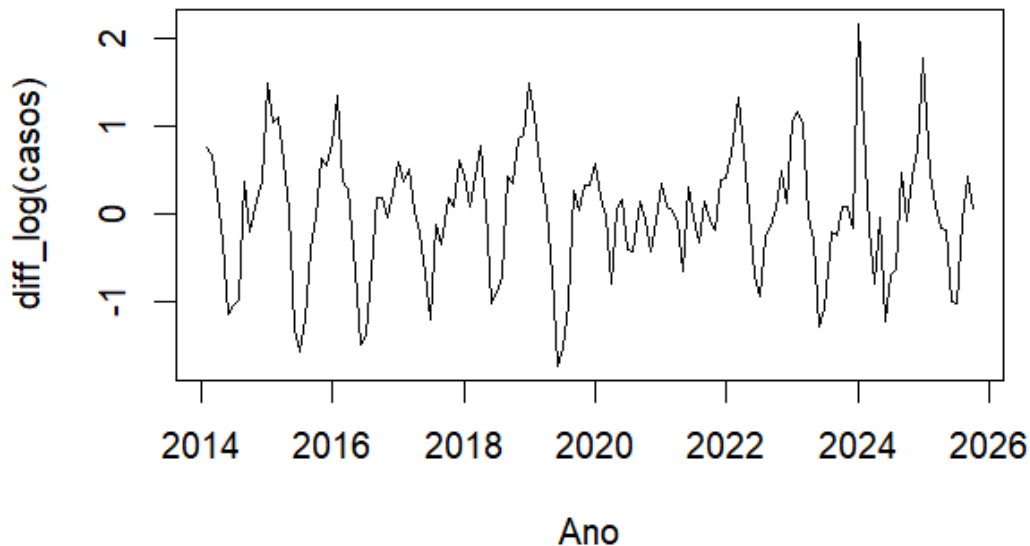


Figura 4.7: Série logarítmica após uma diferenciação simples ($d = 1$).

As funções de autocorrelação (FAC) e de autocorrelação parcial (FACP) da série diferenciada, apresentadas na Figura 4.8, serviram de base para a identificação das ordens dos modelos candidatos. Observa-se que a FAC deixou de apresentar o decaimento lento característico de

séries não estacionárias, sugerindo que a diferenciação regular reduziu substancialmente a dependência associada à tendência. Entretanto, permanecem picos significativos em defasagens múltiplas de 12 meses, o que indica a persistência de um padrão sazonal anual. Na FACP, destaca-se um coeficiente significativo na primeira defasagem, seguido por poucos valores relevantes nas demais, concentrando a dependência temporal nos períodos mais recentes. Em conjunto, esses resultados reforçam a necessidade de incorporar componentes sazonais ao modelo e serviram de base para a especificação inicial dos modelos SARIMA candidatos.

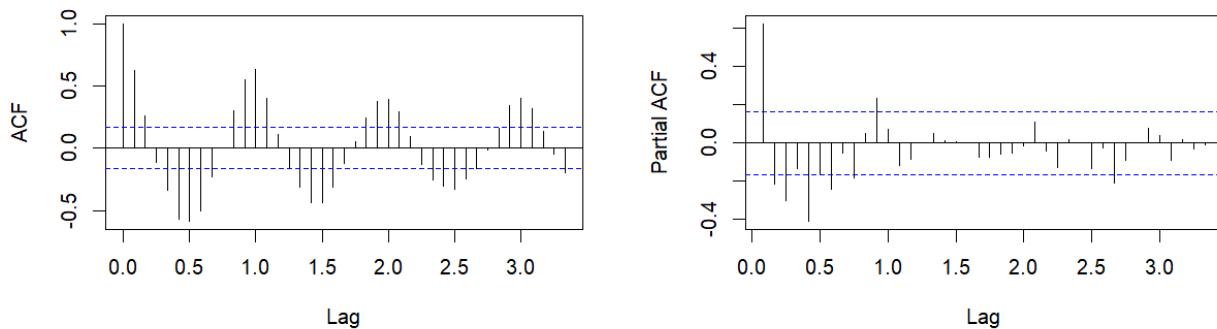


Figura 4.8: FAC e FACP da série logarítmica diferenciada, utilizadas na identificação das ordens do modelo.

Registra-se, ainda, que a série não apresentou meses sem notificações ao longo de todo o período analisado, sendo 28 o menor número mensal de casos observado. Essa característica viabiliza o emprego do MAPE como métrica de avaliação preditiva, uma vez que seu cálculo se torna inviável na presença de valores nulos.

4.2 TESTES DE HIPÓTESE

A Tabela 4.1 apresenta os resultados dos testes de hipótese aplicados à série logarítmica de casos de dengue.

Tabela 4.1: Resultados dos testes de hipótese aplicados à série.

Teste	Série	p-valor	Conclusão
Cox-Stuart	Log dos casos	0,017	Evidência de tendência
Cox-Stuart	Série diferenciada	0,905	Tendência removida após diferenciação
Fisher	Série diferenciada	$3,57 \times 10^{-19}$	Sazonalidade significativa

Os resultados indicaram presença de tendência na série logarítmica original, posteriormente removida após a diferenciação simples, subsidiando a adoção de $d = 1$. O teste de Fisher apontou sazonalidade significativa, corroborando a utilização de diferenciação sazonal com período anual ($D = 1$, $s = 12$).

4.3 SELEÇÃO DO MODELO SARIMA

A identificação das ordens dos modelos candidatos foi conduzida a partir da análise gráfica das funções de autocorrelação (FAC) e autocorrelação parcial (FACP) da série diferenciada, complementada pelo uso da função `auto.arima`, que indicou o modelo SARIMA(1, 1, 0)(0, 1, 1)₁₂ como referência inicial.

A partir dessa referência, foram ajustados sete modelos candidatos, cujos resultados comparativos são apresentados na Tabela 4.2.

Tabela 4.2: Comparativo de critérios de informação entre os modelos SARIMA candidatos.

Modelo	AIC	BIC	Coefficientes
M1 – SARIMA(1, 1, 0)(0, 1, 1)	166,38	174,96	Todos significativos
M2 – SARIMA(0, 1, 1)(0, 1, 1)	168,46	177,04	Todos significativos
M3 – SARIMA(1, 1, 1)(0, 1, 1)	168,20	179,64	ar1 e ma1 não signif.
M4 – SARIMA(1, 1, 0)(1, 1, 1)	163,17	174,61	Todos significativos
M5 – SARIMA(0, 1, 1)(1, 1, 1)	165,35	176,79	Todos significativos
M6 – SARIMA(2, 1, 0)(0, 1, 1)	168,03	179,47	ar2 não significativo
M7 – SARIMA(0, 1, 2)(0, 1, 1)	166,92	178,36	ma2 limítrofe

O modelo M4, SARIMA(1, 1, 0)(1, 1, 1)₁₂, foi selecionado como modelo de referência por apresentar os menores valores de AIC e BIC entre os candidatos avaliados, com coeficientes estatisticamente significativos ao nível de 5%. Os coeficientes estimados para o modelo M4 são apresentados na Tabela 4.3.

Tabela 4.3: Coeficientes estimados do modelo M4.

Parâmetro	Estimativa	Erro Padrão	z	p-valor
ar1	0,3572	0,0818	4,368	< 0,001
sar1	0,2210	0,0942	2,347	0,019
sma1	-0,9999	0,1367	-7,315	< 0,001

Nota: ar1 = coeficiente autorregressivo de ordem 1 (componente não sazonal); sar1 = coeficiente autorregressivo sazonal de ordem 1; sma1 = coeficiente de médias móveis sazonal de ordem 1.

4.3.1 DIAGNÓSTICO DE RESÍDUOS – MODELO M4

O diagnóstico de resíduos do modelo M4 indicou ausência de autocorrelação nas primeiras defasagens. O teste de Box-Pierce aplicado à defasagem inicial resultou em $p = 0,879$, não havendo evidências para rejeitar a hipótese de ruído branco nesse ponto. O comportamento dos resíduos ao longo do tempo e sua distribuição são apresentados nas Figuras 4.9 e 4.10.

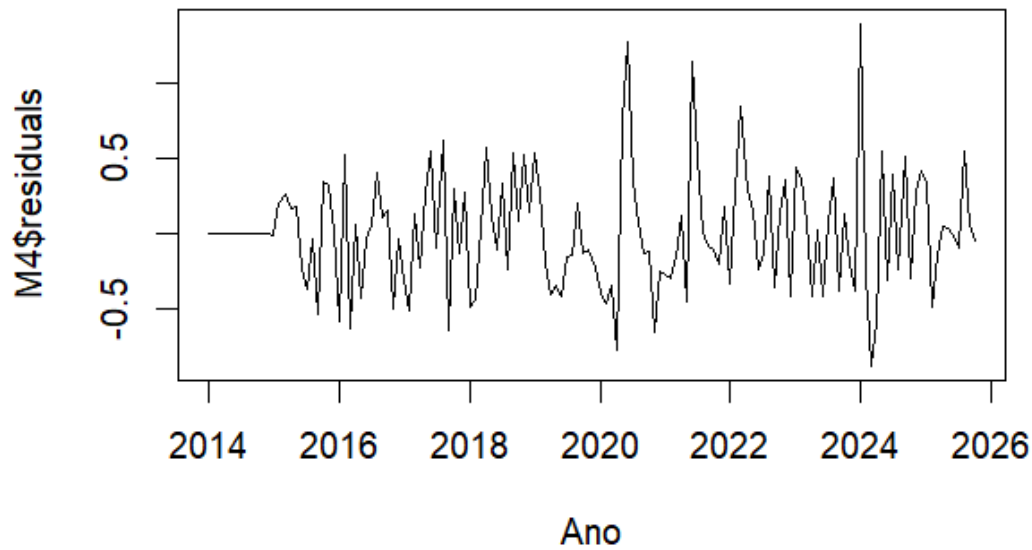


Figura 4.9: Resíduos do modelo M4 SARIMA(1, 1, 0)(1, 1, 1)₁₂ ao longo do tempo.

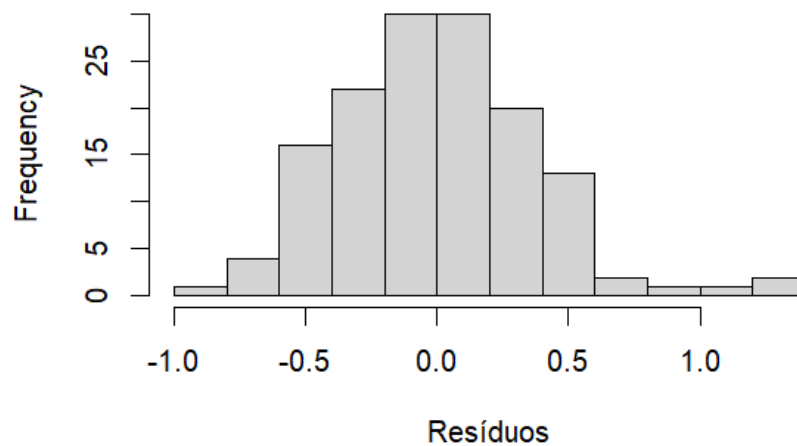


Figura 4.10: Histograma dos resíduos do modelo M4.

Entretanto, ao se avaliar defasagens mais longas por meio do teste de Ljung-Box, observou-se autocorrelação residual estatisticamente significativa na defasagem sazonal (lag 12: $p = 0,017$) e em defasagens superiores (lag 24: $p = 0,002$), indicando que o modelo não capturou integralmente a estrutura de dependência da série. Esse padrão pode ser observado nas funções de autocorrelação dos resíduos, apresentadas na Figura 4.11, e no periodograma acumulado da Figura 4.12.

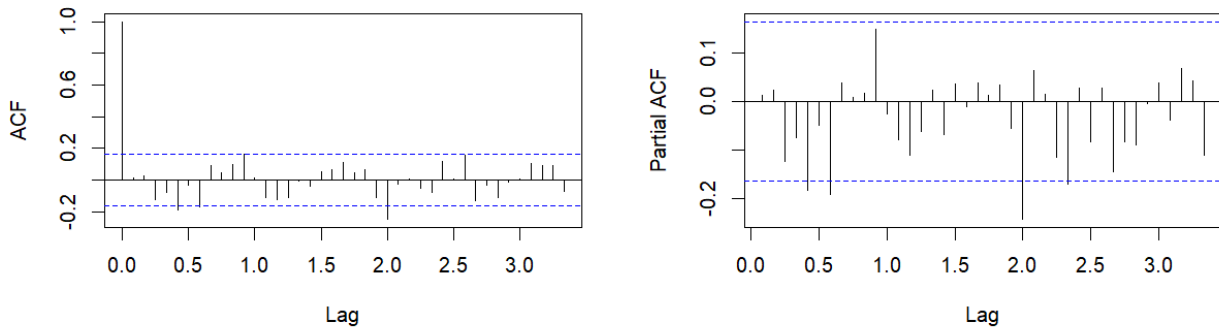


Figura 4.11: FAC e FACP dos resíduos do modelo M4.

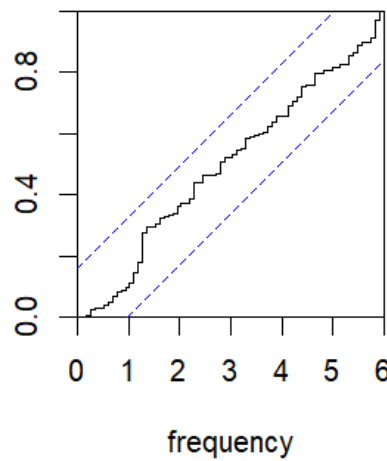


Figura 4.12: Periodograma acumulado dos resíduos do modelo M4.

Com o objetivo de investigar essa autocorrelação remanescente, foram testadas diversas estruturas sazonais alternativas, variando-se as ordens autorregressiva e de médias móveis, sazonais e não sazonais. Em nenhuma das configurações avaliadas a autocorrelação residual nas defasagens longas foi eliminada, e as estruturas que apresentaram leve melhora nesse aspecto introduziam coeficientes não significativos, comprometendo a parcimônia do modelo. Esse resultado sugere que a autocorrelação residual observada pode estar associada, em grande medida, ao comportamento atípico da série no período do surto epidêmico de 2025, um evento de magnitude extrema e difícil de ser plenamente representado por modelos lineares, mais do que a uma inadequação evidente da especificação adotada. O Erro Percentual Absoluto Médio (MAPE) no período de treino foi de 5,354%.

4.3.2 VALIDAÇÃO POR PREVISÃO ROLLING – MODELO M4

Em virtude do comportamento atípico observado na série durante o período de 2025, a validação do modelo M4 foi conduzida por meio da metodologia de previsão *rolling* (um passo

à frente), conforme descrito na seção de Metodologia. Os resultados da validação para o período de novembro de 2025 a abril de 2026 são apresentados na Tabela 4.4.

Tabela 4.4: Validação *rolling* do modelo M4 SARIMA.

Mês	Real	Previsto	IC Inf. (95%)	IC Sup. (95%)	Dentro do IC
Nov/2025	1.356	2.014	879	4.614	Sim
Dez/2025	728	1.857	811	4.250	Não
Jan/2026	913	1.615	698	3.740	Sim
Fev/2026	985	1.201	519	2.777	Sim
Mar/2026	1.541	1.197	519	2.761	Sim
Abr/2026	1.514	1.596	694	3.671	Sim

O modelo M4 apresentou MAPE de 55,03% e RMSE de 628,77 casos no período de validação, com 5 das 6 observações reais contidas dentro do intervalo de confiança de 95%. O elevado valor do MAPE reflete a dificuldade inerente de prever a magnitude exata dos casos no período imediatamente posterior ao maior surto histórico da série. O ano de 2025 registrou 53.266 casos, valor aproximadamente o dobro do segundo maior surto da série, ocorrido em 2019, com 30.779 casos. Ainda assim, o modelo acompanhou parcialmente o movimento de redução dos casos nos meses subsequentes ao surto, mantendo os valores reais dentro do intervalo de confiança na maior parte do período avaliado.

4.4 INCORPORAÇÃO DE VARIÁVEIS CLIMÁTICAS – MODELO SARIMAX

4.4.1 FUNÇÃO DE CORRELAÇÃO CRUZADA

A análise da função de correlação cruzada (CCF) entre as variáveis climáticas e a série logarítmica de casos de dengue evidenciou a existência de correlações estatisticamente significativas em diferentes defasagens temporais para ambas as variáveis, com sinais tanto positivos quanto negativos. Esse padrão reflete a forte sazonalidade compartilhada entre clima e casos, o que faz com que as correlações brutas se manifestem em múltiplas defasagens. A temperatura destacou-se por apresentar correlações de maior magnitude, incluindo valores significativos em defasagens em torno de quatro a seis meses, faixa compatível com a existência de um intervalo temporal entre as condições climáticas e a posterior alteração na incidência da doença. Diante da presença de correlações em múltiplas defasagens, tornou-se necessária uma avaliação sistemática das diferentes defasagens candidatas nos modelos SARIMAX, apresentada na seção seguinte. A Figura 4.13 apresenta as funções de correlação cruzada obtidas.

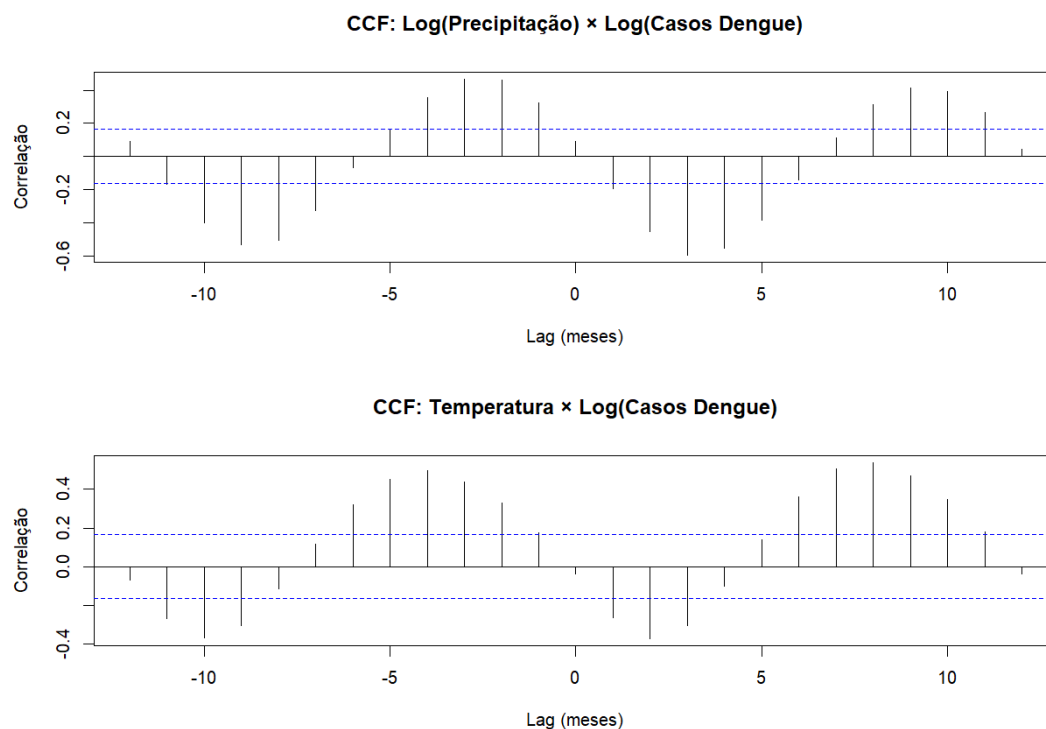


Figura 4.13: Função de correlação cruzada (CCF) entre as variáveis climáticas e a série logarítmica de casos de dengue.

4.4.2 SELEÇÃO DO LAG POR BUSCA SISTEMÁTICA

A partir dos lags apontados pela CCF, foi conduzida uma busca sistemática considerando defasagens de 1 a 8 meses, testando-se temperatura e precipitação separadamente e em conjunto, totalizando 24 modelos candidatos. Os resultados das quatro melhores configurações, ordenadas pelo critério BIC, são apresentados na Tabela 4.5.

Tabela 4.5: Melhores configurações da busca sistemática de defasagens.

Configuração	Lag	BIC	p (precip.)	p (temp.)
Só temperatura	5	170,39	—	< 0,05
Só temperatura	4	172,98	—	0,116
Precipitação + temperatura	5	174,13	0,303	0,013
Só precipitação	8	173,21	0,139	—

Nota: os valores de BIC e de p apresentados nesta tabela referem-se à etapa de seleção da defasagem, na qual todos os candidatos foram ajustados sobre a mesma amostra de 134 observações. Os valores do modelo final reajustado são apresentados a seguir.

A primeira e a segunda linhas da Tabela 4.5 representam configurações com apenas a temperatura como variável exógena, mas em defasagens distintas. Na configuração com lag 5, o coeficiente associado à temperatura é estatisticamente significativo e o BIC é o menor entre todas as combinações testadas. Na configuração com lag 4, embora a temperatura também seja

única covariável, o coeficiente não é significativo ao nível de 5% e o BIC é mais elevado, o que indica pior equilíbrio entre ajuste e parcimônia.

A configuração com temperatura defasada em 5 meses apresentou, portanto, o menor valor de BIC entre todas as configurações testadas na etapa de seleção ($BIC = 170,39$, sobre a amostra de 134 observações comum a todos os candidatos), com coeficiente estatisticamente significativo ao nível de 5%. A variável de precipitação não foi significativa em nenhuma configuração isolada, e sua inclusão conjunta com a temperatura elevou o BIC do modelo, indicando que a configuração mais parcimoniosa, apenas com a temperatura defasada, é a mais adequada.

4.4.3 MODELO SARIMAX FINAL

A configuração selecionada foi o modelo $SARIMAX(1, 1, 0)(1, 1, 1)_{12}$ com temperatura defasada em 5 meses, cujos coeficientes estimados são apresentados na Tabela 4.6.

Tabela 4.6: Coeficientes estimados do modelo SARIMAX com temperatura defasada em 5 meses.

Parâmetro	Estimativa	Erro Padrão	z	p-valor
ar1	0,3664	0,0832	4,405	< 0,001
sar1	0,2288	0,0974	2,350	0,019
sma1	-0,9999	0,1389	-7,201	< 0,001
temp_lag5	-0,1292	0,0598	-2,160	0,031

Nota: ar1 = coeficiente autorregressivo de ordem 1 (componente não sazonal); sar1 = coeficiente autorregressivo sazonal de ordem 1; sma1 = coeficiente de médias móveis sazonal de ordem 1; temp_lag5 = coeficiente associado à temperatura defasada em cinco meses.

O coeficiente negativo associado à temperatura defasada em cinco meses indica que, no modelo ajustado, aumentos na temperatura média observada cinco meses antes estão associados a reduções na série logarítmica de casos no mês corrente, mantida a estrutura autorregressiva e sazonal do modelo. Embora esse sinal possa parecer contraintuitivo à primeira vista, ele é compatível com a literatura que aponta relações não lineares entre temperatura e dengue, nas quais temperaturas excessivamente elevadas podem reduzir a sobrevivência do vetor ou deslocar temporalmente as condições mais favoráveis à transmissão. Esse resultado está em consonância com os achados de Santos et al. [20] e Leite [12], que também identificaram associação inversa entre a temperatura média e a incidência de dengue em seus respectivos estudos.

Uma vez definida a defasagem de cinco meses na etapa de seleção, o modelo SARIMAX final foi reajustado utilizando todas as observações válidas para essa defasagem, totalizando 137 observações. Para que a comparação com o modelo SARIMA puro (M4) fosse metodologicamente válida, o modelo M4 foi igualmente reajustado sobre essa mesma amostra efetiva de 137 observações. Nessa base comum, o modelo SARIMAX apresentou AIC de 159,76 e BIC de 173,86, enquanto o modelo M4 apresentou AIC de 162,33 e BIC de 173,61. Observa-se, por-

tanto, que a inclusão da temperatura defasada reduziu o AIC do modelo, indicando ganho de ajuste, ao passo que, pelo critério BIC, mais rigoroso na penalização de parâmetros adicionais, os dois modelos apresentaram desempenho praticamente equivalente, com leve vantagem para o modelo mais parcimonioso.

Quanto ao diagnóstico de resíduos, o modelo SARIMAX apresentou comportamento análogo ao do modelo SARIMA. Foi observada ausência de autocorrelação na defasagem inicial (Box-Pierce: $p = 0,804$), mas autocorrelação residual significativa nas defasagens sazonais e mais longas (Ljung-Box lag 12: $p = 0,014$; lag 24: $p = 0,002$). Esse padrão, semelhante ao observado no modelo M4, sugere que a autocorrelação remanescente está associada a características da própria série, em especial ao surto atípico de 2025, mais do que a uma característica específica de um dos modelos. Os resíduos do modelo e seu periodograma acumulado são apresentados nas Figuras 4.14 e 4.15. O MAPE no período de treino foi de 5,421%.

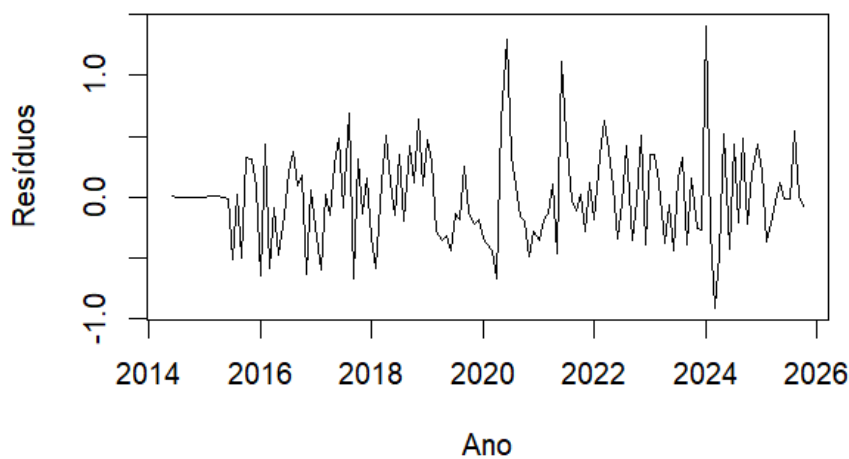


Figura 4.14: Resíduos do modelo SARIMAX (temperatura defasada em cinco meses) ao longo do tempo.

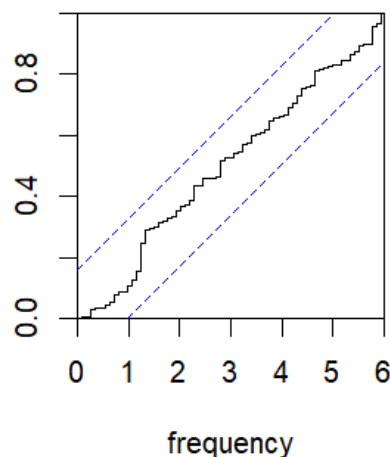


Figura 4.15: Periodograma acumulado dos resíduos do modelo SARIMAX.

4.4.4 VALIDAÇÃO POR PREVISÃO ROLLING – MODELO SARIMAX

A validação do modelo SARIMAX seguiu a mesma metodologia de previsão *rolling* aplicada ao modelo M4, incorporando, a cada iteração, a temperatura observada cinco meses antes do mês previsto. Os resultados são apresentados na Tabela 4.7.

Tabela 4.7: Validação *rolling* do modelo SARIMAX.

Mês	Real	Previsto	IC Inf. (95%)	IC Sup. (95%)	Dentro do IC
Nov/2025	1.356	1.862	809	4.285	Sim
Dez/2025	728	1.849	804	4.250	Não
Jan/2026	913	1.772	761	4.124	Sim
Fev/2026	985	1.475	633	3.435	Sim
Mar/2026	1.541	1.182	509	2.746	Sim
Abr/2026	1.514	1.418	612	3.287	Sim

O modelo SARIMAX apresentou MAPE de 60,79% e RMSE de 661,91 casos no período de validação, com 5 das 6 observações reais contidas dentro do intervalo de confiança de 95%, resultado equivalente ao obtido pelo modelo SARIMA puro.

4.5 COMPARATIVO FINAL ENTRE OS MODELOS

A Tabela 4.8 apresenta o comparativo entre o modelo SARIMA de referência (M4) e o modelo SARIMAX final, considerando os principais critérios avaliados ao longo deste trabalho.

Tabela 4.8: Comparativo final entre os modelos M4 SARIMA e SARIMAX.

Critério	M4 SARIMA(1, 1, 0)(1, 1, 1)	SARIMAX + temp lag 5
AIC	162,33	159,76
BIC	173,61	173,86
MAPE (treino)	5,462%	5,421%
MAPE (validação)	55,03%	60,79%
RMSE (validação)	628,77	661,91
Observações dentro do IC 95%	5 de 6	5 de 6
Significância da covariável	–	$p = 0,031$
Box-Pierce (resíduos)	$p = 0,879$	$p = 0,804$

Nota: os valores de AIC, BIC e MAPE de treino referem-se ao ajuste de ambos os modelos sobre a mesma amostra efetiva de 137 observações, assegurando a comparabilidade dos critérios de informação.

Do ponto de vista aplicado, o modelo SARIMAX apresentou melhor ajuste aos dados históricos segundo o critério AIC e identificou associação estatisticamente significativa entre a

temperatura defasada em cinco meses e a incidência de dengue em Uberlândia-MG. Esse resultado reforça que variáveis climáticas, em especial a temperatura, podem contribuir para antecipar períodos de maior risco, oferecendo subsídios para ações de vigilância e planejamento de campanhas de controle vetorial com antecedência.

Por outro lado, pelo critério BIC, mais rigoroso na penalização de parâmetros adicionais, os dois modelos apresentaram desempenho muito próximo, com leve vantagem para o SARIMA por ser mais parcimonioso. No período de validação, o modelo SARIMA também apresentou desempenho ligeiramente superior em termos de MAPE e RMSE, o que o torna uma opção mais estável para uso rotineiro em previsões operacionais, especialmente em cenários de surto recente e elevada incerteza.

Em conjunto, esses resultados indicam que os modelos desempenham papéis complementares na prática em saúde pública. O SARIMA, mais simples e ligeiramente superior em previsão, é adequado para o acompanhamento contínuo da série e para apoiar decisões de curto prazo. O SARIMAX, por incorporar informação climática, é particularmente útil para explorar cenários e compreender como variações de temperatura se relacionam com a dinâmica dos casos, contribuindo para o planejamento antecipado de intervenções e para a integração entre vigilância epidemiológica e monitoramento ambiental.

5. CONSIDERAÇÕES FINAIS

Este trabalho analisou o comportamento da dengue em Uberlândia-MG entre 2014 e 2026, empregando modelos de séries temporais SARIMA e SARIMAX para descrever a dinâmica da doença e verificar em que medida as variáveis climáticas, consideradas com defasagem, ajudam a explicar a variação dos casos.

A série mostrou-se fortemente sazonal e heterocedástica, com surtos que se repetem em intervalos regulares, porém com intensidades bastante distintas. O episódio de 2025 destacou-se de todos os anteriores e configurou o maior surto de toda a série analisada, contribuindo para a presença de tendência estatisticamente significativa nos dados. A magnitude desse surto mostra como eventos epidêmicos extremos podem alterar as propriedades de uma série temporal e influenciar a escolha do modelo mais adequado para representá-la.

Entre os modelos SARIMA avaliados, o $\text{SARIMA}(1, 1, 0)(1, 1, 1)_{12}$ foi o que reuniu melhor ajuste e coeficientes significativos ao nível de 5%, tendo sido adotado como modelo de referência. A inclusão da temperatura defasada em cinco meses, identificada pela função de correlação cruzada e confirmada na busca sistemática de defasagens, melhorou o ajuste do modelo segundo o critério AIC e apresentou coeficiente significativo. O sinal negativo desse coeficiente indica que temperaturas médias mais altas cinco meses antes tendem a se associar a uma incidência menor no mês corrente. Ainda que pareça contraditório à primeira vista, esse comportamento encontra respaldo na literatura, que descreve uma relação não linear entre temperatura e dengue, e coincide com o que Santos et al. [20] e Leite [12] observaram em seus estudos, ambos também apontando associação inversa com a temperatura. Já a precipitação não se mostrou significativa em nenhuma das configurações testadas, o que sugere que boa parte de seu efeito sazonal já é absorvida pela própria estrutura sazonal do modelo.

Na etapa de validação, feita por previsão um passo à frente, o SARIMA teve desempenho um pouco melhor que o SARIMAX em termos de MAPE e RMSE, embora os dois modelos tenham apresentado erros de previsão elevados no período. Mais do que apontar um “vencedor”, esse resultado sugere que os modelos cumprem papéis distintos e complementares. O SARIMA mostrou-se mais estável para prever, enquanto o SARIMAX permitiu medir e evidenciar a influência da temperatura sobre os casos. O erro de previsão alto, presente em ambos, não invalida a modelagem, mas revela um limite conhecido desse tipo de abordagem. Modelos que aprendem apenas com o histórico da própria série têm dificuldade em acompanhar transições bruscas entre regimes, como a que ocorreu entre o pico intenso do início de 2025 e o retorno da doença a patamares mais habituais nos meses seguintes.

O diagnóstico dos resíduos revelou outra limitação que merece registro. Nas primeiras defasagens, os resíduos dos dois modelos comportaram-se como ruído branco. Porém, ao se examinarem defasagens mais longas por meio do teste de Ljung-Box, verificou-se autocorrelação residual significativa nas componentes sazonais. Foram testadas diferentes estruturas sazonais na tentativa de eliminá-la, mas nenhuma delas resolveu o problema sem introduzir coeficientes não significativos e comprometer a parcimônia do modelo. Isso sugere que a autocorrelação remanescente está mais ligada ao comportamento atípico da série no ano do surto do que, necessariamente, a uma especificação inadequada. Esse achado abre espaço para que trabalhos futuros explorem alternativas capazes de lidar melhor com eventos extremos, como modelos de intervenção ou formulações não lineares. Outras extensões possíveis incluem a inclusão simultânea de mais variáveis exógenas, com o devido cuidado quanto à multicolinearidade entre elas, e a comparação dos modelos clássicos aqui empregados com abordagens de aprendizado de máquina.

Em síntese, os resultados mostram que a dinâmica da dengue em Uberlândia-MG está associada não apenas ao comportamento histórico da própria doença, mas também às condições climáticas observadas meses antes da ocorrência dos casos. A identificação de uma defasagem aproximada de cinco meses entre a temperatura e a incidência de dengue sugere a existência de uma janela potencial para ações preventivas e planejamento de medidas de controle vetorial antes da intensificação dos surtos. Embora os modelos avaliados apresentem limitações na representação de eventos epidêmicos extremos, eles contribuem para a compreensão dos mecanismos temporais que influenciam a evolução da doença e podem servir como ferramenta auxiliar para a vigilância epidemiológica no município.

REFERÊNCIAS BIBLIOGRÁFICAS

- [1] Ahdesmaki, M., Fokianos, K. e Strimmer, K.: *Robust regression for periodicity detection in non-uniformly sampled time-course data*. BMC Bioinformatics, 6:117, 2005.
- [2] Akaike, H.: *A new look at the statistical model identification*. IEEE Transactions on Automatic Control, 19(6):716–723, 1974. Disponível em: <https://doi.org/10.1109/TAC.1974.1100705>.
- [3] Andrade Azevedo, M. E. S. de, Bonilha, G. C., Oliveira, T. R. S. de, Crispim, M. J. F., Nogueira, L. S. R., Santi Lima, A. L. de, Araújo, L. M. de e Faria, L. G. P.: *Dinâmica temporal dos casos de dengue e fatores climáticos no Distrito Federal (2013 a 2023)*. Revista Ciências da Saúde Ceuma, 3(1):75, 2025.
- [4] Box, G. E. P. e Cox, D. R.: *An Analysis of Transformations*. Journal of the Royal Statistical Society, Series B, 26(2):211–252, 1964. Disponível em: <https://doi.org/10.1111/j.2517-6161.1964.tb00553.x>.
- [5] Box, G. E. P. e Jenkins, G. M.: *Time Series Analysis: Forecasting and Control*. Holden-Day, San Francisco, 1976.
- [6] Box, G. E. P., Jenkins, G. M., Reinsel, G. C. e Ljung, G. M.: *Time Series Analysis: Forecasting and Control*. John Wiley & Sons, Hoboken, 5ª ed., 2015.
- [7] Box, G. E. P. e Pierce, D. A.: *Distribution of Residual Autocorrelations in Autoregressive-Integrated Moving Average Time Series Models*. Journal of the American Statistical Association, 65, 1970. Disponível em: <https://doi.org/10.1080/01621459.1970.10481180>.
- [8] Cox, D. R. e Stuart, A.: *Some Quick Sign Tests for Trend in Location and Dispersion*. Biometrika, 42(1/2):80–95, 1955. Disponível em: <https://doi.org/10.1093/biomet/42.1-2.80>.
- [9] Fava, V. L.: *Manual de Econometria*. Rel. Téc., In: Vasconcelos, M. A. S.; Alves, D. Editora Atlas, São Paulo, 2000.
- [10] Hyndman, R. J. e Athanasopoulos, G.: *Forecasting: Principles and Practice*. OTexts, Melbourne, 2ª ed., 2018.

- [11] Hyndman, R. J. e Khandakar, Y.: *Automatic time series forecasting: the forecast package for R*. Journal of Statistical Software, 27(3):1–22, 2008.
- [12] Leite, E. S.: *Análise temporal da relação entre dengue e variáveis climáticas na cidade de Uberlândia – MG*. Tese de Doutorado, Universidade Federal de Uberlândia, Programa de Pós-Graduação em Geografia, Uberlândia, 2023. Disponível em: <http://doi.org/10.14393/ufu.te.2023.8041>.
- [13] Ljung, G. M. e Box, G. E. P.: *On a Measure of Lack of Fit in Time Series Models*. Biometrika, 65(2):297–303, 1978. Disponível em: <https://doi.org/10.1093/biomet/65.2.297>.
- [14] Mateus, A. e Caeiro, F.: *An R implementation of several Randomness Tests*. Em *AIP Conference Proceedings*, vol. 1618, pp. 531–534. American Institute of Physics, 2014.
- [15] Moraes, P. L. L. de, Castanha, P. M. S., Nascimento, G. I. L. A. e Montarroyos, U. R.: *Análise temporal da dengue associada a fatores climáticos em Garanhuns, Pernambuco, Brasil, de 2010 a 2019*. Research, Society and Development, 9(12), 2020. Disponível em: <https://doi.org/10.33448/rsd-v9i12.11138>.
- [16] Morettin, P. A. e Toloi, C. M. C.: *Análise de Séries Temporais*. Edgard Blücher, São Paulo, 2ª ed., 2006.
- [17] Priestley, M. B.: *Spectral Analysis and Time Series*. Academic Press, London, 1989.
- [18] R Core Team: *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2024. <https://www.R-project.org/>.
- [19] Schwarz, G.: *Estimating the Dimension of a Model*. The Annals of Statistics, 6(2):461–464, 1978. Disponível em: <https://doi.org/10.1214/aos/1176344136>.
- [20] Silva Santos, D. A. da, Rodrigues, J. Z., Olinda, R. A. de e Goulart, L. S.: *Relação das variáveis climáticas com os casos de dengue em um município do interior de Mato Grosso dos anos 2001 a 2015*. Multitemas, 23(55):5–24, 2018. Disponível em: <https://doi.org/10.20435/multi.v23i55.1742>.
- [21] Souza, I. C. A. de, Vianna, R. P. T. e Moraes, R. M. de: *Modelagem da incidência do dengue na Paraíba, Brasil, por modelos de defasagem distribuída*. Cadernos de Saúde Pública, 23(11):2623–2630, 2007.
- [22] Trapletti, A. e Hornik, K.: *tseries: Time Series Analysis and Computational Finance*, 2022. <https://CRAN.R-project.org/package=tseries>, R package version 0.10-51.
- [23] Viana, D. V. e Ignotti, E.: *A ocorrência da dengue e variações meteorológicas no Brasil: revisão sistemática*. Revista Brasileira de Epidemiologia, 16(2):240–256, 2013. Disponível em: <https://doi.org/10.1590/S1415-790X2013000200002>.

-
- [24] Wickham, H., François, R., Henry, L., Müller, K. e Vaughan, D.: *dplyr: A Grammar of Data Manipulation*, 2023. <https://CRAN.R-project.org/package=dplyr>, R package version 1.1.2.
- [25] Zeileis, A. e Grothendieck, G.: *zoo: S3 Infrastructure for Regular and Irregular Time Series*. Journal of Statistical Software, 14(6):1–27, 2005.
- [26] Zeileis, A. e Hothorn, T.: *Diagnostic Checking in Regression Relationships*. R News, 2(3):7–10, 2002. <https://CRAN.R-project.org/doc/Rnews/>.

A. DADOS UTILIZADOS

As Tabelas A.1, A.2 e A.3 apresentam os dados utilizados na modelagem dos casos de dengue, temperatura média mensal e precipitação total mensal em Uberlândia-MG.

Tabela A.1: Casos mensais confirmados de dengue em Uberlândia-MG.

Ano	Jan	Fev	Mar	Abr	Mai	Jun	Jul	Ago	Set	Out	Nov	Dez	Total Anual
2014	231	491	973	1209	878	279	98	37	53	43	51	74	4.417
2015	332	935	2846	5097	5309	1428	293	87	58	56	106	183	16.730
2016	403	1542	2207	2965	1538	346	86	45	54	65	63	81	9.395
2017	146	213	352	366	294	150	45	40	28	34	37	69	1.774
2018	105	115	178	388	414	149	61	30	46	65	153	374	2.078
2019	1653	5037	8824	9165	4615	805	174	62	81	85	117	161	30.779
2020	288	347	342	154	159	186	124	80	92	87	57	52	1.968
2021	74	81	85	79	41	56	54	39	45	42	35	52	683
2022	78	155	581	1199	1197	599	232	178	154	164	266	298	5.101
2023	860	2758	7594	7466	5534	1537	521	425	332	361	392	334	28.114
2024	2942	7825	6669	3028	2934	858	425	224	355	327	498	1033	27.118
2025	6151	9697	10812	9177	7656	2798	1008	933	1434	1516	1356	728	53.266
2026	913	985	1541	1514	–	–	–	–	–	–	–	–	4.953

Fonte: Sistema de Informação de Agravos de Notificação (SINAN/DATASUS).

Tabela A.2: Temperatura média mensal (°C) em Uberlândia-MG.

Ano	Jan	Fev	Mar	Abr	Mai	Jun	Jul	Ago	Set	Out	Nov	Dez	Média Anual
2014	23,61	23,89	23,05	22,45	20,57	20,37	19,55	21,47	25,03	26,04	23,34	22,78	22,68
2015	25,47	22,95	22,29	22,34	20,05	19,69	20,48	22,03	25,12	28,08	24,51	23,89	23,08
2016	23,51	23,98	23,14	22,71	21,07	18,46	19,82	21,93	24,85	24,62	22,94	23,30	22,36
2017	23,27	22,87	23,38	22,99	21,36	19,71	17,97	22,31	23,58	25,34	22,87	23,23	22,41
2018	23,47	22,99	23,58	21,35	19,71	20,68	20,08	21,69	24,07	24,00	22,43	23,58	22,30
2019	24,09	23,77	22,80	22,37	21,03	18,71	18,49	21,56	25,85	25,38	24,68	23,34	22,84
2020	23,92	22,99	22,65	21,82	18,53	19,68	19,71	20,74	25,26	26,25	24,43	24,07	22,50
2021	24,64	23,98	23,05	21,62	21,28	20,11	18,42	22,45	27,35	24,98	23,95	23,42	22,94
2022	22,67	22,77	23,25	22,85	18,68	19,51	19,84	21,23	24,26	24,76	24,10	22,87	22,23
2023	22,45	23,15	22,82	21,88	19,72	18,69	20,13	23,05	26,80	26,89	27,11	26,79	23,79
2024	25,27	25,04	25,32	24,22	23,46	21,40	21,63	22,11	27,16	26,50	23,36	24,24	24,14
2025	24,44	25,54	25,77	24,20	21,78	20,58	20,25	22,06	25,68	25,90	26,39	24,72	23,94

Fonte: NASA POWER (*Prediction Of Worldwide Energy Resources*).

Tabela A.3: Precipitação total mensal (mm) em Uberlândia-MG obtida da plataforma NASA POWER.

Ano	Jan	Fev	Mar	Abr	Mai	Jun	Jul	Ago	Set	Out	Nov	Dez	Total Anual
2014	81,8	93,2	118,1	86,7	16,4	0,6	49,9	0,6	20,1	54,6	229,8	151,6	903,4
2015	92,4	170,0	202,4	86,7	47,7	15,0	3,4	0,3	68,7	62,6	203,7	179,8	1132,7
2016	323,0	145,3	191,6	6,0	22,3	45,9	0,0	25,4	13,5	127,1	191,7	179,5	1271,3
2017	177,9	143,4	110,4	55,8	70,4	2,4	0,0	0,3	24,3	117,8	233,1	194,7	1130,5
2018	217,0	199,1	94,9	86,7	23,9	0,3	0,0	13,6	45,3	180,7	316,5	206,2	1384,2
2019	86,2	238,3	170,2	117,3	34,1	1,8	0,6	4,0	34,5	80,3	167,1	265,4	1199,8
2020	296,7	277,2	83,4	27,0	16,7	0,3	4,6	0,9	8,1	65,7	56,4	188,5	1025,5
2021	91,1	180,3	88,7	98,4	22,9	19,8	0,0	3,1	9,3	132,4	188,4	206,8	1041,2
2022	274,7	221,2	89,0	38,1	38,8	3,0	0,0	12,1	69,3	61,4	75,0	229,4	1112,0
2023	278,7	153,2	119,7	81,9	16,7	2,7	5,3	26,7	30,6	122,5	84,3	124,9	1047,2
2024	120,0	90,2	191,3	28,2	1,9	0,3	0,3	0,3	0,9	127,1	239,4	144,5	944,4
2025	115,6	41,7	117,5	72,9	2,8	6,0	1,6	3,7	8,4	47,7	48,3	207,1	673,3

Fonte: NASA POWER (*Prediction Of Worldwide Energy Resources*).