

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`

Motivação

Métodos de Ciência de Dados como ferramentas de apoio à Vigilância Epidemiológica de malária na Amazônia Legal Brasileira

Global Grand Challenges

ABOUT PARTNERSHIPS CHALLENGES **AWARDED GRANTS** GRANT OPPORTUNITIES NEWS

Data Science Applied to Epidemiological and Demographic Information as a Strategy to Simulation and Malaria Vigilance Monitoring in the Brazilian Amazon

The objective of this project is to develop data visualization tools and malaria outbreaks spreading simulation-based in machine learning methods, using demographic, epidemiological, climatic and clinical data related to malaria in the Brazilian Amazon. Such tools will be provided through a Web Platform with supporting tools to public administrators; the goal is for it to be ready for use in the short term. All the infrastructure for the creation and provision for a Web Platform had been developed.

[More information about Research on Malaria Prevention, Detection, and Control →](#)

SHARE THIS

INITIATIVE
Grand Challenges Brazil

CHALLENGE
Malaria Elimination

FUNDING DATE
Jun 25, 2020

PRINCIPAL INVESTIGATOR
Luciana Correia Alves

ORGANIZATION
Universidade Estadual de Campinas

LOCATION
Campinas, São Paulo, Brazil

PVD Malária

INICIAL Sobre Dashboards Visualizações Modelos Produções Contato

Sobre a Malária no Brasil

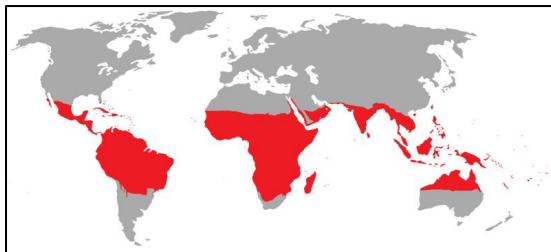
A malária é uma doença grave causada por protozoários do gênero *Plasmodium* que são transmitidos para as pessoas por meio da picada de fêmeas de mosquitos da sub-família Anophelinae e está presente em praticamente todas as regiões tropicais e subtropicais do planeta. Em 2020 foram contabilizados 241 milhões de casos e 627.000 óbitos, sendo que 95% dos casos e 96% dos óbitos se concentraram no continente Africano, onde a infecção por *P. falciparum*, responsável pelas formas mais graves da doença, prevalece (OMS, 2021).

A malária representa um dos principais problemas de saúde pública no mundo em desenvolvimento e continua sendo um grande desafio para o Brasil, que apesar do conhecimento acumulado sobre a doença e dos anos de campanha nacionais para combatê-la, concentra cerca de 34,4% dos casos da doença registrados no continente americano (ONU, 2017). Atualmente, 99% dos casos de malária no Brasil ocorrem na Amazônia, habitat natural para cerca de 50 espécies de anofelinos, dentre os quais aproximadamente 20 são potenciais vetores de malária, destacando-se a espécie *Nyssorhynchus darlingi* como principal vetor no Brasil (Carlos et al. 2019).

[Ler mais](#)

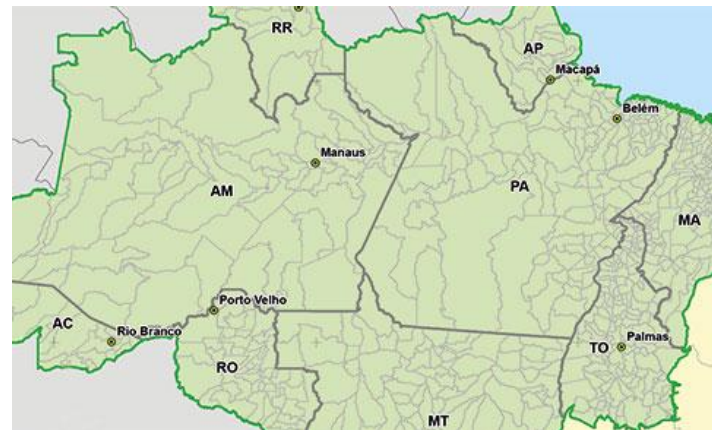
Introdução

- Malária doença infecciosa
 - Presente em países em desenvolvimento com clima subtropical
 - Centenas de milhões de vítimas, milhares delas fatais, todos os anos (WHO, 2021)
 - *240 mi em 2020, com mais de 627 mil mortes no mundo*
 - Doença parasitária (protozoário *Plasmodium*)
 - Transmitida por mosquitos *Anopheles* (*Guia de Tratamento de Malária - MS*).



Introdução

- Brasil possui 2º maior número de casos das Américas (Venezuela em 1º)
 - Principalmente na Amazônia Legal Brasileira
 - Ambiente favorável à procriação do mosquito
- Prever casos e identificar períodos de surtos favorecem planejamento da manejo da doença
 - Dimensionamento de recursos
 - Antecipação de medidas de saúde pública para combate e tratamento



Proposta

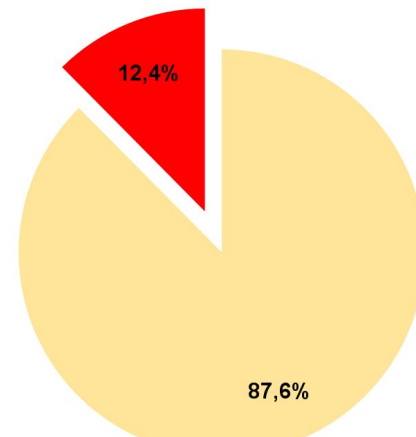
- Construção de série temporal mensal com número de casos
- Criação de um modelo de predição
 - Utilização do método de Validação **WalkFoward**
 - Permite obter melhor previsão a cada passo de tempo (Tran et al., 2021)
 - Seccionados por UFs da Amazônia Legal Brasileira
 - **ARIMA (Baseline)**
 - Método mais utilizado em contextos epidemiológicos (Perone, 2020)
 - Maior facilidade de interpretação dos resultados

Conjunto de dados - SIVEP-Malaria

- Programa Nacional de Controle e Prevenção de Malária
- Monitoramento espacial e temporal de epidemias
- Plataforma de vigilância de malária coleta e disseminação de dados relevantes
- Utilizado para construir indicadores epidemiológicos
 - Número de casos, Índice Parasitário, letalidade
 - Perfil demográfico e socioeconômico

Notificações de 2007 à 2019

- Negativos (25 mi)
- Positivos (360k)



Conjunto de dados - SIVEP-Malaria

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 mm^3
	HEMOPARASI	Resultado do exame para outros parasitas pesquisados
	SINTOMAS	Presença de sintomas
	ESQUEMA	Código do esquema de tratamento utilizado

Pré-processamento e Análise

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

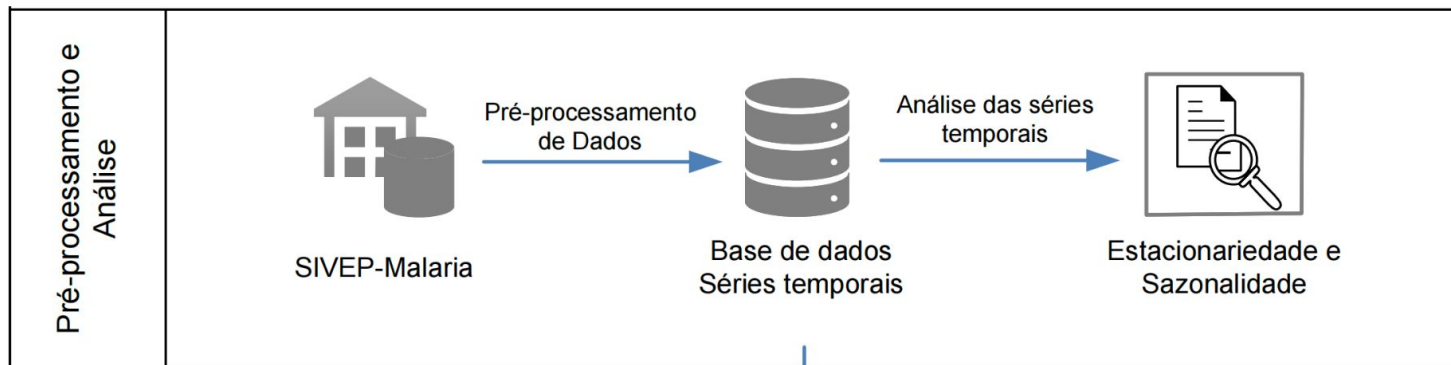


Figura 1. Visão geral das etapas que envolvem pré-processamento da base de dados e análise das séries temporais.

Pré-processamento e Análise (Cenário Unificado, todas UFs)

Série mensal de número de Casos de Malária

- Entre 2007 e 2017 redução de 80%
- Aumento em 2018
- Nova queda em 2019

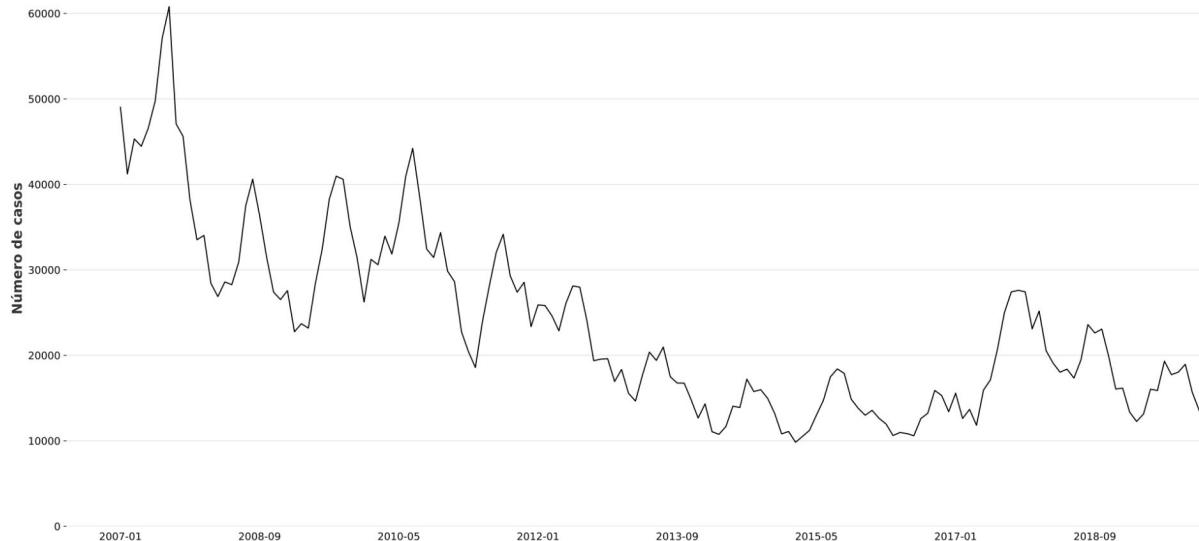
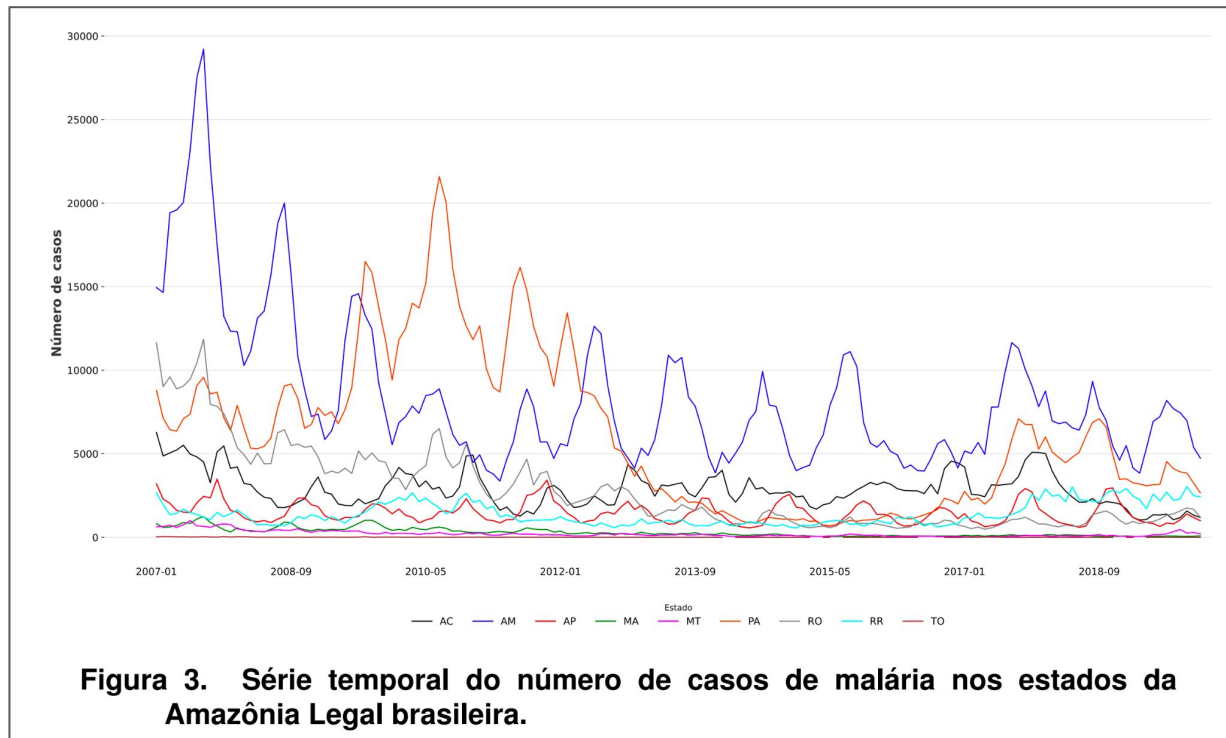


Figura 2. Série temporal unificada do número de casos de malária na região da Amazônia Legal brasileira por ano.

Pré-processamento e Análise (Por UF)

Série mensal de número de Casos de Malária (Por UF)

- 2007-2013: AM e RO maior queda
- 2009-2010: PA aumento importante, com nova queda em 2010-2017



Pré-processamento e Análise

"Dadas as observações das diferenças de representatividade e impacto de cada um dos estados no cenário unificado, o método foi aplicado para a série unificada e também para as séries mensais individuais de cada uma das UF"

Análise de Estacionaridade

- Pressuposto para aplicação de modelos de previsão e métodos de inferência estatística a uma determinada série temporal (Cryer, 1986)

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

**nível de confiança de 95%*

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.

Análise de Estacionaridade

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 .

Análise de Sazonalidade

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.

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

Validação *Walk-Foward*

"Atualiza 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"

- Modelo inicial: 2007/15
- WF a partir de Jan/2016
- Conjunto de testes: 2018/19

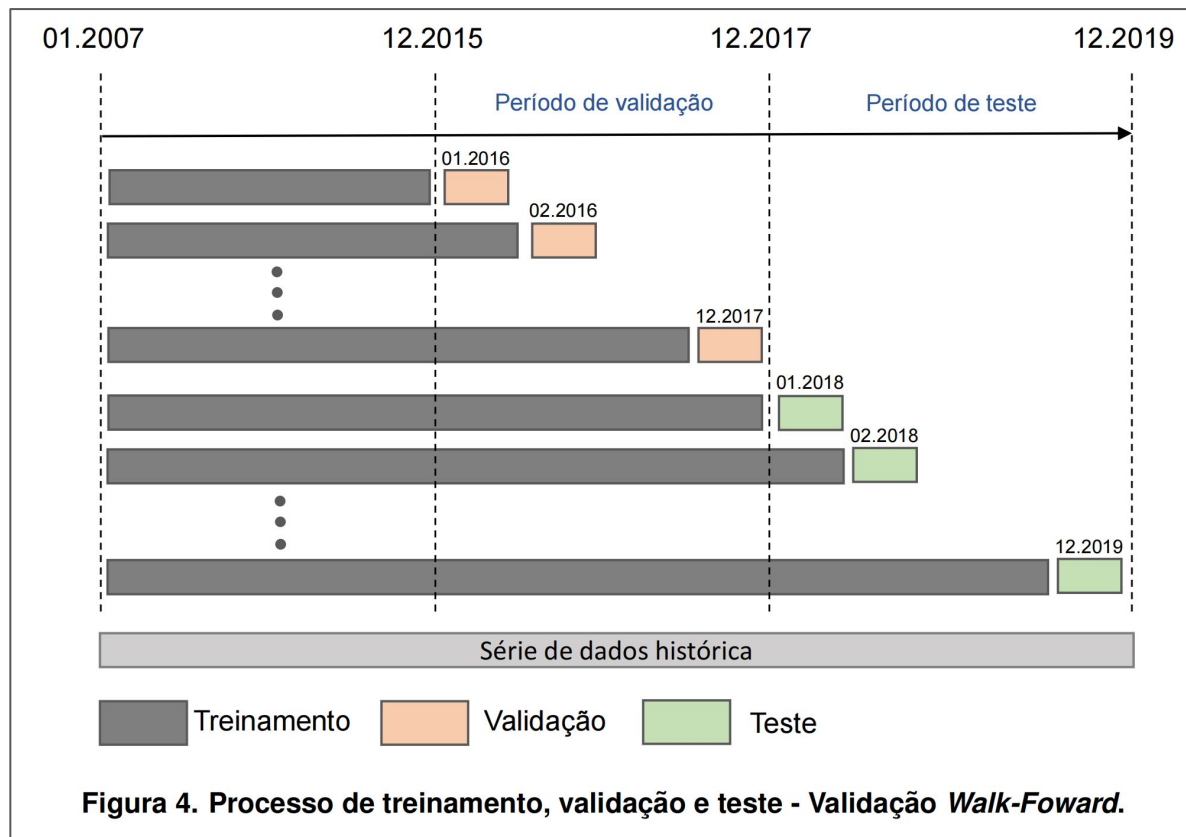


Figura 4. Processo de treinamento, validação e teste - Validação *Walk-Foward*.

Métricas e Resultados

- Modelos escolhidos e ajustados em cada etapa da previsão com a utilização da função **AUTOARIMA** (lib *pmdarima.arima* do *Pyhton*)
 - Identifica parâmetros ótimos para um modelo **ARIMA** com base em métodos estatísticos de diferenciação da série e minimização de erros
- Métricas de avaliação
 - **MAPE** (Mean absolute percentage error)
 - **RMSE** (Root-mean-square deviation)
 - **RMSLE** (Root Mean Squared Log Error)

Métricas e Resultados

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

Conclusão e Trabalhos Futuros

- Método proposto se mostrou adequado para previsões do número de casos de malária, na região estudada, considerando o horizonte de um mês
- Pode ser utilizado por órgãos públicos como ferramenta para planejamento de distribuição de recursos / controle e tratamento (Lima e Laporta, 2021; Ferrão et al., 2021)
- Próximos passos
 - Realizar previsões para diferentes horizontes: 2 meses a um 1 ano
 - Realizar experimentos utilizando redes neurais recorrentes
 - Realizar experimentos utilizando modelos de aprendizado de máquina
 - Incluir outras variáveis já conhecidas como correlacionadas ao problema
 - Característica temporal: índices pluviométricos, temperatura, umidade do ar

Acknowledgment

Fulfillment



Support



Funding



MINISTÉRIO DA
SAÚDE



This work is funded by Bill & Melinda Gates Foundation [ID INV 003970], Brazilian Ministry of Health and Brazilian National Council for Scientific and Technological Development [443048/2019-3].