Comparative evaluation of daily rainfall forecasting models using INMET data

Leonardo Costa de Souza[1], **Iberê Luiz Caldas**[1], **Diego Minatel**[2]

[1]Institute of Physics, University of São Paulo,
São Paulo 05315-970, Brazil

[2]Institute of Mathematics and Computer Science, University of São Paulo,
São Carlos 13566-590, Brazil

{leocsouza, dminatel}@usp.br, ibere@if.usp.br

Abstract. *Accurate rainfall forecasting remains a complex challenge, yet it is vital for agriculture, water resource management, and urban planning. Given Brazil's climatic diversity, understanding how different forecasting models perform across regions is particularly important. Using historical data from 46 selected INMET automatic stations distributed across all five Brazilian geographic regions and spanning 2015 to 2026, we evaluated six machine learning models– Random Forest, XGBoost, LightGBM, MLP, LSTM, and Reservoir Computing– for daily precipitation forecasting, using ARIMAX as a baseline. Our results suggest that the machine learning approaches consistently outperform the statistical baseline across all Brazilian regions, with neural-network-based models achieving the highest overall performance. These findings highlight the potential of nonlinear learning methods for improving daily precipitation forecasting across diverse climatic conditions in Brazil.*

1. Introduction

Precipitation is one of the main components of the Earth's hydrological cycle, directly influencing agriculture, energy production, ecosystems, and the survival of living organisms [Atkinson 2019, Liu 2015]. In the agricultural sector, for example, both the amount and distribution of rainfall strongly affect crop productivity and determine which crops can be successfully cultivated in a given region [Rodrigues et al. 2013]. Precipitation also has a significant impact on urban areas. Excessive rainfall, combined with inadequate urban planning and drainage infrastructure, can trigger floods, damage critical infrastructure, disrupt power supply systems, and cause substantial socioeconomic losses [Atkinson 2019]. Consequently, accurate precipitation forecasting plays a crucial role in supporting agricultural planning, mitigating the impacts of extreme weather events, improving water resource management, and assisting decision-making processes related to environmental conservation and disaster risk reduction.

Despite its importance, precipitation forecasting remains a challenging task due to the highly nonlinear nature of rainfall processes and the large number of interacting atmospheric variables involved [Barrera-Animas et al. 2022, New et al. 2001]. Traditionally, forecasting approaches relied on physically based models that represent atmospheric processes through mathematical formulations, commonly referred to as climate models [Wallace and Hobbs 2006, Burlando et al. 1996, Chow 1971]. Advances in computational resources and the growing availability of meteorological data have enabled the

development of data-driven approaches capable of discovering complex patterns from historical observations [da Silva et al. 2025, Das et al. 2024, Solomatine and Ostfeld 2008]. Such approaches provide a promising alternative for precipitation forecasting in Brazil, where the country's climatic diversity and strong spatial variability in rainfall patterns increase the complexity of accurate prediction.

Several data-driven modeling paradigms have been investigated for precipitation forecasting. Classical statistical models, such as Seasonal Autoregressive Integrated Moving Average (SARIMA) and its exogenous extension (SARIMAX), remain widely used to capture seasonal patterns in meteorological time series [Kumar and Richariya 2024]. Ensemble tree methods have also gained popularity due to their ability to model complex nonlinear relationships with relatively limited parameter tuning. In particular, Random Forest, Extreme Gradient Boosting (XGBoost), and Light Gradient Boosting Machine (LightGBM) have demonstrated competitive performance in precipitation forecasting tasks [Cutler et al. 2012]. Neural network architectures constitute another important class of forecasting models. Recurrent Neural Networks (RNNs), including Long Short-Term Memory (LSTM) and Bidirectional Long Short-Term Memory (BiLSTM) networks, can capture temporal dependencies in sequential weather data. For example, [Barrera-Animas et al. 2022] compared LSTM, BiLSTM, and XGBoost models using hourly rainfall observations and reported superior performance for deep recurrent architectures. Another RNN-based paradigm is Reservoir Computing (RC), commonly implemented through Echo State Networks (ESNs) [Lukoševičius et al. 2012, Jaeger 2001]. RCs have shown promising results in modeling complex and chaotic dynamical systems [Pathak et al. 2017]. However, to the best of our knowledge, their application to precipitation forecasting in Brazil has not yet been investigated.

Given this context, this paper presents a comparative evaluation of precipitation forecasting models across Brazil. To this end, we use daily rainfall records collected from 46 automatic weather stations operated by the National Institute of Meteorology (INMET), covering all five Brazilian geographic regions. Specifically, we evaluate seven forecasting models: (i) ARIMAX, (ii) Random Forest, (iii) XGBoost, (iv) LightGBM, (v) Multilayer Perceptron (MLP), (vi) LSTM and (vii) RC. By assessing these approaches under a common experimental protocol and across diverse climatic conditions, this study provides a broad comparison of their predictive performance in the Brazilian context. Furthermore, the experimental results highlight the potential of Reservoir Computing as a competitive approach for precipitation forecasting, demonstrating its ability to capture complex temporal patterns in rainfall data effectively.

2. Related Work

Precipitation forecasting has traditionally relied on numerical weather prediction and physical climate models that simulate atmospheric dynamics and thermodynamic processes through mathematical formulations [Atkinson 2019, Chow 1971]. More recently, data-driven approaches have gained increasing attention by leveraging historical meteorological records to identify complex patterns associated with rainfall dynamics [Barrera-Animas et al. 2022]. Among these approaches, statistical models such as ARIMA and ARIMAX remain widely used due to their interpretability and effectiveness in modeling temporal dependencies. However, they often struggle to capture the nonlinear dynamics of rainfall, particularly in tropical climates.

These limitations have motivated the adoption of machine learning techniques for precipitation forecasting. In Brazil, supervised learning approaches have been successfully applied to rainfall prediction in the Amazon region, outperforming traditional statistical baselines [Monteiro et al. 2024]. Unsupervised learning methods have also been explored, such as cluster analysis to identify rainfall patterns and estimate the number of rainy days in semiarid regions of the country [Tinôco et al. 2018].

Among supervised learning approaches, MLPs have been applied to precipitation forecasting for several decades with encouraging results [Gardner and Dorling 1998]. More recently, [Barrera-Animas et al. 2022] compared LSTM, BiLSTM, and XGBoost models using hourly rainfall observations in the United Kingdom and found that recurrent neural networks achieved superior predictive performance. In contrast, [da Silva et al. 2025] reported lower forecasting errors for Random Forest than for LSTM-based models when forecasting precipitation from satellite-derived data in Brazil. Similarly, [Domingos et al. 2025] evaluated gradient boosting and deep learning models for seasonal precipitation forecasting in South America and concluded that boosting techniques can achieve competitive predictive accuracy. These findings suggest that the effectiveness of forecasting models depends strongly on data characteristics and the prediction scenario, highlighting the importance of comparative evaluations across different climatic and geographic settings.

Another promising approach is RC, commonly implemented through ESNs. RC has demonstrated strong predictive capabilities in nonlinear dynamical systems and time-series forecasting tasks [Pei and Wang 2023]. Despite these encouraging results, its application to precipitation forecasting in Brazil remains largely unexplored. Motivated by this gap, the present study evaluates several forecasting approaches using rainfall data from different Brazilian regions, with particular emphasis on Reservoir Computing.

3. Methodology

This section describes the methodology adopted to conduct the comparative evaluation of precipitation forecasting models across Brazil. The methodology comprises four main components: dataset, data processing and feature engineering, forecasting models, and evaluation metrics. The following subsections present each component in detail.

3.1. Dataset

We constructed the dataset used in this study from publicly available data collected by INMET's automatic meteorological stations[1]. The observations cover the period from January 1, 2015, to April 30, 2026, with a daily temporal resolution. Initially, we collected data from 659 meteorological stations with records available during the study period. To ensure data quality and reduce the impact of missing values, we retained only stations with less than 30% missing observations. After applying this filtering criterion, the dataset comprised 362 stations distributed across different regions of Brazil. Fig. 1 shows the spatial distribution of these stations and their respective percentages of missing data.

As shown in Fig. 1, the spatial distribution of meteorological stations in Brazil is highly heterogeneous, with station density decreasing substantially toward the North

[1] Available at: https://bdmep.inmet.gov.br

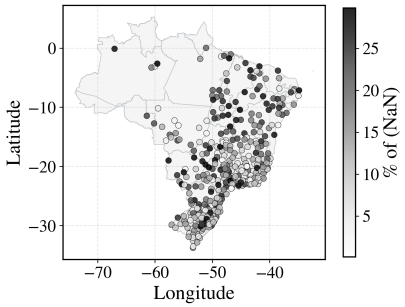


Figure 1. Spatial distribution of the automatic INMET stations retained after applying the 30% missing-data threshold. The color scale indicates the percentage of missing precipitation observations at each station.

and Central-West regions, as previously reported by [Rozante et al. 2010]. This uneven coverage poses challenges for precipitation studies, particularly in areas with sparse observational networks. To mitigate this limitation, researchers often complement ground-based observations with satellite-derived and gauge-corrected precipitation products [Rozante et al. 2020]. However, we rely exclusively on observations collected by INMET stations. Despite its heterogeneous spatial coverage, the INMET meteorological network provides reliable ground-based measurements and constitutes a valuable data source for investigating precipitation dynamics and evaluating forecasting models.

After applying the missing-data filter, we retained two stations from each state, prioritizing those with the highest data completeness (i.e., the lowest proportion of missing data). We excluded Acre, Rio Grande do Norte, and Roraima because none of their stations satisfied the 30% missing-data threshold. We also excluded Amapá because only one station met this criterion. As a result, the final dataset comprises 46 stations, corresponding to the two stations with the most complete records from each of the following 22 states and the Federal District: Alagoas, Amazonas, Bahia, Ceará, Distrito Federal, Espírito Santo, Goiás, Maranhão, Mato Grosso, Mato Grosso do Sul, Minas Gerais, Pará, Paraíba, Paraná, Pernambuco, Piauí, Rio de Janeiro, Rio Grande do Sul, Rondônia, Santa Catarina, São Paulo, Sergipe, and Tocantins.

3.2. Data Processing and Feature Engineering

Each station provides meteorological observations together with geographic metadata, including the station identification code, station name, state, geographic region, latitude, longitude, and altitude. These metadata allow us to uniquely identify each station and stratify observations by state and geographic region. The meteorological observations recorded at each station are used as input variables for the forecasting models. Table 1 summarizes the observations considered in this study.

We first organized the observations chronologically for each station and then handled missing values. For the target variable (Prec), we set missing precipitation values to zero, reflecting the predominance of dry days in most regions and avoiding the introduction of artificial rainfall events.

For the remaining meteorological variables (P, T_{max}, T_{min}, RH_{mean}, RH_{min}, and WS_{mean}), we adopted a hierarchical imputation strategy designed to preserve temporal

Table 1. Description of the meteorological variables

Variable	Description
Date	Observation date in the format YYYY-MM-DD
P	Daily mean atmospheric pressure (mB)
T_{max}	Daily maximum air temperature (°C)
T_{min}	Daily minimum air temperature (°C)
RH_{mean}	Daily mean relative humidity (%)
RH_{min}	Daily minimum relative humidity (%)
WS_{mean}	Daily mean wind speed (m/s)
Prec	Daily accumulated precipitation (mm), used as the target variable in the forecasting task

continuity and seasonal patterns [Wangwongchai et al. 2023, Alejo-Sanchez et al. 2025]. First, we filled gaps of up to three consecutive days through linear interpolation applied independently to each station. We imputed longer gaps using station-specific climatological information by replacing each missing value with the historical mean of the corresponding variable for the same month and day at the same station, computed from all available years in the training data. When no such historical record was available, we used the long-term mean of the variable at that station. As a final fallback, we replaced missing values with the overall mean of the corresponding variable computed from the data.

In addition to the meteorological variables, we derived temporal features from the observation date to capture seasonal and cyclical patterns in precipitation. Specifically, we encoded the month and day of the year using sine and cosine transformations, resulting in four additional variables: $Month_{sin}$, $Month_{cos}$, Day_{sin}, and Day_{cos}. This representation preserves the cyclical nature of temporal variables by ensuring that adjacent periods, such as December and January, remain close in the feature space. For a temporal variable t with period T, the cyclical encoding is defined as

$$\sin(t) = \sin\left(2\pi\frac{t}{T}\right), \qquad \cos(t) = \cos\left(2\pi\frac{t}{T}\right). \tag{1}$$

To further capture temporal dependencies in rainfall dynamics, we incorporated lagged precipitation features. Specifically, we included precipitation values observed 1, 2, 3, 4, 5, 6, 7, 14, 21, and 30 days before the prediction date as predictor variables.

The final feature set comprises 20 variables: P, T_{max}, T_{min}, RH_{mean}, RH_{min}, WS_{mean}, $Month_{sin}$, $Month_{cos}$, Day_{sin}, Day_{cos}, and the ten lagged precipitation features $Prec_{t\text{-}1}$, $Prec_{t\text{-}2}$, $Prec_{t\text{-}3}$, $Prec_{t\text{-}4}$, $Prec_{t\text{-}5}$, $Prec_{t\text{-}6}$, $Prec_{t\text{-}7}$, $Prec_{t\text{-}14}$, $Prec_{t\text{-}21}$, and $Prec_{t\text{-}30}$. These variables serve as inputs to the forecasting models, while daily accumulated precipitation (Prec) is used as the target variable.

3.3. Forecasting Models

We selected seven regression algorithms to perform daily precipitation forecasting. As a statistical baseline, we employed (i) SARIMAX. We also evaluated three ensemble tree methods, namely (ii) Random Forest, (iii) XGBoost, and (iv) LightGBM, as well as three neural-network-based models: (v) MLP, (vi) LSTM and (vi) RC.

Because SARIMAX, Random Forest, XGBoost, LightGBM, MLP, and LSTM are well-established forecasting models in the literature, we provide a brief overview of RC, whose application to precipitation forecasting remains relatively limited. RC is

a recurrent neural network architecture composed of a fixed, randomly connected dynamical reservoir and a trainable linear readout layer [Jaeger 2001]. Unlike conventional recurrent neural networks, RC trains only the output weights, resulting in a simpler training procedure. The reservoir dynamics provide an internal memory mechanism capable of capturing temporal dependencies and nonlinear interactions in time-series data [Lukoševičius et al. 2012]. We included RC in this study because of its strong performance in forecasting chaotic dynamical systems [Griffith et al. 2019] and its recent application to similar meteorological data collected from weather stations in China [Pei and Wang 2023].

3.4. Evaluation Metrics

To evaluate the models described in Section 3.3, we employed two regression metrics: Mean Absolute Error (MAE) and the Root Mean Squared Error (RMSE). Let y_i denote the observed precipitation, $\hat{y}_i$ the predicted precipitation for observation i, and n the total number of observations. The mathematical formulations of these metrics is given by,

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|, \qquad \text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}. \tag{2}$$

In this study, MAE measures the average magnitude of prediction errors in millimetres of precipitation, assigning equal importance to all deviations. RMSE also quantifies prediction errors in millimetres but places greater emphasis on larger errors due to the squared-error term.

4. Experimental Setup

Fig. 2 provides an overview of the experimental setup adopted in this study. First, we applied an additional preprocessing stage to the final dataset described in Section 3. We standardized all input variables and applied a logarithmic transformation to the target variable (Prec) to reduce skewness and stabilize variance.

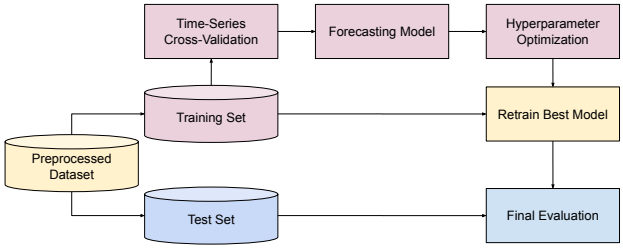


Figure 2. Overview of the adopted experimental setup

Next, we chronologically split the dataset into training and test sets to preserve the temporal structure of the precipitation series. Observations collected between 2015-01-01 and 2024-05-31 composed the training set, whereas observations collected between 2024-06-01 and 2026-04-30 composed the test set.

We optimized the hyperparameters of the forecasting models (Section 3.3) using time-series cross-validation on the training set. Table 2 summarizes the search spaces considered for each model. We adopted a five-fold expanding-window validation scheme. We employed Randomized Search with 100 candidate configurations sampled from predefined search spaces for Random Forest, XGBoost, and LightGBM, whereas MLP, LSTM, and RC used exhaustive grid search. In the MLP experiments, we applied early stopping based on the validation loss to reduce overfitting. For RC, we used a warm-up period before training the readout layer to allow the reservoir dynamics to reach a stable state.

Table 2. Hyperparameter search spaces used for model optimization.

Model	Hyperparameter search space
Random Forest	$\texttt{n_estimators} \in [200, 1000]$, $\texttt{max_depth} \in [3, 100]$, $\texttt{min_samples_split} \in [2, 40]$ $\texttt{min_samples_leaf} \in [1, 20]$, $\texttt{max_features} \in \{\texttt{sqrt}, \texttt{log2}, \texttt{None}\}$
XGBoost	$\texttt{n_estimators} \in [200, 1000]$, $\texttt{learning_rate} \in [0.01, 0.30]$, $\texttt{max_depth} \in [3, 12]$ $\texttt{subsample} \in [0.6, 1.0]$, $\texttt{colsample_bytree} \in [0.6, 1.0]$, $\texttt{min_child_weight} \in [1, 10]$
LightGBM	$\texttt{n_estimators} \in [200, 1000]$, $\texttt{subsample} \in [0.6, 1.0]$, $\texttt{max_depth} \in [3, 15]$, $\texttt{num_leaves} \in [10, 100]$ $\texttt{colsample_bytree} \in [0.6, 1.0]$, $\texttt{learning_rate} \in [0.01, 0.30]$, $\texttt{min_child_samples} \in [10, 50]$
MLP	$\texttt{units} \in \{32, 64, 128, 256\}$, $\texttt{dropout} \in \{0.0, 0.1, 0.2\}$ $\texttt{hidden_layers} \in \{1, 2, 3\}$, $\texttt{batch_size} \in \{32, 64, 128\}$
LSTM	$\texttt{sequences} \in \{7, 14, 30\}$, $\texttt{layers} \in \{1, 2, 3\}$, $\texttt{units} \in \{32, 64, 128\}$ $\texttt{learning_rate} \in \{10^{-2}, 10^{-3}, 10^{-4}, 10^{-4}\}$, $\texttt{dropout} \in \{0.0, 0.2\}$
RC	$\texttt{units} \in \{100, 300, 500, 1000\}$, $\texttt{spectral_radius} \in \{0.5, 0.7, 0.9, 1.1, 1.2\}$ $\texttt{leaking_rate} \in \{0.1, 0.3, 0.5\}$, $\texttt{ridge} \in \{10^{-6}, 10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}\}$

SARIMAX followed a different optimization procedure. We automatically selected the autoregressive (p), differencing (d), and moving-average (q) orders using the Auto-ARIMA algorithm on the training set. We turned off seasonal components because the feature set already explicitly represented seasonal patterns through sine and cosine temporal encodings, effectively resulting in an ARIMAX model. Furthermore, we excluded precipitation lag features from SARIMAX to avoid redundancy with its intrinsic autoregressive structure.

Finally, we selected the best configuration for each model based on the lowest validation RMSE. We then retrained the selected setting using the entire training set and evaluated its predictive performance on the test set using MAE and RMSE, as defined in Section 3.4. Before computing these metrics, we transformed all predictions back to the original precipitation scale.

We implemented all experiments in Python using NumPy, Pandas, Scikit-learn, ReservoirPy, Keras, LightGBM, and XGBoost, with Seaborn and Matplotlib for data visualization. The dataset, source code, and analysis scripts are publicly available at: `https://github.com/leonardo-cSouza/inmet_rainfall_forecasting`.

5. Results and Discussion

Table 3 presents the average RMSE values and corresponding standard deviations obtained for the considered Brazilian federative units. In general, neural-network-based models achieved the best performance. In particular, RC had the lowest average RMSE among all evaluated methods and achieved the best result in 11 of the 23 federative units, ranking second in another 11. MLP and LSTM also performed competitively, each

achieving the best result in at least four federative units. Among the ensemble tree methods, XGBoost yielded the lowest average RMSE, although its advantage over Random Forest and LightGBM was relatively small. In contrast, SARIMAX consistently produced the highest errors and achieved the worst overall performance.

Table 3. Mean RMSE values by Brazilian federative unit (FU), with standard deviations reported in parentheses. Bold and underlined values indicate the best and second-best results, respectively.

FU	SARIMAX	Random Forest	XGBoost	LightGBM	MLP	LSTM	RC
Alagoas	8.41 (0.82)	8.35 (0.67)	8.10 (0.54)	8.05 (0.51)	7.76 (0.58)	8.05 (1.38)	**6.47 (1.32)**
Amazonas	10.47 (0.48)	10.41 (0.83)	**10.01 (1.06)**	10.20 (1.18)	**10.01 (0.77)**	10.73 (1.35)	10.37 (0.85)
Bahia	8.50 (0.88)	7.72 (0.49)	7.62 (1.07)	7.72 (0.87)	7.24 (0.66)	**7.21 (1.39)**	7.39 (0.60)
Ceará	2.70 (2.25)	2.42 (2.22)	2.32 (2.08)	2.39 (2.12)	2.34 (1.86)	2.59 (2.98)	**2.04 (1.72)**
Distrito Federal	10.03 (0.93)	9.24 (0.89)	9.36 (0.97)	9.27 (0.91)	**9.04 (0.73)**	9.79 (2.18)	9.23 (0.90)
Espírito Santo	10.13 (0.19)	9.55 (0.07)	9.29 (0.03)	9.20 (0.22)	**8.94 (0.35)**	9.66 (0.81)	8.97 (0.37)
Goiás	9.85 (1.57)	8.82 (1.17)	8.75 (1.15)	8.82 (1.21)	**8.59 (1.07)**	9.41 (1.09	8.64 (1.27)
Maranhão	7.14 (4.60)	6.94 (4.50)	6.79 (3.96)	6.67 (4.06)	6.22 (4.11)	**5.75 (5.11)**	6.31 (4.54)
Minas Gerais	10.73 (0.98)	9.88 (1.11)	9.98 (0.82)	9.98 (0.90)	**9.21 (0.34)**	9.64 (0.34)	9.55 (0.37)
Mato Grosso	7.46 (4.76)	6.92 (4.35)	6.87 (4.31)	6.91 (4.34)	6.88 (4.33)	6.54 (5.68)	**6.61 (4.09)**
Mato Grosso do Sul	8.18 (1.02)	7.57 (0.83)	7.52 (0.83)	7.55 (0.87)	**7.27 (0.77)**	7.82 (2.28)	7.50 (0.77)
Pará	12.71 (0.52)	11.86 (0.01)	11.85 (0.02)	12.00 (0.05)	11.43 (0.08)	11.68 (0.22)	**10.94 (0.07)**
Paraíba	5.00 (1.23)	4.87 (1.16)	4.66 (1.13)	4.75 (1.11)	4.72 (1.17)	5.12 (1.97)	**4.03 (0.59)**
Pernambuco	2.23 (1.41)	2.15 (1.36)	2.72 (1.93)	2.35 (1.54)	**1.91 (1.33)**	2.04 (2.14)	1.93 (1.09)
Piauí	3.66 (1.74)	3.41 (1.80)	3.40 (1.77)	3.38 (1.76)	3.05 (1.42)	2.87 (1.95)	**3.01 (1.42)**
Paraná	9.99 (2.89)	9.72 (3.03)	9.61 (2.97)	9.67 (3.02)	9.76 (3.05)	10.59 (5.18)	**9.37 (2.59)**
Rio de Janeiro	8.13 (1.62)	7.85 (1.53)	7.77 (1.50)	7.82 (1.54)	7.66 (1.44)	8.47 (1.93)	**7.25 (1.59)**
Rondônia	9.90 (0.63)	8.93 (0.52)	8.75 (0.38)	8.83 (0.47)	8.53 (0.53)	**8.11 (0.88)**	8.71 (0.69)
Rio Grande do Sul	10.94 (1.24)	10.85 (1.16)	10.55 (1.16)	10.63 (1.14)	10.25 (1.06)	12.15 (1.77)	**9.88 (0.62)**
Santa Catarina	7.98 (4.23)	7.10 (4.11)	7.15 (4.10)	7.20 (4.26)	6.88 (3.49)	**6.72 (4.58)**	6.89 (3.66)
Sergipe	8.18 (0.98)	8.06 (1.10)	7.93 (1.06)	7.97 (1.14)	**7.64 (0.98)**	8.63 (1.00)	7.84 (1.19)
São Paulo	9.38 (0.19)	8.76 (0.14)	8.69 (0.03)	8.72 (0.05)	8.50 (0.02)	8.50 (0.98)	**8.31 (0.16)**
Tocantins	8.10 (0.79)	7.62 (0.99)	7.60 (1.04)	7.55 (1.03)	7.63 (0.87)	8.03 (0.51)	**7.44 (1.10)**
Average	8.25 (1.56)	7.78 (1.48)	7.71 (1.47)	7.72 (1.49)	7.45 (1.35)	7.83 (2.07)	**7.33 (1.37)**

From a geographical perspective, the lowest forecast errors were observed in the Northeast region, particularly in Pernambuco, where MLP achieved the lowest average RMSE (1.91 ± 1.33), and in Ceará, where RC achieved an RMSE of 2.04 ± 1.72. In contrast, the highest prediction errors were recorded in Pará. Interestingly, although Pará exhibited relatively large RMSE values, it also presented one of the lowest standard deviations among all federative units. This result suggests a consistent error pattern across the two analyzed stations. Conversely, larger standard deviations indicate greater spatial heterogeneity in rainfall dynamics within a state, suggesting that precipitation patterns differ substantially between the selected INMET stations.

Figures 3 and 4 present the distributions of MAE and RMSE, respectively, across stations grouped by Brazilian region. Both metrics indicate that the Northeast region generally exhibits the lowest median errors. However, its distributions show considerable variability, particularly for RMSE. In contrast, the North and South regions represent the most challenging forecasting environments, yielding the highest median errors (RMSE $\approx$ 9–11; MAE $\approx$ 3.5–5) across the evaluated methods. The Southeast region stands out for producing the most consistent predictions, as evidenced by the narrowest interquartile ranges for both metrics, despite a few lower-bound outliers.

The superior performance observed in the Northeast region can be attributed to its prolonged, well-defined dry seasons, which result in many days without rainfall. Conse-

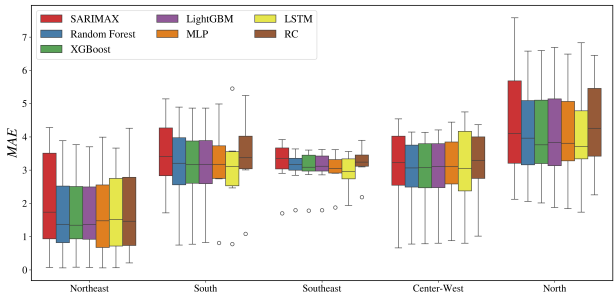


Figure 3. Distribution of MAE across stations grouped by Brazilian region.

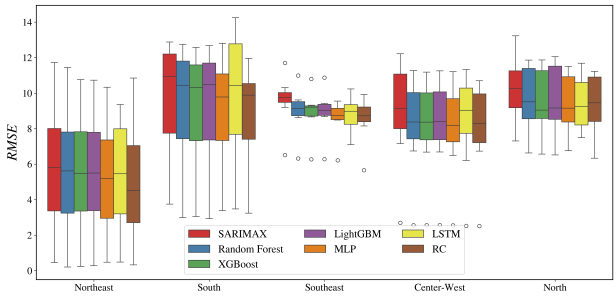


Figure 4. Distribution of RMSE across stations grouped by Brazilian region.

quently, the models can more easily learn this dominant pattern, reducing overall prediction errors and stabilizing the evaluation metrics. In contrast, the North region presented the highest errors, likely due to the predominance of convective rainfall in the Amazon, characterized by intense, sudden, high-volume, and highly localized precipitation events. The South region also exhibited relatively high errors, particularly during extreme rainfall events. Unlike the Northeast, rainfall in the South is distributed throughout the year and is strongly influenced by frequent cold-front passages. The evaluated models do not fully capture this greater day-to-day variability and the frequent shifts in weather patterns. Finally, the Central-West and Southeast regions showed similar intermediate performance. Both regions share a comparable climatic regime, characterized by a well-defined rainy season during summer and a relatively distinct dry winter.

Placing our findings in a broader context, the RMSE values achieved by the best-performing models compare favorably with those reported in recent rainfall forecasting studies. For example, [Wani et al. 2024] reported a minimum RMSE of 14 mm using LSTM for stations in the North-western Himalayas, a value that exceeds even the highest RMSE observed in our experiments (12.71 mm for ARIMAX in Pará). This difference may reflect both the distinct climatic characteristics of the Himalayan region and differences in the observational datasets. For Reservoir Computing, [Pei and Wang 2023] reported mean absolute errors ranging from 1.8 to 4.3 mm across weather stations in China using ESNs; our RC results achieved comparable performance and, in some cases, lower errors in the Northeast region. In contrast to [Barrera-Animas et al. 2022], who found that LSTMs outperformed XGBoost for hourly rainfall forecasting in the United Kingdom, our daily-resolution experiments indicate similar performance for both models, suggesting that temporal resolution may influence their relative advantages. Finally,

[da Silva et al. 2025] reported lower forecasting errors for Random Forest than for LSTM when using satellite-derived precipitation data in Brazil. Our results only partially support this finding, as Random Forest generally underperformed neural-network-based models when trained on ground-station observations.

Overall, our results highlight the inherent difficulty of daily precipitation forecasting. Precipitation is highly intermittent and nonlinear, influenced by atmospheric processes operating across multiple spatial and temporal scales. As a result, part of its variability cannot be fully explained by the local meteorological variables considered in this study. In addition, the dataset exhibits a highly imbalanced distribution, with a predominance of dry days and relatively few intense rainfall events. Model performance may also have been affected by the adopted imputation strategy. Incorporating more sophisticated imputation procedures and complementary data sources, such as National Water Agency records and satellite-based observations, could potentially improve forecasting accuracy.

6. Concluding Remarks

This paper presented a comparative evaluation of daily rainfall forecasting models across Brazil using meteorological data collected from INMET stations. The study considered 46 stations distributed across 22 Brazilian states and the Federal District, selecting two stations per state according to data-quality criteria. To ensure data reliability, we retained only stations with less than 30% missing observations. Using this dataset, we compared a statistical baseline model (SARIMAX) alongside six data-driven forecasting models: Random Forest, XGBoost, LightGBM, MLP, LSTM, and Reservoir Computing.

Our results indicate that nonlinear data-driven models substantially and consistently outperform the linear ARIMAX baseline on the INMET dataset. Among the evaluated approaches, RC and MLP achieved the highest predictive accuracy, with RC obtaining the lowest errors in the Northeast region and the best average performance across states. Despite its theoretical suitability for sequential data, the LSTM was the weakest neural network-based model in several states characterized by predominantly convective or frontal precipitation regimes. Across all regions, forecasting remained most challenging in the North and South of Brazil, where convective rainfall and cold-front-driven precipitation, respectively, generate high day-to-day variability that is not fully captured by the local meteorological predictors considered in this study.

For future work, we highlight the development of multi-stage hybrid frameworks to address the zero-inflation in precipitation data. A promising direction is adopting a two-stage architecture, where a binary classifier first predicts rainfall occurrence and a regression model subsequently estimates precipitation intensity. In addition, future studies could expand the feature space by incorporating variables that describe non-local atmospheric conditions, such as cloud-cover information, indicators of vertical convective development, and geopotential height. These variables may provide additional information on the atmospheric processes driving precipitation and improve forecast accuracy.

Acknowledgments

This study was financed, in part, by the São Paulo Research Foundation (FAPESP), Brasil. Process Numbers #2023/16146-6 and #2024/20417-8.

References

Alejo-Sanchez, L. E., Márquez-Grajales, A., Salas-Martínez, F., Franco-Arcega, A., López-Morales, V., Acevedo-Sandoval, O. A., González-Ramírez, C. A., and Villegas-Vega, R. (2025). Missing data imputation of climate time series: A review. *MethodsX*, 15:103455.

Atkinson, B. (2019). Precipitation. In *Man and Environmental Processes*, pages 23–37. Routledge.

Barrera-Animas, A. Y., Oyedele, L. O., Bilal, M., Akinosho, T. D., Delgado, J. M. D., and Akanbi, L. A. (2022). Rainfall prediction: A comparative analysis of modern machine learning algorithms for time-series forecasting. *Machine Learning with Applications*, 7:100204.

Burlando, P., Montanari, A., and Ranzi, R. (1996). Forecasting of storm rainfall by combined use of radar, rain gages and linear models. *Atmospheric research*, 42(1-4):199–216.

Chow, V. T. (1971). Handbook of applied hydrology.

Cutler, A., Cutler, D. R., and Stevens, J. R. (2012). Random forests. In *Ensemble machine learning*, pages 157–175. Springer.

da Silva, S. T., Milani, L. C., Gabrick, E. C., Iarosz, K. C., Viana, R. L., Caldas, I. L., and Batista, A. M. (2025). Rainfall forecast in brazil using machine learning. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 35(7).

Das, P. K., Sahu, R. L., and Swain, P. C. (2024). Comparative analysis of machine learning models for rainfall prediction. *Journal of Atmospheric and Solar-Terrestrial Physics*, 264:106340.

Domingos, M. C., Santiago Júnior, V. A., Anochi, J. A., Shiguemori, E. H., Costa dos Santos, L. M., Pereira, H. C. S., and Oliveira, A. E. C. (2025). Exploring machine learning, deep learning, and explainable ai methods for seasonal precipitation prediction in south america. *arXiv preprint arXiv:2512.13910*.

Gardner, M. W. and Dorling, S. R. (1998). Artificial neural networks (the multilayer perceptron)—a review of applications in the atmospheric sciences. *Atmospheric Environment*, 32(14-15):2627–2636.

Griffith, A., Pomerance, A., and Gauthier, D. J. (2019). Forecasting chaotic systems with very low connectivity reservoir computers. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 29(12).

Jaeger, H. (2001). The "echo state" approach to analysing and training recurrent neural networks-with an erratum note. *Bonn, Germany: German national research center for information technology gmd technical report*, 148(34):13.

Kumar, D. and Richariya, P. (2024). Leveraging sarimax for accurate rainfall forecasting in india. *International Journal of Innovations in Science, Engineering And Management*, pages 37–47.

Liu, Z. (2015). Comparison of versions 6 and 7 3-hourly trmm multi-satellite precipitation analysis (tmpa) research products. *Atmospheric Research*, 163:91–101.

Lukoševičius, M., Jaeger, H., and Schrauwen, B. (2012). Reservoir computing trends. *KI-Künstliche Intelligenz*, 26(4):365–371.

Monteiro, L. A. F., do Nascimento, F. I. C., de Oliveira-Júnior, J. F., Nunes, D. D., Mendes, D., de Gois, G., Sanches, F. d. O., Wollmann, C. A., Watanabe, M., and Gobo, J. P. A. (2024). Rainfall projections for the brazilian legal amazon: An artificial neural networks first approach. *Climate*, 12(11).

New, M., Todd, M., Hulme, M., and Jones, P. (2001). Precipitation measurements and trends in the twentieth century. *International Journal of Climatology: A Journal of the Royal Meteorological Society*, 21(15):1889–1922.

Pathak, J., Lu, Z., Hunt, B. R., Girvan, M., and Ott, E. (2017). Using machine learning to replicate chaotic attractors and calculate lyapunov exponents from data. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 27(12).

Pei, L. and Wang, K. (2023). Precipitation forecast in china based on reservoir computing. *The European Physical Journal Special Topics*, 232(5):695–702.

Rodrigues, G. C., Martins, J. D., Da Silva, F. G., Carlesso, R., and Pereira, L. S. (2013). Modelling economic impacts of deficit irrigated maize in brazil with consideration of different rainfall regimes. *biosystems engineering*, 116(1):97–110.

Rozante, J. R., Gutierrez, E. R., Fernandes, A. d. A., and Vila, D. A. (2020). Performance of precipitation products obtained from combinations of satellite and surface observations. *International Journal of Remote Sensing*, 41(19):7585–7604.

Rozante, J. R., Moreira, D. S., de Goncalves, L. G. G., and Vila, D. A. (2010). Combining trmm and surface observations of precipitation: technique and validation over south america. *Weather and forecasting*, 25(3):885–894.

Solomatine, D. P. and Ostfeld, A. (2008). Data-driven modelling: some past experiences and new approaches. *Journal of Hydroinformatics*, 10(1):3–22.

Tinôco, I. C. M., Bezerra, B. G., Lucio, P. S., and Barbosa, L. M. (2018). Characterization of rainfall patterns in the semiarid brazil. *Anuário do Instituto de Geociências*, 41(2):397–409.

Wallace, J. M. and Hobbs, P. V. (2006). *Atmospheric Science: An Introductory Survey*, volume 92 of *International Geophysics Series*. Elsevier Academic Press, Amsterdam, 2nd edition.

Wangwongchai, A., Waqas, M., Dechpichai, P., Hlaing, P. T., Ahmad, S., and Humphries, U. W. (2023). Imputation of missing daily rainfall data; a comparison between artificial intelligence and statistical techniques. *MethodsX*, 11:102459.

Wani, O. A., Mahdi, S. S., Yeasin, M., Kumar, S. S., Gagnon, A. S., Danish, F., Al-Ansari, N., El-Hendawy, S., and Mattar, M. A. (2024). Predicting rainfall using machine learning, deep learning, and time series models across an altitudinal gradient in the north-western himalayas. *Scientific Reports*, 14(1):27876.