v.5, n.10, 2024

RECIMA21 - REVISTA CIENTÍFICA MULTIDISCIPLINAR
ISSN 2675-6218

A STUDY OF THE STANDARDIZED PRECIPITATION INDEX (SPI) AND MACHINE LEARNING TECHNIQUES FOR DROUGHT PREDICTION IN THE STATE OF PARAÍBA, BRAZIL

UM ESTUDO DO ÍNDICE PADRONIZADO DE PRECIPITAÇÃO (SPI) E TÉCNICAS DE APRENDIZADO DE MÁQUINA PARA PREDIÇÃO DE SECA NO ESTADO DA PARAÍBA, BRASIL

UN ESTUDIO DEL ÍNDICE ESTANDARIZADO DE PRECIPITACIÓN (SPI) Y TÉCNICAS DE APRENDIZAJE AUTOMÁTICO PARA LA PREDICCIÓN DE SEQUÍAS EN EL ESTADO DE PARAÍBA, BRASIL

Jefferson Vieria dos Santos[1], Viviane Farias Felipe[2], Erika Fialho Morais Xavier[3], Tiago Almeida de Oliveira[4], Jader Silva Jale[5], Silvio Fernando Alves Xavier Junior[6]

e5105736
https://doi.org/10.47820/recima21.v5i10.5736

RECEIVED: 08/23/2024 APPROVED: 09/23/2024 PUBLISHED: 10/02/2024

ABSTRACT
This study aimed to identify and analyze droughts in Paraíba, using the Standardized Precipitation Index (SPI) and machine learning algorithms for predicting SPI for the subsequent years (2020-2021) at six rainfall stations distributed across the mesoregions of Paraíba. The Precipitation data were downloaded from the Global Precipitation Climatology Centre (GPCC) and the National Oceanic and Atmospheric Administration (NOAA) database, covering the period from 1991 to 2019. Three machine learning algorithms were selected based on their ability to fit historical SPI data: Extra Trees Regressor, Gradient Boosting Regressor, and Random Forest Regressor. The applied machine learning models yielded satisfactory results, with the Extra Trees Regressor consistently producing the highest R^2 value across all stations, indicating high data explainability. The predictions were analyzed to determine their accuracy and reliability, providing valuable insights into precipitation variability and drought occurrence in different mesoregions of Paraíba. In conclusion, this study contributed to understanding climate variability and its implications in Paraíba, offering valuable insights into drought occurrence and the importance of adaptive approaches to mitigate adverse impacts. The application of SPI and machine learning techniques proved effective in analyzing and predicting precipitation, providing an objective approach to characterizing drought and rainfall intensity in specific regions.

KEYWORDS: Standardized precipitation index. Drought. Machine learning.

RESUMO
Objetivou-se neste estudo identificar e analisar as secas na Paraíba, utilizando o Índice Padronizado de Precipitação (SPI) e técnicas de modelagem com algoritmos de machine learning para prever o SPI para os anos subsequentes (2020-2021) em seis estações pluviométricas distribuídas nas mesorregiões da Paraíba. Os dados de precipitação foram obtidos a partir do Global Precipitation Climatology Centre (GPCC) e da base de dados da National Oceanic and Atmospheric Administration (NOAA), abrangendo o período de 1991 a 2019. Foram selecionados três algoritmos de *machine learning* com base em sua capacidade de ajuste aos dados históricos de SPI: Extra Trees Regressor, Gradient Boosting Regressor e Random Forest Regressor. Os modelos de *machine learning* aplicados apresentaram resultados satisfatórios, com destaque para o Extra Trees Regressor, que consistentemente produziu o maior valor de R^2 em todas as estações, indicando uma alta explicabilidade dos dados. As previsões foram analisadas para determinar sua precisão e confiabilidade, fornecendo insights valiosos sobre a variabilidade da precipitação e a ocorrência de secas nas diferentes mesorregiões da Paraíba. Em conclusão, este estudo contribuiu para a

[1] Universidade Federal Rural de Pernambuco. Mestrando no programa de Pós Graduação em Biometria e Estatística Aplicada.
[2] Universidade Federal Rural de Pernambuco.
[3] Pesquisadora na CIDACS-FIOCRUZ.
[4] Professor do Departamento de Estatística - UEPB - Universidade Estadual da Paraiba.
[5] Professor do Departamento de Estatística e Informática - UFRPE.
[6] Universidade Estadual da Paraíba.

compreensão da variabilidade climática e de suas implicações na Paraíba, fornecendo insights valiosos sobre a ocorrência de secas e a importância de abordagens adaptativas para mitigar impactos adversos. A aplicação do SPI e técnicas de *machine learning* mostrou-se eficaz na análise e previsão da precipitação, oferecendo uma abordagem objetiva para caracterizar a intensidade das secas e chuvas em determinadas regiões.

PALAVRAS-CHAVE: Índice Padronizado de Precipitação. Seca. Machine Learning.

RESUMEN
Este estudio tuvo como objetivo identificar y analizar las sequías en Paraíba, utilizando el Índice Estandarizado de Precipitación (SPI) y algoritmos de aprendizaje automático para predecir el SPI para los años subsiguientes (2020-2021) en seis estaciones pluviométricas distribuidas por las mesorregiones de Paraíba. Los datos de precipitación fueron descargados del Centro de Climatología de Precipitación Global (GPCC) y de la base de datos de la Administración Nacional Oceánica y Atmosférica (NOAA), cubriendo el período de 1991 a 2019. Se seleccionaron tres algoritmos de aprendizaje automático en función de su capacidad para ajustar los datos históricos del SPI: Extra Trees Regressor, Gradient Boosting Regressor y Random Forest Regressor. Los modelos de aprendizaje automático aplicados produjeron resultados satisfactorios, con el Extra Trees Regressor presentando consistentemente el mayor valor de R^2 en todas las estaciones, lo que indica una alta explicabilidad de los datos. Las predicciones se analizaron para determinar su precisión y confiabilidad, proporcionando valiosas ideas sobre la variabilidad de la precipitación y la ocurrencia de sequías en diferentes mesorregiones de Paraíba. En conclusión, este estudio contribuyó a la comprensión de la variabilidad climática y sus implicaciones en Paraíba, ofreciendo valiosas ideas sobre la ocurrencia de sequías y la importancia de enfoques adaptativos para mitigar impactos adversos. La aplicación del SPI y de técnicas de aprendizaje automático resultó eficaz para analizar y predecir la precipitación, proporcionando un enfoque objetivo para caracterizar la sequía y la intensidad de la lluvia en regiones específicas.

PALABRAS CLAVE: Índice estandarizado de precipitación. Sequía. Aprendizaje automático.

INTRODUCTION

Understanding and predicting hydrometeorological events is essential for the scientific community and society. Extreme precipitation events, such as droughts and floods, significantly impact various economic and social sectors, making it essential to develop accurate and reliable methods for their assessment. However, more than conventional statistical techniques are often required to model and predict these complex phenomena due to the estimation errors inherent to these methods (Hasan *et al.,* 2012).

Due to its recurrence and complexity, drought plays a prominent role in hydrometeorological phenomena. In addition to directly influencing agricultural activities, drought has socioeconomic, environmental, and political implications, affecting several world regions. In particular, arid and semi-arid regions are particularly vulnerable to the devastating effects of drought (Mathbout *et al.,* 2018; Wang *et al.,* 2014). In the Brazilian context, the Northeast region, notably Paraíba, faces prolonged periods of drought, motivating climate and meteorological research.

The intersection between statistics and meteorology has proven fundamental for analyzing these complex phenomena. The application of statistical methods, such as the Standardized

Precipitation Index (SPI), has stood out as a crucial tool in analyzing climate variability, especially in drought. The SPI, which quantifies precipitation deviations at different temporal scales, offers an objective approach to characterize the intensity of droughts and rainfall in certain regions (Macedo *et al.,* 2010; Macedo *et al.,* 2011; De Sousa Guedes *et al.,* 2012; De Farias *et al.,* 2014; Santos *et al.,* 2019; Da Silva *et al.,* 2021).

The interdisciplinarity between statistics and meteorology, combined with mastery of computational tools such as R and Python, is essential for training researchers to face current climate challenges.

In this context, this work aimed to apply the Standardized Precipitation Index (SPI) and do modeling with machine learning techniques and algorithms to predict the SPI for the subsequent year (2020-2021) in the six rainfall stations distributed in the mesoregions of Paraíba. Each station represents a specific mesoregion of the state, namely Mata Paraibana (João Pessoa), Agreste Paraibano (Campina Grande), Borborema (Monteiro and Picuí), and Sertão Paraibano (Patos and Sousa). Analysis of the SPI in these different mesoregions will allow a more detailed understanding of precipitation variability and the occurrence of droughts at different temporal and geographic scales.

METHODOLOGY

Dataset

The data used in this study were obtained from the Global Precipitation Climatology Center(GPCC available at https://psl.noaa.gov/data/gridded/data.gpcc.html), a renowned repositor widely used in climate research, and part of the extensive database from the National Oceanic and Atmospheric Administration (NOAA). These data sources are known for their reliability and comprehensiveness, providing valuable information about climate conditions and precipitation variations over time. The time series of precipitation data covered 1991 to 2019, allowing a comprehensive analysis of climate trends.

Study region

The mesoregions selected for this study in Paraíba exhibit a rich climatic diversity influenced by their geographic location and topographic characteristics. The *Mata Paraibana*, represented by the city of João Pessoa, is characterized by a humid tropical climate with rainfall well distributed throughout the year. The *Agreste Paraibano*, represented by Campina Grande, has a tropical climate with well-defined dry and rainy seasons. The *Borborema region*, represented by Monteiro and Picuí, has a semi-arid tropical climate with scarce and irregular rainfall. Finally, the *Sertão Paraibano*, represented by Patos e Sousa, is characterized by a semi-arid climate, with even scarcer rainfall and concentrated quickly (Paraíba Total, 2021). These different climatic conditions between mesoregions result in notable variability in precipitation patterns throughout the year.

RECIMA21 - REVISTA CIENTÍFICA MULTIDISCIPLINAR
ISSN 2675-6218

Standardized Precipitation Index (SPI)

The Standardized Precipitation Index (SPI) is a statistical tool widely used to evaluate drought conditions or excess rainfall in a specific region. This index is based on analyzing the historical distribution of precipitation over time and the standard deviation of past rainfall data. Using the SPI, it is possible to quantify the intensity, duration, and frequency of water scarcity events or intense rainfall. Initially developed by McKee *et al.,* (1993), the SPI aims to measure the deficit or excess of precipitation on different temporal scales (1, 3, 6, 12, 24, and 48 months). This methodology is widely supported by the World Meteorological Organization (WMO).

When a normal distribution is assumed, the SPI is expressed in standard deviations, representing the discrepancies between observed precipitation records and long-term averages. Since precipitation does not follow a normal distribution, a transformation is applied to fit the data to a normalized probability distribution. This normalization process makes it possible to compare drier and wetter climates equally, thus enabling the analysis of both rainy and dry periods (Fernandes *et al.,* 2009).

This index is calculated using the statistical characteristics of historical precipitation series in different temporal intervals and probability density concepts. Statistical distributions such as Normal and Gamma were considered to adjust the frequency distribution of total precipitation in a specific rainfall series.

The following function represents the Gamma probability distribution:

$$g(x) = \frac{1}{\beta^{\alpha}\,\delta(\alpha)} x^{\alpha-1} e^{\frac{-x}{b}} \qquad (1)$$

In which α > 0 represents the shape parameter, b > 0 is the scale parameter, and x > 0 is the amount of rainfall and $\delta(\alpha)$ is the Gamma function given by the following expression:

$$\delta(\alpha) = \int_0^{\infty} y^{\alpha-1} e^{-y}\, dy \qquad (2)$$

In the formula mentioned, the temporary variable y can vary from 0 to ∞. The coefficients are determined using the maximum likelihood method, as outlined by THOM (1966), and are calculated using the following equations:

(3)

$$A = \ln\ln(\overline{x}) - \frac{\sum_i^n x_i}{n}$$

(4)

$$\hat{\alpha} = \frac{1}{4A}\left(1 + \sqrt{1 + \frac{4A}{3}}\right), \qquad \hat{b} = \frac{\bar{x}}{\hat{\alpha}}$$

The sample size is represented by n, and the average value of precipitation data is denoted as x. The cumulative distribution function of the Gamma distribution is defined as follows:

(5)

$$G(x) = \frac{1}{\delta(\hat{\alpha})} \int_0^x t^{\hat{\alpha}-1} e^{-t} dt$$

The gamma probability density function, g, is not defined for x=0. However, considering that the precipitation data sample may include zero values, the cumulative probability is expressed as follows:

(6)

$$H(x) = q + (1-q)G(x)$$

In Eq(6), "q" represents the probability of occurrence of a zero value, which is calculated as q = m/n. In this context, "m" represents the number of events in which precipitation equals zero, while "n" is the total number of observations. We can simplify statistical analysis by transforming the cumulative probability H(x) into a standardized random variable (Z) with a mean of zero and a standard deviation of one. The variable "Z" corresponds to the Standardized Precipitation Index (SPI) value. According to Abramowitz and Stegun (1965), the relationship between the Gamma and Normal probability distributions is represented by:

(7)

$$Z = -\left(t - \frac{c_0 + c_1 t + c_2 t^2}{1 + d_1 t + d_2 t^2 + d_3 t^3}\right) \quad 0 < H(x) \leq 0{,}5$$

RECIMA21 - REVISTA CIENTÍFICA MULTIDISCIPLINAR
ISSN 2675-6218

$$Z = +(t - \frac{c_0+c_1t+c_2t^2}{1+d_1t+d_2t^2+d_3t^3}\ 0{,}5 < H(x) \leq 1 \tag{8}$$

,

$$t = \sqrt{ln ln(\frac{1}{(H(x))^2})} 0 < H(x) \leq 0{,}5 \tag{9}$$

$$t = \sqrt{ln ln(\frac{1}{(1-H(x))^2})} 0{,}5 < H(x) \leq 1 \tag{10}$$

The Gamma probability density function must be adjusted to the series of monthly precipitation totals to start calculating the SPI. Then, the cumulative probability of occurrence of each monthly total is estimated. Subsequently, the Normal function and the inverse of the Gaussian distribution are applied to this probability to obtain the corresponding SPI value. The drought event is identified when the SPI value becomes negative and reaches the -1 mark. According to the methodology proposed by McKee *et al.,* (1993), this event ends when the SPI returns to positive values. On the index scale, values less than or equal to -2 indicate "extreme drought, " while values greater than or equal to +2 signal conditions of "extreme humidity”.

Table 1- SPI classification

SPI values	Classification
SPI ≥ 2,00	Extremely humid
1,00 ≥ SPI >2,00	Very humid
0,50≥SPI>1,00	Moderately humid
-0,50 > SPI >0,50	Aproximately normal
-1,00 > SPI ≥ -0,50	Moderately dry
-2,00 > SPI ≥ -1,00	Very dry
SPI ≤ -2,00	Extremely dry

Source: FERNANDES et al., (2009)

Modeling

The study was structured in several stages to consider a comprehensive and accurate analysis of the SPI forecast for 2020 and 2021 in the six rainfall stations distributed in the mesoregions of Paraíba. Historical SPI data were initially collected at the six rainfall stations, considering time scales of 9, 12, and 24 months. This data served as the basis for training and testing machine learning models. Using the PyCaret library, three machine-learning algorithms were selected based on their ability to fit historical SPI data. The algorithms chosen were Extra Trees Regressor, Gradient Boosting Regressor, and Random Forest Regressor. After training and tuning, the models were applied to predict the SPI for the year 2020. The predictions were then analyzed to determine their accuracy and reliability, providing valuable insights into precipitation variability and the occurrence of droughts in the different mesoregions of Paraíba. Below, the models used in this study are presented in more detail.

Extra Trees Regressor

The Extra Trees algorithm, or Extremely Randomized Trees, is a machine learning method in the random forest category. It benefits regression and classification tasks by offering a more random approach to creating decision trees (Geurts, 2006).

At each node of the tree, a random selection of a certain number of attributes is made, and among them, the best one is chosen. At the extreme, the method opts for a single attribute and cutoff point randomly at each node, resulting in fully randomized trees whose structures do not depend on the values of the target variable in the training set, as highlighted by Geurts (2006). In other words, in the Extra Trees algorithm, the selection and division of nodes are carried out even more randomly compared to other random forest methods, hence the origin of the term "extremely" in its name. After the random selection of candidate variables for the initial node, the existing values for each variable are divided at random. This procedure is replicated on each subsequent child node until a leaf node is reached. Predictions from all trees are then combined to define the final prediction based on majority voting.

According to Geurts (2006), one of the main advantages of Extra Trees is its ability to improve model accuracy due to its additional randomness. Furthermore, it effectively reduces overfitting, providing more generalized and robust models for various machine-learning tasks.

Random Forest Regressor

The Random Forest algorithm is a class of ensemble methods explicitly developed for decision tree classifiers. It aggregates the predictions made by multiple decision trees, each being produced related to the values of an independent set of random vectors, according to Tan, Steinbach and Kumar (2005).

RECIMA21 - REVISTA CIENTÍFICA MULTIDISCIPLINAR
ISSN 2675-6218

Tan, Steinbach and Kumar (2005) describe that random forests use random vectors generated from a fixed probability distribution to build decision trees, where feature selection and a combination of predictions are made to reduce bias and increase the probability model robustness. They also discuss different strategies for incorporating randomness into the tree growth process and how choosing the number of traits influences the correlation and strength of trees within the forest. Furthermore, the authors point out that the effectiveness of random forests in classification accuracy is comparable to AdaBoost, with additional advantages of robustness and speed.

In Random Forests, each tree is developed during training by randomly selecting samples from the dataset using replacement. A random subset of features is selected to determine the split chosen at each node in the tree. This random approach to tree formation increases the diversity between them, enhancing the robustness and accuracy of the resulting model. After training, each tree in the forest makes a prediction for new inputs. The majority votes of the trees make the final decision in classification tasks or by the average of the predictions in regression tasks.

Gradient Boosting Regressor

Gradient Boosting Regressor is a powerful and efficient machine learning algorithm that uses the boosting technique to optimize prediction accuracy in regression tasks. It builds multiple prediction models sequentially, where each model tries to correct the errors of its predecessor.

Like Random Forest, Gradient Boosting Regressor (GBR) (Friedman, 2001) is a machine-learning technique based on tree averaging. However, instead of training many full high-variance trees that are averaged to avoid overfitting, GBR sequentially adds small trees, each with high bias. In each iteration, the new tree to be added explicitly focuses on the documents responsible for the remaining regression error.

The Gradient Boosting Regressor starts with a base model, which can be simple, and calculates the residuals (differences between predictions and actual values). Then, a new model is built to predict these residuals. This process is repeated several times, and each new model is added to the global model to improve the accuracy of predictions. The algorithm iteratively adjusts to minimize residual error, optimizing the predictive ability of the final model.

RESULTS AND DISCUSSION

Descriptive Analysis

Table 2 presents information on rainfall in six rainfall stations, including geographic and statistical data. João Pessoa has the highest average rainfall, recording an average of 130 millimeters, while Picuí has the lowest average, with 43 millimeters. It is essential to highlight that João Pessoa and Sousa have more significant standard deviations, indicating that these cities experience more significant variations in rainfall. Furthermore, the minimum and maximum values recorded show that rainfall can vary significantly from one month to another in all cities. João Pessoa has a minimum of 0 and a maximum of 640, while Picuí has a minimum of 0 and a maximum of 370.

	Latitude	Longitude	$\bar{x}$	sd	p25	p75	p100
Campina Grande	-7.23072	-35.8817	67	64	15	100	380
João Pessoa	-7.11532	-34.861	130	120	34	190	640
Monteiro	-7.8915	-37.1169	52	63	7.1	79	450
Picuí	-6.54727	-36.3621	43	60	1.1	60	370
Patos	-6.02556	-37.2779	64	88	3.8	94	520
Souza	-6.76782	-38.2099	68	91	3.6	46	590

Table 2 – Descriptive Statistics of Rainfall between 1991-01-01 and 2019-12-01.

$\bar{x}$ = mean; sd=Std deviation; p25=1st quartile(25th percentile); p75= 3rd quartile (75th percentile);

Figure 1 - Histograms of rainfall, A-João Pessoa, B-Campina Grande, C-Picuí, D-Monteiro, E-Patos, F -Souza.

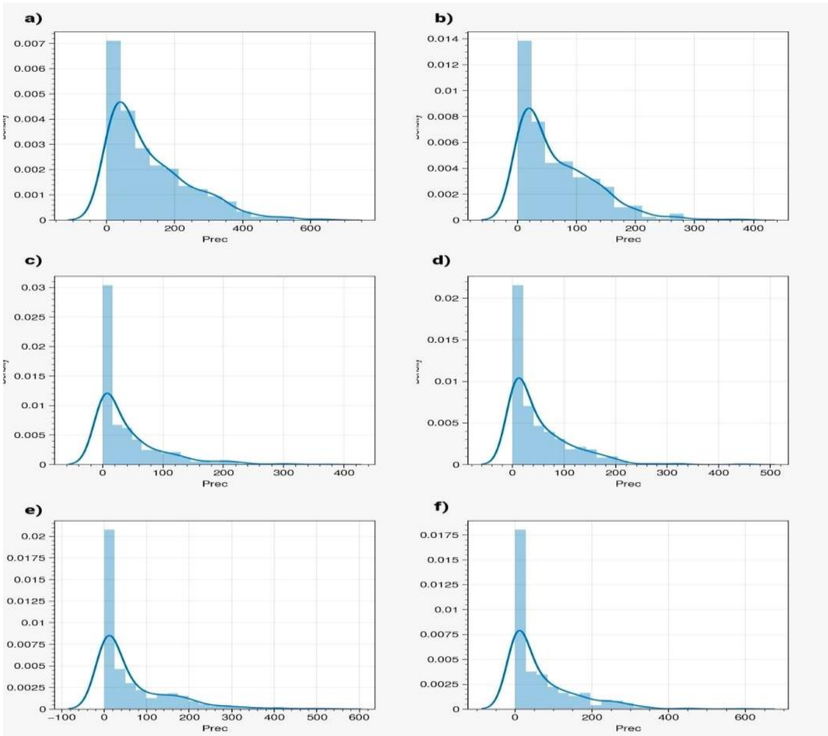
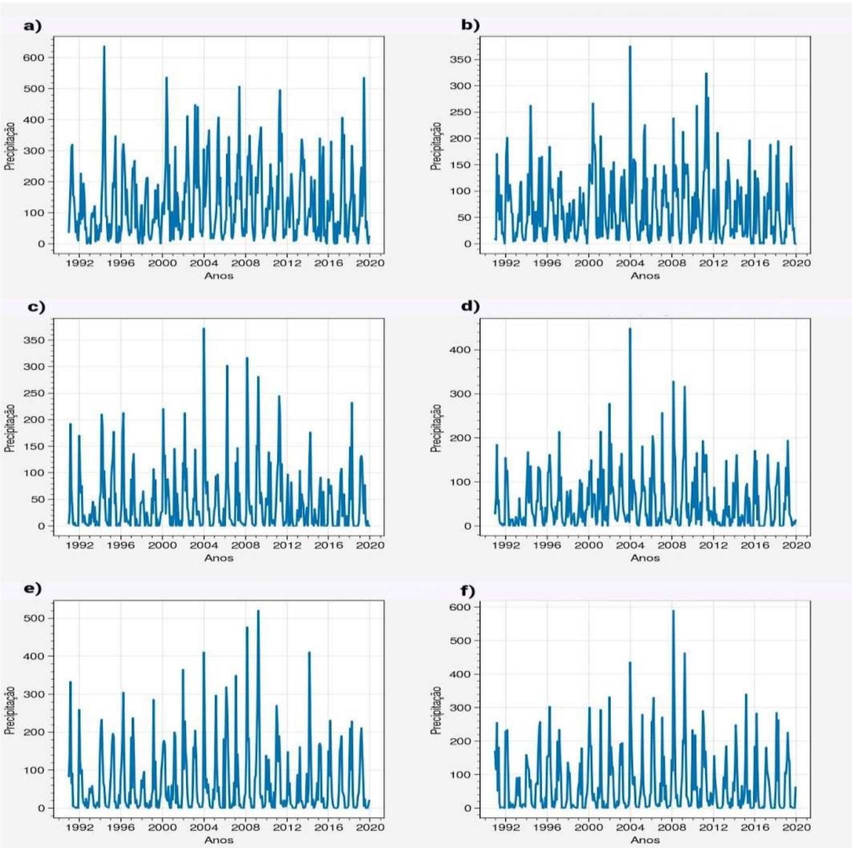


Figure 2 describes a time series illustrating the years of maximum rainfall in the cities subject to this study. Concerning João Pessoa, designated in the graph as (A), it is noteworthy that 1994 and 2019 were characterized by significant rainfall, reaching 636 mm and 536 mm, respectively. Similarly, for Campina Grande, labeled as (B), the peak rainfall years were 2004 and 2011, with precipitation rates of 375 mm and 324 mm, respectively.

Regarding Picuí (C), 2004 and 2008 emerge as the wettest, with precipitation volumes of 371 mm and 316 mm, respectively. In Monteiro (D), data indicate that 2004 and 2007 were marked by high precipitation levels, reaching 448 mm and 328 mm, respectively. As for Patos (E), 2007 and 2009 recorded the highest rainfall, totaling 476 mm and 520 mm, respectively. Finally, in Souza (F), 2008 and 2009 stood out with 589 mm and 462 mm rainfall volumes. In general, these data point to significant temporal variability regarding precipitation rates in the cities analyzed.

Figure 2 - Rainfall time series from 1991-01-01 to 2019-12-01.



Standardized Precipitation Index (SPI)

Through the application of the SPI, rain and drought events were identified on different temporal scales, including periods of 1, 3, 6, 9, 12, and 24 months. Each rainfall data collection station has its distinct characteristics and seasonal variations, which result in periods of abundant or scarce precipitation and variations in the intensity and duration of these periods.

However, even though each season presents particular variations, intense weather events tend to impact a wider area. When these events have a climatic character, they can be observed covering the entire study region, regardless of the temporal scale analyzed.

Figure 3 depicts the behavior of the SPI average with a single-month scale (SPI-1). The analysis reveals a notable monthly fluctuation, characterized by multiple positive and negative peaks of short duration and some of longer duration. Even within this one-month interval, it is possible to notice the presence of sequences of positive values in specific years in the stations analyzed. These sequences indicate the occurrence of rain events that will manifest themselves more clearly on broader temporal scales, meaning that their effects will be perceived later and vary according to the duration at different scales. Similarly, short-term dry periods can be observed on the negative side of the distribution.

Concerning João Pessoa, 2003, 2004, and 2013 presented the highest amounts of rain recorded. However, 2015, 2016, and 2018 were marked by intense droughts in the same region.

In Campina Grande, the years that stood out for the highest volume of rain were 2000, 2004, and 2011. In contrast, 2001, 2012, and 2016 were characterized by more pronounced droughts.

In Picuí, 2000, 2004, and 2008 stood out for the highest amounts of rain recorded, while 1993, 2012, and 2013 presented more severe levels of drought. Analyzing the city of Monteiro, 2004, 2007, and 2009 stood out as the rainiest periods. In contrast, 1993, 2001, and 2012 were notable for their intense droughts.

In the context of Patos, 1992, 2000, and 2009 were characterized by more significant volumes of precipitation. On the other hand, 1991, 2012, and 2013 were marked by periods of more pronounced drought. In Sousa, 2000, 2008, and 1996 presented a greater volume of rain than other years. However, the region was characterized by more severe droughts in 1997, 2016, and 2017. Table 3 displays the maximum and minimum values for each station.

RECIMA21 - REVISTA CIENTÍFICA MULTIDISCIPLINAR
ISSN 2675-6218

Figure 3 - SPI 1

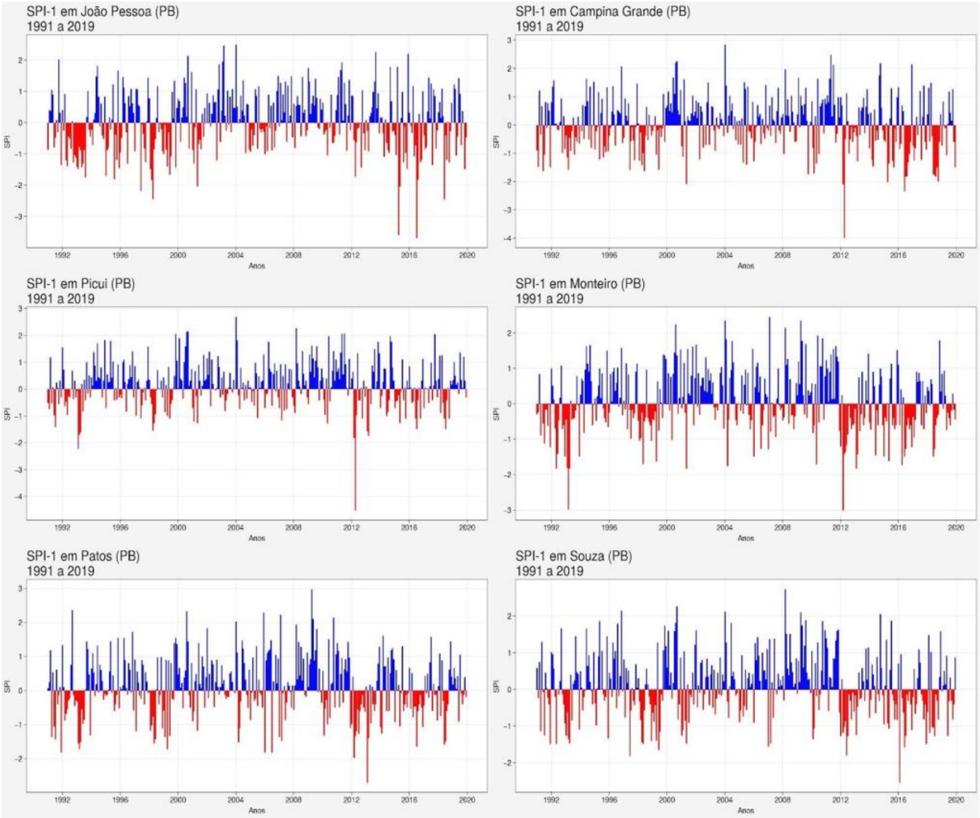


Table 3 - SPI – 1

SPI-1	**Maximum**	**Minimum**
Campina Grande	2.853542	-3.994498
João Pessoa	2.498881	-3.685909
Monteiro	2.458819	-3.004865
Picuí	2.695183	-4.514741
Patos	2.975877	-2.707451
Souza	2.733314	-2.542402

In Figure 4, SPI-3 provides a clearer view of observations made on the monthly scale (SPI-1). It was observed that the previously mentioned events persisted and that the randomness on the smaller scale decreased. An important point to note in this 3-month scale is that it showed the intensity of the events more accurately, accumulating consecutive values of the same signal and expanding the peaks of the most relevant events.

According to Table 4, Campina Grande recorded the maximum value for SPI-3 in 2004. Regarding negative events, Campina Grande also presented the lowest value in 2016.

Figure 4 - SPI-3

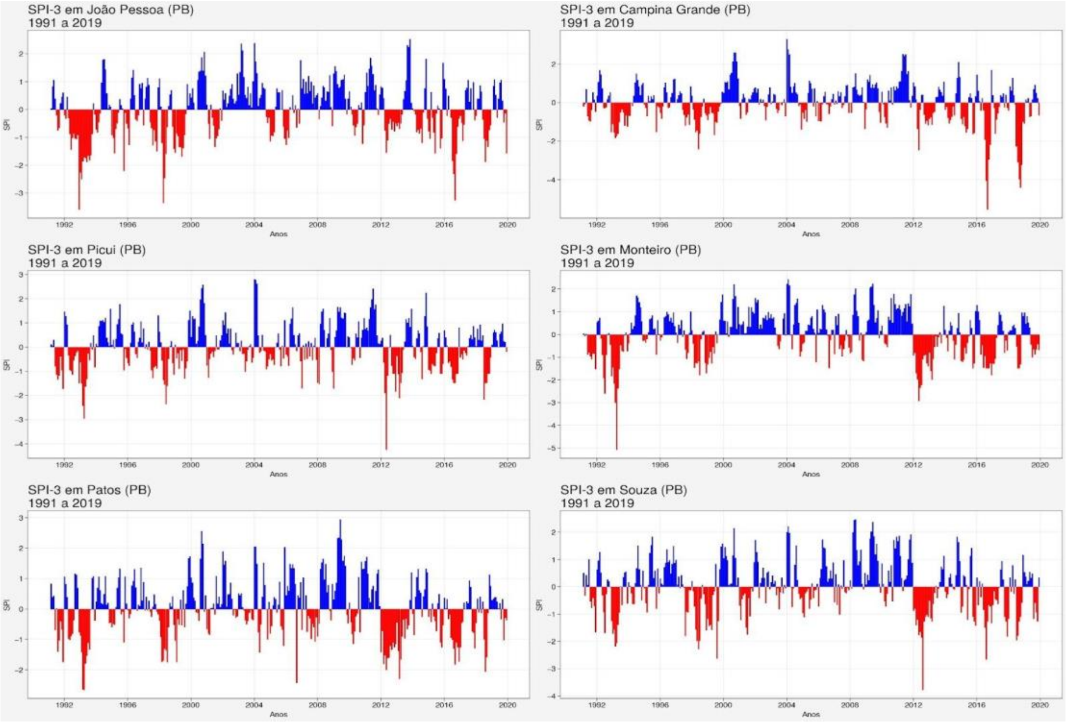


Table 4 - SPI – 3

SPI-3	Maximum	Minimum
Campina Grande	3.316800	-5.553479
João Pessoa	2.539662	-3.585842
Monteiro	2.442202	-5.070721
Picuí	2.816756	-4.243642
Patos	2.952973	-2.658167
Souza	2.478849	-3.776287

Figure 5 displays SPI-6; it was possible to identify the transition period between the rainy and dry periods. In Monteiro, a period of rain remains around an average without prominent peaks, unlike the dry period, which presents significant peaks predominant from 1993 onwards. There are more long rainy periods in the Campina Grande and João Pessoa stations than in the Patos and Souza stations. Likewise, the Patos and Souza stations are more prominent on this scale during the dry period than the others due to their location.

According to Table 5, Campina Grande presented the highest value for the adverse event in 2016. As for SPI-6, the highest positive SPI value recorded was in Patos in 2009.

Figure 5 - SPI-6

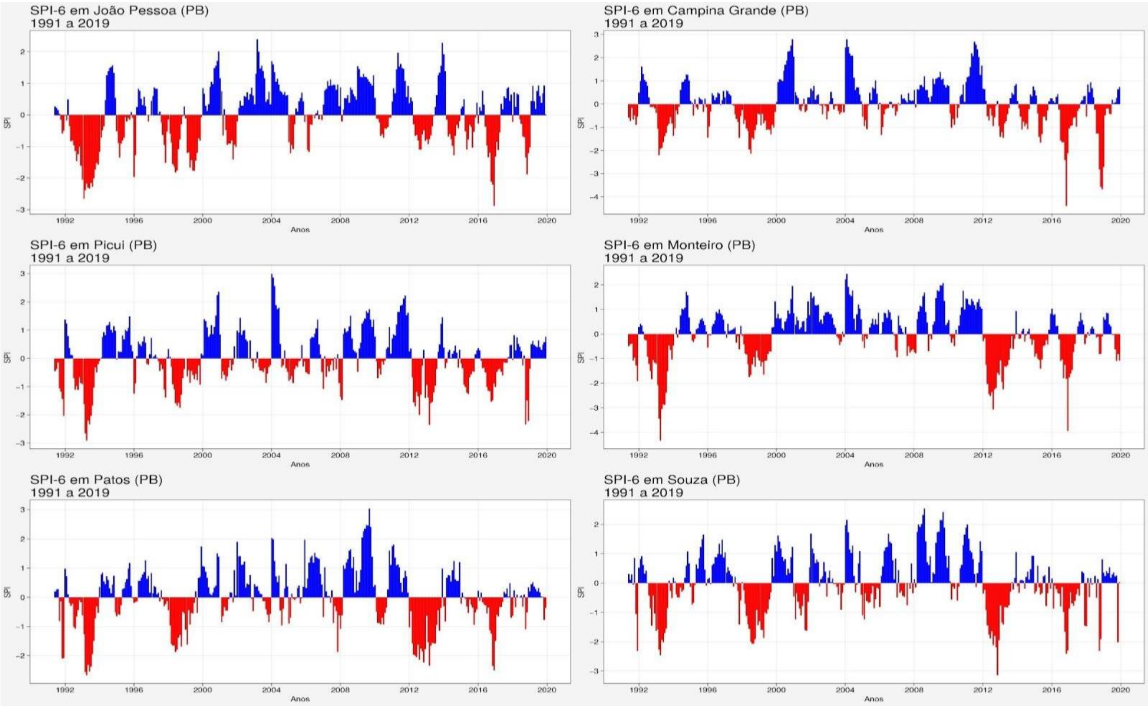


Table 5 - SPI-6

SPI-6	Maximum	Minimum
Campina Grande	2.805822	-4.382853
João Pessoa	2.407535	-2.875667
Monteiro	2.456994	-4.330417
Picuí	2.998535	-2.902734
Patos	3.046565	-2.664731
Souza	2.549074	-3.134210

RECIMA21 - REVISTA CIENTÍFICA MULTIDISCIPLINAR
ISSN 2675-6218

When using the SPI-9 shown in Figure 6, no significant differences were observed about SPI-6 in the regions' transition periods. However, this scale provided a better definition of the oscillation between positive and negative periods, maintaining the abovementioned critical events.

According to Table 6, for records of critical events in SPI-9, Monteiro obtained the highest negative precipitation value. At the same time, Patos presented the highest positive event, the lowest value, and the highest drought peak recorded in Monteiro in 2017.

Figure 6 - SPI-9

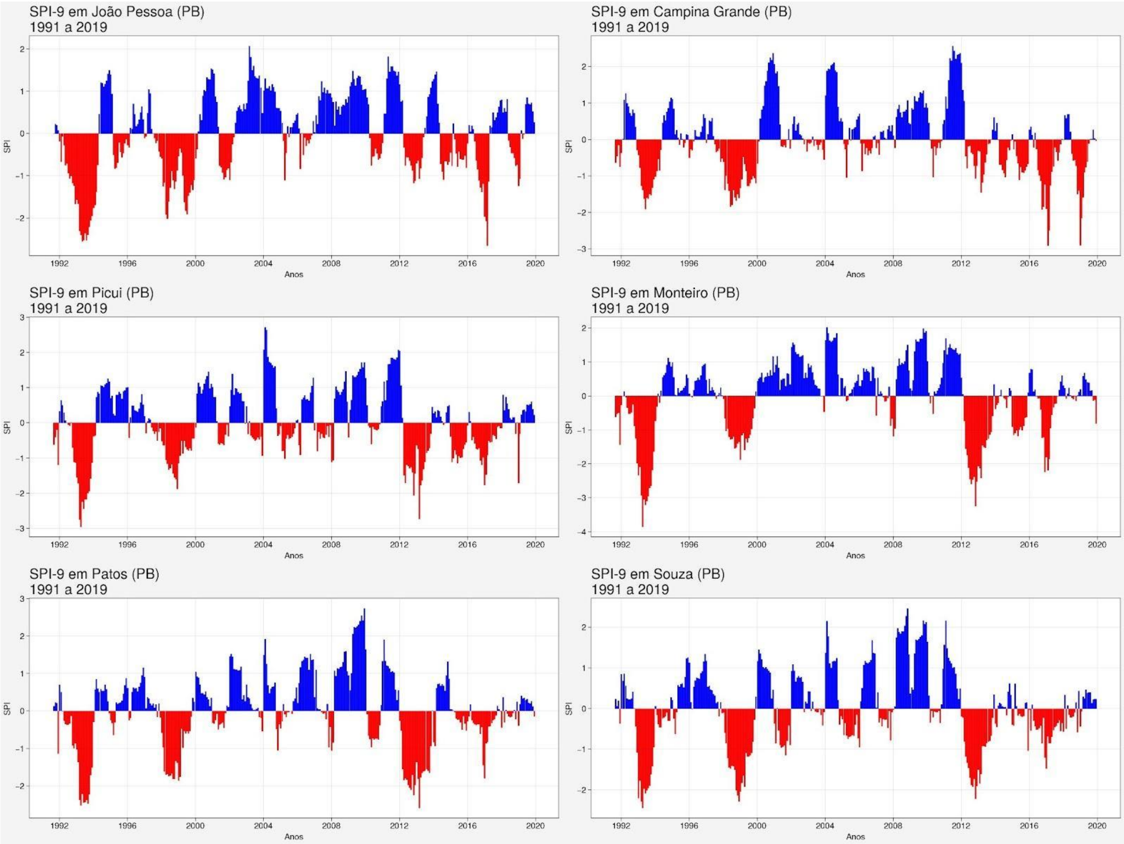


Table 6 - SPI-9

SPI-9	Maximum	Minimum
Campina Grande	2.569257	-2.909613
João Pessoa	2.076713	-2.656554
Monteiro	2.026951	-3.850753
Picuí	2.725501	-2.955701

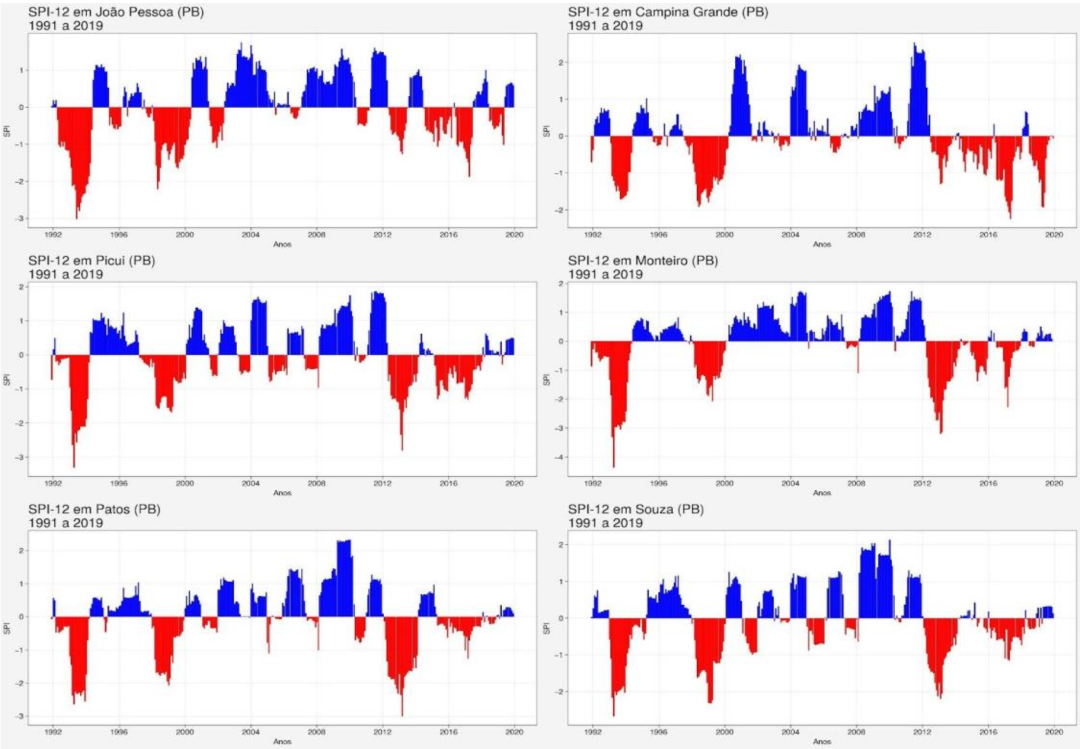
Patos	2.745890	-2.587995
Souza	2.465498	-2.441443

When analyzing the annual scale recorded by SPI-12 in Figure 7 for each season in the state, we can observe the following climate patterns. In João Pessoa, long rainy periods were recorded in 2003 and 2008, except for a significant dry period between 1992, 1998, and 2015. In Campina Grande, there was significant variability with positive and negative peaks, but from the decade of 2011, droughts became more frequent. In Monteiro, the long rainy periods between 2000 and 2012 stand out, while droughts occurred mainly between 1992 and 1994 and prevailed from 2012 to 2017. Picuí presents excellent rainfall variability, emphasizing the years from 1995 to 1997 and 2008 to 2009, dry periods between 1997 and 1999, and 2012 to 2017.

Variations in rainy periods were also observed in the municipalities of Sousa and Patos, located in the state's backlands. In Sousa, the drought periods in 1997 and 2012 stand out, while in Patos, drought became predominant from 2012 onwards.

Table 7 presents the maximum and minimum values recorded for each SPI-12 station in Agreste Paraibano. In Campina Grande, the most significant positive event was recorded in 2011, while Monteiro had the smallest event compared to the other stations. The minimum value regarding adverse drought-related events was recorded in Monteiro in 1993.

Figure 7 - SPI-12



RECIMA21 - REVISTA CIENTÍFICA MULTIDISCIPLINAR
ISSN 2675-6218

Table 7 - SPI-12.

SPI-12	Máximo	Mínimo
Campina Grande	2.546700	-2.246363
João Pessoa	1.751131	-3.014405
Monteiro	1.743142	-4.366061
Picuí	1.876263	-3.310416
Patos	2.329988	-2.999056
Souza	2.140688	-2.673100

The SPI-24 metric is a Standardized Precipitation Index calculated for a 24-month time window, thus providing a comprehensive perspective of water conditions—whether wet or dry—in a specific location. As illustrated in Figure 8, a significant variation in humidity conditions is observed in six selected meteorological stations from 2000 to 2012, predominantly characterized by periods of high humidity. In contrast to this interval, data collected between 2013 and 2019 reveal a tendency for drought conditions to intensify.

In Table 8, the variation range of SPI-24 in different meteorological stations points to significant fluctuations in water conditions. For example, in Campina Grande, the maximum SPI-24 value was recorded as 1.924587, while the minimum was -2.138502. This range suggests that this region experienced both periods of significant wetness and intense drought throughout the analysis period.

Similarly, João Pessoa and Monteiro present maximum and minimum values that indicate dry and wet conditions. The range between maximum and minimum values is extensive in João Pessoa, where SPI-24 varies from 1.800698 to -2.963429, emphasizing the extreme variability of water conditions in the locality. It is also essential to highlight the locations of Patos and Souza, where the highest maximum SPI-24 values were recorded, being 2.518448 and 2.563850, respectively. These values may indicate episodes of exceptionally high humidity.

Figure 8 - SPI-24.

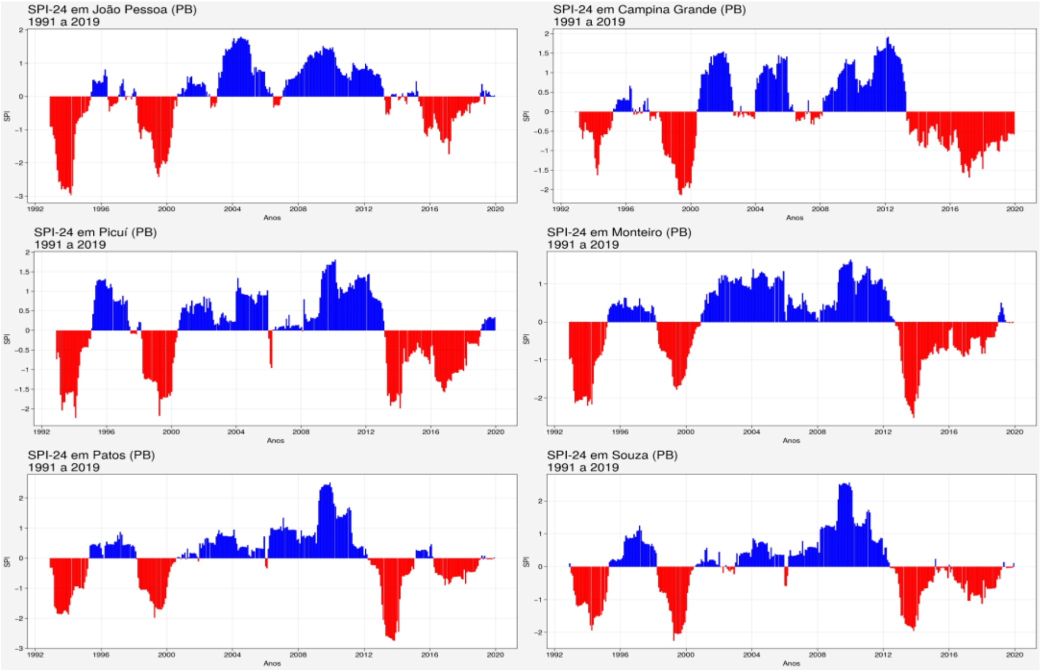


Table 8 - SPI-24

SPI-12	Maximum	Minimum
Campina Grande	1.924587	-2.138502
João Pessoa	1.800698	-2.963429
Monteiro	1.641876	-2.522586
Picuí	1.812801	-2.229101
Patos	2.518448	-2.742311
Souza	2.563850	-2.249191

Machine-Learning Modeling

Analyzing the results in Table 9, obtained by the Machine Learning models applied - Extra Trees Regressor (ET), Gradient Boosting Regressor (GBR), and Random Forest Regressor (RF) - in the six stations (Campina Grande, João Pessoa, Monteiro, Picuí, Patos and Souza), it is possible to discuss the performance metrics (MAE, MSE, RMSE, R^2, RMSLE and MAPE).

In Campina Grande, the ET stood out with an R^2 of 0.7515, indicating that the model could explain 75.15% of the SPI 9 precipitation data variance. This model also presented the lowest value for the error metrics MAE, MSE, and RMSE compared to the other models for that season. On the other hand, RF performed slightly worse than ET but was still relatively solid, with an R^2 of 0.5972. GBR was the model with the lowest performance this season, revealing the highest MAE, MSE, and RMSE and the lowest R^2 value (0.4489) among the three models.

In João Pessoa, ET obtained the best performance among the models with an R^2 of 0.7722, the model that best explains the data variability in João Pessoa. RF and GBR demonstrated similar performance in terms of R^2; however, GBR resulted in an RMSE and MAE slightly higher than rf, indicating that there may have been a more excellent dispersion of errors.

Looking more broadly, ET consistently produced the highest R^2 value across all seasons, indicating a high explainability of the data.RF exhibited consistent predictions, maintaining reasonably stable values for R^2 across all seasons.

The GBF proved to be the most variable model, with a notable difference in its performance between different seasons. It may indicate sensitivity to specific data characteristics or that the model parameters require a more refined adjustment.

For all seasons, the ET demonstrated, in general, lower values for MAE and MSE compared to the other models, which indicates better precision and reliability in the predicted data. ET presented lower RMSE values, indicating better model performance.

Table 9 - Performance Comparison of Machine Learning Models: Extra Trees Regressor (ET), Gradient Boosting Regressor (GBR) and Random Forest Regressor (RF) Applied to Six Meteorological Stations: Campina Grande, João Pessoa, Monteiro, Picuí, Patos and Sousa for the SPI-9.

Municipality	Model	MAE	MSE	RMSE	R^2	RMSLE	MAPE
Campina Grande	**et**	0.3011	0.1963	0.4291	0.7515	0.2022	2.2065
	rf	0.3702	0.3052	0.5392	0.5972	0.2487	3.7448
	gbr	0.4404	0.3626	0.5899	0.4489	0.2804	5.3000
João Pessoa	**et**	0.3024	0.2092	0.4373	0.7722	0.2076	0.9896
	rf	0.3852	0.3147	0.5375	0.6524	0.2469	1.7170
	gbr	0.4641	0.3961	0.6133	0.5566	0.2924	2.6099

Monteiro	**et**	0.2706	0.1629	0.3939	0.8440	0.1864	1.6337
	rf	0.3102	0.2300	0.4705	0.7656	0.2181	1.9983
	gbr	0.3554	0.2683	0.5120	0.7174	0.2314	2.6924
Picuí	**et**	0.3214	0.2194	0.4408	0.7763	0.2176	0.9097
	rf	0.3343	0.2724	0.4837	0.7257	0.2160	0.9919
	gbr	0.4018	0.3303	0.5475	0.6527	0.2573	1.2020
Patos	**et**	0.2772	0.1676	0.3881	0.8377	0.1953	1.0255
	rf	0.3221	0.2225	0.4532	0.7774	0.2238	1.5633
	gbr	0.3761	0.2694	0.5077	0.7211	0.2474	1.4474
Souza	**et**	0.2980	0.1605	0.3979	0.8170	0.2072	2.1041
	rf	0.3060	0.2029	0.4413	0.7644	0.2229	1.4333
	gbr	0.3672	0.2772	0.5149	0.6792	0.2601	1.848

Table 10 shows that the Extra Trees Regressor (ET) demonstrated a robust R2 of 0.8442 in Campina Grande, standing out as the model that best explained the data variability. It also showed robustness in other metrics, presenting the lowest values of MAE, MSE, and RMSE, reflecting its ability to minimize prediction errors. On the other hand, the Gradient Boosting Regressor (GBR) registered greater sensitivity and errors, mainly highlighted by an RMSLE of 0.2678 and a MAPE of 2.0509, indicating that its predictions may have been challenged by outliers or an asymmetric distribution of residuals.

In João Pessoa, ET again presents the best performance in terms of R2 (0.8631), however, it is vital to note a considerable variation in error metrics between the models, with GBR tending less accurate predictions, indicated by its higher MAE and MSE values, for example. Even with the highest R2, et presented a MAPE of 1.0797, implying that the model may still be susceptible to systematic errors and that an additional evaluation of the residuals may be constructive.

In Monteiro, the models reveal error metrics closer to each other, although ET prevails regarding overall fit (R2=0.8895). Here, GBR demonstrated a relative improvement in consistency, although it still presents the highest RMSLE and MAPE among the models.

In Picuí, the ET model stands out with a coefficient of determination (R2) of 0.8143, offering an adequate fit to the data pattern. Even so, when we observe the high MAPE (1.2998), which indicates a certain amplitude in the prediction errors in percentage terms, questions arise about the distribution of the error over the analyzed period. The GBR reveals the most significant volatility in the metrics, suggesting that it may be struggling with noise or less apparent patterns in the data, perhaps requiring a finer tuning of the parameters.

For Patos, the metrics portray a narrative where the et dominates regarding error and fit metrics (R2 of 0.8543). The GBR, presenting the lowest R2 and the highest values for MAE and MSE, indicates that its predictions were generally less accurate and consistent. With a MAPE of 1.4735, the ET still suggests that possible improvements can be explored, perhaps by integrating additional variables or adjusting to treat outliers.

Souza proves to be an interesting case. Although the ET presents a high R2 (0.8609), a look at the MAPE reveals a notable disparity (4.5121).

This divergence between a robust R2 and a high MAPE may indicate that while the model captures the overall variability of the data, it may be struggling to predict certain specific data points – possibly those with high SPI values, which would have a more pronounced impact on MAPE. While exhibiting solid overall performance, the Random Forest Regressor (RF) also exhibits a high MAPE (6.1354), indicating that systematic errors may be at play and warrant further investigation.

Table 10 - Performance Comparison of Machine Learning Models: Extra Trees Regressor (ET), Gradient Boosting Regressor (GBR) and Random Forest Regressor (RF) Applied to Six Meteorological Stations: Campina Grande, João Pessoa, Monteiro, Picuí, Patos and Sousa for the SPI-12.

Municipality	Model	MAE	MSE	RMSE	R2	RMSLE	MAPE
Campina Grande	**et**	0.2733	0.1367	0.3666	0.8442	0.1898	1.0606
	rf	0.3657	0.2789	0.5202	0.6612	0.2308	1.6131
	gbr	0.4329	0.3357	0.5733	0.6133	0.2678	2.0509
João Pessoa	**et**	0.2575	0.1263	0.3479	0.8631	0.1900	1.0797
	rf	0.3995	0.2910	0.5301	0.6846	0.2649	1.3565
	gbr	0.5132	0.4183	0.6411	0.5397	0.3342	1.5097
Monteiro	**et**	0.2294	0.1346	0.3482	0.8895	0.1596	12.5304
	rf	0.2862	0.2282	0.4605	0.8055	0.2142	15.9226
	gbr	0.3452	0.2531	0.4838	0.7867	0.2317	16.1387

RECIMA21 - REVISTA CIENTÍFICA MULTIDISCIPLINAR
ISSN 2675-6218

Picuí	**et**	0.3030	0.1874	0.4179	0.8143	0.2053	1.2998
	rf	0.3462	0.2485	0.4851	0.7421	0.2269	1.8960
	gbr	0.4627	0.3627	0.5852	0.6342	0.2766	2.4618
Patos	**et**	0.2671	0.1509	0.3799	0.8543	0.1966	1.4735
	rf	0.2885	0.2012	0.4388	0.7993	0.2070	1.9101
	gbr	0.3780	0.2860	0.5256	0.7150	0.2554	2.6850
Souza	**et**	0.2583	0.1371	0.3504	0.8609	0.1743	4.5121
	rf	0.2901	0.1983	0.4338	0.7798	0.2090	6.1354
	gbr	0.3640	0.2432	0.4827	0.7345	0.2385	6.2883

According to Table 11, the models for Campina Grande in the context of SPI 24 reflect a high capacity to explain the variance in the data, with the Extra Trees Regressor (ET) once again being the leading model with an R2 of 0.9542. However, the interpretation needs to be careful when observing the MAPE, which is relatively high (5.8258) and points to a significant percentage discrepancy in the predictions of this model. An investigation into the distribution of errors over time and at different SPI magnitudes would be a crucial step to understanding the origin of this discrepancy.

João Pessoa follows a similar narrative, with ET displaying a high R2 (0.9517) but also with a contrasting MAPE (0.9092). It is noted that the models are generally able to capture the patterns in the time series, but this high MAPE may indicate episodes of more expressive errors that deserve attention. The Gradient Boosting Regressor (GBR), while capturing a reasonable proportion of the data variability (R2 of 0.8269), its MAPE metric suggests that the accuracy of its predictions may be problematic in specific periods or SPI ranges.

In Monteiro, the et exhibits a robust R2 of 0.9726, showing strong adequacy to the variations in the SPI 24 data. The other metrics such as MAE, MSE and RMSE are also the lowest among the models, indicating the solidarity of this model in the predictions.

Picuí presents a picture where the et maintains good explainability (R2 of 0.9374). The MAPE of the GBR stands out for being the highest among the models (0.7079), indicating a more pronounced discrepancy in its predictions and worthy of further investigation into how its parameters are being influenced by the patterns in the data and whether a recalibration is necessary. For Patos, the models generally exhibit respectable performance, but the contrast between R2 and MAPE in et (0.9397 and 0.8697, respectively) suggests that the percentage errors in the predictions are not uniform and may be influenced by specific patterns or outliers in the data that are worth exploring further.

Finally, in Souza, while the models have a remarkable ability to map the variability of the data (with et leading with R2 of 0.9401), the MAPE suggests that additional refinements in the modeling or in the selection of variables could provide improvements, minimizing the percentage errors in the predictions.

Table 11 - Performance Comparison of Machine Learning Models: Extra Trees Regressor (ET), Gradient Boosting Regressor (GBR) and Random Forest Regressor (RF) Applied to Six Meteorological Stations: Campina Grande, João Pessoa, Monteiro, Picuí, Patos and Sousa for the SPI-24.

Município	Modelo	MAE	MSE	RMSE	R2	RMSLE	MAPE
Campina Grande	**et**	0.1436	0.0389	0.1920	0.9542	0.1037	5.8258
	rf	0.1958	0.0716	0.2630	0.9165	0.1438	5.8953
	gbr	0.2768	0.1401	0.3631	0.8375	0.1871	7.3998
João Pessoa	**et**	0.1437	0.0485	0.2059	0.9517	0.1086	0.9092
	rf	0.2313	0.1144	0.3325	0.8832	0.1612	2.6106
	gbr	0.3131	0.1756	0.4118	0.8269	0.1966	2.0440
Monteiro	**et**	0.1077	0.0261	0.1565	0.9726	0.0873	0.4166
	rf	0.1794	0.0672	0.2579	0.9306	0.1256	0.4778
	gbr	0.2402	0.1104	0.3286	0.8827	0.1663	0.9735
Picuí	**et**	0.1444	0.0546	0.2256	0.9374	0.1217	0.5456
	rf	0.2167	0.1049	0.3138	0.8788	0.1581	0.5177
	gbr	0.2941	0.1798	0.4180	0.7907	0.2051	0.7079
Patos	**et**	0.1437	0.0536	0.2253	0.9397	0.1158	0.8697
	rf	0.2010	0.0974	0.3071	0.8909	0.1493	1.0662
	gbr	0.2882	0.1698	0.4093	0.8130	0.2023	1.7234
Souza	**et**	0.1431	0.0482	0.2111	0.9401	0.1173	1.2024
	rf	0.1933	0.0909	0.2905	0.8914	0.1438	1.4080
	gbr	0.2678	0.1383	0.3656	0.8308	0.1863	1.3774

Predictions

We evaluated the forecasts related to the SPI, specifically in Campina Grande, João Pessoa, Monteiro, Picuí, Patos, and Sousa, aiming at a deep understanding of climate projections and their implications for 2020 and 2021. As an indicator widely used to characterize variation in precipitation and identify drought and excess rainfall events, the SPI offers us a systematic and enlightening view of the potential scenarios of variability and climate change in the regions analyzed. In this study, we chose to implement the Extra Trees (ET) model for the SPI forecast since it demonstrated a better fit for the data and a remarkable predictive capacity in scenarios of climate complexity. The forecasts for the years 2020 and 2021 were applied.

As shown in the graph in Figure 9, the SPI 9 forecast for Campina Grande shows a persistent drought pattern for 2020 and 2021, with predominantly negative SPI values. This trend indicates periods of drought, where values below -2 are of particular concern, representing extreme drought. This situation may imply water stress, affecting the region's water resources and agricultural production.

In contrast, João Pessoa exhibits relatively wetter conditions throughout both years, with SPI values mostly positive or close to zero. The hydrological trend here is more optimistic, with values indicating normal or wetter conditions that benefit agriculture and other water resource uses.

Monteiro shows a mixed situation, with SPI varying around zero, suggesting conditions fluctuate between dry and wet periods. The tendency for mild to moderate drought in the late years is worth noting, which may be a sign of a seasonal pattern that requires further investigation.

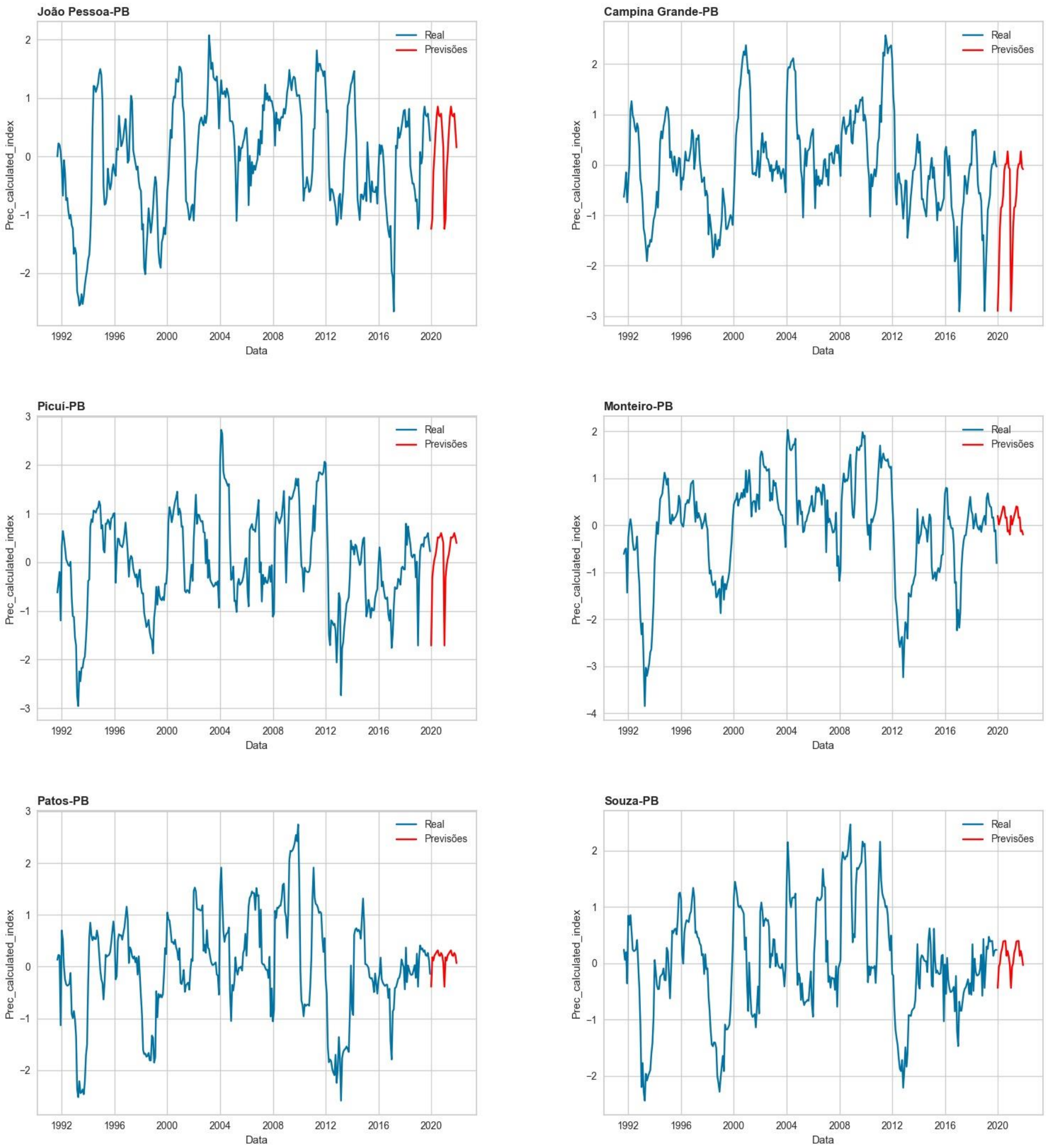
The city of Picuí presents an interesting pattern, starting the year with a significant drought but improving steadily throughout. Here, the early months of each year can be critical for farmers and water resource managers, and mitigation strategies for these periods may be necessary.

Patos has a predominantly positive trend in the SPI, suggesting wet or normal conditions throughout the two years. However, the end of 2021 suggests a trend towards drier conditions, which could indicate potential future problems if this trend continues.

The SPI forecasts in Souza oscillate notably, alternating between dry and wet conditions. This oscillating pattern may indicate significant regional water variability, presenting challenges in water resource management and agricultural planning.

RECIMA21 - REVISTA CIENTÍFICA MULTIDISCIPLINAR
ISSN 2675-6218

Figure 9 - Predictions for the SPI-9 via ET model for 2020, and 2021. Data in blue, predictions in red.



The SPI-12 forecasts provide a detailed perspective on precipitation trends and, consequently, on potential drought or excess precipitation scenarios. According to the graphs in Figure 10, we have the following results:

Campina Grande shows a predominantly negative SPI 12 trend throughout 2020 and 2021, indicating drought conditions. The values presented, especially in the first half of both years, signal a situation of moderate to severe drought, requiring robust strategies for water resource management and the adoption of mitigating measures.

RECIMA21 - REVISTA CIENTÍFICA MULTIDISCIPLINAR
ISSN 2675-6218

The city of João Pessoa presents a different dynamic, with a predominantly positive SPI 12 trend. This result suggests a generally wetter condition and, consequently, less susceptibility to drought periods; on the contrary, possible attention may be focused on managing rainier periods.

Monteiro reveals a consistency in the predominantly positive SPI 12, although with smaller magnitudes than João Pessoa. The stability of the indices suggests relative water security; however, constant monitoring is crucial to ensure resource sustainability.

Similarly, Picuí shows a positive trend in SPI 12, indicating wetter conditions, especially in the second half of the years analyzed. This result indicates a stable water scenario, but resource management must be improved to avoid potential challenges associated with excessive precipitation.

Patos shows a gradual positive evolution in SPI 12 over the months, suggesting a transition to wetter conditions. However, the alternation and proximity of negative indices indicate the need for adaptive strategies to deal with variability in water availability.

The city of Souza reveals a pattern where SPI 12 increases progressively throughout the year, indicating a recovery in water conditions after a drier start to the year. This profile reinforces the importance of water management strategies resilient to seasonal variations.

Figure 10 - Predictions for the SPI-12 via ET model for 2020, and 2021. Data in blue, predictions in red.

Below, we briefly discussed the SPI-24 forecasts for the different weather stations, as seen in the graphs in Figure 11. These provide a comparative analysis and valuable insights into future weather patterns.

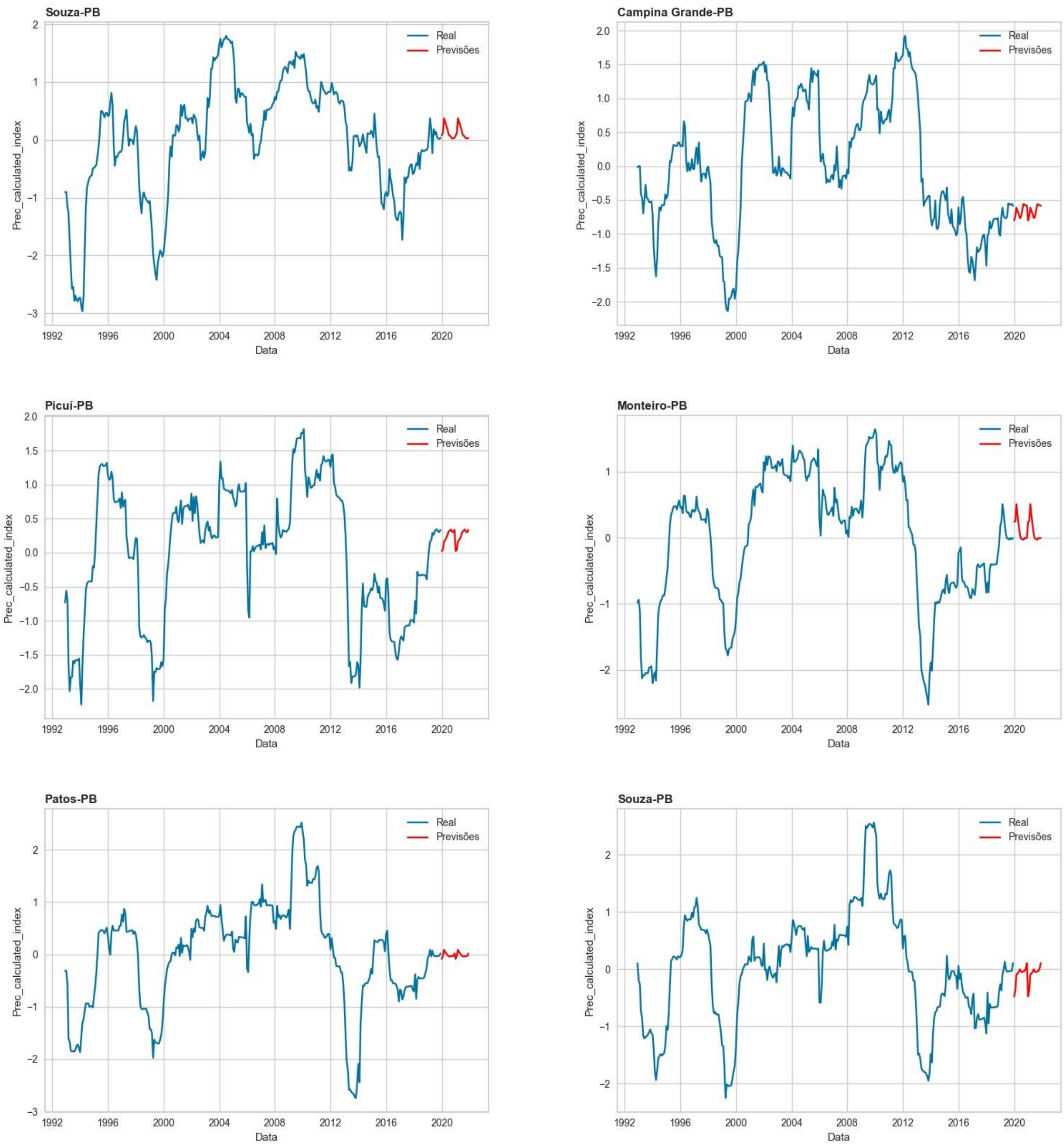
The Campina Grande station consistently presented negative SP-I24 values during 2020 and 2021, indicating periods of drought. The intensity of the drought varies; however, the persistent presence of this pattern suggests relatively dry atmospheric conditions and implies possible challenges in water-dependent activities and water resource management in the region.

The data for João Pessoa indicate a relatively stable and humid condition throughout the observed years, with SPI-24 values oscillating primarily in positive territory. Monteiro presents an interesting trajectory, starting the years with positive SPI-24 indices, indicating wetter periods, and gradually regressing to negative values as the years progress.

Picuí shows a generally positive pattern in SPI-24 throughout the two years, suggesting a trend towards wetter conditions. Patos presents a mixed situation, with SPI-24 values oscillating between negative and positive over the years. This alternation between drier and wetter periods creates a scenario requiring balanced water resource management and adapting activities dependent on climate conditions to this variability. In Souza, SPI-24 oscillates from negative values, indicating drought, to positive values, suggesting wetter periods, over two years.

RECIMA21 - REVISTA CIENTÍFICA MULTIDISCIPLINAR
ISSN 2675-6218

Figure 11 - Predictions for the SPI-12 via ET model for 2020, and 2021. Data in blue, predictions in red.



CONCLUSIONS

In summary, the comprehensive analysis of rainfall and the subsequent study of the Standardized Precipitation Index (SPI) provide essential insights into the climate patterns of a region, revealing a complex interconnection between natural and social factors. Through the data collected and the observations made, a deep understanding of climate variability and its implications emerges, transcending mere statistics that directly impact the lives of communities.

The analysis of averages, standard deviations, and extreme rainfall values in six different rainfall stations provided a diverse geographic overview of climate trends, highlighting João Pessoa with the highest average rainfall and Picuí with the lowest. The visualization of the histograms showed that, despite the prevalence of rainfall with lower values, exceptional high-intensity events are not rare, highlighting the need for resilient planning to deal with extreme variations.

By delving into the time series analysis, it was possible to identify specific years with notable rainfall in each location. The SPI demonstrated its effectiveness in detecting rainfall and drought events, providing a deeper understanding of climate dynamics at different time scales. Years marked by intense rainfall and severe drought it was varied between seasons, highlighting the complexity of regional climate and the importance of adaptive approaches to mitigate adverse impacts.

Across the different scales of the SPI, it became clear that intense climate events can broadly affect a region, regardless of the time scale analyzed. The 1990s emerged as a severe drought, highlighting the need to understand and address historical climate trends to ensure adequate preparation for future events.

Reflecting on the SPI's contribution to detecting extreme droughts, the results obtained reinforce its validity as a reliable analytical tool and predict its continued usefulness for climate risk management. Proactive adaptation and science-based policymaking can alleviate the adverse impacts of droughts and floods, preserving the resilience of communities and ecosystems.

The analysis of climate forecasts, particularly from the SPI, for several municipalities demonstrates the depth and complexity of regional climate and its implications for water management, agriculture, and urban planning. The Extra Trees Regressor (ET) model excelled in its ability to interpret this variability, although challenges persist in forecast accuracy and applicability. These results emphasize the need for a holistic approach to climate modeling, combining technical expertise, climate understanding, and sociopolitical considerations.

REFERENCES

ALFIAN, G.; SYAFRUDIN, M.; FAHRURROZI, I.; FITRIYANI, N. L.; ATMAJI, F. T. D.; WIDODO, T.; BAHIYAH, N.; BENES, F.; RHEE, J. Predicting Breast Cancer from Risk Factors Using SVM and Extra-Trees-Based Feature Selection Method. **Computers**, v. 11, p. 136, 2022. Disponível em: https://doi.org/10.3390/computers11090136. Acesso em: 28 set. 2023.

BREIMAN, L. Random Forests. **Machine Learning**, v. 45, p. 5-32, 2001.

CUTLER, A.; CUTLER, D. R.; STEVENS, J. R. Random Forests. *In:* ZHANG, C.; MA, Y. (Ed.). **Ensemble Machine Learning**: Methods and Applications. [S. l.]: Springer, 2011. cap. 5, p. 157-176. DOI: 10.1007/978-1-4419-9326-7_5.

DA SILVA, A. S. A.; MENEZES, R. S. C.; STOSIC, T. Análise multifractal do índice de precipitação padronizado. **Research, Society and Development**, v. 10, n. 7, p.e24710716535-e24710716535, 2021.

DE FARIAS, A. A.; SOUZA, J. T. A.; De Sousa, F. D. A. S. IDENTIFICAÇÃO E ANÁLISE DE SECAS SEVERAS E EXTREMAS NO MUNICÍPIO DE TAPEROÁ-PB. **Revista Brasileira de Geografia Física,** v. 07, n. 05, especial- VIWMCRHPE, p. 818-826, 2014.

DE SOUZA GUEDES, R. V.; MACEDO, M. J. H.; DE SOUZA, F. D. A. S. Análise espacial de eventos de secas com base no índice padronizado de precipitação e análise de agrupamento. **Brazilian Journal of Environmental Sciences** (Online), v. 23, p. 55-65, 2012.

DIDÁTICA TECH. **Como funciona o algoritmo ExtraTrees**. [S. l.]: Didativa Tech, s. d. Disponível em: https://didatica.tech/como-funciona-o-algoritmo-extratrees/. Acesso em: 28 set. 2023.

FILHO, M. **MAE (Erro Médio Absoluto) Em Machine Learning**. [*S. l.: s. n*.], s. d. Disponível em: https://mariofilho.com/mae-erro-medio-absoluto-em-machine-learning/. Acesso em: 28 set. 2023.

FRIEDMAN, Jerome H. "Greedy Function Approximation: A Gradient Boosting Machine." **The Annals of Statistics**, v. 29, n. 5, p. 1189–1232, 2001. http://www.jstor.org/stable/2699986.

HASAN, H.; RADI, N. A.; KASSIM, S. Modeling of extreme temperature using generalized extreme value (GEV) distribution: A case study of Penang. *In:* **Proceedings of the world congress on engineering,** v. 1, p. 181-186, july 2012.

ICMC JÚNIOR. **Random Forest**. [S. L.]: ICMC Junior, s. d. Disponível em: https://icmcjunior.com.br/random-forest/. Acesso em: Acesso em: 28 set. 2023.

MACEDO, M. J. H.; DE SOUZA GUEDES, R. V.; DE ASSIS SALVIANO, F. D. A. Monitoramento e intensidade das secas e chuvas na cidade de Campina Grande/PB. **Revista Brasileira de Climatologia**, v. 8, 2011.

MACEDO, M. J. H.; DE SOUZA GUEDES, R. V.; DE SOUZA, F. D. A. S.; DA CUNHA DANTAS, F. R. Análise do índice padronizado de precipitação para o estado da Paraíba, Brasil. **Ambiente & Água - An Interdisciplinary Journal of Applied Science**, v. 5, n. 1, p. 204-214, 2010.

MASUI, T. All You Need to Know about Gradient Boosting Algorithm − Part 1. Regression. **Towards Data Science**, *[S. l.],* 20 jan. 2022. Disponível em: https://towardsdatascience.com/all-you-need-to-know-about-gradient-boosting-algorithm-part-1-regression-2520a34a502. Acesso em: 28 set. 2023.

MATHBOUT, S.; LOPEZ-BUSTINS, J. A.; MARTIN-VIDE, J.; BECH, J.; RODRIGO, F. S. Spatial and temporal analysis of drought variability at several time scales in Syria during 1961–2012. **Atmospheric Research**, v. 200, p. 153-168, 2018.

MATPLOTLIB. **Documentação estável**. [S. l.]: MATPLOTLIB, s. d. Disponível em: https://matplotlib.org/stable/index.html. Acesso em: 20 ago. 2023.

MCKEE, T. B.; DOESKEN, N. J.; KLEIST, J. The relationship of drought frequency and duration to time scales. *In*: **Proceedings of the 8th Conference on Applied Climatology**, v. 17, n. 22, p. 179-183, 1993, jan. 1993.

OTCHERE, D. A.; GANAT, T. O. A.; OJERO, J. O.; TACKIE-OTOO, B. N.; TAKI, M. Y. Application of gradient boosting regression model for the evaluation of feature selection techniques in improving reservoir characterisation predictions. **Journal of Petroleum Science and Engineering**, v. 208, Part E, p. 109244, 2022. ISSN 0920-4105. Disponível em: https://doi.org/10.1016/j.petrol.2021.109244. Acesso em: 28 set. 2023.

PANDAS. **Documentação**. [S. l.]: Pandas, s. d. Disponível em: https://pandas.pydata.org/docs/. Acesso em: 20 ago. 2023

PARAÍBA TOTAL. **Geografia**. [S. l.]: Paraíba Total, 2021. Disponível em: https://www.paraibatotal.com.br/a-paraiba/geografia/. Acesso em: 20 ago. 2023.

PYTHON SOFTWARE FOUNDATION. **Documentação Python 3**.x (Português do Brasil). [S. l.]: Python Software Foundation, s. d. Disponível em: https://docs.python.org/pt-br/3/. Acesso em: 20 ago. 2023.

R DEVELOPMENT CORE TEAM. **R:** a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing, 2013. ISBN 3-900051-07-0, 2014.

SANTOS, C. A. G.; BRASIL NETO, R. M.; DA SILVA, R. M.; DOS SANTOS, D. C. Innovative approach for geospatial drought severity classification: a case study of Paraíba state, Brazil. **Stochastic Environmental Research and Risk Assessment**, v. 33, n. 2, p. 545-562, 2019.

SEABORN. **Introdução ao tutorial**. [S. l.]: SEABORN, s. d. Disponível em: https://seaborn.pydata.org/tutorial/introduction.html. Acesso em: 20 ago. 2023.

SODRÉ, G. D. A. **2040:** o aquecimento global gerando impacto econômico na produção de cana no Estado de São Paulo. 2010. Tese (Doutorado) – Fundação Getulio Vargas, São Paulo, 2010.

SVOBODA, M.; HAYES, M.; WOOD, D. **Standardized precipitation index: user guide**. [*S. l.: s. n.*], 2012.

THOM, H. C. A note on the gamma distribution. **Monthly weather review**, v. 86, n. 4, p. 117-122, 1958.

WANG, K. Y.; LI, Q. F.; YANG, Y.; ZENG, M.; Li, P. C.; ZHANG, J. X. Analysis of spatio-temporal evolution of droughts in Luanhe River Basin using different drought indices. **Water Science and Engineering,** v. 8, n. 4, p. 282-290, 2015.

XARRAY. **Documentação estável**. [S. l.]: Xarray, s. d. Disponível em: https://docs.xarray.dev/en/stable/. Acesso em: 20 ago. 2023.