

STATISTICAL PROCESS CONTROL AS A TOOL TO CONTROL AND PREVENT MALARIA EPIDEMICS IN THE LEGAL AMAZON REGION

José Guilherme Santana de SENA¹
Paulo Henrique FERREIRA¹
Rosemeire Leovigildo FIACCONE¹

- **ABSTRACT:** Malaria is still a fatal disease in many countries around the world. Establishing measures to control and prevent this disease has been a worldwide concern since 1950, when the World Health Organization launched a Malaria Eradication Plan. In Brazil, malaria was eliminated in much of the territory, but resisting in an area known as the Legal Amazon. That said, the main scope of this work is to develop statistical control charts that consider the temporal dependency structure in the data and are suitable for the current and future monitoring of malaria cases, in order to detect possible outbreaks or epidemics in states in the Legal Amazon region. The tools presented here could also collaborate in directing control actions and combating the spread of the disease. In particular, we intend to: (i) build a statistical model to predict the occurrence of cases of the disease, which considers the existence of a possible temporal dependency structure between the collected data; (ii) use the Statistical Process Control (SPC) techniques, notably the control charts, to monitor (separately for each form of the disease) cases of malaria of the types *Plasmodium Vivax*, *Plasmodium Falciparum* and *Plasmodium Mista* in the Amazon region; (iii) establish epidemic thresholds based on the obtained control charts. Considering data from 2013 to 2017, this study revealed, among others, that the generalized autoregressive and moving average models with Negative Binomial distribution (Negative Binomial GARMA models) were more efficient, fitting better, compared to Poisson GARMA models, due to the overdispersion existing in the analyzed data.
- **KEYWORDS:** Control chart; GARMA models; statistical monitoring; forecast; overdispersion.

1 Introduction

Malaria is a disease that has attracted global concern for a long time, and is considered one of the greatest causes of fatality in many countries situated in tropical regions, due to the high number of people who are exposed to the risk of infection. In

¹ Universidade Federal da Bahia – UFBA, Departamento de Estatística, CEP: 40.170-110, Salvador, BA, Brasil.
Email: santana.guilherme@outlook.com; paulohenri@ufba.br; fiaccone@ufba.br

Brazil, the majority of malaria cases are in the Amazon region, which comprises the states of Acre, Amazonas, Amapá, Maranhão Mato Grosso, Pará, Rondônia, Roraima and Tocantins, located in an area known as the Legal Amazon (or “Amazônia Legal” in Portuguese) (LAPOUBLE *et al.*, 2015). With the aim of increasing control of the disease, and to support the states and municipalities in this region, the Epidemiological Surveillance Information System (Sivep-Malária), a Ministry of Health (MS) and Department of Health Surveillance (SVS) initiative, was created in 2003, to receive real time notifications.

According to Braz (2005), the greatest challenge in controlling malaria in this region is defining the epidemic threshold of the disease, in order to identify the transmission limits expected for every municipality and/or state, and when they may be exceeded, behaving as an epidemic situation. Answering these questions is essential for the surveillance of malaria. Control charts are the main tools used in Statistical Process Control (SPC) and could play an important role in detecting and, principally, preventing diseases, since the epidemiological situation of a specific deterioration in a certain location, for example, can be monitored through control limits. When correctly applied, control charts are highly efficient in detecting outbreaks, indicating the best way to establish intervention measures, although its application needs to become established among health professionals.

There are studies in literature in which control charts were used to monitor diseases and other health-related aspects. For example, Schilling (1944) used control charts to monitor the number of visits made by San Francisco Nursing Association nurses. Sellick (1993) used SPC (control charts *u*, *c* and *p*) techniques to monitor attribute data, in order to efficiently monitor the occurrences of resistant pathogens in hospital infections. Benneyan (1998) held a discussion on the importance of applying SPC techniques in health contexts, describing some of the main tools, which are control charts *c*, *p* and *np* (SHEWHART, 1931). He confirmed that maintaining a process under statistical control may reduce the nosocomial rate (rate of infections contracted with a patient entering or leaving a hospital). Burns *et al.* (2005) conducted a study on the occupancy of beds at the University of Queensland Hospital, located in Australia, in which they demonstrated that a high occupancy rate indicated an overload for local activities, ranging from the preparation of beds to administrative tasks; and a low rate indicated an idle system, producing the same expenses. With the use of a cumulative sum (CuSum) control chart, patient flow trends could be identified, which resulted in increased efficiency in administering beds, reducing the number of occupancies. Pafiadache *et al.* (2015) made a comparison between the X-MR control charts (for individual measurements and moving ranges, respectively) and the Exponentially Weighted Moving Average (EWMA), for the early detection of monthly cases of hepatitis C notifications in the state of Rio Grande do Sul, with the EWMA being more efficient in identifying periods with a higher number of cases. Zanini *et al.* (2016) conducted a study where the main objective was to describe a number of quality control charts (\bar{X} , *R*, CuSum and EWMA) to follow-up surveillance processes, using indicators obtained at the Hospital Universitário de Santa Maria (Santa Maria University Hospital), in Rio Grande do Sul.

Specifically related to the application to malaria data, we cite the work of Cullen *et al.* (1984), in which they proposed the use of control charts, with the aim of creating an early epidemiological warning system for malaria outbreaks in two regions in the north of Thailand between 1973 and 1981. Four techniques were considered to estimate the

process average (overall, moving, accumulated, and monthly averages) in the control chart construction, based on years without episodes of the epidemic. Braz (2005) conducted a study on the early detection of malaria epidemics in the Legal Amazon region, in which he compared five control charts, constructed according to the Cullen (or Shewhart), Albuquerque, 3rd quartile, tabular CuSum, and Stern & Lightfoot methods, applied to the data obtained from two institutions: the National Malaria Control Program Information System and Sivep-Malaria, between 1996 and 2003. Following an analysis of the results, the author concluded that the 3rd quartile method presented higher rates of real alarms, therefore being the most adequate for the early detection of malaria epidemics in Brazilian municipalities of the Legal Amazon region.

The main objective of this research was to develop control charts that consider the time dependence structure existent in the data, and are adequate for current and future monitoring of *Plasmodium Vivax*, *Plasmodium Falciparum* and *Plasmodium Mista* malaria cases, in order to detect possible outbreaks or epidemics in states in the Legal Amazon region. The statistical tools presented here may also assist with directing control actions and contribute towards combating the spread of the disease. The specific objectives are as follows: (i) to construct a statistical model to predict occurrences of cases of the disease, which considers the existence of a possible time dependence structure in the data collected; (ii) to use SPC techniques, particularly, control charts, to monitor (separately for each form of the disease) the three types of malaria cases mentioned above, in the Amazon region; and (iii) to establish epidemic thresholds, based on the control charts obtained.

The remainder of this article is divided into three sections. The materials and methods used to develop this research are described in Section 2. The discussions and results obtained with the application of the methods cited in Section 2 to a real data set are presented in Section 3. Lastly, the conclusions and prospects for future research are presented in Section 4.

2 Materials and methods

The data (Section 2.1) and main statistical procedures (Sections 2.2 and 2.3) used to develop this research are briefly presented in this section.

2.1 Malaria data

In order to achieve the objectives described in Section 1, a real data set was used, comprising information obtained from two distinct databases. The first contains information on malaria taken from Sivep-Malaria (http://200.214.130.44/sivep_malaria/), considering the period 2013 to 2017. This set is made up of 102,932 observations for the municipalities that form the Legal Amazon region and six variables, which are presented in Table 1.

Table 1 – Variables extracted from Sivep-Malaria

| Variable | Description |
|-------------------------|--|
| <i>codibgeinfeccao</i> | Municipal code of the area where the notification was registered |
| <i>anoinfec</i> | Year of notification |
| <i>semepid</i> | Epidemiological week in which the notification occurred |
| <i>ncasosvivax</i> | Number of cases of occurrence by <i>Plasmodium Vivax</i> |
| <i>ncasosfalciparum</i> | Number of cases of occurrence by <i>Plasmodium Falciparum</i> |
| <i>ncasosmista</i> | Number of cases of occurrence by <i>Mixed Plasmodium</i> |

The second database contains climatic information taken from the National Oceanic and Atmospheric Administration (NOAA) - NCEP/NCAR Reanalysis 1 (<https://www.cpc.ncep.noaa.gov/>), which were used as predictors in the modeling process, since, according to Ferreira (2015), on studying the etiology of vectors of many diseases considered “tropical”, it is noted that high temperatures, high humidity, the length of the summer season, or heat and humidity conditions, among other factors, favor the proliferation of mosquitos which transmit these diseases. The database includes 1,825 observations related to daily collections during the period 2013 to 2017, and three variables, which are presented in Table 2.

We used R software, version 3.5.1 (R CORE TEAM, 2021), to analyze the data.

Table 2 – Variables extracted from NOAA

| Variable | Description |
|----------------|--|
| <i>precip</i> | Average rainfall (in millimeters, mm) |
| <i>tempair</i> | Average air temperature (in Celsius degrees, °C) |
| <i>rhum</i> | Average humidity (in percentage, %) |

2.2 GARMA models

Benjamin *et al.* (2003) describe the class of GARMA models, as an extension of Gaussian autoregressive moving average (ARMA) models, for a more flexible class, indicated for time series with count data and non-Gaussian distributions (such as Poisson and Negative Binomial, for example), in which the response variable has a conditional distribution in the exponential family, given the past of the process.

As an illustration, consider Y_t , with $t = 1, 2, \dots, n$, as the count process response variable, so that the distribution of Y_t belongs to the exponential family. Given the past of the process, the Y_t conditional model could be written in the following way:

$$f(y_t | F_{(t-1)}) = \exp \left\{ \frac{y_t \theta_t - b(\theta_t)}{\phi} + c(y_t, \phi) \right\}, \quad (1)$$

where t represents time; $F_{(t-1)} = (y^{(t-1)}, \mathbf{x}^{(t)})$, with $y^{(t-1)}$ representing the past of the count process, $\mathbf{x}^{(t)}$ a p -dimensional vector of exogenous covariates which were observed at the time t ; θ_t is the canonical parameter; ϕ is the dispersion or scale parameter; $b(\cdot)$ and $c(\cdot)$ are known functions, which vary in accordance with the distribution of probability associated with Y_t .

In addition to the random component, represented by Y_t , there are two essential components in the GARMA model composition: the systematic component and the link function. The systematic component (or linear predictor) is the linear combination of the predictor variables with their respective effects, being expressed by:

$$\eta_t = \beta_0 + \sum_{j=1}^p \beta_j x_j + \tau_t, \quad \text{or also,} \quad \eta_t = \mathbf{x}^T \boldsymbol{\beta} + \tau_t,$$

where $\mathbf{x} = (1, x_1, x_2, \dots, x_p)^T$ is the vector of explanatory variables; in other words, every x_j represents the predictor variable of the model, with $j = 1, 2, \dots, p$; $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_p)^T$ denotes the vector of model coefficients; details of τ_t will be demonstrated below.

In order to relate the linear predictor to the average of Y_t , the link function is used, which is monotonous and distinguishable, and denoted by $g(\cdot)$. Thus,

$$g(\mu_t) = \eta_t = \mathbf{x}^T \boldsymbol{\beta} + \tau_t, \quad (2)$$

where $\tau_t = \sum_{j=1}^p \lambda_j [g(y_{t-j}) - \mathbf{x}_{t-j}^T \boldsymbol{\beta}] + \sum_{r=1}^q \gamma_r [g(y_{t-r}) - \eta_{t-r}]$ represents the autoregressive moving average term.

The estimation of the GARMA model is carried out using the classical method of maximum likelihood; for further details, consult Benjamin *et al.* (2003). The parameters p and q are selected by an analysis of autocorrelation and partial autocorrelation functions, in addition to the Akaike information criterion (AIC) (AKAIKE, 1977), or Bayesian information criterion (BIC) (SCHWARZ, 1978). The `tsglm` package function `tsglm()` (LIBOSCHIK *et al.*, 2017) may be used in R software to adjust the GARMA models.

In the GARMA models, the forecasts h steps ahead (in other words, \hat{y}_{t+h}) are obtained through recursive forecasting by the linear predictor of each model (ANDRADE, 2016), in which non-observed values are replaced by their respective one-step ahead forecasts.

The ideal one-step ahead predictor, \hat{y}_{t+1} for y_{t+1} , is obtained considering the past of the process until moment t , and possible covariates collected at moment $t+1$. Therefore, the conditional expected value, $\hat{y}_{t+1} = E[y_{t+1} | F_{t+1}]$, is obtained through Equation (1). The `predict()` function is used in R software.

2.2.1 Poisson GARMA

A specific case in GARMA models is when we associate Poisson distribution to the random component, which is very common when the variable to be investigated is a count that has an average equal, or very close, to the variance.

In other words, $Y_t | F_{(t-1)}$ is a variable with Poisson distribution, with a μ_t average. As demonstrated in Equation (1), its probability function can be expressed as:

$$f(y_t | F_{(t-1)}) = \exp\{y_t \log(\mu_t) - \mu_t - \log(y_t!)\},$$

where $\phi = 1$, $\theta_t = \log(\mu_t)$, $b(\theta_t) = \exp\{\theta_t\}$ e $c(y_t, \phi) = -\log(y_t!)$. Using the canonical link function, i.e., the logarithmic function, the linear predictor may be represented, as demonstrated in Equation (2), by:

$$\log(\mu_t) = \alpha_0 + \sum_{j=1}^p \lambda_j [\log(y_{t-j}^*)] + \sum_{r=1}^q \gamma_r [\log(y_{t-r}^*) - \log(\mu_{t-r})],$$

where $y_{t-j}^* = \max\{y_{t-j}, c\}$, $0 < c < 1$.

2.2.2 Negative Binomial GARMA

Another specific case of GARMA models takes place when we associate the Negative Binomial (NB) distribution with the random component, which is indicated when the variable to be investigated is an overdispersed count, i.e., with a higher variance than mean.

Let $Y_t | F_{(t-1)}$ be a variable with $NB(k, \mu_t)$ distribution. As demonstrated in Equation (1), the probability function may be expressed as:

$$f(y_t | F_{(t-1)}) = \exp\left\{y_t \log\left(\frac{\mu_t}{\mu_t + k}\right) + k \log\left(\frac{k}{\mu_t + k}\right) + \log\left(\frac{\Gamma(k + y_t)}{\Gamma(1 + y_t)\Gamma(k)}\right)\right\}$$

where $\phi = 1$, $\theta_t = \log\left(\frac{\mu_t}{\mu_t + k}\right)$, $b(\theta_t) = -k \log\left(\frac{k}{\mu_t + k}\right)$ e $c(y_t, \phi) = \log\left(\frac{\Gamma(k + y_t)}{\Gamma(1 + y_t)\Gamma(k)}\right)$. In addition, $\Gamma(s) = \int_0^\infty x^{s-1} e^{-x} dx$ represents the gamma function.

Using the logarithmic link function (canonical), the linear predictor shown in Equation (2), may be represented by:

$$\log\left(\frac{\mu_t}{\mu_t + k}\right) = \alpha_0 + \sum_{j=1}^p \lambda_j [\log(y_{t-j}^*)] + \sum_{r=1}^q \gamma_r [\log(y_{t-r}^*) - \log(\mu_{t-r})],$$

where $y_{t-j}^* = \max\{y_{t-j}, c\}$, $0 < c < 1$.

2.2.3 Multicollinearity

Multicollinearity is a term used in literature to indicate the existence of a correlation between two (or more) predictor variables inserted in a multiple regression model. In many cases, it becomes a problem since it indicates a violation of the assumption of independence between these variables, thereby impacting the parameter estimates. Thus, it is suggested that only the covariates that present a variance inflation factor (VIF) (BELSLEY *et al.*, 1980) of less than 10 are maintained in the analysis (AKINWANDE *et al.*, 2015). The VIF is calculated as shown in Equation (3):

$$VIF_j = \frac{1}{1-R_j^2}, \quad j = 1, 2, \dots, p, \quad (3)$$

where R_j^2 is the coefficient of determination of the model adjusted for the j -th variable, i.e., it represents the proportion of the variability in data (response variable) which is explained by this model.

2.3 Residual control charts and the forecast of new process observations

Two control chart proposals are considered in this research. The first is the control chart for autocorrelated data (MONTGOMERY, 2019), applied to the residuals of the adjusted time series model, which aims to obtain/calculate the control limits and then identify any possible observations outside these thresholds (phase 1 of the analysis). The second proposal involves use of the control chart to monitor any new observations of the response of interest, i.e., with regards to monitoring future process observations (phase 2 of the analysis), which will enable any possible outbreaks or epidemics in the context of application to malaria, and other infectious diseases, to be detected in advance. According to Montgomery (2019), phases 1 and 2 of applying the control charts have different objectives. In phase 1, also called retrospective analysis, the history of the process until the period of interest is analyzed, constructing control limits (called “trial control limits”), to determine if the process was under statistical control during the period of time in which the data was collected, and to see if these estimated limits may be established to monitor future production. On the other hand, in phase 2, also called prospective analysis, the control chart is used to monitor any future observations of the process, with the limits established in phase 1 as its base.

The origin of the residual control chart dates to work by Mandel (1969) and Haworth (1996), which considered the adjustment of simple and multiple linear regression models, respectively. Within the context of autocorrelated data, the use of this SPC tool has been gaining ground and importance in literature (for example, see NOSKIEVICOVÁ, 2009; BISRI and SINGGIH, 2018). The main idea of this technique is to use the residuals of the proposed model as a variable to be monitored. This procedure has a number of advantages, such as: (i) the residuals are uncorrelated; (ii) the control limits obtained are constant; (iii) it is easy to interpret; and (iv) it assists in visualizing the behavior of the series presented.

In their simplest form, the residuals are called “ordinary”, and obtained by the differences between the values observed and those estimated (or predicted) by the model: $e_t = y_t - \hat{y}_t$, for $t = 1, 2, \dots, n$.

Some authors suggest modifying these residuals (standardized and studentized residuals), and working with other types, such as the randomized quantile (PARK *et al.*, 2020) and deviance residuals (SKINNER *et al.*, 2003).

In this research, the variables to be monitored are the standardized Anscombe residuals (ANSCOMBE, 1953; MCCULLAGH and NELDER, 1989), which have approximate normality (even with very small sample sizes) and behavior similar to deviance residuals (cf. PIERCE and SCHAFER, 1986) as the main characteristics. These standardized residuals may be calculated as follows:

$$e_t^{Ansc.p} = \frac{\frac{3}{\hat{\sigma}^2}(1 + y_t \hat{\sigma}^2)^{\frac{2}{3}} - (1 + \hat{y}_t \hat{\sigma}^2)^{\frac{2}{3}} + 3 \left(y_t^{\frac{2}{3}} - \hat{y}_t^{\frac{2}{3}} \right)}{2(\hat{y}_t + \hat{y}_t^2 \hat{\sigma}^2)^{\frac{1}{6}}}, \quad t = 1, 2, \dots, n,$$

where $\hat{\sigma}$ represents the estimate of the standard deviation σ (unknown), and \hat{y}_t the estimated value for the model at moment t .

In the resulting control chart, the central line (CL) is equal to zero, since it is the value expected for the sum of the residuals of a model whose assumptions have not been violated; and the lower and upper control limits (LCL and UCL, respectively) are calculated based on Normal (or Gaussian) distribution. The expressions to obtain these limits are:

$$\text{LCL} = -w * \sigma, \quad \text{CL} = 0 \quad \text{and} \quad \text{UCL} = w * \sigma, \quad (4)$$

where w represents the distance between the control limits and CL ($w = 3$ is usually adopted, which corresponds to the usual three-sigma limits); and σ is the standard deviation (unknown) of the residuals, which may be replaced by the standard deviation $\hat{\sigma}$ calculated (or estimated) from the data available (in particular, it assumes a value equal to 1, since they are standardized residuals). In this case, estimates of the control limits presented in Equation (4) are also obtained.

Described in Montgomery (2019), the control chart to monitor future observations of the process is an empirical procedure, based on the EWMA control chart for autocorrelated data, which combines information on the state of statistical control, and the process dynamic in a single graphic tool. Constructed from the model proposed (adjusted), its CL is defined based on forecasts obtained and the control limits (LCL and UCL) are based on the forecast error variance. Consequently, it is possible to make forecasts on the monitored variable, since the chart is constructed according to the model proposed. Therefore, the occurrence of atypical and/or undesirable phenomena can be predicted. The CL, LCL and UCL estimates can be obtained, as presented in Equation (5):

$$\text{LCL}_{t+h} = \hat{y}_{t+h} - w\hat{\sigma}_{\hat{y}_{t+h}}, \quad \text{CL}_{t+h} = \hat{y}_{t+h} \quad \text{and} \quad \text{UCL}_{t+h} = \hat{y}_{t+h} + w\hat{\sigma}_{\hat{y}_{t+h}}, \quad (5)$$

where t represents the moment at which the observation was collected; h is the number of steps ahead in which the forecast will be carried out; \hat{y}_{t+h} denotes the value (response) estimated by the model at the moment $t + h$; and $\hat{\sigma}_{\hat{y}_{t+h}}$ is the standard deviation estimate of the response predicted at the moment $t + h$.

3. Results and discussion

In this section, we present and discuss the main results obtained from application of the statistical techniques described above (Sections 2.2 and 2.3) to a real data set containing information on notifications of malaria cases in the states of the Legal Amazon region (Section 2.1). In particular, construction of control limits for the proposed charts will be carried out based on a 6σ quality policy, i.e., considering $w = 3$, or also the probability of a type I error (false alarm or false positive) equal to $\alpha = 0.0027$. Therefore, although the process is under statistical control, an out of control sign will be observed for every $1/\alpha = 1/0.0027 \approx 370$ samples, on average (MONTGOMERY, 2019). The ratio (or measure) $1/\alpha$ is well known in literature on SPC as the average run length (ARL). It is important to highlight that in the following analysis, *Falciparum* and *Mista* were considered *Falciparum* malaria, due to the similarity in behavior and form of treatment.

3.1 Descriptive analysis

Initially, a descriptive analysis was carried out, with the aim of studying the behavior of *Vivax* and *Falciparum* malaria notifications, separately, by Legal Amazon state, and also to identify the municipalities which would be maintained to apply the models proposed, since the interest resides in defining an epidemic threshold, and the absence of notifying the disease is not an interesting factor to be taken into consideration. The state of Tocantins was not considered in the analysis process, since there was only one malaria notification in 2015. Therefore, it was removed, since it did not contain the information required to construct the time series.

Table 3 provides the number of malaria cases notified between 2013 and 2017, according to Sivep-Malaria, and the quantity of municipalities that registered occurrences of the disease. A decline in the total number of notifications is observed until 2016, followed by a large increase in 2017; similar behavior is observed for the notifying municipalities. It should be highlighted that during 2017 there was an increase in the incidence of diseases such as dengue, zika and chikungunya, which have the *Aedes aegypti* mosquito as the vector. Another important factor, which is related to the number of malaria cases notified, was the immigration movement encouraged by the new Migration Law proposed through Bill PLS 288/2013, sanctioned on May 24, 2017.

Table 3 - Number of total cases of malaria reported in the Legal Amazon region (period 2013-2017)

| Year | No. Notifications | No. Municipalities |
|------|-------------------|--------------------|
| 2013 | 169,196 | 361 |
| 2014 | 138,925 | 319 |
| 2015 | 137,931 | 288 |
| 2016 | 121,155 | 268 |
| 2017 | 189,191 | 282 |

Source: Sivep-Malaria.

A selection of municipalities within each state in the region (except for Tocantins) was then carried out. The proportion of notifications was calculated in every municipality, in relation to its respective state, and the values obtained were then classified in descending order. The municipalities that jointly accumulated 50% of the cases notified were then selected to continue with the analysis. Table 8 (see Appendix A) presents the municipal codes of the municipalities selected to form the series of notifications of *Vivax* and *Falciparum* malaria cases. It was observed that at least one municipality was maintained for the years under analysis for the states of Acre, Maranhão, Mato Grosso and Rondônia.

Table 4 displays the number of notifications of *Vivax* and *Falciparum* malária cases registered by state and year in the Brazilian Legal Amazon region. From an analysis of the results, it is observed that the state of Maranhão presented the lowest number of cases of the disease, while the state of Amazonas registered the highest number, with this behavior being observed for both types of malaria. With the exception of Rondônia, all the states presented a decline in notifications between 2013 and 2016, followed by an increase in *Vivax* malaria in 2017. It was noted that four states (Acre, Amapá, Maranhão and Roraima) had an increase in the number of *Falciparum* malaria cases in 2016.

Table 4 - Number of total cases of malaria of the types *Vivax* and *Falciparum*, reported in the states of the Legal Amazon region (period 2013-2017)

| State/Year | <i>Vivax</i> | | | | | <i>Falciparum</i> | | | | |
|-------------|--------------|--------|--------|--------|--------|-------------------|-------|-------|-------|-------|
| | 2013 | 2014 | 2015 | 2016 | 2017 | 2013 | 2014 | 2015 | 2016 | 2017 |
| Acre | 13,416 | 11,730 | 10,729 | 15,101 | 16,365 | 6,278 | 5,449 | 3,109 | 3,611 | 4,525 |
| Amapá | 6,373 | 5,535 | 5,105 | 4,322 | 6,692 | 812 | 407 | 245 | 698 | 538 |
| Amazonas | 36,153 | 30,596 | 36,824 | 24,113 | 41,217 | 4,524 | 4,432 | 3,398 | 1,757 | 4,084 |
| Maranhão | 232 | 321 | 92 | 56 | 165 | 28 | 8 | 2 | 5 | 11 |
| Mato Grosso | 539 | 817 | 1,023 | 439 | 460 | 150 | 12 | 10 | 2 | 2 |
| Pará | 8,805 | 3,743 | 3,699 | 6,297 | 17,641 | 3,955 | 1,145 | 811 | 433 | 442 |
| Rondônia | 8,349 | 5,308 | 2,767 | 2,283 | 2,214 | 387 | 854 | 565 | 506 | 489 |
| Roraima | 2,232 | 2,672 | 3,143 | 2,597 | 5,366 | 169 | 221 | 73 | 300 | 75 |

Source: Sivep-Malaria.

Following selection of the municipalities, the data was aggregated by epidemiological weeks (according to the notification calendar provided by the National Notifiable Diseases Information System - SINAN), thereby constructing the time series, with a total of 261 points (or observations) per state, which are displayed in Figures 1 and 2.

As demonstrated in Figure 1, which depicts the time series of *Vivax* malaria cases, the Legal Amazon states do not have identical behavior for this form of the disease, and the following is highlighted: Amazonas and Amapá, which present seasonal behavior during the years studied, with the former experiencing peaks half way during the year (between weeks 20 and 30), while the latter displays peaks at the end of the year (between weeks 40 and 48); Pará and Roraima display little variability in the number of notifications between 2013 and 2016, but register an expressive increase in 2017;

lastly, there was a decline in the number of occurrences registered in Rondônia during this five-year period.

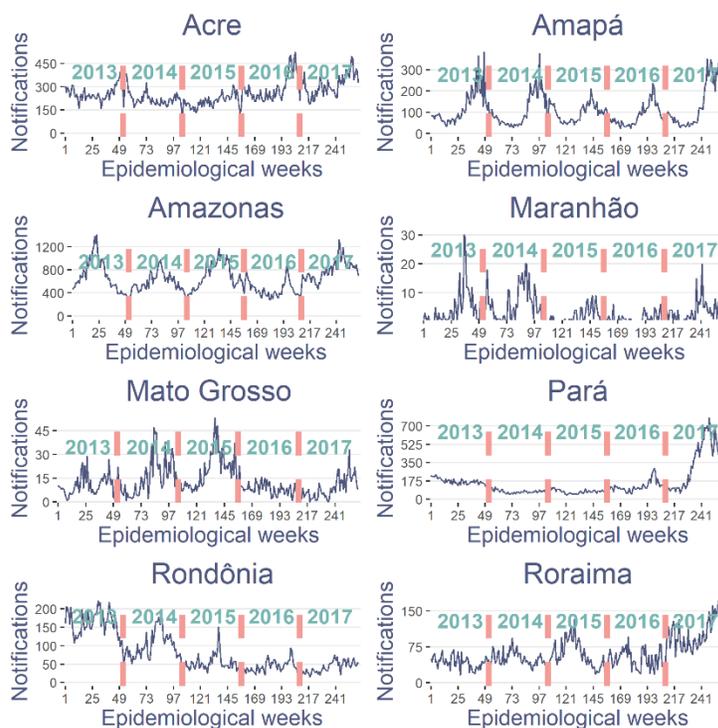


Figure 1 – Time series of *Vivax* malaria cases, by state in the Legal Amazon region (period 2013-2017).

Figure 2 shows the time series of *Falciparum* malaria cases in the states of the region under study, highlighting the following results: the states of Maranhão and Mato Grosso registered very few occurrences and for some weeks (between 2014 and 2017), they did not notify any cases of this form of the disease; in the states of Amazonas and Acre, there was a decline in the number of cases in 2015 and 2016, followed by an increase in 2017.

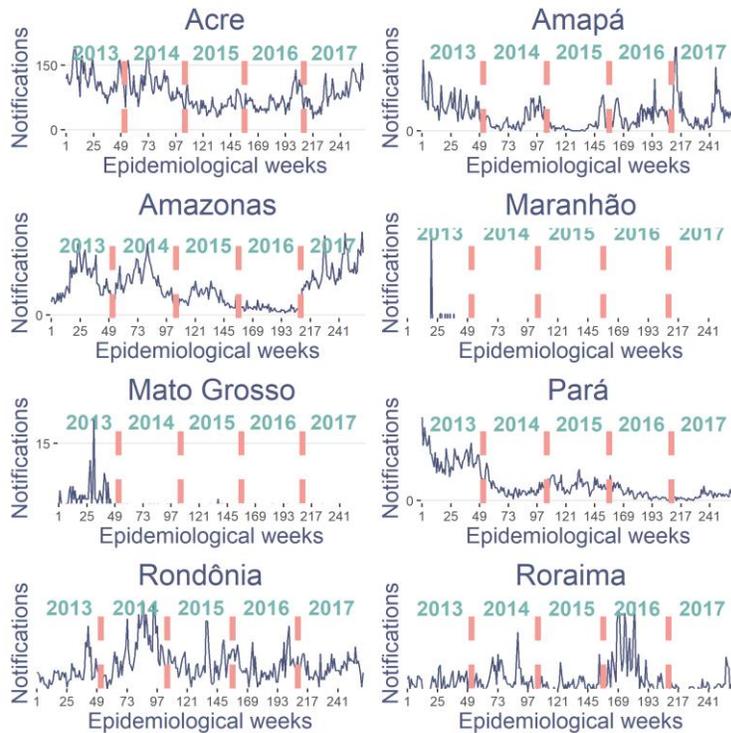


Figure 2 – Time series of *Falciparum* malaria cases, by state in the Legal Amazon region (period 2013-2017).

In a more detailed comparison between Figures 1 and 2, a higher number of notifications for *Vivax* malaria cases is observed in all the states of the Legal Amazon region. This result is expected since this form of the disease occurs on Brazilian territory with greater frequency. For the state of Acre, *Vivax* malaria presents similar behavior during the years analyzed, demonstrating little variability; for *Falciparum* malaria, a decline in the notification of cases for 2015 and 2016 is observed, but this increases once more in 2017. The state of Amapá displays seasonal behavior during the years studied for *Vivax* malaria, while very few occurrences are registered for *Falciparum* during this period. The state of Amazonas displays seasonal behavior for *Vivax* (over this five-year period) and *Falciparum* malaria (only between 2013 and 2014, followed by a decline between 2015 and 2016). Maranhão is the state that presents the lowest number of records for both types of malaria: for *Vivax*, it displays a decline after 2014, while there are no *Falciparum* notifications in many epidemiological weeks (between 2014 and 2017). The states of Mato Grosso and Maranhão also present a decline in *Vivax* malaria between 2016 and 2017, while there are almost no *Falciparum* records of occurrences between 2014 and 2017, considering that no cases of the disease were notified in many epidemiological weeks. In turn, Pará displays the

opposite behavior for the two types of malaria studied: while *Vivax* retains a low variability between 2013 and 2016, followed by an increase in 2017, *Falciparum* registers a decline between 2013 and 2014, maintaining low variability between 2015 and 2017. The state of Rondônia demonstrates a decrease in the number of *Vivax* notifications; there is similar behavior for *Falciparum* over the years analyzed, except for 2014, when there is an increase in the number of cases of this form of the disease. Lastly, the state of Roraima registers an increase in the number of *Vivax* malaria notifications for 2017, while there is an increase in *Falciparum* variability in 2016, when an increase in case notifications also occurs.

3.2 Analysis and modeling the time series

With the assistance of the descriptive analysis carried out in the previous section, we were able to identify that there was an absence of notifications of malaria cases in some states in various epidemiological weeks, for at least one of the forms of the disease studied, causing inflation or an excess of zeros. Therefore, these states were excluded in the final analysis, with only Acre, Amapá, Amazonas and Rondônia remaining. There is adequate treatment for this type of data in literature, considering time series models for excess zeros (or zero inflation) in count data (YANG, 2012; QI *et al.*, 2019), but working with this approach is not within the scope of this study. In addition, due to the limitations of this article, only the results obtained for the state of Amazonas are presented, and those for the other three states can be found in the attachment available at: https://est.ufba.br/sites/est.ufba.br/files/anexo-artigo_tcc.pdf.

The holdout validation technique was utilized, so that observations for the period 2013-2016 are used to form the training group, and the remainder for 2017, to make up the test group. This method has proven useful in the process of model construction and validation (TANTITHAMTHAVORN *et al.*, 2016).

Figure 3 presents the time series of *Vivax* malaria case notifications, and their respective autocorrelation and partial autocorrelation functions for the state of Amazonas (period between 2013 and 2016). The behavior described in these plots indicates the presence of a seasonal component, which was taken into consideration during the modeling process.

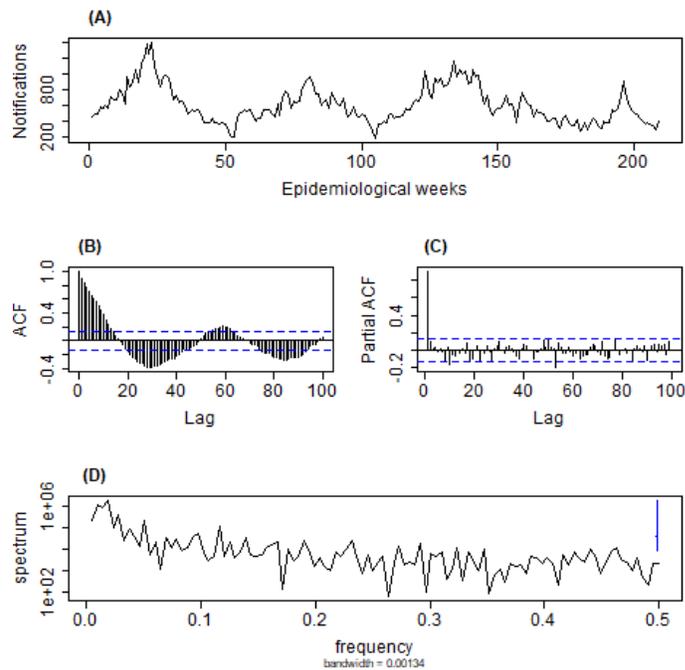


Figure 3 – (A) Time series; (B) autocorrelation plot; (C) partial autocorrelation plot; (D) periodogram, for *Vivax* malaria in the state of Amazonas (period 2013-2016).

Figure 4 displays the behavior of the series of notifications of *Falciparum* malaria cases, and their respective autocorrelation and partial autocorrelation functions. The standard observed in these plots suggests the presence of seasonality, also with *Vivax* malaria. Therefore, this component was taken into consideration in the modeling process.

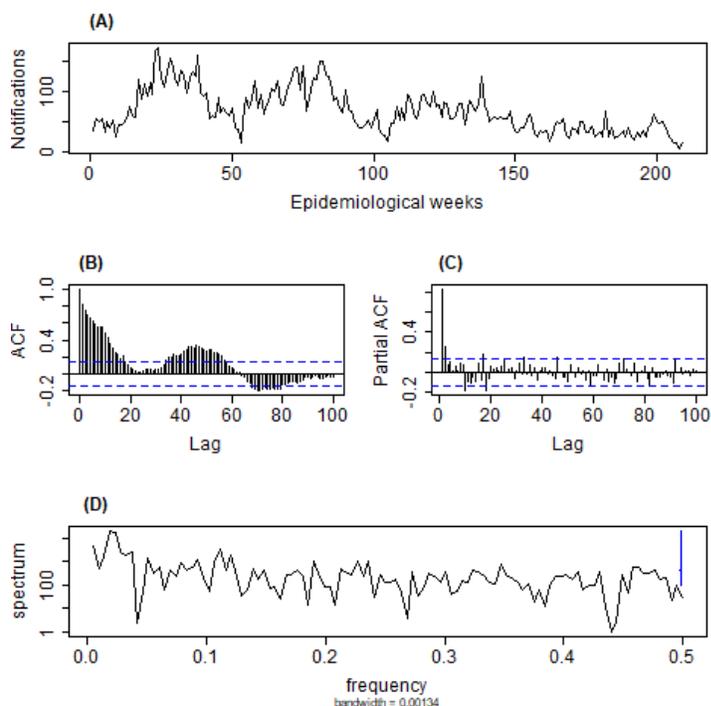


Figure 4 – (A) Time series; (B) autocorrelation plot; (C) partial autocorrelation plot; (D) periodogram, for *Falciparum* malaria in the state of Amazonas (period 2013-2016).

Table 5 presents the AIC measurements calculated to assist in identifying the time terms for the models adjusted to the series of *Vivax* and *Falciparum* malaria cases. The selection, based on the periodogram (Figures 3(D) and 4(D)), and removal of the seasonal effect, were the stages required to detect the time terms. Thus, the ARMA(1,1) and ARMA(2,1) models were selected for *Vivax* and *Falciparum* malaria, respectively, since they have lower AIC values. Nevertheless, the ARMA(1,1) model was also selected for *Falciparum*, since there are fewer parameters to be estimated.

Table 5 – AIC values for the fitted models

| <i>Vivax</i> | | <i>Falciparum</i> | |
|--------------|----------|-------------------|----------|
| Model | AIC | Model | AIC |
| ARMA(1,0) | 2,509.20 | ARMA(1,0) | 1,843.97 |
| ARMA(2,0) | 2,508.16 | ARMA(2,0) | 1,829.91 |
| ARMA(1,1) | 2,507.86 | ARMA(1,1) | 1,825.09 |
| ARMA(2,2) | 2,509.69 | ARMA(2,1) | 1,823.99 |
| ARMA(3,0) | 2,509.83 | ARMA(2,2) | 1,826.89 |

The approach proposed in this study was then applied, based on adjustment of the GARMA model, in which the Poisson and Negative Binomial distributions were considered for the two series of malaria notifications. The following were used as predictor variables: average rainfall (in mm), average humidity (in %), and average air temperature (in °C), in which the multicollinearity between them was tested. Thus, the covariates that presented a VIF higher than 10 were excluded from the analysis, as described in Section 2.2.3.

These variables are influential on the number of malaria cases, since they affect development of the disease vector. However, since they are expressed in distinct units of measurement, there was a need to transform these quantities, in which use of the logarithmic function (base e) was selected, with the value of 0.001 (“offset”) added to the records (purely for mathematical reasons) whose originally observed data was equal to zero. This transformation aims to reduce the effect of different magnitudes between the units of scale of the data observed.

For the two series of malaria cases, the presence of multicollinearity was detected, which led to the removal of the humidity variable from analysis (with a VIF of 93.07 and 30.25 for *Vivax* and *Falciparum*, respectively).

Time and seasonal components were included for both series. The Wald test (WALD, 1943) was applied, where the variables which were statistically significant at the level of 5% for the model to continue, were verified. In addition, the precision of the estimates and residual behavior were also evaluated, with removal of the variables.

Following adjustment of the GARMA models, a diagnostic analysis was carried out, and scoring rules calculated, in order to identify which model better adjusted to the data in each time series. According to Gneiting *et al.* (2007), the model with the lowest scorings is preferable/most appropriate.

Figure 5 displays the results of the diagnostic analysis of the GARMA models adjusted to the data of the series of *Vivax* malaria cases for the state of Amazonas. It is observed that there is no serial correlation that has not been contained by the Negative Binomial GARMA model, as shown in the residual autocorrelation plot (Figure 5(A)). In addition, the Negative Binomial GARMA model was more precise for the estimates obtained, in line with the marginal calibration plot (Figure 5(B)), and there was a better probabilistic calibration of the predictive distribution, considering that the Probability Integral Transform (PIT) histogram (Figure 5(C)-(D)) better addresses uniformity. For further information on the marginal calibration plot and PIT histogram, see Gneiting *et al.* (2007) and Czado *et al.* (2009), respectively. The same behavior was observed in the diagnostic analysis of the GARMA models adjusted to the time series of *Falciparum* malaria cases (see Figure 9, Appendix B).

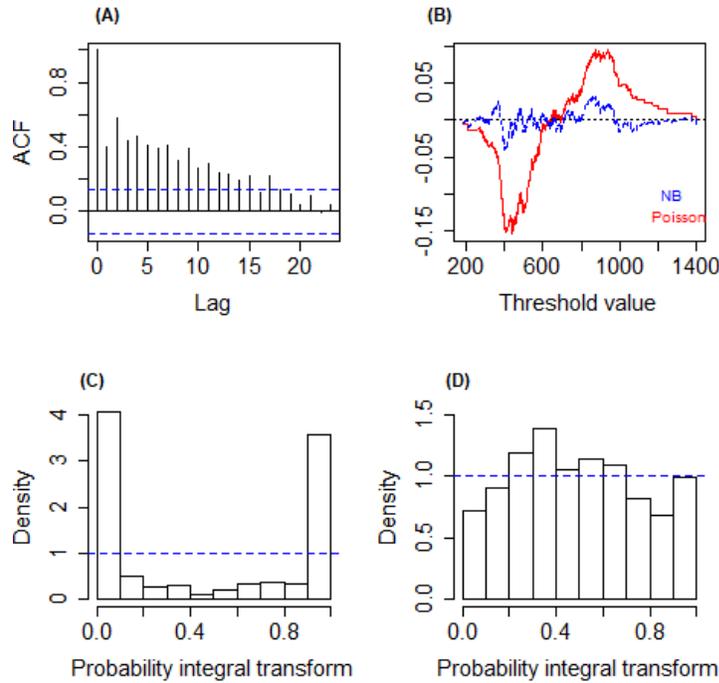


Figure 5 – Diagnostic analysis of the Poisson GARMA and Negative Binomial GARMA models fitted to the series of *Vivax* malaria cases. (A) Autocorrelation plot of residuals; (B) marginal calibration plot; (C) PIT histogram for the adjusted Poisson GARMA model; (D) PIT histogram for the adjusted Negative Binomial GARMA model.

Analyzing the results (scoring rules) presented in Table 6, it is observed that the Negative Binomial GARMA model is most appropriate for both time series of malaria cases, being indicated by six scoring rules for *Vivax*, and all of the rules for *Falciparum*.

Table 6 – Scoring rules for the fitted GARMA models

| Rule | <i>Vivax</i> | | <i>Falciparum</i> | |
|-----------------|--------------|----------|-------------------|--------|
| | Poisson | NB | Poisson | NB |
| Logarithmic | 17.57 | 5.96 | 7.01 | 4.34 |
| Quadratic | 0.00 | 0.00 | 0.00 | -0.02 |
| Spherical | -0.05 | -0.06 | -0.09 | -0.13 |
| <i>Rankprob</i> | 77.48 | 46.77 | 14.58 | 10.69 |
| <i>Dawseb</i> | 32.68 | 10.13 | 11.85 | 6.91 |
| <i>Normsq</i> | 26.30 | 0.97 | 7.70 | 0.98 |
| <i>Sqerror</i> | 17,236.26 | 9,123.38 | 601.32 | 379.85 |

Note: NB = Negative Binomial.

Thus, the estimates (point and interval) of the parameters of the Negative Binomial GARMA models adjusted to the series of *Vivax* and *Falciparum* malaria cases are presented in Table 7. It is observed that the overdispersion coefficient σ^2 , which is related to the dispersion parameter ϕ of the Negative Binomial distribution by $\sigma^2 = 1/\phi$, presented a higher value for *Vivax* malaria when compared with *Falciparum* (37.04 versus 12.05).

Table 7 – Estimation results of the Negative Binomial GARMA models fitted to the series of *Vivax* and *Falciparum* malaria cases

| Parameter | <i>Vivax</i> | | | | <i>Falciparum</i> | | | |
|--------------|--------------|-------|----------|----------|-------------------|-------|----------|----------|
| | Estimate | S.E. | 95% L.L. | 95% U.L. | Estimate | S.E. | 95% L.L. | 95% U.L. |
| α_0 | 1.326 | 0.375 | 0.590 | 2.062 | 0.902 | 0.199 | 0.511 | 1.293 |
| λ_1 | 0.835 | 0.078 | 0.683 | 0.987 | 0.784 | 0.048 | 0.690 | 0.878 |
| γ_1 | -0.047 | 0.095 | -0.234 | 0.139 | - | - | - | - |
| Temperature | 0.005 | 0.044 | -0.081 | 0.090 | - | - | - | - |
| β_{s1} | -0.066 | 0.024 | -0.114 | -0.018 | -0.076 | 0.035 | -0.145 | -0.006 |
| β_{s2} | 0.048 | 0.018 | 0.014 | 0.083 | 0.053 | 0.032 | -0.010 | 0.117 |
| ϕ | 0.027 | - | - | - | 0.083 | - | - | - |

Note: S.E. = Standard Error, L.L. = Lower Limit,

U.L. = Upper Limit, α_0 = Intercept,

λ_1 = autoregressive coefficient,

γ_1 = moving average coefficient,

β_{si} = seasonality coefficient, $i = 1, 2$.

3.3 Monitoring and forecasting new cases of the disease

Following definition of the final GARMA models (Negative Binomial GARMA for both series of malaria cases), the residual control charts of these models were constructed, in order to verify if the processes in question are under statistical control (phase I or retrospective analysis).

Figure 6(A) displays the control chart for the residuals of the Negative Binomial GARMA model adjusted to the series of *Vivax* malaria cases. From this chart, it is verified that the process is under statistical control, considering that no point exceeded the control limits.

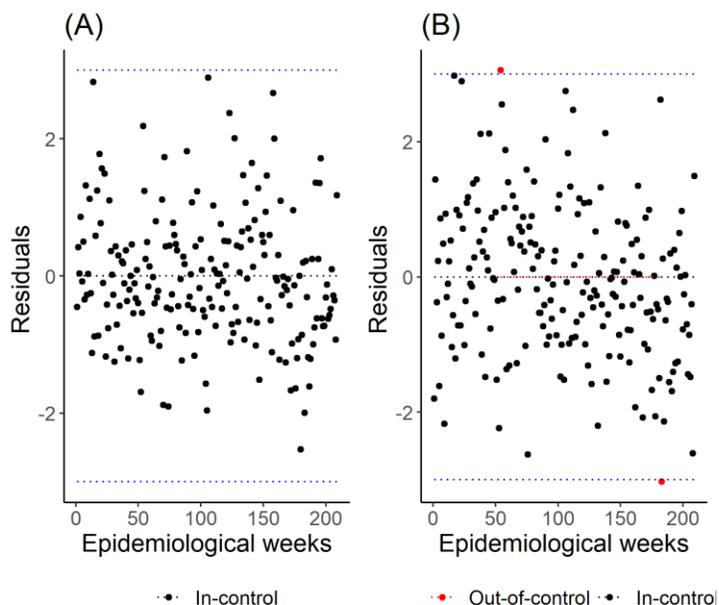


Figure 6 – Control charts for the residuals of the Negative Binomial GARMA models fitted to the malaria cases series of the types: (A) *Vivax*; (B) *Falciparum*.

Figure 6(B) shows the control chart for the residuals of the Negative Binomial GARMA model adjusted to the series of *Falciparum* malaria cases. It is observed that this process is out of statistical control, since one point exceeded the control limits, indicating a possible outbreak of *Falciparum* malaria in the state of Amazonas.

Forecasts (of the number of cases of the two series of the disease) were then made, considering five steps ahead (in other words, five epidemiological weeks). The control limits to monitor future observations of the two processes in question, were calculated according to a quality policy 6σ ($w = 3$ or, equivalently, $\alpha = 0.0027$). In addition, the behavior of these processes was verified, with the insertion of new observations of the test set (2017), compared with the 3rd quartile charts (phase 2 of the analysis, or monitoring). Proposed by Braz (2005) in a study which had the objective of the early detection of malaria epidemics in the Legal Amazon region, this type of control chart is currently used by the MS to monitor and control malaria.

As shown in Figure 7, the process related to the number of *Vivax* malaria cases will be out of statistical control, according to the two methodologies used: the Negative Binomial GARMA model (Figure 7(A)) and 3rd quartile method (Figure 7(B)), considering that some observations exceeded the control limits established. Under the approach of the Negative Binomial GARMA model, only the number of cases in epidemiological week 1 exceeded the UCL, indicating a possible outbreak. In relation to the approach of the 3rd quartile method, the lack of statistical control for epidemiological weeks 1-5 was signaled, which is indicative of a possible future *Vivax* malaria epidemic in the state of Amazonas.

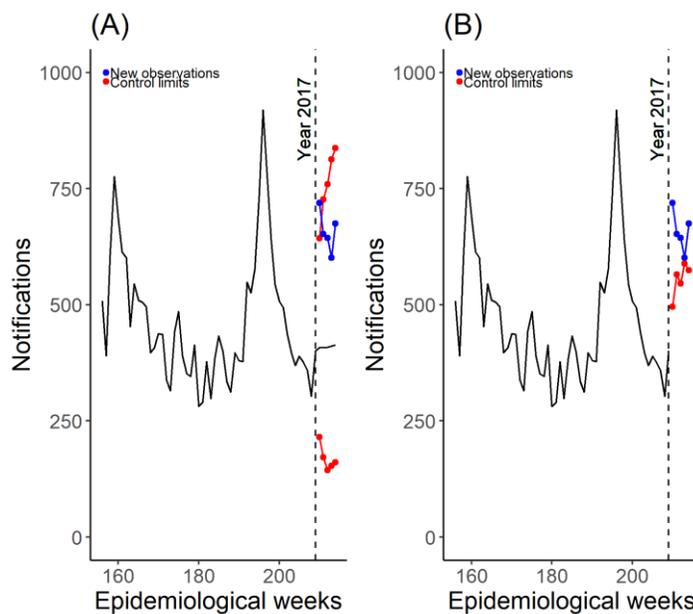


Figure 7 – Control charts for monitoring future observations of the process regarding *Vivax*-type malaria: (A) Negative Binomial GARMA model; (B) 3rd quartile method.

In Figure 8, it is observed that the process related to the number of *Falciparum* malaria cases will be out of statistical control according to the 3rd quartile method. Under the approach of the Negative Binomial GARMA model (Figure 8(A)), the lack of statistical control for the five following epidemiological weeks was not signaled. While in the approach of the 3rd quartile method (Figure 8(B)), epidemiological weeks 1-2 and 5 were detected as a possible future *Falciparum* malaria epidemic in the state of Amazonas.

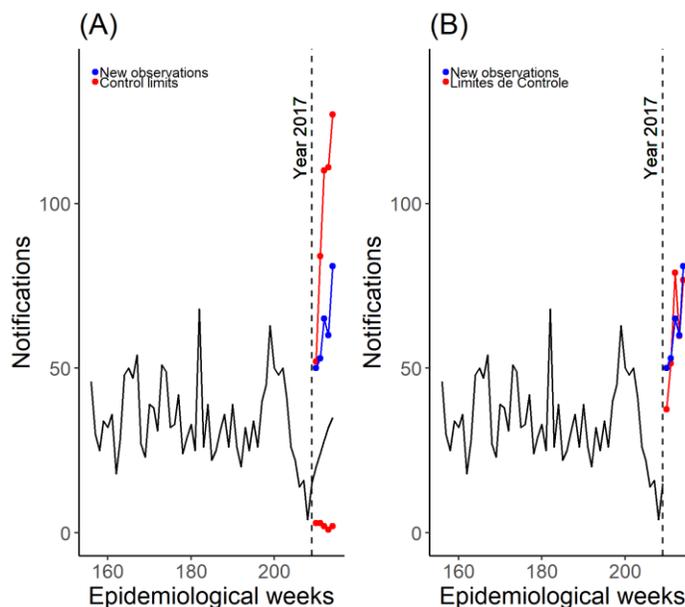


Figure 8 – Control charts for monitoring future observations of the process regarding *Falciparum*-type malaria: (A) Negative Binomial GARMA model; (B) 3rd quartile method.

It is interesting to observe that the chart to monitor future observations of the process, based on the Negative Binomial GARMA model, presents the UCL and LCL, considering that the latter is also of major relevance, since although it does not identify possible outbreaks/epidemics, it may indicate a possible underreporting of cases, since the values observed are lower than that estimated for the LIC. When inserting new observations, it is important to review the model, in order to attain more precise forecasts. In addition, we recommend not forecasting many steps ahead, since the forecasts are obtained based on an estimate of the previous observation, and for every estimate there is an associated error, which tends to rise as the number of forecasts is added, thereby making them less accurate.

Conclusions

In the first instance, modeling for discrete count data with a time dependence structure for the number of *Vivax*, *Falciparum* and *Mista* malaria cases (separately) in the Legal Amazon region of Brazil was addressed in this research. The use of GARMA models for these time series of discrete data (counts) was then proposed, observing that the Negative Binomial distribution adjusted more adequately, when compared to the Poisson distribution, since the former was able to consider the overdispersion present.

Application of the control chart for GARMA model residuals in phase 1 (retrospective analysis), was shown to be appropriate, considering that the process of

interest (related to *Vivax*, *Falciparum* and *Mista* malaria cases) is influenced by explanatory variables, with no great correlation identified between them, together with the time dependence structure considered in the modeling.

It was also observed that the control chart proposed to monitor future observations of the process produced similar results to those of the 3rd quartile (or method) chart, with the latter currently being used by the MS of Brazil to identify malaria outbreaks or epidemics (BRAZ, 2005). It is important to highlight that we do not have a gold standard, i.e., reliable and concrete information on the occurrence, or otherwise, of a malaria outbreak/epidemic in the Legal Amazon region in the weeks analyzed. We should highlight that the proposal based on the GARMA model has the advantage of making forecasts (since it is a probabilistic approach) and, therefore may foresee the emergence of atypical/undesirable phenomena, and contribute towards directing intervention measures, i.e., control actions with a view to combating the spread of the disease. In addition, the control limits of the proposed chart may be used to define epidemic thresholds, as in the 3rd quartile chart. Registering a number of cases of the disease which is higher than the UCL may indicate an outbreak (in the case of a single overshoot), or epidemic (in the case of successive overshoots), while registering a number that is lower than the LCL may indicate possible underreporting.

As suggestions for future study, the following is important: (i) to investigate the performance of the method proposed to monitor new observations of the process, compared with the usual 3rd quartile method, considering a gold standard (known status) for outbreaks and epidemics in the Amazon region; (ii) to carry out numerical simulations using a calibration/tuning process, to improve the performance of the method proposed, investigating the effects associated with estimating the model parameters on the constructed control chart performance in greater detail (JENSEN *et al.*, 2006); (iii) to consider other models for counting processes, such as the state-space models (DURBIN and KOOPMAN, 1997; SHEPHARD and PITT, 1997) and the GARMA model extensions/modifications (CORDEIRO and ANDRADE, 2009; ALBARRACIN *et al.*, 2019); (iv) to evaluate the existence of spatial dependence in the states in the region, and consider this in the modeling process, which is an area that has been underdeveloped in the study of diseases (LYRIO, 2019); and (v) to use multivariate SPC techniques to both model and monitor the occurrence of *Vivax*, *Falciparum* and *Mista* malaria cases.

Acknowledgements

This research was funded by Bahia Research Foundation (FAPESB; n. app 0071/2016). Authors would like to thank reviewers and editor for their comments and suggestions.

SENA, J. G. S.; FERREIRA, P. H.; FIACCONE, R. L. O Controle estatístico de processos como ferramenta para controle e prevenção de epidemias de malária na região da Amazônia Legal. *Braz. J. Biom.*, Lavras, v.40, n.1, p.21-49, 2022.

- *RESUMO: A malária ainda é uma doença fatal em muitos países ao redor do mundo. Estabelecer medidas de controle e prevenção dessa doença é uma preocupação mundial desde 1950, quando a Organização Mundial da Saúde lançou um Plano de Erradicação da Malária. No Brasil, a malária foi eliminada em grande parte do território, porém resistindo em uma área conhecida como Amazônia Legal. Posto isto, o escopo principal deste trabalho consiste em desenvolver gráficos de controle estatístico que considerem a estrutura de dependência temporal existente nos dados e sejam adequados para o monitoramento atual e futuro dos casos de malária, com o intuito de detectar possíveis surtos ou epidemias em estados da região da Amazônia Legal. As ferramentas aqui apresentadas poderiam, ainda, colaborar no direcionamento das ações de controle e no combate à propagação da doença. Em particular, pretende-se: (i) construir um modelo estatístico para previsão de ocorrências de casos da doença, o qual considere a existência de uma possível estrutura de dependência temporal entre os dados coletados; (ii) utilizar as técnicas de Controle Estatístico de Processos (CEP), notadamente os gráficos de controle, para monitorar (separadamente para cada forma da doença) os casos de malária dos tipos Plasmodium Vivax, Plasmodium Falciparum e Plasmodium Mista na região amazônica; (iii) estabelecer limiares epidêmicos com base nos gráficos de controle obtidos. Considerando dados de 2013 a 2017, o presente estudo revelou, dentre outros, que os modelos generalizados autorregressivos e de médias móveis com distribuição Binomial Negativa (modelos GARMA Binomial Negativo) se mostraram mais eficientes, ajustando-se melhor, comparativamente aos modelos GARMA Poisson, devido à superdispersão existente nos dados analisados.*
- *PALAVRAS-CHAVE: Gráfico de controle, modelos GARMA, monitoramento estatístico, previsão, superdispersão.*

References

AKAIKE, H. On entropy maximization principle. In: Krishnaiah, P.R. (Ed.), *Applications of Statistics*, North-Holland, 1977.

AKINWANDE, M. O.; DIKKO, H. G.; SAMSON, A. Variance inflation factor: as a condition for the inclusion of suppressor variable(s) in regression analysis. *Open Journal of Statistics*, v.5, n.07, p.754, 2015.

ALBARRACIN, O. Y. E.; ALENCAR, A. P.; HO, L. L. Generalized autoregressive and moving average models: multicollinearity, interpretation and a new modified model. *Journal of Statistical Computation and Simulation*, v.89, n.10, p.1819-1840, 2019.

ANDRADE, B. S. *GARMA models, a new perspective using Bayesian methods and transformations*. 103f. Tese de Doutorado em Estatística - Universidade Federal de São Carlos, São Carlos-SP, 2016.

- ANSCOMBE, F. J. Contribution to the discussion of H. Hotelling's paper. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, v.15, p.229-230, 1953.
- BELSLEY, D. A.; KUH, E.; WELSCH, R. E. *Regression Diagnostics*. New York: John Wiley & Sons, 1980.
- BENJAMIN, M. A.; RIGBY, R. A.; STASINOPOULOS, D. M. Generalized autoregressive moving average models. *Journal of the American Statistical Association*, v.98, n.461, p.214-223, 2003.
- BENNEYAN, J. C. Statistical quality control methods in infection control and hospital epidemiology - Part I: introduction and basic theory. *Infection Control and Hospital Epidemiology*, v.19, n.3, p.194-214, 1998.
- BISRI, H.; SINGGIH, M. L. Improvement of Shewhart Control Chart for Autocorrelated Data in Continuous Production Process. In: *2018 4th International Conference on Science and Technology (ICST)*. IEEE, p.1-6, 2018.
- BRAZ, R. M. *Detecção precoce de epidemias de malária no Brasil: uma proposta de automação*. 122f. Dissertação (Mestrado em Saúde Pública) - Escola Nacional de Saúde Pública, Rio de Janeiro-RJ, 2005.
- BURNS, C. M.; BENNETT, C. J.; MYERS, C. T.; WARD, M. The use of cusum analysis in the early detection and management of hospital bed occupancy crises. *Medical Journal of Australia*, v.183, n.6, p.291-294, 2005.
- CORDEIRO, G. M.; ANDRADE, M. G. Transformed generalized linear models. *Journal of Statistical Planning and Inference*, v.139, n.9, p.2970-2987, 2009.
- CULLEN, J. R.; CHITPRAROP, U.; DOBERSTYN, E. B.; SOMBATWATTANANGKUL, K. An epidemiological early warning system for malaria control in northern Thailand. *Bulletin of the World Health Organization*, v.62, n.1, p.107-114, 1984.
- CZADO, C.; GNEITING, T.; HELD, L. Predictive model assessment for count data. *Biometrics*, v.65, n.4, p.1254-1261, 2009.
- DURBIN, J.; KOOPMAN, S. J. Monte Carlo maximum likelihood estimation for non-Gaussian state space models. *Biometrika*, v.84, n.3, p.669-684, 1997.
- FERREIRA, M. E. M. C. "Doenças tropicais": o clima e a saúde coletiva. Alterações climáticas e a ocorrência de malária na área de influência do reservatório de Itaipu, PR. *Terra Livre*, v.1, n.20, p.179-192, 2015.

- GNEITING, T.; BALABDAOUI, F.; RAFTERY, A. E. Probabilistic forecasts, calibration and sharpness. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, v.69, n.2, p.243–268, 2007.
- HAWORTH, D. A. Regression control charts to manage software maintenance. *Journal of Software Maintenance: Research and Practice*, v.8, n.1, p.35-48, 1996.
- JENSEN, W. A.; JONES-FARMER, L. A.; CHAMP, C. W.; WOODALL, W. H. Effects of parameter estimation on control chart properties: a literature review. *Journal of Quality Technology*, v.38, n.4, p.349-364, 2006.
- LAPOUBLE, O. M. M.; SANTELLI, A. C. F.; MUNIZ-JUNQUEIRA, M. I. Situação epidemiológica da malária na região amazônica brasileira, 2003 a 2012. *Revista Panamericana de Salud Pública*, v.38, p.300–306, 2015.
- LIBOSCHIK, T.; FOKIANOS, K.; FRIED, R. tscount: An R package for analysis of count time series following generalized linear models. *Journal of Statistical Software*, v.82, n.5, p.1–51, 2017.
- LYRIO, B. C.; MONROY, N. A. J.; MOLINARES, F. A. F. Modelagem espaço-temporal da incidência de dengue na região da Grande Vitória (RGV) - ES. *Revista Brasileira de Biometria*, v.37, n.2, p.135-148, 2019.
- MANDEL, B. J. The regression control chart. *Journal of Quality Technology*, v.1, n.1, p.1-9, 1969.
- MCCULLAGH, P.; NELDER, J. A. *Generalized Linear Models*. 2nd Edition, Chapman and Hall. London, UK, 1989.
- MONTGOMERY D. C. *Introduction to Statistical Quality Control*. 8.ed., New York: John Wiley & Sons, 2019.
- NOSKIEVICOVÁ, D. Statistical analysis of the blast furnace process output parameter using ARIMA control chart with proposed methodology of control limits setting. *Metalurgija*, v.48, n.4, p.281-284, 2009.
- PAFIADACHE, C.; ZANINI, R.; SOUZA, A. M.; ROSA, L. C.; PERIPOLLI, A. Uma Aplicação de Gráficos de Controle Estatístico e Modelos de Previsão De Box- Jenkins para Monitoramento de Sistemas de Saúde no Rio Grande do Sul, Brasil. *Revista ESPACIOS*, v.36, n.17, p.6, 2015.
- PARK, K.; JUNG, D.; KIM, J. M. Control charts based on randomized quantile residuals. *Applied Stochastic Models in Business and Industry*, v.36, n.4, p.716-729, 2020.

PIERCE, D. A.; SCHAFER, D. W. Residuals in generalized linear models. *Journal of the American Statistical Association*, v.81, n.396, p.977-986, 1986.

QI, X.; LI, Q.; ZHU, F. Modeling time series of count with excess zeros and ones based on INAR (1) model with zero-and-one inflated Poisson innovations. *Journal of Computational and Applied Mathematics*, v.346, p.572-590, 2019.

R CORE TEAM. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>. 2021.

SCHILLING, W. Analysis of the data of a public health organization by the control chart method. *Journal of the American Statistical Association*, v.39, n.227, p.311-324, 1944.

SCHWARZ, G. Estimating the dimension of a model. *The Annals of Statistics*, v.6, n.2, p.461-464, 1978.

SELLICK, J. A. The use of statistical process control charts in hospital epidemiology. *Infection Control and Hospital Epidemiology*, v.14, n.11, p.649-656, 1993.

SHEPHARD, N.; PITT, M. K. Likelihood analysis of non-Gaussian measurement time series. *Biometrika*, v.84, n.3, p.653-667, 1997.

SHEWHART, W. A. *Economic Control of Quality of Manufactured Product*. London: Macmillan and Co Ltd, 1931.

SKINNER, K. R.; MONTGOMERY, D. C.; RUNGER, G. C. Process monitoring for multiple count data using generalized linear model-based control charts. *International Journal of Production Research*, v.41, n.6, p.1167-1180, 2003.

TANTITHAMTHAVORN, C.; MCINTOSH, S.; HASSAN, A. E.; MATSUMOTO, K. An empirical comparison of model validation techniques for defect prediction models. *IEEE Transactions on Software Engineering*, v.43, n.1, p.1-18, 2016.

WALD, A. Tests of statistical hypotheses concerning several parameters when the number of observations is large. *Transactions of the American Mathematical Society*, v.54, n.3, p.426-482, 1943.

YANG, M. *Statistical models for count time series with excess zeros*. PhD thesis, University of Iowa, 2012.

ZANINI, R. R.; SOUZA, A. M.; JACOBI, L. F.; PERIPOLLI, A.; DAPPER, S. N.; SILVA, A. C. J. A utilização dos gráficos de controle para acompanhamento de processos de vigilância. *Revista ESPACIOS*, v.37, n.27, p.5, 2016.

Received in 22.10.2020
Approved after revised in 04.06.2021

Appendix A

Table 8 - Codes of the municipalities selected for the analysis of malaria of the types *Vivax* and *Falciparum*

| State / Year | <i>Vivax</i> | | | | | <i>Falciparum</i> | | | | |
|--------------------|--------------|---------|---------|---------|---------|-------------------|---------|---------|---------|---------|
| | 2013 | 2014 | 2015 | 2016 | 2017 | 2013 | 2014 | 2015 | 2016 | 2017 |
| Acre | 1200203 | 1200203 | 1200203 | 1200203 | 1200203 | 1200203 | 1200203 | 1200203 | 1200203 | 1200203 |
| | 1600402 | 1600402 | 1600154 | 1600600 | 1600600 | 1600204 | 1600600 | 1600402 | 1600402 | 1600402 |
| Amapá | 1600303 | 1600600 | 1600055 | 1600402 | 1600204 | | 1600303 | | | |
| | 1600204 | 1600303 | 1600204 | 1600204 | 1600402 | | | | | |
| | 1301407 | 1302405 | 1302603 | 1303809 | 1303809 | 1301407 | 1302405 | 1300201 | 1303601 | 1303601 |
| | 1302603 | 1302603 | 1301803 | 1302603 | 1302603 | 1302405 | 1300201 | 1302405 | 1300201 | 1300409 |
| | 1303809 | 1303809 | 1301407 | 1300409 | 1300409 | 1300201 | 1301407 | 1301803 | 1301654 | |
| Amazonas | 1301803 | 1301407 | 1303809 | 1300201 | 1303601 | 1301803 | 1301654 | 1303601 | 1300409 | |
| | 1302405 | 1300409 | 1302405 | | 1301209 | 1301654 | | 1301654 | | |
| | 1303908 | 1301209 | 1300409 | | 1301654 | | | 1300409 | | |
| | 1300201 | 1301803 | 1303908 | | | | | | | |
| | 1304062 | 1300201 | | | | | | | | |
| | 2109270 | 2100550 | 2100550 | 2102606 | 2114007 | 2112456 | 2112456 | 2111672 | 2112407 | 2105351 |
| Maranhão | 2112456 | 2102606 | 2102606 | 2104677 | 2102606 | | 2109809 | 2104800 | 2109239 | |
| | 2102606 | 2103307 | 2103307 | 2103174 | 2103174 | | | | 2104008 | |
| Mato Grosso | 5103254 | 5103254 | 5103254 | 5103254 | 5103254 | 5103254 | 5103254 | 5103254 | 5103254 | 5106752 |
| | 1503606 | 1503606 | 1503606 | 1500701 | 1501105 | 1503606 | 1503606 | 1500701 | 1503606 | 1503606 |
| Pará | 1503754 | | 1500701 | 1501808 | 1500701 | | | | | |
| | | | | 1507706 | 1505205 | | | | | |
| | | | | 1503606 | | | | | | |
| Rondônia | 1100205 | 1100205 | 1100205 | 1100205 | 1100205 | 1100205 | 1100205 | 1100205 | 1100205 | 1100205 |
| | 1400027 | 1400027 | 1400050 | 1400050 | 1400472 | 1400027 | 1400027 | 1400027 | 1400027 | 1400175 |
| Roraima | 1400506 | 1400506 | 1400472 | 1400027 | 1400175 | 1400472 | | | | |
| | 1400175 | 1400209 | 1400027 | 1400472 | | | | | | |

Source: Sivep-Malaria. To consult the chosen municipalities, access the link: <https://www.ibge.gov.br/explica/codigos-dos-municipios.php>.

Appendix B

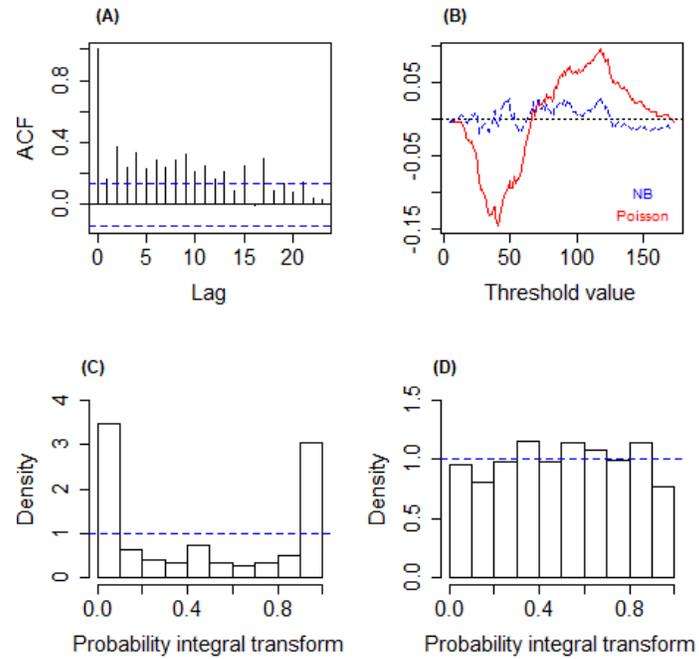


Figure 9 – Diagnostic analysis of the Poisson GARMA and Negative Binomial GARMA models fitted to the series of *Falciparum* malaria cases. (A) Autocorrelation plot of residuals; (B) marginal calibration plot; (C) PIT histogram for the adjusted Poisson GARMA model; (D) PIT histogram for the adjusted Negative Binomial GARMA model.