Research, Society and Development, v. 11, n. 14, e406111436382, 2022
(CC BY 4.0) | ISSN 2525-3409 | DOI: http://dx.doi.org/10.33448/rsd-v11i14.36382

Artificial neural network applied to reported dengue cases in Maceió – Alagoas

Rede neural artificial aplicada aos casos notificados de dengue cases em Maceió – Alagoas

Red neuronal artificial aplicada a casos reportados de dengue en Maceió – Alagoas

Recebido: 11/10/2022 | Revisado: 25/10/2022 | Aceitado: 26/10/2022 | Publicado: 31/10/2022

Iwldson Guilherme da Silva Santos
ORCID: https://orcid.org/0000-0001-5663-9047
Universidade Federal de Campina Grande, Brasil
E-mail: iwldson@gmail.com
José Francisco de Oliveira Júnior
ORCID: https://orcid.org/0000-0001-5280-9600
Universidade Federal de Alagoas, Brasil
E-mail: jose.junior@icat.ufal.br
Isnaldo Isaac Barbosa
ORCID: https://orcid.org/0000-0003-3147-1780
Universidade Federal de Alagoas, Brasil
E-mail: isnaldo@pos.mat.ufal.br
Luis Felipe Francisco Ferreira da Silva
ORCID: https://orcid.org/0000-0001-8667-8100
Universidade Federal de Alagoas, Brasil
E-mail: luis.silva@icat.ufal.br
William Max de Oliveira Romão
ORCID: https://orcid.org/0000-0002-2204-3592
Universidade Federal de Alagoas, Brasil
E-mail: william.romao@icat.ufal.br
Vitória Rejane Marques dos Santos
ORCID: https://orcid.org/0000-0002-9043-3912
Universidade Federal de Alagoas, Brasil
E-mail: vitoria.marques@icat.ufal.br
Kelvy Rosalvo Alencar Cardoso
ORCID: https://orcid.org/0000-0002-4391-8167
Universidade Federal de Alagoas, Brasil
E-mail: kelvy.cardoso@icat.ufal.br
Caroline Cristina da Silva de Andrade
ORCID: https://orcid.org/0000-0002-4050-0945
Universidad de Buenos Aires, Argentina
E-mail: carolineandradeac@gmail.com

Abstract
Dengue is one of the serious public health problems worldwide. The Northeast of Brazil (NEB) has an ideal climate and urban environment for the proliferation of the Aedes mosquito (*aegypti* and *albopictus*), the vector of the disease. The State of Alagoas, especially its capital, has frequent epidemics of the disease. Therefore, the objective of this study is to evaluate the application of Artificial Neural Networks (ANN) in reported cases of dengue (CND) in the administrative regions (AR) of Maceió. The RAs are divided into: AR1, AR2, AR3, AR4, AR5, AR6, AR7 and AR8. The CND were submitted to ANN nonlinear autoregressive (NAR) – (ANN-NAR). The study period was 2011 to 2020. The results obtained from the CND stood out in specific years (2012, 2013, 2017, 2018 and 2020), on the other hand, there were overestimations of the forecasts via ANN. In some ARs there was underreporting and, therefore, it interfered with the forecasts results. The ANN-NAR was validated, as most of the predictions showed a positive correlation and responded to the observed data, except for the ARs with underreporting. The use of ANN is suitable for warning and disease prediction, where such an instrument can be used in preventive actions to control the disease.
Keywords: Dengue; Northeast of Brazil; Artificial intelligence; Urban climate.

Resumo
A dengue é um dos graves problemas de saúde pública mundial. O Nordeste do Brasil (NEB) possui um clima e ambiente urbano ideal para a proliferação do mosquito *Aedes* (*aegypti* e *albopictus*), vetor da doença. O Estado de Alagoas, principalmente a sua capital, tem epidemias da doença de forma frequente. Portanto, o objetivo deste estudo é avaliar a aplicação de Rede Neural Artificial (RNA) nos casos notificados de dengue (CND) nas regiões administrativas (RA) de Maceió. As RAs são divididas em: RA1, RA2, RA3, RA4, RA5, RA6, RA7 e RA8. Os CND foram submetidos a RNA não linear autorregressiva (NAR) – (RNA-NAR). O período de estudo foi de 2011 a 2020. Os resultados obtidos de CND se destacaram em anos específicos (2012, 2013, 2017, 2018 e 2020), por outro lado

houve superestimativas das previsões via RNA. Em algumas RAs houve subnotificações e, por isso interferiu nos resultados das previsões. A RNA-NAR foi validada, visto que a maioria das previsões apresentou correlação positiva e com resposta aos dados observados, exceto as RAs com subnotificações. O uso da RNA é adequado no alerta e previsão da donça, onde tal instrumento pode ser usado em ações preventivas de controle da doença.
Palavras-chave: Dengue; Nordeste brasileiro; Inteligência artificial; Clima urbano.

Resumen
El dengue es uno de los graves problemas de salud pública a nivel mundial. El Noreste de Brasil (NEB) posee clima y ambiente urbano ideales para la proliferación del mosquito Aedes (*aegypti* y *albopictus*), vector de la enfermedad. El Estado de Alagoas, especialmente su capital, tiene frecuentes epidemias de la enfermedad. Por lo tanto, el objetivo de este estudio es evaluar la aplicación de Rede Neuronal Artificial (RNA) en casos notificados de dengue (CND) en las regiones administrativas (RA) de Maceió. Las RAs se dividen en: RA1, RA2, RA3, RA4, RA5, RA6, RA7 y RA8. Los CND fueron sometidos a RNA no lineal autorregresivo (NAR) – (RNA-NAR). El periodo de estudio fue del 2011 al 2020. Los resultados obtenidos de la CND se destacaron en años específicos (2012, 2013, 2017, 2018 y 2020), por otro lado, hubo sobreestimaciones de los pronósticos vía RNA. En algunas RAs hubo subregistro y, por lo tanto, interfirió con los resultados del pronóstico. El RNA-NAR fue validado, ya que la mayoría de las predicciones mostraron correlación positiva y respondieron a los datos observados, excepto los RAs con subregistro. El uso de RNA es adecuado para alerta y predicción de enfermedades, donde dicho instrumento puede usarse en acciones preventivas para controlar la enfermedad.
Palabras clave: Dengue; Noreste de Brasil; Inteligencia artificiales; Clima urbano.

1. Introduction

One of the biggest public health problems worldwide, especially in developing countries, is dengue. Currently, environmental studies and changes are concerned with climate change in the human epidemic, especially with the increase in dengue climate change around the world (Souza, et al., 2021). With the increase in temperature on a global scale, an incidence of pandemic endemics through diseases and transferred by vector, for example, dengue, yellow fever, leptospirosis, and other viral diseases, will increase considerably in the coming decades (Oliveira-Jr, et al., 2019; Lima, et al., 2019; Andrioli, et al., 2020).

Recently, some studies carried out in Brazil have pointed out that climate change can cause a significant increase in the incidence of waterborne diseases, especially dengue (Souza, et al., 2021; Lee, et al., 2021). From the 1990s to the mid-2010s, the incidence of dengue cases has been more frequent in all regions of Brazil (Oliveira-Jr, et al., 2019). Northeast of Brazil (NEB) is distinguished from other Brazilian regions not only for its recurrent droughts and rainfall variability (Correia-Filho, et al., 2019), as well as the existing problems, for example, floods, of irregular garbage and, currently, real estate speculation in the Metropolitan Regions (MR), which causes an increase in epidemic diseases due to changes in land use and occupation (Silva, et al., 2021).

The East of the NEB (ENEB) is a highly dense region and, therefore, with a series of sanitary and structural problems, which increase the incidence of dengue. It is worth mentioning that the mosquito has great urban adaptability, they need reservoirs of water, preferably clean, to lay their eggs, hence the importance of rainfall and maintenance of high temperature for their development. Maceió, capital of Alagoas (AL) located on the ENEB coast, stands out, with a history of dengue occurrence (Santos-Jr & Silva, 2019; Silva, et al., 2021) and therefore needs monitoring and follow-up by public managers. However, the studies carried out to date are limited to comparisons with climatic variables (Almeida & Ribeiro, 2018), use of multivariate statistics (Silva, et al., 2021; Oliveira-Jr, et al., 2019) and none that address the use of Artificial Neural Networks (ANN). Therefore, the objective of this study is to evaluate the application of ANN in reported cases of dengue in Maceió-AL, at the level of administrative regions (AR).

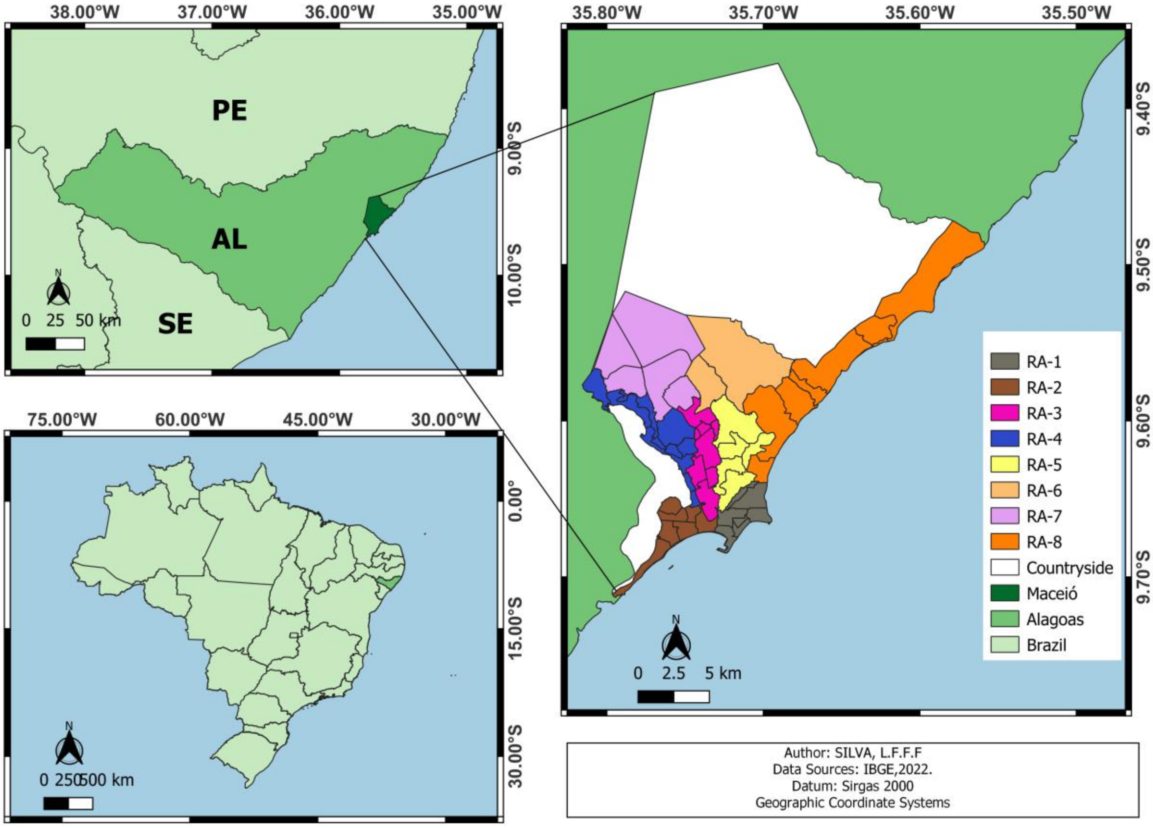
(CC BY 4.0) | ISSN 2525-3409 | DOI: http://dx.doi.org/10.33448/rsd-v11i14.36382

2. Methodology

2.1 Study area

The study area is Maceió, capital of Alagoas, with an area of 503.07 km², divided into 193.34 km² (urban), 285.47 km² (rural) and 23.26 km² (lagoon) - (IBGE, 2020; Oliveira-Jr, et al., 2021). Maceió is in ENEB between latitudes 9.40º to 9.70º S, and longitudes 35.80º to 35.60º W, with altitudes that vary between 0 and 100 m above mean sea level (NMM), located in the Atlantic Forest biome. (Santos, et al., 2022; Batista, et al., 2021), as shown in Figure 1.

Figure 1 - Location of the study area, highlighting the eight RAs in Maceió.



Source: Authors (2022).

According to the Köppen-Geiger classification, it is categorized as tropical hot and humid 'As', with variations in temperature throughout the year. The average temperatures range from 25.0ºC to 26.4ºC, with minimum and maximum of 19.0ºC and 31.0ºC, respectively. The annual rainfall average varies between 1,500 and 2,000 mm, with well-defined seasons: i) dry, from September to February; ii) rainy, from March to August (Correia-Filho, et al., 2019). The current Master Plan for the City of Maceió shows that the urban area consists of 50 neighborhoods within eight RA, as shown in Figure 1 and Table 1 (Oliveira-Jr, et al., 2021).

Table 1 - List of neighborhoods by Administrative Region in the municipality of Maceió.

Administrative Region (AR)	Neighborhoods
AR1	Poço, Jaraguá, Ponta da Terra, Pajuçara, Ponta Verde, Jatiúca e Mangabeiras.
AR2	Centro, Pontal da Barra, Trapiche da Barra, Prado, Levada, Vergel do Lago e Ponta Grossa.
AR3	Farol, Pitanguinha, Pinheiro, Gruta de Lourdes, Canaã, Santo Amaro, Jardim Petrópolis e Ouro Preto.
AR4	Bom Parto, Mutange, Bebedouro, Chã de Bebedouro, Petrópolis, Chã da Jaqueira, Santa Amélia, Fernão Velho e Rio Novo.
AR5	Jacintinho, Barro Duro, Serraria, São Jorge e Feitosa.
AR6	Benedito Bentes e Antares.
AR7	Santos Dumont, Cidade Universitária, Santa Lúcia, Tabuleiro dos Martins e Clima Bom.
AR8	Cruz das Almas, Jacarecica, Guaxuma, Garça Torta, Riacho Doce, Pescaria e Ipioca

Source: Authors (2022).

2.2 Time series of reported dengue cases

The data on Notified Dengue Cases (NDC) from the neighborhoods of Maceió were kindly provided by the State Department of Health (SDH), Superintendence of Health Surveillance (SHS), Management of Surveillance and Control of Communicable Diseases (MSCCD), Advice on Vectors (AOV), Zoonoses and Environmental Factors (ZEF) and Technical Area of Surveillance and Control of Arboviruses (TASCA). Dengue data cover the period from 2011 to 2020.

2.3 Prediction via artificial neural network

Nonlinear autoregressive (NAR) are recurrent or feedback networks, that is, the output values of a layer can feed back to the input layer. The NDC prediction was performed using ANN-NAR, being applied to the NDC time series in the eight ARs of Maceió (Table 1). The initial operation of this ANN is of the Feedforward type, where its information in the input layer always goes towards the output layer. This ANN is characterized by the variable Z, which is the time delay. The Z factor works as a memory that provides current and previous input values (Santos, et al., 2022; Santos, et al., 2020; Samet, et al., 2019).

The ANN-NAR training and learning algorithm used was the Levenberg-Marquardt (LM) which is supervised and aims to adjust all parameters to find a better relationship between the input and target data (Silva, 2018). Samet, et al., (2019), evaluated the prediction performance performed with LM and observed that it significantly outperforms other training algorithms, for example, Bayesian Regularization (BR), Basic Gradient Descent (BGD), One Step Secant (OSS), etc.

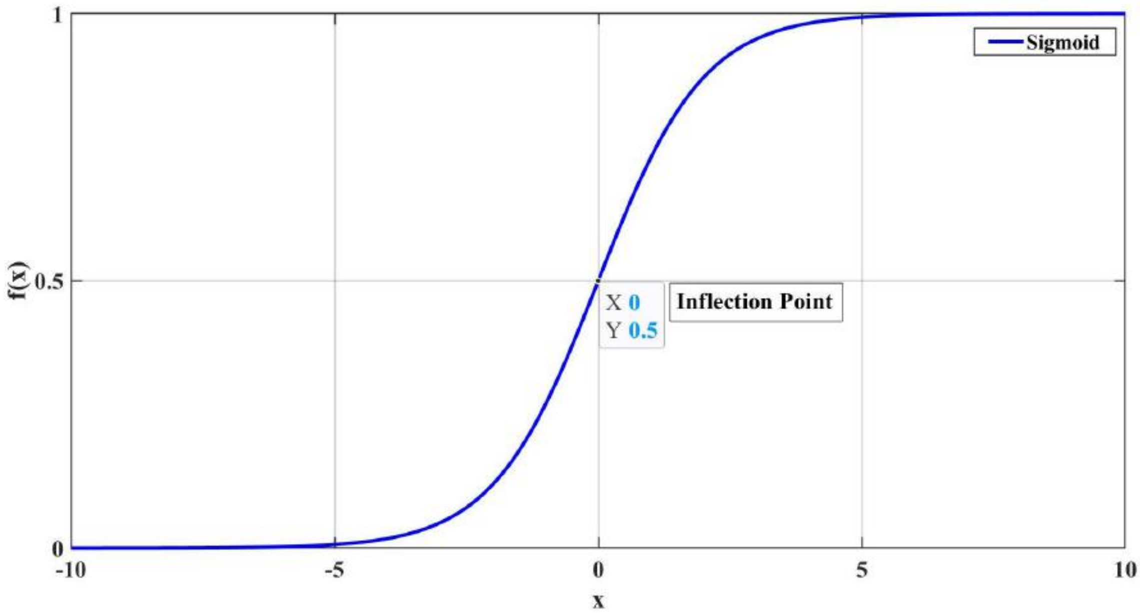
The network performance parameter used was the Mean Square Error (MSE) (Eq. (1)), this algorithm identifies how much the ANN-NAR can reproduce the input/target data throughout the training, testing and validation. The MSE averages the square of the difference between the predicted value $y(t+1)$ and the input/target value $y(t)$. Thus, the best model is the one with the lowest MSE (Silva, 2018).

$$MSE = \frac{1}{N} \cdot \sum_{i=1}^{N} (y(t+1) - y(t))^2 \quad (1)$$

The activation function chosen was the Sigmoid (Eq. (2)), as it does not propagate negative values, since there are no negative NDC records. It is a continuously differentiable and nonlinear activation function with the shape of S that serve as input to the next layer, they reduce the effect of extreme input values, with a better functioning of the ANN-NAR (Santos, et al., 2022). Figure 2 shows the Sigmoid activation function in the Cartesian plane x and f(x) with emphasis on the ordered pair (0,0.5) which is the inflection point of the curve.

$$f(x) = \frac{1}{1 + e^{-x}} \tag{2}$$

Figure 2 - Sigmoid activation function.



Source: Authors (2022).

Table 2 shows the final configuration of the ANN-NAR training phase, the architecture was configured to provide the NDC prediction based on the input/target values in the study period. The configuration of the parameters used was based on studies (Silva 2018; Santos, et al., 2020; Santos, et al., 2022).

Table 2 - Parameters used in the RNA-NAR training phase.

Parameters	Data/values
Entry/target signals	NDC
Signals in the input layer	10
Number of forecasts	10
Training algorithm	Levenberg-Marquardt
Performance algorithm	Mean Squared Error
Seasons	1000
Verification	10
Gradient	0,00001
Training, Testing and Validation	70%, 15% e 15%
Hidden layer	1
Entry delay	1:2
Neurons in the hidden layer	10
Activation function	Sigmoid
output layer	1
Activation function	Linear
Neurons in the output layer	1

Source: Authors (2022).

2.4 Statistical analysis

The evaluation of NDC in Maceió was based on descriptive (DS), exploratory (XS) and bivariate (BS) statistics. In the DS, the accumulated total was used, percentage of the total accumulated of NDC, mean (MA), minimum (Min), 1st quartile (Q1), median (MD), 3rd quartile (Q3), maximum (Max), standard deviation (SD), variance (Var), asymmetry (AC) and kurtosis (KC) coefficients.

In XS it was based on Boxplot to provide a summary view of some DS metrics, e.g., Min, Q1, MD, Q3, Max, interquartile range, data skewness, adjacent lower and upper values, and outliers. In EB, Pearson (PC), Kendall (KC) and Spearman (SC) correlations and error statistics (ES) between observed and predicted were calculated.

The statistical indicators used in the validation of the ANN-NAR were: Mean Error (ME)-(Eq. (3)); Root Mean Square Error (RMSE)-(Eq. (4)); Mean Absolute Percent Error (MAPE)-Eq. (5)); PC-(Eq. (6)); KC-(Eq. (7, 8 and 9)); SC-(Eq. (10)); coefficient of determination (R^2) - (Eq. (11)); Simple Linear Regression (SLR)-(Eq. (12)). All equations are listed below:

$$ME = \frac{1}{N} \cdot \sum_{i=1}^{N} (ANN_i - OBS_i) \tag{3}$$

$$\text{RMSE} = \sqrt{\frac{1}{N} \cdot \sum_{i=1}^{N} (ANN_i - OBS_i)^2} \tag{4}$$

$$MAPE = \frac{1}{N} \cdot \sum_{i=1}^{N} \left| \frac{ANN_i - OBS_i}{OBS_i} \right| \cdot 100 \tag{5}$$

(CC BY 4.0) | ISSN 2525-3409 | DOI: http://dx.doi.org/10.33448/rsd-v11i14.36382

$$PC = \frac{\sum_{i=1}^{N}(ANN_i - \overline{ANN}) \cdot (OBS_i - \overline{OBS})}{\sqrt{\sum_{i=1}^{N}(ANN_i - \overline{ANN})^2} \cdot \sqrt{\sum_{i=1}^{N}(OBS_i - \overline{OBS})^2}} \tag{6}$$

$$KC = \frac{2K}{N(N-1)} \tag{7}$$

$$K = \sum_{i=1}^{N-1} \sum_{j=i+1}^{N} \xi\left(OBS_i, OBS_j, ANN_i, ANN_j\right) \tag{8}$$

$$\xi\left(OBS_i, OBS_j, ANN_i, ANN_j\right) = \begin{cases} 1, & se\left(OBS_i - OBS_j\right)\left(ANN_i - ANN_j\right) > 0 \\ 0, & se\left(OBS_i - OBS_j\right)\left(ANN_i - ANN_j\right) = 0 \\ -1, & se\left(OBS_i - OBS_j\right)\left(ANN_i - ANN_j\right) < 0 \end{cases} \tag{9}$$

$$SC = 1 - \frac{6\sum(ANN_i - OBS_i)^2}{N(N^2 - 1)} \tag{10}$$

$$R^2 = \left(\frac{\sum_{i=1}^{N}(ANN_i - \overline{ANN}) \cdot (OBS_i - \overline{OBS})}{\sqrt{\sum_{i=1}^{N}(ANN_i - \overline{ANN})^2} \cdot \sqrt{\sum_{i=1}^{N}(OBS_i - \overline{OBS})^2}}\right)^2 \tag{11}$$

$$ANN = a \cdot OBS + b \tag{12}$$

Where, the total number of values (N) is equal to ten, the index (i) varies from one to N, ANN_i and OBS_i are the predicted and observed values with their respective indices, $\overline{ANN}$ and $\overline{OBS}$ are the average predicted and observed values. The elements of Kendall's correlation are the indices j such that $i<j$, $\xi(*)$ is the variance of KC, K is the sum of the variances with the respective indices.

The scattering diagram (1:1) between the observed and predicted NDC, being obtained the equations of the lines generated from the (SLR)-(Eq. (12)); (R^2)-(Eq. (11)); and PC-(Eq. (6)). The PC values were classified according to the methodology of Hopkins (2009) - (Table 3). All statistical procedures were performed in Matlab® software (version R2021b – student use).

(CC BY 4.0) | ISSN 2525-3409 | DOI: http://dx.doi.org/10.33448/rsd-v11i14.36382

Table 3 - PC coefficient and its respective classification (Hopkins, 2009).

PC	Classification
0 – 0,1	Very low
0,1 – 0,3	Low
0,3 – 0,5	moderate
0,5 – 0,7	High
0,7 – 0,9	Very tall
0,9 – 1,0	Almost perfect

Source: Hopkins (2009).

3. Results and Discussion

3.1 Statistics applied to NDC

Table 4 shows the DS of the NDC of the ARs of Maceió, with a total of 7,647 cases in the period studied. AR7 stands out with 3,113 cases (41%), the most populous region of the capital and with changes in land use and occupation (Correia-Filho, et al., 2019; Oliveira-Jr, et al., 2019), followed by AR8 with 1,557 cases (20%), region in territorial expansion (Batista, et al., 2021) and AR2 with 1,286 cases (17%) - region with environmental degradation and abandoned by public management for decades (Batista, et al., 2021), being the highest in records in the time series, the exceptions were AR5 with 114 (1%) and AR6 with 141 (2%), both regions with suspected NDC underreporting, respectively.

It is worth mentioning that the greater the NDC record, the greater the variability of the data (Santos, et al., 2016; Silva, et al., 2021) and, especially in areas of environmental degradation, in changes in land use and occupation and with socioeconomic problems (Lee, et al., 2021), as verified in the Var and SD of AR7, AR8 and AR2. Both regions mentioned also had the highest MA and Max of NDC (Table 4). The results obtained in this study were similar to the study by Almeida & Ribeiro (2018), in the municipality of Uberaba in Minas Gerais (MG), followed by Silva (2018) in Pará and Silva, et al., (2021) in Alagoas.

The ARs with the highest positive asymmetric distributions were AR5 and AR6, populated neighborhoods with several changes in land use and occupation in recent decades (Correia-Filho, et al., 2019; Batista, et al., 2021). AR5 and AR6 move away from the symmetrical position along with AR8. The other ARs are within the Pearson symmetry range (-1 to +1) - (Table 4). All ARs used in the study obtained KC >0.263, being platykurtic distribution curves, with a flatter frequency distribution than the normal distribution (Santos, et al., 2018; Silva 2020). The results obtained were similar to Silva (2018) who used ANN to predict new NDC in Pará.

Predictions (PARs) overestimated the NDC in all ARs, except AR5 and AR6. The highest prediction values occurred in PAR7 with 5,262.66 cases (44.63%), followed by PAR8 with 2,972.16 cases (25.20%) and PAR2 with 1,638.99 cases (13.90%), respectively. These forecasts also showed the greatest data variability in statistical metrics, such as Var, SD, and MA similar to ARs. The statistical metrics of PAR5 and PAR6 presented null values due to lack of data. The DS results of predictions were similar to Soares, et al., (2018) in Salvador, Bahia.

Table 4 - Results of the DS applied to the NDC of the ARs and PARs of Maceió, their respective identifiers (ID) in the period 2011 to 2020.

ID	**ATD**	**ATPD**	**Var**	**SD**	**MA**	**Min**	**Q1**	**MD**	**Q3**	**Max**	**CA**	**Ck**
	Dengue.Year^{-1}	%	Dengue^{2}.Year^{-2}	Dengue.Year^{-1}							-	-
AR1	517	7	2.084,23	45,65	51,7	0	11	52	88	113	0,04	1,21
AR2	1.286	17	21.753,82	147,49	128,6	0	21	63,5	252	422	0,92	2,45
AR3	347	5	1.400,46	37,42	34,7	0	4	22	62	113	0,88	2,75
AR4	572	7	2.974,18	54,54	57,2	1	9	42,5	112	135	0,24	1,33
AR5	114	1	1.299,6	36,05	11,4	0	0	0	0	114	2,67	8,11
AR6	141	2	1.988,1	44,59	14,1	0	0	0	0	141	2,67	8,11
AR7	3.113	41	106.423,6	326,23	311,3	0	45	153	714	760	0,46	1,43
AR8	1.557	20	36.112,46	190,03	155,7	0	25	67,5	269	573	1,21	3,23
TOTAL	7.647	100	-	-	-	-	-	-	-	-	-	-
PAR1	577,43	4,90	1.643,61	40,54	57,74	2,38	18,47	74,33	86,41	116,20	-0,18	1,51
PAR2	1.638,99	13,90	22.788,45	150,96	163,90	1,99	22	145,27	267,30	422	0,35	1,67
PAR3	458,41	3,89	539,73	23,23	45,84	15,74	30,25	36,19	69,77	88,80	0,63	2,17
PAR4	882,56	7,48	1.732	41,62	88,26	20,26	51,93	86,81	165,28	142,36	-0,01	1,94
PAR5	0	0	0	0	0	0	0	0	0	0	NaN	NaN
PAR6	0	0	0	0	0	0	0	0	0	0	NaN	NaN
PAR7	5.262,66	44,63	9.112,15	95,46	526,27	359,69	491,64	533,36	602,10	638,32	-0,67	2,30
PAR8	2.972,16	25,20	45.501,03	213,31	297,22	109,54	116,07	192,30	508,52	690,87	0,69	1,98
TOTAL	11.792,21	100	-	-	-	-	-	-	-	-	-	-

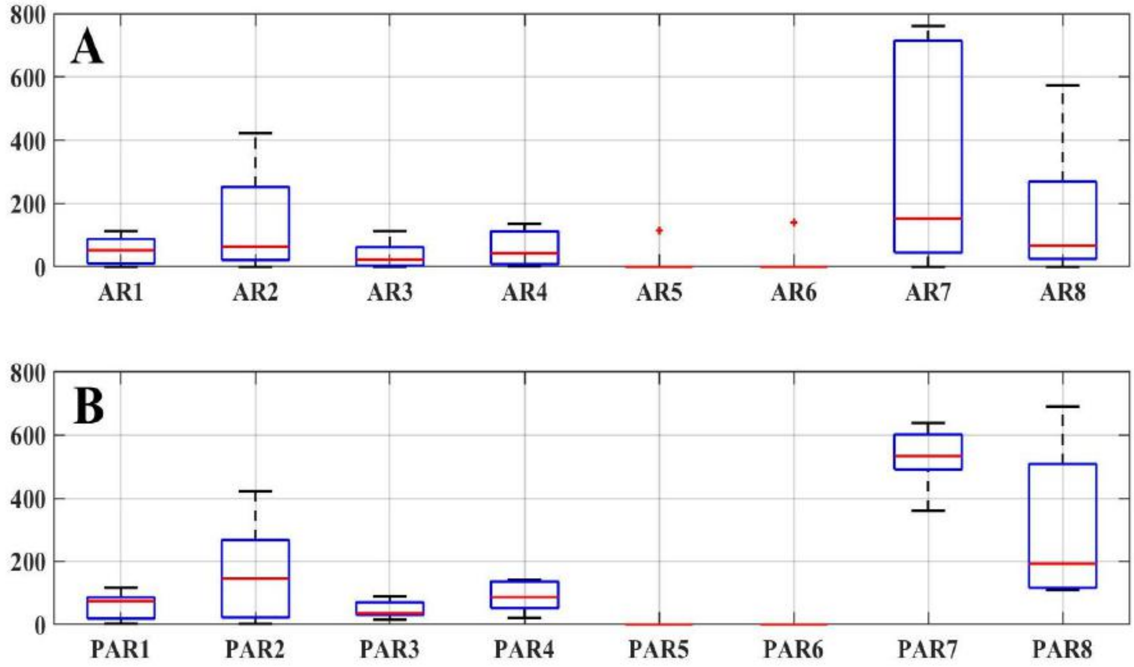
Subtitle: ATD – Accumulated Total CND, TAPD – Total Accumulated Percentage of CND, Var – Variance, SD – Standard Deviation, MA – Arithmetic Mean, Min – Minimum, Q1 – First Quartile, MD – Median, Q3 – Third Quartile, Max – Maximum, CA – Coefficient of Asymmetry, CK – Coefficient of Kurtosis, NaN - Not a Number. Source: Authors (2022).

Figure 3A shows the boxplot of the NDCs in the ARs. AR7 had the highest interquartile range (IQR) of the NDC, followed by the highest Q1, MD, Q3 and Max, together with AR8 and AR2. AR1 was the only one with the MD closest to Q3, which indicated negative asymmetry, different from the other ARs. The AR5 and AR6 were similar, with outliers and the remaining statistical metrics null. It is worth mentioning that the results obtained via boxplot were similar to the studies by Silva, et al., (2021) carried out in Alagoas and Oliveira-Jr, et al., (2019) for Brazil

Figure 3B displays the boxplot of the NDC PARs. The main difference found between PAR7 and AR7 involves the statistical metrics (Min, Q1, MD, Q3, Max), with a decrease in IQR, on the contrary, in PAR8. The PAR5 and PAR6 did not present atypical values, as the AR5 and AR6. PAR1 showed greater negative asymmetry than AR1. The biggest difference between PAR2 and AR2 is MD.

(CC BY 4.0) | ISSN 2525-3409 | DOI: http://dx.doi.org/10.33448/rsd-v11i14.36382

Figure 3 - Boxplot of the NDC of the ARs (A) and PARs (B) of Maceió in the period 2011 to 2020.



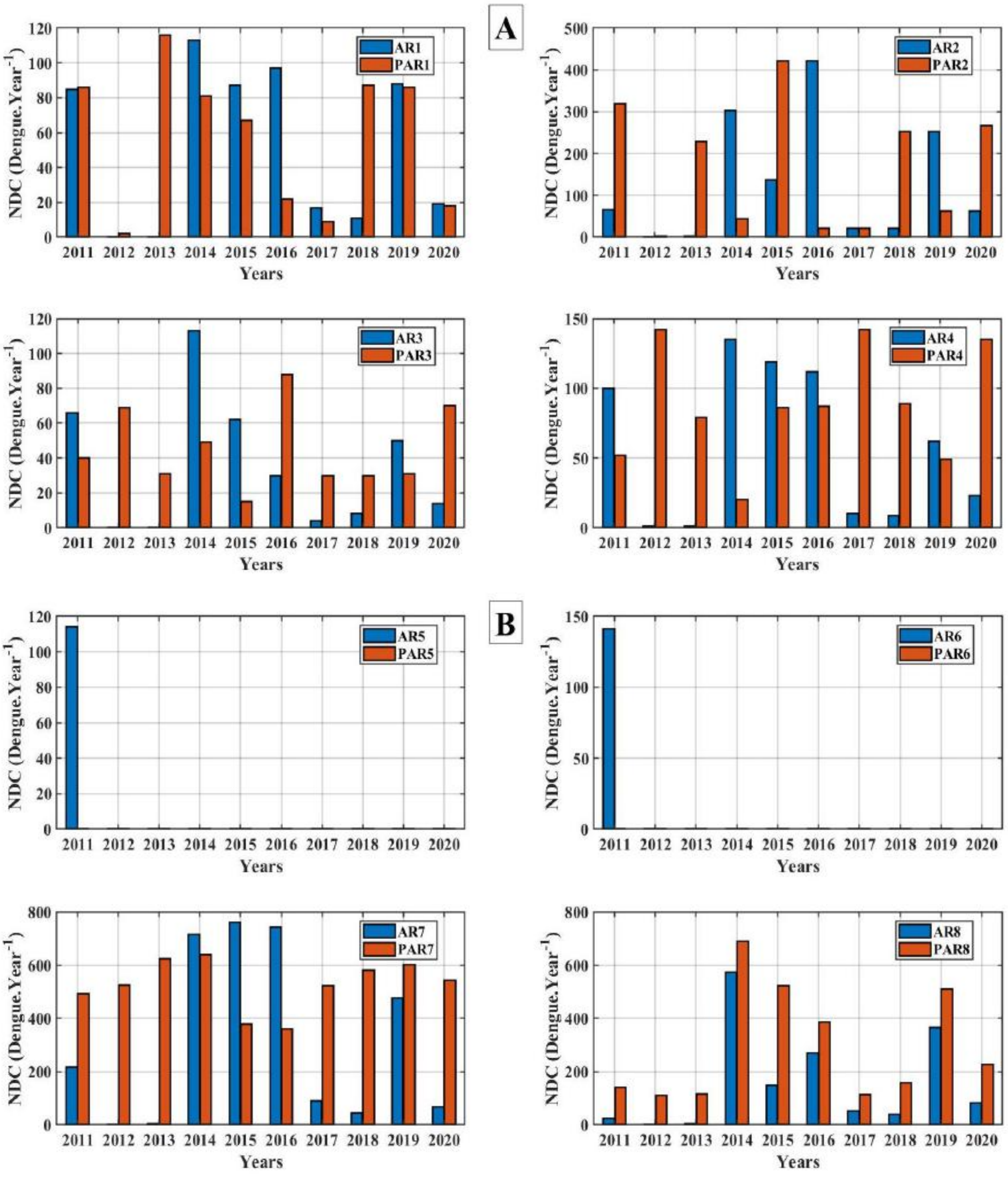
Source: Authors (2022).

3.2 Annual variability of observed versus forecast NDC

Figure 4 displays the annual distribution of observed NDC versus PARs. The patterns of ARs and PARs showed similarity in relation to their annual variability NDC, with emphasis on the years 2012, 2013, 2017, 2018 and 2020, where there was an overestimation of the forecasts in relation to the observed data. It is worth noting that these were years with intense rains and prolonged droughts being influenced by the ENSO phases in the NEB. On the contrary, from the years 2011, 2014, 2015, 2016 and 2019, with the highest accumulated NDC and forecasts were underestimated (Silva, 2018; Silva, et al., 2021). In AR5 and AR6 are the most populous neighborhoods of Maceió (Jacintinho-AR5 and Benedito Bentes-AR6) and these regions may have been neglected in the registration of NDC, that is, underreporting.

(CC BY 4.0) | ISSN 2525-3409 | DOI: http://dx.doi.org/10.33448/rsd-v11i14.36382

Figure 4 - Annual temporal distribution of the NDC of the ARs versus the PARs of Maceió in the period 2011 to 2020.



Source: Authors (2022).

Moraes, et al., (2019) correlated the El Niño/La Niña events with the NDC in the Brazilian Amazon between 2001 and 2012. The El Niño/La Niña events caused a reduction/increase in the local rainfall pattern, being consistent with the lowest/highest number of NDC in most capitals in the region. Oliveira (2019) correlated meteorological variables (rainfall and air temperature) with NDC in Paraíba between 2007 and 2017. The results obtained show that the highest NDC records occurred in the years 2011, 2013 and 2016, followed by a weak to moderate correlation between precipitation and temperature and NDC, with the Paraiba's wood zone being affected by the El Niño that occurred in the years 2007, 2010, 2015 and 2016. Silva, et al., (2021), evaluated the NDC in Alagoas via applied statistics and geospatial location between 2000 and 2015. The results obtained showed that the highest records of NDC occurred in 2010, the year of the largest flood in the state.

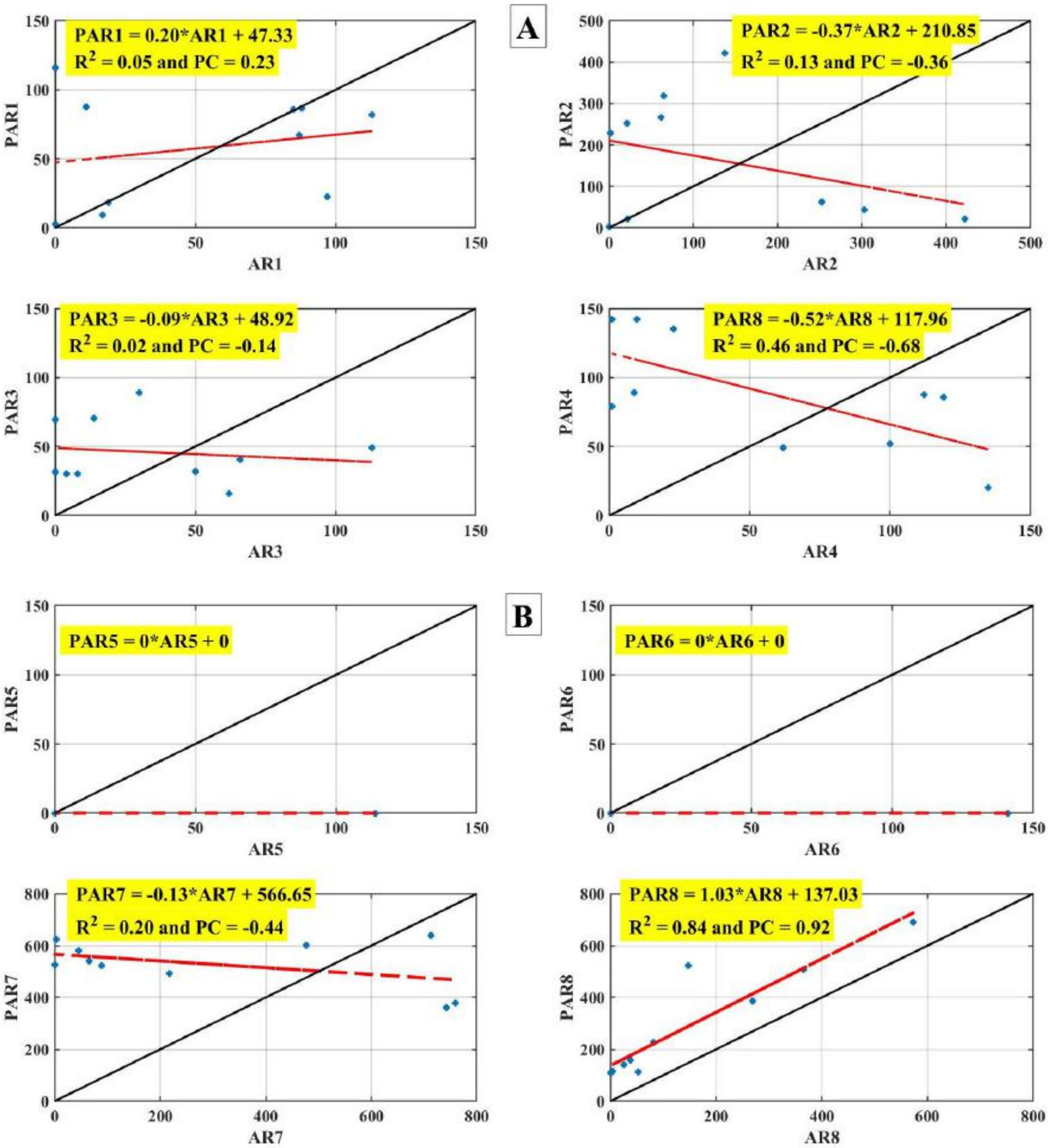
Almeida & Ribeiro (2018) analyzed the dynamics of climate variables (temperature, precipitation, and relative humidity) relating them to the NDC in Uberaba, Minas Gerais, from 2009 to 2016. The results indicated that the 1% increase in temperature, corresponded to an increase of 0.45% in the NDC in the following month, followed by the increase in precipitation impacted by 15% in the NDC in the following month and the increase in relative humidity had a negative impact on the NDC. The highest NDC records occurred in the years 2010, 2012, 2013 and 2016.

Several studies, for example, Almeida and Ribeiro (2018); Santos, et al., (2018); Dantas, et al., (2019); Moraes, et al., (2019); Silva, et al., (2021) suggest the influence of climate on NDC. Recently, Batista, et al., 2021, evaluated the disordered population growth in AR6 and AR7, due to changes in land use and occupation that result in soil sealing and, thus, rains cause flooding in the urban environment and cause diseases of water transmission, for example dengue and leptospirosis (Santos, et al., 2018; Silva, et al., 2021).

3.3 Validation of the ANN-NAR model

Figure 5 shows the scatterplot (1:1) between the NDC of the ARs and PARs. Figure 5A corresponds to AR1, AR2, AR3 and AR4, where the R^2 coefficients ranged from 2 to 46%, categorized as insufficient and low for PARs via ANN-NAR. Pearson's coefficients (r) varied between -0.68 and 0.23 categorized as low (PAR1 and PAR3), moderate (PAR2) and high (PAR4), (Hopkins, 2009) – (Table 2). Some negative correlations occurred between observed and predicted data, mainly due to insufficient data, in this case due to underreporting. The predictions for AR1, AR2, AR3 and AR4 were overestimated. Figure 3B corresponds to AR5, AR6, AR7 and AR8, where the R^2 coefficients ranged between 20 and 84%, that is, the predictive variables (ARs) are explained between 20 and 84% by the ANN-NAR model, with satisfactory use for the study region. The r coefficients varied between -0.44 and 0.92, being categorized as moderate (PAR7) and almost perfect (PAR8), according to (Hopkins, 2009) – (Table 2). PAR5 and PAR6 were compromised, due to omission of NDC records. PAR8 obtained the best coefficients because they are medium-sized neighborhoods with proximity and, thus, facilitates the work of collecting endemics teams.

Figure 5 - Scatter diagram (1:1) of the NDC of the ARs versus the PARs of Maceió in the period from 2011 to 2020.



Source: Authors (2022).

Table 5 displays the summary of the correlation analysis and BS applied in the performance evaluation of the ANN-NAR model. Highlight for lower ME in PAR1 (6.04 cases), followed by PAR3 (11.14 cases) compared to the other PARs, due to the lower number of NDC records. On the contrary, the highest ME in PAR7 (214.97 cases), followed by PAR8 (141.52 cases) and PAR2 (35.30 cases), due to the highest records of NDC.

Table 5 - ES results and correlations between the NDC of the ARs and PARs of Maceió in the period 2011 to 2020.

ID	ME	RMSE	MAPE	PC	KC	SC	R^2
AR1-PAR1	6,04	33,18	29,36	0,23	0,05	0,01	0,05
AR2-PAR2	35,30	205,49	48,69	-0,36	0,11	0,12	0,13
AR3-PAR3	11,14	41,93	37,11	-0,14	0,09	0,07	0,02
AR4-PAR4	31,06	77,66	57,52	-0,68	-0,41	-0,56	0,46
AR5-PAR5	-11,40	11,40	10	*NaN*	*NaN*	*NaN*	*NaN*
AR6-PAR6	-14,10	14,10	10	*NaN*	*NaN*	*NaN*	*NaN*
AR7-PAR7	214,97	382,95	50,39	-0,44	-0,29	-0,38	0,20
AR8-PAR8	141,52	141,52	24,70	0,92	0,78	0,89	0,84

Subtitle: ME - Mean Error (cases), RMSE - Root Mean Square Error (cases), MAPE - Mean Absolute Percent Error (%), PC - Pearson Correlation, KC - Kendall Correlation, SC - Spearman Correlation, R^2 - Coefficient of Determination (%), NaN - Not a Number. Source: Authors (2022).

The result of the RMSE ranged between 11.40 cases (PAR5) and 382.95 cases (PAR7). The RMSE indicates that the larger the NDC record, the greater the variability and degree of dispersion of the data and significantly increased the forecast error. Therefore, the highest prediction errors occurred in PAR7, PAR2 and PAR8, again due to high NDC records. Both the PAR5 and PAR6 predictions showed the lowest statistical errors, due to the smaller NDC records and low data variability.

MAPE reports the percentage of errors in the ANN-NAR model predictions. The results ranged from 10% (PAR5 and PAR6) to 57.52% (PAR4). Again, both PAR5 and PAR6 predictions had the smallest errors. A positive highlight for PAR8 (24.70%) which, even with one of the highest records of NDC, obtained one of the lowest MAPE, with satisfactory prediction.

The correlations (PC, KC and SC) ranged from -0.68 (PC of PAR4) to 0.92 (PC of PAR8). Most predictions showed positive correlations, and the use of the ANN-NAR model is satisfactory. Highlight for PAR4 and PAR7 with negative correlations, followed by PAR1 and PAR3 with correlations close to zero, categorized as very low and low, as shown in Table 2. Again, PAR8 presented the best correlations, in the categories very high and almost perfect, according to (Hopkins, 2009). The R^2 coefficients ranged from 0.02 (PAR3) to 0.84 (PAR8). PAR3 and PAR1 obtained coefficients close to zero, being considered unsatisfactory to explain the predictive variable. It is worth mentioning that PAR8 also presented the highest value of R^2 (Table 5).

4. Conclusion

The use of the ANN-NAR model in the administrative regions of Maceió is satisfactory. Individually, some regions with underreporting problems and fewer NDC records influenced the performance of the ANN-NAR model, in this case, AR5 and AR6. Both regions need better management by public managers in relation to the collection of endemics teams and, consequently, in the treatment and organization of reported data.

The overestimates and underestimates in the scale are due to several factors, in this case, climatic and anthropic, for example, intense and frequent changes in land use and occupation in the city and the occurrence of rains and prolonged droughts that interfere in the dynamics of the dengue vector in Maceió. However, the statistical metrics adopted in the study allow validating the use of the model in an adequate way in the alert and prediction of the disease, which in turn can be an instrument used in preventive actions to control the disease, particularly in neighborhoods with greater population density at the capital.

The ANN-NAR model is applicable in dengue prediction, as it can reproduce the variability of NDC in most ARs in Maceió. A limitation of the model used is to identify the sequence of years with the highest accumulated NDC. To correct this limitation and suggestions for future work, it is necessary to increase the study period, test the ANN-NAR configurations more, use other types of ANN and compare the results.

Acknowledgment

The authors thank the Graduate Program in Meteorology (PPGMET), the Academic Unit of Atmospheric Sciences (UACA), the Center for Technology and Natural Resources (CTRN), the Federal University of Campina Grande (UFCG), the Meteorology Laboratory and the Environment (LAMMA), the Institute of Atmospheric Science (ICAT), the Federal University of Alagoas (UFAL) and the National Council for Scientific and Technological Development (CNPq) for the financial support granted to the first author during the master's course (2017-2019) and PhD (2020-Current). The second author thanks CNPq for the Level 2 Productivity grant under the process number 309681/2019-7.

References

Almeida, F. P. & Ribeiro, F. A. B. S. (2018). Variáveis climáticas e casos notificados de dengue no município de Uberaba, Minas Gerais. *Revista Verde de Agroecologia e Desenvolvimento Sustentável*. 13(5), 644-651. http://dx.doi.org/10.18378/rvads.v13i5.6217.

Andrioli, D. C., Busato, M. A. & Lutinski, J. A. (2020). Spatial and temporal distribution of dengue in Brazil, 1990-2017. *PLoS ONE*. 15(2), 1-13. e0228346. https://doi.org/10.1371/journal.pone.0228346.

Batista, B. A., Correia-Filho, W. L. F., Oliveira-Jr, J. F., Santiago, D. B. & Santos, C. T. (2021). Avaliação da expansão urbana na Cidade de Maceió, Alagoas - Nordeste do Brasil. *Research, Society and Development*. 10(11), 1-14. https://doi.org/10.33448/rsd-v10i11.19537.

Correia-Filho, W. L. F., Santiago, D. B., Oliveira-Jr, J. F. & Silva-Jr, C. A. (2019). Impact of urban decadal advance on land use and land cover and surface temperature in the city of Maceió, Brazil. *Land Use Policy*. 87(1), 1-11. https://doi.org/10.1016/j.landusepol.2019.104026.

Hopkins, W. G. *A new view of statistics: Correlation Coefficient*. New York: Internet Society for Sport Science, 2009. http://www.sportsci.org/resource/stats/correl.html.

IBGE - *Instituto Brasileiro de Geografia e Estatística*. Censo Brasileiro de 2020. IBGE.

Lee, S. A., Economou, T., Barcellos, C., Catão, R., Carvalho, M. S. & Lowe, R. (2021). Effect of climate change, connectivity, and socioeconomic factors on the expansion of the dengue virus transmission zone in 21st century Brazil: an ecological modelling study. *The Lancet Planetary Health*, 5, S14. https://doi.org/10.1016/S2542-5196(21)00098-X.

Lima, M. E. S., Bachur, T. P. R. & Aragão, G. F. (2019). Guillain-Barre syndrome and its correlation with dengue, zika and chikungunya viruses infection based on a literature review of reported cases in Brazil. *Acta Tropica*. 197(105064), 1-4. https://doi.org/10.1016/j.actatropica.2019.105064.

Moraes, B. C., Souza, E. B., Sodré, G. R. C., Ferreira, D. B. S. & Ribeiro, J.B.M. (2019). Sazonalidade nas notificações de dengue das capitais da Amazônia e os impactos do El Niño / La Ninã. *Cadernos de Saúde Pública*. 35(9), 1-7. https://doi.org/10.1590/0102-311X00123417.

Oliveira, A. A. (2019). *Variabilidade Climática e Casos de Dengue em Municípios do Estado da Paraíba, Brasil*. Monografia, Curso de Graduação em Geografia, Universidade Federal de Campina Grande, Campina Grande, Paraíba, 23 p.

Oliveira-Jr, J. F., Gois, G., Silva, E. B., Teodoro, P. E., JohRAN, J. A. & Silva-Jr, C.A. (2019). Non-parametric tests and multivariate analysis applied to reported dengue cases in Brazil. *Environmental Monitoring and Assessment*. 191(473), 1-19. https://doi.org/10.1007/s10661-019-7583-0.

Oliveira-Jr, J. F., Souza, P. H. A., Souza, E. O., Vanderlei, M. H. G. S., Correia-Filho, W. L. F., Santos, C. T. B., Batista, B. A., Santiago, D. B. & Gois, G. (2021). Climatologia da chuva em Maceió: Aspectos Climáticos e Ambientais. *Revista Brasileira de Geografia Física*. 4(4), 2253-2264. https://doi.org/10.26848/rbgf.v14.4.p2253-2264.

Samet, H., Reisi, M. & Marzbani, F. (2019). Evaluation of neural network-based methodologies for wind speed forecasting. *Computers and Eletrical Engineering*. 78(1), 356-372. https://doi.org/10.1016/j.compeleceng.2019.07.024.

Santos, D. A. S., Rodrigues, J. Z., Olinda, R. A. & Goulart, L.S. (2018). Relação das variáveis climáticas com os casos de dengue em um município do interior de Mato Grosso dos anos 2001 a 2015. *Multitemas, Campo Grande, MS*. 23(55), 5-24. https://doi.org/10.20435/multi.v23i55.1742.

Santos, I. G. S., Lyra, R. F. F. & Silva-Jr, R. S. (2020). Comparativo de prognósticos da velocidade do vento utilizando modelo WRF e rede neural artificial. *Revista Brasileira de Meteorologia*. 35(Especial), 1017-1027. https://doi.org/10.1590/0102-77863550103.

Santos, I. G. S., Oliveira-Jr, J. F., Sousa, F. A. S. & Barbosa, I.I. (2022). Previsão de focos de calor na região metropolitana de Maceió utilizando rede neural artificial. *Revista Brasileira de Geografia Física*. 15(5), 2313-2326. 10.26848/rbgf.v15.5.p2313-2326.

Santos, L. L. S., Moura, E. L., Ferreira, J. M., Santos, B. R. C., Santos, A. C. M. & Figueiredo, E. V. M. S. (2016). Análise epidemiológica da dengue em uma população do nordeste. *Revista de Enfermagem UFPE On Line*. 10(6), 1944-1956. https://doi.org/10.5205/1981-8963-v10i6a11205p1944-1956-2016.

Santos-Jr, C. J. & Silva, J. P. (2019). Epidemiologia, fatores climáticos e distribuição espacial da dengue em uma capital do Nordeste do Brasil. *Revista Brasileira de Climatologia*. 25(15), 755-768. http://dx.doi.org/10.5380/abclima.v25i0.69421.

Silva, E. B., Raposo, J. C. S., Oliveira-Jr, J. F., Correia-Filho, W. L. F. & Santiago, D. B. (2021). Diagnóstico dos casos de dengue nas capitais do Nordeste do Brasil entre 2000 e 2017. *Caderno de Geografia*. 31(65), 546-556. https://doi.org/10.5752/P.2318-2962.2021v31n65p546.

Silva, E. T. C. (2020). *Modelos de regressão espacial ajustados a dados de arboviroses (Aedes Aegypti) do estado da Paraíba: influência de fatores* socioeconômicos. Dissertação de Mestrado em Saúde Pública, Universidade Estadual da Paraíba, Campina Grande, Paraíba, 76 p.

Silva, S. D., Oliveira-Jr, J. F., Correia-Filho, W. L. F., Barros, H. G., Souza, E. O., Santiago, D. B., Silva, E. B. & Silva, M. B. (2021). Dinâmica dos casos notificados de dengue em Alagoas: Geoespacialização e Estatística Aplicada. *Research, Society and Development*. 10(15), 1-16. http://dx.doi.org/10.33448/rsd-v10i15.22990.

Silva, W. R. S. (2018). *Metodologia de Monitoramento de Epidemias: Uma Abordagem Baseada em Redes Neurais Artificiais*. Dissertação de Mestrado em Engenharia Elétrica, Universidade Federal do Pará, Belém, Pará, 91 p.

Soares, A. P. M. R., Alves, P. H. P., Martins, I. C., Barreto, L. C. & Carvalho, F. O. (2018). Prognóstico da incidência de casos de dengue na cidade de Salvador-Bahia, utilizando a transformada de wavelet discreta em conjunção com redes neurais artificiais. *Interfaces Científicas - Saúde e Ambiente*. 6(3), 53-62. https://doi.org/10.17564/2316-3798.2018v6n3p53-62.

Souza, A., Abreu, M. C. & Oliveira-Jr, J. F. (2021). Impact of climate change on human infectious diseases: Dengue. *Brazilian Archives of Biology and Technology*. 64(1), 1-14, e21190502. https://doi.org/10.1590/1678-4324-2021190502.