EEET ECOLOGICAL ENGINEERING
& ENVIRONMENTAL TECHNOLOGY

Ecological Engineering & Environmental Technology 2023, 24(9), 271–290
https://doi.org/10.12912/27197050/174146
ISSN 2719-7050, License CC-BY 4.0

Received: 2023.10.03
Accepted: 2023.10.17
Published: 2023.11.15

Deep Learning Approach for Runoff Prediction – Evaluating the Long-Short-Term Memory Neural Network Architectures for Capturing Extreme Discharge Events in the Ouergha Basin, Morocco

Nourelhouda Karmouda[1*], Tarik Bouramtane[1], Mounia Tahiri[1], Ilias Kacimi[1], Marc Leblanc[1,2], Nadia Kassou[1]

[1] Geosciences, Water and Environment Laboratory, Faculty of Sciences, Mohammed V University in Rabat, Avenue Ibn Batouta, Rabat, Morocco

[2] Hydrogeology Laboratory, UMR EMMAH, University of Avignon, Avignon, France

* Corresponding author's e-mail: nourelhouda_karmouda@um5.ac.ma

ABSTRACT

Rainfall-runoff modeling plays a crucial role in achieving efficient water resource management and flood forecasting, particularly in the context of increasing intensity and frequency of extreme meteorological events induced by climate change. Therefore, the aim of this research is to assess the accuracy of the Long-Short-Term Memory (LSTM) neural networks and the impact of its architecture in predicting runoff, with a particular focus on capturing extreme hydrological discharges in the Ouergha basin; a Moroccan Mediterranean basin with historical implications in many cases of flooding; using solely daily rainfall and runoff data for training. For this purpose, three LSTM models of different depths were constructed, namely LSTM 1 single-layer, LSTM 2 bi-layer, and LSTM 3 tri-layer, their window size and hyperparameters were first tuned, and on seven years of daily data they were trained, then validated and tested on two separate years to ensure the generalization on unseen data. The performance of the three models was compared using hydrogram-plots, Scatter-plots, Taylor diagrams, and several statistical metrics. The results indicate that the single-layer LSTM 1 outperforms the other models, it consistently achieves higher overall performance on the training, validation, and testing periods with a coefficient of determination R-squared of 0.92, 0.97, and 0.95 respectively; and with Nash-Sutcliffe efficiency metric of 0.91, 0.94 and 0.94 respectively, challenging the conventional beliefs about the direct link between complexity and effectiveness. Furthermore, all the models are capable of capturing the extreme discharges, although, with a moderate underprediction trend for LSTM 1 and 2 as it does not exceed -25% during the test period. For LSTM 3, even if its underestimation is less pronounced, its increased error rate reduces the confidence in its performance. This study highlights the importance of aligning model complexity with data specifications and suggests the necessity of considering unaccounted factors like upstream dam releases to enhance the efficiency in capturing the peaks of extreme events.

Keywords: Morocco, hydrological modeling, deep neural network, runoff prediction, Ouergha, extreme events, long-short-term memory (LSTM).

INTRODUCTION

Rainfall-runoff modeling is of primordial importance to ensure optimal water resource management and flood mitigation (Young and Liu, 2015). It is particularly crucial in the context of increasing intensity and frequency of extreme meteorological events induced by climate change (Clarke et al., 2022). According to Tabari et al., (2020) there is an intensification of extreme precipitation, which converges towards more extreme floods across different climatic regions. These floods are considered a global natural hazard with high costs and risks (Blöschl et al., 2019), which emphasizes the valuable need to anticipate and accurately predict extreme runoff. Nonetheless, runoff prediction remains a challenging task due to its complex and nonlinear nature, necessitating robust models and enhanced accuracy (Aqnouy et al., 2021; Man et al., 2022). Over the past decades, several hydrological models (stochastic, conceptual, or physically based models) have been usually considered foundational for rainfall-runoff simulation (Bahremand and De Smedt, 2010, 2008; Tingsanchali, 2000) However, they are subject to various challenges, including the significant requirement for diverse data types, the heterogeneity of the natural systems and the complexities in representing nonlinear dependencies, and capturing extreme events.

With the advances in computing capacity, the need for improved accuracy, and reduced complexity, machine learning techniques have emerged as a promising horizon. Thus, an increasing tendency toward deep learning has been noticed with a particular interest in Artificial Neural Networks (ANNs), and it was extensively utilized by researchers in a variety of applications (Alardhi et al., 2023; Babu et al., 2022; Bashayreh et al., 2021; Chen et al., 2023; Oni et al., 2022). In the field of hydrology, ANNs derive their strength from their adaptability and ability to perceive complex and intricate connections between the variables, which is essential for simulating the inherent complexity and non-linearity of the hydrological systems (Govindaraju and Rao, 2000; Wu and Chau, 2011). Furthermore, they have the capability not only to generalize from varied historical training data but also to produce robust predictions even under changing conditions and without requiring in-depth comprehension of the underlying hydrological dynamics (Bouramtane et al., 2023; Rajaee et al., 2019). Numerous research have previously focused on the potential of ANNs for predicting and forecasting runoff under various conditions by using different variables, time steps, and architectures; for instance, Hsu et al., (1995) highlighted the effectiveness of ANNs in modeling the rainfall-runoff dynamics of the Leaf River, against linear and conventional conceptual models. In their comparison of the Soil and Water Assessment Tool SWAT and ANN, Demirel et al., (2009) demonstrated the superior accuracy of the latter, specifically in the prediction of peak flows within the Pracana basin in Portugal. Expanding the scope, Juan et al., (2017) used ANNs to predict runoff fluctuations in the Three-River Headwater Region (TRHR) on the Qinghai-Tibet Plateau with the context of climate change, confirming the utility and validity of ANNs despite data and parameter limitations. Moreover, Xiang et al., (2020) found that the Recurrent Neural Networks model; a subset of neural networks adept at processing sequential data and recognizing temporal dependencies (Sherstinsky, 2020); outperforms regression models in estimating hourly runoff in two USA watersheds. Further, the study of Zema et al., (2020) underscored the efficiency of Multi-Layer Perceptron architecture (MLP) which is a feedforward ANN, in predicting the hydrological behaviors across diverse soil conditions in Southeast Spain, emphasizing their efficacy in controlling runoff and soil erosion, especially in fire-affected areas.

While (ANNs) have proven their ability to capture the hydrological response even without previous knowledge of the catchment's physical characteristics (Haykin, 1999), the Long-Short-Term Memory models (LSTMs) which are the improved architecture of the Recurrent Neural Network (RNN) (Hochreiter and Schmidhuber, 1997) take this ability one step further by cleverly capturing also the sequential and long-term dependencies (Yu et al., 2019). The LSTM distinct design that incorporates memory cells together with the output and forget gates (Gers et al., 2000)**,** allows data to be stored and recalled for extended periods, thereby accurately accounting for past hydrological events while simultaneously capturing current ones. This feature is particularly relevant due to the intricate temporal dynamics of hydrological processes, where the impact of prior rainfall events can manifest in river flows several days or weeks later, depending on the unique characteristics of the catchment region. However, as with any advanced modeling technique, LSTM networks are not exempt from challenges that can profoundly influence their performance, such as the identification of the best architecture, the selection of the window sizes (time step), and the tuning of the hyperparameters (neurons number, batch size, epochs, etc). Recently, many studies have been made to explore the efficiency of LSTM for hydrological purposes; Fang et al., (2017) were the first authors that have explored LSTM techniques in the field of hydrology to extend the coverage of Soil Moisturization Active Passive (SMAP) data, they have found LSTM is to be effective in removing bias, correcting moisture climatology, and capturing extremes. Besides, Both in their research Hu et al., (2018), and Mao et al., (2021) have used the ANN and LSTM for rainfall-runoff simulation, their findings indicate that the performance of LSTM surpasses that of ANN, especially in daily time steps. Furthermore, Fan et al., (2020) tested the impact of the window size on the performance of LSTM, and compared LSTM with ANN and SWAT, their results indicate that ANN and SWAT have almost the same capacity to reproduce runoff, but LSTM outperforms both of them, the reason why they suggest LSTM as the best alternative in case of catchments with a lack of topographical data.

To extend this knowledge, this study aims to assess the potential impact of the LSTM architecture on runoff prediction, with a particular focus on capturing extreme occurrences using only daily rainfall and runoff data for training. For this purpose, three LSTM models of varying depths, ranging from single-layer to multi-layer, were built. Our approach adheres to a rigorous methodology, including the adjustment of the window size, the fine-tuning of the hyperparameters using the grid-search method with the ADEM optimizer(Bergstra and Bengio, 2012; Kingma and Ba, 2017), and the splitting of the used data into training, validation, and test sets, which is essential to ensure the robustness of the models and the good generalization to unseen data and thus the prediction of extreme runoff with their best reliability.

The case study area of this paper, is the Ouergha basin, a Moroccan Mediterranean River basin. It's the principal contributor to the Al Wahda dam the largest reservoir of the country and the main regulator of the floods in the Sebou floodplain. On many occasions, the Ouergha River basin has been the major cause of downstream floods causing high material damage and human casualties as in 2009 and 2010. The choice of this region is therefore significant, as through this study we aim to make a substantial contribution to the understanding of how machine learning, specifically LSTM networks, can be utilized in predicting extreme hydrological events. The overall aim is to provide hydrologists and decision-makers with a more accurate and reliable tool for managing water resources and mitigating the risks associated with extreme weather events.

MATERIALS AND METHODS

Study area

The study area is located in central northern Morocco (Fig. 1a), between latitudes 34.379°, 35.139° North and longitudes 3.906°, 5.371° West (Fig. 1b). The Ouergha is the main tributary of the Sebou River (Combe, 1975), and the major contributor to the inflow of the Al Wahda dam. This river drains the southern side of the Rif mountains and traverses a distance of 300 km, featuring a 20 km segment upon entering the Sebou flood plain.

This study narrows its focus to the flood genesis zone upstream of the Al Wahda dam, an area that constitutes approximately 80% of the total basin and covers 6190 km^2 of lands with elevations ranging from 145 to 2450 m. It's equipped with 4 dams (Al Wahda, Asfalou, Bouhouda, and Sahla) and 11 meteorological stations, 4 of them also function as discharge gauging stations (Fig. 1c). The topography within this region is notably complex, marked by rugged terrains. Approximately 86% of the area possesses a slope greater than 12%, while only 3% features a gentler slope of 3% or lower. In addition, the basin is mainly of clay soils, Mesozoic shale, and marl formations which contribute to the watershed's low permeability. Consequently, the area exhibits limited water retention capacity, resulting in high runoff rates, and fast response. Further, Senoussi et al. (1999) affirm that the Ouergha region, which is characterized by a Mediterranean climate is the most humid area in Morocco, where rainfalls principally occur from October to April. This leads to a hydrological regime marked by high winter flow rates. Notably, between late 2008 and early 2011, the Ouergha basin experienced unusual precipitation height, resulting in discharge rates that surpassed 3000 m^3/s. According to Msatef et al. (2018), the estimated return period of discharges that exceeds 2500 m^3/s, as calculated using the Gumbel distribution, is more than 1000 years (T > 1000), assigning these events as extreme. Consequently, these exceptional runoff volumes, surpassed the storage capacities of the regional reservoirs, leading to riverbank inundations in the Sebou floodplain (Fig. 2).

Datasets

To achieve the aim of this study, we used only two variables rainfall and runoff, their time series were provided by the Direction of Research and Water Planning (DRWP) of Morocco Rainfall data were collected from 11 meteorological stations situated within the Ouergha basin (Fig. 1c), while runoff data were obtained from Al Wahda dam's hydrometric station. The daily dataset spans from January 2003 to December 2010

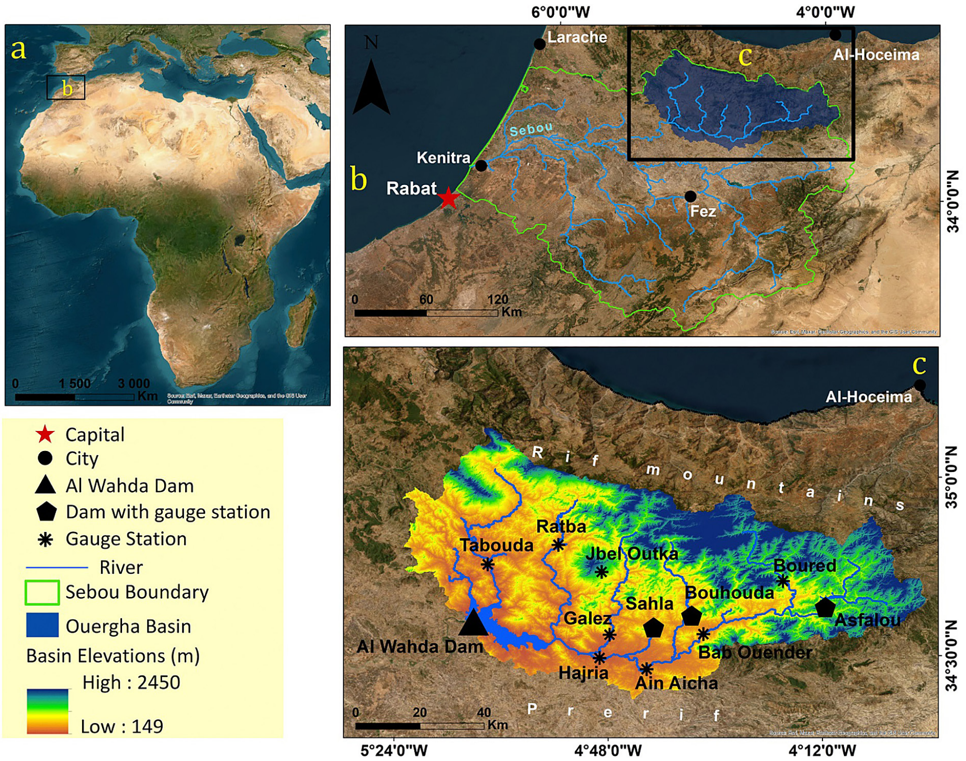
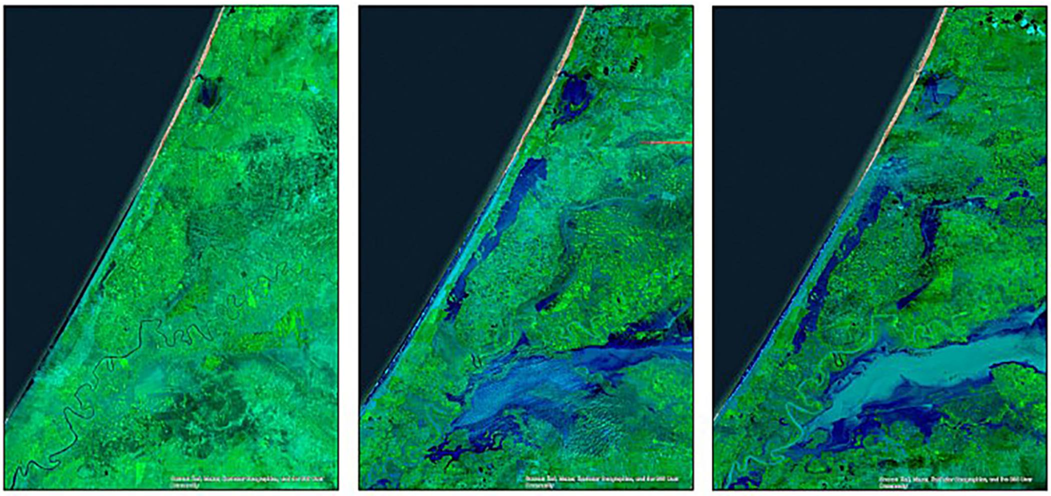


Figure 1. (a), (b) Location of the study area; (c) the study area elevation, dams, and gauge stations



December 2009 January 2010 March 2010

Figure 2. Landsat 5 TM images, false-color, tracking inundations of the Sebou floodplain in 2010

and resumes briefly from late 2012 until the end of 2013. Missing data are minimal, accounting for less than 2% at only three stations (Asfalou, Boured, and Jbel Outka) (Table 1). Depending on the data characteristics of each station, the gaps were filled using the convincible statistical imputation or spatial interpolation methods, such as linear regression method or inverse distance weight method respectively. In addition, the intra-stations and the runoff-stations linear correlations are visually represented in a heatmap (Figure 3). Generally, the precipitation measurements show strong intra-station agreement. Most stations also demonstrate a satisfactory correlation with runoff, with the exceptions of Ain Aicha, Bouhouda, and Sahla. These particular stations have weak correlations, with coefficients of 0.46, 0.39, and 0.47, respectively. To assess the statistical significance of these correlations, t-tests were employed with a significance level set at α=0.05. The resulting

Table 1. Summary, and characteristics of Ouergha's stations

N°	Station	Missing data %	Z (m)	N°	Station	Missing data %	Z (m)
1	Ain Aicha	0	250	7	Hajria	0	191
2	Asfalou	1.02	658	8	Jbel Outka	2.05	1091
3	Bab Ouender	0	392	9	Ratba	0	295
4	Bouhouda	0	487	10	Sahla	0	370
5	Boured	0.20	840	11	Tabouda	0	182
6	Galez	0	251				

p-values for all stations were lower than 0.05 and tended toward zero. This suggests that these correlations are statistically significant and unlikely to have occurred by chance. The near-zero p-values for Ain Aicha, Bouhouda, and Sahla affirm that these weak correlations are statistically validated rather than anomalous. Due to these weak yet statistically significant correlations, we opted to exclude data from Ain Aicha, Bouhouda, and Sahla to maintain the study's overall consistency. Further, the preprocessing conducted on this data is the Standardization; which is usually recommended in machine learning, and definitely needed when variables have different units. Thus, we used the z-score normalization (equation 1), it enables to rescale data to a mean of 0 and a standard deviation of 1 (Shanker et al., 1996), which enhances numerical stability and accelerates learning.

$$z = \frac{x - \mu}{\sigma} \qquad (1)$$

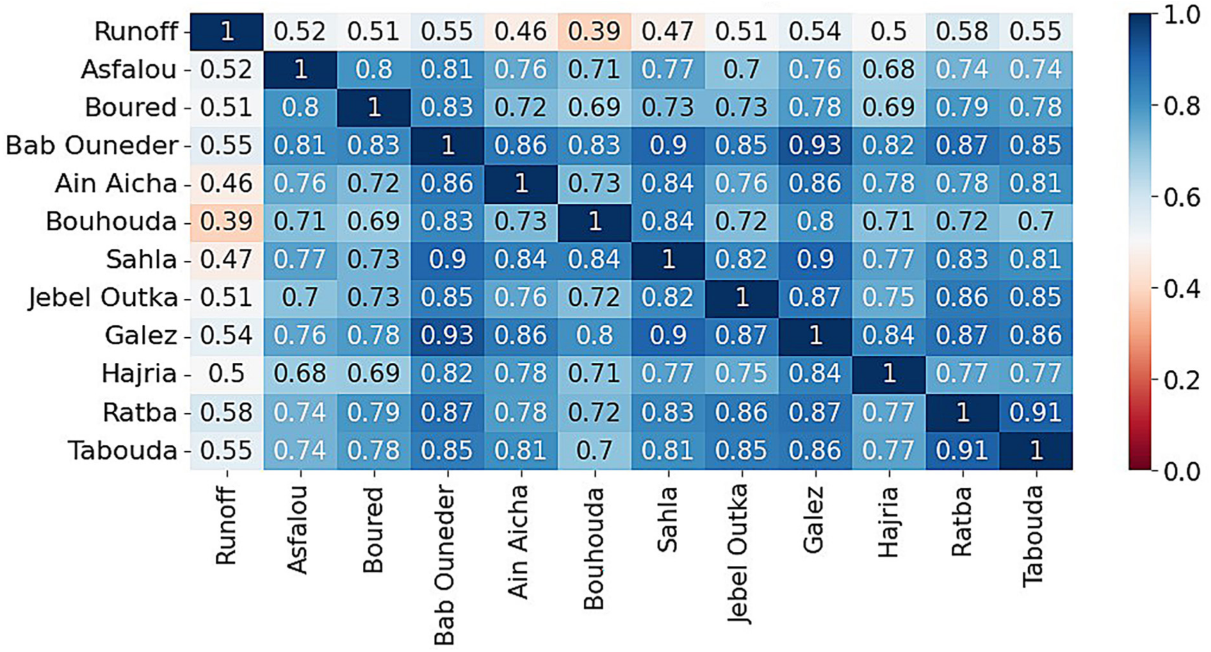
where: x – the value to be transformed to z, μ – the mean and σ – the standard deviation of the data set.

Finally, several descriptive statistics of the data sets including mean values, medians, maxima, minima, standard deviation and coefficient of variation are listed in Table 2.

Deep learning model

Long short-term memory (LSTM)

Recurrent Neural Networks(RNN), are simple neural systems that utilize loops to process sequences of inputs (Williams and Zipser, 1989, Sherstinsky, 2020). These loops facilitate the transfer of information from one layer to another and provide the RNN with a memory capability, which allows the network to store past computations and exhibit dynamic temporal behavior



	Runoff	Asfalou	Boured	Bab Ouneder	Ain Aicha	Bouhouda	Sahla	Jebel Outka	Galez	Hajria	Ratba	Tabouda
Runoff	1	0.52	0.51	0.55	0.46	0.39	0.47	0.51	0.54	0.5	0.58	0.55
Asfalou	0.52	1	0.8	0.81	0.76	0.71	0.77	0.7	0.76	0.68	0.74	0.74
Boured	0.51	0.8	1	0.83	0.72	0.69	0.73	0.73	0.78	0.69	0.79	0.78
Bab Ouneder	0.55	0.81	0.83	1	0.86	0.83	0.9	0.85	0.93	0.82	0.87	0.85
Ain Aicha	0.46	0.76	0.72	0.86	1	0.73	0.84	0.76	0.86	0.78	0.78	0.81
Bouhouda	0.39	0.71	0.69	0.83	0.73	1	0.84	0.72	0.8	0.71	0.72	0.7
Sahla	0.47	0.77	0.73	0.9	0.84	0.84	1	0.82	0.9	0.77	0.83	0.81
Jebel Outka	0.51	0.7	0.73	0.85	0.76	0.72	0.82	1	0.87	0.75	0.86	0.85
Galez	0.54	0.76	0.78	0.93	0.86	0.8	0.9	0.87	1	0.84	0.87	0.86
Hajria	0.5	0.68	0.69	0.82	0.78	0.71	0.77	0.75	0.84	1	0.77	0.77
Ratba	0.58	0.74	0.79	0.87	0.78	0.72	0.83	0.86	0.87	0.77	1	0.91
Tabouda	0.55	0.74	0.78	0.85	0.81	0.7	0.81	0.85	0.86	0.77	0.91	1

Figure 3. Heatmap of the intra-stations and the runoff-stations linear correlation

Table 2. Descriptive statistics of datasets

Data	Mean	Med	Max	Min	STDV	CV %	Mean	Med	Max	Min	STDV	CV %	Mean	Med	Max	Min	STDV	CV %
	Training (01/01/2003 - 31/12/2009)						Validation (01/01/2010 - 01/12/2010)						Test (01/12/2012 - 31/12/2013)					
Runoff	59.47	12.72	2908.73	0	164.15	276.02	269.23	34.31	3361.71	0.06	550.27	204.39	124.8	27.61	2106.66	0.21	266.71	213.7
Ain Aicha	1.18	0	72.5	0	4.47	378.5	2.5	0	67.2	0	7.5	300.18	1.58	0	53.2	0	5.67	358.47
Asfalou	1.67	0	112.2	0	6.07	363.17	3.73	0	150.8	0	12.05	323.55	2.49	0	87.1	0	8.49	341.56
Bab Ounder	1.72	0	96	0	6.13	355.7	3.97	0	120.7	0	11.99	302.11	2.46	0	83.4	0	8.39	340.44
Bouhouda	1.65	0	122.4	0	6.07	367.63	3.04	0	130.8	0	11.05	363.11	1.86	0	65.8	0	6.86	369.05
Boured	1.42	0	81	0	5.44	384.03	3.79	0	136.8	0	12.4	327.62	2.45	0	73.1	0	8.16	333
Galez	1.86	0	95.3	0	6.61	356.02	3.84	0	114.8	0	11.84	308.4	2.39	0	74.6	0	8.36	349.66
Hajria	1.81	0	78.6	0	6.58	364.42	3.66	0	102.7	0	11.29	308.62	2.12	0	64	0	7.21	339.12
Jbel Outka	3.55	0	220	0	12.4	349.79	7.07	0	179.6	0	19.81	280.26	4.92	0	130.6	0	15.5	315.22
Ratba	2.55	0	140.2	0	8.71	341.87	5.86	0	125.1	0	16.44	280.38	3.36	0	100	0	10.78	320.74
Sahla	2	0	108.8	0	6.87	343.79	3.64	0	115.3	0	12.17	334.56	1.95	0	71.3	0	7.21	369.27
Tabouda	1.54	0	79.5	0	5.54	359.95	4.17	0	123	0	12.51	300.34	2.37	0	75	0	8.18	345.12

(Sharma et al., 2022). Despite this innovative architecture, RNNs face certain challenges such as the exploding and vanishing gradient problems while learning long-term dependencies (Hochreiter, 1991; Hochreiter et al., 2001). To overcome these limitations, the LSTM architecture was conceived by (Hochreiter and Schmidhuber, 1997) as an advancement over the fundamental structure of RNNs.

LSTM networks are composed of an input layer, memory cells, and an output layer, aiming to model complex dependencies in sequential data over extended durations with enhanced efficiency and robustness. The improvement of LSTM has been progressively refined by numerous researchers (Kawakami, 2008), through the integration of three distinct gates into the memory cell: the forget gate, the input gate, and the output gate. Together, these gates interact to manage the storage and retrieval of information within the network. This enables precise control of data flow and memory usage, selectively preserving essential information across longer sequences, and thus allowing the network to disregard insignificant information (Gers et al., 2000). The technical insights, including the equations that vectorize and describe the update of the memory cells in the LSTM layer at every time-step t, are detailed (Fischer and Krauss, 2018).

LSTM setup

In this study, three models, each with varying complexities and numbers of LSTM layers, were implemented in TensorFlow using the Keras library in Python. "LSTM 1" is the simplest model, containing a single LSTM layer that functions as both input and hidden layer, followed by a dropout layer to prevent the model from overfitting. Then, a dense output layer is added to conclude the model. "LSTM 2" builds upon this architecture, incorporating two LSTM layers, two corresponding dropout layers, and two dense layers. "LSTM 3", the most complex model in this study, adds an additional LSTM and dense layer to the structure of LSTM 2. Further details regarding the models' parameters, such as the number of units and the dropout rate, are presented in the hyperparameters tuning section.

Tuning procedure

Window size tuning

In hydrological modeling using LSTM, the window size or time step is crucial for runoff prediction and extreme event identification (Gao et al., 2020). It defines the range of historical data considered by the algorithm and influences the model's ability to capture temporal patterns in the data (Liu et al., 2021).

In this study, we optimized the window size for LSTM 1,2 and 3 models to achieve concise runoff simulations. To address time constraints and computational limitations, we fixed other hyperparameters, such as units, epochs, batch size, and dropout rate. Subsequently, specific window sizes ranging from 1 to 30 days were evaluated. Window sizes of 1–5 days capture short-term data fluctuations, such as daily weather changes and storms, while 10–25 days reveal underlying runoff trends. A 30-day window identifies monthly trends. The optimal window size with the best model performance will define which variability impacts the most the hydrological regime of the Ouergha basin and then will be selected for the following phases of this study.

Hyperparameters tuning

Training an LSTM network involves managing both learnable parameters and hyperparameters. The learnable parameters are updated during training, based on a specific loss function like MSE and using back-propagation (Yin et al., 2022). While hyperparameters, such as units (Hidden Neurons), batch size (number of samples in each training iteration), epochs (the total number of training cycles), and dropout rate (a mechanism to prevent overfitting), are either fixed by users or determined through hyperparameter tuning.

According to Goodfellow et al., 2016, hyperparameter tuning can be defined as the systematic search for the optimal set of hyperparameters that results in the minimization of a predefined loss function on a specific dataset, thereby enhancing the efficacy of the model in question. Therefore, to achieve the optimal performance of the LSTM models in predicting hydrological patterns, we employed a gridsearchCV methodology, which is a commonly used approach for hyperparameters optimization (Bergstra and Bengio, 2012) over four iterative rounds of tuning. Each round involved adjusting four key hyperparameters: number of units, batch size, number of epochs, and dropout rate. In the first round, an extensive grid was set up to explore a wide range of hyperparameter combinations. Based on the performance metrics, subsequent rounds incrementally narrowed the grid around the most promising values. For instance, the initial round tested units ranging from 100 to 1000, batch sizes from 64 to 365, epochs from 15 to 60, and dropout rates from 0.1 to 0.4. The best-performing combination guided the grid settings for the next round. The final round of tuning identified an optimal set of hyperparameters for each variant of the LSTM models. This grid-based, iterative approach enabled the fine-tuning of the models to accurately capture the complexities intrinsic to hydrological processes.

Several additional details about the model configurations are noteworthy: To guarantee reproducibility, random seeds were fixed across all experiments. A fixed learning rate of 0.01 was used, as we observed that its variation had no significant impact on model performance in our specific case. Optimization and training were performed using the ADAM optimizer, which is a version of the stochastic gradient method (Kingma and Ba, 2017), together with a Mean Squared Error (MSE) as a loss function.

Study procedure

In machine learning, the data is divided into three sets: training, validation, and test sets. The training set is the sample of data used to fit the model by adjusting its internal parameters, and the validation set fine-tunes the model hyperparameters and gives an initial assessment of performance (Ripley, 1996). It uses a data set with known samples that the model has not been trained on (Xu and Goodacre, 2018). In the past, it was widely assumed that the evaluation of a model's performance based on validation results provided an unbiased measure of overall efficiency. However, researches including the study of (Westerhuis et al., 2008)**,** has challenged this assumption, suggesting that it may not always be accurate. (Harrington, 2017) has similarly demonstrated that splitting data into training and validation sets (considered in this context as both validation and test sets) could lead to an imprecise evaluation of a model's robustness (Xu and Goodacre, 2018). These findings underscore the necessity of employing an independent blind test set, never previously used in either training or validation periods (Alpaydin, 2010), to avoid any potentially biased assessment of the model's performance. In this paper, the dataset is divided as follows: 78% for training, 11% for validation, and 11% for testing. These segments correspond to the periods from 1/1/2003 to 12/31/2009, 1/1/2010 to 12/ 31/2010, and from 12/1/2012 to 11/30/2013, respectively.

Evaluation metrics

The model's ability to accurately predict the watershed's hydrological behavior under both ordinary and extreme conditions is evaluated using the metrics listed in Table 3; the performance rating of each of them is listed in Table 4. The coefficient of determination R^2 assesses the degree of agreement between simulated and observed runoff (Moriasi et al., 2007). The Root Mean Square Error-observations standard deviation ratio (RSR), is calculated as the ratio of the RMSE and standard deviation of measured data to qualify what is considered to be a low RMSE (Singh et al., 2004). The Percent bias (PBIAS), is a metric that gauges the model's tendency to either over-predict or under-predict observed values (Gupta et al., 1999). The Nash-Sutcliffe Efficiency (NSE) is a widely used metric in hydrology, often serving as an objective function. According to Servat and Dezetter (1991), it provides a holistic measure of a hydrograph's goodness of fit. A model is considered robust when the predicted runoff is closely aligned with observed data, leading to an NSE value approaching 1 (Kouassi et al., 2013).

RESULTS

Tuning results

Window size variation results

As presented in the methodology section, we examined the impact of various time window sizes (1, 2, 3, 4, 5, 10, 15, 20, 25, 30) days, on each LSTM model results. The models' performance in reproducing runoff was assessed using RMSE and NSE metrics. The results of this tuning are presented in Figure 4. It is worth noting that the RMSE metric is considerably higher in the validation than in the training sets, this disparity must be attributed to the varying means between the two sets ($\bar{Q}_{\text{training}}$ = 59.47 m^3/s and $\bar{Q}_{\text{validation}}$ = 269.23 m^3/s), which is caused by the extreme flow rates present in the validation set. During the training, all models displayed optimal performance with one-day window size, recording the lowest RMSE of 58.14 m^3/s, and achieving a high NSE efficiency of 0.87. However, as the window extended from one to five days, a trend emerged: barring a minor spike at the 4-day window for LSTM 3, there was a general decline in models' performance. Beyond the 5-day window, the RMSE plateaued for all models, the NSE continued to decrease achieving its lowest rate 0.73 in the 25-day window size for LSTM 1 and 2, and 0.71 in the 20-day window size for LSTM 3. In the validation, the superior performance of the

Table 3. List of the statistical metrics (Moriasi et al., 2015)

Statistical metric	Equation	Value range	Perfect value
Coefficient of determination (R^2)	$R^2 = \left(\frac{\sum_{i=1}^{n}(Q_o - \overline{Q_o})(Q_s - \overline{Q_s})}{\sqrt{\sum_{i=1}^{n}(Q_o - \overline{Q_o})^2}\ \sqrt{\sum_{i=1}^{n}(Q_s - \overline{Q_s})^2}}\right)^2$	$[0, 1]$	1
Root mean square error (RMSE)	$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(Q_s - Q_0)^2}{n}}$	$[0, +\infty]$	0
RMSE observations standard deviation ratio (RSR)	$RSR = \frac{\sqrt{\sum_{i=1}^{n}(Q_o - Q_s)^2}}{\sqrt{\sum_{i=1}^{n}(Q_o - \overline{Q_s})^2}}$	$[0, +\infty]$	0
Percent bias (PBIAS)	$PBIAS = \frac{\sum_{i=1}^{n}(Q_0 - Q_s)}{\sum_{i=1}^{n} Q_o} \times 100$	$[-\infty, +\infty]$	0
Nash-Sutcliffe (NSE)	$NSE = 1 - \left[\frac{\sum_{i=1}^{n}(Q_o - Q_s)^2}{\sum_{i=1}^{n}(Q_o - \overline{Q_o})^2}\right]$	$[-\infty, 1]$	1

Note: *Qo* – the observed flow, the *Qs* – the predicted flow, *n* – the total number of the observation.

Table 4. Performance measurements for stream flow simulation (Moriasi et al., 2015, 2007)

Statistical metric	Unsatisfactory	Satisfactory	Good	Very good
R^2	R^2 <0.50	0.50 ≤ R^2<0.70	0.70≤R^2<0.80	≥0.80
RSR	RSR > 0.7	0.6 < RSR ≤ 0.7	0.5 < RSR ≤ 0.6	0 < RSR ≤ 0.5
PBIAS	PBIAS ≥ ±25	±10 ≤ PBIAS < ±25	±5 ≤ PBIAS < ±10	PBIAS < ±5
NSE	NSE ≤ 0.5	0.5 < NSE ≤ 0.60	0.60 < NSE ≤ 0.80	0.80 < NSE ≤ 1

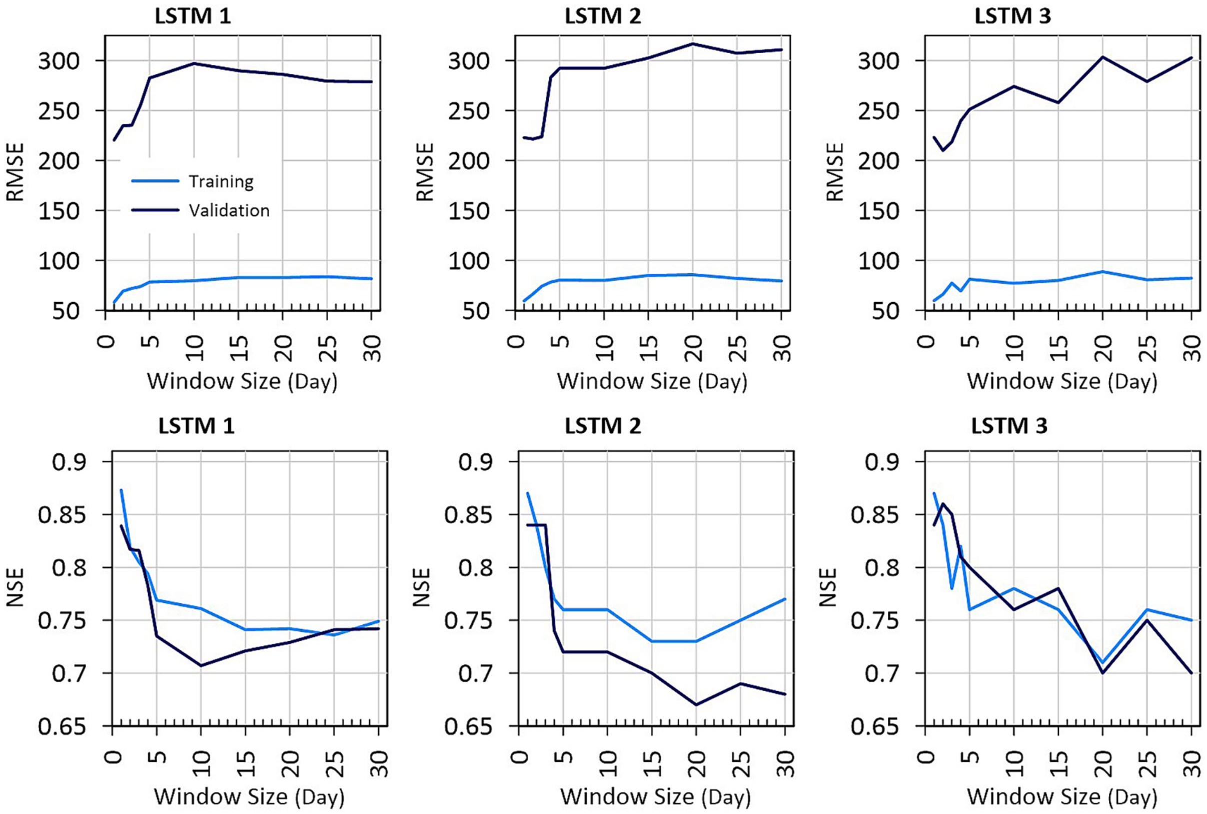


Figure 4. Comparison of RMSE and NSE for window size impact on training and validation (LSTM 1, 2, 3)

one-day window for LSTM 1 is reaffirmed with an NSE of 0.84 and the lowest error rates of 220 m^3/s. These metrics inversely evolve in the subsequent windows as RMSE increases and NSE decreases. For LSTM 2, the initial three window sizes (1, 2, and 3 days) emerge as optimal; all boasting an NSE of 0.84 and minimized RMSE of 222.86, 221.44, and 223.77 m^3/s respectively. In contrast, the validation for LSTM 3 reveals that the two-day window has the best NSE of 0.86 and the lowest validation error of 209.99 m^3/s.

Hyperparameters tuning results

Each LSTM model underwent four rounds of hyperparameter tuning. Table 5 illustrates the

Table 5. Hyperparameter tuning rounds and results for LSTM Models 1, 2, and 3

Data	Tunned values					Results			
	Round	Units	Batch size	Epochs	Dropout	Units	Batch	Epochs	Dropout
LSTM 1	1	100, 500, 1000	64,128,365	15,30,60	0.1 to 0.4	1000	365	15	0.3
	2	800,1000,1200	128,365	10, 15, 20	0.1 to 0.4	1200	365	15	0.4
	3	1150,1200, 1250	365	14,15,16	0.1 to 0.4	1250	365	14	0.3
	4	1240,1250,1260	365	14,15	0.1 to 0.4	1260	365	14	0.3
LSTM 2	1	100, 500, 1000	64,128,365	15,30,60	0.1 to 0.4	1000	365	30	0.4
	2	800,1000,1200	125, 365	25,30,35	0.1 to 0.4	800	365	30	0.4
	3	700, 800, 900	365	30	0.1 to 0.4	800	365	30	0.4
	4	790, 800, 810	365	30	0.1 to 0.4	800	365	30	0.4
LSTM 3	1	100, 500, 1000	64,128,365	16,32,64	0.1 to 0.4	100	128	32	0.2
	2	100,150,200	128, 256	22,32,42	0.1 to 0.4	150	128	22	0.3
	3	125,150,175	128	22, 32	0.1 to 0.4	150	128	22	0.4
	4	100,150	64,128	22,32	0.1 to 0.4	150	128	22	0.1

different sets of hyperparameters used during the tuning rounds for each LSTM model. It illustrates also the hyperparameters optimal configurations fixed by GridsearchCV using the ADEM optimizer. In the optimization process, each LSTM model exhibited distinct preferences for the number of units. LSTM 1, being a single-layer model, tended to prefer around 1260 units. The two-layer LSTM 2 found an optimal configuration with approximately 800 units per layer. The LSTM 3, with its three-layer architecture, consistently opted for a lower unit count of 150. For the batch size, LSTM 1 and 2 constantly chose the larger size of 365. Furthermore, LSTM 3 favored a smaller batch size of 128. Regarding epochs tuning, LSTM 1 has stabilized at 14 epochs beyond this, there might not be a significant improvement or there could be a risk of overfitting. LSTM 2 stabilized on 30 epochs, and for LSTM 3 the 22 epochs were found to be optimal. Finally, for the dropout, LSTM 1 and 2 often opted for higher values of 0.3 or 0.4 but LSTM 3 showed a broader optimal range from 0.1 to 0.4. Based on these findings, it becomes evident that as the rounds of tuning progress, there is a clear pattern of convergence on specific hyperparameters. This implies that the tuning process is efficiently narrowing down to the optimal configuration. We relied on the results of the fourth round and set those optimized hyperparameters for the training, validation, and testing periods of the models. The subsequent section provides the assessment of the models' results for each dataset.

The performance of the models

Training period

Evaluating the training performance of a model is of great importance. Good results during this period are an indicator that the model has effectively captured underlying patterns in the training period. Hence, if a model cannot achieve a good

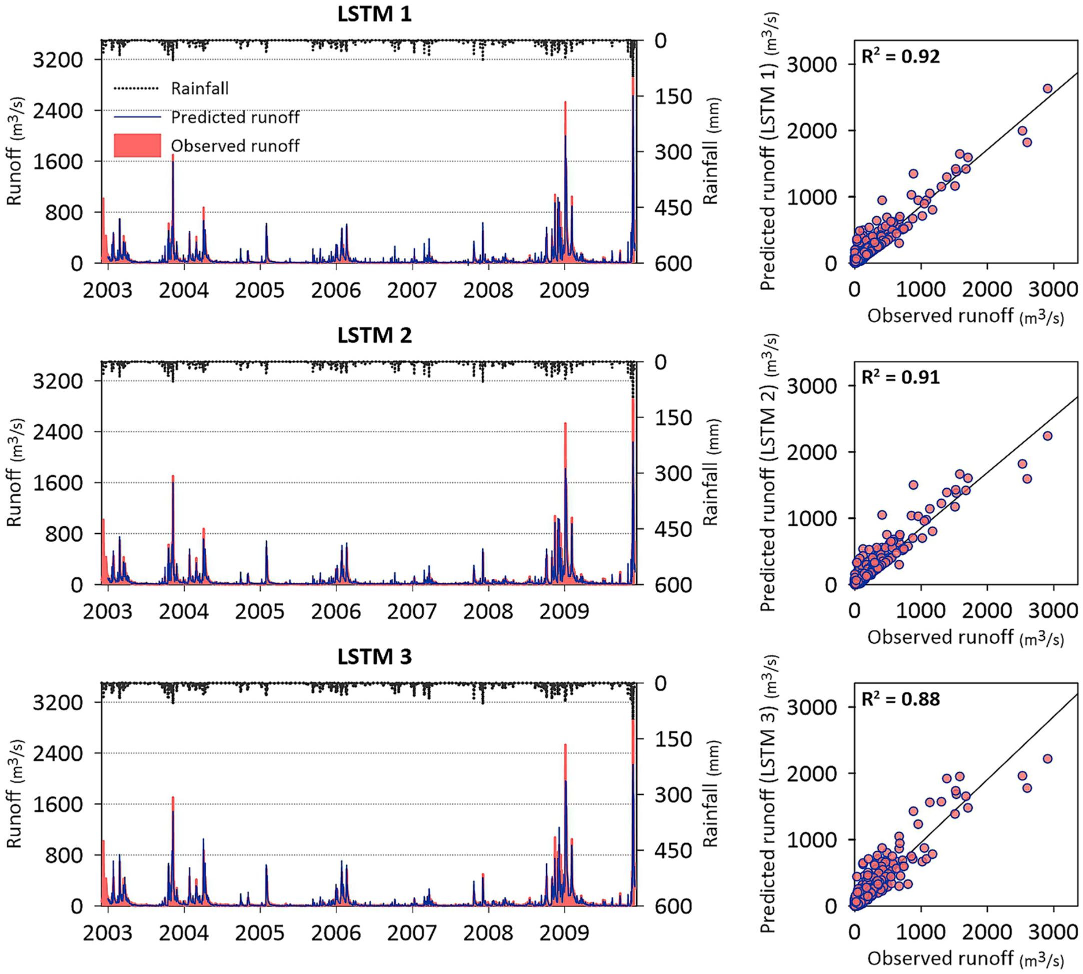


Figure 5. Graphical comparison of observed and predicted runoff by LSTM models 1, 2, and 3 during the training period

performance, its capability on unseen data becomes questionable. Based on this principle, this section presents an evaluation of the training results using graphical analysis and metrics detailed previously. According to Figure 5, the overall pattern depicts that the three models have a reasonable understanding of the system dynamics. It's clear that the predicted runoff ranges from low flow or base flow conditions to very high values indicating intense to extreme discharge episodes.However, differences in their performances emerge at peaks of significant anomalies. There's an evident trend across the models to underestimate the most pronounced anomalies, specifically the discharges of the year 2009. The underestimation rate does not only vary from one model to another but also varies with each anomaly. In terms of correlation, while all models display excellent performance, LSTM 1 stands out with a determination coefficient R^2 of 0.92, followed by LSTM 2 and LSTM 3, with R^2 of 0.91 and 0.88 respectively. All the statistical metrics from the training period for each model are graphically presented in Figure 6.

The results show that LSTM 1 emerges as the most accurate model, capturing 92% of the variability in observed discharge and registering the lowest error ratio of 0.3, equivalent to 48 m^3/s. Although LSTM 2 and 3 also perform proficiently but slightly lower than LSTM 1, with R^2 values of 0.91 and 0.88, respectively. Furthermore, all the studied models exhibit excellent NSE scores that closely reflect their R^2 values, as well as RSR indices below the 0.5 thresholds are considered excellent. However, in terms of PBIAS, LSTM 1 and 2 show minor overestimations of 2.87% and 2.7 %, respectively, whereas LSTM 3 has a more substantial overestimation of 7.28 %. When it comes to capturing extreme runoff anomalies exceeding 2000 m^3/s, all models generally tend to an underestimation pattern, whereas LSTM 2 is beyond the satisfactory threshold since the underestimation is -29% (Fig. 6b). All these result metrics confirm that LSTM 1 consistently outperforms the other two models during the training period. The subsequent section will extend this assessment to the validation period.

Validation period

Reliable performance in the validation period, particularly in accurately reproducing extreme runoff events, not only confirms the training results but also serves as an initial assessment of the model's generalization ability, setting the period for the final evaluation in the testing period. For this purpose, the graphical fit and correlation between observed and predicted runoff by LSTM models 1, 2, and 3 during the validation period are exhibited in Figure 7.

The graphs indicate that the predictions from the three models exhibit a close alignment with the actual runoff data, particularly during periods of low runoff, demonstrating their good performance under conditions of watershed stability or predictable behavior. In addition, the models capture the seasonal trends in the data with varying degrees of accuracy. Furthermore, the three models generally capture the trend during high runoff periods, but they often fail to accurately capture the peaks of extreme anomalies. This limitation in the performances of the three models is more graphically evident during the months of January, March, and December 2010 (Fig. 7).

The statistical comparisons of these findings are displayed in Figure 8. LSTM 1 and LSTM 2 notably excel with high R^2 values of 0.97 and

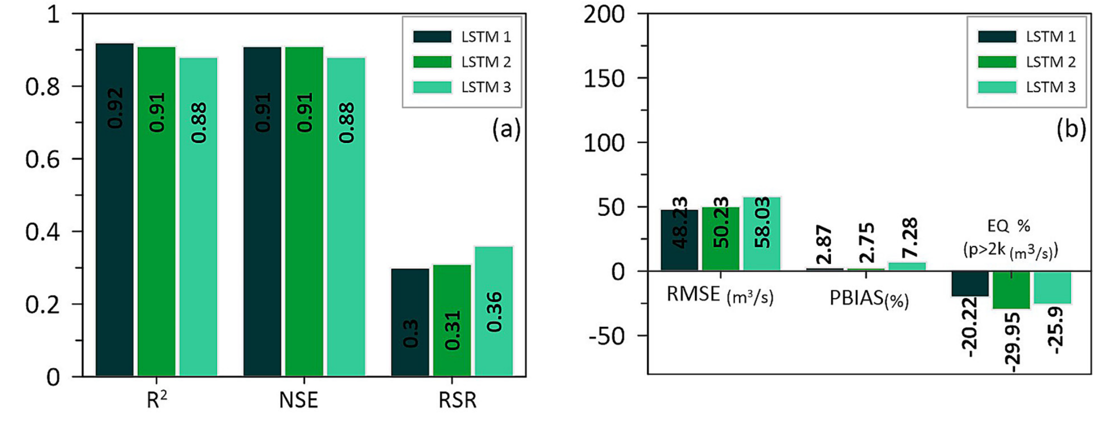


Figure 6. Statistical metrics assessment of LSTM Models 1, 2, and 3 during the training period

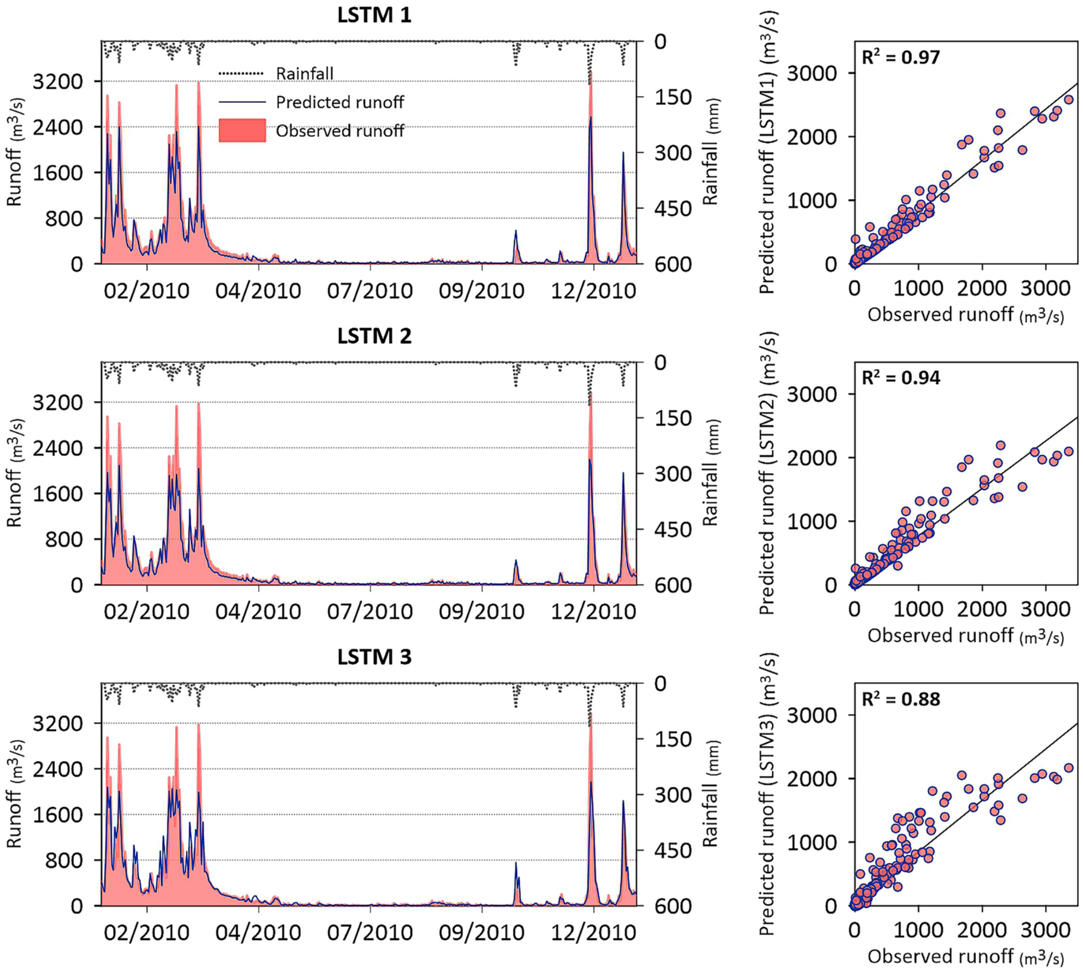


Figure 7. Graphical comparison of observed and predicted runoff by LSTM models 1, 2, and 3 during the validation period

0.94, respectively. This high performance is further supported by their NSE scores of 0.94 and 0.90 respectively, which also indicate an excellent fit to the observed data. In contrast, LSTM 3 falls behind in accurately capturing runoff variability, despite still performing well with an NSE of 0.87. In terms of the RSR metric, all three models show good results. However, LSTM 1 continues to emerge as the most accurate, closely followed by LSTM 2, while LSTM 3 displays a relatively higher error ratio. Overall, these metrics collectively suggest LSTM 1 as the most reliable model for runoff prediction during the validation period.

The PBIAS metrics during the validation period show a shift from the training results. All models exhibit an overall tendency to underestimate runoff, with LSTM 3 showing the least bias at -3.03 (Fig. 8b). In stark contrast, LSTM 1 and LSTM 2 have considerably higher biases of -15.26 and -16.49, respectively. This divergence becomes even more noteworthy when focusing on extreme runoff events exceeding 2000 m^3/s. LSTM 1 manages to keep its underestimation below the acceptable 25% threshold, with a rate of 19.88%. On the other hand, both LSTM 2 and LSTM 3 surpass this threshold, signaling potential challenges in accurately predicting extreme runoff scenarios.

Test period

The test period covers a year of dataset and comprises a combination of three flow types: the base flow, the high runoff, and two extreme discharge events. The graphical fit and correlation between the observed and predicted runoff by LSTM models 1, 2, and 3 during the test period are presented in Figure 9. It’s visible that the predicted runoff during this period aligns with the results of the validation, as all the models capture variability, seasonality, and trend of the runoff data.

In fact, both LSTM 1 and LSTM 2 adeptly reproduce the low flow variability. However, a

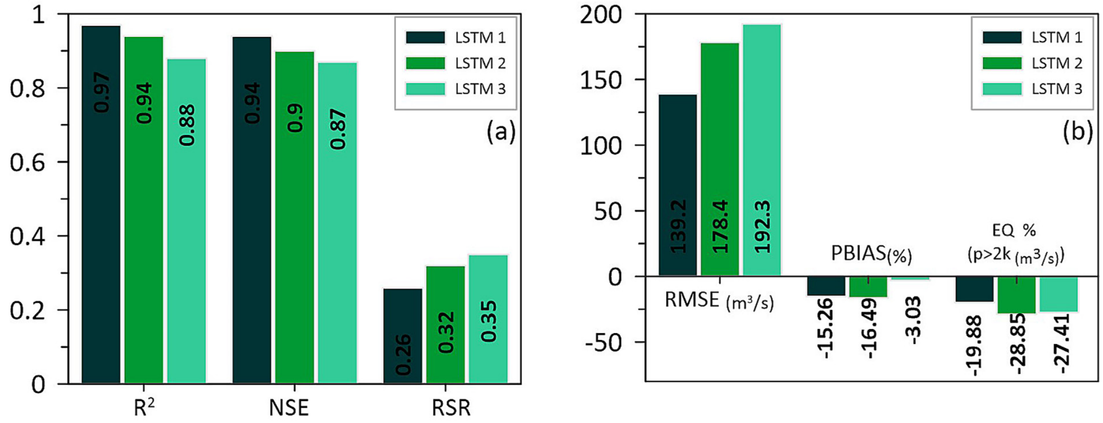


Figure 8. Statistical metrics assessment of LSTM Models 1, 2, and 3 during the validation period

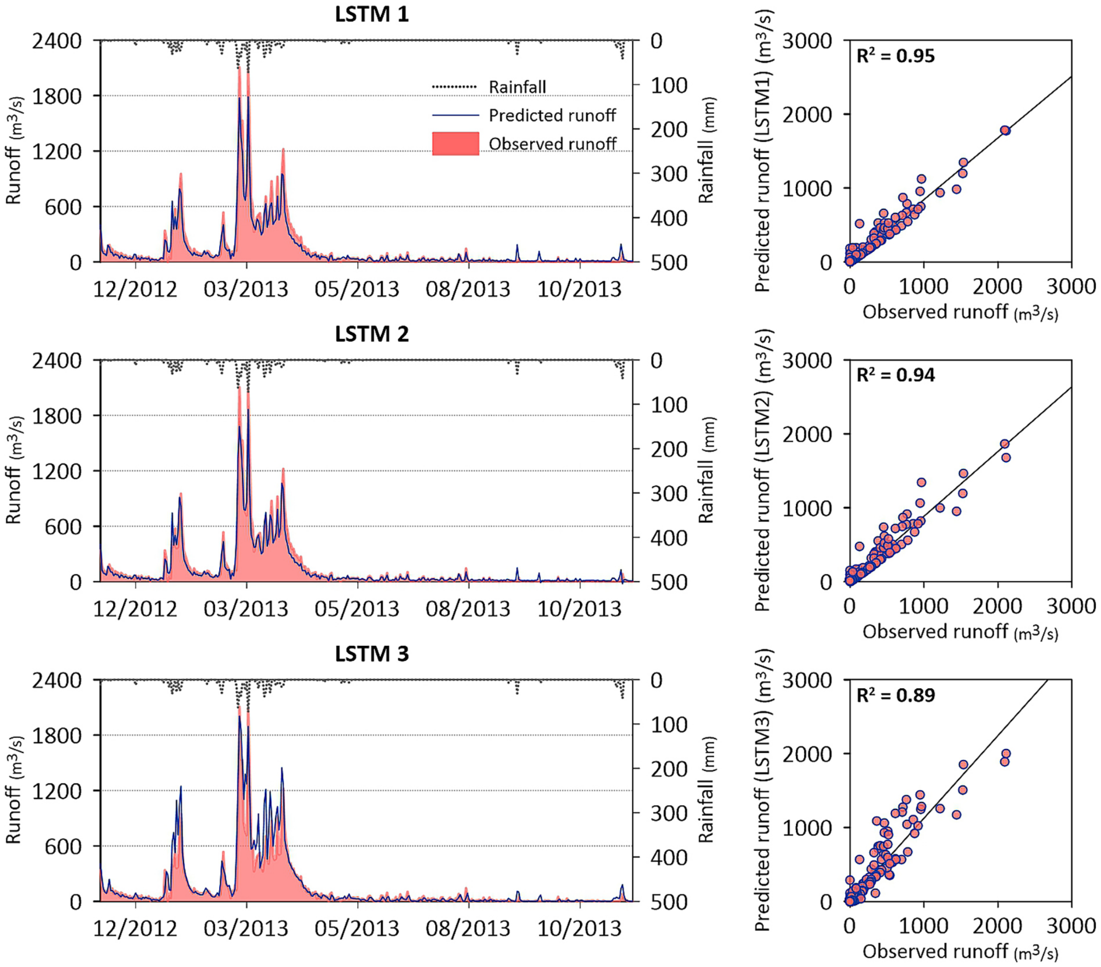


Figure 9. Graphical comparison of observed and predicted runoff by LSTM models 1, 2, and 3 during the test period

divergence appears between their performances from late 2012 to early March 2013 and throughout April. During these intervals, while LSTM 2 seems to offer a more accurate simulation of the high runoff, LSTM 1 tends to slightly underestimate it. On the other hand, in the two extreme discharge events, both models capture the extreme anomalies with almost the same level of underestimation. LSTM 3 presents a distinct pattern. It's the best in simulating the two extreme discharge peaks even though with a slight underestimation. However, there is an evident temporal lag with the low flow peaks, coupled with an overestimation of most high runoff.

Regarding the correlation pattern, the runoff predicted by LSTM 1 visibly demonstrates the best correlation, closely followed by LSTM 2 which exhibits nearly identical agreement. Nevertheless, scatter plots for LSTM 2 reveal that runoff exceeding 1200 m^3/s is less correlated than those below this threshold. LSTM 3 also shows a strong correlation, but it's narrowing to overestimate most runoff surpassing 600 m^3/s. The statistical metrics displayed in Figure 10, offer additional insight into the overall performance of the three models in the test period.

The strong linear correlation between the observed and predicted runoff, identified in the previous periods, persists with the test period. Here, LSTM 1 and LSTM 2 account for the maximum runoff variability with an R^2 of 0.95 and 0.94, respectively (Fig. 10a). This very good performance is also proven by an NSE score of 0.94 for both models. Conversely, while LSTM 3 exhibits a minor lag, it also achieves a good correlation coefficient of 0.89 and an NSE score of 0.82, thus affirming its ability to capture hydrological patterns during the test period. Further, it became evident that LSTM 1 and 2 outperformed LSTM 3 in terms of error metrics. The smallest error ratio observed in this study is less than 0.26 and was recorded for LSTM 1 and 2, which is equivalent to error magnitudes of 69 m^3/s and 68 m^3/s, respectively. LSTM 3, exhibits a broader error ratio with an RSR of 0.42, corresponding to an error magnitude of 166 m^3/s. Moreover, there is a clear tendency for LSTM 1 and 2 to underestimate not only typical runoff with biases of -9% and -5% respectively; but also, extreme discharges with biases of up to -15% for discharges above 2000 m^3/s, which is consistent with the graphical observations made above. In contrast, LSTM 3 generally overestimates runoff but achieves the closest predictions for extreme discharge events.

The results from this period have further clarified the comparative outcomes of the three models. They also demonstrate a strong consistency with the findings from the validation period and reaffirm that the three models have been successfully trained and tuned but with different degrees of performance. In fact, both LSTM 1 and LSTM 2 emerge as strong competitors and show almost similar performances of accuracy and reliability, which is reflected in closely aligned metrics such as R^2 and Nash, their only divergence resides in the LSTM 1's slightly higher negative bias. However, the LSTM 3 model possesses a distinct characteristic which is the ability to accurately capture the peaks of extreme discharge events considered a notable strength. Nevertheless, it tends to overestimate typical runoff, a tendency confirmed by its significant positive bias.

Assessment of the models' overall performance

To identify the best model performance the Taylor diagram was used (Fig. 11). This diagram visually compares modeled to observed data using three key metrics: standard deviation (radial distance from the origin), correlation coefficient (angular position), and RMSE (contour lines). In the diagram, the reference point represents the observed position on the x-axis, based on its standard deviation. It provides a consolidated view of

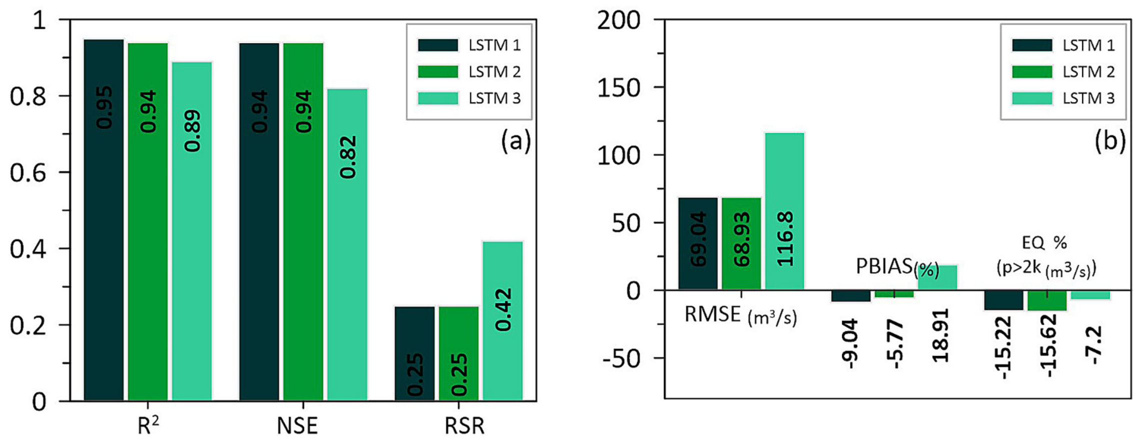


Figure 10. Statistical metrics assessment of LSTM Models 1, 2, and 3 during the test period

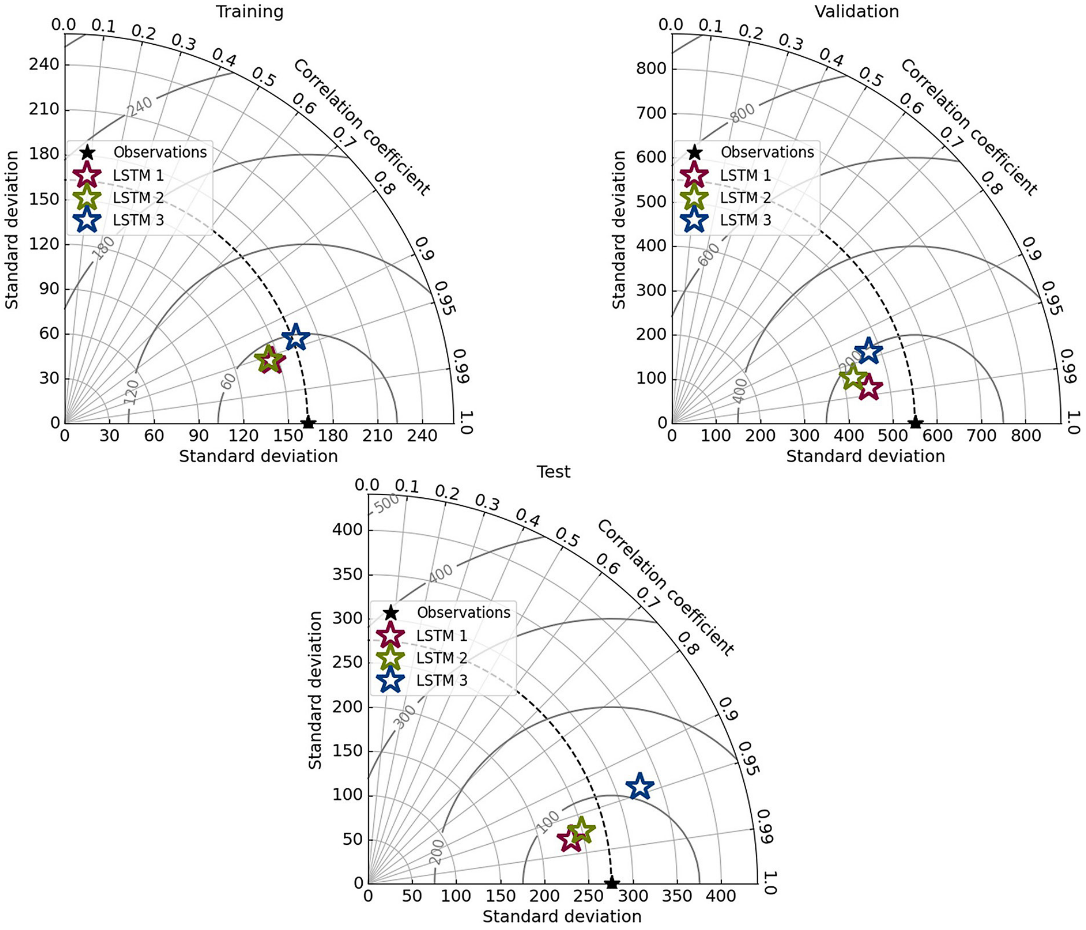


Figure 11. Comparison between the LSTM 1, 2, and 3 results cross the three periods using a Taylor diagram

model performance, allowing easy assessment of model accuracy and variability. It is noteworthy that a model's proficiency is defined by its proximity to the reference point.

The diagram indicates that LSTM 1 and 2 consistently maintain the highest correlation coefficients across all periods, suggesting their predictions closely align with observed runoff values. While they exhibit fewer errors, implying precision and consistency, they are somewhat less effective in capturing the full range of runoff fluctuations, particularly during the validation and test periods. Comparison between the LSTM 1, 2, and 3 results cross the three periods using a Taylor diagram. However, LSTM 3 also demonstrates a strong correlation but excels in capturing a broader range of runoff variations, especially in the training and validation periods. Due to increased errors, LSTM 3 is the least precise model across all periods. Therefore, while LSTM 3 may be more adept at predicting a wider range of hydrological variations, its predictions should be interpreted with caution due to the higher likelihood of inaccuracies. Overall, both LSTM 1 and LSTM 2 achieved notable results throughout the study. Nevertheless, LSTM 1 demonstrated a visibly superior performance during the validation and testing period, exhibiting a lower rate of errors and a stronger correlation, thus indicating it to be the optimal model.

DISCUSSION

While LSTMs are designed to capture long-term dependencies in time series data (Li and Wang, 2022), the number of layers or the depth within the network can have significant implications on how well these dependencies are captured. In this study, since the first experimentation, the three models have revealed a divergence in defining the suitable window size for capturing the hydrological pattern of the studied basin. The models' level of complexity has led to a different window size preference, likely due to their varying degrees of sensitivity toward major and underneath patterns within data. While the single layer LSTM 1 may primarily focus on the major patterns (Xu et al., 2021), more complex models, such as LSTM 3 with multiple layers, may be more adept at understanding complex temporal patterns spread out over longer periods. Nevertheless, deeper models are more prone to overfitting (Mejia Cajica et al., 2021), as they might capture not only the underlying patterns but also noise in the training data. The Ouergha basin; known for its rugged impermeable lands, low-density canopy, and brief concentration time (Karmouda et al., 2023), appears to naturally align with a one-day window size theoretically sufficient to capture the runoff variability and is practically defined as preferable for the single-layer LSTM 1 and the bilayer LSTM 2. Otherwise, LSTM 3, with its three-layer structure, shows intriguing performance nuances. Although it excels with a one-day window in training, a two-day window proves superior during validation. This could be a reflection of subtle patterns ignored by the less complex models, or a highlight of the model's ability to capture both short-term and longer hydrological patterns, making it versatile. However, the complex architecture of LSTM 3 brings an elevated risk of overfitting, particularly with shorter windows (Mejia Cajica et al., 2021). Thus, balancing performance and overfitting considerations, a two-day window was selected for LSTM 3.

This analysis aligns with the tuning results, revealing a notable diversity in the optimized hyperparameters. It became clear that each model can reach an optimization of the learning process, focusing on either depth or width. For LSTM 1, possessing a substantial number of units (1260) might enhance its ability to identify a diverse set of features. In contrast, LSTM 2, with a fewer number of units per layer (800), seems to strike a balance between temporal dependencies and feature extraction. LSTM 3, with its consistent preference for the lower unit count of 150, suggests a reliance on depth for feature extraction over individual layer width. Regarding the batch size, LSTM 1 and 2 are tuned with a larger batch size (365), likely aiming to stabilize gradient estimations and facilitate model convergence. In opposition, LSTM 3 opted for a smaller batch size of 128, possibly to allow more frequent weight updates in its complex architecture. In terms of epoch tuning, LSTM 1 plateaued at 14 epochs, indicating that extending beyond this might either offer marginal improvements or introduce overfitting risks. LSTM 2, which was set at 30 epochs, implies that its two-layer structure may necessitate extended iterations for optimal convergence, as opposed to a single-layer LSTM. Meanwhile, for LSTM 3, the 22 epochs seem to represent a harmony between training time and convergence.

When considering dropout, both LSTM 1 and 2 leaned towards elevated values, probably as a preventive measure against overfitting, given their substantial unit sizes. In contrast, LSTM 3 demonstrated a broader optimal range, spanning from 0.1 to 0.4, reflecting its sensitivity to different regularization needs in its three-layer structure. Overall, the substantial reduction in the number of units from LSTM 1 (1260) to LSTM 3 (150) may reveal each model's unique approach to managing the specific characteristics of the dataset. This leads to varying degrees of accuracy in predicting both regular and extreme runoff, which underscores the complexity of defining an optimal model architecture and leads us to examine the consistency and effectiveness of each model. The main results of this study indicate several insights. Despite its simpler architecture, the consistent performances of LSTM 1 across all periods emphasize the importance of a balanced architecture in the context of hydrological modeling. These findings highlight that a simple and well-designed model can potentially outperform those with more layers. Also, the efficiency of LSTM 1 may reflect the characteristics of the dataset, suggesting that a model's complexity must align with the underlying dynamics within the hydrological data.

Additionally, the LSTM 2 exhibits almost similar performance characteristics and is particularly adept at capturing a wide spectrum of hydrological patterns. Its performance metrics indicate reliable and robust behavior across different periods, making it an attractive option when LSTM 1 is not effective. However, the performance gap in LSTM 3 despite its three-layer architecture, adjusted hyperparameters, and 2-day window size likely results from the complex interaction of factors in the machine learning tuning. The 2-day window size, although optimal during tuning, may not align with the overall needs of the hydrological modeling task or may have been influenced by an overfitting of validation data. Moreover, the inherent complexity of LSTM 3 might not correspond to the nature of the data, leading to suboptimal performance. This underscores the importance of a holistic approach to model selection, considering hyperparameters in conjunction with each other, the specific characteristics of the data, and the hydrological modeling needs. Further, although few studies have utilized single-layer LSTM for predictions and their results contrast with ours (Berhich et al., 2020 and Salman et al., 2018), a significant number of research across various domains suggests that two-layer LSTMs are effective for prediction (Xu et al., 2021; Yin et al., 2022; Zia and Zahid, 2019), which support the finding of this study. Finally, the detection of extreme discharge events poses a challenge across all models. A consistent tendency towards underprediction, as evident in the EQp values, emphasizes the difficulties encountered by the models in effectively capturing the outliers within the distribution. While testing the models, this underestimation is deemed acceptable for LSTM 1 and 2, since it does not exceed -25%. With LSTM 3, this underestimation is even less pronounced but with a marked trend for error. This observation has a serious impact on hydrological forecasting, especially in scenarios where the precise prediction of extreme events is essential.

The root cause of this behavior may be linked to the training data lacking sufficient extreme samples (Sahraei et al., 2021), impeding the models' ability to adequately learn these patterns. Rahimzad et al., (2021) also support this finding and reported that earlier research, such as (Damavandi et al., 2019; Jimeno-Sáez et al., 2018) has noted similar behavior with models based on neural networks. It can be also related to other unaccounted variables influencing the discharge, such as upstream dams' releases, especially that the Ouergha basin is equipped with 3 upstream dams. This study provides an interesting insight into the potential of LSTM neural networks in predicting runoff. It emphasizes the importance of aligning model complexity with data specifications and also suggests the necessity to consider unaccounted factors like upstream dam releases to enhance the efficiency in capturing the peaks of extreme events.

CONCLUSIONS

This study has provided several key insights into the performance of LSTM models and the influence of their architecture on the accuracy of runoff prediction in the Ouergha basin. First and foremost, LSTM models exhibit a robust ability to accurately simulate river discharge and identify extreme events using only chronological rainfall and runoff data for training. In fact, the high data requirements can create significant obstacles for hydrologists and policymakers. The overall performance of LSTM in this study, suggests that these deep learning models offer a potential alternative to the conventional distributed models, which are data-intensive, requiring variables like soil type, land use, and climatic conditions such as evaporation and temperature.

Second, the different models' architecture showed varying preferences for window sizes, revealing a nuanced relationship between the complexity of the model's architecture and its ability to detect different hydrological patterns. For example, while single and bi layer LSTMs favored a one-day window, the tri-layer LSTM 3 leaned towards a two-day window. Further, the consistent excellent performance of the simplest model, LSTM 1, challenges the traditional notion that performance efficacy is directly proportional to complexity. LSTM 2, with its balanced approach, excels in environments requiring a fine-tuned relationship between temporal dependencies and feature extraction. Both these models are well-adapted to the Ouergha basin characterized by rugged impermeable lands, low-density canopy, and brief concentration time; also, they exhibit the highest level of accuracy in simulating runoff variability over extended time intervals, which implies their potential applicability in assessing the future implications of various climate change scenarios on runoff patterns. In contrast, although LSTM 3 displays adequate performance, it lags behind the simpler models in the overall accuracy. However, its ability to minimize underprediction in extreme events, positions it as the preferable choice for the forecast of eventual flood events.

Finally, the intricacies of deep learning optimization demand a careful alignment of model complexity and depth with data specifications, which is essential to avoid overfitting or the under-detection of the hydrological underlying dynamics. Moreover, it is important to take into account additional variables such as water releases from upstream dams, as these could profoundly affect the accuracy of extreme discharge predictions.

REFERENCES

1. Alardhi, S., Al-Jadir, T., Hasan, A., Jaber, A., Al Saedi, L., 2023. Design of Artificial Neural Network for Prediction of Hydrogen Sulfide and Carbon Dioxide

Concentrations in a Natural Gas Sweetening Plant. Ecol. Eng. Environ. Technol. 24, 55–66. https://doi.org/10.12912/27197050/157092

2. Alpaydin, E., 2010. Introduction to Machine Learning, fourth edition. MIT, USA.
3. Aqnouy, M., Messari, J.E.S.E., Bouadila, A., Morabbi, A., Benaabidate, L., Al-Djazouli, M.O., 2021. Modeling of Continuous and Extreme Hydrological Processes Using Spatially Distributed Models MERCEDES, VICAIR and VISHYR in a Mediterranean Watershed. Ecol. Eng. Environ. Technol., 22, 9–23. https://doi.org/10.12912/27197050/132098
4. Babu, S.R., Varma, K.P.V.K., Mohan, K.S.S., 2022. Artificial Neural Network Technique for Estimating the Thermo-Physical Properties of Water-Alumina Nanofluid. Ecol. Eng. Environ. Technol., 23, 97–106. https://doi.org/10.12912/27197050/145583
5. Bahremand, A., De Smedt, F., 2010. Predictive Analysis and Simulation Uncertainty of a Distributed Hydrological Model. Water Resour Manage, 24, 2869–2880. https://doi.org/10.1007/s11269-010-9584-1
6. Bahremand, A., De Smedt, F., 2008. Distributed Hydrological Modeling and Sensitivity Analysis in Torysa Watershed, Slovakia. Water Resour Manage, 22, 393–408. https://doi.org/10.1007/s11269-007-9168-x
7. Bashayreh, E., Manasrah, A., Alkhalil, S., Abdelhafez, E., 2021. Estimation of Water Disinfection by Using Data Mining. Ecol. Eng. Environ. Technol., 22, 109–116. https://doi.org/10.12912/27197050/132088
8. Bergstra, J., Bengio, Y., 2012. Random Search for Hyper-Parameter Optimization. Mach. Learn 281–305.
9. Berhich, A., Belouadha, F.-Z., Kassiri, A.E., 2020. Single and Multilayer LSTM Models for Positive COVID-19 Cases Prediction: Proceedings of the 2nd International Conference on Advanced Technologies for Humanity., 27–34. https://doi.org/10.5220/0010426900270034
10. Blöschl, G., et al., 2019. Changing climate both increases and decreases European river floods. Nature 573, 108–111. https://doi.org/10.1038/s41586-019-1495-6
11. Bouramtane, T., Leblanc, M., Kacimi, I., Ouatiki, H., Boudhar, A., 2023. The contribution of remote sensing and input feature selection for groundwater level prediction using LSTM neural networks in the Oum Er-Rbia Basin, Morocco. Frontiers in Water, 5. https://doi.org/10.3389/frwa.2023.1241451
12. Chen, F., Tiwari, S., Mohammed, K.S., Huo, W., Jamróz, P., 2023. Minerals resource rent responses to economic performance, greener energy, and environmental policy in China: Combination of ML and ANN outputs. Resources Policy, 81, 103–307. https://doi.org/10.1016/j.resourpol.2023.103307
13. Clarke, B., Otto, F., Stuart-Smith, R., Harrington, L., 2022. Extreme weather impacts of climate change: an attribution perspective. Environ. Res: Climate 1, https://doi.org/10.1088/2752-5295/ac6e7d
14. Combe, M., 1975. Le bassin Gharb-Maamora et les petits bassins septentrionaux des oueds Dradère et Souieire., in: Ressources En Eau Du Maroc, Notes et Mem. Serv. Géol. Maroc, pp. 93–145.
15. Damavandi, H.G., Shah, R., Stampoulis, D., Wei, Y., Boscovic, D., Sabo, J., 2019. Accurate prediction of streamflow using long short-term memory network: A case study in the Brazos river basin in Texas. International Journal of Environmental Science and Development, 10, 294–300. https://doi.org/10.18178/ijesd.2019.10.10.1190
16. Demirel, M.C., Venancio, A., Kahya, E., 2009. Flow forecast by SWAT model and ANN in Pracana basin, Portugal. Advances in Engineering Software, 40, 467–473. https://doi.org/10.1016/j.advengsoft.2008.08.002
17. Fang, K., Shen, C., Kifer, D., Yang, X., 2017. Prolongation of SMAP to Spatio-temporally Seamless Coverage of Continental US Using a Deep Learning Neural Network. Geophysical Research Letters 44. https://doi.org/10.1002/2017GL075619
18. Fischer, T., Krauss, C., 2018. Deep learning with long short-term memory networks for financial market predictions. European Journal of Operational Research, 270, 654–669. https://doi.org/10.1016/j.ejor.2017.11.054
19. Gao, S., Huang, Y., Zhang, S., Han, J., Wang, G., Zhang, M., Lin, Q., 2020. Short-term runoff prediction with GRU and LSTM networks without requiring time step optimization during sample generation. Journal of Hydrology, 589, 125188. https://doi.org/10.1016/j.jhydrol.2020.125188
20. Gers, F.A., Schmidhuber, J., Cummins, F., 2000. Learning to forget: continual prediction with LSTM. Neural Comput 12, 2451–2471. https://doi.org/10.1162/089976600300015015
21. Goodfellow, I., Bengio, Y., Courville, A., 2016. Deep Learning; MIT Press, Cambridge, MA, USA.
22. Govindaraju, R.S., Rao, A.R. (Eds.), 2000. Artificial Neural Networks in Hydrology, Water Science and Technology Library. Springer Netherlands, Dordrecht. https://doi.org/10.1007/978-94-015-9341-0
23. Gupta, H.V., Sorooshian, S., Yapo, P.O., 1999. Status of Automatic Calibration for Hydrologic Models: Comparison with Multilevel Expert Calibration. Journal of Hydrologic Engineering 4, 135–143. https://doi.org/10.1061/(ASCE)1084-0699(1999)4:2(135)
24. Harrington, P., 2017. Multiple versus single set validation of multivariate models to avoid mistakes. Crit

Rev Anal Chem 33–46.

25. Haykin, S., 1999. Neural Networks: A Comprehensive Foundation, 2nd edition. ed. Prentice-Hall, Englewood Cliffs, New Jersey, USA.
26. Hochreiter, S., 1991. Untersuchungen zu dynamischen neuronalen netzen.
27. Hochreiter, S., Bengio, Y., Frasconi, P., Schmidhuber, J., 2001. Gradient flow in recurrent nets: the difficulty of learning long-term dependencies. IEEE Press, A Field Guide to Dynamical Recurrent Neural Networks. A Field Guide to Dynamical Recurrent Neural Networks.
28. Hochreiter, S., Schmidhuber, J., 1997. Long Short-Term Memory. Neural Computation 9, 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735
29. Hsu, K., Gupta, H.V., Sorooshian, S., 1995. Artificial Neural Network Modeling of the Rainfall-Runoff Process. Water Resources Research, 31, 2517–2530. https://doi.org/10.1029/95WR01955
30. Hu, C., Wu, Q., Li, H., Jian, S., Li, N., Lou, Z., 2018. Deep Learning with a Long Short-Term Memory Networks Approach for Rainfall-Runoff Simulation. Water, 10, 1543. https://doi.org/10.3390/w10111543
31. Jimeno-Sáez, P., Senent-Aparicio, J., Pérez-Sánchez, J., Pulido-Velazquez, D., 2018. A Comparison of SWAT and ANN Models for Daily Runoff Simulation in Different Climatic Zones of Peninsular Spain. Water, 10, 192. https://doi.org/10.3390/w10020192
32. Juan, C., Genxu, W., Tianxu, M., Xiangyang, S., 2017. ANN Model-Based Simulation of the Runoff Variation in Response to Climate Change on the Qinghai-Tibet Plateau, China. Advances in Meteorology 2017, e9451802. https://doi.org/10.1155/2017/9451802
33. Karmouda, N., Naïma, E.A., Kacimi, I., Mahe, G., Bouramtane, T., Brirhet, H., Idrissi, A., Kassou, N., 2023. Hydrological Modelling of Extreme Events in Ouergha Mediterranean Basin, Northern Morocco, Using a Deterministic Model and Gridded Precipitations. Iraqi Geological Journal, 56, 1–20. https://doi.org/10.46717/igj.56.2B.1ms-2023-8-10
34. Kawakami, K., 2008. Supervised Sequence Labelling with Recurrent Neural Networks. (Ph.D. Thesis). Technical University of Munich, Munich, Germany.
35. Kingma, D.P., Ba, J., 2017. Adam: A Method for Stochastic Optimization San Diego.
36. Kouassi, A.M., Koffi, Y.B., Kouamé, K.F., Lasm, T., 2013. Application d'un modèle conceptuel et d'un modèle de réseaux de neurones artificiels à la simulation des débits annuels dans le bassin versant du N'zi-Bandama (Côte d'Ivoire) 09, 64–76.
37. Li, M., Wang, Z., 2022. LSTM-augmented deep networks for time-variant reliability assessment of dynamic systems. Reliability Engineering & System Safety, 217, 108014. https://doi.org/10.1016/j.ress.2021.108014
38. Liu, Y., Zhang, T., Kang, A., Li, J., Lei, X., 2021. Research on Runoff Simulations Using Deep-Learning Methods. Sustainability, 13, 1336. https://doi.org/10.3390/su13031336
39. Man, Y., Yang, Q., Shao, J., Wang, G., Bai, L., Xue, Y., 2022. Enhanced LSTM Model for Daily Runoff Prediction in the Upper Huai River Basin, China. Engineering, S2095809922002806. https://doi.org/10.1016/j.eng.2021.12.022
40. Mao, G., Wang, M., Liu, J., Wang, Z., Wang, K., Meng, Y., Zhong, R., Wang, H., Li, Y., 2021. Comprehensive comparison of artificial neural networks and long short-term memory networks for rainfall-runoff simulation. Physics and Chemistry of the Earth, Parts A/B/C, 123, 103026. https://doi.org/10.1016/j.pce.2021.103026
41. Mejia Cajica, F.A., García Henao, J.A., Barrios Hernández, C.J., Riveil, M., 2021. High Performance Computing: 7th Latin American Conference, CARLA 2020, Cuenca, Ecuador, September 2–4, 2020. https://doi.org/10.1007/978-3-030-68035-0
42. Moriasi, D.N., Arnold, J.G., Liew, M.W., Bingner, R.L., Harmel, R.D., Veith, T.L., 2007. Model Evaluation Guidelines for Systematic Quantification of Accuracy in Watershed Simulations. Transactions of the ASABE, 50, 885–900. https://doi.org/10.13031/2013.23153
43. Moriasi, D.N., Gitau, M.W., Daggupati, P., 2015. Hydrologic and Water Quality Models: Performance Measures and Evaluation Criteria. Trans. ASABE, 58, 1763–1785. https://doi.org/10.13031/trans.58.10715
44. Msatef, K., Benaabidate, L., Bouignane, A., 2018. Hydrological and hydroclimatic regimes in the Ouergha watershed. E3S Web Conf., 37, 1–11. https://doi.org/10.1051/e3sconf/20183704001
45. Oni, O.M., Aremu, A.A., Oladapo, O.O., Agboluaje, B.A., Fajemiroye, J.A., 2022. Artificial neural network modeling of meteorological and geological influences on indoor radon concentration in selected tertiary institutions in Southwestern Nigeria. Journal of Environmental Radioactivity, 251–252, 106933. https://doi.org/10.1016/j.jenvrad.2022.106933
46. Rahimzad, M., Moghaddam Nia, A., Zolfonoon, H., Soltani, J., Danandeh Mehr, A., Kwon, H.-H., 2021. Performance Comparison of an LSTM-based Deep Learning Model versus Conventional Machine Learning Algorithms for Streamflow Forecasting. Water Resour Manage, 35, 4167–4187. https://doi.org/10.1007/s11269-021-02937-w
47. Rajaee, T., Ebrahimi, H., Nourani, V., 2019. A review of the artificial intelligence methods in groundwater level modeling. Journal of Hydrology, 572, 336–351. https://doi.org/10.1016/j.jhydrol.2018.12.037

48. Ripley, B.D., 1996. Pattern Recognition and Neural Networks Cambridge, MA, USA.

49. Sahraei, A., Chamorro, A., Kraft, P., Breuer, L., 2021. Application of Machine Learning Models to Predict Maximum Event Water Fractions in Streamflow. Frontiers in Water 3.

50. Salman, A.G., Heryadi, Y., Abdurahman, E., Suparta, W., 2018. Single Layer & Multi-layer Long Short-Term Memory (LSTM) Model with Intermediate Variables for Weather Forecasting. Procedia Computer Science, 135, 89–98. https://doi.org/10.1016/j.procs.2018.08.153

51. Senoussi, S., Agoumi, A., Yacoubi, M., Fakhraddine, A., Sayouty, E.H., Mokssit, A., Chikri, N., 1999. Changements climatiques et ressources en eau Bassin versant de I'Ouergha (Maroc). Hydroécol. Appl. 11, 163–182. https://doi.org/10.1051/hydro:1999007

52. Servat, E., Dezetter, A., 1991. Selection of calibration objective fonctions in the context of rainfall-ronoff modelling in a Sudanese savannah area. Hydrological Sciences Journal, 36, 307–330. https://doi.org/10.1080/02626669109492517

53. Shanker, M., Hu, M.Y., Hung, M.S., 1996. Effect of data standardization on neural network training. Omega 24, 385–397. https://doi.org/10.1016/0305-0483(96)00010-2

54. Sharma, D.K., Chatterjee, M., Kaur, G., Vavilala, S., 2022. 3 - Deep learning applications for disease diagnosis, in: Gupta, D., Kose, U., Khanna, A., Balas, V.E. (Eds.), Deep Learning for Medical Applications with Unique Data. Academic Press, pp. 31–51. https://doi.org/10.1016/B978-0-12-824145-5.00005-8

55. Sherstinsky, A., 2020. Fundamentals of Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) network. Physica D: Nonlinear Phenomena 404, 132306. https://doi.org/10.1016/j.physd.2019.132306

56. Singh, J., Knapp, H.V., Demissie, M., 2004. Hydrologic Modeling of the Iroquois River Watershed Using HSPF and SWAT. Illinois State Water Survey Contract, 08, 1–24.

57. Tingsanchali, T., 2000. Forecasting model of chao phraya river flood levels at bangkok., in: International Conference on Chao Phraya Delta. Citeseer, Bangkok, Thailand.

58. Westerhuis, J., Hoefsloot, H., Smit, S., Vis, D., Smilde, A., van Velzen, E., van Duijnhoven, J., van Dorsten, F., 2008. Assessment of PLSDA cross validation. Metabolomics, 81–9.

59. Wu, C.L., Chau, K.W., 2011. Rainfall–runoff modeling using artificial neural network coupled with singular spectrum analysis. Journal of Hydrology, 399, 394–409. https://doi.org/10.1016/j.jhydrol.2011.01.017

60. Xiang, Z., Yan, J., Demir, I., 2020. A Rainfall-Runoff Model With LSTM-Based Sequence-to-Sequence Learning. Water Resources Research, 56. https://doi.org/10.1029/2019WR025326

61. Xu, Y., Yao, L., Xu, P., Cui, W., Zhang, Z., Liu, F., Mao, B., Wen, Z., 2021. Load Forecasting Method for Building Energy Systems Based on Modified Two-Layer LSTM, in: 2021 3rd Asia Energy and Electrical Engineering Symposium. pp. 660–665.

62. Yin, H., Wang, F., Zhang, X., Zhang, Y., Chen, J., Xia, R., Jin, J., 2022. Rainfall-runoff modeling using long short-term memory-based step-sequence framework. Journal of Hydrology, 610, 127901. https://doi.org/10.1016/j.jhydrol.2022.127901

63. Young, Liu, 2015. Prediction and modelling of rainfall-runoff during typhoon events using a physically-based and artificial neural network hybrid model. Hydrol. Sci. ,J 60, 2102–2116.

64. Yu, Y., Si, X., Hu, C., Zhang, J., 2019. A Review of Recurrent Neural Networks: LSTM Cells and Network Architectures. Neural Comput, 31, 1235–1270. https://doi.org/10.1162/neco_a_01199

65. Zema, D.A., Lucas-Borja, M.E., Fotia, L., Rosaci, D., Sarnè, G.M.L., Zimbone, S.M., 2020. Predicting the hydrological response of a forest after wildfire and soil treatments using an Artificial Neural Network. Computers and Electronics in Agriculture, 170, 105280. https://doi.org/10.1016/j.compag.2020.105280

66. Zia, T., Zahid, U., 2019. Long short-term memory recurrent neural network architectures for Urdu acoustic modeling. Int J Speech Technol, 22, 21–30. https://doi.org/10.1007/s10772-018-09573-7