

Situação: O preprint foi publicado em um periódico como um artigo
DOI do artigo publicado: <https://doi.org/10.1590/s1679-49742021000100007>

Avaliação de modelos de predição para ocorrência de malária no estado do Amapá, 1997-2016: um estudo ecológico

Marcos Venicius Malveira de Lima, Gabriel Zorello Laporta

<https://doi.org/10.1590/SciELOPreprints.1322>

Este preprint foi submetido sob as seguintes condições:

- O autor submissor declara que todos os autores responsáveis pela elaboração do manuscrito concordam com este depósito.
- Os autores declaram que estão cientes que são os únicos responsáveis pelo conteúdo do preprint e que o depósito no SciELO Preprints não significa nenhum compromisso de parte do SciELO, exceto sua preservação e disseminação.
- Os autores declaram que a pesquisa que deu origem ao manuscrito seguiu as boas práticas éticas e que as necessárias aprovações de comitês de ética de pesquisa estão descritas no manuscrito, quando aplicável.
- Os autores declaram que os necessários Termos de Consentimento Livre e Esclarecido de participantes ou pacientes na pesquisa foram obtidos e estão descritos no manuscrito, quando aplicável.
- Os autores declaram que a elaboração do manuscrito seguiu as normas éticas de comunicação científica.
- Os autores declaram que o manuscrito não foi depositado e/ou disponibilizado previamente em outro servidor de preprints ou publicado em um periódico.
- O autor submissor declara que as contribuições de todos os autores estão incluídas no manuscrito.
- O manuscrito depositado está no formato PDF.
- Os autores declaram que caso o manuscrito venha a ser postado no servidor SciELO Preprints, o mesmo estará disponível sob licença [Creative Commons CC-BY](https://creativecommons.org/licenses/by/4.0/).
- Caso o manuscrito esteja em processo de avaliação ou sendo preparado para publicação mas ainda não publicado por um periódico, os autores declaram que receberam autorização do periódico para realizar este depósito.

Submetido em (AAAA-MM-DD): 2020-10-21

Postado em (AAAA-MM-DD): 2021-04-13

Como citar este artigo:

Lima MVM, Laporta GZ. Avaliação de modelos de predição para ocorrência de malária no estado do Amapá, 1997-2016: um estudo ecológico. Epidemiol Serv Saúde [preprint]. 2020 [citado 2020 out 7]:[21 p.]. Disponível em: <https://doi.org/10.1590/s1679-49742021000100007>

Artigo original

Avaliação de modelos de predição para ocorrência de malária no estado do Amapá, 1997-2016: um estudo ecológico*

Evaluation of prediction models for the occurrence of malaria in the state of Amapá, Brazil, 1997-2016: an ecological study

Evaluación de modelos de predicción para la aparición de la malaria en el estado del Amapá, Brasil, 1997-2016: un estudio ecológico

Marcos Venicius Malveira de Lima^{1,2} - orcid.org/0000-0002-0332-2721

Gabriel Zorello Laporta¹ - orcid.org/0000-0001-7412-9390

¹Centro Universitário Saúde ABC, Faculdade de Medicina, Setor de Pós-Graduação, Pesquisa e Inovação, Santo André, SP, Brasil

²Secretaria de Estado de Saúde do Acre, Diretoria de Ações Programáticas e Vigilância em Saúde, Rio Branco, AC, Brasil

*Estudo financiado com recursos do ‘Projeto Acre – Saúde na Amazônia Ocidental’, estabelecido entre a Secretaria de Estado de Saúde do Acre e o Centro Universitário Saúde ABC (Convênio nº 007/2015). Gabriel Zorello Laporta foi contemplado com Bolsa de Produtividade em Pesquisa, pelo Conselho Nacional de Desenvolvimento

Científico e Tecnológico (CNPq) do Ministério da Ciência, Tecnologia, Inovações e Comunicações (MCTIC): Processo nº 307432/2019-0.

Endereço para correspondência:

Marcos Venicius Malveira de Lima – Secretaria de Estado de Saúde do Acre, Rua Benjamim Constant, nº 907, Centro, Rio Branco, AC, Brasil. CEP: 69914-220

E-mail: marcos.malveira@ac.gov.br

Recebido em 09/04/2020

Aprovado em 02/09/2020

Editora associada: Lúcia Rolim Santana de Freitas - orcid.org/0000-0003-0080-2858

Resumo

Objetivo. Avaliar a capacidade preditiva de diferentes modelos de série temporal de casos de malária no estado do Amapá, Brasil, no período 1997-2016. **Métodos.** Estudo ecológico de séries temporais com casos de malária registrados no Amapá. Foram utilizados dez modelos estatísticos determinísticos ou estocásticos para simulação e teste em horizontes de previsão de 3, 6 e 12 meses. **Resultados.** O teste inicial mostrou que a série é estacionária. Os modelos determinísticos apresentaram melhor desempenho do que os modelos estocásticos. O modelo ARIMA apresentou erros absolutos menores do que 2% na escala logarítmica e erros relativos 3,4-5,8 vezes menores em relação ao modelo nulo. A predição de casos futuros de malária nos horizontes de 6 e 12 meses de antecedência foi possível. **Conclusão.** Recomenda-se o uso de modelo ARIMA para a previsão de cenários futuros e para a antecipação do planejamento nos serviços de saúde dos estados da Região Amazônica.

Palavras-chave: Estudos de Séries Temporais; Malária; Técnicas de Apoio para a Decisão; Monitoramento Epidemiológico; Previsões.

Abstract

Objective. To evaluate predictive power of different time-series models of malaria cases in the state of Amapá, Brazil, in the period 1997-2016. **Methods.** This is an ecological study of time series with malaria cases registered in the state of Amapá. Ten

deterministic or stochastic statistical models were used for simulation and testing in 3, 6, and 12 month forecast horizons. **Results.** The initial test showed that the series is stationary. Deterministic models performed better than stochastic models. The ARIMA model showed absolute errors of less than 2% on the logarithmic scale and relative errors 3.4-5.8 times less than the null model. The prediction of future cases of malaria in the horizons of 6 and 12 months in advance was possible. **Conclusion.** It is recommended the use of the ARIMA model to predict future scenarios and to anticipate planning in state health services in the Amazon Region.

Keywords: Time Series Studies; Malaria; Decision Support Techniques; Epidemiological Monitoring; Forecasting.

Introdução

A malária é uma doença evitável e curável; entretanto, no ano de 2018, ocorreram cerca de 228 milhões de casos da doença no mundo.^{1,2} A previsão oportuna de casos de malária permite a alocação de recursos adequados para o controle da doença e planejamento de sua eliminação.³ A iniciativa denominada ‘The Malaria Eradication Research Agenda’ analisou diversos aspectos importantes para a eliminação global da malária. A modelagem preditiva de casos foi sugerida como ferramenta de auxílio ao setor da Vigilância em Saúde, para o planejamento de ações de controle da infecção.^{4,5} Séries temporais mensais de novos casos podem ser estatisticamente ajustadas em funções matemáticas, utilizando-se de programas computacionais.⁶⁻⁸ A série temporal pode ser decomposta sob a ótica de três componentes básicas: (i) a componente sazonal, que representa o padrão cíclico da doença ao longo do tempo; (ii) a componente linear, entendida como uma tendência de aumento ou decréscimo linear da doença ao longo do tempo; e (iii) a componente estocástica, referente aos fatores intervenientes que afetam a série temporal sem padrão específico.

Modelos estatísticos de série temporal podem ser utilizados para a previsão de casos futuros. O modelo autorregressivo com média móvel integrada (ARIMA) é um método pioneiro para descrição e previsão de série temporal.⁹ O modelo de suavização exponencial (ETS) representa uma alternativa ao ARIMA.¹⁰ Modelos de suavização

exponencial para ajustar padrões sazonais complexos (TBATS e BATS) são métodos considerados mais eficientes que o ARIMA.¹¹ Outra alternativa encontra-se no modelo que divide a componente sazonal em subcomponentes (STLM).^{12,13}

Os modelos descritos anteriormente são determinísticos e possuem estrutura estatística para decompor e ajustar as componentes sazonal e linear de série temporal; porém, não conseguem estimar a componente estocástica. Nesse sentido, foram propostas abordagens computacionais de aprendizado de máquina, para quantificar o efeito desta terceira componente: modelo estrutural (StructTS), modelo de redes neurais (NNETAR) e modelos com aprendizado de máquina (ELM, MLP) são exemplos dessas abordagens.^{9,14-16} Os modelos podem ser comparados com um modelo nulo, definido pelo valor constante da última observação.¹⁷

A premissa deste trabalho é a de que a vigilância em saúde pode utilizar modelos preditivos de série temporal, para prever o impacto da malária em determinado estado do Brasil. A incidência de casos de malária no estado do Amapá entre 2015 e 2018 foi, em média, de 17 casos por 1000 habitantes, uma das maiores do país quando comparada às observadas para os estados do Acre (37/1000 hab.), Amazonas (17/1000 hab.) e Roraima (19/1000 hab.) no mesmo período. É importante identificar a série temporal futura de casos de malária para o planejamento de medidas de controle. O presente estudo teve por objetivo avaliar a capacidade preditiva de diferentes modelos de série temporal dos casos de malária no estado do Amapá.

Métodos

Aplicou-se uma abordagem estatística e computacional direcionada ao serviço de saúde em nível estadual, complementar às atividades de controle da malária no Brasil como um todo, acorde com os termos pactuados junto à Organização Mundial da Saúde (OMS) para a eliminação da doença.²

Trata-se de estudo ecológico de séries temporais, utilizando o número de casos de malária registrados no estado do Amapá no período de 1997 a 2016.

O Amapá constitui uma das mais importantes regiões endêmicas da malária no Brasil (Figura 1). Seu clima predominante, segundo a classificação climática de Köppen-

Geiger, é tropical de monção, ou seja, quente e bastante úmido com índice de pluviosidade médio de 3.300mm anuais. A maior parte do território do estado (73%; 97.000km²) é coberta por vegetação nativa. Em 2019, a população amapaense somava 830 mil hab., distribuídos entre 16 municípios.

Os dados de malária (casos autóctones, identificados por microscopia em lâmina positiva para plasmódio – decorrente da técnica de gota espessa) foram obtidos a partir dos seguintes sistemas de informações da Secretaria de Vigilância em Saúde do Ministério da Saúde (SVS/MS): Sistema de Informação do Programa Nacional de Controle da Malária (SISMAL), sobre o período 1997-2003; e Sistema de Informação de Vigilância Epidemiológica – Malária (SIVEP-Malária), sobre o período 2003-2016. Foram utilizadas três variáveis quantitativas discretas, transformadas em logaritmo natural:¹⁸

- a) número de casos mensais de malária, de janeiro de 1997 a dezembro de 2015 (a variável foi utilizada para ajustar os parâmetros dos modelos estatísticos);
- b) valores estimados, de janeiro a dezembro de 2016 (variável de previsão, foi utilizada no teste de capacidade preditiva dos modelos estatísticos); e
- c) número de casos mensais de malária, de janeiro a dezembro de 2016 (variável de teste, foi comparada com a variável de previsão).

A abordagem estatística utilizada baseou-se em modelos de séries temporais. O primeiro procedimento foi testar se a série temporal era ou não estacionária com o teste de Dikey-Fuller, aumentado ao nível de significância de 5%, sendo que a premissa de estacionariedade foi assumida no uso dos modelos estatísticos, descritos mais adiante.

A série temporal abrangeu 240 meses, de janeiro de 1997 a dezembro de 2016, dividida em dois períodos: período de treino, de janeiro de 1997 a dezembro de 2015; e período de teste, de janeiro a dezembro de 2016. O número de casos mensais de malária no período de treino foi utilizado para ajustar cada um dos modelos estatísticos e estimar os parâmetros das componentes temporais (sazonalidade; tendência linear; efeito estocástico). O número de casos mensais de malária no período de teste foi utilizado para se comparar aos valores estimados pelos modelos estatísticos. O teste foi feito com três horizontes de tempo de previsão: 12 meses de antecipação (janeiro a dezembro de 2016); seis meses de antecipação (julho a dezembro de 2016); e três meses de antecipação (outubro a dezembro de 2016). O resultado de cada teste foi utilizado para avaliar a capacidade preditiva dos modelos.

Os modelos estatísticos utilizados foram:

- a) modelos determinísticos (ARIMA;⁹ ETS;¹⁰ TBATS;¹¹ BATS;¹¹ STLM^{12,13});
- b) modelos estocásticos (StructTS;¹⁴ NNETAR;⁹ ELM;¹⁵ MLP¹⁶); e
- c) modelo nulo.¹⁷

Foram considerados três critérios para avaliação da capacidade preditiva dos modelos estatísticos:

- a) Erro percentual absoluto médio (MAPE), em que A_t é o valor real e F_t é o valor da previsão. Os valores são somados para cada ponto previsto no tempo, e o resultado dividido pelo número de pontos n . Multiplicar por 100% torna o MAPE um erro de percentagem, de acordo com a seguinte fórmula:

$$MAPE = \frac{100\%}{n} \sum_{t=1}^n \left| \frac{A_t - F_t}{A_t} \right|$$

- b) Escala de MAPE relativo, em que o MAPE do modelo nulo é dividido pelos valores de MAPE dos demais modelos. Se o resultado dessa divisão é menor ou igual a 1, o modelo é classificado como ruim; se maior que 1 e menor ou igual a 2, o modelo é classificado como de baixa capacidade de previsão; e se é maior do que 2, o modelo apresenta capacidade de previsão confiável.
- c) Coeficiente de incerteza (Theil's U). Trata-se de uma medida de precisão relativa, que visa penalizar modelos estatísticos com desvios elevados em relação ao valor médio. Valores abaixo de 1 significam que a capacidade preditiva é confiável.¹⁷

Uma capacidade preditiva considerada aceitável, portanto, foi definida por (i) menor valor de MAPE, (ii) valores superiores a 2 para a escala MAPE do modelo nulo e (iii) valores abaixo de 1 para Theil's U.

As análises foram realizadas no ambiente computacional R, mediante um *script* de programação corrigido e validado, possível de ser reproduzido pelas equipes de vigilância dos municípios e estados do país.¹⁹

Resultados

O total de número de casos de malária foi de 403.832, à média de 1.771 casos por mês e mediana de 1.518 casos/mês, com o primeiro quartil de 1.021 casos e o terceiro quartil de 2.079 casos – desvio-padrão (DP) = 918 casos. O valor do teste de Dikey-Fuller aumentado mostrou que a série temporal é estacionária (valor do teste = -5,352; p-valor <0,01). O número de casos mínimo foi de 487 (maio de 2014), enquanto o máximo foi de 5.944 (outubro de 2000), consistentemente com a natureza sazonal da malária no estado (Tabela 1).

Na Tabela 2, são apresentadas as performances dos modelos de acordo com os critérios de avaliação da capacidade preditiva e os três horizontes de tempo de previsão selecionados. Nos horizontes de 12 e seis meses de antecipação, todos os modelos determinísticos (ETS, ARIMA, STLM, BATS e TBATS) mostraram capacidade preditiva confiável. Entretanto, nenhum dos modelos estocásticos apresentou confiabilidade na previsão de casos futuros de malária nesses horizontes de antecipação. Surpreendentemente, o horizonte de três meses de antecipação se mostrou desafiador para todos os modelos. Afinal, nenhum modelo foi considerado confiável para a realização de previsão de casos de malária com três meses de antecipação (Figura 2).

Discussão

Os modelos determinísticos mostraram-se confiáveis para a previsão de número de casos mensais de malária nos próximos 12 ou seis meses de antecipação, no estado do Amapá. Um resultado que pode ser interpretado da seguinte forma: a série temporal em estudo possui características que possibilitaram o melhor desempenho dos modelos determinísticos, isto porque séries temporais com (i) forte componente sazonal, (ii) relativamente baixa tendência linear (i.e., estacionariedade temporal) e (iii) efeito estocástico pequeno ou nulo são ajustadas satisfatoriamente, por modelos determinísticos, a exemplo do ARIMA. Os modelos estatísticos com capacidade de detecção de efeitos estocásticos podem, teoricamente, apresentar melhor desempenho, comparados aos determinísticos, no caso de séries temporais com (i) ausência de sazonalidade, (ii) alta tendência linear e (iii) efeitos estocásticos presentes.²⁰ Entretanto, nenhum dos modelos avaliados apresentou resultados confiáveis para a antecipação de

número de casos mensais de malária no horizonte de três meses. A previsão de malária em períodos futuros curtos também resultou insatisfatória para outro estudo, no qual foram aplicados modelos determinísticos em distritos do Sri Lanka, entre 1972 e 2005.¹⁸ Aquele estudo desencorajou o uso de modelos estatísticos destinados à previsão casos de malária em período proximal futuro de um mês de antecedência, não obstante o prazo do estudo do Sri Lanka ser ainda mais curto que o menor prazo estimado aqui.

Uma limitação da abordagem do trabalho em tela encontra-se na impossibilidade de previsão de casos de malária em períodos curtos de antecedência, como o de 3 meses. Outra limitação reside no fato de, ao utilizar o estado do Amapá como unidade de análise, informações na escala local serem perdidas. As vantagens da presente abordagem, entretanto, estão na possibilidade de previsão de casos de malária com 12 ou seis meses de antecipação e no uso de modelos estatísticos determinísticos. Estes modelos, de melhor entendimento para os gestores, tornam mais viável a implementação dessa ferramenta nos serviços de saúde.²¹

O uso potencial de técnicas de séries temporais em estudos epidemiológicos, vigilância de doenças e previsão de surtos para malária têm sido explorado em diversos trabalhos.^{18,22-24} Por exemplo, um estudo com delineamento similar, realizado no norte da Tailândia no período de 1999 a 2004, encontrou que modelos determinísticos permitem previsão futura de malária e dengue com antecipação de um a quatro meses, a ponto de seus autores sugerirem o uso desses modelos para a alocação de recursos no controle e prevenção dessas doenças.²² Outro estudo com modelos determinísticos, realizado no Sudão, de 2009 a 2013, evidenciou que a capacidade preditiva dos modelos utilizados variou de acordo com cada estado (daquele país).²³ Modelos determinísticos utilizados para a previsão de malária no Afeganistão, entre 2005 e 2015, resultaram em previsões confiáveis para horizontes de 12 até quatro meses de antecipação.²⁴

Considerando-se que, entre os modelos determinísticos, o ARIMA se apresenta como o mais aplicado na literatura e possui característica de previsão para doenças afetadas pela sazonalidade,²²⁻²⁶ recomenda-se sua implementação como protocolo de previsão de casos mensais de malária em horizontes de longo prazo – 12 ou seis meses – na escala do estado da Amazônia brasileira.

Contribuição dos autores

Lima MVM contribuiu com a concepção e delineamento do artigo, análise e interpretação dos dados e redação da primeira versão do manuscrito. Laporta GZ contribuiu com a análise e interpretação dos dados e revisão crítica do manuscrito. Ambos autores aprovaram a versão final e são responsáveis por todos os aspectos do trabalho, incluindo a garantia de sua precisão e integridade.

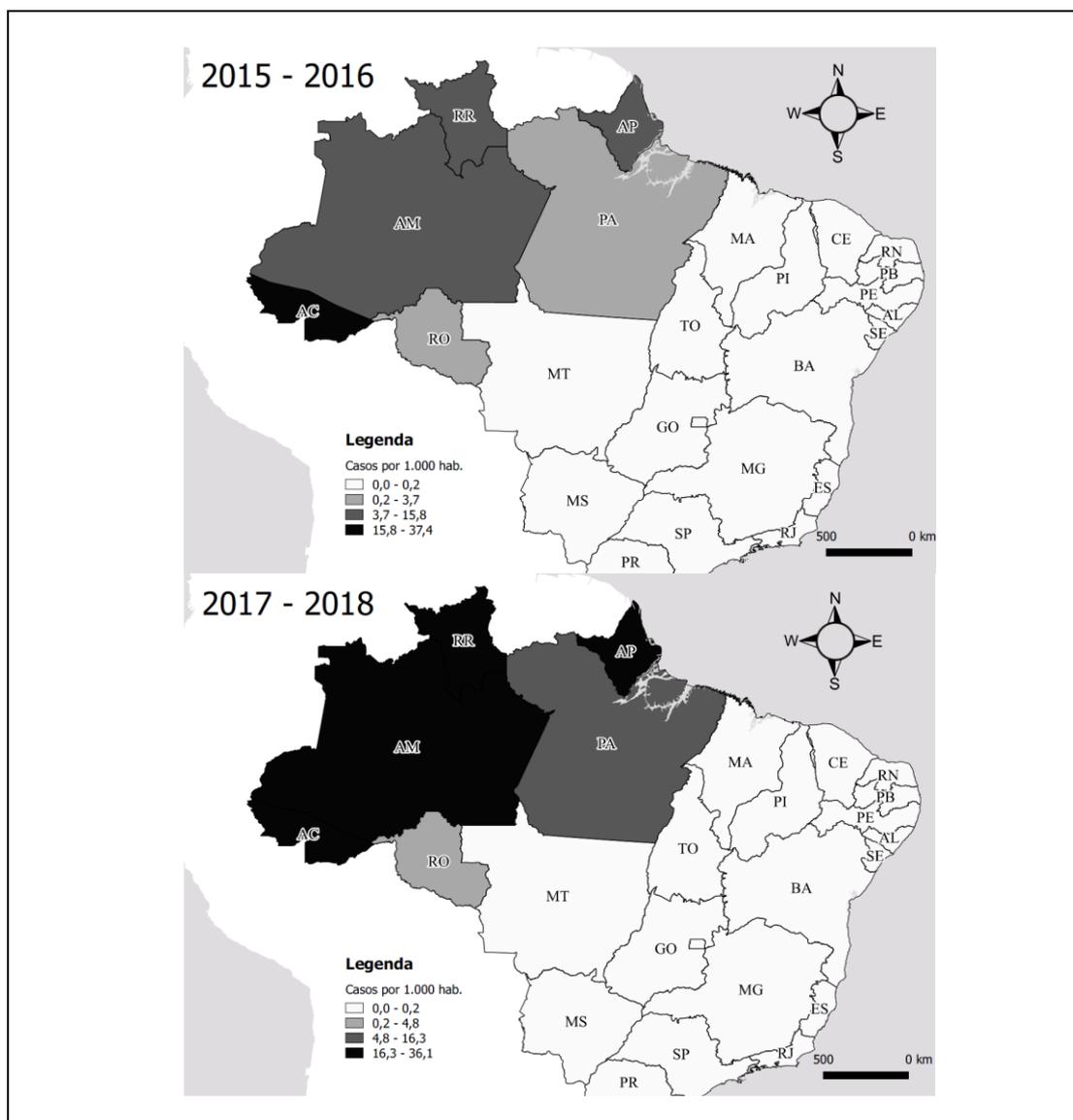
Referências

1. World Health Organization - WHO. World malaria report 2019 [Internet]. Geneva: World Health Organization; 2019 [cited 2020 Sep 28]. 232 p. Available from: <https://www.who.int/publications/i/item/world-malaria-report-2019>
2. Organização Mundial da Saúde - OMS. Estratégia técnica mundial para o paludismo 2016-2030 [Internet]. Geneva: Organização Mundial da Saúde; 2015 [citado 2020 set 28]. 30 p. Disponível em: https://apps.who.int/iris/bitstream/handle/10665/176712/9789248564994_por.pdf?sequence=6
3. Gomez-Elipse A, Otero A, Van Herp M, Aguirre-Jaime A. Forecasting malaria incidence based on monthly case reports and environmental factors in Karuzi, Burundi, 1997-2003. *Malar J* [Internet]. 2007 Sep [cited 2020 Sep 28];6(129):1-10. Available from: <https://doi.org/10.1186/1475-2875-6-129>
4. The malERA Consultative Group on Modeling. A research agenda for malaria eradication: modeling. *PLoS Med* [Internet]. 2011 Jan [cited 2020 Sep 28];8(1). Available from: <https://doi.org/10.1371/journal.pmed.1000403>
5. The malERA Consultative Group on Modeling. A research agenda for malaria eradication: monitoring, evaluation, and surveillance. *PLoS Med* [Internet]. 2011 Jan [cited 2020 Sep 28];8(1). Available from: <https://doi.org/10.1371/journal.pmed.1000400>
6. Hyndman RJ, Khandakar Y. Automatic time series forecasting: the forecast package for R. *J Stat Softw* [Internet]. 2008 Jul [cited 2020 Sep 28];27(3):C3. Available from: <http://dx.doi.org/10.18637/jss.v027.i03>
7. Hyndman JR, Anne KB, Ralph SD, Simone G. A state space framework for automatic forecasting using exponential smoothing methods. *Int J Forecast* [Internet]. 2002 Jul-Sep [cited 2020 Sep 28];18(3):439-54. Available from: [https://doi.org/10.1016/S0169-2070\(01\)00110-8](https://doi.org/10.1016/S0169-2070(01)00110-8)
8. Livera AM, Hyndman RJ, Snyder RD. Forecasting time series with complex seasonal patterns using exponential smoothing. *J Am Stat Assoc* [Internet]. 2011 Jan [cited 2020 Sep 28];106(496):1513-27. Available from: <https://doi.org/10.1198/jasa.2011.tm09771>
9. Dokumentov A, Hyndman RJ. STR: a seasonal-trend decomposition procedure based on regression. Monash Econometrics and Business Statistics Working Papers [Internet]. 2015 [cited 2020 Sep 28]. Available from: <https://robjhyndman.com/publications/str/>

10. Cleveland R, Cleveland W, McRae J, Terpenning I. STL: a seasonal-trend decomposition procedure based on Loess (with discussion). *J Off Stat* [Internet]. 1990;6(1):3-73. Available from: <https://www.wessa.net/download/stl.pdf>
11. Durbin J, Koopman SJ. *Time series analysis by state space methods*. Oxford University Press; 2013.
12. Crone SF, Kourentzes N. Feature selection for time series prediction - A combined filter and wrapper approach for neural networks. *Neurocomputing* [Internet]. 2010 Jun [cited 2020 Sep 28];73(10-12):1923-36. Available from: <https://doi.org/10.1016/j.neucom.2010.01.017>
13. Kourentzes N, Barrow DK, Crone SF. Neural network ensemble operators for time series forecasting. *Expert Syst Appl* [Internet]. 2014 Jul [cited 2020 Sep 28];41(9):4235-44. Available from: <https://doi.org/10.1016/j.eswa.2013.12.011>
14. Hyndman RJ, Athanasopoulos G. *Forecasting: principles and practice*. 2nd ed. [S.l.]: OTexts; 2018.
15. Morettin PA, Toloi CM. *Análise de séries temporais: modelos lineares univariados*. 3rd ed. São Paulo: Blucher; 2018.
16. Antunes JLF, Cardoso MRA. Uso da análise de séries temporais em estudos epidemiológicos. *Epidemiol Serv Saúde* [Internet]. 2015 jul-set [cited 2020 set 28];24(3):565-76. Disponível em: <http://dx.doi.org/10.5123/S1679-49742015000300024>
17. Helfenstein U. Box-jenkins modelling of some viral infectious diseases. *Stat Med* [Internet]. 1986 Jan-Feb [cited 2020 Sep 28];5(1):37-47. Available from: <https://doi.org/10.1002/sim.4780050107>
18. Briët OJT, Vounatsou P, Gunawardena DM, Galappaththy GNL, Amerasinghe PH. Models for short term malaria prediction in Sri Lanka. *Malar J* [Internet]. 2008 May [cited 2020 Sep 28];7(76). Available from: https://github.com/MVMLima/artigo_RESS_malaria
19. Lima MVM. *Protocolo de programação em ambiente R e banco de dados disponíveis em repositório de acesso público* [Internet]. [S.l.]: GitHub; 2020 [citado 2020 ago 5]. Disponível em: https://github.com/MVMLima/artigo_RESS_malaria
20. Baquero OS, Santana LMR, Chiaravalloti-Neto F. Dengue forecasting in São Paulo city with generalized additive models, artificial neural networks and seasonal autoregressive integrated moving average models. *PLoS One* [Internet]. 2018 Apr [cited 2020 Sep 28];13(4):1-12. Available from: <https://doi.org/10.1371/journal.pone.0195065>
21. World Health Organization - WHO. *Malaria epidemics: forecasting, prevention, early detection and control*. Lesyin: World Health Organization; 2004 [cited 2020 Sep 28]. 48 p. Available from: <https://apps.who.int/iris/handle/10665/70073>
22. Sriwattanapongse W, Khanabsakdi S. Modeling and forecasting malaria and dengue hemorrhagic fever incidence and prevalence in Northern Thailand. *J Math Syst Sci* [Internet]. 2011 [cited 2020 Sep 28];1(1):52-9. Available from: <https://www.semanticscholar.org/paper/Modeling-and-Forecasting-Malaria-and-Dengue-Fever-Sriwattanapongse-Khanabsakdi/b5de80ebdda6ce67bd6e5cdf666ea2270e98d>
23. Hussien HH, Eissa FH, Awadalla KE. *Statistical methods for predicting malaria*

- incidences using data from Sudan. *Malar Res Treat* [Internet]. 2017 [cited 2020 Sep 28];4205957. Available from: <https://doi.org/10.1155/2017/4205957>
24. Anwar MY, Lewnard JA, Parikh S, Pitzer VE. Time series analysis of malaria in Afghanistan: using ARIMA models to predict future trends in incidence. *Malar J* [Internet]. 2016 Nov [cited 2020 Sep 28];15(1):566. Available from: <https://doi.org/10.1186/s12936-016-1602-1>
 25. Tohidinik HR, Mohebbali M, Mansournia MA, Kalhori SRN, Akbarpour MA, Yazdani K. Forecasting zoonotic cutaneous leishmaniasis using meteorological factors in eastern fars province, Iran: a sarima analysis. *Trop Med Int Health* [Internet]. 2018 Aug [cited 2020 Sep 28];23(8):860-9. Available from: <https://doi.org/10.1111/tmi.13079>
 26. Padilha MAO, Melo JDO, Romano G, Lima MVM, Alonso WJ, Sallum MAM, et al. Comparison of malaria incidence rates and socioeconomic environmental factors between the states of Acre and Rondônia: a spatio temporal modelling study. *Malar J* [Internet]. 2019 Sep [cited 2020 Sep 28];18(306). Available from: <https://doi.org/10.1186/s12936-019-2938-0>

Tabelas e Figuras



AC: Acre.
 AM: Amazonas.
 AP: Amapá.
 PA: Pará.
 RO: Rondônia.
 RR: Roraima.

Figura 1 – Incidência média (por 1000 hab.) de casos confirmados de malária em estados do Brasil

Tabela 1 – Número de casos de malária confirmados por mês e ano, Amapá, janeiro de 1997 – dezembro de 2016

Ano	Janeiro	Fevereiro	Março	Abril	Mai	Junho	Julho	Agosto	Setembro	Outubro	Novembro	Dezembro	Total
1997	1.997	1.520	1.023	1.096	1.030	1.364	2.108	2.708	3.506	3.592	2.814	2.110	24.868
1998	1.633	1.863	1.566	1.524	1.773	1.899	2.005	2.534	3.001	1.795	1.929	1.047	22.569
1999	1.077	1.124	994	661	809	801	1.594	3.472	4.435	4.963	5.570	3.146	28.646
2000	2.473	2.731	1.632	1.269	1.563	1.714	2.377	4.336	3.767	5.944	4.574	2.898	35.278
2001	2.861	2.510	2.019	1.513	1.638	1.760	1.920	2.496	2.466	2.588	1.731	985	24.487
2002	1.422	1.250	849	671	593	649	971	1.953	2.330	2.174	2.144	1.251	16.257
2003	963	854	704	690	755	953	1.622	1.601	2.118	2.456	2.258	1.677	16.651
2004	1.949	1.972	1.453	987	1.014	1.126	1.408	1.954	1.848	2.731	2.387	1.841	20.670
2005	1.872	2.524	2.214	1.331	1.517	1.546	2.252	3.133	3.416	3.615	2.681	1.958	28.059
2006	2.505	1.500	1.231	1.101	1.582	1.637	2.403	3.393	3.431	4.371	3.750	2.386	29.290
2007	2.527	1.743	1.560	1.314	1.254	1.241	1.801	2.147	2.036	3.119	1.914	1.319	21.975
2008	1.217	957	823	770	872	745	960	1.121	1.721	2.099	2.144	1.702	15.131
2009	1.558	1.133	955	860	1.049	1.036	1.096	1.501	1.816	1.783	1.541	1.176	15.504
2010	1.466	1.143	994	730	892	970	1.384	1.390	1.292	1.553	2.072	1.502	15.388
2011	1.119	872	830	723	923	941	1.348	2.254	2.324	2.639	3.115	1.910	18.998
2012	1.685	1.257	1.040	693	863	917	1.272	1.344	1.234	1.619	1.925	1.432	15.281
2013	1.675	1.401	939	842	710	717	918	1.272	1.445	2.121	2.004	1.250	15.294
2014	1.057	806	599	516	487	516	646	1.201	1.844	2.163	2.199	1.521	13.555

2015	1.504	1.097	767	597	558	635	971	1.234	1.705	1.906	1.606	1.078	13.658
2016	1.138	1.040	667	532	580	673	962	1.296	1.585	1.554	1.321	925	12.273

Tabela 2 – Comparação de desempenho de modelos estatísticos em acordo com os critérios de avaliação de modelos e os tempos de previsão de cada horizonte de antecipação para os casos mensais de malária do Amapá

Tempo de previsão no horizonte de antecipação de 12 meses				Tempo de previsão no horizonte de antecipação de 6 meses				Tempo de previsão no horizonte de antecipação de 3 meses			
Modelos	MAPE ^a	Escala de MAPE ^a relativo	Theil's U	Modelos	MAPE ^a	Escala de MAPE ^a relativo	Theil's U	Modelos	MAPE ^a	Escala de MAPE ^a relativo	Theil's U
ETS	1,25	3,76	0,40	ETS	0,76	11,28	0,29	NNETAR	2,47	1,44	0,64
ARIMA	1,39	3,38	0,47	ARIMA	1,47	5,83	0,56	ETS	2,87	1,24	0,82
TBATS	1,57	2,99	0,52	TBATS	1,60	5,36	0,58	TBATS	3,28	1,09	0,91
ELM	1,64	2,87	0,50	STLM	1,99	4,31	0,73	Nulo	3,56	1,00	1,45
STLM	1,63	2,88	0,57	NNETAR	2,16	3,97	0,76	STLM	3,96	0,90	1,17
MLP	1,66	2,83	0,55	BATS	2,24	3,83	0,93	ARIMA	4,36	0,82	1,27
BATS	1,81	2,60	0,68	StructTS	3,57	2,40	1,29	BATS	4,67	0,76	1,34
NNETAR	3,29	1,43	1,01	MLP	5,24	1,64	1,96	MLP	5,74	0,62	2,08
Nulo	4,70	1,00	1,63	ELM	6,97	1,23	2,53	ELM	5,82	0,61	2,20
StructTS	10,15	0,46	3,75	Nulo	8,57	1,00	2,88	StructTS	10,44	0,34	3,43

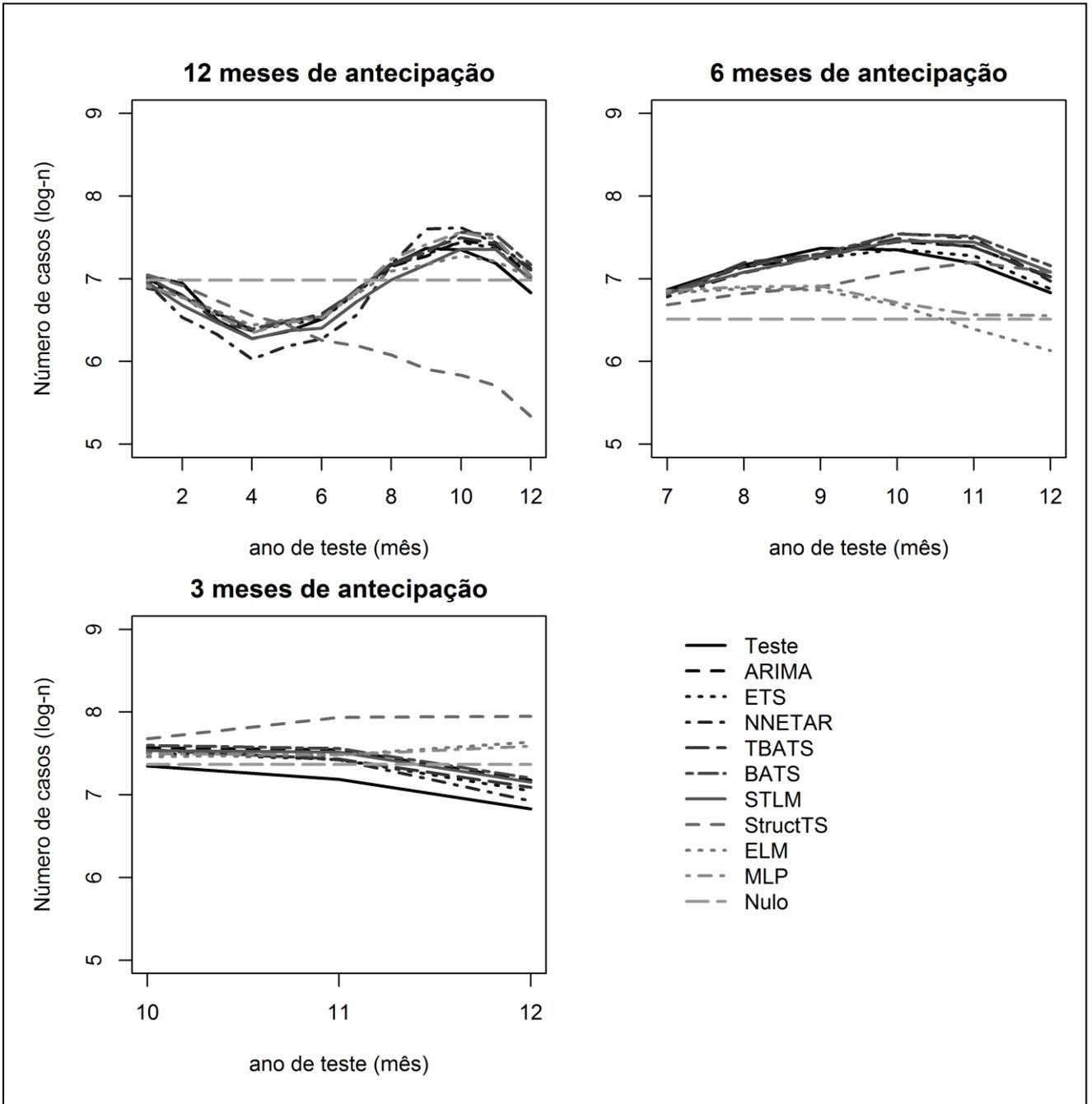
a) MAPE: erro percentual absoluto médio.

Notas:

ARIMA, ETS, TBATS, BATS e STLM: modelos determinísticos.

StructTS, NNETAR, ELM e MLP: modelos estocásticos.

Nulo: modelo nulo.



Notas:

ARIMA, ETS, TBATS, BATS e STLM: modelos determinísticos.

StructTS, NNETAR, ELM e MLP: modelos estocásticos.

Nulo: modelo nulo.

Figura 2 – Capacidade preditiva dos modelos estatísticos considerando-se a comparação entre a variável de teste e os valores estimados para cada modelo, segundo três horizontes de antecipação

**Material Suplementar - O desenvolvimento do *script* de Programação em
Ambiente R seguiu estes passos:**

Passo 1: Instalar os pacotes necessários para a análise e previsão das séries temporais

```
Install.packages("forecast", "TSA", "tseries", "seasonal", "fma", "nnfor")

#Carregamento dos pacotes
library(forecast)
library(TSA)
library(tseries)
library(fma)
library(nnfor)
```

Passo 2: Abrir o banco de dados e determinar o diretório de trabalho

```
setwd("C:/Users/Desktop/Artigo Protocolo")
load("C:/Users/Desktop/Artigo Protocolo/Amapá.Rdata")
```

Passo 3: Criar a série temporal para o horizonte de previsão de 12 meses

Para executar os modelos de previsão em 'R', é preciso converter os dados em um objeto de série temporal, procedimento feito na primeira linha de código abaixo. Os argumentos 'start' e 'end' especificam o tempo da primeira e o tempo da última observação, respectivamente. O argumento 'frequency' especifica o número de observações por unidade de tempo.

```
Set.seed(123)
x.h12 <- ts(malaria$AP, start = c(1997,1), frequency = 12)
x.h12 = log(x.h12)
test_x.h12 <- window(x.h12, start=c(2016,1))
x.h12 <- window(x.h12, end=c(2015,12))
```

Passo 4: Fazer a modelagem e previsão conforme a função a seguir

```
models.h12 <- list(mod_arima = auto.arima(x.h12, ic='aicc', stepwise=FALSE),
  mod_exp = ets(x.h12, ic='aicc', restrict=FALSE),
  mod_neural = nnetar(x.h12, p=12, size=25),
  mod_tbats = tbats(x.h12, ic='aicc', seasonal.periods=12),
  mod_bats = bats(x.h12, ic='aicc', seasonal.periods=12),
  mod_stl = stlm(x.h12, s.window=12, ic='aicc', robust=TRUE, method='ets'),
  mod_sts = StructTS(x.h12),
  mod_elm = elm(x.h12),
  mod_mlp = mlp(x.h12)
)
```

Passo 5: Fazer a previsão e apresentar os melhores modelos

```
acc.h12 <- lapply(forecasts.h12, function(f){
  accuracy(f, test_x.h12)[2,drop=FALSE]
})
acc.h12 <- Reduce(rbind, acc.h12)
row.names(acc.h12) <- names(forecasts.h12)
acc.h12 <- acc.h12[order(acc.h12[, 'MASE']),]
round(acc.h12, 2)
```

Passo 6: Repetir os passos 4 e 5 para o horizonte de seis e três meses

Passo 7: Fazer os gráficos para as previsões realizadas e repetir para os horizontes de previsão de seis e três meses

```

s.realMT <- log(malária$AP)

#h12 plot()

forecast_h12_arima <- c(forecasts.h12$mod_arima$fitted,
forecasts.h12$mod_arima$mean)
forecast_h12_exp <- c(forecasts.h12$mod_exp$fitted, forecasts.h12$mod_exp$mean)
forecast_h12_neural <- c(forecasts.h12$mod_neural$fitted,
forecasts.h12$mod_neural$mean)
forecast_h12_t bats <- c(forecasts.h12$mod_t bats$fitted,
forecasts.h12$mod_t bats$mean)
forecast_h12_bats <- c(forecasts.h12$mod_bats$fitted,
forecasts.h12$mod_bats$mean)
forecast_h12_stl <- c(forecasts.h12$mod_stl$fitted, forecasts.h12$mod_stl$mean)
forecast_h12_sts <- c(forecasts.h12$mod_sts$fitted, forecasts.h12$mod_sts$mean)
forecast_h12_elm <- c(forecasts.h12$mod_elm$fitted,
forecasts.h12$mod_elm$mean)
forecast_h12_mlp <- c(forecasts.h12$mod_mlp$fitted,
forecasts.h12$mod_mlp$mean)
forecast_h12_naive <- c(forecasts.h12$naive$fitted, forecasts.h12$naive$mean)

png("Fig_AP_malaria.png", units="mm", height=170, width=170, res=600)
par(mar=c(4,4,2.2,0.1))
par(mfrow=c(2,2))

st <- 1:12

plot(st, s.realMT[229:240], type = "l", lwd = 2, ylim = c(4,10), ylab = "Número de
casos (log-n)", xlab = "2017 (mês)", main="A")
points(st,forecast_h12_arima[229:240], type = "l", lwd = 2, col = 2, lty = "dashed")
points(st,forecast_h12_exp[229:240], type = "l", lwd = 2, col = 3, lty = 3)
points(st,forecast_h12_neural[229:240], type = "l", lwd = 2, col = 4, lty = 4)
points(st,forecast_h12_t bats[229:240], type = "l", lwd = 2, col = 5, lty = 5)
points(st,forecast_h12_bats[229:240], type = "l", lwd = 2, col = 6, lty = 6)
points(st,forecast_h12_stl[229:240], type = "l", lwd = 2, col = 7, lty = 7)
points(st,forecast_h12_sts[229:240], type = "l", lwd = 2, col = 8, lty = 8)
points(st,forecast_h12_elm[224:235], type = "l", lwd = 2, col = 9, lty = 9)
points(st,forecast_h12_mlp[224:235], type = "l", lwd = 2, col = 10, lty = 10)
points(st,forecast_h12_naive[229:240], type = "l", lwd = 2, col = 11, lty = 11)

#Legenda
plot(229:240, s.realMT[229:240],type='n',axes=F,ann=F)
legend(229,7, legend=c("Real", "Arima", "Exponencial", "Neural", "TBATS",
"BATS", "STL", "STS", "ELM", "MLP", "Naive"),
lty=c(1,2,3,4,5,6,7,8,9,10,11),col=c(1,2,3,4,5,6,7,8,9,10,11), lwd=2, bty="n")

dev.off()

```