Rainfall and runoff time-series trend analysis using LSTM recurrent neural network and wavelet neural network with satellite-based meteorological data: case study of Nzoia hydrolog…

1 author:

Yashon Ouma
Moi University
37 PUBLICATIONS **1,088** CITATIONS

Some of the authors of this publication are also working on these related projects:

Project Surface water quality and long-term impacts of LULC in Botswana's Limpopo River Basin. US NAS/PEER/ORD View project

Project Mobile Mapping Systems for Pavement Condition Mapping and Evaluation. AvH Germany/HFT Stuttgart View project

Rainfall and runoff time-series trend analysis using LSTM recurrent neural network and wavelet neural network with satellite-based meteorological data: case study of Nzoia hydrologic basin

Yashon O. Ouma[1,2] · Rodrick Cheruyot[1] · Alice N. Wachera[1]

Abstract

This study compares LSTM neural network and wavelet neural network (WNN) for spatio-temporal prediction of rainfall and runoff time-series trends in scarcely gauged hydrologic basins. Using long-term in situ observed data for 30 years (1980–2009) from ten rain gauge stations and three discharge measurement stations, the rainfall and runoff trends in the Nzoia River basin are predicted through satellite-based meteorological data comprising of: precipitation, mean temperature, relative humidity, wind speed and solar radiation. The prediction modelling was carried out in three sub-basins corresponding to the three discharge stations. LSTM and WNN were implemented with the same deep learning topological structure consisting of 4 hidden layers, each with 30 neurons. In the prediction of the basin runoff with the five meteorological parameters using LSTM and WNN, both models performed well with respective R^2 values of 0.8967 and 0.8820. The MAE and RMSE measures for LSTM and WNN predictions ranged between 11–13 m^3/s for the mean monthly runoff prediction. With the satellite-based meteorological data, LSTM predicted the mean monthly rainfall within the basin with $R^2 = 0.8610$ as compared to $R^2 = 0.7825$ using WNN. The MAE for mean monthly rainfall trend prediction was between 9 and 11 mm, while the RMSE varied between 15 and 21 mm. The performance of the models improved with increase in the number of input parameters, which corresponded to the size of the sub-basin. In terms of the computational time, both models converged at the lowest RMSE at nearly the same number of epochs, with WNN taking slightly longer to attain the minimum RMSE. The study shows that in hydrologic basins with scarce meteorological and hydrological monitoring networks, the use satellite-based meteorological data in deep learning neural network models are suitable for spatial and temporal analysis of rainfall and runoff trends.

Keywords Rainfall and runoff trend analysis · Deep learning · LSTM neural network · Wavelet neural network (WNN) · Satellite-based meteorological data · Nzoia River basin

Introduction

In sustainable water resources management, the accurate modelling of hydrological processes at watershed scales is a significant contributing factor. In particular, predictions of rainfall and runoff trends are important for different water resource planning such as in irrigation, flood control, structural design, and eco-hydrological services [1]. In most countries, catchment basins are sparsely gauged without accurate and adequate rainfall and runoff measurements [2]. This contributes to higher uncertainty in attempts to predict hydrological responses in such areas [3]. Rainfall and runoff can be characterized as random stochastic processes related to complex physical factors within catchments. Because of the spatial and temporal variabilities within watersheds, the patterns and number of variables required for the modelling of rainfall and runoff presents a complex hydrologic problem [4, 5].

Generally, the forecasting of time-series data depends on the sequence being modelled and can have different dimensional spatial dependencies. In the prediction of rainfall and runoff time-series trends, physical and conceptual models

✉ Yashon O. Ouma
yashon_o@hotmail.com

1 Department of Civil and Structural Engineering, Moi University, Eldoret 30100, Kenya

2 Department of Civil Engineering, University of Botswana, PBag 0061, Gaborone, Botswana

مدينة الملك عبدالعزيز
KACST للعلوم والتقنية

🜀 Springer

are traditionally utilized [6]. While conceptual models are considered to be suitable for daily timescale analysis, physical models can be used for daily and sub-daily timescale predictions. Because of the timescale dependencies, the physical and conceptual models are considered as unsuitable for accurate prediction of rainfall and runoff particularly where there is lack of high resolution spatial [7, 9]. Furthermore, these models require physical parameters which limit their application in the prediction of sequence data with unknown or limited quasi-periodic dynamics [10–12, 50].

Among the statistical methods, the Autoregressive Moving Average (ARMA) and its invariants as Autoregressive Integrated Moving Average (ARIMA) and nearest-neighbor methods have been used for rainfall and runoff predictions. Nevertheless, the accuracy of the statistical methods depends on the quality of the input data and can only satisfactorily describe time-series data that exhibit non-stationary behaviors within and across seasons [45]. To improve on the rainfall and runoff prediction results, data-driven approaches have been proposed. This is attributed to the ability to approximate the inherent patterns and dynamics of series data without the knowledge of the parameters, and to take into consideration the stochasticity in observation and system noise. In particular, ANNs have been explored and preferred for rainfall and runoff time-series trend analysis [13]. To predict rainfall and runoff, [14] proposed to use ANN and fuzzy logic, while ANN and ARIMA models were adopted in [15]. Other studies, e.g. [16], utilized Radial Basis Function (RBF) network and empirical model decompositions in the prediction of rainfall. The above studies reported that the approaches did not adequately capture the seasonal decomposition and the inherent cyclical fluctuations in rainfall and runoff time sequence data. Towards improving ANN performance in hydrological predictions for Kentucky River catchment, [17] developed an ANN-based training using genetic algorithms for streamflow magnitude predictions. For the case study of data sparse Malaprabha River Basin in India, [18] developed a modular neural network to capture variable intensities in rainfall and runoff simulations. The results from [18] were superior as compared to the methods in [6, 17, 19].

Despite the aforementioned results, the main drawback of the conventional feedforward ANNs is their tendency to lose significant information on the sequential order of the input data during training. This is attributed to the vanishing gradient effect which occurs as the number of layers increases [13]. Further, the applicability of a single ANN in hydrological phenomena modelling may not reliably capture the localized temporal and seasonal dynamics of rainfall and runoff [18, 20]. Thus, because of the inherent seasonality and non-linear characteristics of rainfall and runoff time-series data, hybrid models such as: wavelets and least squares Support Vector Machines [40]; wavelet transform

and artificial neural network hybrid models [46]; wavelet-artificial neural network and comparison with adaptive neurofuzzy inference system [48], and singular spectral analysis and discrete wavelet transform in hybrid models [55] have been recommended in their simulation and prediction.

The problem of displaying long-term dependencies in time-series data implies that the desired output at time t depends on input value that occurred at an earlier time $\tau << t$. As such, the dynamical neural system for such a task should be able to learn to store information for an arbitrary duration (memory) for the minimization of noise corruption. Because the typical feedforward network is not sufficiently powerful to discover contingencies spanning long temporal distances, it easily suffers from vanishing gradient effect as the number of layers increases. RNNs are most suited to store long-term time-series data with different temporal scales. However, simple RNNs that depend on the largest eigenvalue of the state-update matrix may have gradients which either increase or decrease exponentially over time. Long short-term memory (LSTM) RNN [25], was developed to improve on the conventional RNN models. LSTM-RNN uses input, output and forget gates to achieve a network that can maintain state and propagate gradients in a stable fashion over long time spans. These networks have been shown to outperform deep feedforward neural networks on a variety of tasks [57]. Due to such capability, LSTM has been applied for rainfall and runoff predictions in [13, 21].

To take into account the non-stationarity in the assimilation of rainfall and runoff time-series data, this study proposes to compare the wavelet neural network (WNN) and the LSTM recurrent neural network. The comparison is based on the fact that in data showing persistence structure within the series, data-driven models can be considered to be more appropriate in accounting for their sequence dependency, non-stationarity and non-linearity. Recent investigations have demonstrated that WNN [23] and LSTM [13, 21] as data-driven models, can overcome the constraints of time-series modelling and are suitable for taking into account the quasi-periodicities in rainfall and runoff predictions.

Specifically, wavelet transform (WT) can be used to analyze the data signal details through signal decomposition into time–frequency domains. Adopting discrete wavelet transformation (DWT), the rainfall and runoff series data can be estimated into independent data with periodicity [24, 25, 39]. Further, in temporal sequence predictions, DWT can infer the normally hidden time–frequency information in time-series data. This study thus proposes the wavelets coupled ANN towards improving the rainfall and runoff prediction model performance as demonstrated in forecasting streamflows with more reliable results [40, 41]. The wavelet-neural network model is proven to be superior to the conventional ANN and statistical regression models in rainfall and runoff prediction in different case studies by [48, 51, 52].

Compared to WNN, LSTM is capable of dynamically incorporating predecessor or past learning experience due to internal recurrence. LSTM is also considered to be more powerful computationally and topologically more reasonable as compared to the conventional feedforward neural networks without internal states [25]. With this ability, LSTM can automatically project the inherent properties in time-series data for accurate simulation and approximation of the chaotic series. LSTM is also suitable where there is a long delay and accounts for time-series signals with low- and high-frequencies. Compared to the conventional ANN and statistical models such as ARMA and ARIMA, LSTM is capable of robustly learning information contained in time-series data, and can effectively capture the variability of time-series data [23, 42, 43].

To understand the significance of the two neural network models in rainfall and runoff trend analysis, this study explores the implementation of WNN and LSTM in rainfall and runoff trend characterization and predictions within a hydrologic basin with scarce meteorological and hydrological monitoring network. From literature, comparisons on the advantages of WNN and LSTM have not been carried out especially in rainfall and runoff hydrological applications in data scarce basins. Because of lack of observed meteorological data, this study also evaluates the significance of satellite-based meteorological data including rainfall, temperature, relative humidity, wind speed and solar radiation as input data in rainfall and runoff trend characterization and prediction in the Nzoia River basin in Kenya.

Materials and methods

Study area characterization

The Nzoia River Basin forms part of the larger Lake Victoria basin. The basin is situated within latitudes 1°30′ N and 0°30′ S and longitude 34°00′ E and 35°45′ E, with an approximate area of 12,700 km². The elevation ranges between 1100 m and 4000 m AMSL (Fig. 1) [44]. The lower parts of the basin have a flat terrain with a slope mainly ranging between 2 and 6%, whereas the upper part is hilly with more rugged terrain as depicted in Fig. 1. Further details on the study area land-use and climatic characteristics can be found in [27, 44].

In the prediction modelling, the rainfall and streamflow stations are independently modelled with respect to the containing sub-basin as depicted in Fig. 1. Each discharge station was treated as a pour point and sub-basins were delineated to include all the streams draining to each discharge station. In Fig. 1, the spatial distributions of the discharge stations and the ground and satellite meteorological stations are shown.
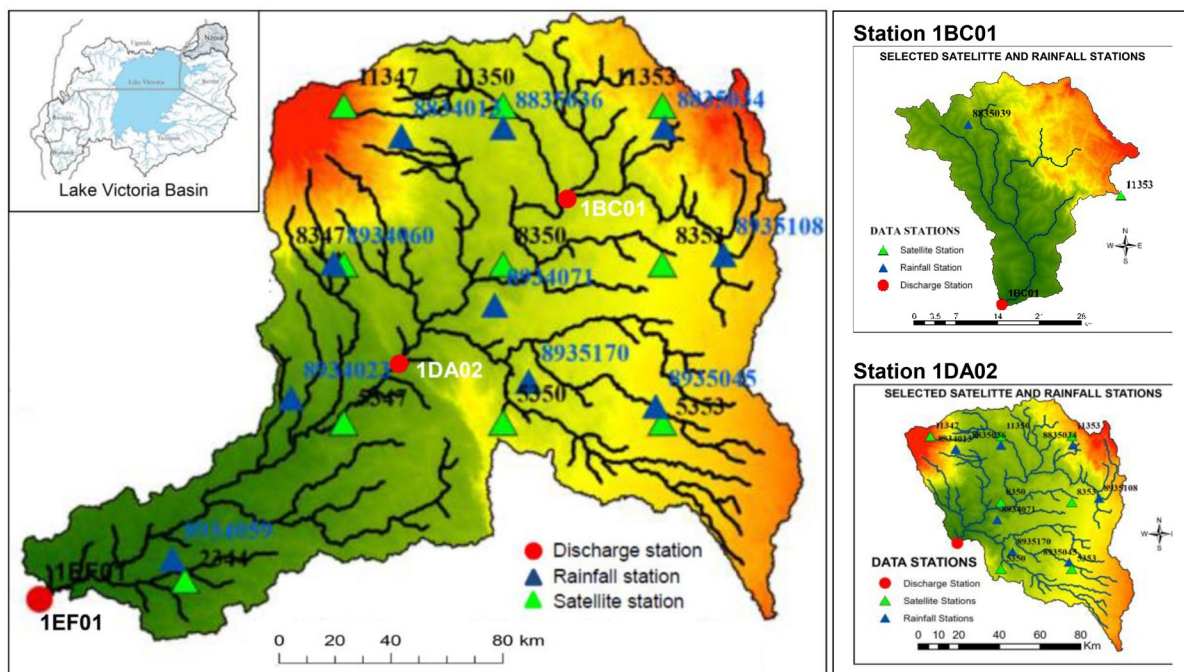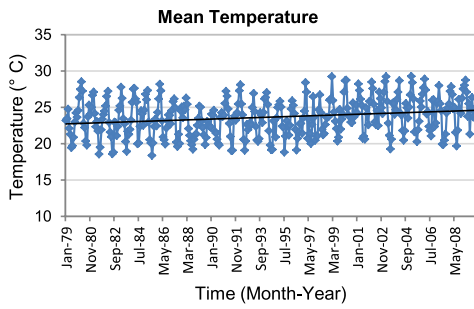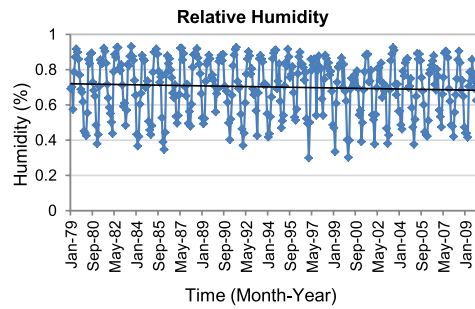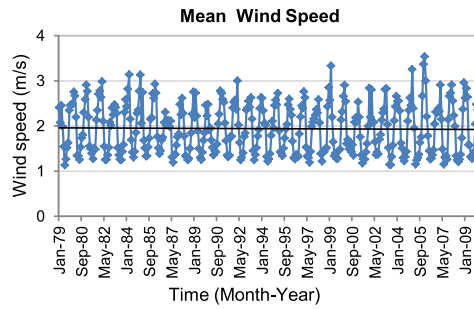

Fig. 1 Location of Nzoia River basin within the Lake Victoria basin and the rainfall and discharge measurement stations. The main pour point is station 1EF01 and sub-basin discharge stations are 1BC01 and 1DA02
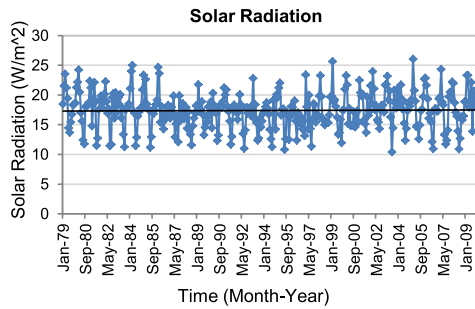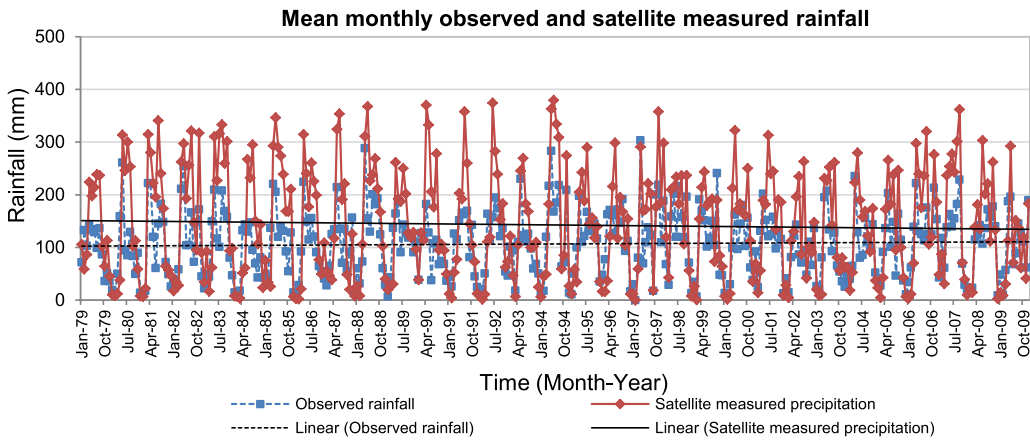
(a) Mean Temperature

(b) Relative Humidity

(c) Mean Wind Speed

(d) Solar Radiation

(e) Mean monthly observed and satellite measured rainfall

Observed rainfall • Satellite measured precipitation • Linear (Observed rainfall) • Linear (Satellite measured precipitation)
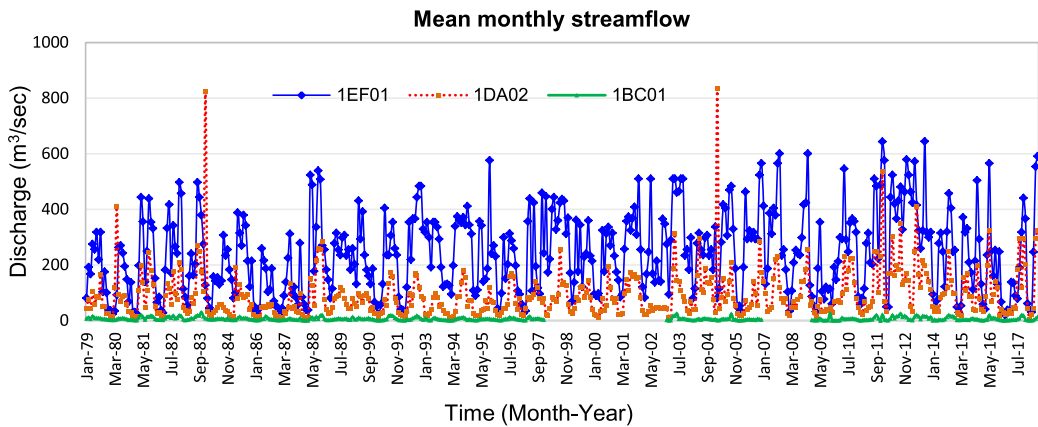
(f) Mean monthly streamflow

1EF01 • 1DA02 • 1BC01

◀Fig. 2 Mean monthly satellite-based: **a–d** temperature, humidity, wind speed and solar radiation; **e** ground-observed and satellite measured rainfall, and **f** streamflow for stations 1EF01, 1DA02 and 1BC01

Data

The basin was divided into upper (1BC01), middle (1DA02) and lower (1EF01) sub-basins according to the locations of the discharge measurement stations (Fig. 1). The in situ rainfall data and satellite-based meteorological data were aggregated to monthly averages for the 30 years (1980–2009) of study.

The satellite meteorological data were downloaded from the National Centers for Environmental Prediction Climate Forecast System Reanalysis (NCEP-CFSR) website (https://climatedataguide.ucar.edu/climate-data/climate-forecast-system-reanalysis-cfsr). The graphical variability of the mean monthly satellite-based and measured meteorological parameters and streamflow are presented in Fig. 2. Notably, from the meteorological data, a linear trend analysis shows increase in temperature and solar radiation, with corresponding decrease in relative humidity, wind speed and received precipitation within the basin (Fig. 2a–d). The trends in the precipitation from the *in-situ* measurements and from satellite data have a marginal difference with the satellite data overestimating the observed precipitation (Fig. 2e).

Table 1 presents a summary of the mean monthly statistical descriptions of the satellite-based meteorological data and the observed rainfall and streamflow data. Comparing the observed and satellite-based meteorological data, it is noted that the satellite-based rainfall overestimated the measured mean monthly rainfall by approximately 36 mm, and with twice the standard deviation. The streamflow volume is seen to vary according to the location of the pour point and size of the sub-basin.

Methods

This section introduces the neural network models structure, implementation and validation approach. Figure 3 presents a summary of the implementation strategy and processing flow for the prediction of rainfall and runoff trends with the proposed WNN and LSTM models. In Fig. 3, the missing rainfall data (RF) were interpolated using Inverse Distance Weighting (IDW) method. The deterministic IDW was used to interpolate the rainfall points since the missing data was less than 5%.

In general, time-series data forecasting can be represented as a 2D problem, for example as a $P \times Q$, phenomenon tensor $Y \in R^{P \times Q \times k}$ with k measurements. The spatial observations can be represented in a time-dimension with T-time steps as $Y_{1:T}$. In the forecasting modelling, with $Y_{1:T}$ as the previous observations, the future ΔT sequence is defined by the time interval $T(1 + \Delta T)$ which is estimated as $\widehat{Y}_{T+1:T+\Delta T}$. The forecasting task in this case is defined as $\widehat{Y}_{T+1:T+\Delta T} = \underset{Y_{T+1:T+\Delta T}}{\arg \max} \, p(Y_{T+1:T+\Delta T} | Y_{1:T})$, which is the mostly likely predicted sequence.

LSTM recurrent neural networks

RNNs are a special neural network designed for understanding the temporal and dynamic sequences in series data [28–31]. As compared to feedforward networks that pass the data forward from input to output, RNNs have feedback loops where output data can feedback into the input at some point before feedforward again for further processing and final output. The advantage of the RNN model is in the ability to model sequence time-series data such that each sample can be assumed to be dependent on previous ones. The feedback connections in RNNs provide the memory of previous activations, thus allowing the model to iteratively learn the dynamics of the sequential data in time-steps [42].

Though RNNs exhibit powerful capability for modelling of complex and non-linear time-series data [32], the conventional RNNs suffers from diminishing gradient, especially in the backpropagation iterative learning process. Thus, conventional RNNs may not adequately learn from longer time-lag data with dependencies [32]. To solve this problem, [25] proposed the LSTM algorithm (Fig. 3). The LSTM have been proven to be more suitable in the simulation of sequence-based problems with long-term dependencies [34, 41].

In the LSTM-RNN model, the memory blocks contain the input gate, output gate and forget gate that replace the hidden units. The gates are responsible for the control the internal operations on the network. Despite different LSTM variants being proposed, a comparative analysis shows that the standard LSTM is still the most significant [35, 42], and thus adopted and evaluated in this study. Figure 4 shows a sample representation of a single LSTM architecture as adopted in the current study. In addition to the memory gates, the fundamental components of the LSTM network comprise of the cell state, sigmoid and the tan*h* activation functions as depicted in Fig. 4. The inclusion and removal of information to the cell state are regulated by the gates. The sigmoid activation functions in the gates multiplies inputs by values of [1], and determines the data to be included or removed.

Let c_t be the input sum at time step t, the LSTM updates for time step i at given inputs x_t, h_{t-1}, and c_{t-1} are given as in Eqs. 1–5 [25], that describe the algorithms of a typical LSTM layers.

$$f_t = \sigma\left(W_{xf} \cdot x_t + W_{hf} \cdot h_{t-1} + W_{cf} \cdot c_{t-1} + b_f\right) \tag{1}$$

Table 1 Descriptive statistics of the mean monthly observed and satellite-based meteorological data and streamflow

Parameters	Min	Max	Median	Mean	SD	CV (%)	SE
Satellite-based parameters (1980–2009)							
Precipitation (mm)	0.1	379.3	127.2	142.9	102.7	69	5.30
Max. Temperature (°C)	25.5	37.1	31.6	31.4	2.3	7	0.12
Min. Temperature (°C)	12.1	17.7	15.8	15.6	0.9	6	0.05
Relative Humidity (%)	0.3	0.9	0.6	0.6	0.1	19	0.01
Wind Speed (m/s)	1.6	3.9	2.4	2.5	0.5	19	0.02
Solar Radiation (W/m^2)	14.0	27.2	18.6	19.1	2.7	14	0.13
Observed rainfall data (1980–2009)							
Precipitation (mm)	0	303.0	110.4	106.6	61.0	60	3.20
Observed discharge data per station (1980–2009)							
1BC01	0.5	30.4	5.4	6.1	4.6	80	0.34
1DA01	9.0	836.1	59.7	82.2	82.4	100	4.30
1EF01	20.13	900.3	192.1	214.7	152.1	71	7.90

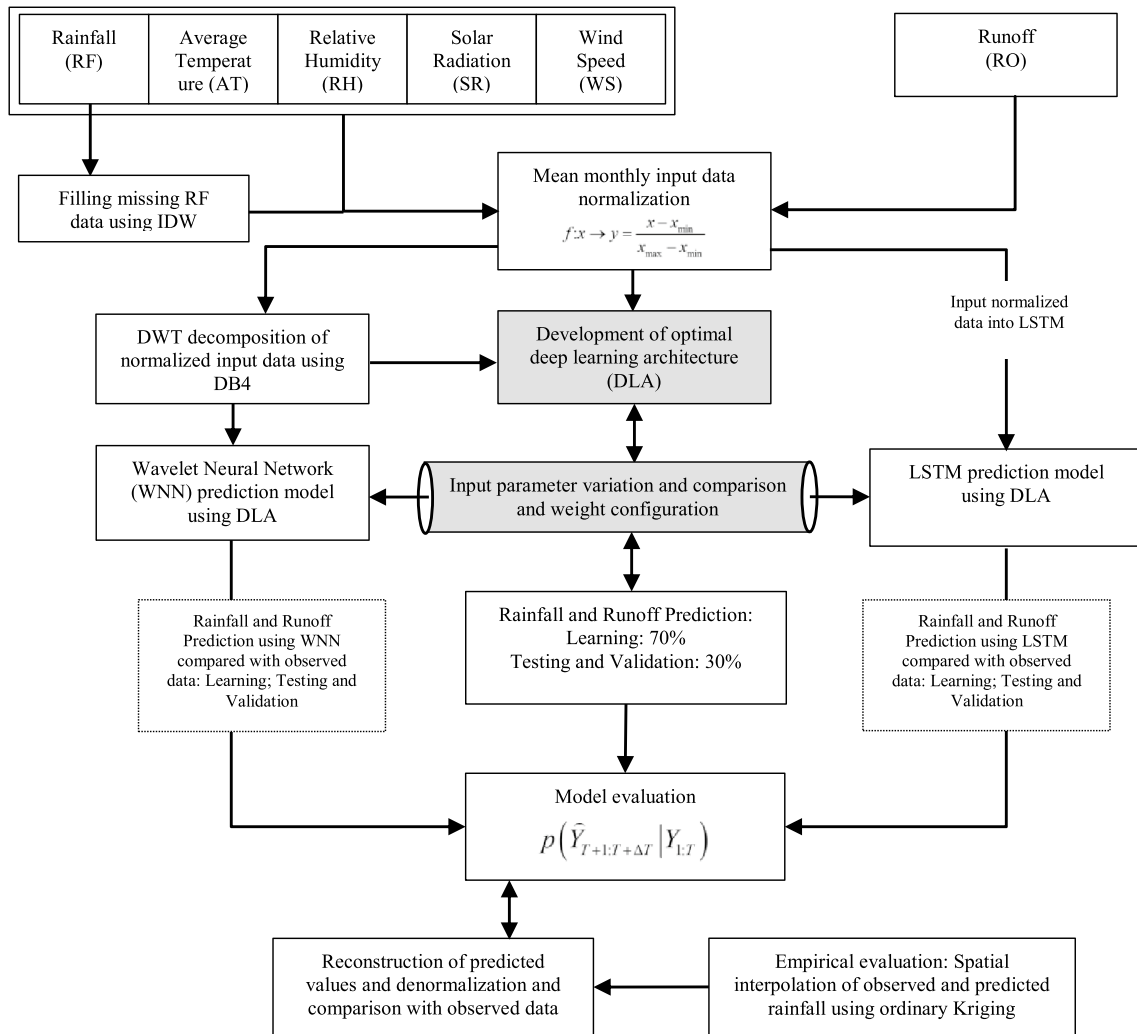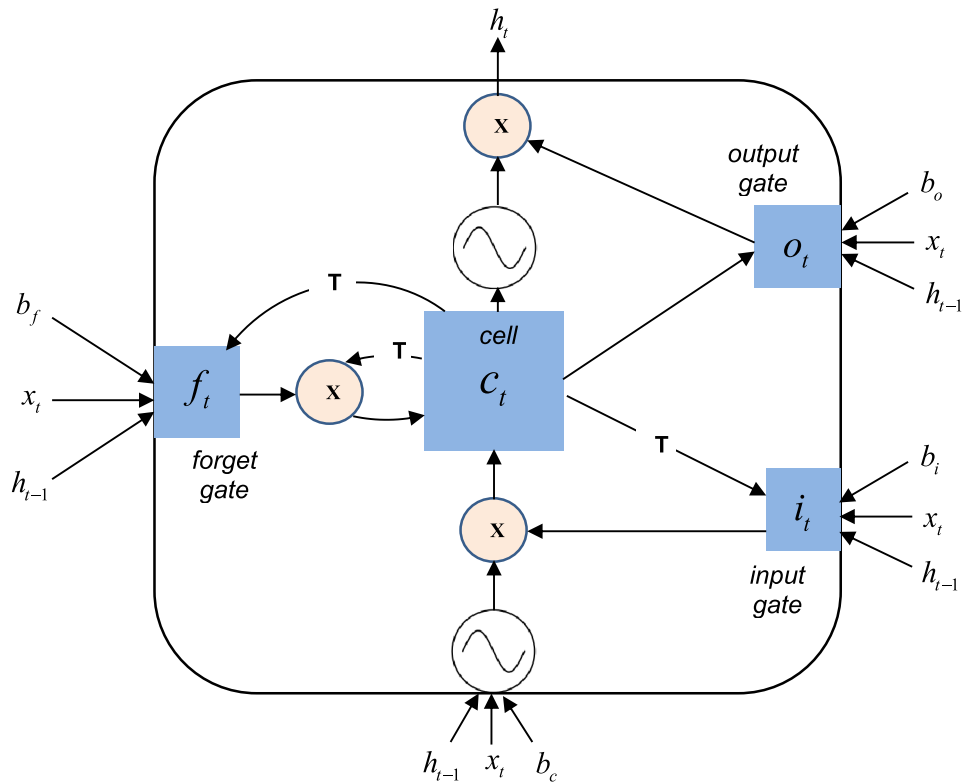*SD* standard deviation, *CV* coefficient of variation, *SE* standard error

Fig. 3 Processing flow for the prediction of rainfall and runoff using LSTM and WNN

Fig. 4 RNN layer structure for LSTM implementation memory cell at the time step t

$$i_t = \sigma\left(W_{xi} \cdot x_t + W_{hi} \cdot h_{t-1} + W_{ci} \cdot c_{t-1} + b_i\right) \tag{2}$$

$$c_t = i_t \cdot \tanh\left(W_{xc} \cdot x_t + W_{hc} \cdot h_{t-1} + b_c\right) + f_t \cdot c_{t-1} \tag{3}$$

$$o_t = \sigma\left(W_{xo} \cdot x_t + W_{ho} \cdot h_{t-1} + W_{co} \cdot c_t + b_o\right), \tag{4}$$

$$h_t = o_t \cdot \tanh\left(c_t\right), \tag{5}$$

where: σ = non-linearity sigmoid function, i_t = input gate, W = weight matrix, x_t = time t input, c_t = cell state, h_t = time t output, f_t = forget gate [1], o_t = output gate, h_{t-1} = hidden state vector of the previous time step, and b_i = input bias vector [25].

During model training, the LSTM network is optimized using the backpropagation algorithm, and the structure of the LSTM hidden layer unit comprises of addition and multiplication operations, and several active layers. This minimizes the RNN drawback of gradient vanishing problems. The implementation of the LSTM is detailed in [13, 21].

Wavelet neural networks

Wavelet transform (WT) WT of a signal is the representation of the data in terms of the time–frequency domain. Through the transformation, the noise components are removed and the signals are decomposed into high- and low-frequency through the high-pass and low-pass operations. In the trend analysis of time-series data, wavelet transform is useful for the effective capture of the inherent and hidden characteristics and trends, as well as in detecting localized and non-stationary of the events. It is proposed in this study that WT is capable of detecting the non-stationarity phenomena in rainfall and runoff data by representing the original signal in low- and high-frequency data components.

As detailed in [25, 36], a mother wavelet function is constructed for the wavelet function. If $\varphi(t)$ is an integrable square function, with $\varphi(t) = L^2(R)$, and if its Fourier transform $\Psi(\omega)$ satisfies the compatibility condition (Eq. 6):

$$\int_R \frac{|\Psi(\omega)|^2}{\omega} d\omega < \infty, \tag{6}$$

then $\varphi(t)$ is the mother wavelet. Translation (τ) and scale (a) factors of WT are made so that we get function $\varphi_{a,\tau}(t)$ as in Eq. 7:

$$\varphi_{a,\tau}(t) = a^{1/2}\Psi\left(\frac{t-\tau}{a}\right), a > 0, \tau \in R, \tag{7}$$

$\Psi_{a,\tau}(\omega)$ = continuous wavelet, and the inner product of the input signal $x(t)$ and $\varphi(t)$ is calculated as in Eq. 8, and its Fourier transform time-domain in Eq. 9.

$$f_x(a, \tau) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} x(t)\varphi^* \left(\frac{t - \tau}{a} \right) dt, \qquad (8)$$

$$f_x(a, \tau) = \frac{\sqrt{a}}{2\pi} \int_{-\infty}^{\infty} X(\omega)\Psi^*(a\omega)e^{j\omega\tau} d\omega. \qquad (9)$$

In the implementation of WT, the success is based on the selected mother wavelet. In hydrological time-series modelling, the Daubechies (DAUB-N) wavelets have proven to be more effective due to the orthonormality properties and balance between information quality conservation and abundance [37, 38, 46]. DWT was used in this study to decompose the rainfall and runoff data using the Daubechies level 4 (DB4), since DB4 is able to minimize the noise but does not oversmoothen the signal [37]. More details on the development and implementation of DWT are in presented our previous studies [25, 36, 47].

To accomplish the characterization and detection of localized phenomena of non-stationary time-series data, the first step is the decomposition of the measured discharge $\{D_{d1}(t), D_{d2}(t), \ldots, D_{di}(t), D_a(t)\}$ and rainfall $\{R_{d1}(t), R_{d2}(t), \ldots, R_{di}(t), R_a(t)\}$ to multi-frequent data, where $D_{di}(t)$ and $R_{di}(t)$ are the resulting DWT details and $D_{ai}(t)$ and $R_{ai}(t)$ represent the approximation of time-series discharge (D) and rainfall (R). The detail i defines the ith approximation level of the decomposed data.

WT-neural network (WNN) model Inheriting the properties of wavelets and ANN, the topology of WNN is based on feedforward backpropagation (multilayer perceptron) network, with the mother wavelet acting as the hidden-layers transfer function. The WNN network topology is shown in Fig. 5a and the implementation strategy in Fig. 5b. The feedforward multilayer perceptron with input layer, hidden layers and output layer was adopted for the rainfall and runoff signals decomposed by the wavelet transform into approximation [$D_{ai}(t)$ and $R_{ai}(t)$] and detail [$D_{di}(t)$ and $R_{di}(t)$] coefficients.

Before selecting the DB4, levels 1–10 were compared by trial-and-error and level 4 was adopted on the basis of the size of the validation data [53–55]. From the wavelet decomposition, several sub-series from the original data were obtained as input variables to the feedforward MLP (Fig. 5a). Figure 5c, d shows sample inputs following DB4 decomposition of the original max temperature and rainfall for station 8,934,023. Sample station 8,934,023 is chosen as it represents the middle elevation of the catchments area. The input includes the level 4 approximation of the original signal and the 4-level details (D1–D4).

From the input vector, hidden layer output is determined as in Eq. 10:

$$h(j) = h_j \left[\frac{\sum_{i=1}^{n} \omega_{ij} x_i - b_j}{a_j} \right], \qquad (10)$$

where $h(j)$ = node; j = hidden layer nodes; h_j = mother wavelet; w_{ij} = input and hidden layer connecting weight; b_j = translation factor, and a_j = scaling factor for h_j.

Adopting the DB4 [25, 36], the output layer is determined as in Eq. 11:

$$y(k) = \sum_{i=1}^{k} \omega_{jk} \times h_j \left[\frac{\sum_{i=1}^{n} \omega_{ij} x_i - b_j}{a_j} \right] = \sum_{i=1}^{k} \omega_{jk} h(j). \qquad (11)$$

The updating of the WNN weights and the wavelet function parameters is as the following steps:

Step 1: WNN prediction error $E(W)$ computation:

$$E(W) = e = \sum_{1}^{k} \left[y_t(k) - y(k) \right]^2, \qquad (12)$$

where $y(k)$ = prediction output value and $y_t(k)$ = target output.

Step 2: WNN weights update and variation of wavelet according to the prediction e:

$$\begin{aligned} \omega_{n \cdot k}^{(i+1)} &= \omega_{n \cdot k}^{(i)} + \Delta \omega_{n \cdot k}^{(i+1)} \\ a_k^{(i+1)} &= a_k^{(i)} + \Delta a_k^{(i+1)} \\ b_k^{(i+1)} &= b_k^{(i)} + \Delta b_k^{(i+1)} \end{aligned} \qquad (13)$$
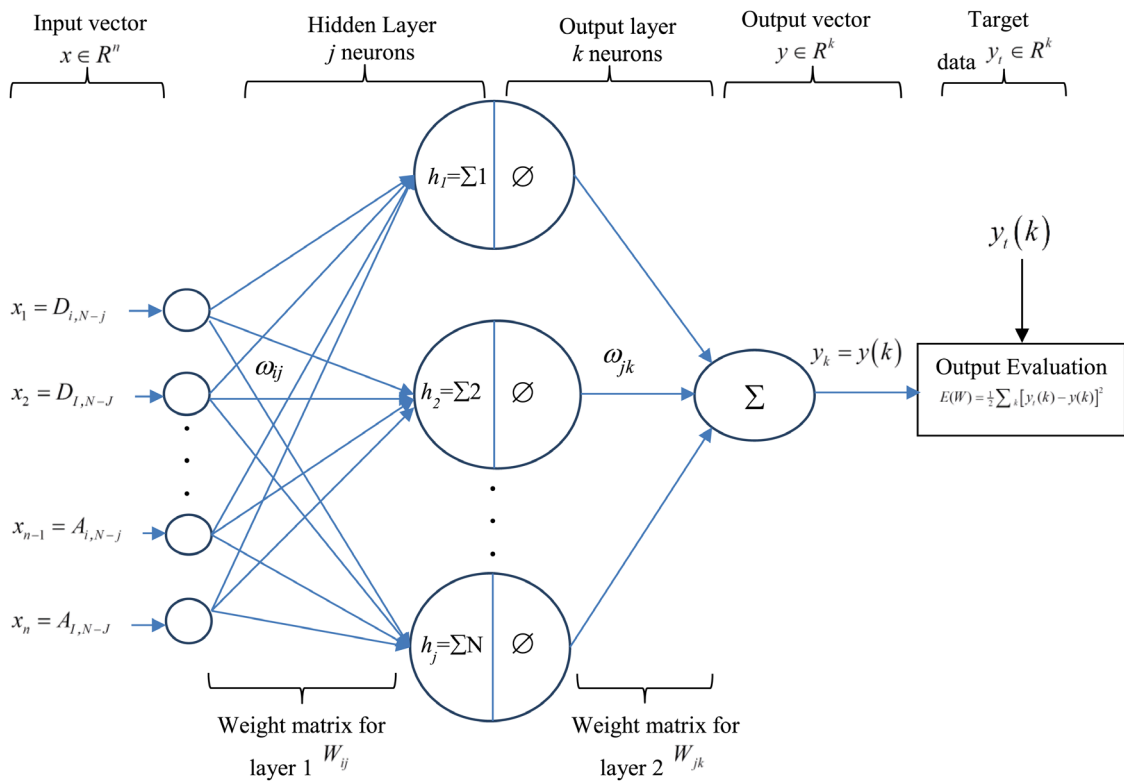
where $\Delta \omega_{n.k}^{(i+1)}$, $\Delta a_k^{(i+1)}$ and $\Delta b_k^{(i+1)}$ are calculated by prediction error of the network:

$$\begin{aligned} \Delta \omega_{n \cdot k}^{(i+1)} &= -\eta \frac{\partial e}{\partial \omega_{n,k}^{(i)}} \\ \Delta a_k^{(i+1)} &= -\eta \frac{\partial e}{\partial a_k^{(i)}} \\ \Delta b_k^{(i+1)} &= -\eta \frac{\partial e}{\partial b_k^{(i)}}, \end{aligned} \qquad (14)$$
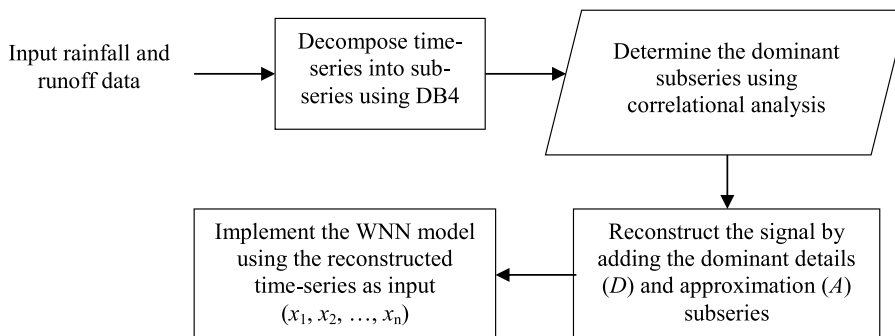
with η as the network learning rate.

The training of the network comprises of the following steps [48]:

1. Data pre-processing: normalized data division into training (70%), testing (15%) and validation (15%) datasets. The validation is part of training the model and updating the parameters. It utilizes part of datasets to validate and update the model parameters after each training epoch.
2. Network initialization: random initialization of weights, translation, translation and scale factor, and the learning rate $\{\omega_{ij}$ and $\omega_{jk}, b_k, a_k, \eta\}$.
3. Network training: training, prediction and prediction error e estimation between output and expected value.

Input vector
$x \in R^n$

Hidden Layer
j neurons

Output layer
k neurons

Output vector
$y \in R^k$

Target
data $y_t \in R^k$

(a)

(b)

Fig. 5 **a** Wavelet-based feedforward multilayer perceptron (MLP) neural network layer structure. n is the input vector data which comprise of the wavelet details (D) and approximation (A) as: $x_1 = D_{i,N-j}$, ...$x_n = A_{I,N-J}$. **b** Model development for wavelet coupled artificial neural network (WNN). $[x_i]$ and $[y_i]$ are respectively the input and output predicted vector. **c, d** Original and DB4 approximation (low-frequency) and detail coefficients (high-frequency), respectively, for mean monthly temperature and precipitation at station 8,934,023

4. Weights updating: parameter and network weights update depending on magnitude of e.
5. Network testing: use test dataset for network reliability testing, else iterate to Step 3.

Input data quantification and normalization

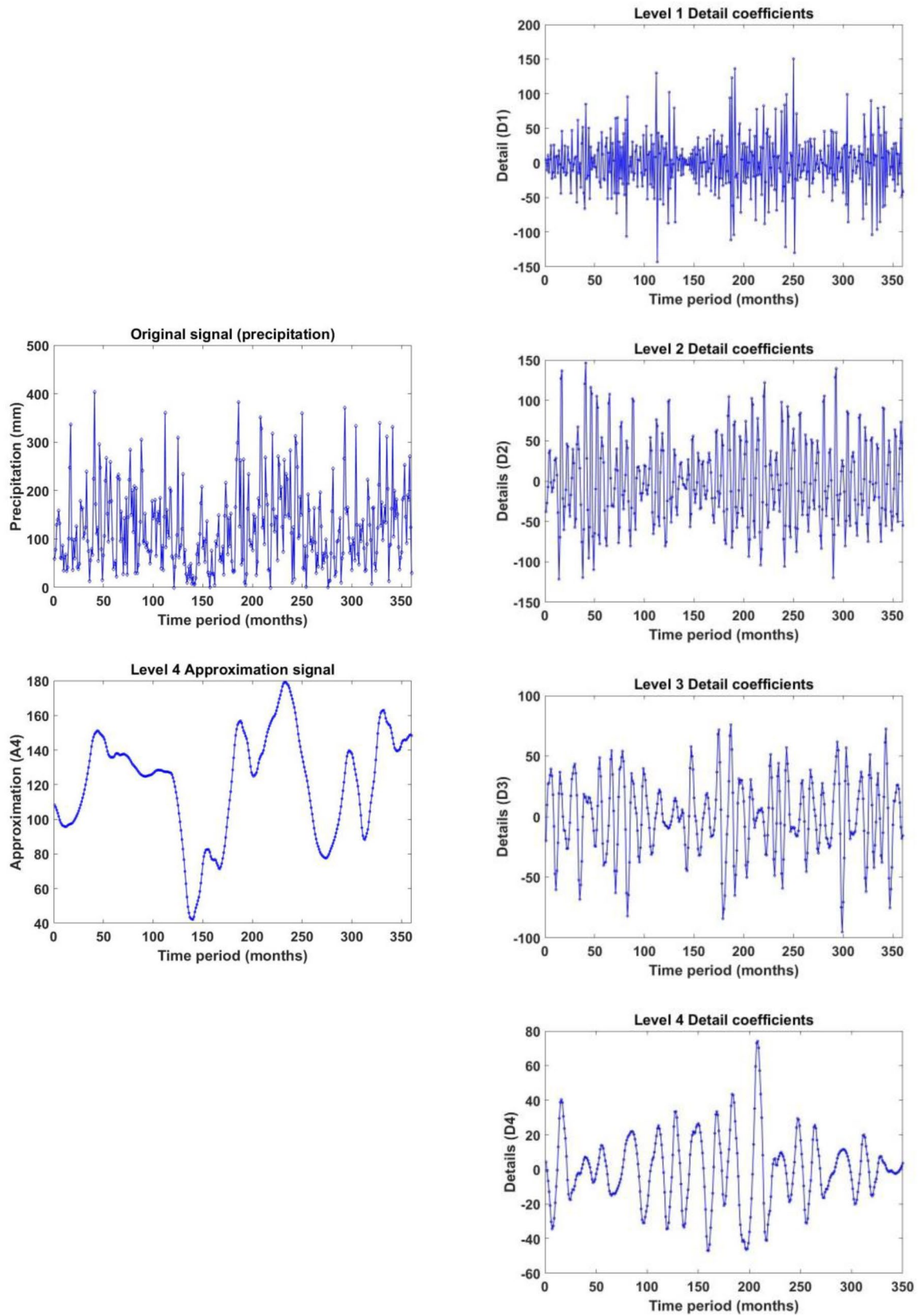The [1] normalization based on *min–max* predefined boundary method (Eq. 15), was used to linearly transform the original data and to maintain the inherent relationships within the respective datasets [56]. After the test output, denormalization is carried out to be able to relate the prediction output with the observed data.

$$f : x \to y\left(\frac{x - x_{\min}}{x_{\max} - x_{\min}}\right), \tag{15}$$

(c)
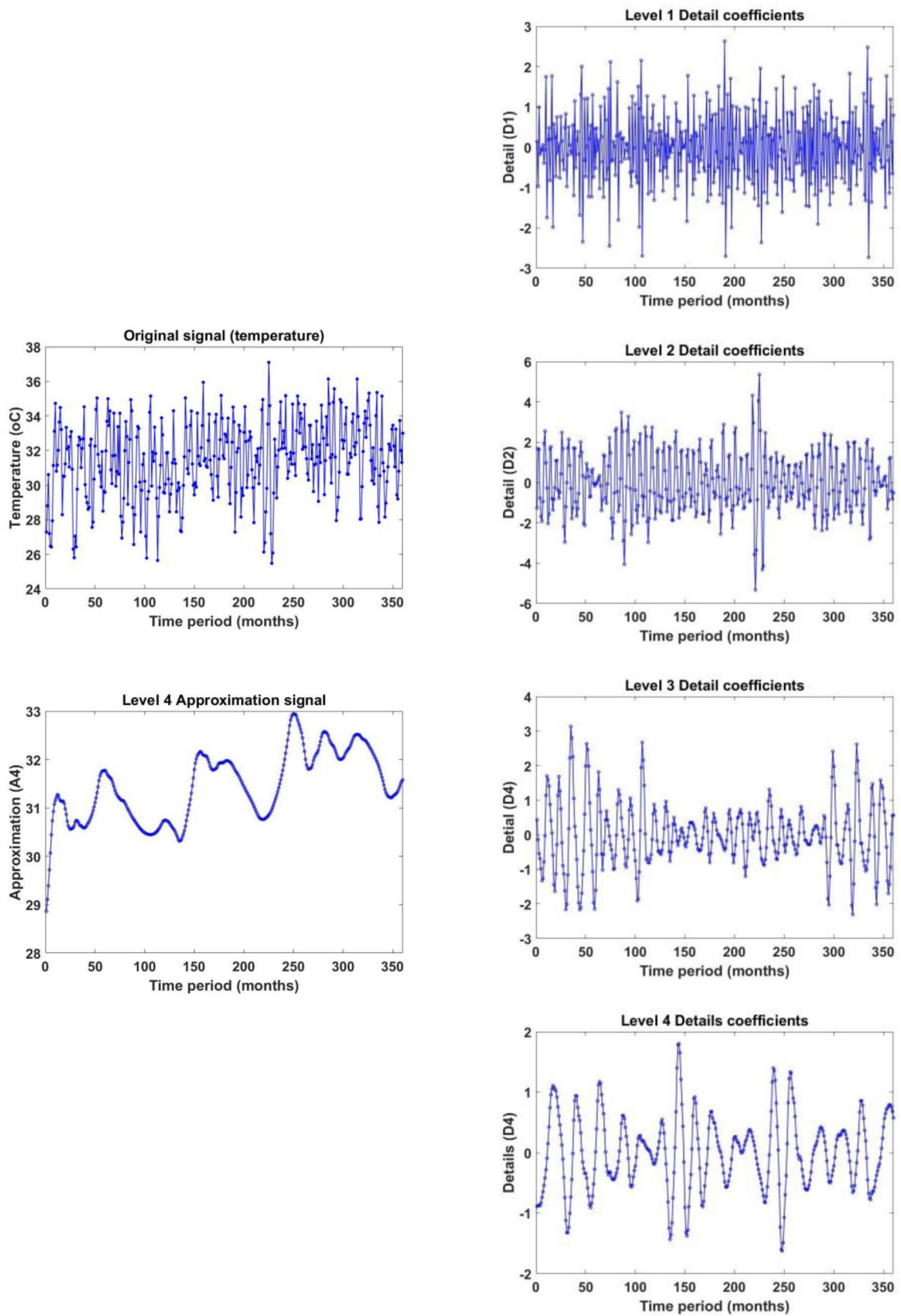
Fig. 5 (continued)

Fig. 5 (continued)

where $x, y \in R^n$, $x_{\min} = \min(x)$, $x_{\max} = \max(x)$ and $x =$ input data. The values are converted into the range [1] through the normalization, that is $x_{\max} \in [0, 1]$, $i = 1, 2,...,n$.

Metrics for model performance evaluation

To compare and evaluate the models, the following statistical measures were used R^2, *RMSE* and *MAE* (Eqs. 16–18):

$$R^2 = \frac{\left(\sum_{i=1}^{n} \left(P_i - \overline{P} \right) \left(P_i' - \overline{P'} \right) \right)^2}{\sum_{i=1}^{n} \left(P_i - \overline{P} \right)^2 \sum_{i=1}^{n} \left(P_i' - \overline{P'} \right)^2}, \tag{16}$$

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{n} \left(P - P_i' \right)^2}{n}} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} e_i^2}, \tag{17}$$

$$\text{MAE} = \frac{\sum_{i=1}^{n} \left| P_i - P_i' \right|}{n} = \frac{1}{n} \sum_{i=1}^{n} \left| e_i \right|, \tag{18}$$

where $P_i =$ observed data $P_i' =$ simulated data; $\overline{P} =$ mean observed data; $\overline{P'} =$ mean simulated data and $e =$ model errors.

Results and discussions

Neural network optimum architecture for rainfall and runoff prediction

The construction of LSTM and WNN model architectures comprises of the creation of the topology of the deep learning network which is significantly determined by the hidden layer neurons and the selection of the optimal training parameters. Through trial-and-error input parameter combinations, the output performance of hidden layers is determined using R^2, MAE and RMSE (Table 2). To determine the optimal architecture in rainfall and runoff predictions, the hidden layers were varied from 1 to 5 layers, with 30 neurons in each layer and for each sub-basin. The training and validation results from LSTM and WNN models are summarized in Table 2, and are based on all the five input parameters in the entire basin.

The prediction results in Table 2 using R^2 evaluation shows that the performance of the LSTM and WNN models improved with increase in the number hidden layers and corresponding neurons. Notably, after the fourth hidden layer, there was observed to be a consistent decrease in the prediction performance for the models. The LSTM prediction of rainfall as measured with R^2 increased from 0.6448 with 1 hidden layer of 30 neurons, to a maximum of 0.8610 with a topology of 4 hidden layers with 30 neurons each. At one hidden layer, WNN marginally outperformed the LSTM by about 6%, however with four hidden layers LSTM performed better than the WNN by 8% as measured in terms of R^2. Similar patterns in the results were also observed for runoff prediction with the tested network topological structures as in Table 2. The MAE and RMSE performance metrics showed that with increase in neurons, WNN tended to marginally minimize the prediction errors in comparison with the LSTM neural network model. It is considered that both WNN and LSTM are capable of predicting the rainfall and runoff trends, however with deep learning structure, LSTM marginally outperformed the WNN.

Notably in Table 2, MAE for rainfall prediction using LSTM with three hidden layers received the lowest error, while the opposite observations are obtained when using RMSE evaluation and when using WNN with two hidden layers. Similarly for runoff predictions using LSTM with four hidden layers and WNN with three hidden layers, the inverse variabilities in MAE and RMSE with increase in

Table 2 Performance statistics of LSTM and WNN hidden layer architectures for rainfall and runoff simulations

ANN model	Number of hidden layers	Rainfall			Runoff		
		R^2	MAE (mm)	RMSE (mm)	R^2	MAE (m^3 s^{-1})	RMSE (m^3 s^{-1})
LSTM	1	0.6448	13.3589	29.1334	0.6321	18.2446	15.7903
	2	0.7561	11.4248	26.3519	0.7468	12.0772	15.2249
	3	0.8534	8.3901	25.0677	0.8900	10.5237	14.5134
	4	0.8610	13.5769	20.5492	0.9249	10. 2213	13.0459
	5	0.8577	13.9505	20.5833	0.8981	11.7404	13.0520
WNN	1	0.7219	33.4401	27.6993	0.8209	12.5991	16.4821
	2	0.7404	21.0580	26.4530	0.8243	13.1103	17.5491
	3	0.7187	28.2397	26.5166	0.8561	11.4854	17.1507
	4	0.7825	24.4877	25.4206	0.8579	12.3735	15.2128
	5	0.7812	25.9110	26.0488	0.8447	12.9546	15.2312

hidden layers are observed. These observations could be attributed in part to the fact that for the same data, the differences in the observed RMSE and MAE may arise when the error distributions in the data are biased or non-Gaussian. Further, while the MAE gives the same weight to all errors, the RMSE penalizes variance as it gives errors with larger absolute values more weight than errors with smaller absolute values. This could be the cause of the varied prediction results in Table 2 for rainfall and runoff using LSTM and WNN.

To further assess the difference in performance between LSTM and WNN in rainfall and runoff trend prediction in terms of computing time, the variations of the epochs and the RMSE as the standard deviation of the prediction residuals is presented in Fig. 6. In the prediction of rainfall, RMSE converged to a minimum of 14.55 mm in 31st epochs for LSTM and at 15.17 mm in 37th epochs using WNN (Fig. 6a). Further training of the networks towards loss minimization and to obtain possible higher accuracy resulted in an increase in RMSE and reaches a saturation point after 43 epochs with WNN.

In the prediction of runoff, WNN is observed to consistently overestimate the mean runoff within the basin with a minimum RMSE of 15.17 m^3/s at 34th epochs (Fig. 6b). Using LSTM, the minimum RMSE is at approximately 13.12 m^3/s was achieved between the 25th–35th epochs. The results further confirm that both models are suitable for the prediction of rainfall and runoff trends with reasonable computing time.

Figure 6c, d shows the accuracy performance the LSTM and WNN models calculated for 50 iteration epochs using training and validation datasets for prediction of rainfall and runoff for the entire basins. In general, the model accuracies on the training and validation datasets increases after each
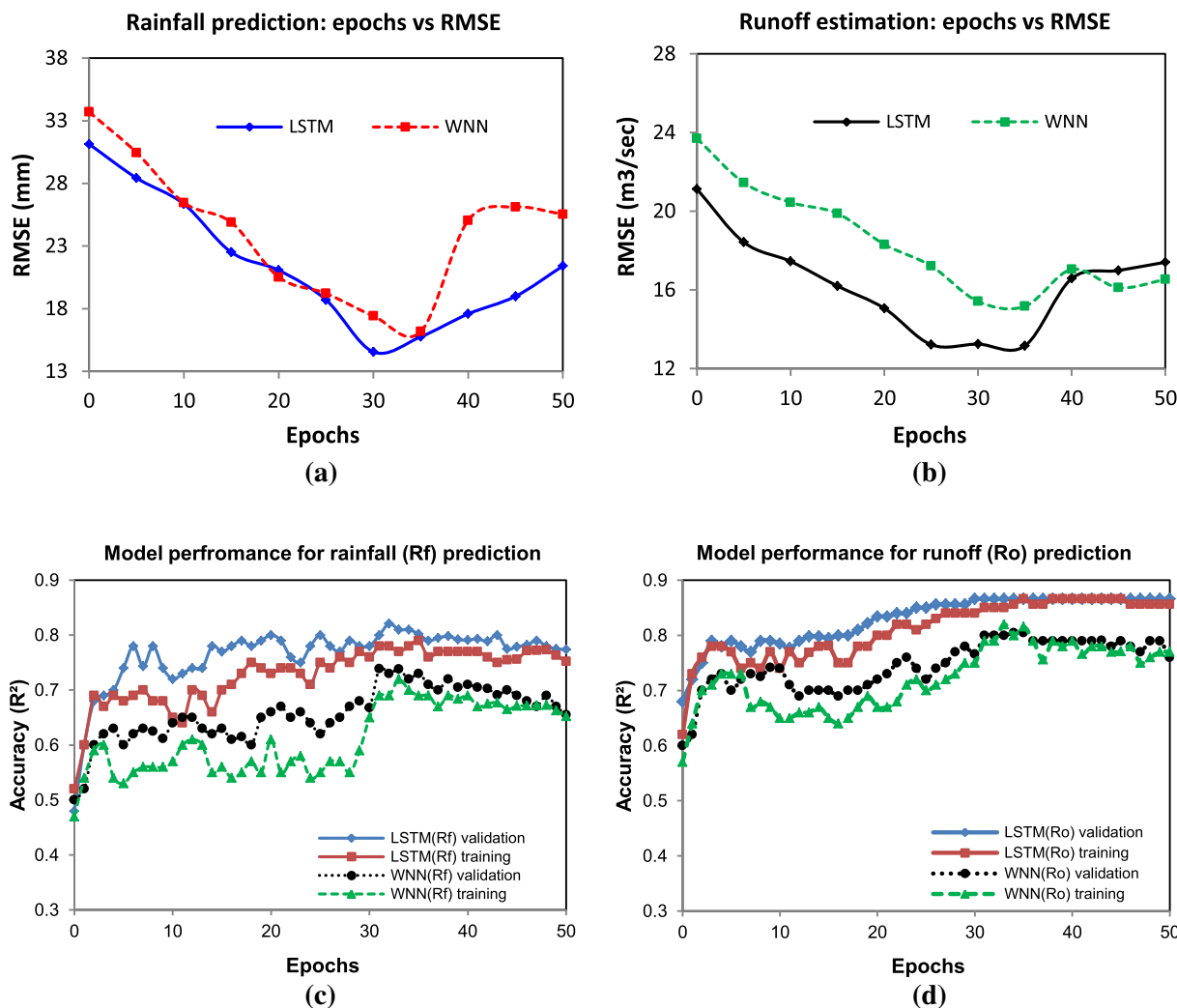

Fig. 6 **a**, **b** RMSE vs epoch for model learning with LSTM and WNN models in predicting the mean monthly rainfall and runoff. **c**, **d** Accuracy performance for the prediction of rainfall and runoff using LSTM and WNN models

iteration with fluctuations, which could be attributed to some randomnesses in the network. As the model trains during the first pass through the data, both training and validation accuracy increases indicating that the model is learning the structure of the rainfall and runoff prediction data as well as the temporal correlations of the time-series. In the first and consecutive iterations, the validation accuracy did not increase significantly and always higher than the training accuracy. This indicates that the network did not overfit the training data and accurately generalized to the unseen validation data. The optimal accuracy was obtained between 30 and 40 epochs with LSTM being higher WNN in predicting rainfall and runoff respectively by approximately 10% and 5%. The prediction of runoff was consistently recorded with a higher R^2 accuracy as compared to rainfall using both WNN and LSTM (Fig. 6c, d).

Runoff prediction with LSTM and WNN

Runoff prediction results using the LSTM and WNN models

Adopting the optimal four hidden-layer configuration, the runoff prediction results at the three discharge stations (1BC01, 1DA02 and 1EF01) using the LSTM model is presented in Fig. 7. Similarly, the runoff prediction results using WNN for the three discharge measurement stations are shown in Fig. 8. For both models, the input consisted of the mean monthly satellite-based meteorological datasets.

A statistical comparison of the runoff prediction results from LSTM and WNN are presented in Table 3 in terms of R^2, MAE and RMSE. Except for station IBC01 where WNN predicted the runoff with the R^2 of 0.7820, both models predicted the runoff at the three stations with R^2 greater than 0.80 using the five input parameters as input (Fig. 9). This is evidenced in the 30-year prediction accuracy with the MAE and RMSE of less than 13 m^3/s for both models. In overall for the entire basin, it is observed that LSTM marginally outperformed the WNN model, with MAE $= 11.1452$ m^3/s and RMSE $= 12.1933$ m^3/s at the basin outlet 1EF01. In practical applications, it is conclusive that both models can be used in the prediction of runoff in data scarce basins.

A comparison of the goodness-of-fit for the prediction of streamflow runoff with the two models as presented in Fig. 9 shows that for the three stations, the use of satellite data to predict streamflow is acceptable as the R values were more than 85%. The accuracy of streamflow prediction is observed to increases with increase in the number of prediction stations within the sub-basins.

Performance of individual meteorological factors in runoff prediction

To determine the significance and accuracy of the contributions of the satellite-based meteorological parameters in the prediction of runoff, each of the five parameters were used as independent inputs with runoff as output. The comparative output results are presented in Fig. 10 representing the runoff predictions for the three discharge stations.

The results in Fig. 10 show that rainfall is the highest contributing indicator variable in runoff prediction with $R^2 > 0.8$ for the three discharge stations. This confirms the fact that the amount of rainfall that remains after storage, infiltration, interception, evaporation and transpiration contributes to runoff. The least contributing meteorological factor is the relative humidity with R^2 ranging between 0.6 and 0.65 for the three discharge stations using LSTM and WNN. The rest of the parameters, average temperature, wind speed and solar radiation estimated the runoff in the three stations with R^2 ranging between 0.675 and 0.80, with temperature performing better than wind speed and solar radiation. Conclusively, as the input increases from rainfall to all the datasets (1–5) and the hidden layers increased from 1 and 4, the accuracy of runoff prediction is observed to increase for the model training, testing and validation by up to 10%. Further investigations on the explanation of the basis of the individual predictions by comparing the contribution of each feature to each prediction using a unified approach such as SHapley Additive exPlanations (SHAP) [57] is recommended.

Rainfall prediction using LSTM and WNN

Performance of individual meteorological parameters for rainfall prediction

In evaluating the significance of the satellite-based meteorological parameters in rainfall trend prediction, Table 4 presents the performance results of basin mean monthly rainfall prediction using the different meteorological parameters. Satellite-based precipitation is observed to be the most significant predictor in estimating the measured rainfall with $R^2 > 0.8$ and the least MAE and RMSE errors measures. This is contributed to by the fact that for medium sized and climatically homogenous basins like the Nzoia Basin, the climate factors tend to be replicated throughout the catchment area with minimal variabilities. As such the occurrence of rainfall at one station is generally an indication of rainfall also being recorded at a near distant station within the basin.

Temperature is the second best predictor for rainfall prediction, and the effect of temperature on rainfall arises from the fact that increased temperature leads to increased evaporation, an accelerated rate of the hydrological cycle
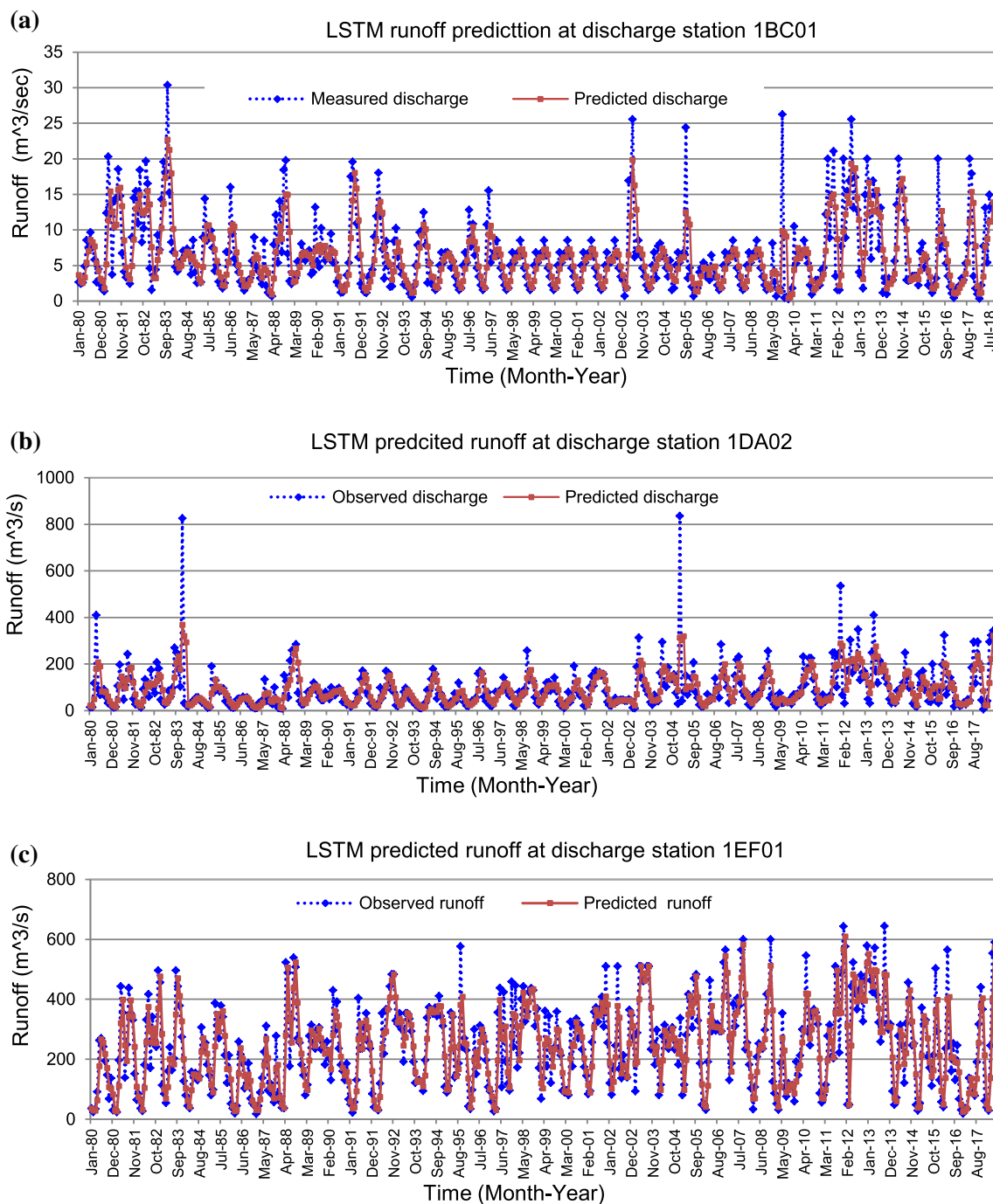
(a)

(b)

(c)

Fig. 7 Observed and LSTM predicted discharge at stations: **a** 1BC01, **b** 1DA02 and **c** 1EF01

and more precipitation especially during the wet season. Humidity, wind speed and solar radiation are consecutively ranked as in Table 4 with nearly the same contribution effects on rainfall prediction, implying that they are highly correlated within the basin in terms of their contribution in rainfall prediction. This is also attributed to the size of the basin and the fact that the climate factors are nearly similar within the basin.

Rainfall prediction with combined satellite-based meteorological data

Results of the mean monthly predicted rainfall for the four stations using the five meteorological datasets are

(a)

(b)

(c)
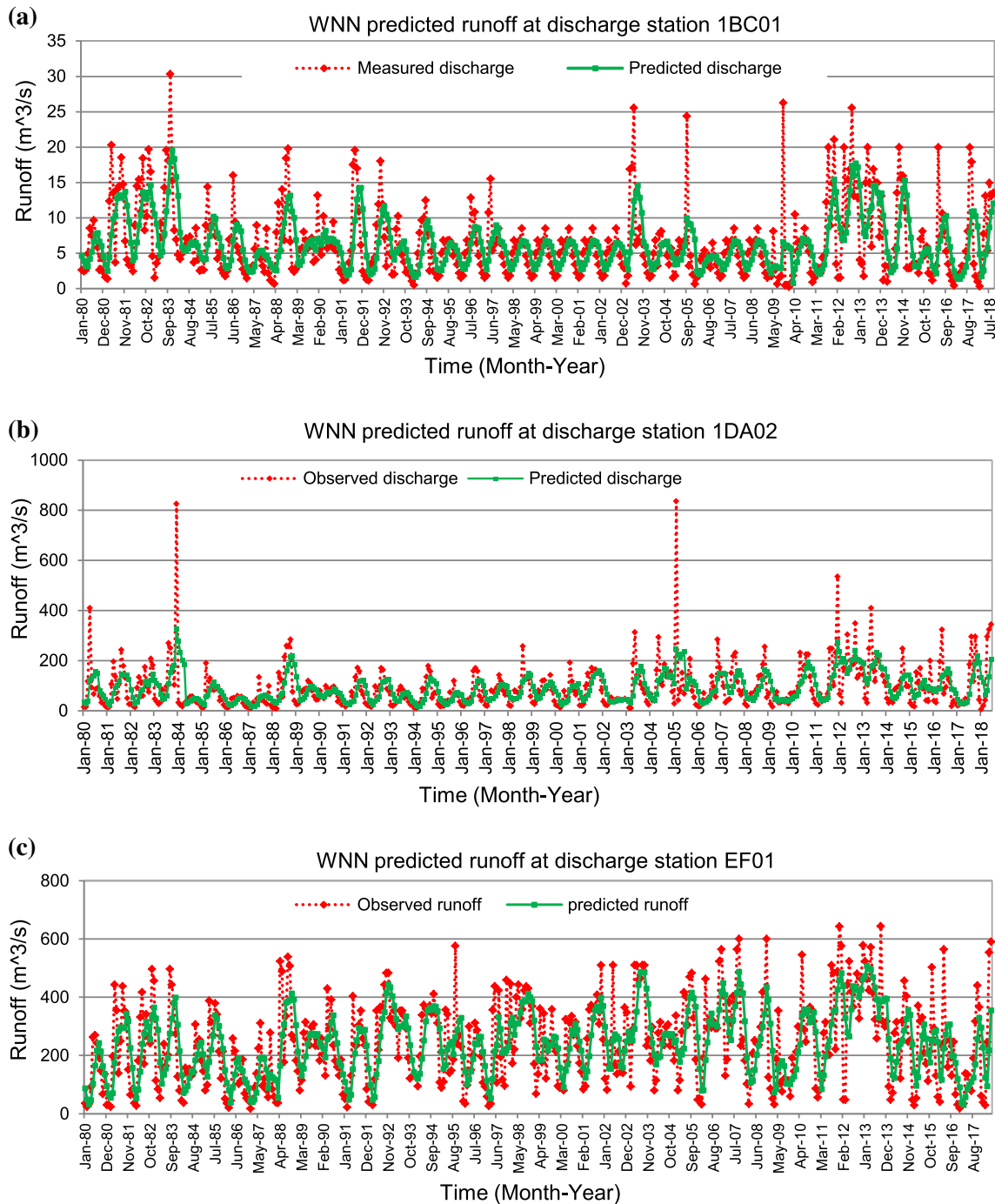
Fig. 8 Observed and predicted discharge at stations: **a** 1BC01, **b** 1DA02, and **c** 1EF01 using the WNN model

Table 3 Performance evaluation for runoff prediction with LSTM and WNN using the meteorological parameters

Station ID	LSTM model			WNN model		
	R^2	MAE ($m^3 s^{-1}$)	RMSE ($m^3 s^{-1}$)	R^2	MAE ($m^3 s^{-1}$)	RMSE ($m^3 s^{-1}$)
1BC01	0.8017	12.0890	12.1450	0.7820	12.6490	12.8793
1DA02	0.8330	12.0971	13.1318	0.8812	12.5637	12.7465
1EF01	0.8967	11.1452	12.1933	0.8053	12.4917	12.3040

Springer
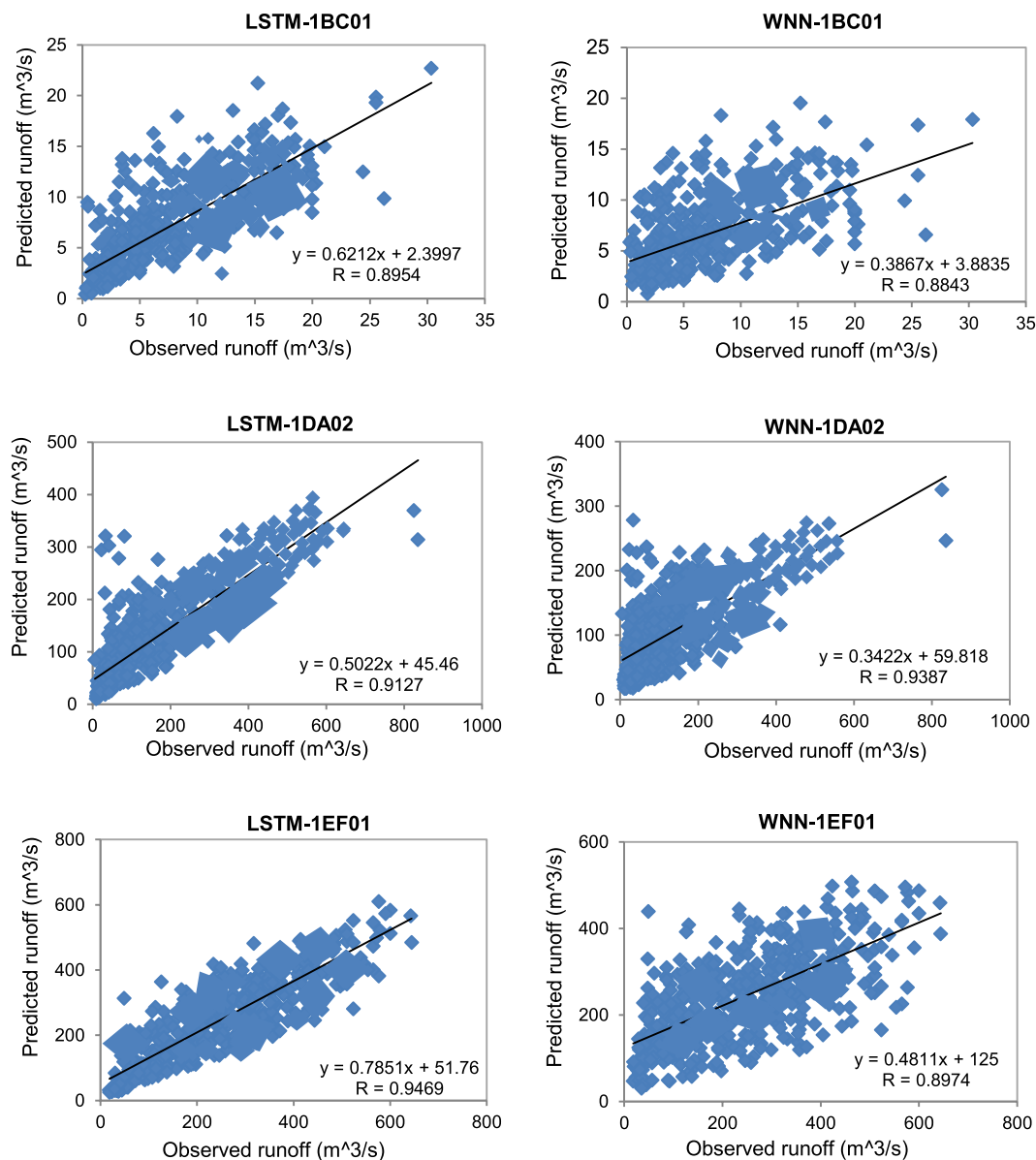
Fig. 9 Comparative regression models for the prediction of runoff with LSTM and WNN at the discharge stations 1BC01, 1DA02 and 1EF01

respectively presented in Figs. 11 and 12 for LSTM and WNN. The results are samples of distributed gauge stations representing high (station 8,835,034), mid (stations 8,934,071 and 8,934,023) and low (stations 8,934,059) elevation areas.

The coefficients of determination for the four representative stations 8,835,034, 8,934,071, 8,934,023 and 8,934,059 are presented in Fig. 13. The comparative performance between the two models imply that the LSTM predicted the basin mean rainfall with higher $R^2 = 0.8610$ for the ten stations, while WNN's prediction was at $R^2 = 0.7825$. The higher accuracy prediction results at individual rainfall

stations could be attributed to continuous and accurate gauge data.

In addition to the statistical performance evaluations, graphical comparisons of the observed and modelled runoff and rainfall with LSTM and WNN in Figs. 7, 8 and 11, 12, respectively, shows that the LSTM results matched closely in spatial position and trend with the observed data as compared to the WNN results. Further, the regression lines for runoff and rainfall in Figs. 9 and 13, respectively, shows that LSTM modelled the parameters closer to the 45° line of fit as compared to the WNN. The slightly lower performance from WNN results could be attributed to the feedforward ANN used in the training the input signals.
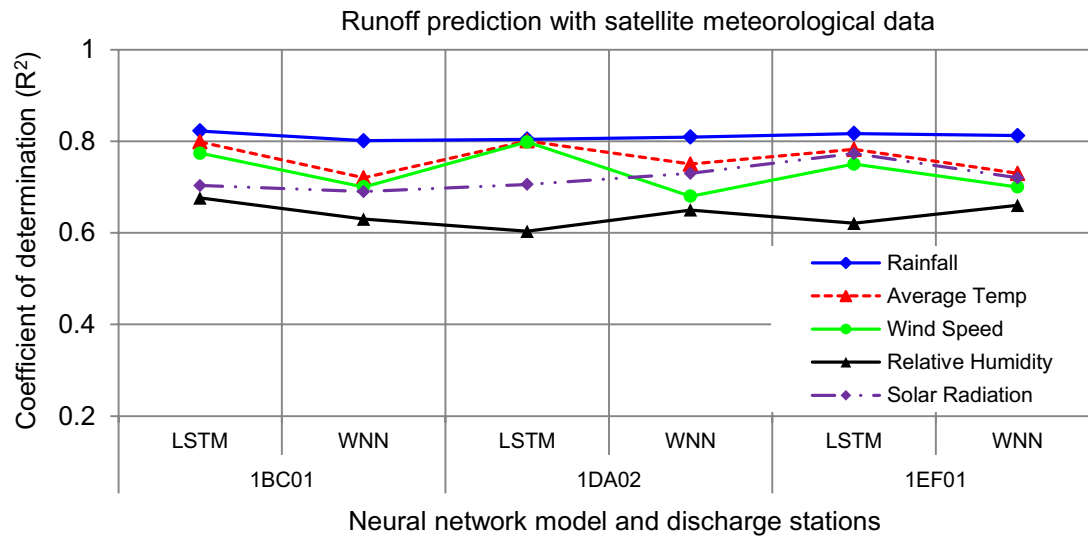
Fig. 10 Significance of individual meteorological factors in runoff predictions for discharge stations 1BC01, 1DA02 and 1EF01 using LSTM and WNN

Table 4 Performance of individual meteorological parameters for rainfall prediction

ANN model	Meteorological input parameter	Predicted rainfall		
		R^2	MAE (mm)	RMSE (mm)
LSTM	Rainfall	0.7824	9.0264	17.2264
	Average temperature	0.6938	10.9790	20.1791
	Relative humidity	0.6729	10.1826	20.2338
	Wind speeds	0.6551	9.2838	18.2254
	Solar radiation	0.6604	11.2005	21.2700
WNN	Rainfall	0.7013	11.9301	14.6411
	Average temperature	0.6490	10.0717	15.3029
	Relative humidity	0.6136	12.4593	15.6574
	Wind speeds	0.6050	11.8406	14.9878
	Solarradiation	0.6038	12.4955	16.1606

The study results on the prediction of rainfall and runoff trends show that LSTM and wavelet-based neural networks are able to overcome the timescale conversion problems in time-series data analysis for accurate forecasting [8, 9], as they are capable of capturing the quasi-periodic signals in long-term rainfall and runoff data which are also characterized by cyclical fluctuations with inherent noise [10–12]. The LSTM and WNN are considered to be superior to the conventional ANN models since they appear conserve the crucial information input data sequence order because of the deep learning process in the hidden layers [13, 21–23, 45].

According to [24], WNN as a data-driven model is able to take into account the non-stationarity in the assimilation of rainfall and runoff time-series data as they account for sequence dependency, periodicity and non-linearity in such data [24, 25, 39–41]. LSTM on the other hand can dynamically incorporate past learning experience due to internal

recurrence, thus presenting a powerful internal state for accurate learning and predictions in data with long delay and mixed frequencies [23, 26, 42, 43].

The distribution of the measured and predicted rainfall from the results in Figs. 11 and 12 for the year 1999 were spatially interpolated using ordinary Kriging [33], and the results presented in Fig. 14. The year 1999 is chosen because it had the most continuous measured precipitation in all the gauge stations within the basin, thus suitable for comparative analysis. Figure 14a presents the observed mean monthly precipitation in 1999 and the results from LSTM and WNN are respectively presented in Fig. 14b, c. It is observed that LSTM has the ability to accurately infer the long-term patterns in the 30-year rainfall data in most parts of the basin. As compared to the LSTM, WNN tended to overestimate the higher precipitation values and underestimate the lower precipitations. The results in Fig. 14 also illustrates that despite the good
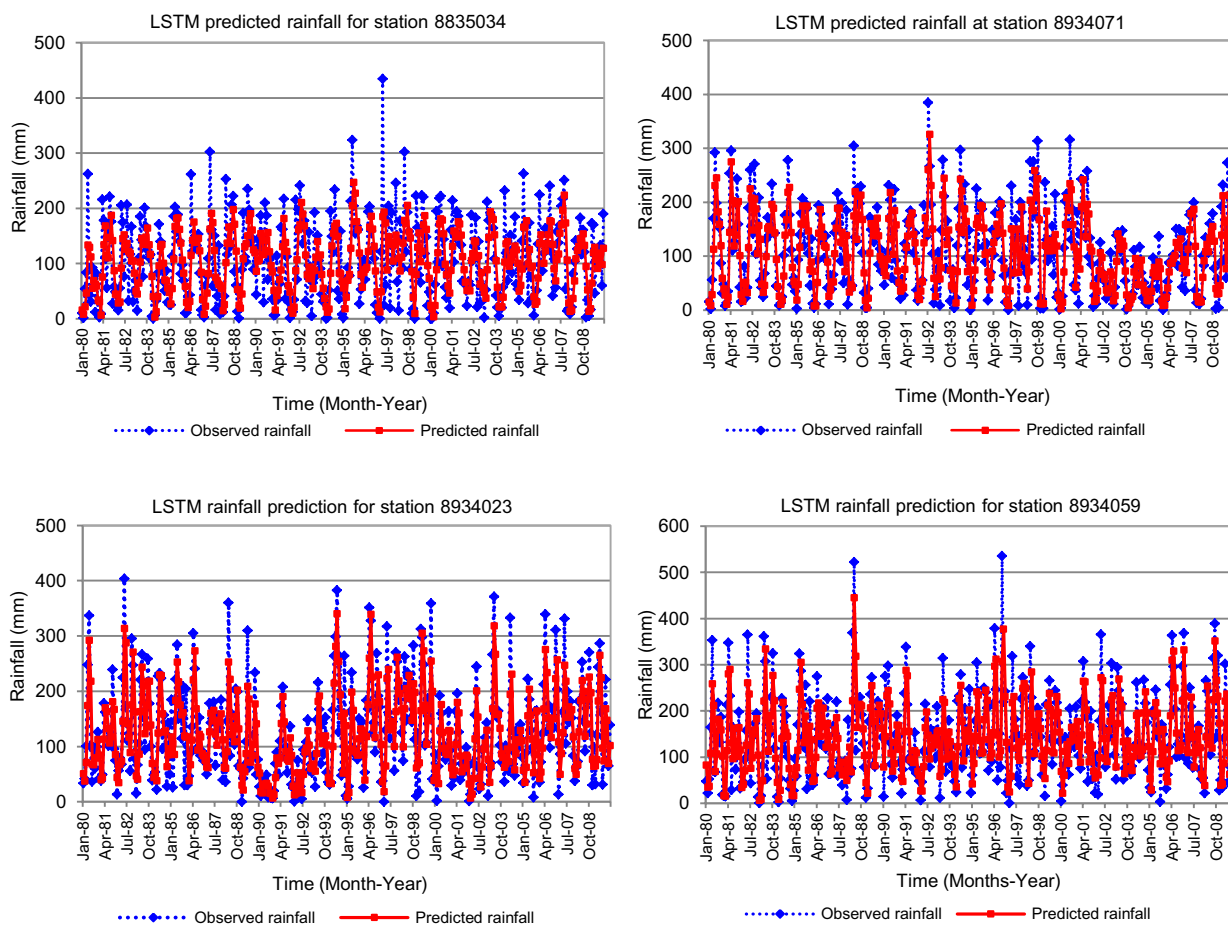
Fig. 11 Observed and LSTM predicted rainfall for four gauge stations within the basin

statistical evaluation results, the spatial representation of the phenomenon gives a more insightful area-based comparison of the results.

In the prediction of rainfall and runoff in hydrologic basins with scarce data, the LSTM model performed marginally better than the wavelet-based neural network model. Both the models displayed the capability to learn the inherent temporal dynamics in the time-series data, and also to capture the seasonality in the quasi-periodic rainfall and runoff data. The results show that optimized effectively, LSTM and WNN can resolve the non-stationarity and non-linearity problems associated with trend analysis of rainfall and runoff data.
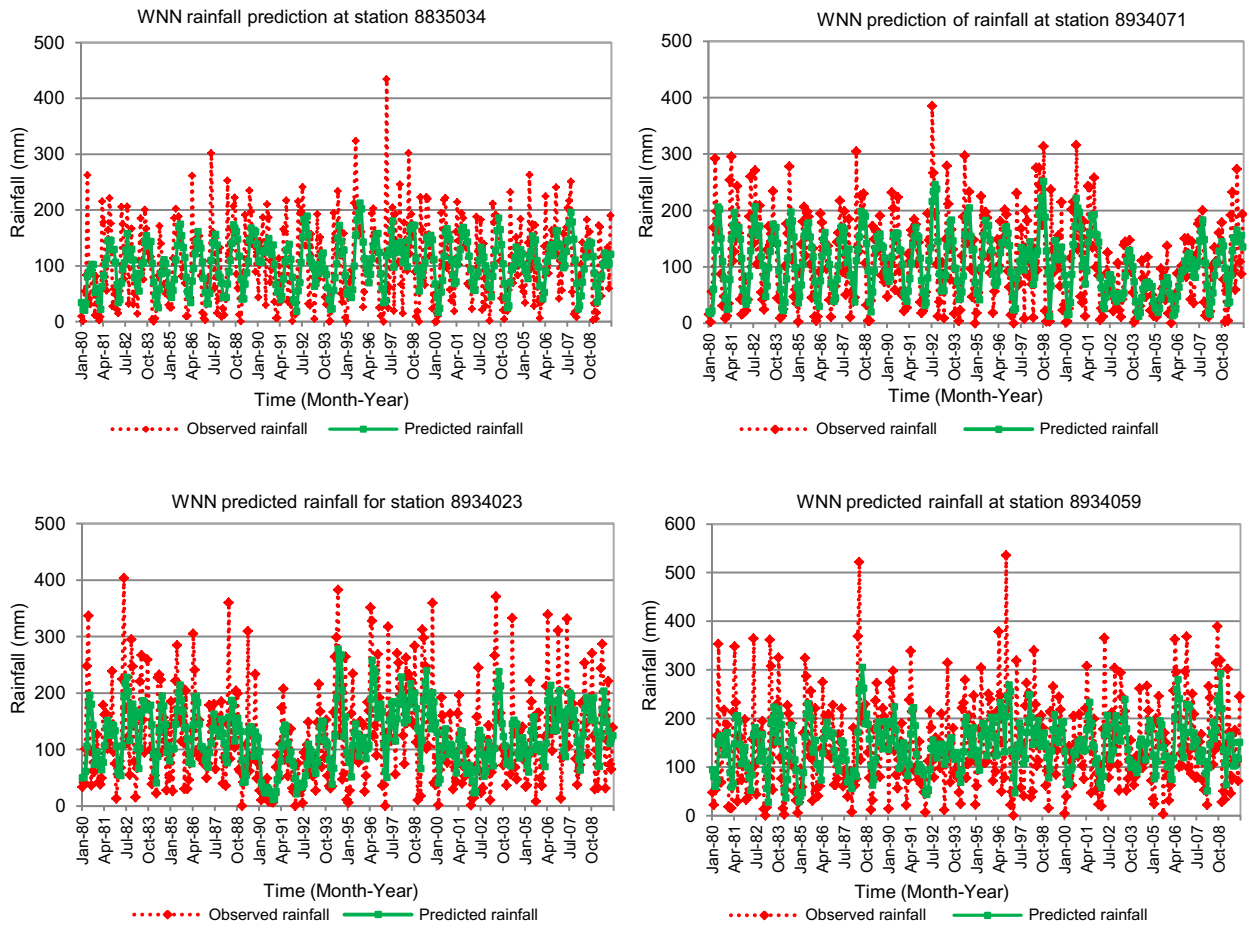
Fig. 12 Observed and predicted rainfall with WNN for four stations within the basin

Fig. 13 Regressions between the observed and predicted rainfall with WNN for four stations within the basin
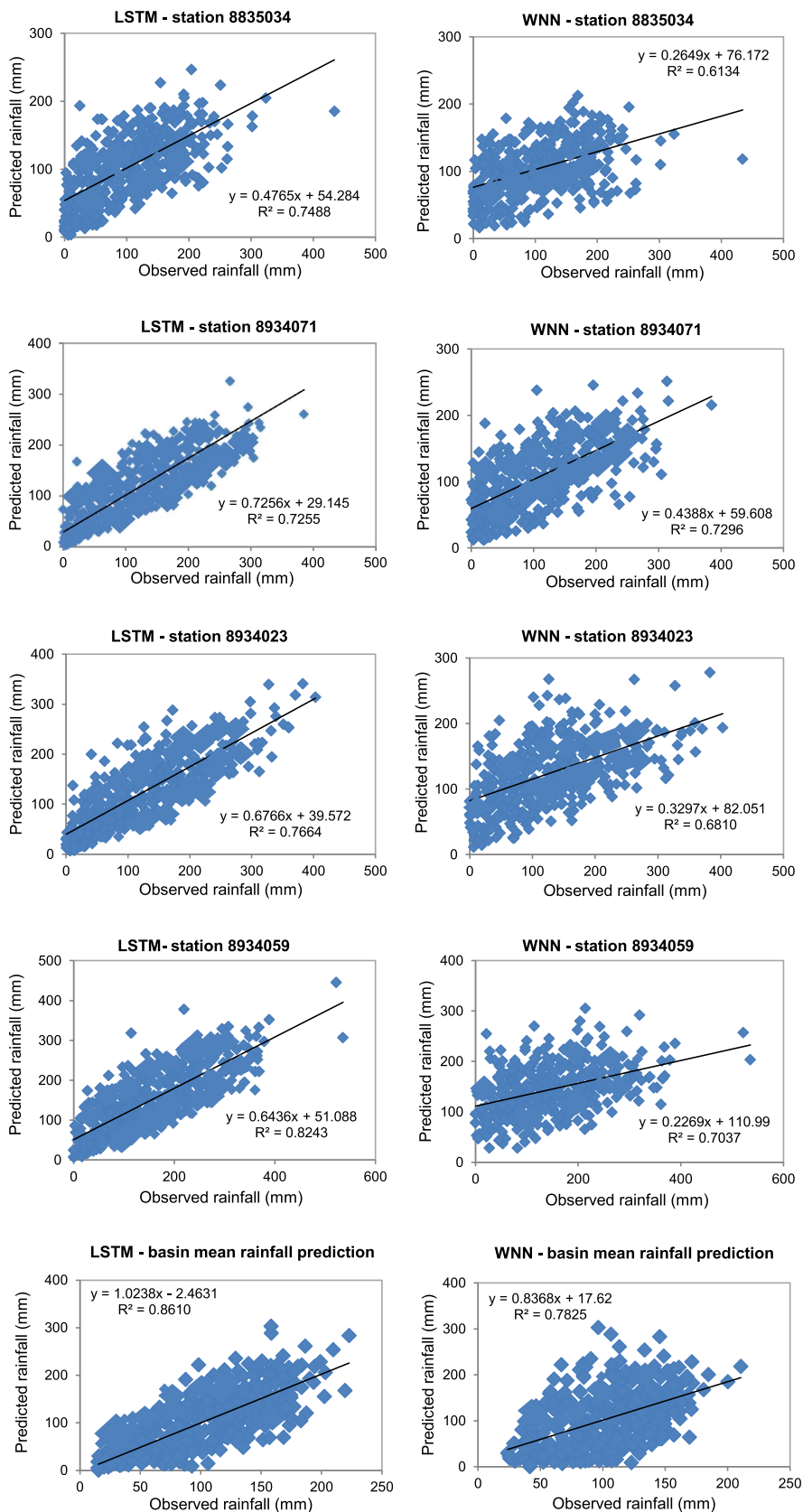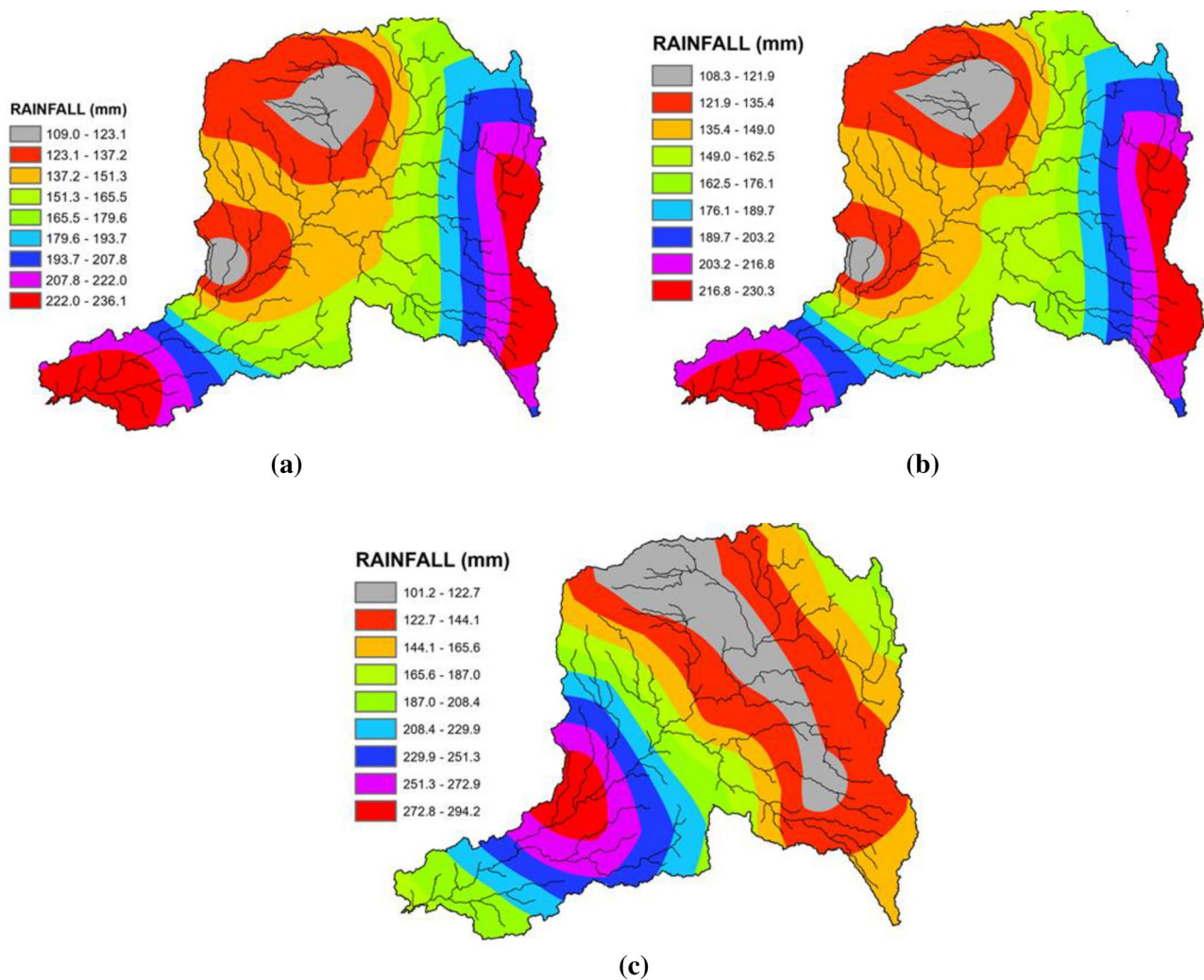
Fig. 14 Spatial distribution of **a** observed mean monthly rainfall in 1999, and the corresponding predicted rainfall from LSTM **b** and WNN)
models

Conclusions

This study presented a performance evaluation of recurrent neural network implementation based on LSTM architecture and wavelet neural network for rainfall and runoff trend analysis in sparsely gauged river basin. With 30 years (1980–2009) of measured rainfall and streamflow data, and satellite-based meteorological data comprising of precipitation, average temperature, relative humidity, wind speed and solar radiation from 10 satellite stations, the two neural network models were compared for rainfall and runoff prediction in the Nzoia hydrologic river basin in Kenya.

With the same optimal neural network topological structure of 4 hidden layers each consisting of 30 neurons, both LSTM and WNN models predicted runoff with average R^2 value of 0.8 for all the 3 stations, except station 1BC01 using WNN. The RMSE and MAE metrics from both models in

runoff prediction was achieved at less than 13 m^3/s for the 30-year study period. The evaluation of the significance of each meteorological parameter in the prediction of runoff showed rainfall as the most significant input parameter, followed by temperature, and solar radiation as the least contributing factor. Best results were obtained by including all the parameters in the prediction model. In the forecasting of rainfall, LSTM gave the best predictive results with $R^2 = 0.8610$ for the average monthly basin rainfall from the ten stations, with satellite-based precipitation being the best rainfall predictor. WNN estimated the mean basin rainfall with $R^2 = 0.7825$ using the five satellite data. At the sub-basins scale, it was observed that the performance of the models improved with increase in the number if input parameters and number of data stations.

The study results shows that for catchments with scarce and low quality hydrological and meteorological data observations, use of satellite data in optimized LSTM and wavelet

neural network models can be relied upon for the prediction of rainfall and runoff trends. It is recommended that similar studies be carried out with the inclusion of basin physical characteristics such as elevation, slope and flow accumulation as training inputs to determine the significance of the physical watershed characteristics in rainfall and runoff predictions.

Declarations

References

1. Cibin R, Trybula E, Chaubey I, Brouder SM, Volenec JJ (2016) Watershed-scale impacts of bioenergy crops on hydrology and water quality using improved SWAT model. GCB Bioenergy 8(4):837–848. https://doi.org/10.1111/gcbb.12307
2. Jeong D, Kim Y (2005) Rainfall-runoff models using artificial neural networks for ensemble streamflow prediction. Hydrol Process 19(19):3819–3835. https://doi.org/10.1002/hyp.5983
3. Sivapalan M et al (2003) IAHS Decade on Predictions in Ungauged Basins (PUB), 2003–2012: shaping an exciting future for the hydrological sciences. Hydrol Sci J 48(6):867–880. https://doi.org/10.1623/hysj.48.6.857.51421
4. Hrachowitz M et al (2013) A decade of Predictions in Ungauged Basins (PUB)–a review. Hydrol Sci J 58(6):1198–1255. https://doi.org/10.1080/02626667.2013.803183
5. Parkes BL, Wetterhall F, Pappenberger F, He Y, Malamud BD, Cloke HL (2013) Assessment of a 1-hour gridded precipitation dataset to drive a hydrological model: a case study of the summer 2007 floods in the Upper Severn. UK Hydrol Res 44(1):89–105. https://doi.org/10.2166/nh.2011.025
6. Wu CL, Chau KW (2013) Prediction of rainfall time-series using modular soft computing methods. Eng Appl Artif Intell 26(3):997–1007. https://doi.org/10.1016/j.engappai.2012.05.023
7. Chen JC, Jan CD, Huang WS (2013) Characteristics of rainfall triggering of debris flows in the Chenyulan watershed. Taiwan Nat Hazards Earth Syst Sci 13(4):1015–1023. https://doi.org/10.5194/nhess-13-1015-2013
8. Pollacco JAP, Ugalde JMS, Angulo-Jaramillo R, Braud I, Saugier B (2008) A linking test to reduce the number of hydraulic parameters necessary to simulate groundwater recharge in unsaturated soils. Adv Water Resour 31(2):355–369. https://doi.org/10.1016/j.advwatres.2007.09.002
9. Wen J, Han P-F, Wang X-S (2019) Lake level dynamics exploration using deep learning, artificial neural network, and multiple linear regression techniques. Environ Earth Sci 78:222. https://doi.org/10.1007/s12665-019-8210-7
10. Sang YF (2013) A review on the applications of wavelet transform in hydrology time-series analysis. Atmos Res 122:8–15. https://doi.org/10.1016/j.atmosres.2012.11.003
11. Wood EF et al (2011) Hyper-resolution global land surface modelling: Meeting a grand challenge for monitoring Earth's terrestrial water. Water Resour Res 47(5):W05301. https://doi.org/10.1029/2010WR010090
12. Wesemann J, Herrnegger M, Schulz K (2018) Hydrological modelling in the anthroposphere: predicting local runoff in a heavily modified high-alpine catchment. J Mt Sci 15(5):921–938. https://doi.org/10.1007/s11629-018-4979-1
13. Kratzert F, Klotz D, Brenner C, Schulz K, Herrnegger M (2018) Rainfall-Runoff modelling using Long-Short-Term-Memory (LSTM) networks. Hydrol Earth Syst Sci 22:6005–6022. https://doi.org/10.5194/hess-22-6005-2018
14. Wong KW, Wong PM, Gedeon TD, Fung CC (2003) Rainfall prediction model using soft computing technique. Soft Comput 7(6):434–438. https://doi.org/10.1007/s00500-002-0232-4
15. Somvanshi VK, Pandey OP, Agrawal PK, Kalanker NV, Prakash MR, Ramesh C (2006) Modelling and prediction of rainfall using artificial neural network and ARIMA techniques. J Indian Geophys Union 10:141–151
16. Liu X, Zhang A, Shi C, Wang H (2009) Filtering and multi-scale RBF prediction model of rainfall based on EMD method. In Proceedings of the 2009 1st International Conference on Information Science and Engineering (ICISE), Nanjing, China, December 26–28, pp. 3785–3788. https://doi.org/10.1109/ICISE.2009.592
17. Jain A, Srinivasulu S (2006) Integrated approach to model decomposed flow hydrograph using artificial neural network and conceptual techniques. J Hydrol 317(3–4):291–306. https://doi.org/10.1016/j.jhydrol.2005.05.022
18. Zhang B, Govindaraju S (2000) Prediction of watershed runoff using Bayesian concepts and modular neural networks. Water Resour Res 36:753–762. https://doi.org/10.1029/1999WR900264
19. Corzo G, Solomatine DP (2007) Base flow separation techniques for modular artificial neural network modelling in flow forecasting. Hydrol Sci J 52(3):491–507. https://doi.org/10.1623/hysj.52.3.491
20. Nelson M, Hill T, Remus T, O'Connor M (1999) Time-series forecasting using neural networks. Should the data be deseasonalized first? J Forecast 18:359–367
21. Xiang Z, Yan J, Demir I (2020) A rainfall-runoff model with LSTM-based sequence-to-sequence learning. Water Resourc Res. https://doi.org/10.1029/2019WR025326
22. Cui Z, Ke R, Wang Y (2018) Deep Bidirectional and Unidirectional LSTM Recurrent Neural Network for Network-wide Traffic Speed Prediction. arXiv:1801.02143.

23. 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

24. Gürsoy O, Engin SN (2019) A wavelet neural network approach to predict daily river discharge using meteorological data. Meas Control 52:599–607. https://doi.org/10.1177/0020294019827972

25. Ouma YO, Tetuko J, Tateishi R (2008) Analysis of co-occurrence and discrete wavelet transform textures for differentiation of forest and non-forest vegetation in very-high-resolution optical-sensor imagery. Int J Remote Sens 29(12):3417–3456. https://doi.org/10.1080/01431160701601782

26. Hochreiter S, Schmidhuber J (1997) Long short-term memory. Neural Comput 9(8):1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735

27. Awange JL, Saleem A, Sukhadiya RM, Ouma YO, Kexian H (2019) Physical dynamics of Lake Victoria over the past 34 years (1984–2018): Is the lake dying? Sci Total Environ 65:199–218. https://doi.org/10.1016/j.scitotenv.2018.12.051

28. Rumelhart DE, Hinton GE, Williams RJ (1985) Learning internal representations by error propagation. California University San Diego La Jolla Inst for Cognitive Science, No. ICS-8506

29. Carriere P, Mohaghegh S, Gaskar R (1996) Performance of a virtual runoff hydrographic system. Water Resourc Plann Manag 122:120–125

30. Kumar DN, Raju KS, Sathish T (2004) River flow forecasting using recurrent neural networks. Water Resour Manag 18:143–161. https://doi.org/10.1023/B:WARM.0000024727.94701.12

31. Hsu K, Gupta HV, Soroochian S (1997) Application of a recurrent neural network to rainfall-runoff modelling. Proc., Aesthetics in the Constructed Environment, ASCE, New York, pp. 68–73

32. Ma X, Tao Z, Wang Y, Yu H, Wang Y (2015) Long short-term memory neural network for traffic speed prediction using remote microwave sensor data. Transp Res Part C Emerg Technol 54:187–197. https://doi.org/10.1016/j.trc.2015.03.014

33. Bengio Y, Simard P, Frasconi P (1994) Learning long-term dependencies with gradient descent is difficult. IEEE Trans Neural Netw 5(2):157–166. https://doi.org/10.1109/72.279181

34. Chang FJ, Lo YC, Chen PA, Chang LC (2015) Shieh MC multi-step-ahead reservoir inflow forecasting by artificial intelligence techniques. Springer International Publishing, Cham, pp 235–249

35. Greff K, Srivastava RK, Koutník J, Steunebrink BR, Schmidhuber J (2017) LSTM: a search space odyssey. IEEE Trans Neural Netw Learn Syst 28(10):2222–2232. https://doi.org/10.1109/TNNLS.2016.2582924

36. Ouma YO, Hahn M (2016) Wavelet-morphology based detection of incipient linear cracks in asphalt pavements from RGB camera imagery and classification using circular Radon transform. Adv Eng Inform 30:481–499. https://doi.org/10.1016/j.aei.2016.06.003

37. Rajaee T (2011) Wavelet and ANN combination model for prediction of daily suspended sediment load in rivers. Sci Total Environ 409:2917–2928. https://doi.org/10.1016/j.scitotenv.2010.11.028

38. Liu Y, Sergey F (2013) Seismic data analysis using local time-frequency decomposition. J Geophys Prospect 61(3):516–525. https://doi.org/10.1111/j.1365-2478.2012.01062.x

39. Kim TW, Valdes JB (2003) Nonlinear model for drought forecasting based on a conjunction of wavelet transforms and neural networks. J Hydrol Eng 8(6):319–328. https://doi.org/10.1061/(ASCE)1084-0699(2003)8:6(319)

40. Shabri A (2015) A hybrid model for stream flow forecasting using wavelet and least Squares support vector machines. Jurnal Teknologi 73:89–96

41. Belayneh A, Adamowski J (2013) Drought forecasting using new machine learning methods. J Water Land Devel 18(1):3–12. https://doi.org/10.2478/jwld-2013-0001

42. Liu M, Huang Y, Li Z, Tong B, Liu Z, Sun M, Jiang F, Zhang H (2020) The applicability of LSTM-KNN model for real-time flood forecasting in different climate Zones in China. Water 12:440. https://doi.org/10.3390/w12020440

43. Xiang Z, Yan J, Demir I (2020) A rainfall-runoff model with LSTM-based sequence-to-sequence learning. Water Resour Res 56(1):1–17. https://doi.org/10.1029/2019WR025326

44. Ouma YO, Owiti T, Kipkorir E, Kibiiy J, Tateishi R (2012) Multi-temporal comparative analysis of TRMM-3B42 satellite estimated rainfall with surface gauge data at basin scales: daily, decadal and monthly evaluations. Int J Remote Sens 33(12):7662–7684. https://doi.org/10.1080/01431161.2012.701347

45. Poornima S, Pushpalatha M (2019) Prediction of rainfall using intensified LSTM based recurrent neural network with weighted linear units. Atmosphere 10(11):668. https://doi.org/10.3390/atmos10110668

46. Guimarães-Santos CA, Silva GBLD (2014) Daily streamflow forecasting using a wavelet transform and artificial neural network hybrid models. Hydrol Sci J 59(2):312–324. https://doi.org/10.1080/02626667.2013.800944

47. Chou CM (2014) Random modelling of daily rainfall and runoff using a seasonal model and wavelet denoisin. Math Probl Eng. https://doi.org/10.1155/2014/917365

48. Solgi A, Nourani V, Pourhaghi A (2014) Forecasting daily precipitation using hybrid model of wavelet-artificial neural network and comparison with adaptive neurofuzzy inference system (case study: Verayneh station Nahavand). Adv Civil Eng. https://doi.org/10.1155/2014/279368

49. Ouma YO, Okuku CO, Njau EN (2020) Use of artificial neural networks and multiple linear regression model for the prediction of dissolved oxygen in rivers: case study of hydrographic basin of River Nyando. Kenya Complexity. https://doi.org/10.1155/2020/9570789

50. Doycheva K, Horn G, Koch C, Schumann A, König M (2017) Assessment and weighting of meteorological ensemble forecast members based on supervised machine learning with application to runoff simulations and flood warning. Adv Eng Inform 33:427–439. https://doi.org/10.1016/j.aei.2016.11.001

51. Komasi M (2007) Modelling rainfall-runoff model using a combination of wavelet-ANN. Tabriz University, Tabriz

52. Adamowski J, Sun K (2010) Development of a coupled wavelet transform and neural network method for flow forecasting of non-perennial rivers in semi-arid watersheds. J Hydrol 390(1):85–91. https://doi.org/10.1016/j.jhydrol.2010.06.033

53. Ramana RV, Krishna B, Kumar SR, Pandey NG (2013) Monthly rainfall prediction using wavelet neural network analysis. Water Resour Manag 27:3697–3711. https://doi.org/10.1007/s11269-013-0374-4

54. Wu C, Chau K, Li Y (2009) Methods to improve neural network performance in daily flows prediction. J Hydrol 372:80–93. https://doi.org/10.1016/j.jhydrol.2009.03.038

55. Du K, Zhao Y, Lei J (2017) The incorrect usage of singular spectral analysis and discrete wavelet transform in hybrid models to predict hydrological time-series. J Hydrol 552:44–55. https://doi.org/10.1016/j.jhydrol.2017.06.019

56. Panda SK, Jana PK (2015) Efficient task scheduling algorithms for heterogeneous multi-cloud environment. J Supercomput 71:1505–1533. https://doi.org/10.1007/s11227-014-1376-6

57. Wang S, Feng J, Liu G (2013) Application of seasonal time-series model in the precipitation forecast. Math Comput Model 58:677–683. https://doi.org/10.1016/j.mcm.2011.10.034

58. Lundberg SM et al (2020) From local explanations to global understanding with explainable AI for trees. Nat Mach Intell 2:2522–5839