



Intelligent daily rainfall prediction for early warning using deep learning and satellite data: application to Bouaflé and Zuénoula stations, Ivory coast

Satti J. R. Kamenan¹, Ta M. Youan^{3,4}, Miessan G. Adja⁵, Sandona I. Soro¹, and Amani M. Kouassi²

¹Laboratory of Geographic Sciences, Civil Engineering and Geosciences, Polytechnic Doctoral School STI, UMRI Engineering Sciences and Techniques, National Polytechnic Institute Félix HOUPHOUËT-BOIGNY, Yamoussoukro, BP 1093, Ivory Coast

²School of Mines and Geology, National Polytechnic Institute Félix HOUPHOUËT-BOIGNY, Yamoussoukro, BP 1093, Ivory Coast

³University Center for Applied Remote Sensing Research (CURAT), Félix Houphouët-Boigny University, Abidjan, 01 B.P.V 34, Ivory Coast

⁴Training and Research Unit in Earth Sciences and Mineral Resources (UFR-STRM), Félix Houphouët-Boigny University of Abidjan-Cocody, Abidjan, 01 B.P.V 34, Ivory Coast

⁵Life and Earth Sciences Department (SVT), Higher Normal School (ENS), Abidjan, 01 B.P.V 34, Ivory Coast

Correspondence: Satti J. R. Kamenan (sattijhon@gmail.com)

Received: 12 September 2025 – Revised: 30 March 2026 – Accepted: 20 April 2026 – Published: 8 May 2026

Abstract. Recurrent flooding in the Marahoué region, particularly in Bouaflé and Zuénoula, underscores the need for reliable and operational tools to anticipate hydrological risks and support early warning systems. This study presents a rainfall forecasting framework based on Long Short-Term Memory (LSTM) neural networks, integrating satellite-derived precipitation products and reanalysis-based atmospheric variables to predict daily rainfall at the Bouaflé and Zuénoula stations at $t + 1$, $t + 3$, and $t + 7$ d lead times. The performance of the LSTM model was systematically evaluated and compared with commonly used reference models, namely Random Forest (RF), Extra Trees (ET), and XGBoost (XGB), using standard statistical metrics (R^2 , NSE, Pearson correlation coefficient R , normalized RMSE, and MAE). The results show that the LSTM model consistently outperforms the reference models across all forecasting horizons and at both study stations. At short and medium lead times ($t + 1$ and $t + 3$), LSTM exhibits strong predictive skill, with R^2 and NSE values exceeding 90 %, indicating an accurate representation of daily rainfall variability. Although performance decreases at the seven-day horizon due to increasing uncertainty and challenges in capturing extreme events, LSTM remains more robust than tree-based models, whose accuracy degrades markedly with increasing lead time. These findings confirm the relevance of LSTM-based approaches for rainfall forecasting and early warning applications in flood-prone regions. Future work will focus on integrating additional atmospheric predictors and applying advanced hyperparameter optimization techniques to improve long-term forecast reliability.

1 Introduction

Precipitation encompasses all meteoric waters, or hydrometeors, that fall to the Earth's surface, either in liquid form (rain) or solid form (snow, hail, sleet) (L'Hôte, 1993). In Africa, particularly in Ivory coast, precipitation consists exclusively of rainfall, whose intensity during the rainy season often causes severe flooding, leading to economic damages and loss of human lives. This hydroclimatic reality is well known in the Marahoué River Basin, where the localities of Bouaflé and Zuénoula are frequently affected by floods triggered by torrential rains during the rainy season, thereby threatening economic activities and human safety. Thus, accurate and quantitative near real-time rainfall forecasting can support the formulation of effective measures and help prevent human losses and material damages caused by flooding. The issue of rainfall prediction has long been a subject of scientific debate. Traditional approaches relying on statistical techniques have been used to quantitatively estimate rainfall. However, the complexity of rainfall processes, particularly their nonlinearity, makes prediction a challenging task (Wu and Chau, 2013). To address this, alternative methods have been developed to reduce nonlinearity, including singular spectrum analysis, empirical mode decomposition, wavelet analysis (Xiang et al., 2018; Gan et al., 2019), among other mathematical techniques. Nevertheless, the mathematical and statistical models employed often require high computational power (Singh and Borah, 2013) and may also be time-consuming.

Today, with advances in computer science, particularly artificial intelligence combined with hydrological modeling and meteorological satellite products integrated with efficient algorithms, new opportunities have emerged for achieving high-quality rainfall forecasts in near real time. Artificial intelligence, through machine learning, provides powerful algorithms capable of analyzing historical meteorological data to estimate future conditions, thus facilitating automated decision-making. Regarding satellite-derived meteorological products, the GPM IMERG program offers high-resolution temporal data suitable for global-scale monitoring, as demonstrated by recent studies (Simanjuntak et al., 2022). In fact, Simanjuntak et al. (2022) applied Long Short-Term Memory (LSTM), a deep learning algorithm, together with Random Forest (RF) models, to develop an early warning system for rainfall estimation in Indonesia, using data from the Himawari-8 geostationary satellite and GPM IMERG at 10 min intervals. Despite the great importance of having rainfall data at fine temporal resolutions, very few studies have been conducted in Ivory coast, and to date, no research has focused on near real-time rainfall prediction in any of its river basins. This study, therefore, complements previous works carried out in Ivory coast, particularly in the Marahoué Region, with the aim of improving existing approaches for the establishment of early warning systems for floods and rainfall forecasting. Its ambition is to explore deep learning tech-

niques combined with GPM IMERG satellite data for near real-time rainfall prediction at different stations within the Marahoué River Basin.

2 Presentation of the study area

The study area considered in this research is located between longitudes 5°32' and 6°19' W and latitudes 6°45' and 7°36' N. It corresponds to a portion of the Marahoué River basin (Fig. 1) and encompasses the departments of Bouaflé and Zuénoula in central-western Ivory coast. The topography is characterized by gently undulating plateaus, with elevations ranging from 200 to 400 m, interspersed with low-lying areas that are regularly affected by seasonal flooding. The climate is of a tropical sub-equatorial type, featuring two rainy seasons (April–July and September–October) and two dry seasons (August and November–March) (Assoko, 2022). Average temperatures vary between 25 and 28 °C, while annual precipitation

ranges from 1200 to 1600 mm. These climatic conditions, combined with a dense hydrographic network dominated by the Marahoué River, favor frequent flooding, particularly impacting agricultural zones and riverside communities. Geologically, the region rests on granitoid formations and Birimian series (Tagini, 1971; Kouamé et al., 2017). Economic activities, as highlighted by Peltre-Wurtz and Steck (1979) and Dje Bi (2015), are mainly centered on subsistence agriculture (yam, cassava, rice, maize), cash crops (coffee, cocoa), livestock farming (cattle, sheep, goats), and artisanal gold mining, which is particularly active in the Bouaflé department (Denis, 2016; Kouadio et al., 2018). However, the recurring floods severely disrupt production systems, increase the vulnerability of rural communities, and pose significant challenges for land use planning and territorial management.

3 Data and Materials

The data used in this study comprise both satellite-based climatic data and in-situ observations. Ground-based rainfall measurements were provided by the Airport, Aeronautical, and Meteorological Operations and Development Company (SODEXAM). The rainfall database covers daily records from 1 January 1961 to 31 December 2022 for the 10 operational stations in the Marahoué watershed (Bouaflé, Dianra, Kani, Kébi, Kongasso, Mankono, Morondo, Sarhala, Worofla, and Zuénoula). Special attention was given to the Bouaflé and Zuénoula stations, which served as the primary sites for the study. The historical records from these two stations include gaps, with missing data rates ranging from 7.29 % to 8.04 %. In addition, airborne data were collected from GPM IMERG and MERRA-2. GPM IMERG is a precipitation algorithm maintained by NASA and JAXA (Zubieta et al., 2017) that generates a relatively long satellite-based rainfall record of more than 20 years, with high spatial

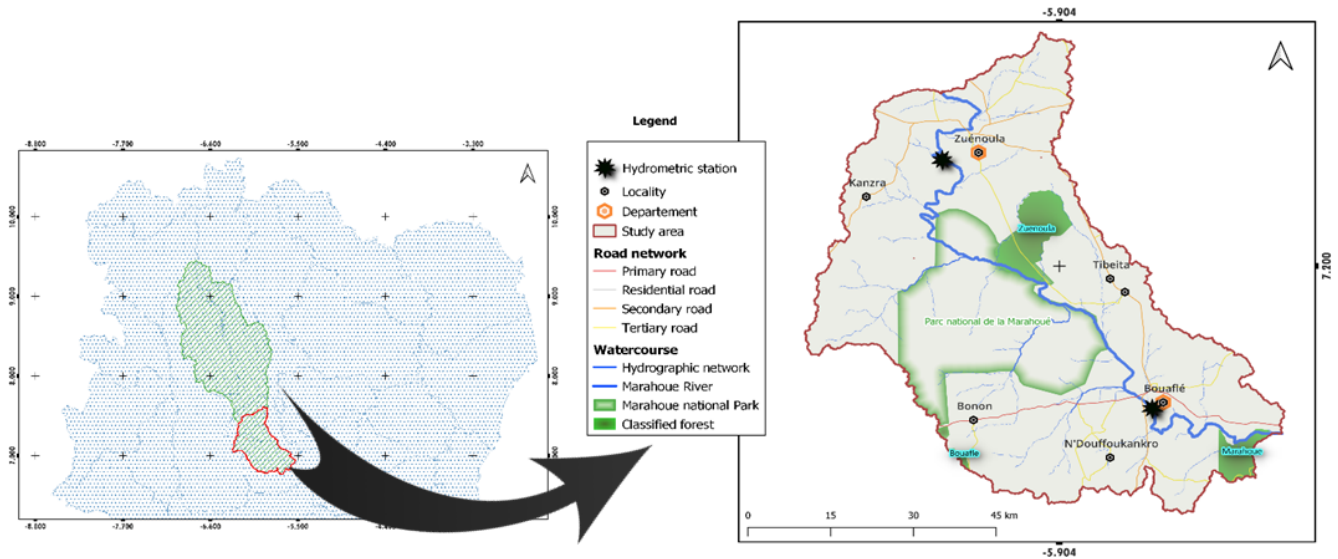


Figure 1. Location of the study area.

resolution (0.1°) and a temporal resolution of approximately 30 min, covering the entire globe. GPM IMERG data, available at <https://pmm.nasa.gov/data-access/downloads/> (last access: 5 March 2024), were compiled at three-hour intervals and then aggregated to daily resolution for the study period 2000–2024. These data were acquired and preprocessed using the Google Earth Engine (GEE) platform, a cloud-based computing environment dedicated to the management, processing, and analysis of large satellite and climate data repositories, in order to fill gaps in the in-situ rainfall time series. MERRA-2 (Modern-Era Retrospective analysis for Research and Applications, Version 2) provides a reanalyzed set of atmospheric data produced by the Goddard Earth Sciences Data and Information Services Center (GES DISC), affiliated with NASA's Goddard Space Flight Center (GSFC). This reanalysis integrates satellite observations, ground-based measurements, and numerical weather prediction models through data assimilation techniques to generate spatially and temporally consistent atmospheric fields. Several variables, including hourly mean temperatures at 2 m a.g.l., as well as essential atmospheric parameters such as 50 m wind, sea-level and surface pressure, and other atmospheric pressures (Molod et al., 2015), were aggregated to daily time steps over the period from 1 January 1983 to 28 April 2024. These data were collected via Google Earth Engine to complement the training dataset for the developed models.

4 Methodology and model development

The methodological approach adopted in this study (Fig. 2) is structured in two main steps: data preprocessing and the application of deep learning models to estimate near real-time rainfall. The first step involves data acquisition and prepara-

tion using Google Earth Engine (GEE) via Google Colab. The second step describes the implementation of the models for rainfall estimation, the identification of relevant variables, and the validation of the performance of the developed models.

4.1 Data Preprocessing

The methodological framework implemented in this study is based on two successive and complementary steps: prior validation of the GPM IMERG satellite product, followed by quality control of the in-situ data at the scale of a network of 10 stations. Daily precipitation data from GPM IMERG were compared with ground-based observations over the 2000–2017 period for all 10 operational stations within the watershed. The correlation coefficient (R) was calculated for each station to assess the temporal consistency between the two data sources, and scatter plots were generated and fitted to provide visual validation. This step was intended solely to verify the overall reliability of the satellite product before its potential use as an auxiliary source during data correction.

Quality control was subsequently applied to the daily time series from the 10-station network, which provides the basis for the spatial consistency assessment. The stations of Bouaflé and Zuénoula, later used for model development, are part of this network but were not the only stations considered during the outlier detection phase. The adopted approach combines statistical quality control with spatial consistency verification, following recommended by Durre et al. (2010). At the individual station level, the interquartile range (IQR) method, according to Tukey (1977), was applied to identify observations exhibiting unusual deviations from the station-specific distribution. A value x was considered atypical if $x > Q_3 + 1.5 \times \text{IQR}$ or $x < Q_1 - 1.5 \times \text{IQR}$, where Q_3

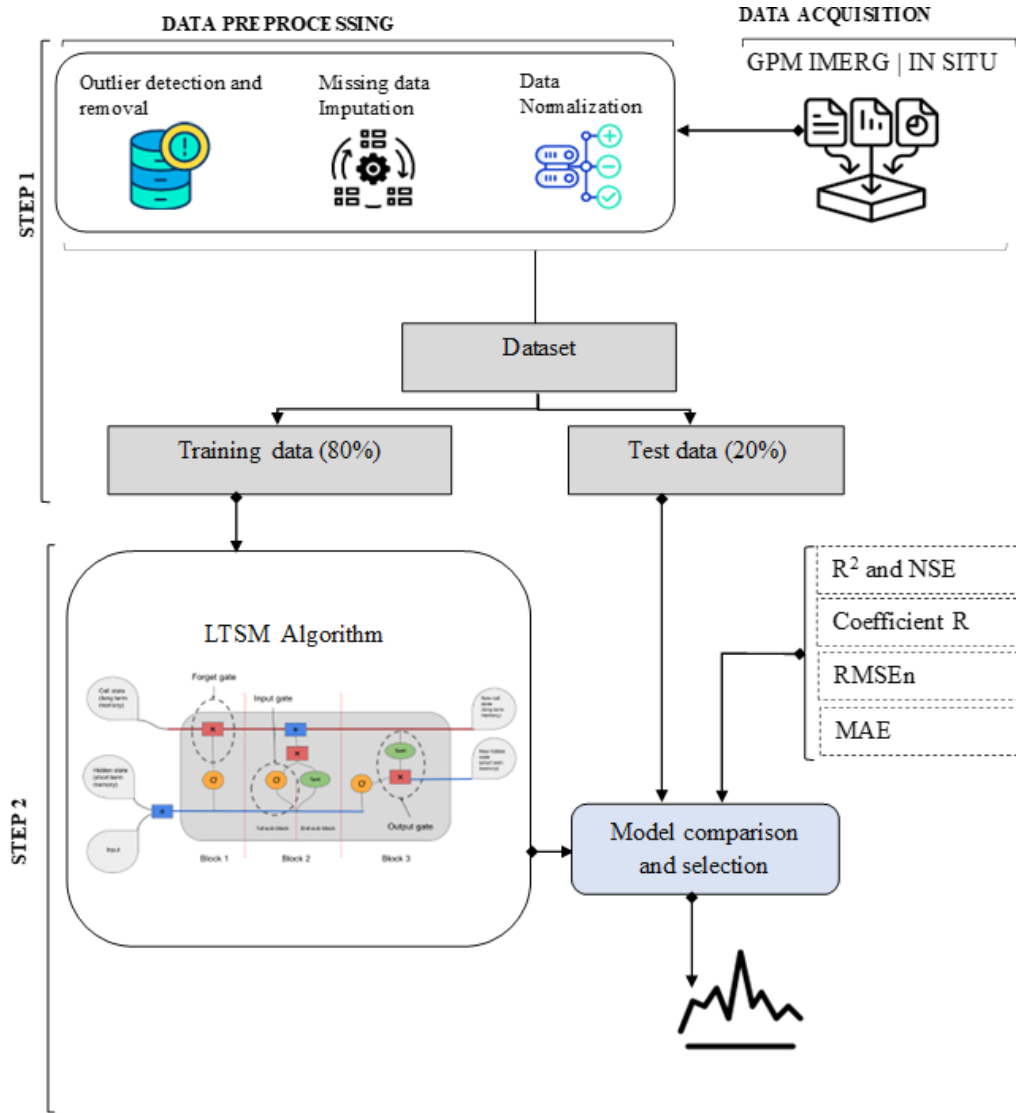


Figure 2. Methodological framework for near real time rainfall forecasting.

and Q_1 denote the first and third quartiles, respectively, and $IQR = Q_3 - Q_1$

Given the potentially localized nature of precipitation processes, a spatial consistency check based on the median and the Median Absolute Deviation (MAD) was implemented to avoid removing physically plausible extreme events. For a given day j , an observation $x_{i,j}$ at station i was compared to the network distribution and considered spatially inconsistent if $|x_{i,j} - \tilde{x}_j| > k \times MAD_j$ where \tilde{x}_j and MAD_j represent the network-wide median precipitation and the corresponding median absolute deviation for day j , respectively, with $MAD_j = \text{median}(|x_{i,j} - \tilde{x}_j|)$. The threshold k was analyzed for values between 2 and 3.5 in order to assess the sensitivity of detection rates and verify the robustness of the results.

Finally, an observation was classified as an outlier only if it was simultaneously atypical according to the IQR test and spatially inconsistent relative to the network. High-precipitation events that were spatially coherent were retained, even when exceeding statistical thresholds, to prevent the removal of meteorologically plausible extremes. Observations confirmed as erroneous were corrected or replaced only when necessary, using GPM IMERG as an auxiliary data. After applying this quality control protocol, the corrected dataset was used to develop predictive models specifically for the Bouaflé and Zuénoula stations. This dataset was normalized to a range between 0 and 1 using the formula in Eq. (1) as described by Hayder et al. (2023). This final operation performed on these data allows for an accelerated

learning process (Jingwei et al., 2022).

$$\hat{X} = \frac{X_i - \min(X)}{\max(X) - \min(X)} \quad (1)$$

where X_i is the input value, $\max(X)$ and $\min(X)$ refer, respectively, to the maximum and minimum values within the dataset X .

4.2 LSTM Model

The Long Short-Term Memory (LSTM) model constitutes an advanced architecture of the traditional Recurrent Neural Network (RNN), initially proposed by Hochreiter and Schmidhuber (1997) to address the intrinsic limitations of standard RNNs, particularly their difficulty in capturing long-term dependencies and retaining information across extended sequences (Le et al., 2019). Its originality lies in the integration of specific components, referred to as gates, which regulate the selective transmission, storage, and forgetting of information, thereby enabling an efficient management of temporal dependencies (Gers et al., 1999). Unlike conventional RNNs, the LSTM relies on a sequence of recurrently connected memory cells, whose architecture enhances both the stability of the learning process and the accuracy of predictions (Hayder et al., 2023). The operational mechanism of the LSTM is schematically represented in Fig. 3, adapted from Zhao and Obonyo (2020). According to these authors, the LSTM includes two forms of memory: Long-term memory (c_t): information retained over time, stored in the cell state at time t ; Short-term memory (h_t): information used immediately to perform the current task, represented by the hidden state at time t . The LSTM model operates as follows:

- *Forget gate f_t* . The model first determines which information from the previous cell state c_t should be retained or forgotten, based on the newly received input data;
- *Input gate i_t* . It then identifies the candidate new information c_t to be added to the long-term memory;
- *Cell state update c_t* . The new cell state is updated by combining the retained portion of the previous state with the new candidate information, modulated by the forget gate and the input gate;
- *Output gate o_t* . Finally, the model decides which part of the updated cell state should be exposed as output through the hidden state h_t , guided by the output gate.

4.3 Identification of daily rainfall predictors and dataset

The selection of predictors was carried out based on their availability as well as those used in previous studies by Sagna (2021) and Endalie et al. (2022). Special attention was given

to climatic factors, including temperature, atmospheric pressure, wind speed and direction, specific humidity, as well as sea surface temperature and temperature at 2 m a.g.l. Additionally, Integrated Water Vapor (IWV) data, although relevant for this type of study, were not available at a sufficient spatial resolution for the study area. A summary of these predictors is presented in Table 1. First, a correlation analysis was performed among the explanatory variables by computing the pairwise correlation matrix. Predictors exhibiting a correlation coefficient below 60 % with other predictors were retained, ensuring the selection of variables with minimal multicollinearity while preserving independent information for the LSTM modeling. Following this phase, the selected predictors were used as input variables for a Random Forest model to simulate daily precipitation. The relative importance of each predictor was then evaluated using the variable importance measures provided by the Random Forest algorithm. This analysis was used as a preliminary covariate selection step to identify the most informative predictors prior to training the LSTM model, thereby limiting redundant information, and improving model generalization. Variables with significant importance were considered relevant for daily precipitation forecasting due to their notable contribution to the model's accuracy.

4.4 Model construction and daily rainfall estimation

The methodological approach relied on the use of the Long Short-Term Memory (LSTM) model, which was implemented and optimized according to its architectural parameters within the PyTorch environment, widely used in deep learning applications (Xu, 2022). The input to the LSTM model consists of a three-dimensional tensor comprising the batch size, the input sequence length, and the number of input channels. The model output is also a tensor including the batch size and the sequence length, with a single output channel (1), corresponding to the predicted precipitation at the $t + 1$, $t + 3$, and $t + 7$ forecasting horizons (t expressed in days). Accordingly, three (3) distinct LSTM models were developed to perform the modeling at each of these forecasting horizons. To limit overfitting, several learning parameters were rigorously defined, including the number of hidden layers, the number of neurons per layer, the batch size, the number of iterations, the learning rate, and the temporal window length, which determines the data sampling interval. The optimization process consisted of adjusting one parameter at a time while keeping the others constant. For instance, the learning rate was initially set to 0.001, a value commonly adopted in the literature (Granet et al., 2018; Alavoine et al., 2024), and an automatic reduction of this rate was applied after five (05) epochs. The optimal value of each parameter was selected based on the minimization of the loss function and the improvement of fitting accuracy during the training process. When no improvement in the validation loss was observed after ten (10) epochs, an early stopping mechanism

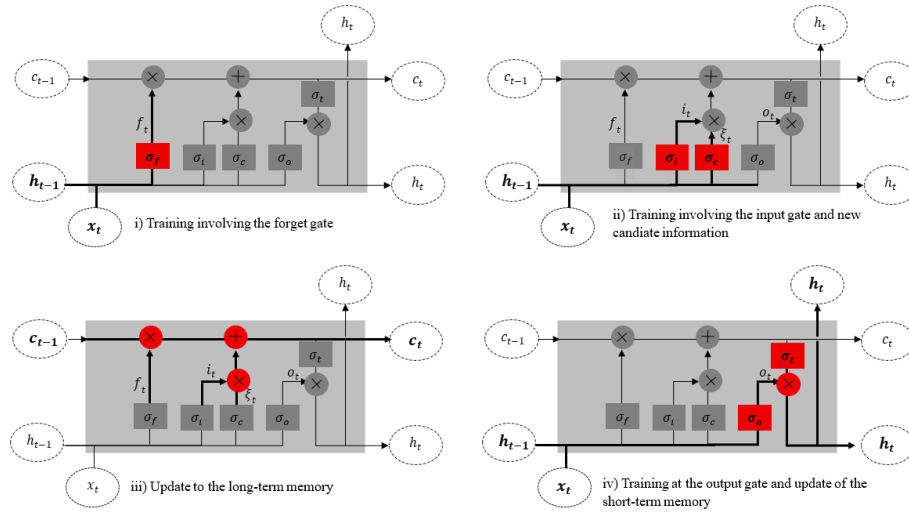


Figure 3. Operating process of the LSTM Model.

Table 1. Predictors identified for rainfall forecasting.

Predictors	Description
CLDPRS	Cloud top pressure
PS	Surface pressure
OMEGA500 (500 hPa)	Zonal wind
H250 (250 hPa) H500 (500 hPa) H850 (850 hPa)	Height
Q250 (250 hPa) Q500 (500 hPa) Q850 (850 hPa)	Specific humidity
T250 (250 hPa) T500 (500 hPa) T850 (850 hPa)	Air temperature
U250 (250 hPa) U500 (500 hPa) U850 (850 hPa)	Easterly wind
V250 (250 hPa) V500 (500 hPa) V850 (850 hPa)	Northward wind
Temp_min	Minimum air temperature at 2 m a.g.l.
Temp_max	Maximal air temperature at 2 m a.g.l.
Pre_Moy_7j	The cumulative rainfall of the previous 7 d

was activated. The Adam optimizer was used to update the network weights and biases in order to mitigate overfitting. All these steps were implemented using Python code executed on the Google Colab platform. To evaluate the performance of the LSTM model, it was compared with three reference models commonly used in hydrology, namely Ran-

dom Forest (RF), Extra Trees (ET), and Extreme Gradient Boosting (XGB). The optimization of these reference models was performed using a grid search approach, in which the values of each hyperparameter were tested across a pre-defined set of values. Once optimized, these reference models were also used to predict precipitation at the $t + 1, t + 3,$

and $t + 7$ forecasting horizons. After the model design phase, all developed models were trained on the normalized rainfall datasets from the Bouafé and Zuenoula stations, using a supervised learning approach based on labeled input–output pairs. Cross-validation was incorporated to optimize model hyperparameters and to assess the robustness and generalization capability of the trained models. The dataset used in this study consists of daily observations spanning the period from 1 January 1984 to 28 April 2024, corresponding to a total of 15 094 time steps. For modeling and rigorous evaluation purposes, the data were divided into two distinct subsets. Observations from 1 January 1984 to 22 January 2020 (time steps 1–13 585) were used for model training, while those from 23 January 2020 to 28 April 2024 (time steps 13 586–15 094) were reserved for independent model validation.

4.5 Model validation

4.5.1 Statistical indicators

In order to rigorously assess the quality and predictive performance of the developed models, a set of widely recognized statistical indicators commonly employed in predictive modeling and machine learning was adopted. These metrics were used not only to evaluate model accuracy during training and validation phases, but also to guide hyperparameter tuning and determine the optimal model configuration. The evaluation framework incorporated three error-based measures: Root Mean Square Error (RMSE) (Willmott, 1981), Normalized Root Mean Square Error (RMSE_n) (Loague and Green, 1991), and Mean Absolute Error (MAE) (Kassam, 1977; Fouotsa Manfouo et al., 2023; Renteria-Mena et al., 2024), together with efficiency and correlation-based criteria, namely the Nash–Sutcliffe Efficiency (NSE), the coefficient of determination (R^2) (Wright, 1921), and the Pearson correlation coefficient (R) (Pearson, 1909). During hyperparameter optimization, candidate models were compared using all these indicators. Particular emphasis was placed on RMSE as the primary selection criterion because of its sensitivity to large deviations, which is critical for daily rainfall modeling where extreme events strongly influence predictive reliability. MAE was also considered as a complementary metric, providing an assessment of the average magnitude of errors without disproportionately penalizing large deviations. When discrepancies occurred among the metrics, the configuration with the lowest RMSE was selected as the optimal model. MAE, NSE, and R^2 were used as complementary metrics to ensure consistent and acceptable model performance. These metrics were implemented in a Python environment, ensuring automated execution and consistent computation during both training and validation phases. The corresponding mathematical formulations of these indicators are presented below.

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^N (\hat{y}_i - y_i)^2}{N}} \quad (2)$$

$$\text{RMSE}_n = 100 \times \frac{\text{RMSE}}{\bar{O}} \quad (3)$$

$$\text{MAE} = \frac{\sum_{i=1}^N |y_i - \hat{y}_i|}{N} \quad (4)$$

$$R^2 (\%) = (100) \times \left[1 - \frac{\sum_{i=1}^N (\hat{y}_i - y_i)^2}{\sum_{i=1}^N (y_i - \bar{y})^2} \right] \quad (5)$$

$$\text{NSE} (\%) = (100) \times \left[1 - \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N (y_i - \bar{y})^2} \right] \quad (6)$$

$$R (\%) = (100) \times \left[\frac{\sum_{i=1}^N (y_i - \bar{y})(\hat{y}_i - \bar{\hat{y}})}{\sqrt{\sum_{i=1}^N (y_i - \bar{y})^2 \sum_{i=1}^N (\hat{y}_i - \bar{\hat{y}})^2}} \right] \quad (7)$$

where y_i is the observed data, \hat{y}_i is the predicted data, and N is the length of the dataset, \bar{y} is the mean of the observed data, and $\bar{\hat{y}}$ is the mean of the predicted data.

According to Hussain et al. (2023), a normalized RMSE (RMSE) lower than 10 % indicates excellent model performance, while a value between 10 % and 20 % reflects very good performance. When RMSE ranges between 20 % and 30 %, the model is considered moderately accurate, whereas values exceeding 30 % denote poor reliability. Furthermore, model performance improves as the MAE (Mean Absolute Error) approaches zero, which demonstrates higher predictive accuracy (Chicco et al., 2021). Regarding the coefficient of determination (R^2) and the Nash–Sutcliffe Efficiency (NSE), values above 90 % characterize an excellent model, those between 80 % and 90 % a very satisfactory model, those between 60 % and 80 % a satisfactory model, while values below 60 % indicate poor performance (Bodian et al., 2012). Finally, a value of R greater than or equal to 80 reflects a very strong correlation between variables, a value between 50 and 80 indicates a strong correlation, whereas a value below 50 suggests a weak correlation.

4.5.2 Quality of the rainfall estimates

In addition to the statistical performance criteria, a residual analysis was carried out to further evaluate the quality of rainfall estimates generated by the best-performing model and to assess their consistency with observed field data. Residuals were computed as the difference between observed and simulated values (Bodian, 2011). A statistical evaluation was then performed following the recommendations of Lek et al. (1996). According to these authors, residual analysis involves verifying their stationarity with respect to the predicted variable and ensuring that their mean

is statistically zero, which would indicate that the estimates are unbiased. Stationarity was examined through a graphical analysis of the residuals plotted against the model outputs. It was also deemed necessary to analyze the residuals according to classes of predicted values in order to evaluate potential variations with rainfall intensity. The distribution of residuals was assessed using the Kolmogorov–Smirnov test implemented in Python, with the following hypotheses: H0: the residuals follow a normal distribution and H1: the residuals do not follow a normal distribution. This test was performed to verify whether the residuals satisfy the normality assumption, which is part of the underlying assumptions of the model, and to identify potential deviations such as skewness or extreme values.

5 Results

5.1 Validation of GPM data and gap filling of in situ data

The Figs. 4 and 5 show the different scatter plots and the corresponding fittings performed for each station. The correlation coefficient was calculated, and a linear regression equation was determined.

Figures 4 and 5 present scatter plots comparing daily rainfall estimates from the satellite-based GPM IMERG product with in situ rain gauge observations at the Bouaflé and Zuénoula stations, respectively. At the Bouaflé station (Fig. 4), a strong positive correlation is observed between the two datasets, with a correlation coefficient of 0.80, indicating that GPM IMERG reproduces approximately 80 % of the variability in the in-situ rainfall observations. Similarly, at the Zuénoula station (Fig. 5), the correlation coefficient reaches 0.83, reflecting a high level of consistency between satellite-based estimates and ground-based measurements. Overall, these results demonstrate that GPM IMERG rainfall estimates are in good agreement with in situ observations at both stations, capturing more than 80 % of the observed rainfall variability. Based on this level of agreement, GPM IMERG data were considered sufficiently reliable and were therefore used to replace missing values in the ground-based rainfall records prior to model development. The quality control, based on a combination of the interquartile range (IQR) test and a spatial consistency check applied to the network of ten stations in the Marahoué basin, identified 689 outliers out of a total of 29 510 observations from the Bouaflé and Zuénoula stations, representing 2.34 %. This relatively low percentage indicates a generally good data quality and is consistent with the proportions commonly reported in climatological studies. Observations confirmed as erroneous after spatial verification were corrected only when necessary, using the previously validated GPM IMERG estimates as an auxiliary source, while spatially coherent extreme events were retained.

5.2 Selection of significant variables

Figure 6 presents the correlation matrix for twenty-one (21) variables, namely CLDPRS, PS, OMEGA500, H250, H500, H850, Q250, Q500, Q850, T250, T500, T850, U250, U500, U850, V250, V500, V850, Temp_min, Temp_max, and Pre_Moy_7j. Analysis of this matrix (Fig. 6) reveals a clear division of the variables into two distinct groups. Group 1 includes variables with correlation coefficients below 0.60, namely Pre_Moy_7j, OMEGA500, U500, Q250, T250, V500, V850, T500, U850, U250, Q850, Q250, V250, and Temp_min. The remaining predictors, with correlation coefficients above 0.60, form Group 2, which was excluded from the dataset used for modeling. For the subsequent analysis, variables from Group 1 (Fig. 7) were selected as input features for the Random Forest model to assess their relative importance.

Figure 8 presents the importance levels of Group 1 variables as evaluated by the Random Forest algorithm. The analysis of this figure indicates that most variables in Group 1 had a considerable impact on the model, with significance scores ranging from 0.04 to 0.18, except for V250 and Temp_min, whose importance levels were below 0.04. To construct a set of predictors with an acceptable level of importance, only variables with importance scores greater than 0.04 were retained for daily rainfall forecasting, as they were considered significant for the study. These variables include Pre_Moy_7j, OMEGA500, U500, Q250, T250, V500, V850, T500, U850, U250, Q850, and Q250.

5.3 Model performance

Before presenting the model performance results, it is important to note that all evaluation metrics are detailed in Sect. 4.5. The RMSE was used as the primary metric for selecting the optimal models, followed by normalized RMSE and MAE, while the R^2 and NSE coefficients were considered as complementary indicators to confirm the overall consistency of model performance.

5.3.1 Model optimization

The developed models were optimized according to several hyperparameters (Table 2). For Random Forest (RF) and Extra Trees (ET), the maximum number of features was determined using the “sqrt” option for the max_features parameter. In this configuration, the number of candidate predictors randomly selected at each split is equal to the square root of the total number of input features (\sqrt{p}), where p denotes the total number of predictors. The models were configured with 300 trees and a maximum depth of 20, while all other hyperparameters were retained at their default settings, indicating consistency in modeling the complex relationships of rainfall processes. For XGBoost (XGB), the optimal configuration included 200 trees, a maximum depth of 6, a learning rate of 0.1, and subsampling of observations and features set

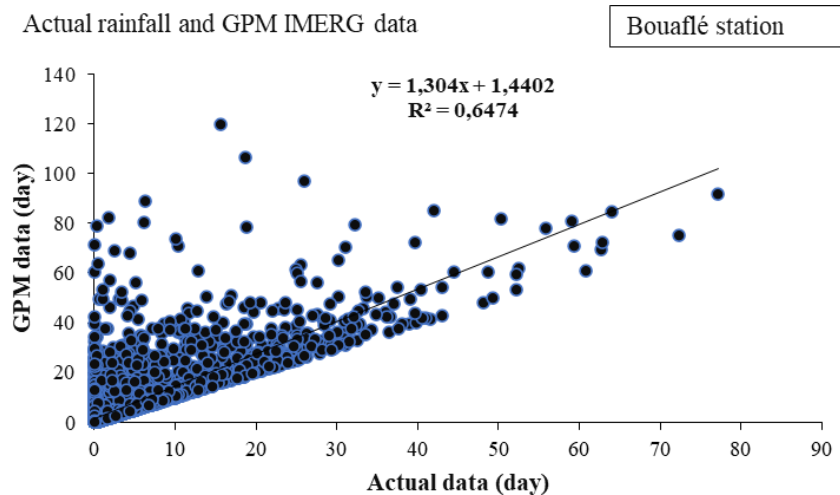


Figure 4. Scatter plot of GPM and in situ rainfall at Bouaflé station (2000–2017).

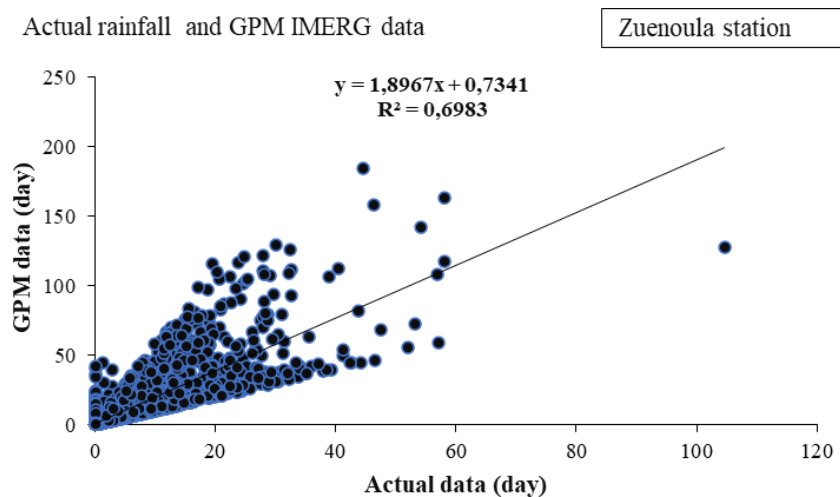


Figure 5. Scatter plot of GPM and in situ rainfall at Zuenoula station (2000–2017).

at 0.8, promoting generalization through regularization. The LSTM model was optimized with a 3-layer architecture, 100 neurons per layer, a regularization rate of 0.1, a learning rate of 0.001, and a batch size of 32, enabling the capture of complex temporal dependencies and hydro-climatic trends. All optimized models were then used to estimate daily precipitation at the Bouaflé and Zuenoula stations.

5.3.2 Model performance

The performance results of the developed models Long Short-Term Memory (LSTM), Extra Trees (ET), Random Forest (RF), and eXtreme Gradient Boosting (XGB) during the Training and Validation phases at the respective forecasting horizons $t + 1$, $t + 3$, and $t + 7$ are reported in Tables 3, 4, and 5. In addition to these results, scatter plot representations

for each station, calculated using the best-performing model, are provided in Figs. 9 and 10.

The combined analysis of the results (Tables 3, 4, and 5) highlights distinct behaviors depending on the model and the forecast horizon. At the $t + 1$ horizon, the LSTM model stands out as the most accurate. It exhibits the lowest RMSE, a normalized RMSE below 10 %, and an MAE under 1 mm, reflecting excellent precision. These performances are further supported by high R^2 and NSE values (≥ 95 %) for both stations, demonstrating the model's ability to faithfully reproduce the daily precipitation dynamics at Bouaflé and Zuenoula. At the $t + 3$ horizon, the LSTM maintains its high performance. The RMSE remains stable, with a normalized RMSE below 10 % and a low MAE, indicating the robustness of the model despite the extended forecast horizon. R^2 values range between 88 % and 90 %, confirming the consistency of the predictions and the model's capacity to cap-

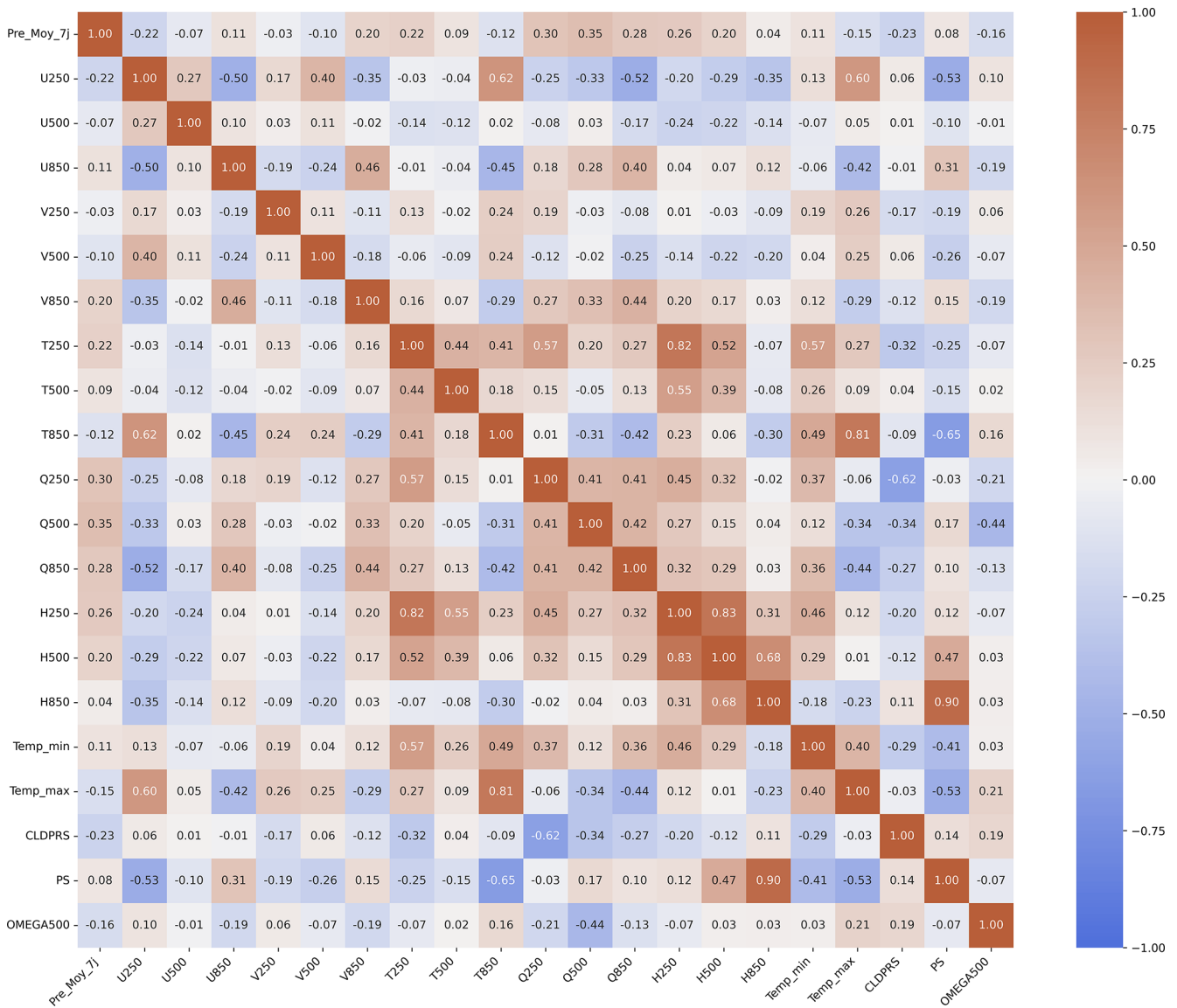


Figure 6. Pearson correlation matrix of all predictor variables.

ture daily variability. At the $t + 7$ horizon, the LSTM remains the only model providing statistically reliable forecasts. It maintains a controlled RMSE, a normalized RMSE below 10 %, and MAE values between 1.141 and 1.26 mm, while R^2 values remain between 70 % and 81 %. These results show that the LSTM can still provide acceptable forecasts, unlike other models (RF, ET, XGB), whose performances degrade significantly at this horizon. The RF and ET models, although effective at the $t + 1$ horizon with relatively low RMSE and R^2 and NSE values between 75 % and 80 %, see their performance gradually decrease as the horizon increases. At $t + 3$, their RMSE rises significantly (normalized RMSE 10 %–16 %) and MAE increases, while R^2 often drops below 65 %. By $t + 7$, their performance becomes very limited, with very high RMSE and normalized RMSE,

MAE exceeding 7 mm, and NSE values sometimes below 40 %. The XGB model is the least performing across all horizons, with high RMSE, higher normalized RMSE compared to other models, and high MAE, reflecting its limited ability to capture daily precipitation variability. Its R^2 values range from 37 % to 70 %, confirming its inability to accurately reproduce the time series dynamics. These results are primarily explained by the LSTM’s ability to model long-term temporal dependencies and the nonlinearities inherent in daily precipitation series, in contrast to ensemble models (RF, ET, XGB), whose tree-based architectures cannot efficiently capture the sequential structure of the data. Finally, the graphical representation of simulated versus observed data by the LSTM at the $t + 1$ horizon (Figs. 9 and 10) shows a point cloud aligned along a straight line during both training and

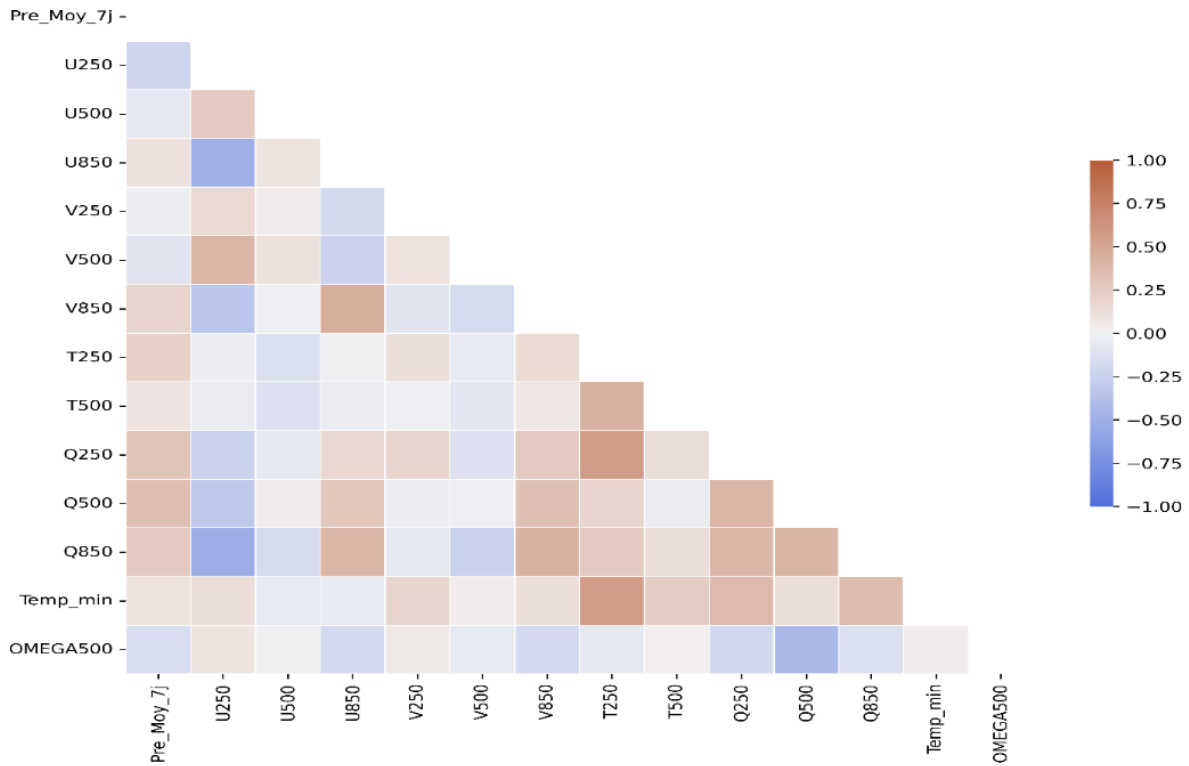


Figure 7. Pearson correlation matrix of the selected predictor variables.

Table 2. Optimal values of the hyperparameters of the developed models.

No.	Model	Parameters	Range	Best values
1	Random Forest (RF)	Maximum number of features	“sqrt”, “log2”	“sqrt”
		Number of trees	100, 200, 300	300
		Maximum tree depth	10, 15, 20	20
		Minimum samples to split a node	2, 5, 10	2
		Minimum samples per leaf	1, 2, 4	1
2	Extra Trees (ET)	Maximum number of features	“sqrt”, “log2”	“sqrt”
		Number of trees	100, 200, 300	300
		Maximum tree depth	10, 15, 20	20
		Minimum samples to split a node	2, 5, 10	2
		Minimum samples per leaf	1, 2, 4	1
3	XGBoost (XGB)	Number of trees	100, 200, 300	200
		Maximum depth	3, 6, 9	6
		Learning rate	0.01, 0.1, 0.2	0.1
		Fraction of samples used	0.8, 0.9, 1.0	0.8
		Fraction of features per tree	0.8, 0.9, 1.0	0.8
4	LSTM	Number of layers	1, 2, 3, 4	3
		Number of hidden neurons	100, 200, 300	100
		Regularization rate (dropout)	0, 0.1, 0.2	0.1
		Learning rate	0.001, 0.0001, 0.1	0.001
		Batch size	10, 20, 32, 64	32

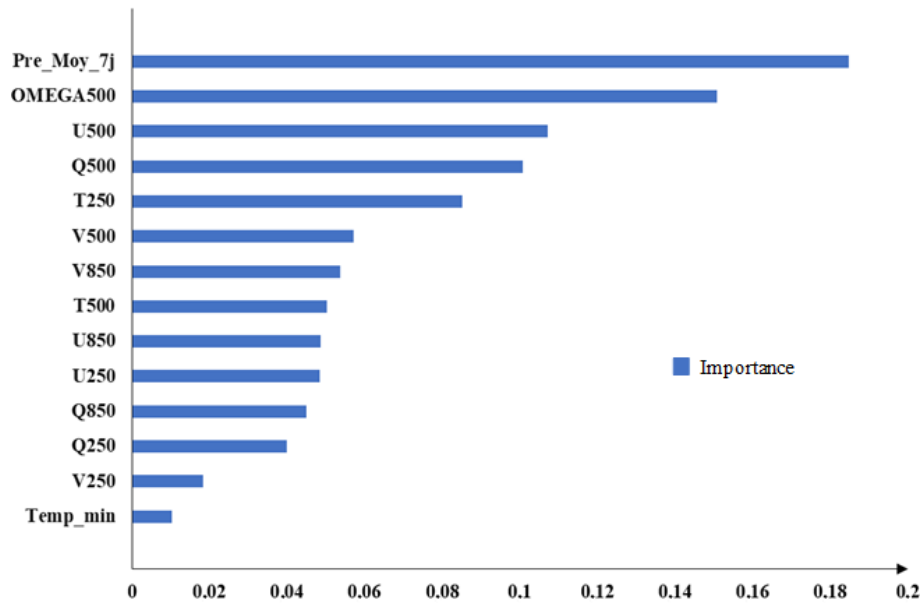


Figure 8. Variable importance according to the Random Forest algorithm.

Table 3. Model performance at the $t + 1$ forecasting horizon.

Models	Stations	Training phase					Validation phase				
		R^2	NSE	MAE	RMSE _n	R	R^2	NSE	MAE	RMSE _n	R
ET _{t+1}	Bouafle	78.26	78.26	7.663	12.03	89.50	79.89	79.89	8.419	11.47	88.35
RF _{t+1}	Bouafle	75.90	75.90	7.699	12.66	88.58	76.11	76.11	8.442	12.18	88.64
XGB _{t+1}	Bouafle	67.75	67.75	8.184	14.65	87.09	69.23	69.23	7.512	11.84	88.62
LSTM _{t+1}	Bouafle	96.00	96.00	0.670	9.34	97.00	95.00	96.00	0.852	9.12	97.00
ET _{t+1}	Zuenoula	75.67	75.67	4.687	14.17	88.76	78.85	78.85	4.427	10.88	89.02
RF _{t+1}	Zuenoula	76.33	76.33	5.419	13.98	88.56	79.19	79.19	4.405	10.26	89.41
XGB _{t+1}	Zuenoula	69.19	69.19	4.727	15.95	86.22	73.11	73.11	3.684	11.27	88.43
LSTM _{t+1}	Zuenoula	95.00	95.00	0.61	8.96	97.00	95.00	95.00	0.694	9.35	97.00

validation. This alignment underscores the strong correlation between RMSE, MAE, and the actual observations. A few points deviate from the main trend, reflecting LSTM estimation errors and confirming the calculated normalized RMSE and MAE values.

The pluviograms shown in Figs. 11 and 12 were obtained during the Training and Validation phases of the LSTM model at $t + 1$, $t + 3$, and $t + 7$. The analysis of the pluviograms (Figs. 11 and 12) compared to observations in Bouaflé and Zuenoula highlights a progressive deterioration in performance as the forecasting horizon increases. In Bouaflé, during the training phase, the $t + 1$ forecast generally reproduces the observed rainfall dynamics well, particularly the timing of peaks and dry periods, although some maxima are slightly underestimated or overestimated. At $t + 3$, the agreement remains satisfactory for moderate events, but a greater dispersion is observed around intense peaks, reflecting increased errors and a loss of precision in

amplitude estimation. At $t + 7$, the discrepancies become more pronounced; although the model is still able to capture the occurrence of major rainfall events, their intensity is estimated more irregularly. During the validation phase, this trend becomes even more evident. At $t + 1$, the model maintains a good ability to detect extreme events, even if some high peaks show noticeable differences. At $t + 3$, the variability of the estimates increases, particularly for heavy rainfall events, where both underestimations and significant overestimations are observed. At $t + 7$, the dispersion is the most pronounced, with maximum intensities reproduced more irregularly, especially during extreme events. In Zuenoula, during training, the $t + 1$ forecast shows very good agreement with observations. Rainfall peaks, including intense episodes, are generally well reproduced in terms of both timing and amplitude, despite a few slight occasional overestimations. At $t + 3$, the overall rainfall dynamics remain well captured, but greater variability is observed around

Table 4. Model performance at the $t + 3$ forecasting horizon.

Models	Stations	Training phase					Validation phase				
		R^2	NSE	MAE	RMSE _n	R	R^2	NSE	MAE	RMSE _n	R
ET _{t+3}	Bouafle	63.03	63.03	7.916	15.682	79.41	64.67	64.67	7.7118	21.31	79.69
RF _{t+3}	Bouafle	55.13	55.13	8.784	17.277	78.29	53.62	53.62	8.3392	27.90	79.67
XGB _{t+3}	Bouafle	40.90	40.90	7.136	19.830	64.74	43.33	43.33	7.8814	28.15	61.02
LSTM _{t+3}	Bouafle	90.00	90.00	0.921	9.51	95.00	83.00	83.00	0.998	10.00	92.00
ET _{t+3}	Zuenoula	62.61	62.61	8.735	17.569	79.36	63.84	63.84	8.2427	23.39	77.79
RF _{t+3}	Zuenoula	55.36	55.36	7.448	19.197	78.23	58.56	58.56	8.0638	26.80	76.35
XGB _{t+3}	Zuenoula	43.54	43.54	7.698	21.590	66.71	44.86	44.86	7.502	22.34	75.69
LSTM _{t+3}	Zuenoula	88.00	88.00	0.923	9.55	94.00	81.00	81.00	0.999	10.00	92.00

Table 5. Model performance at the $t + 7$ forecasting horizon.

Models	Stations	Training phase					Validation phase				
		R^2	NSE	MAE	RMSE _n	R	R^2	NSE	MAE	RMSE _n	R
ET _{t+7}	Bouafle	58.92	58.92	8.927	16.532	76.86	56.98	56.98	8.2326	27.42	76.11
RF _{t+7}	Bouafle	43.21	43.21	7.839	19.439	66.15	52.72	52.72	6.8623	23.26	65.34
XGB _{t+7}	Bouafle	37.62	37.62	6.169	20.372	61.85	34.98	34.98	6.9864	27.18	65.57
LSTM _{t+7}	Bouafle	78.00	78.00	1.234	12.53	85.00	81.00	81.00	1.164	10.78	84.00
ET _{t+7}	Zuenoula	35.91	35.91	9.768	23.003	67.24	38.55	38.55	9.3191	23.16	65.29
RF _{t+7}	Zuenoula	31.77	31.77	8.538	23.736	56.84	40.08	40.08	8.4865	24.85	54.43
XGB _{t+7}	Zuenoula	41.34	41.34	8.715	22.007	64.83	47.25	47.25	8.6505	23.03	69.21
LSTM _{t+7}	Zuenoula	70.00	70.00	1.26	12.57	84.00	77.00	77.00	1.141	10.68	84.00

the maxima, with more noticeable discrepancies for moderate to strong events. At $t + 7$, the differences become more apparent, with peak intensities sometimes attenuated or, conversely, amplified, indicating a progressive loss of precision in the estimation of extremes. During validation, this trend becomes more pronounced. The $t + 1$ forecast retains satisfactory skill in detecting major events, including heavy rainfall, although some extreme peaks show significant amplitude differences. At $t + 3$, divergences become more frequent, particularly for very intense episodes, where significant underestimations or overestimations are observed. At $t + 7$, dispersion is the most pronounced, with uncertain reproduction of extreme maxima. Overall, in Zuenoula as in Bouafle, the results confirm robust performance at short lead time ($t + 1$), still acceptable accuracy at medium lead time ($t + 3$), and a more marked degradation at longer lead time ($t + 7$), particularly for extreme rainfall events, which is consistent with the expected behavior of multi-horizon forecasting models.

5.4 Quality of the rainfall estimates

The results of the quality analysis of the rainfall estimates are presented in accordance with the methodological approach detailed in Sect. 4.5.2.

5.4.1 Residual analysis

The analysis of Table 6 indicates the following:

- *At the Bouafle station.* During the training phase, predicted precipitation ranges from 0 to 170.99 mm, with a mean of 2.72 mm. The mean residual, representing the average difference between observed and predicted precipitation, is approximately -0.05 mm, indicating that the model reproduces the observed in situ precipitation very closely on average. During the validation phase, predicted precipitation ranges from 0 to 123.1 mm, with a mean of 5.76 mm. Residuals vary between -12.2 and 13.7 mm, demonstrating that the LSTM model at $t + 1$ reproduces observed values accurately and captures the same temporal trends as the observed data.
- *At the Zuenoula station.* During training, predicted precipitation has a mean of 2.12 mm, very close to the observed mean of 2.24 mm. The minimum and maximum predicted values are 0.00 and 183.75 mm, respectively, which are comparable to the observed extremes of 0.00 and 190.84 mm. The mean residual is approximately 0.12 mm, and 75 % of predicted precipitation values have residuals below 0.00 mm. During validation, a similar trend is observed, with residuals ranging from -23.92 to 12.27 mm and a mean of -0.24 mm, reflecting minor deviations of the LSTM model at $t + 1$. These

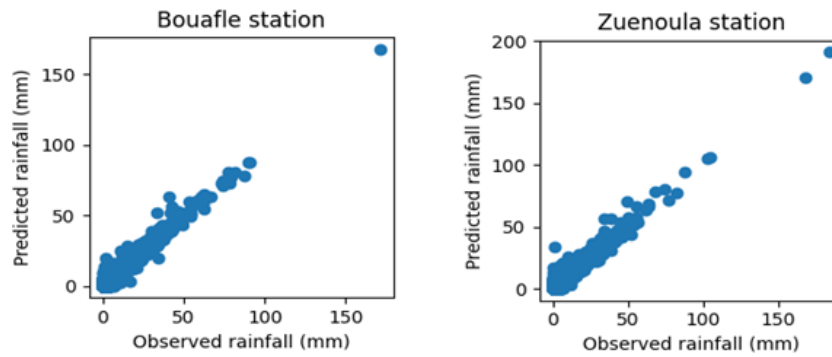


Figure 9. Scatter plots during the training phase of the LSTM model at $t + 1$.

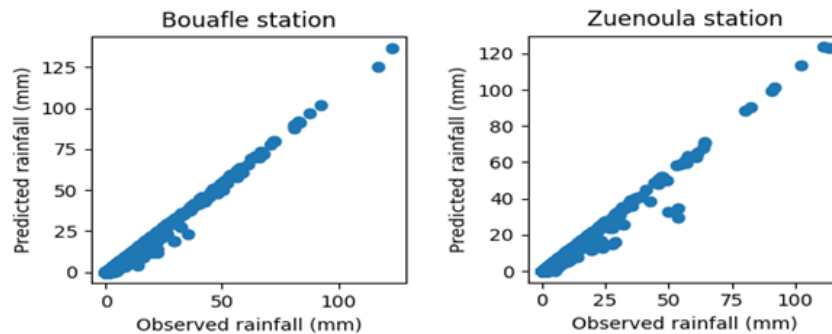


Figure 10. Scatter plots during the validation phase of the LSTM model at $t + 1$.

results indicate that the LSTM model at $t + 1$ accurately reproduces observed precipitation during the training phase.

5.4.2 Normality test of rainfall residuals

The results of the Kolmogorov–Smirnov tests performed on the residual rainfall values during the training and validation phases are presented in Table 7.

The analysis of Table 7 indicates that during the training phase, the Kolmogorov–Smirnov (KS) test values for each station range between 0.3289 and 0.3321. All corresponding p -values are 0.0000, which is below the significance level of $\alpha = 0.05$. These results suggest rejection of the null hypothesis (H_0) and acceptance of the alternative hypothesis (H_1) for both study stations, indicating that the residuals computed during the training phase do not follow a normal distribution.

During the validation phase, the KS test values for each station range from 0.3116 to 0.3269, with corresponding p -values of 0.0000, also below the significance threshold of $\alpha = 0.05$. These results again indicate rejection of (H_0) and acceptance of (H_1), showing that the residuals during validation are not normally distributed.

In summary, residuals obtained from the various training and validation phases of the LSTM model at $t + 1$ do not follow a normal distribution. Nevertheless, analyzing the dis-

persion of residuals remains necessary to assess the behavior and accuracy of the model's predicted data.

5.4.3 Relationship between predicted rainfall and residuals

The analysis of Figs. 13 and 14 indicates that at the Bouafé and Zuenoula stations, during the training phase, most residual values are clustered around zero. Furthermore, the relationship between residuals and predicted precipitation demonstrates a good fit for low precipitation values, which correspond to the majority of points concentrated near zero. In contrast, for high precipitation events, larger errors are observed, resulting in significant dispersion of points in each plot. These observations suggest that the model tends to underestimate or overestimate certain high-intensity precipitation events at both stations.

During the validation phase, residuals also show a strong clustering around zero; however, dispersion remains evident for higher precipitation values. This distribution reflects a continued tendency of the model to under- or overestimate some of the intense precipitation events. Overall, the LSTM model at $t + 1$ proves to be the most effective neural network approach for forecasting daily rainfall time series at the Bouafé and Zuenoula stations within the study area. It reproduces, in most cases, more than 95 % of the observed precipitation at the t -step forecast. Predicted precipitation follows

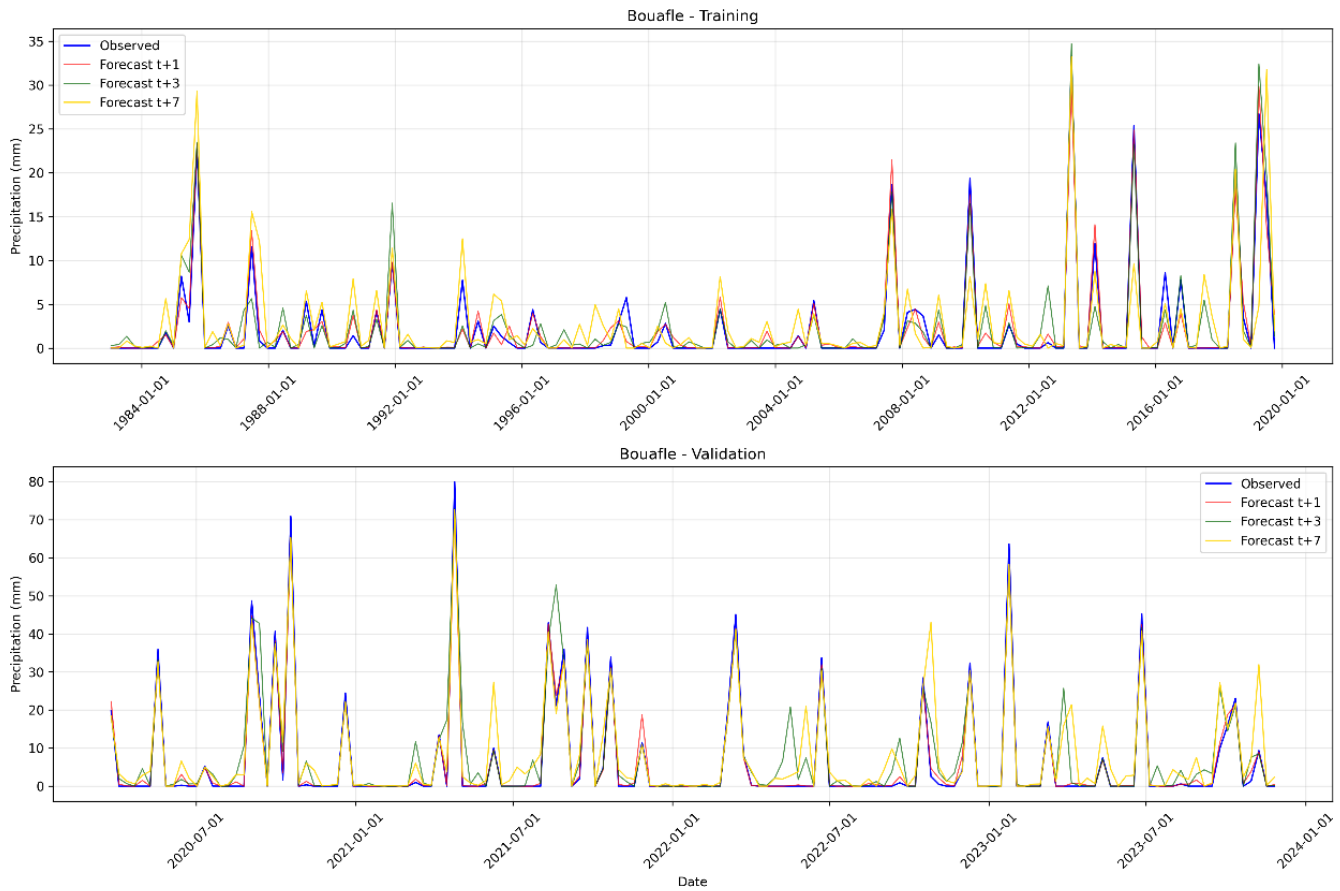


Figure 11. Rainfall charts of the Bouaflé station.

Table 6. Statistical parameters of predicted and observed rainfall.

		Training			Validation		
		Observed	Predicted	Residuals	Observed	Predicted	Residuals
Bouaflé	Nb	13 578	13 578	13 578	1502	1502	1502
	Av	2.67	2.72	-0.05	5.63	5.76	-0.14
	Std	6.88	6.74	1.37	14.53	13.47	1.74
	Min	0	0	-14.01	0	0	-12.2
	25 %	0	0	-0.27	0	0	-0.5
	50 %	0	0.13	-0.04	0	0.2	0
	75 %	1.77	2.28	0	2.33	3.63	0
	Max	167.47	170.99	22.39	136.7	123.1	13.7
Zuenoula	Nb	13 578	13 578	13 578	1502	1502	1502
	Av	2.24	2.12	0.12	3.86	4.10	-0.24
	Std	6.45	5.97	1.45	11.31	10.72	1.71
	Min	0	0	-8.35	0	0	-23.92
	25 %	0	0	-0.05	0	0	-0.48
	50 %	0	0	0	0	0.16	0
	75 %	0.87	1.42	0	1.50	2.46	0
	Max	190.84	183.75	32.60	123.47	113.04	12.27

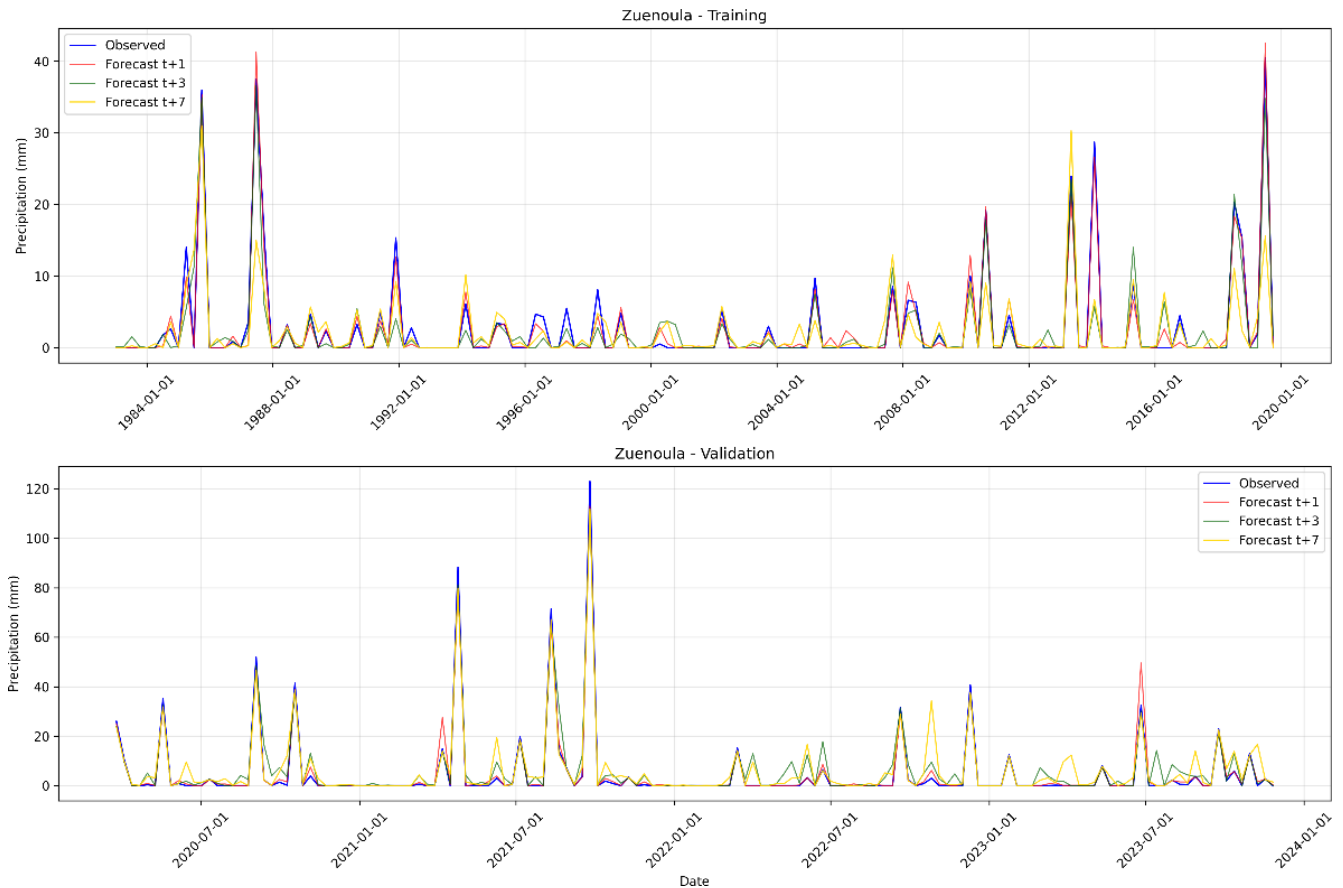


Figure 12. Rainfall charts of the Zuénoula station.

Table 7. Kolmogorov–Smirnov (KS) test of residuals calculated by LSTM model at $t + 1$.

Station	Training			Validation		
	Calculated K–S	<i>P</i> -value	Normally distributed?	Calculated K–S	<i>P</i> -value	Normally distributed?
Bouafle	0.3321	0.0000	No	0.3136	0.0000	No
Zuénoula	0.3289	0.0000	No	0.3116	0.0000	No

similar temporal patterns and variability as the observed data. Nevertheless, the model exhibits a tendency to slightly overestimate or underestimate extreme precipitation peaks.

5.4.4 Relationship between observed rainfall and residuals

A detailed analysis of the residuals (Figs. 15 and 16) reveals the presence of bias across all precipitation classes, both at Bouaflé and Zuénoula, during training and validation. A class-wise examination shows that the errors are not randomly distributed around zero. At Bouaflé, during training, low precipitation events (0–26 mm) exhibit a slight positive residual trend for some categories, indicating underes-

timation. In contrast, for intermediate classes (26–31 mm), residuals are predominantly negative, reflecting a tendency to overestimate moderate precipitation. This trend becomes more pronounced for high precipitation events (> 31 mm), where negative residuals dominate and the error magnitude increases significantly. Thus, the model exhibits an intensity-dependent bias: underestimation for low values and increasing overestimation as precipitation intensity rises. At Zuénoula, this behavior is even more marked during training. Several low-intensity points deviate more from zero, while residuals for intermediate classes become strongly negative. For high precipitation events, overestimation is pronounced, with large deviations. This indicates a bias that increases with precipitation intensity, suggesting poor calibration of

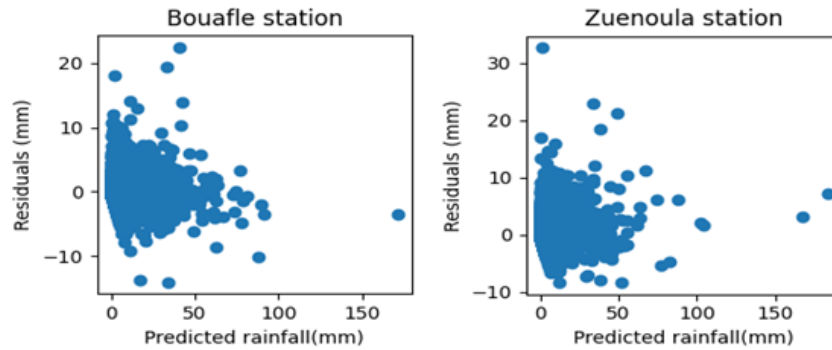


Figure 13. Distribution of residuals with respect to predicted rainfall during the Training phase.

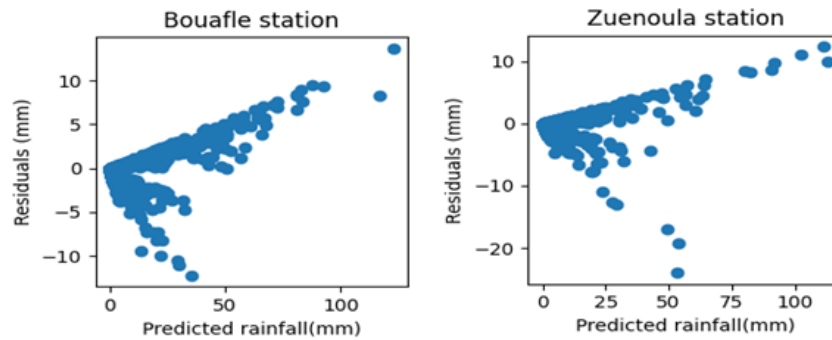


Figure 14. Distribution of residuals with respect to predicted rainfall during the Validation phase.

the model for extreme events. During validation, these biases persist, confirming that they are not merely a training artifact. At Bouafle, low precipitation events remain slightly underestimated, whereas intermediate and high classes are mostly overestimated, with greater dispersion than in training. At Zuenoula, the bias structure is similar: high precipitation events are consistently overestimated, and errors increase with intensity, highlighting the model's persistent difficulty in accurately reproducing out-of-sample extremes.

In summary, although the overall mean of residuals may approach zero, this masks compensatory biases between classes. The model exhibits a directional error dependent on precipitation intensity, characteristic of a nonlinear calibration issue, where low, medium, and high rainfall are not reproduced with the same accuracy. Despite these limitations, the results show overall satisfactory performance for predicting low to medium precipitation, which constitutes the majority of rainfall occurrences at both stations. The consistency observed between training and validation phases reinforces the model's robustness for operational use, particularly in early warning contexts where accurate reproduction of ordinary rainfall is essential.

6 Discussion

Several models, including Random Forest (RF), Extra Trees (ET), XGBoost (XGB), and Long Short-Term Memory

(LSTM), were developed to predict daily precipitation observed at the Bouafle and Zuenoula rainfall stations. This approach required the establishment of three distinct forecasting frameworks corresponding to time steps $t + 1$, $t + 3$, and $t + 7$, in order to analyze model performance across longer temporal sequences. The results highlight marked differences in performance among the tested forecasting models (ET, RF, XGB, and LSTM) and a strong dependence on the forecasting horizon. Overall, the LSTM model emerges as the most robust and highest-performing approach for predicting daily precipitation at the Bouafle and Zuenoula stations, regardless of the time horizon considered. At the short-term horizon ($t + 1$), the high performance of the LSTM characterized by R^2 and NSE values exceeding 95 % and low errors (MAE and normalized RMSE) demonstrates excellent reproduction of daily precipitation dynamics. Comparable results were reported by Hakizimana (2023) in Burundi, with determination coefficients between 0.84 and 0.98 and Nash indices ranging from 0.74 to 0.95, as well as by Endalie et al. (2022) in Ethiopia, who reported an NSE of 0.81 and a Pearson correlation coefficient of 0.9972 (99.72 %) for daily precipitation forecasts. The model's ability to faithfully reproduce observed precipitation, confirmed in this study by the strong alignment of scatter plots and the coherence of hyetographs, suggests good generalization of the network and an absence of significant overfitting, in line with observations by Gers et al. (1999) on long short-term memory

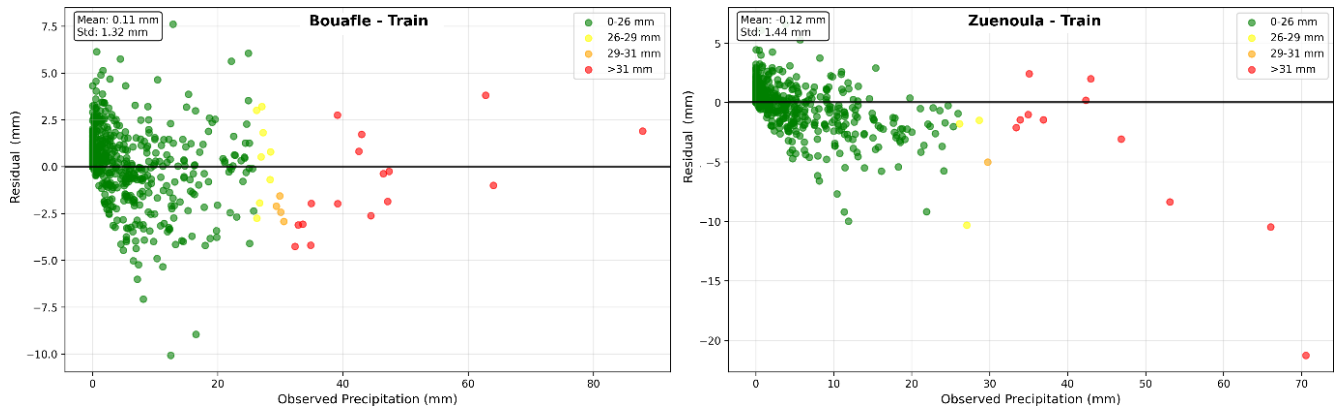


Figure 15. LSTM model residuals as a function of observed rainfall, grouped by rainfall intensity classes in training phase.

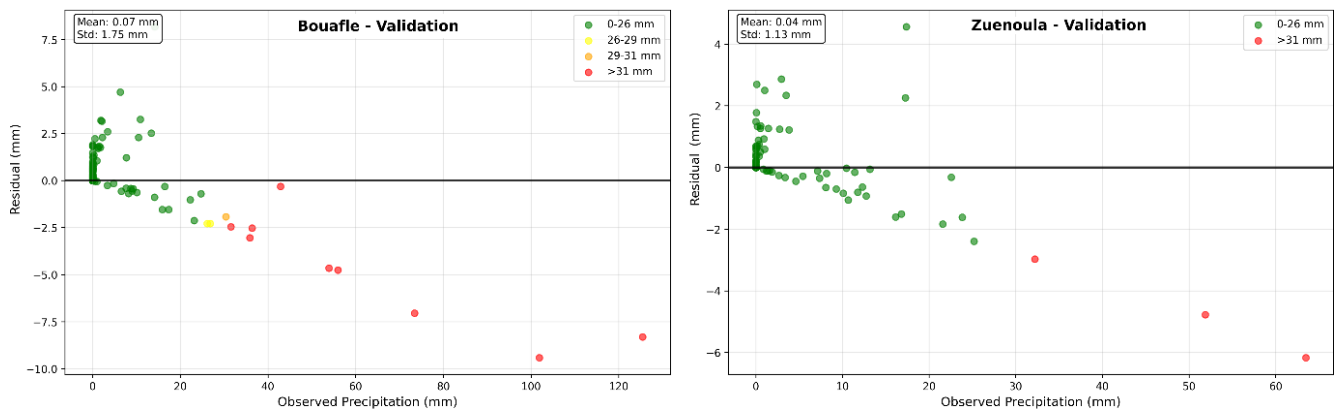


Figure 16. LSTM model residuals as a function of observed rainfall, grouped by rainfall intensity classes in validation phase.

networks. In comparison, ET and RF models show satisfactory performance only at this horizon, while XGB remains less accurate, indicating a lower capacity of tree-based models to capture local rainfall variability, as reported in several comparative studies in hydrology. As the forecasting horizon increases ($t + 3$ and $t + 7$), a gradual degradation of performance is observed across all models, a phenomenon commonly reported in the hydrometeorological time series literature. However, this degradation is much less pronounced for LSTM, which maintains statistically acceptable levels of accuracy up to $t + 7$, unlike decision-tree-based models (ET, RF, and XGB), whose performance becomes limited at medium and long-term horizons. These findings are consistent with results reported by Hao and Bai (2023), and Waqas et al. (2024). This difference can be attributed to the intrinsic capacity of LSTM to model long-term temporal dependencies and complex nonlinearities, characteristic of daily precipitation series. As shown by Hochreiter and Schmidhuber (1997) and later by Gers et al. (1999), the LSTM architecture mitigates issues of gradient explosion or vanishing, enabling efficient learning of long-range dependencies, whereas ensemble models primarily treat observations in a static manner. Residual analysis provides additional insights into fore-

cast quality. The non-normality of residuals, as evidenced by the Kolmogorov–Smirnov test, reveals a systematic structuring of errors across precipitation classes. This non-normality primarily reflects asymmetry and non-uniform dispersion of the residuals rather than purely random behavior. Moderate to high precipitation events are predominantly associated with negative residuals, indicating a tendency of the model to overestimate these events, whereas some low-intensity events exhibit slight underestimation. This directional organization of errors highlights an intensity-dependent bias and suggests a nonlinear calibration issue, as the different precipitation classes are not reproduced with the same level of accuracy. The increasing dispersion of residuals at higher intensities underscores the persistent difficulty of the model in accurately representing extreme events, which remains a well-documented limitation of LSTM models applied to daily precipitation series (Endalie et al., 2022). This behavior, frequently observed in Deep Learning models applied to precipitation, is attributable to the rarity, irregularity, and strong skewness of intense rainfall, which limits adequate representation during training. Analysis by rainfall intensity classes confirms that the LSTM model performs particularly well for light to moderate precipitation, which constitutes the

majority of occurrences in the study area. In contrast, for heavy rainfall, errors increase, reflecting a structural limitation of the model rather than mere overfitting, since this behavior is observed in both Training and Validation phases. Recent studies similarly indicate that LSTM networks perform very well for reproducing ordinary rainfall regimes but face limitations in representing extreme events. In particular, Baste et al. (2025) showed that LSTM models struggle to extrapolate extreme precipitation in a physically realistic manner, especially when observed intensities exceed the range covered by training data. Similarly, Sham et al. (2025) highlighted a marked tendency of LSTM networks to smooth high-intensity rainfall, resulting in increasing deviations between observed and predicted precipitation during extreme events. These findings confirm that, although LSTM networks are particularly effective for modeling light to moderate rainfall, their capacity to accurately reproduce extreme precipitation remains limited. To mitigate these limitations, several approaches can be considered: (i) applying post-prediction bias correction methods to adjust systematic overestimations or underestimations; (ii) enriching the training dataset with extreme precipitation events, whether historical or synthetic, to better represent rare occurrences; (iii) incorporating additional meteorological variables or exogenous covariates to better capture factors influencing extreme rainfall; (iv) exploring hybrid models combining LSTM with physical or statistical models specifically designed for extreme event modeling; and (v) optimizing the LSTM architecture by increasing the number of layers or neurons, or by employing advanced regularization techniques to reduce asymmetry and heteroscedasticity in the residuals.

7 Conclusion

Recurrent flooding in the Marahoué region, particularly in the cities of Bouaflé and Zuénoula, highlights the pressing need for reliable and operational tools to anticipate hydrological risks and enhance early warning capabilities. In response to this challenge, this study proposed a data-driven rainfall forecasting framework based on Long Short-Term Memory (LSTM) neural networks combined with satellite-derived precipitation data. Three forecasting configurations were developed to predict daily rainfall at $t + 1$, $t + 3$, and $t + 7$ horizons, and their performance was systematically compared with commonly used ensemble-based reference models, namely Random Forest (RF), Extra Trees (ET), and XGBoost (XGB). The results demonstrate that the LSTM model consistently outperforms the reference models across all forecasting horizons and both study stations. At the short-term horizon ($t + 1$), LSTM exhibits excellent predictive skill, with R^2 and NSE values exceeding 95 % and low error metrics (MAE and normalized RMSE), indicating an accurate reproduction of daily rainfall dynamics. At the medium-term horizon ($t + 3$), LSTM maintains strong performance,

confirming its robustness in capturing temporal dependencies beyond immediate lead times. Although a decline in accuracy is observed at the long-term horizon ($t + 7$), the LSTM remains the only model achieving statistically acceptable performance, whereas tree-based models experience a pronounced degradation. The comparative analysis confirms that the superior performance of LSTM is primarily attributable to its ability to model long-term temporal dependencies and nonlinear relationships, which are intrinsic characteristics of daily precipitation processes. In contrast, ensemble tree-based models, which treat observations in a largely static manner, show limited capacity to generalize as the forecasting horizon increases. Residual analysis further indicates that LSTM predictions are largely unbiased for light to moderate rainfall events, which dominate the rainfall regime in the Marahoué basin. However, increased dispersion and systematic underestimation are observed for high-intensity precipitation, reflecting structural limitations of deep learning models when confronted with rare and highly skewed extreme events. Despite these limitations, the overall findings confirm that LSTM-based models constitute a robust and reliable foundation for rainfall forecasting and early warning systems in data-scarce and flood-prone regions such as the Marahoué basin. Their strong performance at short and medium lead times is particularly relevant for operational flood preparedness and risk mitigation. Future work should focus on improving the representation of extreme rainfall events through the integration of additional atmospheric and climatic predictors, such as Convective Available Potential Energy (CAPE), Convective Inhibition (CIN), total column water vapor, radar reflectivity, infrared brightness temperature, and regional sea surface temperatures, particularly over the Atlantic Ocean, the Gulf of Guinea, and ENSO-related regions (e.g., Niño 3.4). Furthermore, the application of advanced hyperparameter optimization techniques and hybrid modeling strategies could further enhance long-term forecast reliability and support the development of more effective early warning systems in West Africa.

Code and data availability. The Python implementations of the models used in this study are available from <https://doi.org/10.5281/zenodo.19669875> (Kamenan, 2026).

Author contributions. SJRK designed the study, developed and implemented the prediction models, conducted numerical simulations, performed data preprocessing and curation, prepared the figures, carried out the formal analysis, and wrote the original draft of the manuscript. TMY, MGA, SIS, and AMK supervised the research, contributed to the interpretation and discussion of the results, and critically reviewed and edited the manuscript. All authors discussed the results and approved the final version of the manuscript.

Competing interests. The contact author has declared that none of the authors has any competing interests.

Disclaimer. Publisher's note: Copernicus Publications remains neutral with regard to jurisdictional claims made in the text, published maps, institutional affiliations, or any other geographical representation in this paper. The authors bear the ultimate responsibility for providing appropriate place names. Views expressed in the text are those of the authors and do not necessarily reflect the views of the publisher.

Special issue statement. This article is part of the special issue "Artificial intelligence and machine learning in climate and weather science research". It is not associated with a conference.

Acknowledgements. The authors would like to express their sincere gratitude to SODEXAM as well as to the local authorities of Bouaflé and Zuénoula for providing the various datasets used in this study.

Review statement. This paper was edited by Lyndsay Shand and reviewed by two anonymous referees.

References

- Alavoine, N., Laperrière, G., Servan, C., Ghannay, S., and Rosset, S.: New semantic task for the French speech understanding corpus MEDIA, in: Proceedings of JEP-TALN-RECITAL 2024, 470–480, <https://aclanthology.org/2024.jeptalnrecital-jep.48/> (last access: 10 May 2025), 2024 (in French).
- Assoko, A. S. V.: Design of hydrological extremes forecasting tools in tropical West Africa: the case of the Marahoué River Basin in Côte d'Ivoire, PhD thesis, Félix Houphouët-Boigny National Polytechnic Institute, Yamoussoukro, Côte d'Ivoire, 240 pp., <https://www.edp.inphb.ci/ecoleDocPolytech/uploads/publications/publication%20these/umri%2068/ASSOKO%20Adoa%20Victoire.pdf> (last access: 10 August 2024), 2022 (in French).
- Baste, S., Klotz, D., Acuña Espinoza, E., Bardossy, A., and Loritz, R.: Unveiling the limits of deep learning models in hydrological extrapolation tasks, *Hydrol. Earth Syst. Sci.*, 29, 5871–5891, <https://doi.org/10.5194/hess-29-5871-2025>, 2025.
- Bodian, A.: Rainfall–runoff modeling approach for regional knowledge of water resources: application to the upper Senegal River Basin, PhD thesis, Cheikh Anta Diop University, Dakar, Senegal, 288 pp., <https://doi.org/10.13140/RG.2.2.30749.13289>, 2011 (in French).
- Bodian, A., Dezetter, A., and Dacosta, H.: Contribution of rainfall–runoff modeling to the understanding of water resources: application to the upper Senegal River Basin, *Climatologie*, 9, 109–125, <https://doi.org/10.4267/climatologie.223>, 2012 (in French).
- Chicco, D., Warrens, M. J., and Jurman, G.: The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE and RMSE in regression analysis evaluation, *PeerJ Comput. Sci.*, 7, e623, <https://doi.org/10.7717/peerj-cs.623>, 2021.
- Denis, G.: Artisanal gold mining in Côte d'Ivoire: the persistence of an illegal activity, *Eur. Sci. J.*, 12, 18–36, <https://doi.org/10.19044/esj.2016.v12n3p18>, 2016 (in French).
- Dje Bi, D. D.: Assessment and planning of water resources in the Marahoué River Basin (center-west Côte d'Ivoire), Master's thesis, Université Nangui Abrogoua, 83 pp., <https://www.scrip.org/reference/referencespapers?referenceid=3037461> (last access: 12 January 2022), 2015 (in French).
- Durre, I., Menne, M. J., Gleason, B. E., Houston, T. G., and Vose, R. S.: Comprehensive automated quality assurance of daily surface observations, *J. Appl. Meteorol. Clim.*, 49, 1615–1633, <https://doi.org/10.1175/2010JAMC2375.1>, 2010.
- Endalie, D., Haile, G., and Taye, W.: Deep learning model for daily rainfall prediction: case study of Jimma, Ethiopia, *Water Suppl.*, 22, 3448–3461, <https://doi.org/10.2166/ws.2021.391>, 2022.
- Fouotsa Manfouo, N. C., Potgieter, L., Watson, A., and Nel, J. H.: A comparison of the statistical downscaling and long short-term memory artificial neural network models for long-term temperature and precipitation forecasting, *Atmosphere*, 14, 708, <https://doi.org/10.3390/atmos14040708>, 2023.
- Gan, Z., Zou, F., Zeng, N., Xiong, B., Liao, L., Li, H., Luo, X., and Du, M.: Wavelet denoising algorithm based on NDOA compressed sensing for fluorescence image of microarray, *IEEE Access*, 7, 13338–13346, <https://doi.org/10.1109/ACCESS.2019.2891759>, 2019.
- Gers, F. A., Schmidhuber, J., and Cummins, F.: Learning to forget: continual prediction with LSTM, in: Proceedings of ICANN 1999, 850–855, <https://doi.org/10.1049/cp:19991218>, 1999.
- Granet, A., Morin, E., Mouchère, H., Quiniou, S., and Viard-Gaudin, C.: Neural decoder for the transcription of historical handwritten documents, in: Proceedings of JEP-TALN-RECITAL 2018, 183–195, <https://aclanthology.org/2018.jeptalnrecital-long.14/> (last access: 12 January 2022), 2018 (in French).
- Hakizimana, J. C.: Development of a simplified rainfall prediction model in Burundi, Master's thesis, Africa Center of Excellence for Water and Sanitation (C2EA), Abomey, Benin, 67 pp., <https://doi.org/10.13140/RG.2.2.18370.59841>, 2023.
- Hao, R. and Bai, Z.: Comparative study for daily streamflow simulation with different machine learning methods, *Water*, 15, 1179, <https://doi.org/10.3390/w15061179>, 2023.
- Hayder, I. M., Al-Amiedy, T. A., Ghaban, W., Saeed, F., Nasser, M., Al-Ali, G. A., and Younis, H. A.: An intelligent early flood forecasting and prediction leveraging machine and deep learning algorithms with advanced alert system, *Processes*, 11, 481, <https://doi.org/10.3390/pr11020481>, 2023.
- Hochreiter, S. and Schmidhuber, J.: Long short-term memory, *Neural Comput.*, 9, 1735–1780, <https://doi.org/10.1162/neco.1997.9.8.1735>, 1997.
- Hussain, T., Anothai, J., Nualsri, C., Ata-Ul-Karim, S. T., Duangpan, S., Hussain, N., and Ali, A.: Assessment of CSM–CERES–Rice as a decision support tool in the identification of high-yielding drought-tolerant upland rice genotypes, *Agronomy*, 13, 423, <https://doi.org/10.3390/agronomy13020432>, 2023.
- Jingwei, H., Wang, Y., Zhou, J., and Tian, Q.: Prediction of hourly air temperature based on CNN–

- LSTM, *Geomat. Nat. Haz. Risk*, 13, 1962–1986, <https://doi.org/10.1080/19475705.2022.2102942>, 2022.
- Kamenan, S. J. R.: KAMENAN/rainfall-prediction-deep-learning-civ: First release – Rainfall prediction model (v1.0), Zenodo [code, data set], <https://doi.org/10.5281/zenodo.19669875>, 2026.
- Kassam, S.: The mean-absolute-error criterion for quantization, in: *Proceedings of ICASSP 1977*, 632–635, <https://doi.org/10.1109/ICASSP.1977.1170242>, 1977.
- Kouadio, A. C., Kouassi, K., and Assi-Kaudjhis, J. P.: Artisanal gold mining, food availability, and land competition in the gold-producing areas of the Bouaflé department, *Tropicultura*, 36, 369–379, 2018 (in French).
- Kouamé, K. A., Koudou, A., Sorokoby, V. M., Kouamé, K. F., and Kouassi, A. M.: Relationship between surface and groundwater flows in the upper Bandama River Basin, Côte d’Ivoire, *Larhyss J.*, 29, 137–152, 2017 (in French).
- Le, X.-H., Ho, H. V., Lee, G., and Jung, S.: Application of long short-term memory neural network for flood forecasting, *Water*, 11, 1387, <https://doi.org/10.3390/w11071387>, 2019.
- Lek, S., Dimopoulos, I., Derraz, M., and El Ghachtoul, Y.: Modeling the rainfall–runoff relationship using artificial neural networks, *Rev. Sci. Eau*, 9, 319–331, 1996 (in French).
- Loague, K. and Green, R. E.: Statistical and graphical methods for evaluating solute transport models: overview and application, *J. Contam. Hydrol.*, 7, 51–73, [https://doi.org/10.1016/0169-7722\(91\)90038-3](https://doi.org/10.1016/0169-7722(91)90038-3), 1991.
- L’Hôte, Y.: Measurement and study of precipitation in hydrology, University of Montpellier II and ORSTOM, https://horizon.documentation.ird.fr/exl-doc/pleins_textes/divers11-08/010020119.pdf (last access: 18 April 2023), 1993 (in French).
- Molod, A., Takacs, L., Suarez, M., and Bacmeister, J.: Development of the GEOS-5 atmospheric general circulation model: evolution from MERRA to MERRA2, *Geosci. Model Dev.*, 8, 1339–1356, <https://doi.org/10.5194/gmd-8-1339-2015>, 2015.
- Pearson, K.: Determination of the coefficient of correlation, *Science*, 30, 23–25, <https://doi.org/10.1126/science.30.757.23>, 1909.
- Peltre-Wurtz, J. and Steck, B.: Influence of a development company on the rural environment: cotton and draft-based farming in the Bagoué region (Northern Côte d’Ivoire), ORSTOM, https://horizon.documentation.ird.fr/exl-doc/pleins_textes/pleins_textes_6/b_fdi_35-36/39896.pdf (last access: 12 March 2023), 1979 (in French).
- Renteria-Mena, J. B., Plaza, D., and Giraldo, E.: Multivariate hydrological modeling based on long short-term memory networks for water level forecasting, *Information*, 15, 358, <https://doi.org/10.3390/info15060358>, 2024.
- Sagna, D.: Rainfall modeling using artificial intelligence or machine learning in Casamance, Master’s thesis, Université Assane Seck de Ziguinchor, Sénégal, 57 pp., <http://rivieresdusud.uasz.sn/xmlui/handle/123456789/1283> (last access: 9 August 2024), 2021 (in French).
- Sham, F. A. F., El-Shafie, A., Jaafar, W. Z. B. W., Adarsh, S., Sherif, M., and Ahmed, A. N.: Improving rainfall forecasting using deep learning data fusing model approach for observed and climate change data, *Sci. Rep.*, 15, 27872, <https://doi.org/10.1038/s41598-025-13567-2>, 2025.
- Simanjuntak, F., Jamaluddin, I., Lin, T.-H., Siahaan, H. A. W., and Chen, Y.-N.: Rainfall forecast using machine learning with high spatiotemporal satellite imagery every 10 minutes, *Remote Sens.*, 14, 5950, <https://doi.org/10.3390/rs14235950>, 2022.
- Singh, P. and Borah, B.: Indian summer monsoon rainfall prediction using artificial neural network, *Stoch. Env. Res. Risk A.*, 27, 1585–1599, 2013.
- Tagini, B.: Structural outline of Côte d’Ivoire, *Essai de géotectonique régionale*, SODEMI, Abidjan, 302 pp., https://documentation-beauvais.unilasalle.fr/index.php?lvl=notice_display&id=9673 (last access: 14 April 2020), 1971 (in French).
- Tukey, J. W.: *Exploratory Data Analysis*, Addison-Wesley, Reading, MA, <https://doi.org/10.1002/bimj.4710230408>, 1977.
- Waqas, M., Humphries, U., Hlaing, P. W., Wangwongchai, A., and Dechpichai, P.: Advancements in daily precipitation forecasting: a deep dive into hybrid methods in the tropical climate of Thailand, *MethodsX*, 12, 102757, <https://doi.org/10.1016/j.mex.2024.102757>, 2024.
- Willmott, C. J.: On the validation of models, *Phys. Geogr.*, 2, 184–194, <https://doi.org/10.1080/02723646.1981.10642213>, 1981.
- Wright, S.: Correlation and causation, *J. Agric. Res.*, 20, 557–585, 1921.
- Wu, C. L. and Chau, K. W.: Prediction of rainfall time series using modular soft computing methods, *Eng. Appl. Artif. Intel.*, 26, 997–1007, <https://doi.org/10.1016/j.engappai.2012.05.023>, 2013.
- Xiang, C., Han, G., Zheng, Y., Ma, X., and Gong, W.: Improvement of CO2-DIAL signal-to-noise ratio using lifting wavelet transform, *Sensors*, 18, 2362, <https://doi.org/10.3390/s18072362>, 2018.
- Xu, W.: Stock price prediction based on CNN–LSTM model in the PyTorch environment, in: *Proceedings of the Conference*, 1272–1276, https://doi.org/10.2991/978-94-6463-036-7_188, 2022.
- Zhao, J. and Obonyo, E.: Convolutional long short-term memory model for recognizing construction workers’ postures from wearable inertial measurement units, *Adv. Eng. Inform.*, 46, 101177, <https://doi.org/10.1016/j.aei.2020.101177>, 2020.
- Zubieta, R., Getirana, A., Espinoza, J. C., Lavado-Casimiro, W., and Aragon, L.: Hydrological modeling of the Peruvian–Ecuadorian Amazon Basin using GPM-IMERG satellite-based precipitation dataset, *Hydrol. Earth Syst. Sci.*, 21, 3543–3555, <https://doi.org/10.5194/hess-21-3543-2017>, 2017.