Time-series forecasting models: An application for climatological parameters in the city of Belém, Pará, Brazil

Douglas Matheus das Neves Santos[1], Yuri Antônio da Silva Rocha[1], Danúbia Leão de Freitas[1], Paulo Roberto Estumano Beltrão Júnior[1], Paulo Cerqueira dos Santos Junior[2], Glauber Tadaiesky Marques[2], Otavio Andre Chase[2], Pedro Silvestre da Silva Campos[2].

[1]Graduating in Environmental Engineering and Renewable Energies at the Federal Rural University of the Amazon - UFRA, campus Belém, Pará, Brazil.
[2]Professor at the Federal Rural University of the Amazon - UFRA, Belém campus, Pará Brazil, Cyberspace Institute - ICIBE.

Abstract

Statistical and mathematical models of forecasting are of paramount importance for the understanding and study of databases, especially when applied to data of climatological variables, which enables the atmospheric study of a city or region, enabling greater management of the anthropic activities and actions that suffer the direct or indirect influence of meteorological parameters, such as precipitation and temperature. Therefore, this article aimed to analyze the behavior of monthly time series of Average Minimum Temperature, Average Maximum Temperature, Average Compensated Temperature, and Total Precipitation in Belém (Pará, Brazil) on data provided by INMET, for the production and application forecasting models. A 30-year time series was considered for the four variables, from January 1990 to December 2020. The Box and Jenkins methodology was used to determine the statistical models, and during their applications, models of the SARIMA and Holt-Winters class were estimated. For the selection of the models, analyzes of the Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), Autocorrelation Correlogram (ACF), and Partial Autocorrelation (PACF) and tests such as Ljung-Box and Shapiro-Wilk were performed, in addition to Mean Square Error (NDE) and Absolute Percent Error Mean (MPAE) to find the best accuracy in the predictions. It was possible to find three SARIMA models: (0,1,2) (1,1,0) [12], (1,1,1) (0,0,1) [12], (0,1,2) (1,1,0) [12]; and a Holt-Winters model with additive seasonality. Thus, we found forecasts close to the real data for the four-time series worked from the SARIMA and Holt-Winters models, which indicates the feasibility of its applicability in the study of weather forecasting in the city of Belém. However, it is necessary to apply other possible statistical models, which may present more accurate forecasts.

Index Terms- Time series, Forecasting, Meteorology, SARIMA, Holt-Winters.

INTRODUCTION

Time series can be conceptualized as a collection of observations ordered over time, in a specific interval, explaining behaviors of variables (SILVA; GUIMARÃES; TAVARES, 2008) of one or more data sets, in which the model applications mathematicians and statisticians can explain the dynamics of the phenomena that occur in nature (SILVA; GUIMARÃES; TAVARES, 2003).

Such a mechanism makes possible the technical and scientific monitoring of databases generated and stored in the multiple scientific and professional sectors. In this context, understanding the data sequence can help understand the causes and consequences of their behavior. Therefore, the need to add statistical knowledge to generate mathematical models that resemble the original data over time emerges, and then, based on this synthetic model, make predictions that will be close to the accurate data.

Applied to climatological factors, the analysis of time series plays a significant role in the success or failure of many enterprises, since the study of anticipating how the climate varies allows better management of agriculture, water resources, and fishing activity, in addition to the possibility of relevant contribution in fields of transport, supply, tourism, and leisure (SILVA; GUIMARÃES; TAVARES, 2008).

More specifically, concerning water resources, they are invited, and their uses are the most diverse, such as agriculture and industrial water supply. If they are not used correctly, they tend to cause major problems, especially in arid and semi-arid regions that deal more constantly with scarcity (HEYDARI et al., 2018). In addition, according to what is already commonly debated, the repercussion of the inflationary drought from the water supply to the socioeconomic impacts that result from it, it is worth mentioning that with the greater frequency of the same need for water, it tends to rise more and more (MISHRA; SINGH et al., 2012).

Air temperature is an indicator that has the role of informing whether the air could be cold or hot on a numerical scale, and this variable has an impact not only on living things like plants and animals but also on a wide range of other meteorological indicators, such as wind speed, precipitation, and relative humidity (PENG CHEN et al., 2018). Therefore, such applicability further highlights the importance of understanding the behavior of temperature time series.

It is justified that climate change results in changes in meteorological dynamics, such as rainfall, air temperature, and relative air humidity (SANTOS et al., 2010).

Therefore, further investigations of statistical tools are required to improve decision-making in multiple areas. Therefore, the present article sought to analyze the behavior of the monthly average time series of Average Minimum Temperature, Average Maximum Temperature, Average Compensated Temperature, and Total Precipitation in the city of Belém (Pará, Brazil), for the production and application of forecast models.

METHODOLOGY

Study area

The capital of Pará, Belém (figure 1), has an estimated population of 1499641 people (IBGE, 2020), with an area of the territorial unit of 1059.466 km² (IBGE, 2019) and is located in the Amazon biome (IBGE, 2019). Belém is located in the equatorial zone, on the banks of Guajará Bay and Rio Guamá (DIAS; DA

CRUZ VALENTE; FERNANDES, 2020), which have their hydrodynamic, hydrobiogeochemical and physical-chemical aspects (temperature; pH; dissolved oxygen) under the seasonal influence of the rainfall, marine regime and the flow of affluent rivers (SARMENTO, 2019; SANTOS, 2019).

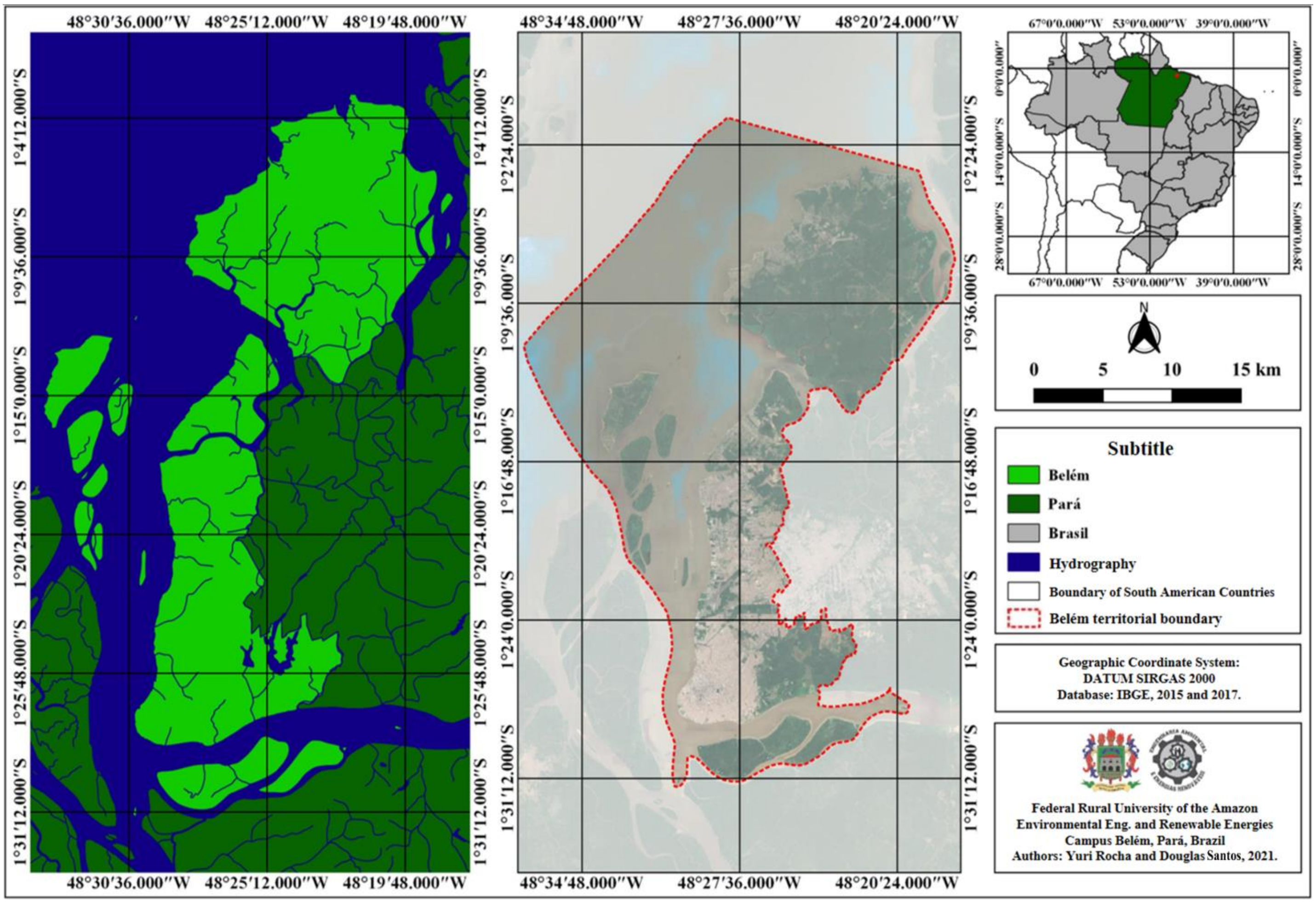

Figure 1 – Location map of the city of Belém (Pará, Brazil).

Source: The authors, 2021.

Climatological aspects of the study area

The climate of Belém is of the Af type, humid tropical (according to the Köppen classification), which points to a rainy or equatorial forest equatorial climate, characterized by its ample accumulated annual precipitation (DIAS; DA CRUZ VALENTE; FERNANDES, 2020). The annual seasonality of Belém is characterized by a rainy period – PC (January to May), the transition period from rainy to less rainy - TCM (June and July), less rainy – MC (August to October), and a transition interval from the minor rainy season to the rainy season – TMC (November and December) - and this seasonal dynamic also influences other atmospheric and climatological aspects, such as temperature and relative humidity.

These aspects can show that the seasonal dynamics of rainfall in Belém, in the course of the annual idiosyncrasy, raises the Intertropical Convergence Zone - ITCZ with mesoscale effects, sea breeze, local effects (such as land breezes) and trade winds being that such as the city has high temperatures, strong convection, unstable air and high humidity, favoring the formation of convective clouds (BASTOS et al., 2002).

Data acquisition

The data used for Average Compensated Temperature (°C), Average Maximum Temperature (°C), Average Minimum Temperature (°C) and Total Precipitation (mm) of Belém/PA, were acquired on the virtual portal of the National Meteorological Institute - INMET (<https://portal.inmet.gov.br/>), made available by the Meteorological Database for Teaching and Research (BDMEP), which owns all copyrights on them, such data being used solely for research by the present article.

Sample period

The 31-year time series was considered for the four variables (figure 2), from January 1990 to December 2020. In addition, the last four months (January, February, March and April) of the year 2021 were also considered, months available during the preparation of the work.

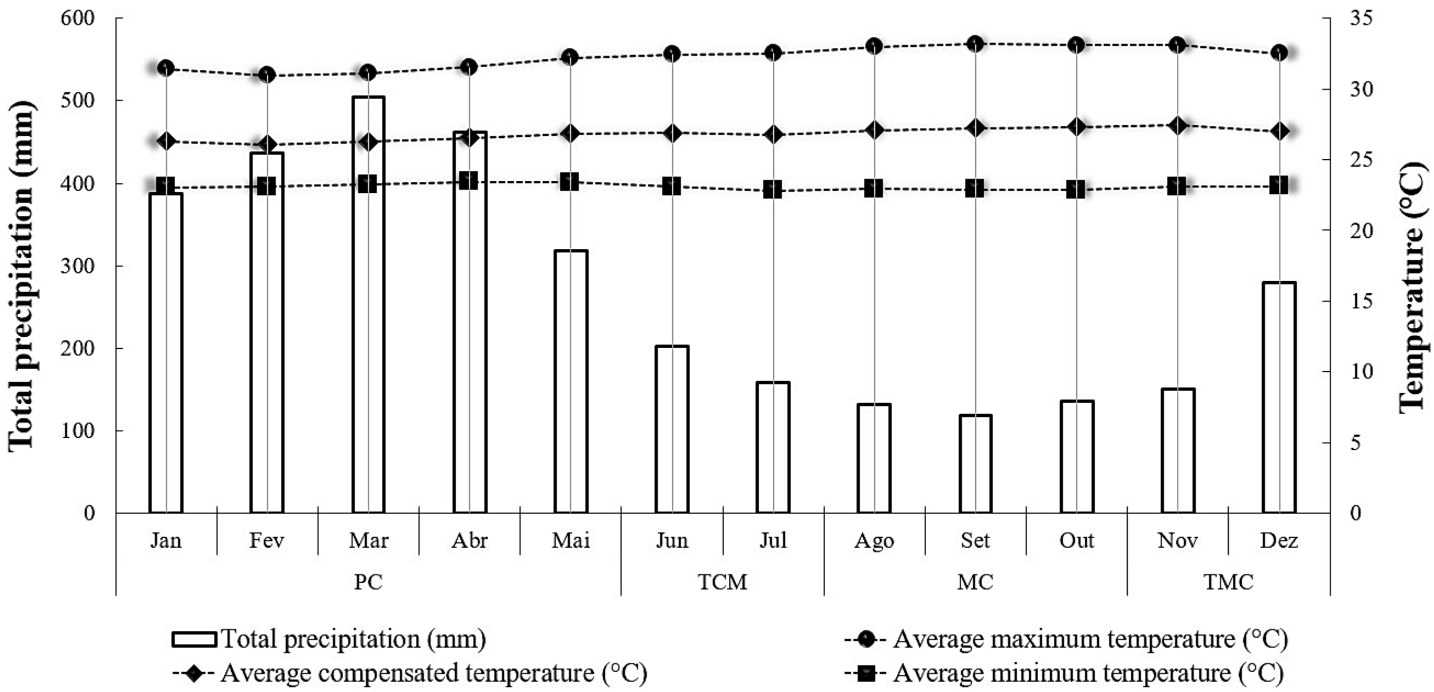

Figure 2 – Average monthly historical series (1990-2020) of precipitation (mm), average maximum temperature (°C), average compensated temperature (°C), and average minimum temperature (°C) of Belém/PA.

Source: INMET, 2020.

Prepared by: The authors, 2021.

In this period, the cumulative annual precipitation (historical series 1990 - 2020) presented an average value of 3287.58 ± 143.60 mm.year-1, ranging between 504.59 ± 133.64 mm and 119.11 ± 44.15 mm. The average maximum temperature varied between 33.19 ± 0.76 ° C and 30.97 ± 0.87 ° C, while the average compensated temperature varied between 27.41 ± 0.45 ° C and 26.07 ± 0, 51 ° C. The mean minimum temperature ranged from 23.44 ± 0.42 ° C and 22.81 ± 0.50 ° C.

Analytical procedures

First, the Box and Jenkins methodology was used to determine the best statistical models. Such a method can determine the appropriate model from a multi-stage filter. In dynamic model adjustments, theoretical analysis can provide us with a good estimate of the model to be used and provide good starting points for the numerical identification of its own parameters (BOX et al., 2015).

During the application of the methodology in question, models of the SARIMA and Holt-Winters class were estimated to analyze which would best apply in the series and produce efficient forecasts.

SARIMA Methodology

Several series have a seasonal behavior over the period studied, which repeatedly happens over some time. Moreover, following this perception, multiple generalized models by Box and Jenkins was created so that it was possible to deal with seasonality along with the trend, this model being known as SARIMA, which has operators of different orders (BOX; JENKINS, 1994) and it is responsible for dealing with these fluctuations that can happen in the time series over time (EHLERS, 2005). The equation says the SARIMA Model:

$$\varphi(B)\Phi(B^s)Wt = \theta(B)\Theta(B^s)\epsilon_t$$

According to Silva, Guimarães and Tavares (2008):

1. $\varphi(B)$ ... $(1 - \varphi B)$ is equivalent to the autoregressive variable of order p;
2. $\Phi(B^s)$ ... $(1 - \Phi B^s)$ equals the P autoregressive seasonal variable;
. W_t ... $(1 - B)(1 - B^s)$ is used to indicate simple and seasonal differentiation;
4. $\theta(B)$ ... $(1 + \theta B)$ equals the moving average variable of order q;
5. $\Theta(B^s)$ ... $(1 + \Theta B^s)$ equals the seasonal moving average variable of order Q;
6. ϵ_t is equivalent to a purely random process with zero mean and variance σ_e^2.

In this work, the seasonality of the series occurs annually.

For the selection of the best SARIMA model for the time series, one of the ways of choosing is by analyzing the Autocorrelation Correlograms (ACF) and Partial Autocorrelation (PACF) and analysis of AIC and BIC, the idea is based on selecting the model that has the lower information criterion.

It is possible to perform the calculation for the prediction of a model of the multiplicative seasonal type in a similar way to the model of the ARIMA type (p, d, q) in three ways: (1) Using the difference equation, (2) Using the form of random shocks and (3) Using the inverted form (MORETTIN & TOLOI, 2006). For the sake of understanding, one can take as an example the equation of a SARIMA model (0,1,1) x (0,1,1) 12 for an example of prediction, according to MORETTIN & TOLOI (2006):

$$(1 - B)(1 - B^{12})Zt = (1 - \theta B)(1 - \Theta B12)at,$$

And developing the present equation, we will have that the forecast at time t + h, according to MORETTIN & TOLOI (2006), will be:

$$Z_{t+h} = \mathrm{Z}_{t+h-1} + \mathrm{Z}_{t+h-12}\mathrm{Z}_{t+h-13} + a_{t+h} - \theta a_{t+h-1} - \Theta a_{t+h-12} + \Theta a_{t+h-13},$$

From this, the minimum NDE forecast, which was performed at the origin t, according to MORETTIN & TOLOI (2006), will be:

$$Z_t(\mathrm{h}) = [\mathrm{Z}_{\mathrm{t+h-1}}] + [\mathrm{Z}_{\mathrm{t+h-12}}] + \ldots + \theta\Theta \mathrm{a}_{\mathrm{t+h-13}}$$

Information Criterion

For the work in question, the Akaike Information Criterion (AIC), a tool proposed by Akaike (1974), and the Bayesian Information Criterion (BIC), a tool proposed by Schwarz (1978), were considered. The following equation represents the AIC:

$$AIC = -2\ log\ maximum\ likelihood + 2m,$$

The m is defined based on the number of variables in the model (autoregressive, moving averages, seasonal autoregressive, and seasonal moving averages). The BIC, on the other hand, is known to penalize the inclusion of extra parameters, even more, being its formula:

$$BIC = -2\ log\ maximum\ likelihood + m\ log\ n$$

The n refers to the number of observations in the considered time series. Furthermore, the information criteria will increase as the model's variables increase. Therefore, the main idea is to select the model SARIMA with the fewest possible variables.

Holt-Winters Methodology

The Holt-Winters methodology was used in the present work. This method is an improvement made by Winters (WINTERS, 1960) based on the work of Holt (HOLT, 2004) to be able to work with time series that may have seasonality and tendency (LIMA et al., 2015). And in this article it was used the exponential smoothing model of the Holt-Winters additive type in the precipitation time series. The equation for additive seasonality is said, according to MORETTIN & TOLOI (2006), by:

$$Z_t = \mu_t + T_t + F_t + a_t,$$

Prediction models can be developed considering an additive or multiplicative seasonal effect. It is important to emphasize that if the amplitude of the seasonal pattern is characterized as independent of the level, an additive type model may fit better (WINTERS, 1960). It is worth mentioning that when using the multiplicative effect, it is considered that the seasonal pattern is concordant in size at the local seasonally adjusted average level (CHATEFIELD, 1978). Such model for additive seasonal series is explained from three variables $\bar{Z}_t$, Tt and Ft, the level, trend and seasonal index at time t respectively, and for such variables smoothing constants A, C and D are estimated (MORETTIN & TOLOI, 2006). For the series in question, the equations for additive seasonality were adopted, which are said by the following equations according to MORETTIN & TOLOI (2006):

$$\hat{F}_t = D(Z_t - \bar{Z}_t) + (1 - D)\hat{F}_{t-s},\ 0 < D < 1$$
$$\bar{Z}_t = A(Z_t - \hat{F}_{t-s}) + (1 - A)(\bar{Z}_{t-1} + \hat{T}_{t-1}),\ 0 < A < 1$$
$$\hat{T}_t = C(\bar{Z}_t - \bar{Z}_{t-1}) + (1 - C)\hat{T}_{t-1},\ 0 < C < 1$$

As for the forecast case, for certain h periods ahead of the series, the equation used, according to MORETTIN & TOLOI (2006), is said by:

$$\hat{Z}_t(\mathrm{h}) = \bar{Z}_t + \mathrm{h}\hat{T}_t + \hat{\mathrm{F}}_{\mathrm{t+h-s}}, h = 1, 2, \dots, s,$$
$$\hat{Z}_t(\mathrm{h}) = \bar{Z}_t + \mathrm{h}\hat{T}_t + \hat{\mathrm{F}}_{\mathrm{t+h-2s}}, h = s+1, \dots, 2s,$$

The Holt-Winters Additive model can also adjust its predictions as new data are also emerging to the SARIMA model, with the adjustment equation for $\hat{F}_t, \bar{Z}_t \, e \, \hat{T}_t$, according to MORETTIN & TOLOI (2006), said by:

$$\hat{F}_{t+1} = D(Z_{t+1} - \bar{Z}_{t+1}) + (1 - D)\hat{F}_{t+1-s},$$
$$\bar{Z}_{t+1} = A(Z_{t+1} - \hat{F}_{t+1-s}) + (1 - A)(\bar{Z}_t + \hat{T}_t),$$
$$\hat{T}_{t+1} = C(\bar{Z}_{t+1} - \bar{Z}_t) + (1 - C)\hat{T}_t,$$

And finally, the new forecast for the value said Z_{t+h}, according to MORETTIN & TOLOI (2006), will be said as:

$$\hat{Z}_{t+1}(h-1) = \bar{Z}_{t+1} + (h-1)\hat{T}_{t+1} + \hat{F}_{t+1+h-s}, \quad h = 1, \dots, s+1,$$
$$\hat{Z}_{t+1}(h-1) = \bar{Z}_{t+1} + (h-1)\hat{T}_{t+1} + \hat{F}_{t+1+h-2s}, \quad h = s+2, \dots, 2s+1,$$

ACF and PACF correlograms

As for the analysis of the residues of the selected models, one of the tests used was the correlogram of the ACF and PACF, whose objective is aimed at verifying the existence of a correlation between the residues, whose equation used for the ACF test is:

$$r_k = \frac{\sum_{t=1}^{n-k}(x_t - \bar{x})(x_{t+k} - \bar{x})}{\sum_{t=1}^{n}(x_t - \bar{x})^2},$$

From the equation, the confrontation of values of $\hat{r}_k$ with the limits of $\pm 2/\sqrt{n}$ will be able to inform a general indication of the possibility of breaking the white noise paper in a_t (MORETTIN & TOLOI, 2006). In addition, the present work also used the partial autocorrelation function, dictated by the following equation:PACF:

$$(k) = Corr(X_{k+1} - P_{\overline{sp}\{1, X_1, \dots, X_k\}} X_{k+1}, X_1 - P_{\overline{sp}\{1, X_1, \dots, X_k\}} X_1 \qquad k \geq 2,$$

In this equation, the variable $\alpha(k)$is known as the partial autocorrelation in lag k (BROCKWELL & DAVIS, 1991).

Ljung-Box Test

In addition, a second test on the residues known as Ljung-Box was used, the equation being proposed by Ljung & Box (1978). The formula used is said by:

$$\tilde{Q} = n(n+2)\sum_{k=1}^{K}(n - k)^{-1} \, r_k^2(\widehat{a}),$$

From it, it is expected that the modified statistic will have, in an approximate way, the average $\mathrm{E}[\tilde{\mathrm{Q}}] \approx \mathrm{K} - \mathrm{p} - q$ with an X^2 $(K - \mathrm{p} - \mathrm{q})$ distribution (BOX et al, 2015).

Normality analysis

Then, the last test performed on the residues was to verify the normality of the residues. For this procedure, plots such as the histogram were plotted to verify that most values are present in the average value of all of them and adapt to the normality curve created. Furthermore, the quantile-quantile graph was generated to verify whether the model residues behave within a generated line and if they have less expressive "tails". It is noteworthy that the term tails refers to the points at the beginning and end of the line generated in the graph that tends to move away from the line, whose general purpose is to check if this distance is not so significant. Finally, a Shapiro-Wilk test was performed, such an equation was proposed by Shapiro & Wilk (1965). The tool of the present test is said by:

$$W = \frac{R^2\hat{\sigma}^2}{C^2S^2}$$

After verifying the model's residuals, they were superimposed with their respective time series to find out if they were not moving away from the actual data. Then, a forecast test was carried out with the last 12 months of the four series to check if the forecasts were approaching the original data. For this, a comparative and accuracy test was carried out.

The free software R (R Core Team (2019)) was used to perform the statistical tests used in this article, where the packages "forecast", "fBasics", "astsa" and "lmtest" were used to make the adjustments and analyzes necessary and generate forecasts.

RESULTS AND DISCUSSION

Statistics of the original data

Basic statistical tests (Table 1) were applied to all the time series to obtain the first understanding of the data and how they behaved in these thirty-one years and four months considered. The maximum and minimum value of the data was considered to understand the maximum and minimum point of oscillation of the data. With the median, it was possible to visualize the midpoint of all, but the main one was the Coefficient of Variation (C.V.), used to understand the "leakage" of the data concerning the average. From this, it is possible to verify low C.V for the first three series. However, in the precipitation, it is possible to observe a more significant oscillation in the series; this fact is justifiable due to the interval of the time series values.

Table 1 – General Data Statistics

Tests	Compensated temperature (°C)	Minimum temperature (°C)	Maximum temperature (°C)	Total precipitation (mm)
Minimum value	25,332857	21,667742	29,400000	8,200000
Maximum value	28,350000	24,680000	35,358065	931,100000
Median	26,814667	23,104947	32,293334	241,650000
Average	26,807665	23,083195	32,262825	276,015691
Standard deviation	0,635455	0,507094	1,110715	163,320000
C. V.	2,370422788	2,196810277	3,442708442	59,17054911

Source: The authors, 2021.

Decomposition of the original time series

Soon after, the series was decomposed (Figures 3, 4, 5, and 6) so that it could be better to observe the components of trend, seasonality, randomness, and the data itself varying over time.

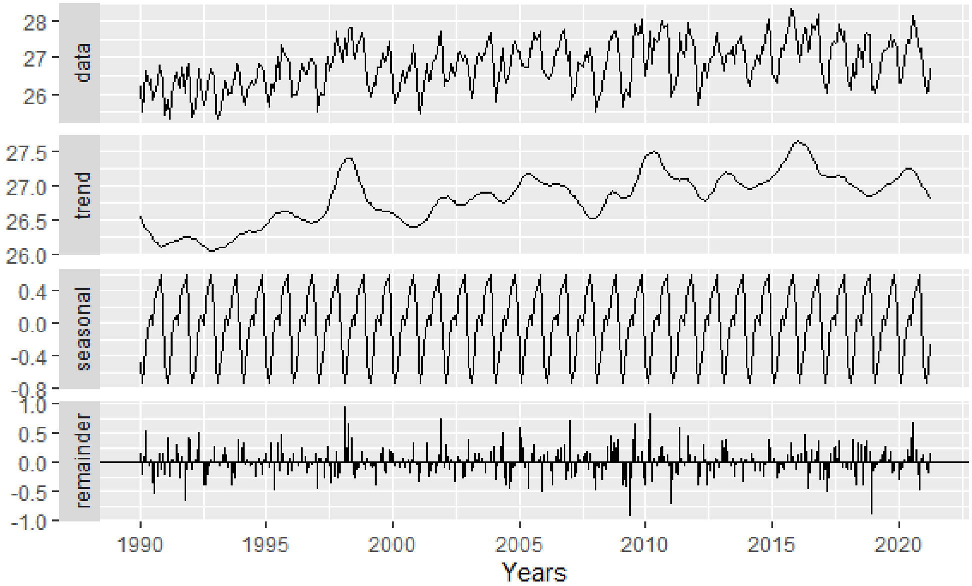

Figure 3 – Decomposition of the Average Compensated Temperature (°C) Time Series.

Source: The authors, 2021.

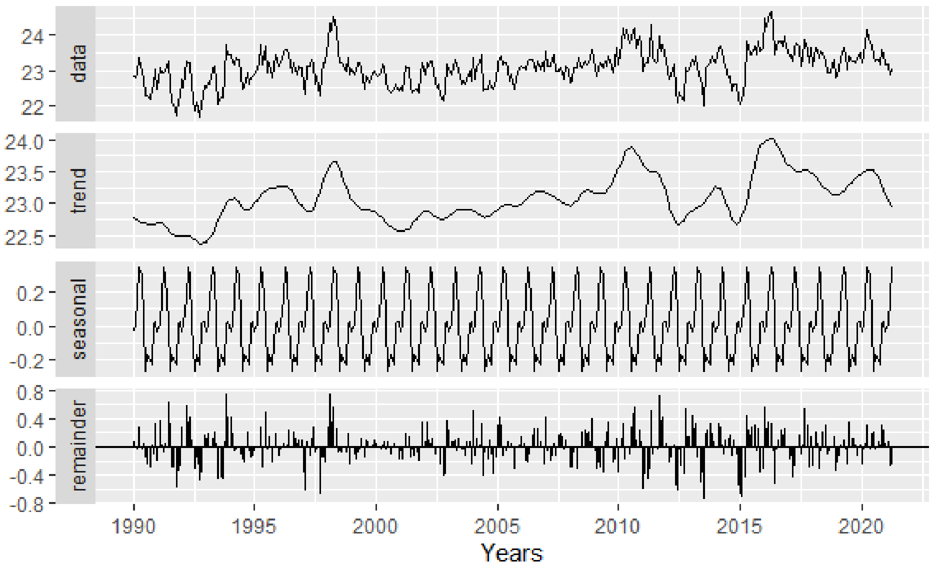

Figure 4 –Decomposition of the Average Minimum Temperature (°C) Time Series.

Source: The authors, 2021.

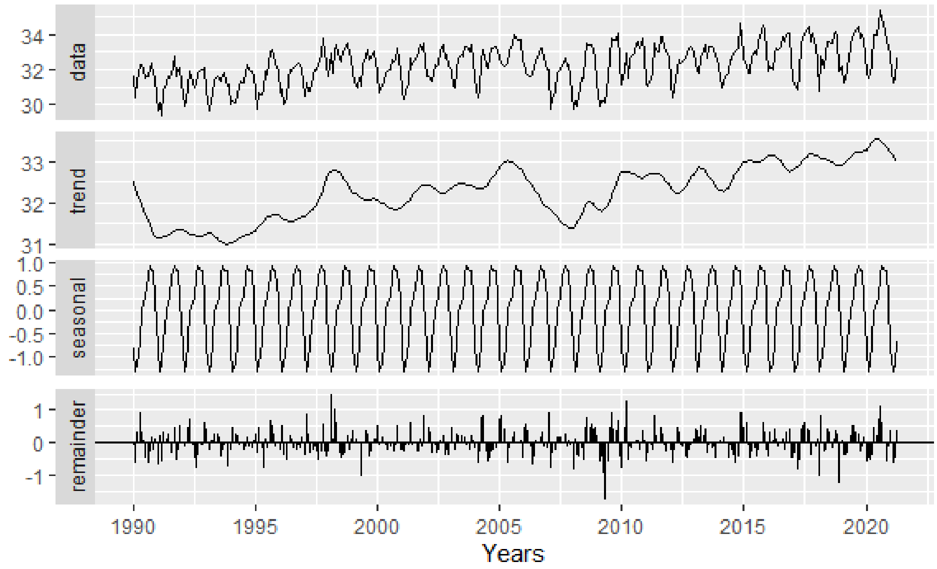

Figure 5 –Decomposition of the Maximum Average Temperature (°C) Time Series.

Source: The authors, 2021.

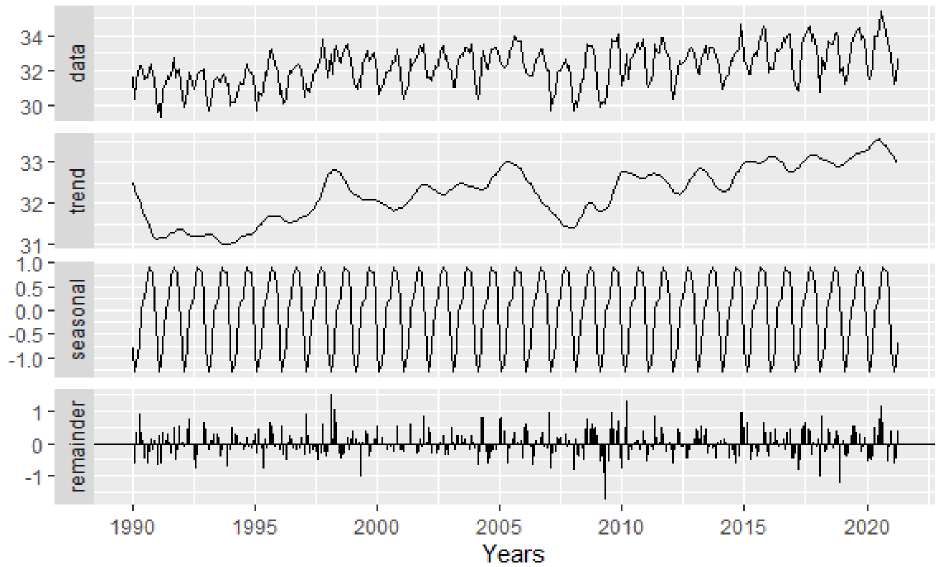

Figure 6 –Decomposition of the Total Monthly Precipitation (mm) Time Series.

Source: The authors, 2021.

Selected statistical models

As previously mentioned, the series was divided into 364 pieces of data, which were used for tests to find the model; and the last twelve were used to analyze the forecast. The evaluations made it possible to find three SARIMA models (Table 2) and one Holt-Winters model (Table 3) that had an excellent ability to adjust the residuals and made predictions close to the actual data.

Table 2 – Characteristics of the Chosen Models

Time series	Model	Coefficients	Standard Error	Critério de Informação
Average compensated temperature (°C)	SARIMA(0,1,2)(1,1,0)[12]	θ = -0.4296 θ = -0.1882 Φ = -0.5094	θ = 0.0545 θ = 0.0591 Φ = 0.0470	AIC = 327.75 BIC = 343.19
Average minimum temperature (°C)	SARIMA(1,1,1)(0,0,2)[12]	φ = 0.7317 θ = -0.9832 Θ = 0.1328 Θ = 0.0954	φ = 0.0395 θ = 0.0110 Θ = 0.0550 Θ = 0.0472	AIC = -2073.89 BIC = -2054.41
Average maximum temperature (°C)	SARIMA(0,1,2)(1,1,0)[12]	θ = -0.4502 θ = -0.1666 Φ = -0.4772	θ = 0.0527 θ = 0.0593 Φ = 0.0479	AIC = 632.8 BIC = 648.24

Fonte: Os autores, 2021.

Table 3 – Holt Winters Model

Time series	Model	Coefficients	Information criteria	MSE
Total precipitation (mm)	Holt-Winters with additive seasonality	Alfa (α): 0.0754911638 Beta (δ): 0.0001000030 Gamma (γ): 0.0001000474	AIC = 2852.235 BIC = 2918.487	6.329684

Fonte: Os autores, 2021.

For the SARIMA models, the "auto.arima" function was used. It selects the most suitable ARIMA model from the best values of AIC and BIC (HYNDMAN et al., 2021). However, it was necessary to adjust 1 non-seasonal differentiation for Compensated and Maximum Temperature so that the tests on residues were within the necessary standards. In addition, for Minimum Temperature it was necessary to perform the transformation of the series. After the tests had been done on it, outliers were observed, which could compromise the quality of the adjustment of the model residues, mainly the normalization of themselves. From there, the logarithmic transformation of the same was carried out.

Subsequently, tests were carried out on the Total Precipitation series using the “auto.arima” function, however it was not possible to find a model based on the same that would adjust the residuals to the standard necessary for subsequent series overlapping and forecasting. Bearing this in mind, the Holt-Winters methodology was applied to the total precipitation time series through the “hw” function of the R software forecast package using the “NULL” command for the smoothing constant variables (Alpha, Beta and Gamma), and the “optimal” command was considered for the “initial” argument. This command is used to select the initial state values from their optimization with the smoothing parameters through the exponential smoothing state space model, “ets” function (HYNDMAN et al., 2021). Finally, when making the forecasts, it was seen that in August and November 2020, the actual data exceeded the limit of 95% lower and higher, respectively, however, in the other months the forecast was within limits and close to the actual data. For the test in the Holt-Winters model, it was necessary to perform the square root transformation in the original total precipitation time series to adjust the residues. In addition, the information criterion was evaluated and the sum of squares of errors was evaluated.

Result of correlograms

After choosing the models, the ACF and PACF correlogram test was first performed (Figures 7, 8, 9, and 10) for each of them, where it was noted that most of the LAGS had correlations close to zero, passing slightly from the limit a few times, denoting that the series are random.

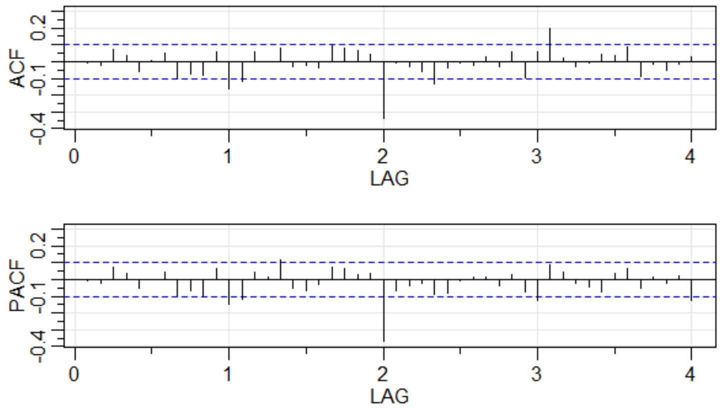

Figure 7 – Average Compensated Temperature ACF and PACF graphs, after 1 simple differentiation and 1 seasonal differentiation.

Source: The authors, 2021.

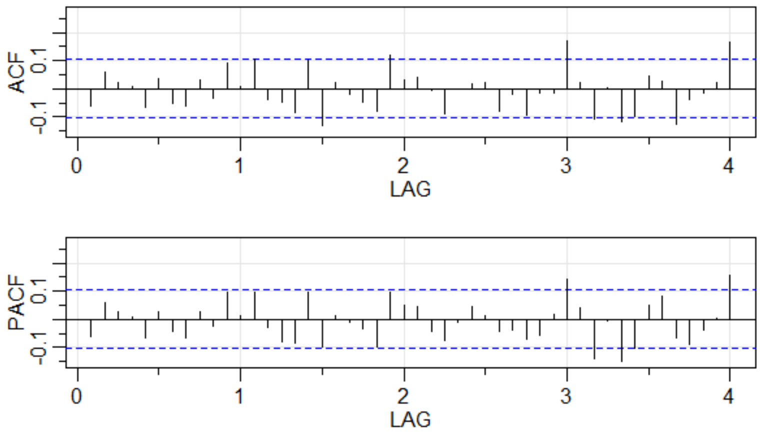

Figure 8 – ACF and PACF plots of Average Minimum Temperature, after 1 simple differentiation.

Source: The authors, 2021.

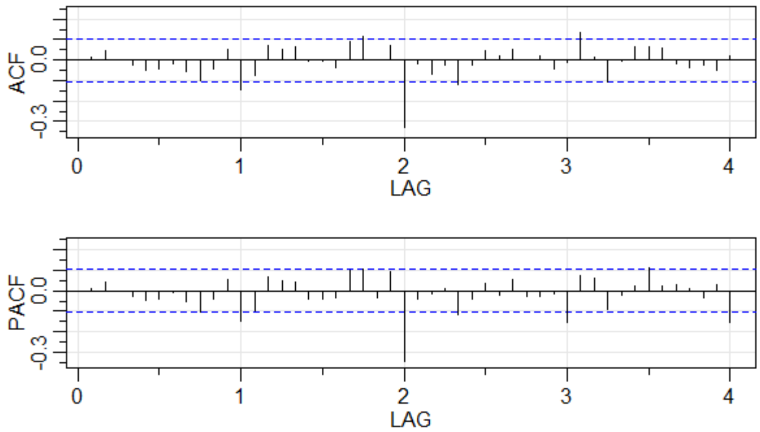

Figure 9 – Charts of ACF and PACF of Maximum Average Temperature, after 1 simple differentiation and 1 seasonal differentiation.

Source: The authors, 2021.

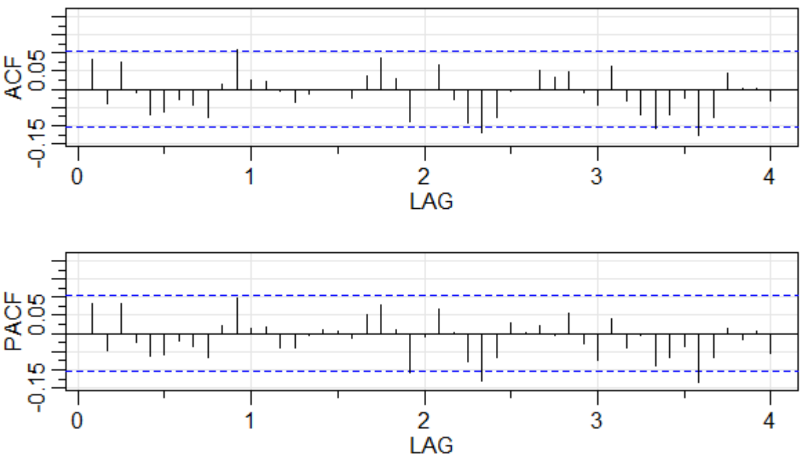

Figure 10 – ACF and PACF plots of Total Precipitation.

Source: The authors, 2021.

Ljung-Box test result

The Ljung-Box test (table 4) was then performed with p values considered up to LAG 10. For this step, all models successfully had p-values above α = 0.05, the usual significance level adopted for the present work. It is worth mentioning that, according to BOX et al. (2015), in addition to considering the $r_k(\hat{a})'s$ individually, it is often necessary to look at whether the first 10-20 autocorrelations of the $\hat{a}_t$ point out the inadequacy of the model.

Table 4 – Ljung-Box test.

Lag	Compensated temperature	Minimum temperature	Maximum temperature	Total Precipitation
1	0,8659596	0,2491515	0,8281611	0,1181128
2	0,9205506	0,2605253	0,6698201	0,2172545
3	0,5263067	0,4152187	0,8489793	0,1705027
4	0,6112497	0,5766696	0,9054526	0,2831883
5	0,565595	0,4915774	0,8581444	0,2329022
6	0,6885933	0,557314	0,8595163	0,2174984
7	0,6862264	0,5442932	0,9141013	0,2867059
8	0,3458412	0,4986679	0,8688544	0,3259196
9	0,2867442	0,5655185	0,5751556	0,2503889
10	0,201488	0,6166538	0,6053071	0,3221984

Source: The authors, 2021.

Result of normality

Finally, the last step of the tests on the residues refers to checking their normality and, as mentioned previously, visual tests were performed with Histograms with the normal curve and quantile-quantile graphs, in addition to the usual Shapiro-Wilk test with $W_\alpha = 0.05$ (Table 5), and for all models normality was present.

Table 5 – Shapiro-Wilk Test.

Shapiro-Wilk Test	Compensated temperature	Minimum temperature	Maximum temperature	Total precipitation
p-valor	0,7144	0,05845	0,09812	0,2686

Source: The authors, 2021.

Superposition of time series

Then, graphs were generated (Figures 11, 12, 13, and 14) where the time series of the real data and the model built were superimposed. Comparing the original series and the SARIMA and Holt-Winters models showed that they managed to follow the original data well, following their trend and seasonality.

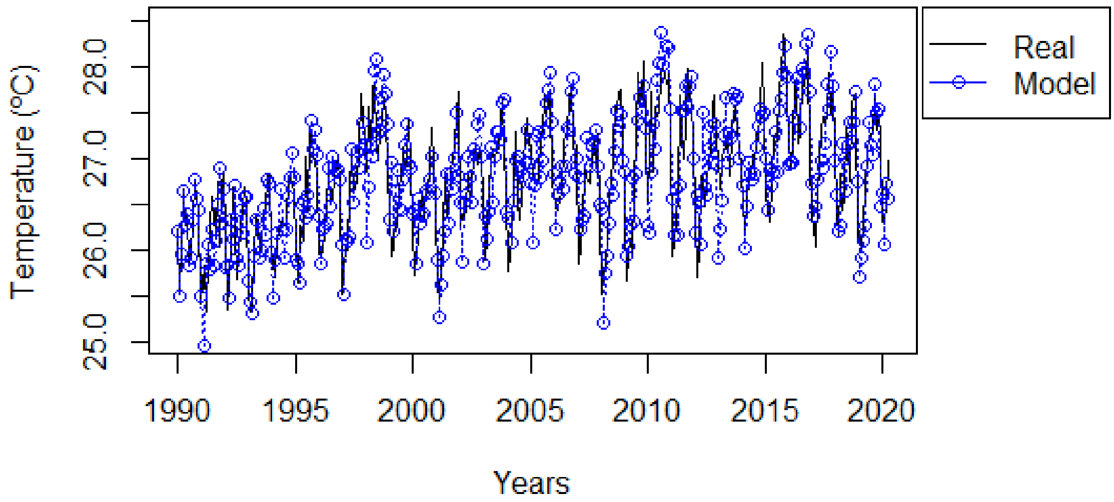

Figure 11 – Superimposition of the Average Compensated Temperature Time Series with their model.

Source: The authors, 2021.

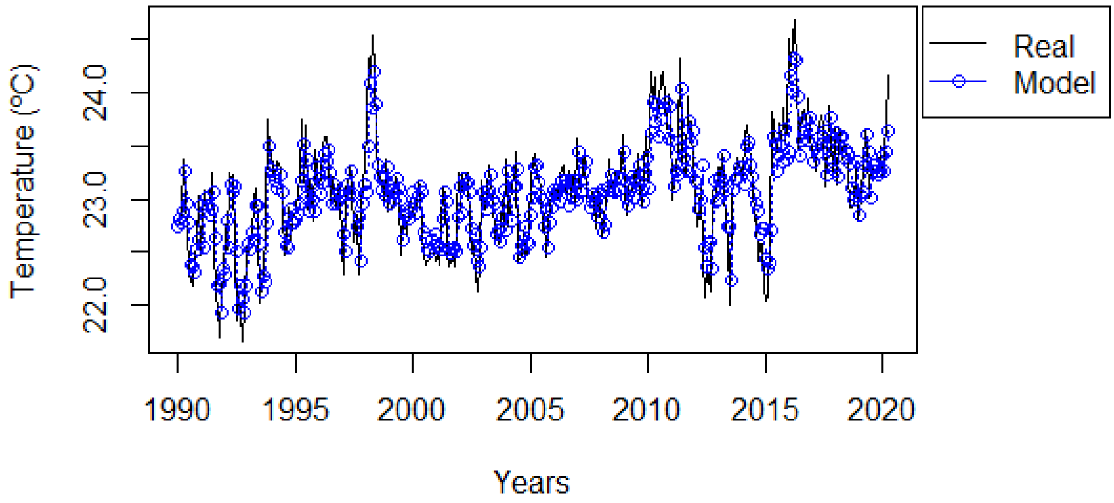

Figure 12 – Superimposition of the Average Minimum Temperature Time Series with their model.

Source: The authors, 2021.

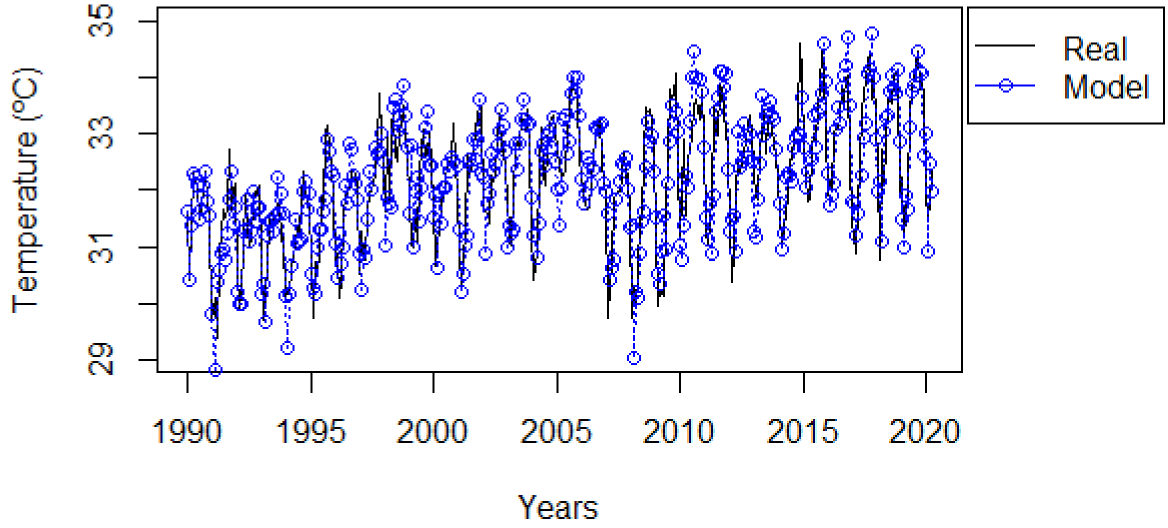

Figure 13 – Superposition of the Maximum Average Temperature Time Series with their model.

Source: The authors, 2021.

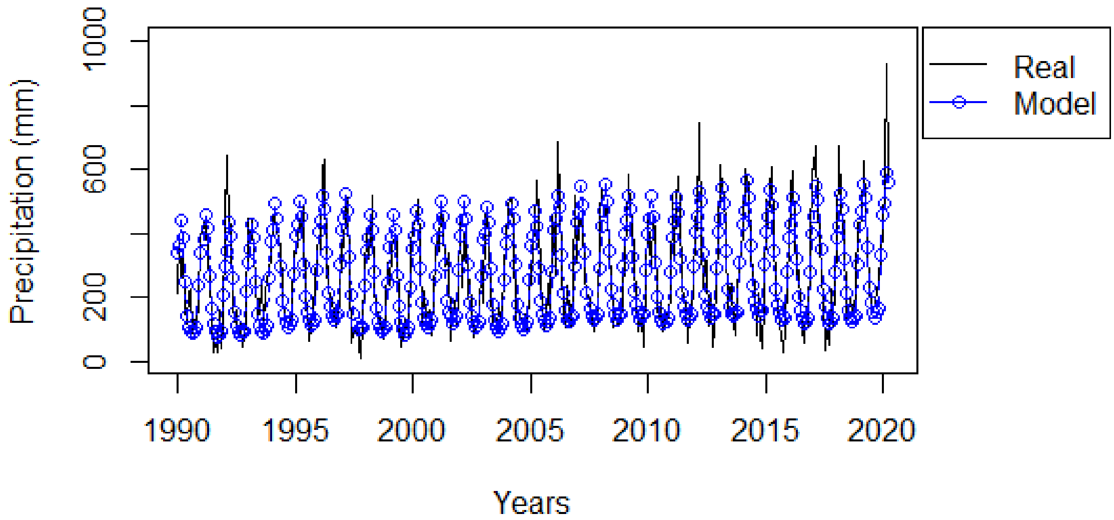

Figure 14 – Superimposition of the Total Monthly Precipitation Time Series with their model.

Source: The authors, 2021.

Accuracy test results

Accuracy tests were performed to verify the prediction's quality with the validation time, which refers to the last thirteen months (table 6). The tests were based on the Root Mean Square Error (RMSE) and the Mean Absolute Percentage Error (MAPE); the test statistics are based on the following assumption: the lower the value of the calculation found, the better the quality of the predictions. The equations used are based on the following formulas:

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

and

$$MAPE = \frac{\sum_{t=1}^{n} \left| \frac{(y_t - \hat{y}_t)^2}{y_t} \right|}{n} * 100,$$

where y_t equals the original series, $\hat{y}_t$ equals the forecast data and n equals the amount of data in the original series.

Table 6 – Values of the accuracy tests for the predictions.

	Compensated temperature	Minimum temperature	Maximum temperature	Total precipitation
RMSE	0.4294168	0.2939587	0.5737297	137.2083
MAPE	1.316479	0.8851493	1.438552	39.7

Source: The authors, 2021.

Result of forecasts

Next, the forecast data were tabulated along with each time series's validation time (tables 7, 8, 9 and 10) and their respective graphs were generated (figures 15, 16, 17 and 18). Besides, the lower and upper limits of 95% were included, showing that, although a prediction did not fall exactly close to the real value, it is included within the forecast limits for the model considered in the present article. From the values found, it was noted that the forecast came close to the actual data, especially in the temperature series, erring approximately 1ºC above or below in some months. Similar forecast values were found in the study by Barbosa et al. (2015) for average temperature using monthly data and the SARIMA model.

In another study, Peng Chen et al. (2018) in turn used the SARIMA model (1.1.1) (1.0.1) [12] for average monthly temperature in Nanjing, located in China, which was able to adjust to the tests on residues adequately and made good predictions when compared to the real data, showing once again the good functioning of models of the SARIMA class for monthly temperature time series.

Similar results were also found in the article by Asamoah-Boaheng (2014) using a SARIMA model (2,1,1) (1,1,2) [12] for average monthly temperature in Ashanti, region of Ghana, from the analysis of AIC, AICc and BIC to select it from a series of neighboring models and together with the analysis of residues from tests such as ACF, Ljung-Box and QQ Plot.

Unlike the use of Holt-Winters in the present article, the work of Afrifa-Yamoah, Saeed and Karim (2016) used a SARIMA model (0,0,0) (1,1,1) [12] to predict monthly rains in Brong Ahafo in the Ghana region. Besides, SARIMA model was able to adapt to the tests carried out, and in its forecast, the model was well adapted to the real data with just one month, the original data exceeded the upper limit of 95%.

Table 7 – Forecast for Average Compensated Temperature (°C).

	Forecast (°C)	Real temperature (°C)	Lower limit 95%	Upper limit 95%
May/20	27,10757	27,48710	26,36070	27,85444
Jun/20	27,48668	27,42000	26,62685	28,34651
Jul/20	27,40648	27,65548	26,50050	28,31246
Aug/20	27,21903	28,16645	26,26914	28,16891

Sep/20	27,80404	27,80000	26,81218	28,79589
Oct/20	27,70770	27,38200	26,67559	28,73982
Nov/20	27,75918	27,18000	26,68831	28,83005
Dec/20	26,87665	27,26710	25,76838	27,98491
Jan/21	26,46712	26,62839	25,32268	27,61155
Feb/21	26,35608	26,02143	25,17658	27,53558
Mar/21	26,60288	26,09226	25,38932	27,81643
Apr/21	27,02458	26,69000	25,77791	28,27126

Fonte: The authors, 2021.

Table 8 – Forecast for Minimum Average Temperature (°C).

	Forecast (°C)	**Real temperature (°C)**	**Lower limit 95%**	**Upper limit 95%**
May/20	23,71190	23,77419	23,24632	24,18680
Jun/20	23,63157	23,73333	23,08466	24,19143
Jul/20	23,39578	23,46452	22,78346	24,02456
Aug/20	23,25195	23,25806	22,58054	23,94332
Sep/20	23,32022	23,29667	22,58943	24,07465
Oct/20	23,29438	23,16000	22,51141	24,10459
Nov/20	23,32200	23,60000	22,48856	24,18632
Dec/20	23,45658	23,47742	22,57144	24,37643
Jan/21	23,45049	23,17419	22,52118	24,41815
Feb/21	23,43271	23,14286	22,46186	24,44552
Mar/21	23,48888	22,88387	22,47523	24,54825
Apr/21	23,64533	23,01667	22,58584	24,75453

Fonte: Os autores, 2021.

Table 9 – Forecast for Maximum Average Temperature (°C).

	Forecast (°C)	**Real temperature (°C)**	**Lower limit 95%**	**Upper limit 95%**
May/20	33,29111	33,96452	32,13696	34,44525
Jun/20	33,91468	33,92333	32,59757	35,23179
Jul/20	33,95086	34,54194	32,56145	35,34027
Aug/20	34,07800	35,35807	32,61987	35,53613
Sep/20	34,38855	35,02667	32,86480	35,91230
Oct/20	34,11506	34,20000	32,52839	35,70172
Nov/20	34,12065	33,59667	32,47348	35,76783
Dec/20	33,01640	33,40968	31,31086	34,72194

Jan/21	32,20200	32,79032	30,44003	33,96398
Feb/21	31,57707	31,21786	29,76041	33,39372
Mar/21	31,74316	31,42581	29,87342	33,61289
Apr/21	32,37835	32,67667	30,45700	34,29971

Fonte: The authors, 2021.

Table 10 – Forecast for Total Monthly Precipitation (mm).

	Forecast (°C)	**Real temperature (°C)**	**Lower limit 95%**	**Upper limit 95%**
May/20	398,4725	444,4	222,56601	625,2453
Jun/20	268,0012	273,9	127,98907	459,1704
Jul/20	213,8831	133,8	91,25675	387,9578
Aug/20	186,9595	54,3	73,73709	351,9226
Sep/20	168,1070	146,6	61,85736	326,3902
Oct/20	181,2138	264,0	69,66454	345,0904
Nov/20	190,3375	502,1	75,11466	358,1822
Dec/20	357,9636	213,6	189,78202	579,0623
Jan/21	484,5113	328,3	284,03917	738,1965
Feb/21	529,3443	614,1	318,08581	794,1127
Mar/21	618,8983	423,1	387,72683	903,8774
Apr/21	568,1491	501,2	347,24945	843,1548

Fonte: The authors, 2021.

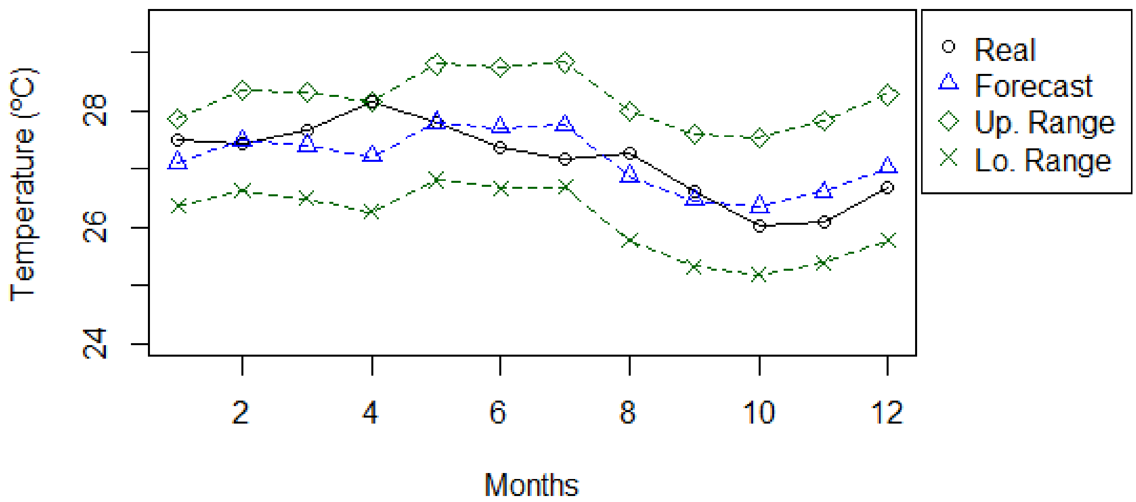

Figure 15 – Comparison of the forecast with actual values for Average Compensated Temperature (°C).

Source: The authors, 2021.

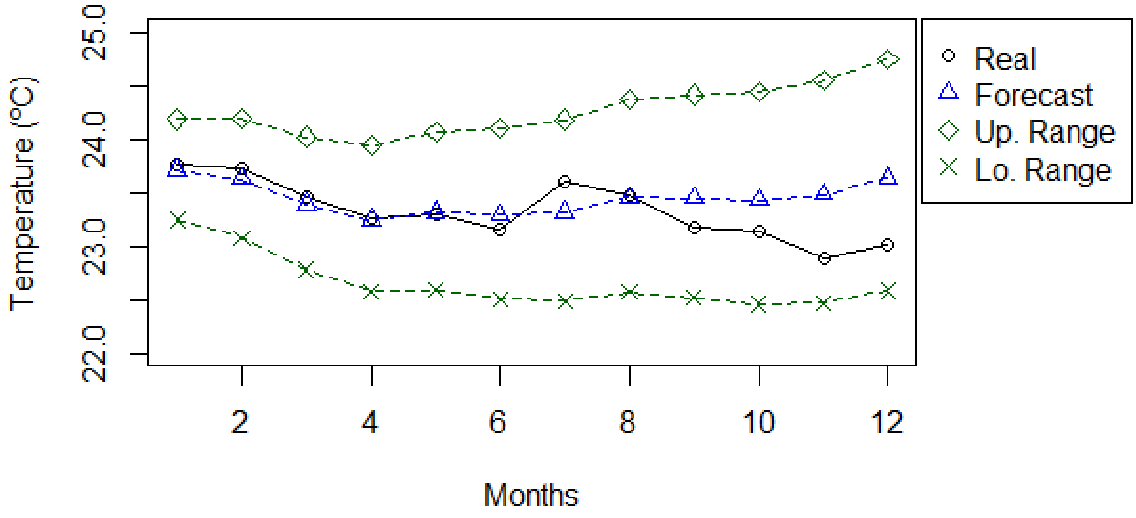

Figure 16 – Comparison of the forecast with actual values for Minimum Average Temperature (°C).

Source: The authors, 2021.

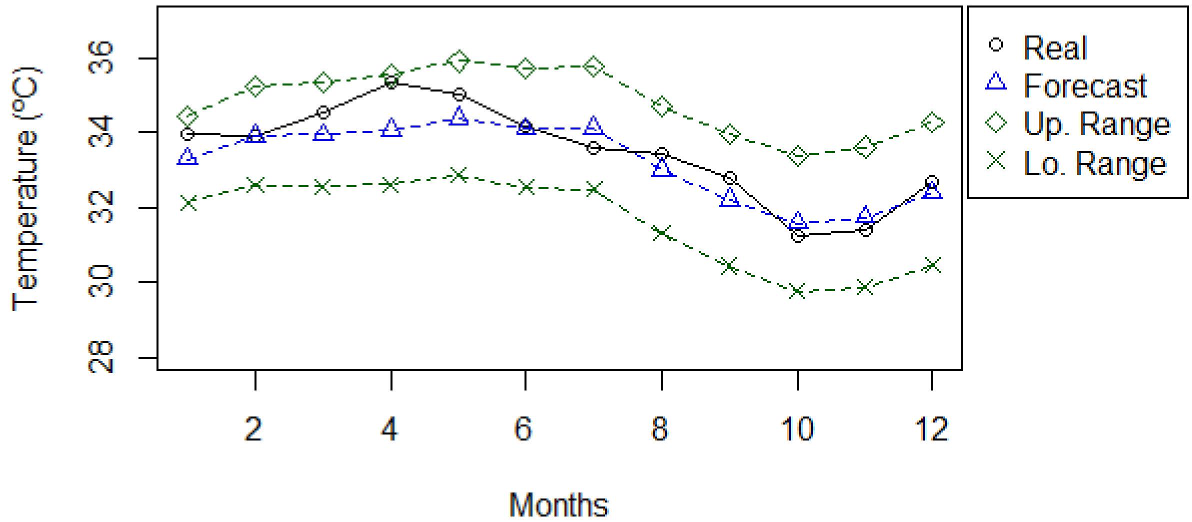

Figure 17 – Comparison of the forecast with actual values for Maximum Average Temperature (°C).

Source: The authors, 2021.

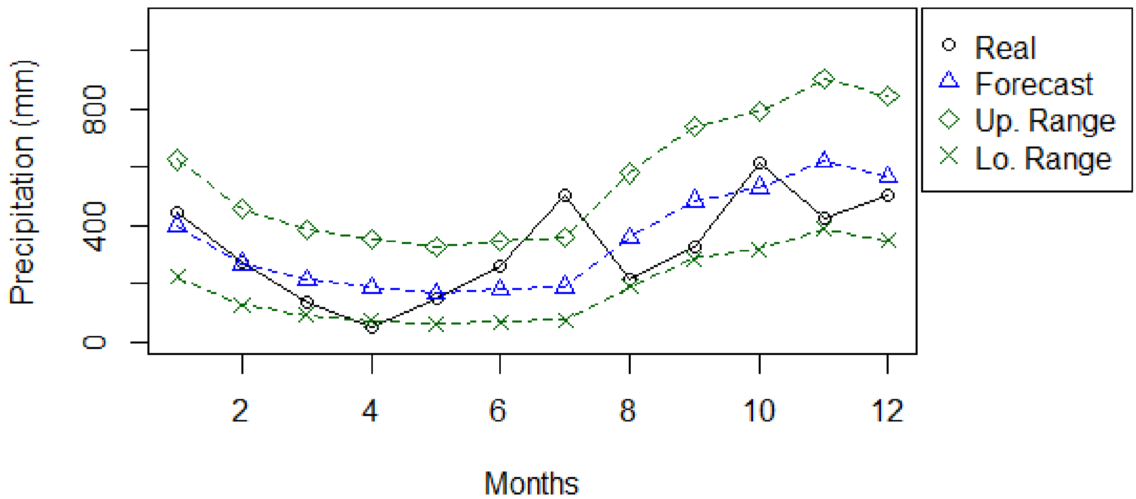

Figure 18 – Comparison of the forecast with actual values for Total Monthly Precipitation (mm).

Source: The authors, 2021.

CONCLUSION

The present article found forecast models that apply well to the original time series, managing to make predictions close to the real data for the four-time series worked from SARIMA and Holt-Winters models. Furthermore, as well as the real data, the forecasts showed the same seasonality, with few divergence values.

Thus, considering that the city of Belém has a very accentuated climatological seasonality, the use of models presented here can be used in the local climatological studies of the city. The use of the model can indicate how the following periods of the subsequent months that constitute the seasonal periods will be (rainy season; transition from rainy season to minor rainy season; less rainy season; transition from minor rainy season to the rainy season). The use of such statistical models will allow: the collective and individual urban transport activities in areas adjacent to the water bodies (Guamá River and Guajará Bay) of the city to have more effective traffic management; there is a public risk alert for the movement of riverside dwellers from the island region of Belém to the city, as well as for commercial or tourist transport vessels and fishing activities; there is a public alert for local farmers about the climatological dynamics; there is an alert to public administration regarding urban and rural locations that are likely to be at greater risk of flooding, especially those places where few or a large number of people reside.

However, it is emphasized that the application of other possible statistical models should be considered to find the possibility of more accurate predictions.

REFERENCES

Akaike, Hirotugu. A new look at the statistical model identification. IEEE Transactions on Automatic Control, v. 19, n. 6, p. 716–723, 1974. DOI:10.1109/tac.1974.1100705.

Afrifa-Yamoah, Ebenezer; Saeed, Bashiru I. I.; Karim, Azumah. Sarima Modelling and Forecasting of Monthly Rainfall in the Brong Ahafo Region of Ghana. World environment. v. 6, p. 1-9, 2016. DOI: 10.5923/j.env.20160601.01.

Asamoah-Boaheng, Michael. Using SARIMA to forecast monthly mean surface air temperature in the Ashanti Region of Ghana. International Journal of Statistics and Applications. v. 4, n. 6, p. 292-299, 2014. DOI: 10.5923/j.statistics.20140406.06.

BARBOSA, E. C.; SÁFADI, T.; NASCIMENTO, M.; NASCIMENTO, A. C. C.; SILVA, C. H. O.; MANULI, R. C. Box & Jenkins methodology to forecasting of average monthly temperature of Bauru city (SP). Rev. Bras. Biom., São Paulo, v.33, n.1, p.104-117, 2015.

BASTOS, T. X.; PACHECO, N. A.; NECHET, D.; SÁ, T. D. A. Aspectos climáticos de Belém nos últimos cem anos. Embrapa Amazônia Oriental-Documentos (INFOTECA-E), 2002.

Box, George E. P.; Jenkins, Gwilym M.; Reinsel, Gregory C.; Ljung, Greta M.. Time series analysis: forecasting and control. John Wiley & Sons, 2015.

Box, George E. P.; Jenkins, Gwilym M.. Time series analysis: forecasting and control. HOLDEN-DAY, 1996.

Brockwell, Peter J.; Davis, Richard A.. Time series: theory and methods. 2ª ed. Springer Series in Statistics, 1991.

CHATFIELD, Chris. The Holt-winters forecasting procedure. Journal of the Royal Statistical Society: Series C (Applied Statistics), v. 27, n. 3, p. 264-279, 1978.

Chen, Peng; Niu, Aichen; Liu, Duanyang; Jiang, Wei; Ma; Bin. Time series forecasting of temperatures using SARIMA: An example from Nanjing. IOP Conference Series: Materials Science and Engineering 394 052024, v. 394, 2018. DOI:10.1088/1757-899X/394/5/052024.

DIAS, L. C.; DA CRUZ VALENTE, A. M.; FERNANDES, L. L. Análise e correlação de variáveis climatológicas com os fenômenos climáticos e a urbanização na Cidade de Belém, no Estado do Pará, região Norte do Brasil. Research, Society and Development, v. 9, n. 8, p. e972986790-e972986790, 2020.

Heydari, Mohammad; Ghadim, Hamed Benisi; Rashidi, Mahmood; Noori, Mohammad. Application of Holt-Winters Time Series Models for Predicting Climatic Parameters (Case Study: Robat Garah-Bil Station, Iran). Polish Journal of Environmental Studies, vol. 29, no. 1, 2020, pp. 617-627. DOI:10.15244/pjoes/100496.

Holt, Charles C. Forecasting seasonals and trends by exponentially weighted moving averages. International journal of forecasting, v. 20, n. 1, p. 5-10, 2004.

Hyndman R, Athanasopoulos G, Bergmeir C, Caceres G, Chhay L, O'Hara-Wild M, Petropoulos F, Razbash S, Wang E, Yasmeen F (2021). forecast: Forecasting functions for time series and linear models. R package version 8.14, https://pkg.robjhyndman.com/forecast/.

Instituto Nacional de Metereologia (INMET) – Banco de Dados Metereológicos do INMET. Disponível em: <https://bdmep.inmet.gov.br/>. Acesso em: 17 de fevereiro de 2021.

Instituto Brasileiro de Geografia e Estatística – IBGE. População: população estimada, Belém/PA. 2020. Disponível em: https://cidades.ibge.gov.br/brasil/pa/belem/panorama. Acesso em: 21/02/2021.

Instituto Brasileiro de Geografia e Estatística – IBGE. Território e Ambiente: Área da unidade territorial, Belém/PA. 2019. Disponível em: https://cidades.ibge.gov.br/brasil/pa/belem/panorama. Acesso em: 21/02/2021.

LIMA, Marcos Bruno Santos Pereira; Santos, Wagner Barbosa; Droguett, Enrique Lopez; Diniz, Helder Henrique Lima; Santos, Rita de Cassia Barbosa. Aplicação do modelo de previsão de demanda Holt-Winters em uma Regional de corte e dobra de aço. XXXV Encontro Nacional de Engenharia da Produção. Fortaleza-CE, 2015.

Instituto Brasileiro de Geografia e Estatística – IBGE. Território e Ambiente: Bioma, Belém/PA. 2019. Disponível em: https://cidades.ibge.gov.br/brasil/pa/belem/panorama. Acesso em: 21/02/2021.

MISHRA, Ashok K.; Singh, Vijay P.. Simulating hydrological drought properties at different spatial units in the United States based on wavelet–Bayesian regression approach. Earth Interactions. v. 16, n. 17, p. 1-23, 2012. DOI: 10.1175/2012EI000453.1

Morettin, P. A.; Toloi, C. M. C.. Análise de series temporais. 2ª ed. São Paulo: Egard Blucher, 2006.

R Core Team (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

SANTOS, D. N. dos; DA SILVA, V. de P. R.; SOUSA, F. de A. S.; SILVA, R. A. e. Estudo de alguns cenários climáticos para o Nordeste do Brasil. Revista Brasileira de Engenharia Agrícola e Ambiental, v. 14, n. 5, p. 492-500, 2010.

SANTOS, K. E. A. Dinâmica biogeoquímica do Estuário Guajarino. Trabalho de Conclusão de Curso (TCC). Graduação em Engenharia de Pesca. Universidade Federal Rural da Amazônia – UFRA, campus Belém/PA. 2019.

SARMENTO, I. C. C. Dinâmica do nitrato, amônio e nitrogênio total dissolvido no Estuário Guajarino. Trabalho de Conclusão de Curso (TCC). Graduação em Engenharia Ambiental e Energias Renováveis. Universidade Federal Rural da Amazônia – UFRA, campus Belém/PA. 2019.

Schwarz, Gideon. Estimating the dimension of a model. The Annals of Statistics. v. 6, n. 2, p. 461-464, 1978. DOI: 10.2307/2958889.

Shapiro, Samuel Sanford; Wilk, Martin B.. An analysis of variance test for normality (complete samples). Biometrika, v. 52, n. 3/4, p. 591-611, 1965. DOI:10.2307/2333709

SILVA, J. W. da; GUIMARÃES, E. C.; TAVARES, M. Variabilidade temporal da precipitação mensal e anual na estação climatológica de Uberaba-MG. Ciência e Agrotecnologia, v. 27, n. 3, p. 665-674, 2003.

SILVA, M. IS; GUIMARÃES, E. C.; TAVARES, M. Previsão da temperatura média mensal de Uberlândia, MG, com modelos de séries temporais. Revista Brasileira de Engenharia Agrícola e Ambiental, v. 12, n. 5, p. 480-485, 2008.

Winters, Peter R. Forecasting sales by exponentially weighted moving averages. Management science, v. 6, n. 3, p. 324-342, 1960. DOI: 10.2307/2627346

AUTHORS

First Author – Douglas Matheus das Neves Santos, Graduating in Environmental Engineering and Renewable Energies at the Federal Rural University of the Amazon - UFRA, campus Belém, Pará, Brazil. douglasneves23@hotmail.com

Second Author – Yuri Antônio da Silva Rocha, Graduating in Environmental Engineering and Renewable Energies at the Federal Rural University of the Amazon - UFRA, campus Belém, Pará, Brazil. yuriantonio2010@hotmail.com

Third Author – Danúbia Leão de Freitas, Graduating in Environmental Engineering and Renewable Energies at the Federal Rural University of the Amazon - UFRA, campus Belém, Pará, Brazil. nubiafreitas010@gmail.com

Fourth Author – Paulo Roberto Estumano Beltrão Júnior, Graduating in Environmental Engineering and Renewable Energies at the Federal Rural University of the Amazon - UFRA, campus Belém, Pará, Brazil. paulobeljr@gmail.com

Fifth Author – Paulo Cerqueira dos Santos Junior, PhD in Statistics, and professor at the Federal Rural University of the Amazon - UFRA, Belém campus, Pará Brazil, Ciberespacial Institute – ICIBE. paulo.cerqueira@ufra.edu.br

Sixth Author – Glauber Tadaiesky Marques, PhD in Physics, and professor at the Federal Rural University of the Amazon - UFRA, Belém campus, Pará Brazil, Ciberespacial Institute – ICIBE. glauber.marques@ufra.edu.br

Seventh Author – Otavio Andre Chase, PhD in Electrical Engineering, and professor at the Federal Rural University of the Amazon - UFRA, Belém campus, Pará Brazil, Ciberespacial Institute – ICIBE. otavio.chase@ufra.edu.br

Eighth Author – Pedro Silvestre da Silva Campos, PhD in Agricultural Sciences, and professor at the Federal Rural University of the Amazon - UFRA, Belém campus, Pará Brazil, Ciberespacial Institute – ICIBE. pedro.campos@ufra.edu.br

Correspondence Author – Douglas Matheus das Neves Santos, douglasneves23@hotmail.com, douglasneves233@gmail.com, +55(91)98379-6475.