

MODELOS DE PREVISÃO PARA AVALIAÇÃO DE CASOS DE MALÁRIA NA AMAZÔNIA LEGAL BRASILEIRA

Natalia Santos¹, Everton Silva¹, Carlos Beluzo^{1,2}, Luciana C. Alves²

Instituto Federal de São Paulo (IFSP), Campinas - SP

Instituto de Filosofia e Ciências Humanas (IFCH)

Núcleo de Estudos de População "Elza Berquó" (NEPO)

Universidade Estadual de Campinas (UNICAMP), Campinas - SP

r.natalia@aluno.ifsp.edu.br, everton.silva@ifsp.edu.br

cbeluzo@gmail.com, lcalves@unicamp.br

Abstract. *Malaria is an infectious disease that affects thousands of people every year around the world. The region known as the Legal Brazilian Amazon, that is composed by nine states in the North, North-east and Central-West regions are the most affected in the national territory. This paper performed an analysis of atemporal series of case numbers of malaria occurred in this region, with the goal of identify and train statistical models, capable of forecast the incidence of the disease in the future (1 month), based on historical data. The data that was used were obtained from SIVEP-malaria (Epidemiological Surveillance Information System for malaria) which is responsible for collecting and disclosing informations about malaria in Brazil and contains records of more than three million of notifications between the years of 2006 and 2019. The results of the Work-Foward validation method for ARIMA models have been shown good to predict cases in the region as well as a whole and also in each of the states separately, except for the state of Tocantins that has a very low incidence of infections.*

Keywords: *malaria, time-series, arima, legal amazon.*

Resumo. *A malária é uma doença infecciosa que acomete milhares de pessoas todos os anos ao redor do mundo. A região conhecida como Amazônia Legal brasileira, composta por nove estados das regiões norte, nordeste e centro-oeste, é a mais afetada no território nacional. Este trabalho realizou uma análise da série temporal do número de casos de malária ocorridos nesta região, a fim de identificar e treinar modelos estatísticos capazes de prever a incidência da doença no futuro (1 mês), baseando-se em dados históricos. Os dados utilizados foram obtidos do SIVEP-Malária (Sistema de Informação de Vigilância Epidemiológica da Malária) que é responsável pela coleta e divulgação de informações sobre a Malária no Brasil e contém registro de mais de três milhões de notificações entre os anos de 2006 a 2019. Os resultados da utilização do método de Validação Walk-Foward para modelos ARIMA se mostraram satisfatórios para a previsão de casos na região como um todo e, também, em cada um dos estados separadamente, com exceção do estado de Tocantins que possui uma incidência muito baixa de infecções.*

Palavras-chave: *malaria, série temporal, arima, amazônia legal.*

1. Introdução

A malária é uma doença infecciosa presente majoritariamente em países em desenvolvimento e de clima subtropical e que faz centenas de milhões de vítimas, sendo milhares delas fatais, todos os anos (WHO, 2021). É uma doença parasitária, transmitida pela picada da fêmea do mosquito *Anopheles* infectada pelo protozoário do gênero *Plasmodium*¹.

Com o intuito de controlar a disseminação e as mortes causadas pela doença, a Organização Mundial da Saúde determinou, em 1993, a estratégia global para controle da malária (WHO, 1993) e passou a fazer um acompanhamento mais rigoroso dos números de casos reportados. Apesar dessa estratégia ter trazido bons resultados na redução de ocorrências, números divulgados pela própria Organização Mundial de Saúde - OMS no Relatório Mundial da Malária de 2021, ainda apontaram mais de 240 milhões de casos da doença em 2020 e mais de 627 mil mortes em todo o mundo (WHO, 2021).

De acordo com este relatório, o Brasil é o segundo país com maior número de casos de malária das Américas, ficando atrás apenas da Venezuela. A transmissão da doença em território nacional ocorre principalmente na região da Amazônia legal, a qual é composta pelos Estados do Amazonas (AM), Acre (AC), Roraima (RR), Amapá (AP), Pará (PA), Tocantins (TO), Rondônia (RO), Mato Grosso (MT) e Maranhão (MA), que são regiões que apresentam ambiente favorável a procriação do mosquito transmissor.

Realizar a previsão do número de casos e identificar possíveis períodos de surtos podem auxiliar o planejamento da tomada de decisão, manejo da doença, dimensionamento de recursos e antecipação de medidas de saúde pública para combate e tratamento da doença.

Diante disso, este trabalho propõe a utilização do método de Validação Walk-Foward, que permite obter a melhor previsão a cada passo de tempo através da atualização dos dados e dos parâmetros do modelo de forma otimizada (Tran et al., 2021), para realizar essa previsão do número de casos de malária para cada um dos estados que compõem a Amazônia legal brasileira para os anos de 2018 e 2019, utilizando, no método, o ajuste de modelos Autorregressivos Integrados de Médias Móveis (ARIMA).

2. Revisão da literatura

Entre os métodos de previsão conhecidos, modelos que envolvem análises de séries temporais estão entre os mais utilizados para esse fim. A análise preditiva de uma série temporal se resume em avaliar os valores históricos de um dado indicador com o objetivo de identificar um padrão de comportamento periódico (diário, semanal, mensal, semestral, anual, etc.) que possa ser usado como referência para expectativa de comportamento dos períodos futuro. Entre os métodos estatísticos utilizados para esse tipo de modelagem preditiva, os modelos da classe ARIMA podem ser considerados como os mais usados no contexto de epidemiologias (Perone, 2020).

Apesar de, atualmente, novos métodos de modelagem para previsão com séries temporais estarem sendo propostos, a maioria deles são baseados em processos de *deep-learning*, os quais são bastante complexos de serem interpretados e podem não funcionar

¹https://www.gov.br/saude/pt-br/centrais-de-conteudo/publicacoes/publicacoes-svs/malaria/guia_tratamento_malaria_2nov21_isbn_site.pdf. Acesso em 22/08/2022

para aplicações na vida real em áreas críticas, nas quais o domínio dos agentes de causa e efeito são indispensáveis, como é, no caso desse estudo, a saúde (Rojat et al., 2021). Nesse cenário, os modelos ARIMA ganham espaço ao combinarem a qualidade dos resultados obtidos, a uma maior facilidade de interpretação e aplicação em relação à procedimentos computacional e matematicamente mais complexos.

Os trabalhos de Purwa et al. (2014) e Wang et al. (2009) trazem comparações da aplicação de modelos ARIMA e métodos de IA. No primeiro, o modelo ARIMA apresenta melhores resultados do que aqueles mais elaborados, enquanto, no segundo, apesar de não obter resultados superiores em comparação aos demais, o ARIMA, diante de sua simplicidade, se mostra satisfatório em relação às métricas de performance estipuladas e capaz de produzir bons resultados de previsão.

Jere and Moyo (2016), destacam a simplicidade matemática e eficiência na descrição dos dados como um dos pontos positivos da aplicação dos modelos ARIMA. Os autores utilizaram um $ARIMA(1,0,0)$ ² para prever o número de casos de malária no distrito de Kabwe, na Zâmbia, para crianças com idade entre 1 e 5 anos.

A seguir é apresentada uma breve revisão de outros trabalhos que utilizaram a aplicação de modelos ARIMA na área da saúde e serviram de fundamentação para este artigo.

2.1. Aplicação de modelos ARIMA na área da saúde

Com início em fevereiro de 2020, a pandemia de COVID-19 causou milhares de mortes ao redor do planeta. Abolmaali and Shirzaei (2021) compararam a performance de quatro modelos de previsão para o número de casos da doença nas quatro regiões com maior número de infectados em agosto de 2020 e concluíram que o modelo ARIMA foi o que obteve melhores previsões.

A doença causada pelos vírus do tipo influenza, popularmente conhecida como gripe, também já foi alvo de estudos de previsão utilizando modelos ARIMA. He and Tao (2018) aplicaram diversas variações dessa classe de modelo para previsão do número de casos de gripe em crianças na cidade de Wuhan, na China, e concluíram que os modelos ARIMA podem ser usados para prever a taxa de diferentes tipos de vírus da gripe.

Sarpong (2013) utilizou esse mesmo tipo de modelo para prever o índice de mortalidade materna no Hospital Universitário *Okomfo Anokye Kumasi*, em Gana, e mostrou que o modelo $ARIMA(1,0,2)$ é adequado para ser usado para prever as taxas trimestrais de mortalidade materna desse hospital.

2.1.1. Estudos sobre Malária

Lima and Laporta (2021), após realizarem a comparação de diversos tipos de modelos (determinísticos, estocásticos e nulo) concluíram que os modelos ARIMA foram os que

²Os modelos ARIMA tradicionais possuem 3 parâmetros que definem sua ordem. Por convenção da literatura, esses parâmetros são p (parâmetro auto-regressivo), d (parâmetro de diferenciação da série) e q (parâmetro de média móvel) e a notação utilizada para identificar a ordem de um modelo é (p,d,q) . Desse modo, quando diz-se que um modelo ARIMA tem ordem $(1,0,0)$, isso significa que o seu parâmetro auto-regressivo tem ordem 1 e os demais, ordem 0.

se mostraram mais confiáveis para a previsão de número de casos mensais de malária para um horizonte de 12 meses de antecipação, no estado do Amapá.

Anokye et al. (2018) fizeram uso de um modelo ARIMA(1,1,2) para fazer a previsão mensal do número de infectados na cidade de Kumasi, em Gana. Os autores compararam o modelo em questão com um modelo quadrático e concluíram que o primeiro mostrou melhores resultados para as previsões mensais e é recomendado para uso da direção metropolitana de saúde da cidade e por pesquisadores que desejam monitorar e controlar casos de malária na região estudada.

Ferrão et al. (2021) utilizaram da metodologia *Box-Jenkins* para definição dos parâmetros e ajustou um modelo ARIMA (2,2,1)(1,1,1)[52] para os dados de ocorrência de malária no distrito de Sussundenga, em Moçambique.

Além disso, segundo Wolfarth-Couto et al. (2019) e Wolfarth-Couto et al. (2020), a incidência de casos de malária está intimamente relacionada a fatores climáticos sazonais como precipitação e nível de água dos rios. Alemu et al. (2011) apontou que outros fatores ambientais, também com características sazonais, como a temperatura e umidade do ar, afetam o ciclo de vida do vetor transmissor da doença causando impacto direto no número de ocorrências. Ferrão et al (2017) também destaca as componentes cíclicas e sazonais dos surtos de malária em combinação com fatores climáticos.

3. Análises Preliminares

A Figura 1 apresenta as etapas desenvolvidas antes da aplicação do método proposto. Nessa seção, serão apresentados mais detalhes sobre cada uma delas.

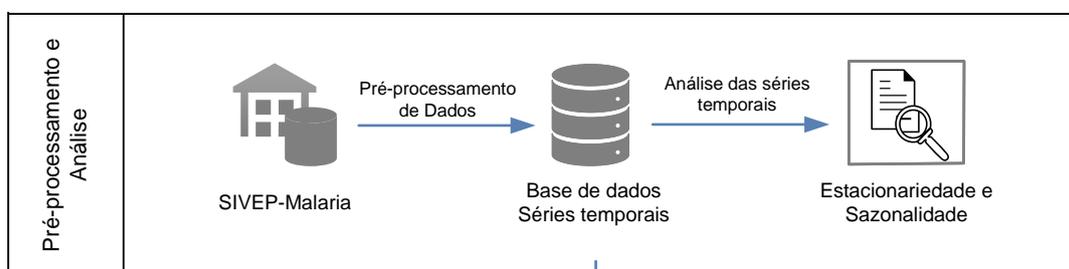


Figura 1. Visão geral das etapas que envolvem pré-processamento da base de dados e análise das séries temporais.

3.1. SIVPEP-Malária

Os dados utilizados no presente estudo são oriundos da plataforma SIVPEP-Malária, que é o sistema do Ministério da Saúde responsável pelo armazenamento e distribuição dos dados sobre a doença (Beluzo et al., 2021).

Esse sistema, que foi implantado em 2003, estabeleceu a separação entre a área com maior concentração de casos no território nacional, que é conhecida como Amazônia Legal brasileira e o restante do território nacional. A região da Amazônia Legal brasileira, que é alvo de interesse desse artigo, é responsável por 99% dos casos autóctones, ou seja, que se originam na mesma região onde são notificados, no país (Carlos et al., 2019).

Devido ao alto potencial endêmico característico da doença, os casos de malária devem, obrigatoriamente, ser reportados ao SIVPEP-Malária, tanto na rede pública, como

na privada, através do formulário padrão³, que contém dados sobre a unidade notificante, o paciente, o local da infecção, o exame e o tratamento, quando aplicáveis (Beluzo et al., 2021).

3.2. Pré-processamento de dados

A base de dados original possui um total de 29.414.406 registros de notificações no período de 2007 a 2019, da Amazônia Legal Brasileira. Cada registro é identificado exclusivamente pelos campos-chave número de notificação (COD_NOTI), data da notificação (DT_NOTIF) e município notificador (MUN_NOTIF). No total são 40 variáveis e a Tabela 1 apresenta a descrição de algumas delas, sendo classificadas em três grandes grupos: dados administrativos, informações do paciente e dados epidemiológicos e laboratoriais.

Tabela 1. Descrição da base de dados SIVEP-Malaria - Adaptada de Beluzo et al. (2021)

Grupo	Variável	Descrição
Dados Administrativos	COD_NOTI	Número da notificação
	DT_NOTIF	Data da notificação
	MUN_NOTIF	Código do município que está notificando
	UF_NOTIF	UF do paciente de acordo com o código do IBGE
	TIPO_LAM	Exame coletado de forma ativa ou passiva
Dados do Paciente	ID_PACIE	Idade do paciente
	SEXO	Sexo do paciente
	GESTANTE	Tempo de gestação (para mulheres gestantes)
	NIV_ESCO	Nível de escolaridade do paciente
	RACA	RaçaCor do paciente
	COD_OCUP	Principal atividade exercida pelo paciente
Dados Epidemiológicos / Laboratoriais	VIVAX	Paciente tratado para malária <i>vivax</i> nos últimos 60 dias
	FALCIPARUM	Paciente tratado para malária <i>falciparum</i> nos últimos 40 dias
	EXAME	Tipo do exame realizado (Gota espessa/Esfregaço ou Teste rápido)
	RES_EXAME	Espécie do plasmódio detectado
	QTD_CRUZ	Quantidade de parasitemia em cruzes)
	QTD_PARA	Quantidade de parasitos por <i>mm</i> ³
	HEMOPARASI	Resultado do exame para outros parasitas pesquisados
	SINTOMAS	Presença de sintomas
ESQUEMA	Código do esquema de tratamento utilizado	

Neste trabalho foi utilizada a base de dados pré-processada por Beluzo et al. (2021). O pré-processamento consiste em manter apenas os registros notificados e confirmados como infectados, totalizando 3.676.991 registros - 12,5% do conjunto de dados inicial. No supracitado trabalho está disponível uma descrição mais detalhada do pré-processamento realizado.

3.3. Análise descritiva das séries temporais

A Figura 2 mostra o número de casos de malária por mês na região da Amazônia Legal brasileira entre os anos de 2007 e 2019.

Ao analisar brevemente a figura 2, observa-se que entre os anos de 2007 e 2017, a incidência de casos de malária na região teve uma redução de aproximadamente 80% (de

³https://www.gov.br/saude/pt-br/centrais-de-conteudo/publicacoes/publicacoes-svs/malaria/folder_orientacoes-para-o-preenchimento-do-sivep-malaria.pdf. Acesso em 22/08/2022

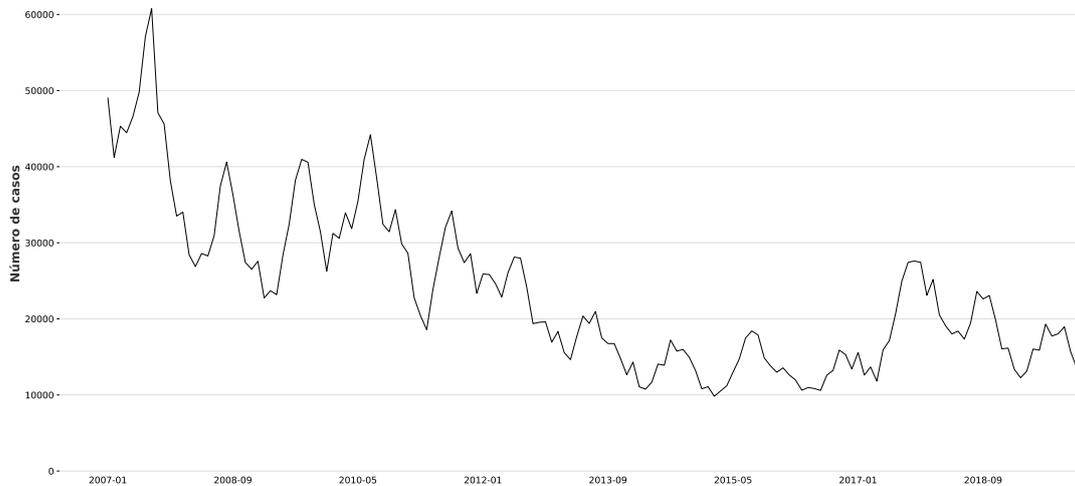


Figura 2. Série temporal unificada do número de casos de malária na região da Amazônia Legal brasileira por ano.

55 mil casos para 10 mil) e, apesar do aumento de 2018, a tendência de queda voltou a ser observada no ano de 2019.

A fim de entender a representatividade e impacto de cada um dos estados no cenário unificado, foi realizada, além da análise unificada do número de casos de malária em todos estados que compõe a Amazônia Legal Brasileira (denominada, aqui, por "Série Unificada"), a análise individual para cada um dos estados, e suas séries mensais são apresentadas na Figura 3.

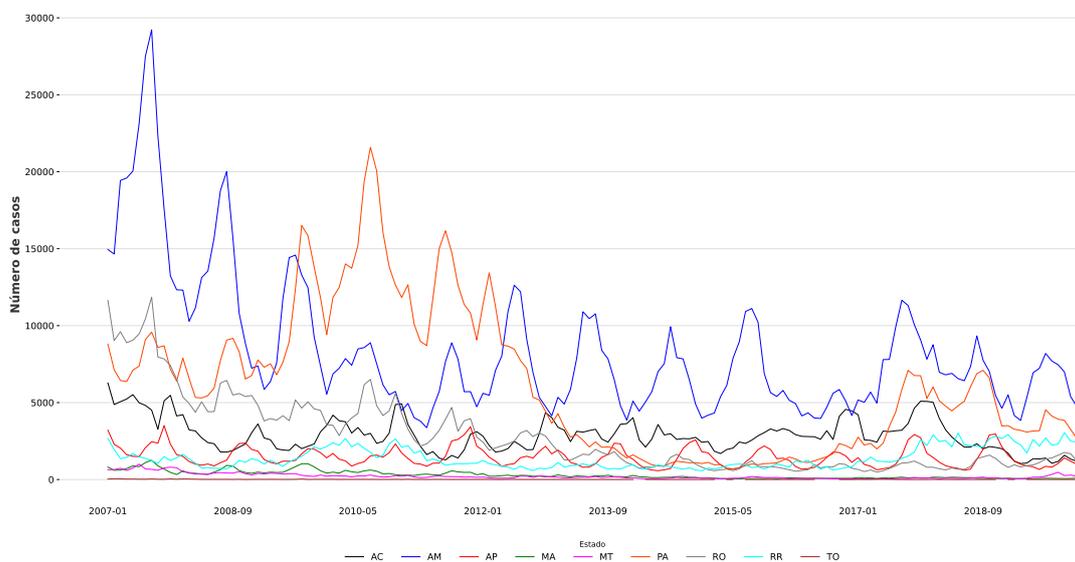


Figura 3. Série temporal do número de casos de malária nos estados da Amazônia Legal brasileira.

A análise conjunta das informações das Figuras 2 e 3 mostra que os estados do Amazonas e Rondônia são os que apresentam maior redução do número de casos entre 2007 e 2013 e, conseqüentemente, os que mais contribuem para a redução desse índice

na região como um todo. O estado do Pará, por sua vez, apesar de andar na direção contrária e sofrer um aumento na quantidade de notificações de malária nos anos de 2009 e 2010, também apresenta diminuição desse índice de 2010 a 2017 e contribui para queda observada na Figura. 2

Desse modo, dadas as observações das diferenças de representatividade e impacto de cada um dos estados no cenário unificado, o método de previsão escolhido foi aplicado não apenas para a série unificada, mas também para as séries mensais individuais de cada um dos estados.

3.4. Políticas públicas para combate da malária

A redução do número de casos de malária apontada nas análises anteriores, tanto para os estados com maior quantidade de notificações quanto para a série unificada, a partir de 2007, é resultado de intensos esforços governamentais para combate da doença.

O Programa Nacional de Prevenção e Controle da Malária (PNCM), desde sua criação em 2003, investiu fortemente na divulgação de políticas de saúde e direcionamento técnico para prevenção, diagnóstico, tratamento e combate à malária e alavancou esforços para a descentralização da vigilância epidemiológica através da capacitação dos funcionários e agentes de saúde municipais e estaduais. Entretanto, foi a partir de 2005 que foram, de fato, implantadas políticas e diretrizes de maior impacto como a implementação de novos esquemas de tratamento, controle de qualidade e monitoramento do desempenho do diagnóstico, testes de diagnóstico rápido em áreas de difícil acesso e sistema de detecção e alerta de epidemias. (Lapouble et al., 2015)

Em 2006, ano anterior ao início da queda do número de casos de malária observada na sessão acima, foi publicado, pelo ministério da saúde, o manual de Ações de Controle da Malária para profissionais da saúde (MS, 2006), que traz orientações de conduta frente a situações de surto e informações sobre tratamento e combate à doença.

Além disso, em dezembro desse mesmo ano, a Secretaria de Vigilância em Saúde (SVS) publicou a Portaria 47⁴, a partir da qual a instalação de qualquer assentamento e/ou empreendimento na região amazônica fica condicionada a existência de um plano de controle da malária na região. Essa medida é importante pois em uma análise histórica temos atividades como a ocupação desordenada das periferias das grandes cidades e os assentamentos de reforma agrária e extração mineral como fatores desencadeadores das epidemias dessa doença na região. (Lapouble et al., 2015)

3.4.1. Estacionariedade

Conforme escrito por Cryer (1986), para se aplicar modelos de previsão e métodos de inferência estatística à uma determinada série temporal, é necessário que esta satisfaça suposições previamente estabelecidas e, dentre essas suposições, a de estacionariedade tem posição de destaque.

Em resumo, uma série temporal é dita estacionária de suas medidas estatísticas, tais como, média, desvio padrão e autocorrelações são constantes ao longo do tempo. Ou seja, as leis probabilísticas que geram o processo, não mudam com o passar do tempo.

⁴https://www.normasbrasil.com.br/norma/portaria-47-2006_194740.html. Acesso em 22/08/202

Para avaliar a estacionariedade da séries de cada estado e também da série unificada, foi realizado o *Teste de raiz unitária de Dickey-Fuller aumentado*, o qual testa as seguintes hipóteses:

H_0 : existe raiz unitária (a série **não** é estatisticamente estacionária).

H_1 : não existe raiz unitária (a série é estatisticamente estacionária).

Os resultados do teste podem ser vistos na Tabela 2:

Tabela 2. P-valores resultantes do Teste de raiz unitária de Dickey-Fuller aumentado

Estado	AC	AM	AP	MA	MT	PA	RO	RR	TO	Série Unificada
P-valor	0,01	0,01	0,05	0,52	0,00	0,41	0,15	0,19	0,00	0,39

Dados os p-valores mostrados acima, temos que: (1) as séries temporais de AC, AM, MT e TO são estacionárias ao nível de significância de 1%; (2) as séries temporais do AP são estacionárias ao nível de significância de 5%; (3) as séries temporais de MA, PA, RO, RR e a série unificada **não** são estacionárias.

Neste trabalho foi considerado o nível de confiança de 95% de confiança, ou seja, as séries que não tiveram a hipótese de estacionariedade rejeitada ao nível de significância de 5%, foram consideradas estatisticamente estacionárias.

A abordagem aqui utilizada para tornar estacionária as séries não estacionárias, se baseia em trabalhar com as diferenças da série ao invés dos seus valores reais. Colocando em termos matemáticos, temos:

$$\Delta y_t = y_t - y_{t-1} \quad (1)$$

onde y_t é o número de casos registrados no momento t.

3.4.2. Sazonalidade

De acordo com Fisman (2007), a sazonalidade, que é um movimento periódico de alta (ou baixa) na incidência de doenças ocasionado em decorrência da passagem das estações do ano ou outros períodos determinados de tempo, é uma característica de muitas doenças infecciosas que apresentam alto grau de relevância para a saúde pública.

Para testar se as séries do número de casos de malária para os estados da Amazônia Legal brasileira apresentam comportamento sazonal, foi utilizada a função *check_seasonality* do pacote *darts.utils.statistics* do Python, a qual utiliza os valores das autocorrelações para identificar a presença, ou não, de sazonalidade nos dados.

A análise identificou que tanto a série unificada quanto as séries de número de casos por estado, com exceção do Tocantins, apresentam comportamento sazonal. O horizonte de sazonalidade, em meses, para cada uma das séries testadas é descrito na tabela 3.

Tabela 3. Horizonte sazonal, em meses, para os estados da região da Amazônia Legal brasileira

Estado	AC	AM	AP	MA	MT	PA	RO	RR	TO	Série Unificada
Horizonte Sazonal	11	12	12	12	36	10	23	84	-	12

4. Aplicação do Método para previsão

Para realização das previsões foi utilizado o método conhecido como *Validação Walk-Foward* (Figura 4), que consiste em atualizar a base de dados a medida que novas informações ficam disponíveis e treinar novamente o modelo incluindo esses dados no conjunto de treino para realizar novas previsões.

Em termos práticos, os dados de 2007 a 2015 foram usados para treinar o primeiro modelo e gerar uma previsão para o mês seguinte, janeiro de 2016. Em seguida, o dado referente ao mês de janeiro de 2016 foi incluído na base de treinamento e, com os dados atualizados, foi ajustado um novo modelo para realização da previsão para fevereiro de 2016 e, assim, sucessivamente.

As observações dos anos de 2007 e 2015 foram utilizadas para treino do modelo de previsão, as de 2016 e 2017 para validação do método e as de 2018 e 2019, para teste. A Figura 4 ilustra os processos de treinamento, validação e teste.

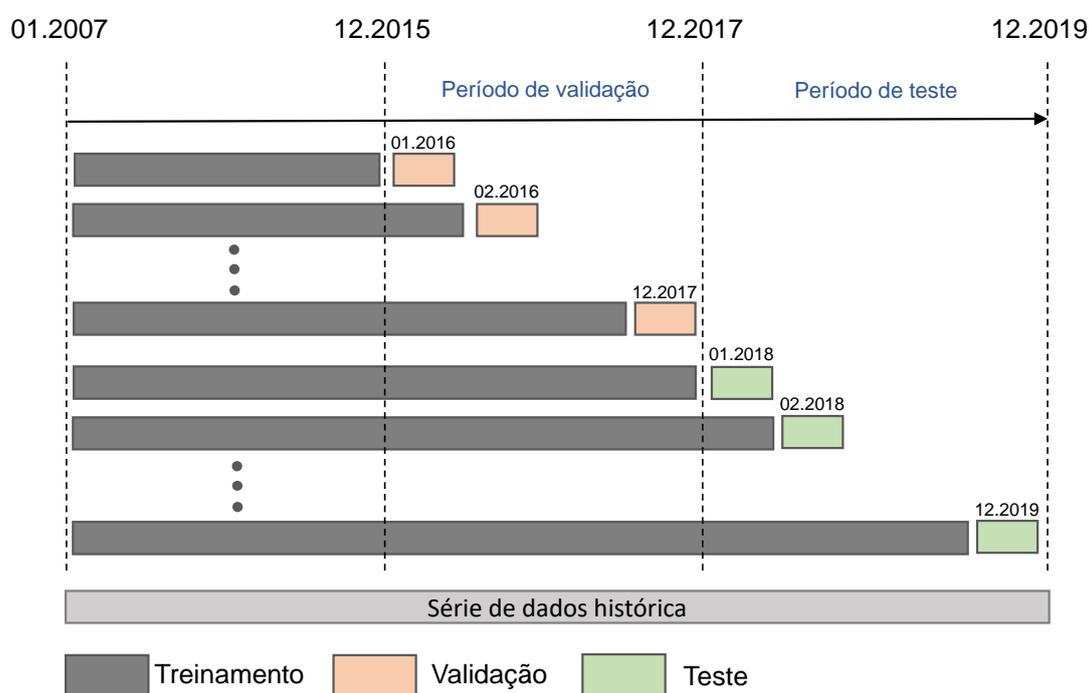


Figura 4. Processo de treinamento, validação e teste - Validação Walk-Foward.

Os modelos escolhidos e ajustados em cada etapa da previsão foram selecionados com a utilização da função `AUTOARIMA`, do pacote `pmdarima.arima` do Python, que procura identificar parâmetros ótimos para um modelo ARIMA com base em métodos estatísticos de diferenciação da série e minimização de erros.

Para as séries consideradas não estacionárias, o parâmetro de diferenciação da série foi iniciado em 1, ou seja, com a primeira diferença já aplicada, enquanto para as séries consideradas estacionárias, esse valor foi definido como 0, que representa a não diferenciação da série.

5. Análise dos Resultados

A análise dos resultados consiste em avaliar a capacidade preditiva dos modelos criados. Para isso existem métricas que consideram o valor real e o previsto. Neste trabalho utilizamos as métricas listadas abaixo:

- **MAPE** (*Mean absolute percentage error*): mostra a porcentagem de erro em relação aos valores reais. Quanto menor o MAPE, melhor o modelo se ajusta a um conjunto de dados.

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

- **RMSE** (*Root-mean-square deviation*): nos informa a distância entre os valores previstos e os valores observados em um conjunto de dados, em média. Quanto menor o RMSE, melhor o modelo se ajusta a um conjunto de dados. Como o cálculo eleva o erro ao quadrado, previsões distantes do real aumentam o valor da métrica muito rapidamente, ou seja, a métrica penaliza previsões distantes da real.

$$RMSE = \sqrt{\left(\frac{1}{n} \right) \sum_{t=1}^n (\hat{y}_t - y_t)^2} \quad (3)$$

- **RMSLE** (*Root Mean Squared Log Error*): possui cálculo e interpretação similar ao do RMSE, porém a utilização do logaritmo tende a evitar a penalização de diferenças elevadas entre valor predito e real quando os valores da série são de escalas muito grandes (o que acontece nesse estudo para estados com altos índices de incidência de malária).

$$RMSLE = \sqrt{\left(\frac{1}{n} \right) \sum_{t=1}^n (\log(y_t + 1) - \log(\hat{y}_t + 1))^2} \quad (4)$$

A análise dos resultados, tanto no conjunto de validação quanto no de teste, mostrou que o método escolhido produz bons resultados na previsão do número de casos de malária para quase todos os estados, com exceção do estado do Tocantins, que apesar de satisfazer a condição de estacionariedade, apresenta uma incidência de casos muito pequena com menos de 10 casos registrados para 104 dos 149 meses considerados. Já os estados que obtiveram melhores resultados para o conjunto de teste foram Amazonas (AM) e Pará (PA), os quais também aparecem, em grande parte do período, como os estados com maior número de casos (ver Figura 2).

Além disso, para a série unificada, o método de previsão escolhido também se mostrou eficiente, sendo, inclusive, melhor para esse caso do que para qualquer um dos estados separadamente. Os resultados obtidos são mostrados na Tabela 4.

As linhas destacadas em azul e negrito, mostram os estados para os quais o resultado do método aplicado se mostrou melhor no conjunto de validação, em comparação ao de teste. Já a linha em vermelho destaca o estado do Tocantins para o qual não se obteve um bom resultado do ajuste nem para o conjunto de validação, nem para o de teste.

Tabela 4. Métricas de avaliação de erro para previsão do número de casos de malária nos estados da Amazônia Legal brasileira

Estado	MAPE		RMSE		RMSLE	
	VALIDAÇÃO	TESTE	VALIDAÇÃO	TESTE	VALIDAÇÃO	TESTE
AC	0,13	0,27	519,35	536,92	0,02	0,07
AM	0,19	0,11	1528,95	837,56	0,05	0,02
AP	0,11	0,18	216,26	267,87	0,02	0,04
MA	0,26	0,2	32,64	30,36	0,15	0,1
MT	0,37	0,29	24,11	64,75	0,31	0,2
PA	0,23	0,1	670,55	522,43	0,13	0,02
RO	0,25	0,15	253,43	175,91	0,17	0,03
RR	0,18	0,13	265,38	396,12	0,05	0,02
TO	0,79	1,16	6,27	3,77	0,64	0,51
Série Unificada	0,12	0,08	2094,31	1759,78	0,02	0,01

6. Conclusão e Trabalhos Futuros

O método de previsão proposto se mostrou adequado para gerar previsões do número de casos de malária, na região estudada, considerando o horizonte de um mês e pode ser utilizado, conforme sugerido por Lima and Laporta (2021) e Ferrão et al. (2021), pelos órgãos públicos como ferramenta para planejamento de distribuição de recursos para controle e tratamento da infecção.

Para trabalhos futuros, vamos incorporar à nossa pesquisa os seguintes tópicos:

- Realizar previsões para diferentes horizontes: 2 meses a um 1 ano;
- Comparar o método atual com modelos de redes neurais recorrentes e modelos de aprendizado de máquina (regressão logística, árvores de decisão, etc) através de criação *lag features*;
- Criar e avaliar modelos que levem em consideração outras variáveis de característica temporal, como por exemplo índices pluviométricos, temperatura, umidade do ar, etc.

Referências

- Abolmaali, S. and Shirzaei, S. (2021). A comparative study of sir model, linear regression, logistic function and arima model for forecasting covid-19 cases. *AIMS Public Health*, 8(4):598.
- Alemu, A., Abebe, G., Tsegaye, W., and Golassa, L. (2011). Climatic variables and malaria transmission dynamics in jimma town, south west ethiopia. *Parasites & vectors*, 4(1):1–11.
- Anokye, R., Acheampong, E., Owusu, I., and Isaac Obeng, E. (2018). Time series analysis of malaria in kumasi: Using arima models to forecast future incidence. *Cogent social sciences*, 4(1):1461544.
- Beluzo, C. E., Arruda, N. M., de Souza Maia, V., and Alves, L. C. (2021). Data visualization for epidemiological and demographic data for malaria surveillance in the brazilian amazon, 2007-2019. *medRxiv*.

- Carlos, B. C., Rona, L. D., Christophides, G. K., and Souza-Neto, J. A. (2019). A comprehensive analysis of malaria transmission in Brazil. *Pathogens and Global Health*, 113(1):1–13.
- Cryer, J. D. (1986). *Time series analysis*, volume 286. Springer.
- Ferrão, J. L., Earland, D., Novela, A., Mendes, R., Tungadza, A., and Searle, K. M. (2021). Malaria temporal variation and modelling using time-series in Sussundenga district, Mozambique. *International Journal of Environmental Research and Public Health*, 18(11):5692.
- Fisman, D. N. (2007). Seasonality of infectious diseases. *Annu. Rev. Public Health*, 28:127–143.
- He, Z. and Tao, H. (2018). Epidemiology and ARIMA model of positive-rate of influenza viruses among children in Wuhan, China: A nine-year retrospective study. *International Journal of Infectious Diseases*, 74:61–70.
- Jere, S. and Moyo, E. (2016). Modelling epidemiological data using Box-Jenkins procedure. *Open Journal of Statistics*, 6(2):295–302.
- Lapouble, O. M. M., Santelli, A. C. F., Muniz-Junqueira, M. I., et al. (2015). Situação epidemiológica da malária na região amazônica brasileira, 2003 a 2012. *Revista Panamericana de Salud Pública*, 38:300–306.
- Lima, M. V. M. d. and Laporta, G. Z. (2021). Avaliação de modelos de previsão para ocorrência de malária no estado do Amapá, 1997-2016: um estudo ecológico. *Epidemiologia e Serviços de Saúde*, 30:e2020080.
- MS (2006). *Ações de controle da malária: manual para profissionais de saúde na atenção básica*. Ministério da Saúde Brasília.
- Perone, G. (2020). An ARIMA model to forecast the spread and the final size of COVID-2019 epidemic in Italy. *MedRxiv*.
- Purwa, T., Nafngiyana, U., and Suhartono, S. (2014). Comparison of ARIMA, transfer function and VAR models for forecasting CPI, stock prices, and Indonesian exchange rate: Accuracy vs. explainability. *Statistics*, 2020.
- Rojat, T., Puget, R., Filliat, D., Del Ser, J., Gelin, R., and Díaz-Rodríguez, N. (2021). Explainable artificial intelligence (XAI) on timeseries data: A survey. *arXiv preprint arXiv:2104.00950*.
- Sarpong, S. A. (2013). Modeling and forecasting maternal mortality; an application of ARIMA models. *International Journal of Applied*, 3(1):19–28.
- Tran, T. N., Phuc, D. T., et al. (2021). Grid search of multilayer perceptron based on the walk-forward validation methodology. *International Journal of Electrical and Computer Engineering*, 11(2):1742.
- Wang, W.-C., Chau, K.-W., Cheng, C.-T., and Qiu, L. (2009). A comparison of performance of several artificial intelligence methods for forecasting monthly discharge time series. *Journal of Hydrology*, 374(3-4):294–306.
- WHO (1993). *A global strategy for malaria control*. World Health Organization.
- WHO (2021). *World malaria report 2021*. World Health Organization.
- Wolfarth-Couto, B., Filizola, N., and Durieux, L. (2020). Padrão sazonal dos casos de malária e a relação com a variabilidade hidrológica no estado do Amazonas, Brasil. *Revista Brasileira de Epidemiologia*, 23:e200018.
- Wolfarth-Couto, B., Silva, R. A. d., and Filizola, N. (2019). Variabilidade dos casos de malária e sua relação com a precipitação e nível d'água dos rios no estado do Amazonas, Brasil. *Cadernos de Saúde Pública*, 35.