Environmental Processes (2022) 9:53
https://doi.org/10.1007/s40710-022-00602-x

ORIGINAL ARTICLE

Improving Monthly Rainfall Forecast in a Watershed by Combining Neural Networks and Autoregressive Models

Albenis Pérez-Alarcón[1,2] · **Daniel Garcia-Cortes**[3] · **José C. Fernández-Alvarez**[1,2] · **Yoel Martínez-González**[3]

Received: 21 March 2022 / Accepted: 5 August 2022 / Published online: 24 August 2022

Abstract

The main aim of the rain forecast is to determine rain occurrence conditions in a specific location. This is considered of vital importance to assess the availability of water resources in a basin. In this study, several methods are analyzed to forecast monthly rainfall totals in hydrological basins. The study region was the Almendares-Vento basin, Cuba. Based on Multi–Layer Perceptron (MLP), Convolutional Neural Network (CNN) and Long Short–Term Memory (LSTM) neural networks, and Autoregressive Integrated Moving Average (ARIMA) models, we developed a hybrid model (ANN+ARIMA) for rainfall prediction. The input data were the one year lagged rainfall records in gauge stations within the basin, sunspots, the sea surface temperature and time series of nine climate indices up to 2014. The predictions were also compared with the rainfall records of a gauge station network from 2015 to 2019 provided by the Cuban National Institute of Hydraulic Resources. Based on several statistical metrics such as mean absolute error, Pearson correlation, BIAS, Nash–Sutcliffe efficiency and Kling–Gupta efficiency, the CNN model showed higher ability to forecast monthly rainfall. Nevertheless, the hybrid model was notably better than individual models. Overall, our findings have proved the reliability of using the hybrid model to predict rainfall time series for water management and can be extensively applied to this sort of application. In addition, this work proposes a new approach to enhance the planning and management of water availability in watershed for agriculture, industry and population through improving rainfall forecasting.

Article Highlights

- Convolutional Neural Network model is able to forecast monthly rainfall amounts.
- Our methodology allows the models to learn the seasonal variations of the rainfall.
- The hybrid model is skillful to forecast rainfall time series for water management.
- The findings are promising to enhance water management systems.
- The method can be easily applied to predict rainfall in other watersheds.

Keywords rainfall forecast · artificial neural networks · ARIMA models · Almendares-Vento basin

Extended author information available on the last page of the article

1 Introduction

Basin water resource planning and management require accurate data on rainfall. Understanding and modelling rainfall is one of the hydrologic cycle most complicated issues because of the complexity of both, the atmospheric processes triggering rainfall and the significant variation range of scales in space and time (Hung et al. 2009; Sumi et al. 2012). Indeed, rainfall forecast poses a great challenge in operational hydrometeorology, regardless of several advances in weather forecasting in the last decades (Li and Lai 2004; Hong 2008; Majumdar et al. 2021).

There are two well-known approaches to forecast precipitation (Luk et al. 2001). The former involves studying the rainfall process by applying the underlying physical laws using mathematical equations for hydrological process (Solomatine and Ostfeld 2008). The latter is substantiated by the pattern recognition methodology, which tries to recognize precipitation patterns, taking into account their characteristics in historical data, which are used to predict the rainfall evolution (Ghumman et al. 2011; Diodato and Bellocchi 2018; Ridwan et al. 2021). By such contribution, the latter is considered appropriate since the immediate priority is to predict monthly rainfall at certain spots inside a basin.

The improvements in pattern recognition methodologies have led to the extensive use of machine learning (ML) tools to solve several problems in many research areas, including spatial prediction of several natural hazards, i.e., flooding (Band et al. 2020a), landslides (Moradi et al. 2019), wildfires (Watson et al. 2019), tropical cyclones (Li et al. 2017; Yang et al. 2020; Yuan et al. 2021) and storm surge (Bai et al. 2022). Ghazvinei et al. (2018) applied extreme ML to describe the sugarcane growth and, consequently, the improvement of agricultural production. Choubin et al. (2019) proposed several models based on ML for predicting earth fissures. Likewise, ML algorithms (e.g., boosted regression tree, random forest, parallel random forest, regularized random forest, randomized trees) have also been applied to evaluate the flash flood (e.g., Band et al. 2020a) and gully erosion (e.g., Band et al. 2020b) susceptibility modelling. Fernández-Alvarez et al. (2019) calculated the power of an intraocular lens to be implanted after cataract surgery by applying a multilayer perceptron (MLP) neural network. Recently, Shabani et al. (2020) used several ML methods (Gaussian Process Regression, K-Nearest Neighbors, Random Forest and Support Vector Regression) to predict evaporation. Note that evaporation is a complex and nonlinear phenomenon and one of the most critical components of the hydrological cycle. While Wang et al. (2017a,b) found that the MLP model fed by regional data was skillful for predicting monthly evaporation in different climate regions of China, Ghorbani et al. (2018) noted that the hybrid MLP-Firefly Algorithm performed better. Recently, Fang and Shao (2022) and Li et al. (2022) applied the Long Short-Term Memory (LSTM) for predicting rainfall-runoff in the Han and Elbe rivers basins, respectively.

The literature reveals that ML has gained popularity for water management, rainfall prediction and solving hydrological issues (Ridwan et al. 2021; Barrera-Animas et al. 2022). Several authors (Teschl et al. 2007; Krasnopolsky and Lin 2012; Nastos et al. 2013; Hardwinarto et al. 2015; Kashiwao et al. 2017; Vathsala and Koolagudi 2017; Chao et al. 2018; Haidar and Verma 2018; Benevides et al. 2019; Anochi et al. 2021; Narejo et al. 2021; Ridwan et al. 2021; Sun et al. 2021; Venkatesh et al. 2021) have utilized artificial neural networks (ANNs) (Anochi et al. 2021), to perform rainfall forecasts. Ridwan et al. (2021) used NN regression to estimate the rainfall data in Tasik Kenyir, Terengganu; and Anochi

et al. (2021), using a supervised ANN model, proposed a new approach to predict seasonal precipitations over South America. Besides, Krasnopolsky and Lin (2012), based on data learning using the ANN technique developed a multi-model cluster approach to improve short-term rainfall forecasting over the continental United States, while Teschl et al. (2007) enhanced weather radar forecast with feed-forward ANNs. Following a similar approach, Benevides et al. (2019) performed an hourly intense rainfall forecasting using the time series feed-forward ANN, integrating meteorological data and Global Navigation and Positioning System (GNSS). Similarly, Zhao et al. (2021) utilized the precipitable water vapour predicted by the GNSS as input data for a supervised learning algorithm for hourly rainfall forecasting. Kashiwao et al. (2017) applied the MLP ANN to predict local rainfall in Japan's areas with data from the Japan Meteorological Agency. Likewise, Haidar and Verma (2018) applying one-dimensional deep Convolutional Neural Network (CNN) proposed a new forecasting method to estimate monthly rainfall for a chosen area in eastern Australia. Furthermore, Lee et al. (2018) developed a model using ANN for rainfall forecasting in South Korea's Geum River Basin. Proposals by Haidar and Verma (2018) and Lee et al. (2018) stand out as novel since they use climate indices to predict monthly accumulated precipitation. Narejo et al. (2021) found that the temporary Deep Belief Network (DBN) model outperforms the CNN particularly on the forecast of rainfall time series, but it requires more computational resources than other deep learning architectures. Additionally, Samad et al. (2020) utilized an LSTM model to forecast rainfall in subsequent days. By using satellite data from the NASA Global Precipitation Measurement, Gamboa-Villafruela et al. (2021) trained a convolutional LSTM architecture for precipitation nowcasting. With a similar approach, Bhuiyan et al. (2020) and Derin et al. (2020) used rainfall observation from satellite to train a random forest and neural network for improving the precipitation estimation in water resources applications.

On the other hand, stochastic modelling techniques are also common for forecasting time series with hydrological applications. Several authors (e.g. Wang et al. 2013; Mahmud et al. 2017; Hernández et al. 2017; Bonakdari et al. 2019; Ebtehaj et al. 2019) have employed Autoregressive Integrated Moving Average (ARIMA) models to perform precipitation forecasts, but these models show a mild ability to catch major nonlinear features of rainfall series (Narejo et al. 2021). Nevertheless, the advantages of ANN models to carry out a nonlinear mapping between inputs and outputs provide a useful alternative for rainfall forecasting at short and long terms (Ali et al. 2020; Wu et al. 2010; Dounia et al. 2014; Nourani et al. 2019). Authors (e.g., Zhang 2003; Aladag et al. 2012; Moeeni and Bonakdari 2017) have shown that the combination of both ANNs and ARIMA models can increasingly trigger forecasting accuracy much more effectively than if they were set apart from each other (Dodangeh et al. 2020). Indeed, hybrid models of ANNs and ARIMA have been widely used in several applications, i.e., for estimating PM_{10} pollution (Wongsathan and Seedadan 2016), modelling rainfall-runoff process (Nourani et al. 2011), predicting water quality (Faruk 2010) and fuelwood prices (Koutroumanidis et al. 2009), and determining the annual energy cost budget (Jeong et al. 2014). Overall, both ANNs and ARIMA models turn out especially applicable when: (a) mathematical simulation of the physical phenomena is either too difficult or impossible; and (b) needed parameters for mathematical simulations cannot be described with adequate precision (Xiaojian and Quan 2009). These situations arise quite often in water resources management problems, modeling of rainfall-runoff process, flood forecasting, etc., turning ANNs and ARIMA models into viable options to be used.

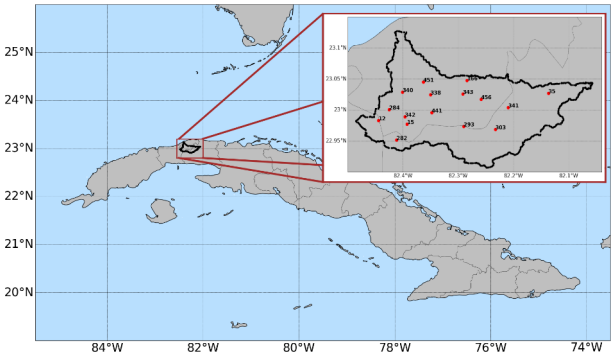

Fig. 1 Almendares-Vento basin. The red point and black numbers represent the station gauge locations and the station IDs, respectively

The accurate rainfall forecast for hydrological watershed management is a challenge nowadays due to its spatial and temporal variability (Xu et al. 2014). Therefore, given the findings of previous works (e.g., Zhang 2003; Haidar and Verma 2018; Lee et al. 2018; Bonakdari et al. 2019; Ebtehaj et al. 2019; Dodangeh et al. 2020), this study aimed to develop a hybrid model based on ANNs (MLP, CNN and LSTM) and ARIMA models for predicting monthly rainfall totals over hydrological basins. A skillful model for forecasting monthly rainfall will be useful for planning water management to mitigate the impact of dry periods. Furthermore, an accurate forecast over a year will also help decision-makers to decide on water availability for agriculture, industry and the population in general. As the study region, we selected the Almendares-Vento catchment, considered one of the most important basins in Cuba because it provides freshwater to Havana city.

2 Materials and methods

2.1 Study Area

The study area spans the Almendares-Vento basin, which is located in the western region of Cuba, on the northern slope of the country (Fig. 1), and is also recognized by the National Center for Hydrographic Basins (CNCH, Spanish acronym) as a basin of national interest (García-Fernández and Díaz 2017). The Almendares watershed covers an area of 422 km^2. Its hydrographic network comprises intermittent streams, while its main river is the Almendares with a length of 49.8 km (Rivera 2009). The Vento underground basin is divided into two large regions: karstic and non-karstic. Gentle slopes and undulating plains predominate and a system of sub-horizontal, monocline blocks and a well-marked stratification stand out (Valcarce-Ortega et al. 2007).

2.2 Data

Rainfalls registered in gauge stations of the Almendares-Vento basin and climatic indices from different sources were collected as monthly predictors of rainfall. Rainfall data used were retrieved from the National Institute of Hydraulic Resources (INRH, Spanish acronym) of Cuba. The gauge station spatial distribution is disclosed in Fig. 1. The compiled rainfall data included unavailable values. Those values were completed by data from nearby

Table 1 Statistics of each rainfall predictor. Niño indices, AMO and SST (in ºC) and rainfall (in mm). (σ: standard deviation, μ: mean value. Rainfall ID-N is referred to the rainfall in the gauge station with the identification number N in Fig. 1)

Attribute	Source	Minimum	Maximum	μ	Median	σ	Period
ID-10	INRH	0.0	702.0	122.0	99.95	105.5	1952–2013
ID-12	INRH	0.0	884.0	130.0	105.95	112.2	1971–2014
ID-15	INRH	0.0	822.0	132.0	103.05	118.5	1906–2014
ID-166	INRH	0.0	801.0	119.0	92.80	100.8	1981–2014
ID-284	INRH	0.0	434.0	128.0	90.95	115.2	2000–2014
ID-293	INRH	0.0	526.0	116.0	90.45	100.4	1993–2006
ID-338	INRH	0.0	765.0	115.0	91.65	104.1	1978–2014
ID-340	INRH	0.0	494.0	112.0	92.10	93.13	1981–2014
ID-341	INRH	0.0	503.0	113.0	87.55	98.38	2001–2014
ID-342	INRH	0.0	593.0	134.0	102.5	116.0	2001–2014
ID-343	INRH	0.0	892.0	109.0	81.25	100.1	1981–2014
ID-441	INRH	0.0	498.0	123.0	101.65	107.1	1997–2014
ID-451	INRH	0.0	558.0	120.0	91.10	105.2	1995–2014
ID-456	INRH	0.0	546.0	112.0	84.25	101.3	1995–2014
NAO	CRU	-4.7	6.66	0.05	0.05	1.714	1906–2014
Niño 1.2	GCOS	18.7	29.6	23.2	22.88	2.284	1906–2014
Niño 3.0	GCOS	22.8	29.2	25.7	25.72	1.196	1906–2014
Niño 3.0	GCOS	22.8	29.2	25.7	25.72	1.196	1906–2014
Niño 3.4	GCOS	26.5	29.8	28.4	28.44	0.589	1906–2014
Niño 4.0	GCOS	24.4	29.1	26.9	27.0	0.866	1906–2014
AMO	PSL	-0.5	0.60	0.0	-0.036	0.212	1906–2014
SOI	BOM	-3.4	4.07	0.0	-0.045	1.087	1906–2014
Sunspots	SIDC	0.0	359.0	91.0	76.65	71.76	1906–2014
SST	NOAA	19.2	25.9	22.3	22.08	1.928	1906–2014

weather station or average. To produce all data, predictors depicted in Table 1 were utilized. Rainfall, sea surface temperature (SST), sunspot and climatic indices time series were included as input characteristics.

The model parameter dependence on climatic features plays an important role in the response of the catchments to climatic variability (Sivapalan et al. 2011). Thus, following Haidar and Verma (2018) and Lee et al. (2018), several climatic indices that modulate climatic conditions in the North Atlantic Ocean were used to predict the variability of rainfall in the Almendares-Vento basin. The North Atlantic Oscillation (NAO) index (Jones et al. 1997) was extracted from the Climatic Research Unit (CRU). The NAO is one of the greater variability modes of the Northern Hemisphere atmosphere in the Atlantic Ocean (Hurrell 1995). The NAO is often deemed as the normalized pressure difference between a station on Iceland and one on the Azores. Niño 1.2, Niño 3.0, Niño 3.4 and Niño 4.0 indices (Rayner et al. 2003) were extracted from the Global Climate Observing System (GCOS). The Atlantic Multi-decadal Oscillation (AMO) is a consistent way of natural change happening in the North Atlantic Ocean. It is based on the SST average anomalies in the North Atlantic basic, frequently over 0–70º N (Enfield et al. 2001). The Southern Oscillation Index (SOI) can be defined as a standardized index, which is supported by the observed differences in sea level pressure between Tahiti and Darwin, Australia. The SOI is a large-scale fluctuation measure of air pressure that occurs in the tropical Pacific, between the eastern and western part, dur-

ing La Niña and El Niño phenomena. The sunspot values were compiled from the Solar Influences Data Analysis Center (SIDC) and the SST averaged over the Atlantic Ocean between 5 and 50 ºN and 5-100 ºW was taken from the Centennial Time Scale (COBE SST2) database (Hirahara et al. 2014) at the National Oceanic and Atmospheric Administration (NOAA). In the dataset the daily SST field is made up of a trend's total, interannual variability, and every day shifts, taking into account in situ SST and sea ice concentration observations. All these climatic indices were considered as predictors of the monthly rainfall in the Almendares-Vento basin. The source, minimum, maximum, mean, median, and standard deviation (σ) of each weather parameter are described in Table 1. Figure SM1 in Supplementary Material (SM) also shows the time series for each rainfall predictor.

2.3 Artificial Neural Networks

ANNs are mathematical models, which harness learning algorithms led by the intellect to save information (Keijsers 2010). Therefore, the ANNs are algorithm clusters made up of computational components, denoted as neurons, which get signals from the environment or from other neurons, and then change those signals and produce an output signal, which could be spread to the environment or to another neuron (Fernández-Alvarez et al. 2019). The first neural network called a perceptron was introduced in 1957 (Hung et al. 2009). It contained a single input layer and the outputs were obtained directly from the inputs through the weighted connections. The MLP neural network was developed in 1960. This ANN gradually turned into one of the most broadly used for solving various problems (Velasco et al. 2019; Ren et al. 2021). The MLP is a feed-forward network that comprises one or more hidden layers (Vathsala and Koolagudi 2017). Mostly, the signals are transmitted within the network from input to output in one direction. Every neuron's output does not influence the neuron itself because of the lack of a loop. The MLP power stems precisely from non-linear activating functions (Rhee and Shin 2018). The learning process is carried out by means of a supervised method where the willing output must be foreseen to actualize the internal connections weights among the layers (Hussain et al. 2020). Connection weights are constantly changed in agreement with the calculated error until the error gradient achieves an adequate minimum value, which implies that the latter output is near the goal (Ebecken 2011; Panchal et al. 2011; Hossain et al. 2013; Popescu et al. 2009). MLPs are usually trained with the back-propagation algorithm.

The CNN is a profound neural network initially created for image study. A convolutional neural networks power stems from a particular sort of layer known as the convolutional layer. Hence, the structure of a convolutional neural network turns out a multi-layered feed-forward neural network, generated by stacking a lot of hidden layers on top of one another in a row. These layers are generally divided into three types: convolutional, pooling and fully connected (Sakib et al. 2018).

The LSTM neural network was suggested by Hochreiter and Schmidhuber (1997), and is widely used today because of its superior performance in accurate modelling of both short and long-term data. LSTMs are particularly created to recall long-term dependencies. As a frequent repeated neural network, it contains a self-loop. The distinction between LSTM and frequent networks is the inner architecture. The most noteworthy element of the LSTM is the cell state that discloses the information throughout the whole chain with a few linear interactions (Fathi and Shoja 2018).

2.4 Autoregressive Integrated Moving Average (ARIMA) Models

One of the most frequently used automatic prediction algorithms are ARIMA models (Tseng et al. 2002; Aguado-Rodríguez et al. 2016). The ARIMA models serve to predict simple series of a single variable, in which the forecasts are just based on previous values of the analyzed variable. The ARIMA models can be used to make short-term forecasts because most of them attach more importance to the late past than to the faraway past.

Box et al. (1994) developed the classic methodology that uses time series for generating models (e.g., ARMA, ARIMA) to obtain predictions. According to Hyndman and Khandakar (2008) and Aguado-Rodríguez et al. (2016), a common obstacle when using ARIMA models for prediction is that the order selection process is generally considered subjective and difficult to apply. The authors recommend to all those interested in delving into the details of the method to consult the specialized literature on the subject (Kotu and Deshpande 2019; Hyndman and Athanasopoulos 2018).

2.5 Methodology

The dataset generated consists of time series created from gauge station records, as shown in Table 1. Additionally, evaluating the ANN model skill to forecast the monthly rainfall, the rainfall records of some gauge stations (ID-12, ID-15, ID-284, ID-338, ID-441, ID-451, see Fig. 1) in the Almendares-Vento basin and the climatic indices from 2015 to 2019 were used. It is worth noting that the selection criteria to use these gauge stations was based on the availability of the rainfall data from 2015 to 2019. About 120 characteristics that represent the predictors backward values up to one natural year backward were generated to estimate the monthly rainfall totals for the next calendar year.

2.5.1 Parameter Tuning

For the training of ANN models, we applied a similar algorithm to that previously described by Haidar and Verma (2018). To sum up, the dataset was partitioned into three distinct parts for training, validation and testing. Some researchers (e.g., Wu et al. 2010) have used 50% of the data for training and remaining 50% for testing, whereas others (e.g., Venkatesh et al. 2021) used 70% and 30% as training and testing data, respectively. In this work, after several tests, we partitioned the dataset in 70% to learn the weights and biases of neurons in each ANN model, 15% for validating procedure to find the proper architecture during training and the remaining 15% for testing. The training and validating steps are critical phases to optimize the model parameters and also prevent overfitting (Chen et al. 2016; Sadeghi et al. 2020). In addition, the inclusion of check point helped identify when a better performance over the validation dataset was attained in order to save the network weights.

To determine the optimal ANNs architecture and ARIMA configuration the trial-and-error method was carried out. The MLP neural network was configured with two layers: the input and output layers (Fig. 2a). The selected architecture for CNN is made up of two convolutional layers, average pooling and fully connected layers (Fig. 2b). Additionally, the LSTM network was configured with a single LSTM layer and two fully connected layers (Fig. 2c). Dropout (a regularization technique) was added to the CNN and LSTM architectures (Srivastava et al. 2014; Haidar and Verma 2018). In all implemented ANNs, we used

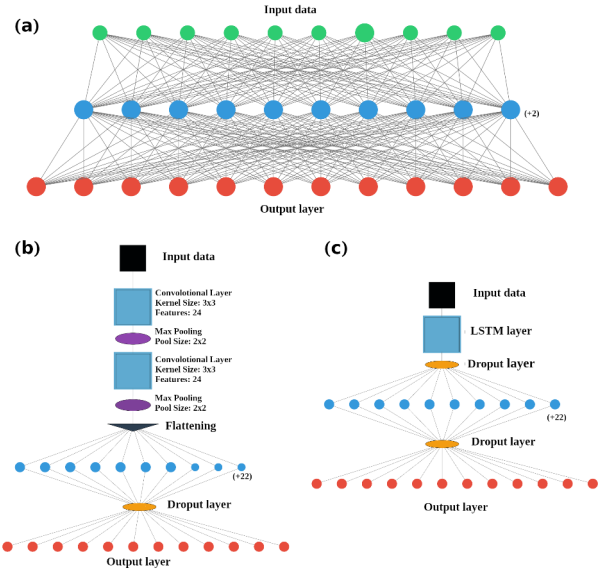

Fig. 2 Artificial Neural Networks architecture for: **(a)** Multi–Layer Perceptron (MLP); **(b)** Convolutional Neural Network (CNN); and **(c)** Long Short-Term Memory (LSTM)

the Rectified Linear Unit (ReLU) as activation function. The superior performance of ReLU is commonly believed to come from sparsity (Glorot et al. 2011; Sun et al. 2015). Formally, in the ReLU activation function, y is equal to x when x is greater than or equal to zero, and null when x is less than zero. Furthermore, the neural network was trained by reducing the mean absolute error (MAE). That is, the MAE was used as loss function for training. Additionally, we used the Adaptive Moment Estimation Algorithm or so-called Adam Optimizer (Kingma and Ba 2017) for deep learning. According to Kingma and Ba (2017), such method is computationally useful, very suitable for issues that are large in terms of data, adequate for non-stationary purposes and has few memory requirements.

Furthermore, an exploratory analysis of rainfall data in gauge stations reveals certain seasonal frequency in the period analyzed. Data were filtered and smoothed with the Fast Fourier Transform (FFT) algorithm, in order to estimate the seasonal frequency. Indeed, Fig. 3 displays the temporal progression of the smoothed mean rainfall in Almendares-Vento basin, which is characterized by 48 ~ 60 months (4 ~ 5 years) of seasonal frequency. Thus, there is a succession of wet and dry periods of approximately 4–5 years each one. These results were utilized to configure the ARIMA error model seasonally integrated with seasonal moving average model.

The rainfall time series of each gauge station were processed in a numerical experiment, with and without differentiating (in ARIMA model, D = 1 and D = 0, respectively). The better results (not shown) were obtained for D = 1, considering moving average terms at lags 1, 2, …, 12; therefore, the configuration applied for all gauge stations was ARIMA(0, 1, 12) with a seasonal frequency of 48 and 60 months, respectively.

The Python Keras package (Chollet 2015) installed on top of Tensorflow framework was employed to carry out and train the CNN, MLP and LSTM neural networks, while the ARIMA function in MATLAB was configured with the rainfall dataset of each gauge station as ARIMA (0,1,12), with a seasonal frequency of 48 (ARIMA4) and 60 (ARIMA5) months, respectively.

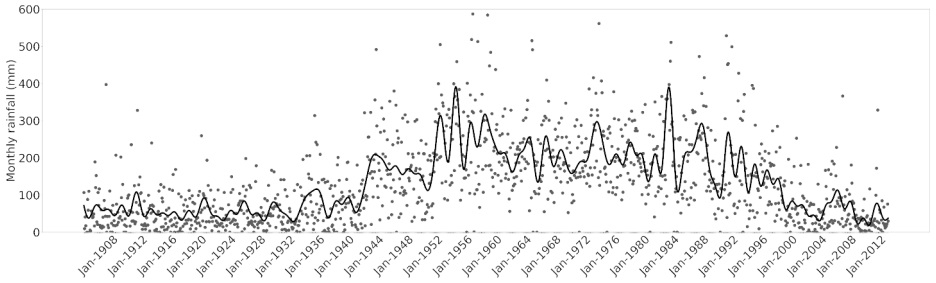

Fig. 3 Smoothed mean rainfall in Almendares-Vento basin by applying the Fast Fourier Transform algorithm

2.5.2 Performance Measurements

To evaluate the precision of the models, some statistics, which have been broadly employed in rainfall forecasting tasks were estimated: MAE (Eq. 1), BIAS (Eq. 2), Root Mean Square Error (RMSE, Eq. 3), Pearson correlation coefficient (r_p, Eq. 4), Nash-Sutcliffe efficiency (NS, Eq. 5) and coefficient of variation (CV, Eq. 6):

$$MAE = \frac{\sum_{i=1}^{n} |x_i - y_i|}{n} \tag{1}$$

$$BIAS = \frac{\sum_{i=1}^{n} (x_i - y_i)}{n} \tag{2}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n} (x_i - y_i)^2}{n}} \tag{3}$$

$$r_p = \frac{\sum_{i=1}^{n} (x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n} (x_i - \overline{x})^2 (y_i - \overline{y})^2}} \tag{4}$$

$$NS = 1 - \frac{\sum_{i=1}^{n} (x_i - y_i)^2}{\sum_{i=1}^{n} (y_i - \overline{y}_i)^2} \tag{5}$$

$$CV = \frac{\sigma_x}{\overline{x}} \tag{6}$$

In all equations y_i is the observed value, x_i is the predicted value, $\overline{y}$ and $\overline{x}$ are the average of observed and predicted values, respectively; n is the number of elements in the dataset, $1 \leq i \leq n$; σ_y, σ_x represents the standard deviation of observed and predicted values, respectively. The RMSE and MAE values fluctuate between 0 and ∞. The nearest to 0 a value is, the more suitable the model forecast will be. Pearson correlation values range between −1 and +1. The closer the value is to 1, the better the forecast is. Nevertheless, to determine the statistical significance of the Pearson correlation coefficient, we applied the two-tailed sig-

nificance test (Weathington et al. 2012). The NS values range between -∞ and 1. Negative NS values denote that the forecasting model does not prove to be a higher predictor than the mean of the measured values. According to Wu et al. (2010), NS is a good alternative as a "goodness-of-fit" because it is sensitive to differences in the observed and forecasted means and variances. Additionally, NS = 0 is regularly used as a benchmark to differentiate 'good' from 'bad' models (Houska et al. 2014).

For calibrating and evaluating hydrological models, the Kling–Gupta efficiency (KGE) (Gupta et al. 2009; Kling et al. 2012) has been recently used to summarize the model performance (Knoben et al. 2019). Therefore, to gain a complete overview of the ability of each ANN and ARIMA model to predict the monthly rainfall in the Almendares-Vento basin, we used the KGE index, which is defined according to following Eq. (7):

$$KGE = 1 - \sqrt{(r_p - 1)^2 + \left(\frac{\sigma_x}{\sigma_y} - 1\right)^2 + \left(\frac{\overline{x}}{\overline{y}} - 1\right)^2} \tag{7}$$

Similar to the NS efficiency, KGE = 1 indicates a perfect agreement. Several authors have used positive KGE values as indicative of "good" model simulations, while negative KGE values indicates "bad" simulations (see references in Knoben et al. 2019). Conversely, Knoben et al. (2019), pointed out that KGE values higher than −0.41 indicates that a model enhances over the mean flow benchmark, but this criterion was applied to the flow in a basin. Therefore, in this study we assumed that higher KGE values suggest that the model is more skillful to predict the monthly rainfall. The KGE efficiency has been widely applied to evaluate the model performance in predicting rainfall time series (e.g., Thiemig et al. 2013; Towner et al. 2019; Gebremichael et al. 2022; Shahid et al. 2021; Girihagama et al. 2022; Li et al. 2022).

Additionally, to better evaluate the relative improvement of one model relative to another, we utilized the KGE skill score (KGE_{SS}), defined according to Eq. 8, in which KGE_a and KGE_b are the scores for the model of interest and the comparative or baseline model, respectively:

$$KGE_{SS} = \frac{KGE_a - KGE_b}{1 - KGE_b} \tag{8}$$

Positive KGE_{SS} indicates improved skill, while a negative score represents that the model of interest performs worse than the baseline. The KGE_{SS} was previously used with this goal by Towner et al. (2019).

The flowchart shown in Fig. 4 summarizes all the steps followed in this work to identify the models with the highest performance to predict the monthly rainfall in the study region and, accordingly, to generate the rainfall forecast. This schematic representation could be a useful guide for future applications of this methodology in other hydrological basins.

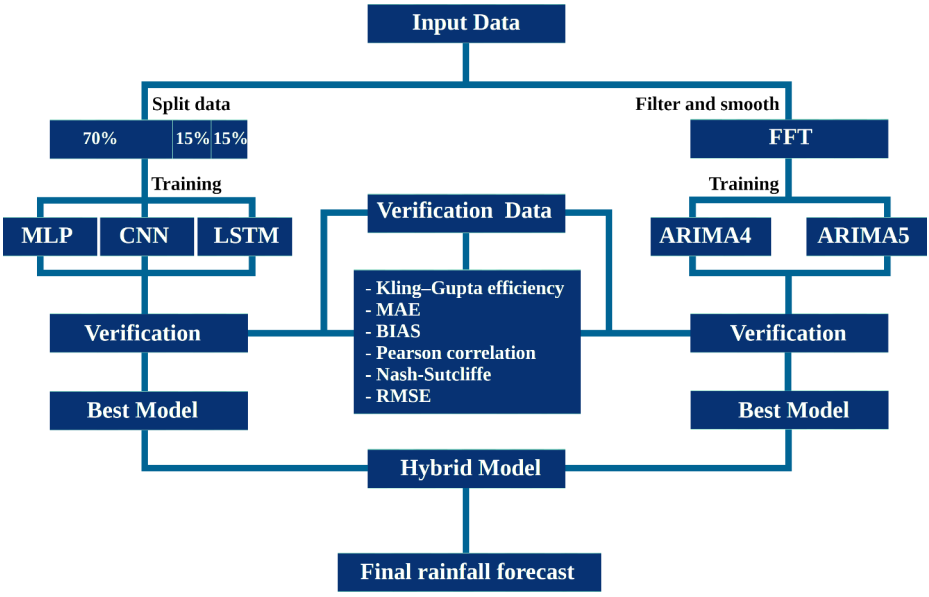

Fig. 4 Schematic representation of the methodology followed in this work. MLP: Multi-Layer Perceptron, CNN: Convolutional Neural Network, LSTM: Long Short-Term Memory, FFT: Fast Fourier Transform, ARIMA4 and ARIMA5: Autoregressive Integrated Moving Average model with a seasonal frequency of 4 and 5 years, respectively

3 Results and Discussion

3.1 ANN Training

The statistics for ANN architectures during the training and testing procedures are depicted in Fig. 5. The loss function above the training database had not been very different from the one over the validating dataset, except for LSTM (Fig. 5c), which compared with CNN (Fig. 5a) and MLP (Fig. 5b) has a larger separation, but with similar monotony. These results suggest that the networks did not overfit in the training phase (Chollet 2015). According to Barrera-Animas et al. (2022), overfitting between observed and forecasted values is one of the major drawbacks in rainfall forecasting. Table 2 presents the MAE, BIAS, RSME, r_p and NS obtained for each ANN model during the training and testing steps. In the training phase, the best weights were obtained on 2898, 995 and 1805 epochs for the CNN, MLP and LSTM models, respectively. The statistics shown in Table 2 suggest that CNN predicted monthly rainfall during the training and validation procedures in good agreement with the observed rainfall.

Every time that an ANN was trained, the training time was approximately 5, 13 and 19 min for MLP, CNN and LSTM architectures, respectively. Nevertheless, the training time depends on the available computing resources. In this work, we used a computer with 32 CPUs and 128 GB of RAM. It is worth noting that the time required for the training process also depends on the learning rate, which controls how much to change the model in response to the estimated error each time the model weights are updated. A small learning rate leads to a long training process, whereas a value too large may result in sub-optimal

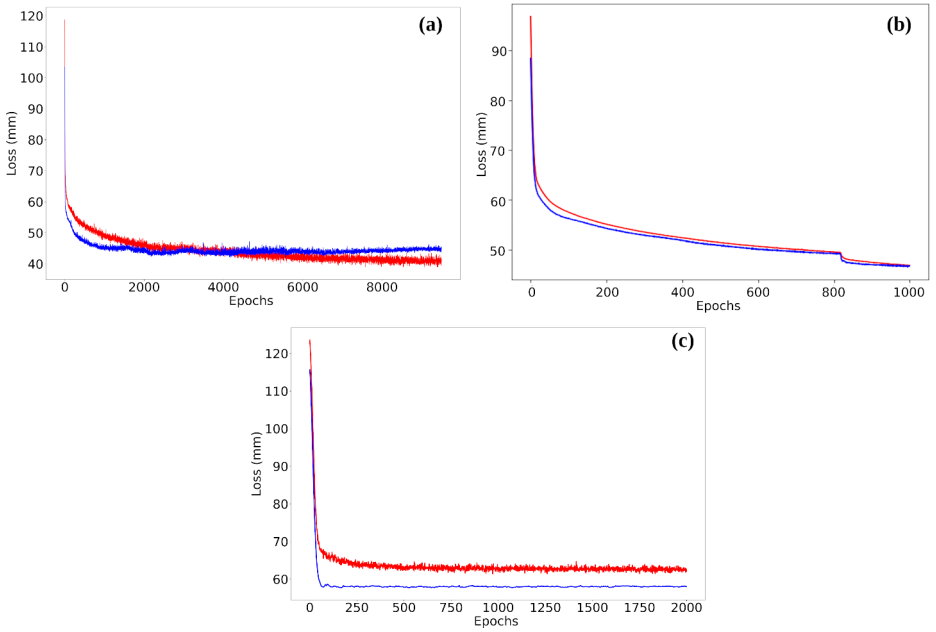

Fig. 5 Loss function while training the neural networks: **(a)** Convolutional Neural Network (CNN), **(b)** Multi–Layer Perceptron (MLP), and **(c)** Long Short-Term Memory (LSTM). The red and blue curves denote losses during training and validations phase, respectively

Table 2 Statistics for each ANN model during the training and testing phases. The epoch denotes when the best loss values were obtained during the training. Mean absolute error (MAE), BIAS and Root Mean Square Error (RMSE) are given in mm; and Pearson correlation coefficient (r_p) and Nash-Sutcliffe efficiency (NS) are dimensionless. Pearson correlation coefficient higher than 0.27 are statistically significant ($p<0.05$)

ANN	Epoch	Training					Testing				
		MAE	BIAS	RSME	r_p	NS	MAE	BIAS	RSME	r_p	NS
CNN	2898	30.50	-11.90	47.17	0.91	0.81	42.93	-5.39	49.28	0.80	0.63
MLP	995	57.472	-14.12	83.45	0.66	0.42	56.14	-7.33	86.15	0.65	0.42
LSTM	1805	59.43	-16.58	86.57	0.63	0.38	58.04	-9.15	86.12	0.64	0.41

learning weights. Furthermore, the differences observed in the training time of the three neural networks were due to the differences in complexity between them. As the trial-and-error method was applied, the total training time to determine the model with the most adjusted weights will also depend on the number of times each ANN architecture was trained. However, once the model that best fits the accumulated monthly rainfall during the training phase has been found, the predictions are obtained relatively quickly.

3.2 Comparative Analysis of the Forecasting Models

The rainfall data in the Almendares-Vento basin and climate indices from January 2015 to December 2019 were used to assess the ability of ANN and ARIMA models, as noted above. The validation of model performance was achieved based on several statistical metrics and

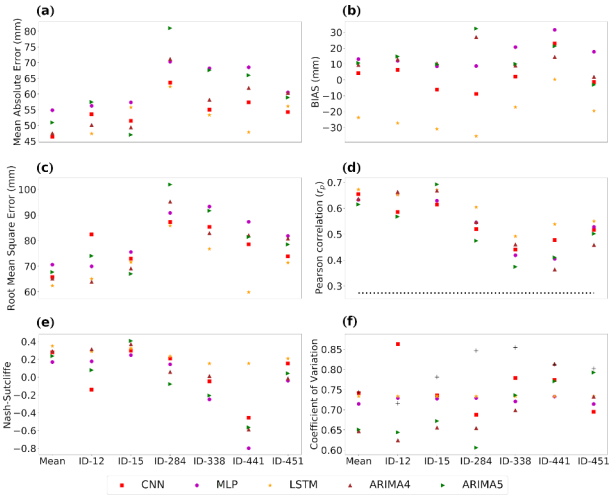

Fig. 6 Comparative analysis of the statistics of each forecasting model: **(a)** Mean Absolute Error, **(b)** BIAS, **(c)** Root Mean Square Error, **(d)** Pearson correlation coefficient, **(e)** Nash-Sutcliffe coefficient and **(f)** Coefficient of variation. In **(d)** the horizontal black dashed line indicates the threshold (r_p =0.27) for statistically significant r_p at 95% significance level. In **(f)** the marker "+" denotes the coefficient of variation of observed data in each gauge station

model performance criteria. Figure 6 and Table SM1 show the statistics for each predicting model. Apparently, the LSTM model exhibits the best performance by analyzing the mean rainfall over the whole basin and in the previously selected gauge stations. In agreement with Kim and Bae (2017), Kim and Won (2018) and Kumar et al. (2019), the LSTM performs better in terms of the RSME, MAE and NS. However, and although several studies (e.g., Kim and Bae 2017; Liu et al. 2017; Kratzert et al. 2018; Kim and Won 2018; Kumar et al. 2019) suggest the superiority of LSTM models, we found in this study that it is the worst to predict the peaks of maximum rainfall, as shown in Fig. 7 (for the mean rainfall) and Figure SM2 (for the monthly total precipitation in each gauge station). The main reason for such behaviour is that the number of heavy rainfall events was scarce in the data, and thus, it was difficult for ANN models, and particularly for the LSTM, to learn such features. Besides, very few rain events might not activate neurons (Zhang et al. 2021). As the LSTM model has the capability of learning long-term dependencies, probably the size of the input data for training was too small for the optimal fit of the model weights. It is important to remark that Fig. 7 also shows that the model simulated rainfalls preserved the seasonality.

Moreover, there is consistency in the results between precipitation time-series predicted by CNN, MLP and ARIMA models and the observed series, but the CNN model shows better performance. This result agrees with the findings of Haidar and Verma (2018), who noted the ability of CNN models to predict rainfall time series. From Fig. 6 (see also Table SM1), the CNN predicted the mean rainfall with the lowest MAE (46.41 mm), BIAS (4.22 mm) and RSME (65.78 mm), and it also showed a skill in predicting the monthly rainfall on the gauge stations. It must be highlighted that the CNN showed a statistically significant Pearson correlation coefficient ($p<0.05$) between the predicted and observed rainfall in both mean rainfall and gauge stations. Furthermore, the CNN model predicted the mean rainfall with the least difference between the coefficient of variation of the observed and simulated data. Overall, the CNN model has the capacity to cope with the time series data with non-stationarity and seasonal feature of rainfall.

To assess the proper functioning of each model to predict the monthly rainfall, we computed the KGE index over the mean rainfall and over the rainfall in selected gauge stations. By definition, KGE includes several metrics as the Pearson correlation coefficient, the

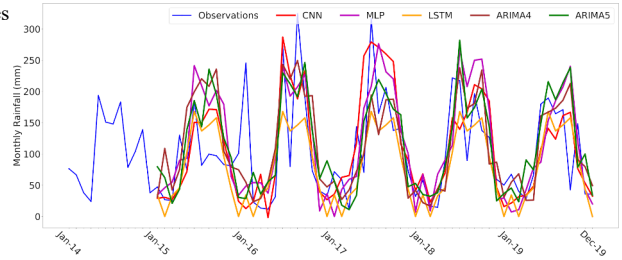

Fig. 7 Mean observed rainfall values and the ANN and ARIMA models outputs in the whole basin from January 2015 to December 2019

standard deviation and the mean (Eq. 7), therefore, it provides a more precise estimation of each model accuracy (Osuch et al. 2015; Towner et al. 2019). As noted above, higher KGE values are indicative of better model performance. The validation results by applying KGE index are shown in Fig. 8. The KGE values of mean rainfall predictions ranged from 0.53 (for the LSTM model) to 0.65 (for the CNN), which confirms the ability of CNN to estimate the monthly mean rainfall over the study region. Additionally, our analysis indicates that in general, ARIMA models are better predicting rainfall in selected gauge stations. It must be noted that the KGE value for LSTM is higher than all in the station gauge ID-441. Based on these findings, the CNN and ARIMA4 models exhibit the best performance again. The ability of the CNN model could be attributed to its filter to capture certain recurrent patterns, and then, try to forecast the future values. Apparently, this property of the CNN model favours their applications on time series forecasting (Fotovatikhah et al. 2018; Moazenzadeh et al. 2018; Chong et al. 2020). The overall low KGE values in station gauges ID-284, ID-338, ID-441, and ID-451 can be attributed to the bias magnitude of model predictions. This behavior agrees with Gebremichael et al. (2022), who found that the KGE value tends to decrease when the model overestimates the precipitation. It is worth noting that the data used from stations ID-12 and ID-15 represent 10.5% and 26.4% of the dataset, while the data for the remaining stations account for 8.8% (ID-338), 4.6% (ID-451), 4.1% (ID-441) and 3.4% (ID-284). Therefore, the best performance of all models for the mean rainfall in the whole basin and gauge stations ID-12 and ID-15 could be attributed to more data availability during the training process. Accordingly, the ANN and ARIMA models better learned the rainfall variability in stations ID-12 and ID-15 than in the other stations. This hypothesis agrees with Abiodun et al. (2018) and Hijazi et al. (2020), who noted that ANNs usually need big databases for training the network to accomplish substantial prediction accuracy.

In order to compare the performance of the models in terms of estimating the mean monthly rainfall, the KGE skill score was also used (Eq. 8). This parameter makes it possible to assess the relative improvement achieved when using one model compared to another, taking into account that the denominator of the KGEss expression measures the difference between the KGE value obtained for a perfect fit (1.0) and the KGE obtained for the comparative base model, i.e., the maximum possible improvement that would be possible to achieve, and the numerator measures the difference between the KGE of the model to be compared with the KGE obtained for the comparative base model, i.e., the actual improvement in fit achieved by using another model. The maximum possible value of the KGEss is 1, assuming that the model selected for comparison with the base model perfectly fits the data with a KGE equal to 1. In this way, the magnitude of the KGEss value is directly proportional to the degree of improvement of the mean monthly rainfall estimation with the

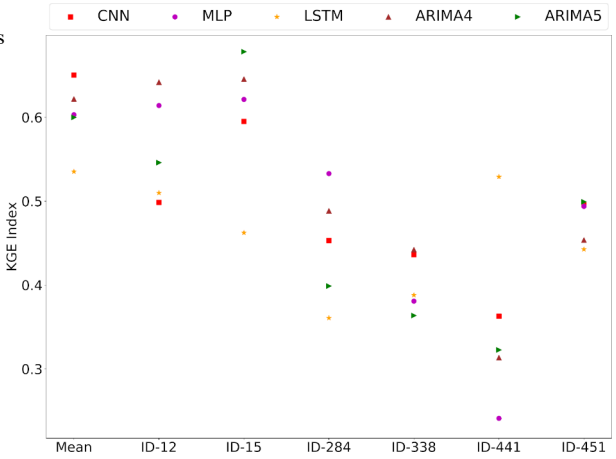

Fig. 8 Kling–Gupta efficiency (KGE) for ANN and ARIMA models to predict the mean rainfall over the Almendares-Vento basin and in gauge stations from January 2015 to December 2019

reference model with respect to the base model of the comparison. The other issue that must be taken into account is that the positive value of KGEss indicates that the model which is compared with the base model has a better prediction performance.

Figure 9 displays the KGE skill score of all models compared against each other in predicting the mean rainfall over the whole basin. Simple inspection of the figure reveals that CNN outperformed all models, followed by ARIMA4>MLP>ARIMA5>LSTM. The improvement ranges from 0.25 between CNN and LSTM models to 0.008 between MLP and ARIMA5 models. Note that CNN improvement over ARIMA4, the second highest performing model, was 0.075, a value that represents 27% of improvement relative to the minimum improvement achieved between two models in the range.

The KGE skill score for predicting the rainfall in the station gauges (Figure SM3) confirmed the results shown in Fig. 8. ARIMA4 showed the highest improvement in station ID-12, ARIMA5 in ID-15, MLP in ID-284, LSTM in ID-441 and ARIMA5 in ID-451, while ARIMA4 and CNN were better at predicting the monthly rainfall time series in ID-338. Interestingly, the LSTM improvement over the other models at station ID-441 ranged from 0.26 to 0.38, which means that the performance of LSTM model in this gauge station is much better than the rest of the models. Overall, the KGE skill score confirmed that CNN and ARIMA4 performed better.

To investigate further, we analyzed the ability of each model to predict the rainfall amounts in each month from 2015 to 2019. Table 3 shows the RMSE, BIAS and KGE values of the mean rainfall series predicted for each month of the year. In terms of RSME, the CNN model is better in February, April, June and November, the LSTM performs best in March, July, September and October, while ARIMA4 is most skillful in January, August and December and ARIMA5 in May and July. In terms of BIAS, CNN (August, October and November), MLP (February, April, May and June) and LSTM (March, July and September) show the best performance. The Wilcoxon Signed Rank Test revealed that the CNN model has the overall best performance in terms of the RMSE, confirming the results discussed above. Additionally, CNN exhibits improved forecast capacity regarding KGE in six months of the year (February, April, June, September, October and November), while the ARIMA4 shows the better prediction for August and December. Likewise, ARIMA5 is better for predicting in May and July, and finally, the LSTM shows the better prediction

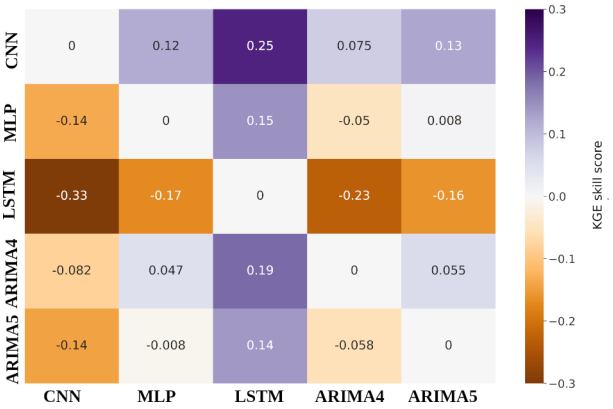

Fig. 9 Kling–Gupta efficiency skill score (KGS) in predicting the mean rainfall over the Almendares-Vento basin from January 2015 to December 2019

of rainfall only in March. Similar results were found for the monthly rainfall prediction each month alone at every gauge station (Tables SM2 to SM7). It is outstanding that the worst performance of all models occurred in January, which can be related to the outlier of 240 mm registered for that month (Fig. 7). It is well known that the weather of January 2016 was described as extremely wet, because of 99.7 mm of average rainfall was recorded for the Cuban archipelago (212% of the historical value). Specifically, 178.1 mm (339%) for the western region and 298.3 mm (424%) for Havana. This unusual behavior was conditioned by the presence of a strong El Niño/Southern Oscillation (ENSO) event in the equatorial Pacific Ocean (González-Pedroso and Estévez 2016). On average, according to these findings, the CNN model performs better than all other models.

From the statistical indices shown in Table 3, a hybrid model based on the ANN and ARIMA models was developed in order to combine the forecasting model efficiencies, as previously suggested by Zhang (2003) and Haidar and Verma (2018). In this contribution, for each month, the maximum KGE value was selected as the decision criterion. This strategy aims at increasing computational efficiency by growing correlations and minimizing errors during the forecasting process. The temporal variations of the observed and predicted rainfall by the hybrid model are presented in Fig. 10 and SM4. Both Figures reveal that the hybrid model fits better the observed rainfall values than the prediction by the individual models. Precisely, the results of hybrid model implementation are presented in Table 4. The statistical evaluation shows high accuracy of hybrid model based on RMSE, BIAS, r_p, NS and KGE indices. The KGE records for the mean rainfall and selected gauge stations are notably higher than 0.5 and the NS values are always positive, suggesting the ability of the hybrid model in predicting monthly rainfall in the study region. Following the breakdown of KGE values into four benchmark categories (Kling et al. 2012), the performance of the hybrid model could be classified from Intermediate ($0.5 \leq KGE < 0.75$) to Good ($KGE \geq 0.75$). This approach was previously applied by Thiemig et al. (2013) for evaluating the satellite-based precipitation products and Towner et al. (2019) for assessing the performance of global hydrological models in capturing peak river flows in the Amazon basin. On the other hand, Xu et al. (2020) found NS values ranging from 0.35 to 0.75 in the runoff prediction in the Hun River basins, while Bhagwat and Maity (2012) previously achieved NS values from 0.58 to 0.68. Overall, the hybrid model better captured particular features of the data set than individual models, in agreement with previous works (e.g., Zhang 2003; Aladag et al. 2012; Dodangeh et al. 2020).

Table 3 RMSE, BIAS and KGE values for the mean rainfall series estimated for each month of the year from 2015 to 2019. The best statistics for each month are highlighted in bold text. The best model to predict the rainfall each month was selected based on the maximum value of the KGE index

Months	CNN			MLP			LSTM			ARIMA4			ARIMA5			Best Model
	RMSE	**BIAS**	**KGE**	**RMSE**	**BIAS**	**KGE**	**RMSE**	**BIAS**	**KGE**	**RMSE**	**BIAS**	**KGE**	**RMSE**	**BIAS**	**KGE**	
Jan	107.83	-64.57	-0.850	97.814	-66.53	-0.196	97.779	-63.56	-0.778	**87.560**	-54.03	**-0.116**	99.312	**-48.90**	-0.851	ARIMA4
Feb	**5.3051**	1.2871	**0.9229**	22.391	**-0.431**	0.1056	36.338	-33.10	-0.289	41.316	15.791	-0.740	33.747	8.2278	-0.701	CNN
Mar	36.749	29.034	-1.797	39.548	37.050	-1.628	**16.002**	**14.464**	**-0.278**	21.845	18.126	-0.684	31.390	20.580	-2.221	LSTM
Apr	**41.382**	-31.86	**0.5549**	41.865	**-16.30**	0.2743	63.561	-42.51	-0.314	52.789	-34.20	0.1490	65.760	-30.72	-0.677	CNN
May	95.000	16.840	0.1281	84.019	**-2.980**	-0.305	73.478	-12.78	-0.609	67.583	22.574	0.1855	**62.293**	14.152	**0.2797**	ARIMA5
Jun	**47.409**	-37.03	**0.6129**	62.777	**-14.22**	0.0633	85.319	-68.44	-0.785	58.313	-28.12	0.1438	67.865	-16.10	-0.193	CNN
Jul	94.324	71.054	-0.111	105.54	98.060	-0.035	**44.783**	**20.584**	-0.252	96.864	67.294	-0.900	76.562	67.645	**0.1890**	ARIMA5
Aug	73.419	**-1.232**	0.0651	70.044	14.954	0.0634	90.214	-53.29	-0.923	**60.450**	2.6282	**0.2464**	87.881	2.8052	-0.798	ARIMA4
Sep	88.496	83.333	**0.2092**	119.21	106.88	-0.509	**61.460**	**36.714**	-0.989	109.48	91.356	-0.703	106.23	89.826	-0.323	CNN
Oct	46.957	**3.9184**	**0.0192**	57.997	35.723	-0.572	**35.621**	-18.29	-0.453	63.619	4.7517	-0.666	47.957	15.197	-0.207	CNN
Nov	**12.959**	**2.6076**	**0.7450**	25.562	-9.568	0.1087	22.509	-14.04	-0.317	31.567	8.9735	-0.254	36.006	7.3443	-0.292	CNN
Dec	37.935	-22.40	-0.642	33.994	-25.08	0.0677	56.834	-50.55	-0.334	**17.429**	**-1.684**	**0.5010**	41.620	-1.886	-0.530	ARIMA4

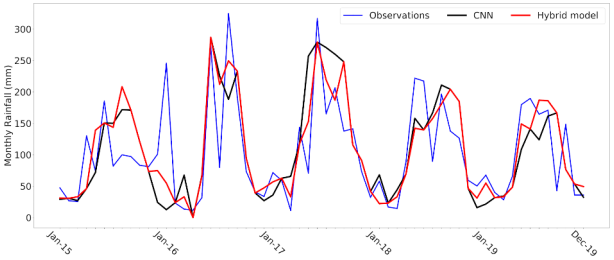

Fig. 10 Mean rainfall observed values, the CNN and the hybrid model outputs in the whole basin from January 2015 to December 2019

Table 4 RMSE, BIAS, r_p, NS and KGE values obtained using the hybrid model for predicting the mean rainfall over the whole basin and at each gauge station

	Mean	ID-12	ID-15	ID-284	ID-338	ID-441	ID-451
RMSE	54.632	58.498	61.772	76.596	73.154	65.434	66.523
BIAS	5.4518	6.1675	0.6445	7.9766	2.8237	14.193	0.2577
r_p	0.7479	0.7294	0.7282	0.6804	0.5684	0.5864	0.6249
NS	0.5027	0.4246	0.4964	0.3907	0.2310	0.1100	0.3127
KGE	0.7413	0.7194	0.7143	0.6669	0.5484	0.5287	0.6127

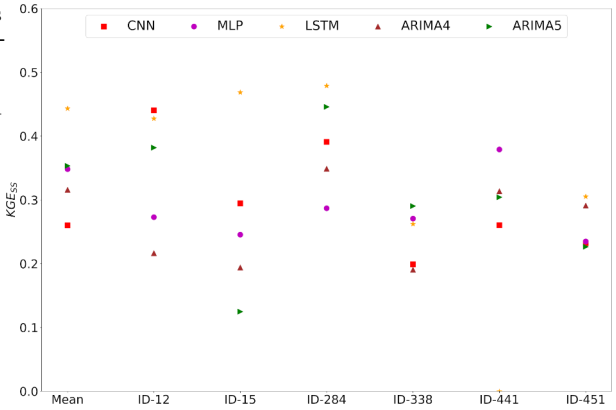

Fig. 11 Hybrid model improvements in the prediction of the monthly rainfall time series in the Almendares-Vento basin from January 2015 to December 2019 based on the Kling–Gupta efficiency skill score (KGS)

Figure 11 highlights the improvements of the hybrid model in forecasting the mean monthly rainfall in the whole Almendares-Vento watershed and in each gauge station. The fact that all KGEss values are positive means that the hybrid model is better predicting the mean monthly rainfall than the rest of the models in the whole basin and in each gauge station, except at station ID-441, where the LSTM model has a similar performance. The improvement ranges from 0.48 between hybrid model and LSTM models at gauge station ID-284 to 0 between the same models at ID-441. For the whole basin the improvement in the mean monthly rainfall prediction ranges from 0.26 with respect to CNN model to 0.44 with respect to LSTM model. This improvement could be considered as significant.

4 Conclusions

We utilized the Multi-Layer Perceptron (MLP), Convolutional Neural Network (CNN), Long Short-Term Memory (LSTM) artificial neural networks (ANNs) models, and the Autoregressive Integrated Moving Average (ARIMA) models in developing a hybrid model (ANN+ARIMA) with the aim to forecast the monthly rainfall totals in hydrological watersheds. The region was the Almendares-Vento basin, which is an important basin in Cuba. The ANN models were trained using the monthly rainfall, sunspots, sea surface temperature and climatic variability modes (NAO, Niño 1.2, Niño 3.0, Niño 3.0, Niño 3.4, Niño 4.0, AMO, SOI) time series of the previous calendar year to predict the rainfall of the next calendar year, while ARIMA models were adjusted using the rainfall seasonality.

Additionally, to assess the ability of each model to predict monthly rainfall in the study region, both ANN and ARIMA models were compared. This study showed in general that the CNN model has a good performance to forecast monthly rainfall amounts in the Almendares-Vento basin. However, better accuracy was reached through a hybridization of ANN and ARIMA models, with maximum monthly KGE as the selection criteria. This approach ensures an efficient performance, as the statistical indices MAE, BIAS, RMSE, r_p, NS, and KGE for mean rainfall and selected gauge stations in the Almendares-Vento basin unfolded.

Although this work was focused on forecasting the monthly rainfall over the Almendares-Vento basin, these results prove the reliability of using the hybrid model (ANN+ARIMA) to predict rainfall time series for water management. Monthly time series of precipitation recorded in gauge stations within the basin and climatic indices used as input data in our methodology also allows the ANN and ARIMA models to learn the seasonal variations of the rainfall, improving the accuracy of prediction and consequently the ability of the hybrid model. Additionally, our method can be easily applied in forecasting rainfall in other hydrological basins.

In summary, this work proposes taking a new detour to enhance the water management system in hydrological basins through improving rainfall forecasting. However, the main limitation for the extensive application of the present approach could be the non-availability of long-term time series of monthly rainfall within a basin. In future studies, the network model structure is going to be assessed and further optimized to accomplish a more accurate prediction.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s40710-022-00602-x.

Acknowledgements The authors thank the National Institute of Hydraulic Resources (INRH, Spanish acronym) of Cuba for providing the rainfall records of each gauge station in the Almendares-Vento basin. The authors also acknowledge the free availability of climatic indices used in this study. Besides, we appreciate the English writing corrections made by Orailly Madruga Rios, professor and translator from Instituto Superior de Tecnologías y Ciencias Aplicadas, Universidad de La Habana.

Authors' contributions A. Pérez-Alarcón: Conceptualization, Data Curation, Methodology, Formal analysis, Software, Investigation, Validation, Visualization, Writing - original draft, Writing - review & editing. D. Garcia-Cortes: Conceptualization, Methodology, Investigation, Writing - review & editing, Supervision. J. C. Fernández-Alvarez: Conceptualization, Data Curation, Methodology, Investigation, Validation, Visualization, Writing - review & editing. Y. Martínez-González: Conceptualization, Data Curation, Methodology, Formal analysis, Software, Investigation, Validation, Writing - original draft, Writing - review & editing, Supervision.

Funding This work was supported by the "Expert system to support decision making during water supply management of the Albear aqueduct" project (grant No. PS113LH001-019) funded by the National Institute of Hydraulic Resources of Cuba.

Data Availability The python source codes developed to train the ANN models and MATLAB codes for ARIMA models are accessible to be downloaded for free at https://github.com/apalarcon/ANNs_train. The climatic indices databases utilized in the article were retrieved online paying no fees. The AMO index can be obtained from the NOAA/Physical Sciences Laboratory (https://www.psl.noaa.gov/data/timeseries/AMO/), the NAO from the Climatic Research Unit (https://crudata.uea.ac.uk/cru/data/nao/nao.dat), the SOI was taken from the Bureau of Meteorology (http://www.bom.gov.au/climate/current/soihtm1.shtml), the mean SST values in the Niño 1.2, Niño 3.0, Niño 4.0 and Niño 3.4 regions were obtained from the Global Climate Observing System (https://psl.noaa.gov/gcoswgsp/), while sunspots were retrieved from the Solar Influences Data Analysis Center (http://sidc.oma.be/silso/datafiles). Additionally, the Centennial Time Scale (COBE SST2) dataset of the National Oceanic and Atmospheric Administration (NOAA) was used, which can be found at https://psl.noaa.gov/data/gridded/data.cobe2.html.

Declarations

Competing Interests The authors have no competing interests to declare that are relevant to the content of this article.

Ethics approval Not applicable.

Consent to participate Not applicable.

Consent for publication Not applicable.

References

Aladag CH, Egrioglu E, Kadilar C (2012) Improvement in forecasting accuracy using the hybrid model of ARFIMA and feed forward neural network. Am J Intell Syst 2(2):12–17. https://doi.org/10.5923/j.ajis.20120202.02

Abiodun OI, Jantan A, Omolara AE, Dada KV, Mohamed NA, Arshad H (2018) State-of-the-art in artificial neural network applications: A survey. Heliyon 4:e00938. https://doi.org/10.1016/j.heliyon.2018.e00938

Aguado-Rodríguez GJ, Quevedo-Nolasco A, Castro-Popoca M, Arteaga-Ramírez R, Vázquez-Peña MA, Zamora-Morales BP (2016) Meteorological variables prediction through arima models. Agrociencia 50:1–13

Ali M, Prasad R, Xiang Y, Yaseen ZM (2020) Complete ensemble empirical mode decomposition hybridized with random forest and kernel ridge regression model for monthly rainfall forecasts. J Hydrol 584:124647. https://doi.org/10.1016/j.jhydrol.2020.124647

Anochi JA, de Almeida VA, de Campos Velho HF (2021) Machine learning for climate precipitation prediction modeling over South America. Remote Sens 13:2468. https://doi.org/10.3390/rs13132468

Band SS, Janizadeh S, Chandra Pal S, Saha A, Chakrabortty R, Melesse AM, Mosavi A (2020a) Flash flood susceptibility modeling using new approaches of hybrid and ensemble tree-based machine learning algorithms. Remote Sens 12(21):3568. https://doi.org/10.3390/rs12213568

Band SS, Janizadeh S, Chandra Pal S, Saha A, Chakrabortty R, Shokri M, Mosavi A (2020b) Novel ensemble approach of deep learning neural network (DLNN) model and particle swarm optimization (PSO) algorithm for prediction of gully erosion susceptibility. Sensors 20(19):5609. https://doi.org/10.3390/s20195609

Bai LH, Xu H (2022) Accurate storm surge forecasting using the encoder–decoder long short term memory recurrent neural network. Phys Fluids 34(1):016601. https://doi.org/10.1063/5.0081858

Barrera-Animas AY, Oyedele LO, Bilal M, Akinosho TD, Delgado JMD, Akanbi LA (2022) Rainfall prediction: A comparative analysis of modern machine learning algorithms for time-series forecasting. Mach Learn Appl 7:100204. https://doi.org/10.1016/j.mlwa.2021.100204

Benevides P, Catalao J, Nico G (2019) Neural network approach to forecast hourly intense rainfall using GNSS precipitable water vapor and meteorological sensors. Remote Sens 11:966. https://doi.org/10.3390/rs11080966

Bhagwat PP, Maity R (2012) Multistep-ahead river flow prediction using LS-SVR at daily scale. J Water Resour Prot 4:528–539. https://doi.org/10.4236/jwarp.2012.47062

Bhuiyan AE, Yang F, Biswas NK, Rahat SH, Neelam TJ (2020) Machine learning-based error modeling to improve GPM IMERG precipitation product over the Brahmaputra River Basin. Forecasting 2:248–266. https://doi.org/10.3390/forecast2030014

Bonakdari H, Moeeni H, Ebtehaj I, Zeynodin M, Mohammadian M, Gharabaghi B (2019) New insights into soil temperature time series modeling: linear or nonlinear? Theore Appl Clim 135:1157–1177. https://doi.org/10.1007/s00704-018-2436-2

Box GE, Jenkins GM, Reinsel GC (1994) Time Series Analysis, Forecasting and Control. Englewood Clifs:197–199

Chollet F (2015) Keras. https://github.com/fchollet/keras. Accessed on 16 June 2022

Chao Z, Pu F, Yin Y, Han B, Chen X (2018) Research on real-time local rainfall prediction based on MEMS sensors. J Sens 2018:6184713. https://doi.org/10.1155/2018/6184713

Chen Y, Jiang H, Li C, Jia X, Ghamisi P (2016) Deep feature extraction and classification of hyperspectral images based on convolutional neural networks. IEEE Trans Geosci Remote Sens 54(10):6232–6251. https://doi.org/10.1109/TGRS.2016.2584107

Chong KL, Lai SH, Yao Y, Ahmed AN, Jaafar WZW, El-Shafie A (2020) Performance enhancement model for rainfall forecasting utilizing integrated wavelet-convolutional neural network. Water Resour Res 34(8):2371–2387. https://doi.org/10.1007/s11269-020-02554-z

Choubin B, Mosavi A, Alamdarloo EH, Hosseini FS, Shamshirband S, Dashtekian K, Ghamisi P (2019) Earth fissure hazard prediction using machine learning models. Environ Res 179:108770. https://doi.org/10.1016/j.envres.2019.108770

Derin Y, Bhuiyan M, Anagnostou E, Kalogiros J, Anagnostou MN (2020) Modeling level 2 passive microwave precipitation retrieval error over complex terrain using a nonparametric statistical technique. IEEE Trans Geosci Remote Sens 59(11):9021–9032. https://doi.org/10.1109/TGRS.2020.3038343

Diodato N, Bellocchi G (2018) Using historical precipitation patterns to forecast daily extremes of rainfall for the coming decades in Naples (Italy). Geosciences 8:293. https://doi.org/10.3390/geosciences8080293

Dodangeh E, Choubin B, Eigdir AN, Nabipour N, Panahi M, Shamshirband S, Mosavi A (2020) Integrated machine learning methods with resampling algorithms for flood susceptibility prediction. Sci Total Environ 705:135983. https://doi.org/10.1016/j.scitotenv.2019.135983

Dounia M, Sabri D, Yassine D (2014) Rainfall–Rain off Modeling Using Artificial Neural Network. APCBEE Proc 10:251–256. https://doi.org/10.1016/j.apcbee.2014.10.048

Ebecken N (2011) An overview on the use of neural networks for data mining tasks. J Braz Neural Netw Soc 9:202–212

Ebtehaj I, Bonakdari H, Gharabaghi B (2019) A reliable linear method for modeling lake level fluctuations. J Hydrol 570:236–250. https://doi.org/10.1016/j.jhydrol.2019.01.010

Enfield DB, Mestas-Nuñez AM, Trimble PJ (2001) The Atlantic multidecadal oscillation and its relation to rainfall and river flows in the continental US. Geophys Res Lett 28:2077–2080. https://doi.org/10.1029/2000GL012745

Fang L, Shao D (2022) Application of long short-term memory (LSTM) on the prediction of rainfall-runoff in Karst area. Front Phys 9:790687. https://doi.org/10.3389/fphy.2021.790687

Faruk DO (2010) A hybrid neural network and SARIMA model for water quality time series prediction. Eng Appl Artif Intell 23(4):586–594. https://doi.org/10.1016/j.engappai.2009.09.015

Fathi E, Shoja BM (2018) Deep neural networks for natural language processing, in: Handbook of Statistics. Elsevier. Vol. 38, pp. 229–316. https://doi.org/10.1016/bs.host.2018.07.006

Fernández-Alvarez JC, Hernández-López I, Cruz-Cobas PP, Cárdenas-Díaz T, Batista-Leyva AJ (2019) Using a multilayer perceptron in intraocular lens power calculation. J Cataract Refract Surg 45:1753–1761. https://doi.org/10.1016/j.jcrs.2019.07.035

Fotovatikhah F, Herrera M, Shamshirband S, Chau KW, Faizollahzadeh Ardabili S, Piran MJ (2018) Survey of computational intelligence as basis to big flood management: Challenges, research directions and future work. Eng Appl Comput Fluid Mech 12(1):411–437. https://doi.org/10.1080/19942060.2018.1448896

Gamboa-Villafruela CJ, Fernández-Alvarez JC, Márquez-Mijares M, Pérez-Alarcón A, Batista-Leyva AJ (2021) Convolutional LSTM Architecture for Precipitation Nowcasting Using Satellite Data. Environ Sci Proc 8(1):33. https://doi.org/10.3390/ecas2021-10340

García-Fernández JM, Díaz JG (2017) A 20 Años de la Creación de los Consejos de Cuenca. In: I Taller de Gestión Integral de Cuencas Hidrográficas. Cubagua, La Habana, Cuba, p. 20

Gebremichael M, Yue H, Nourani V, Damoah R (2022) The skills of medium-range precipitation forecasts in the Senegal River Basin. Sustainability 14:3349. https://doi.org/10.3390/su14063349
Ghazvinei PT, Darvishi HH, Mosavi A, Yusof KBW, Alizamir M, Shamshirband S, Chau KW (2018) Sugarcane growth prediction based on meteorological parameters using extreme learning machine and artificial neural network. Eng Appl Comput Fluid Mech 12(1):738–749. https://doi.org/10.1080/19942060.2018.1526119
Ghorbani MA, Deo RC, Yaseen ZM, Kashani MH, Mohammadi B (2018) Pan evaporation prediction using a hybrid multilayer perceptron-firefly algorithm (MLP-FFA) model: Case study in North Iran. Theor Appl Climatol 133:1119–1131. https://doi.org/10.1007/s00271-010-0225-5
Ghumman A, Ghazaw YM, Sohail A, Watanabe K (2011) Runoff forecasting by artificial neural network and conventional model. Alex Eng J 50:345–350. https://doi.org/10.1016/j.aej.2012.01.005
Girihagama L, Naveed Khaliq M, Lamontagne P, Perdikaris J, Roy R, Sushama L, Elshorbagy A (2022) Streamflow modelling and forecasting for Canadian watersheds using LSTM networks with attention mechanism. Neural Comput Appl. https://doi.org/10.1007/s00521-022-07523-8
Glorot X, Bordes A, Bengio Y (2011) Deep sparse rectifier networks. In: 14th International Conference on Artificial Intelligence and Statistics. JMLR W&CP, pp. 315–323, Fort Lauderdale, FL, USA
González-Pedroso C, Estévez G (2016) Summary of the winter season 2015–2016. Rev Cub Met 22:216–224
Gupta H, Bastidas L, Sorooshian S, Shuttleworth W, Yang Z (2009) Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. J Hydrol 377:80–91. https://doi.org/10.1016/j.jhydrol.2009.08.003
Haidar A, Verma B (2018) Monthly rainfall forecasting using one-dimensional deep convolutional neural network. IEEE Access 6:69053–69063. https://doi.org/10.1109/ACCESS.2018.2880044
Hardwinarto S, Aipassa M et al (2015) Rainfall monthly prediction based on artificial neural network: a case study in Tenggarong Station, East Kalimantan-Indonesia. Proc Comp Sci 59:142–151. https://doi.org/10.1016/j.procs.2015.07.528
Hernández N, Camargo J, Moreno F, Plazas-Nossa L, Torres A (2017) Arima as a forecasting tool for water quality time series measured with UV-Vis spectrometers in a constructed wetland. Tecnol Cienc Agua 8(5):127–139. https://doi.org/10.24850/j-tyca-2017-05-09
Hijazi A, Al-Dahidi S, Altarazi S (2020) A novel assisted artificial neural network modeling approach for improved accuracy using small datasets: Application in residual strength evaluation of panels with multiple site damage cracks. Appl Sci 10. https://doi.org/10.3390/app10228255
Hirahara S, Ishii M, Fukuda Y (2014) Centennial-scale sea surface temperature analysis and its uncertainty. J Clim 27:57–75. https://doi.org/10.1175/JCLI-D-12-00837.1
Hochreiter S, Schmidhuber J (1997) Long short-term memory. Neural Comput 9:1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735
Hong WC (2008) Rainfall forecasting by technological machine learning models. Appl Math Comput 200:41–57. https://doi.org/10.1016/j.amc.2007.10.046
Hossain M, Rahman M, Prodhan UK, Khan M et al (2013) Implementation of back-propagation neural network for isolated Bangla speech recognition. Int J Inf Sci 3(4):1–9. https://doi.org/10.5121/ijist.2013.3401
Houska T, Multsch S, Kraft P, Frede HG, Breuer L (2014) Monte Carlo-based calibration and uncertainty analysis of a coupled plant growth and hydrological model. Biogeosciences 11(7):2069–2082. https://doi.org/10.5194/bg-11-2069-2014
Hung NQ, Babel MS, Weesakul S, Tripathi N (2009) An artificial neural network model for rainfall forecasting in Bangkok, Thailand. Hydrol Earth Syst Sci 13:1413–1425. https://doi.org/10.5194/hess-13-1413-2009
Hurrell JW (1995) Decadal trends in the North Atlantic Oscillation: Regional temperatures and precipitation. Science 269:676–679. https://doi.org/10.1126/science.269.5224.676
Hussain JS, Al-Khazzar A, Raheema MN (2020) Recognition of new gestures using myo armband for myoelectric prosthetic applications. Int J Electr Comput Eng 10(6):2088–8708. https://doi.org/10.11591/ijece.v10i6.pp5694-5702
Hyndman RJ, Athanasopoulos G (2018) Forecasting: principles and practice. Melbourne, Australia, OTexts
Hyndman RJ, Khandakar Y (2008) Automatic Time Series Forecasting: The forecast Package for R. J Stat Softw 27:1–22. https://doi.org/10.18637/jss.v027.i03
Jeong K, Koo C, Hong T (2014) An estimation model for determining the annual energy cost budget in educational facilities using SARIMA (seasonal autoregressive integrated moving average) and ANN (artificial neural network). Energy 71:71–79. https://doi.org/10.1016/j.energy.2014.04.027
Jones PD, Jonsson T, Wheeler D (1997) Extension to the North Atlantic Oscillation using early instrumental pressure observations from Gibraltar and south-west Iceland. Int J Climatol 17(13):1433-1450. https://doi.org/10.1002/(SICI)1097-0088(19971115)17:13<1433::AID-JOC203>3.0.CO;2-P

Kashiwao T, Nakayama K, Ando S, Ikeda K, Lee M, Bahadori A (2017) A neural network-based local rainfall prediction system using meteorological data on the Internet: A case study using data from the Japan Meteorological Agency. Appl Soft Comput 56:317–330. https://doi.org/10.1016/j.asoc.2017.03.015

Keijsers N (2010) Neural Networks. In: Kompoliti K, Metman LV (eds) Encyclopedia of Movement Disorders. Academic Press, Oxford, pp 257–259. https://doi.org/10.1016/B978-0-12-374105-9.00493-7.

Kim HU, Bae TS (2017) Preliminary study of deep learning-based precipitation. J Korean Soc Surv Geod Photogramm Cartogr 35(5):423–430. https://doi.org/10.7848/ksgpc.2017.35.5.423

Kim HY, Won CH (2018) Forecasting the volatility of stock price index: A hybrid model integrating LSTM with multiple GARCH-type models. Expert Syst Appl 103: 25-37. https://doi.org/10.1016/j.eswa.2018.03.002

Kingma DP, Ba J (2017) Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980v9. https://arxiv.org/abs/1412.6980v9

Kling H, Fuchs M, Paulin M (2012) Runoff conditions in the upper Danube basin under an ensemble of climate change scenarios. J Hydrol 424:264–277. https://doi.org/10.1016/j.jhydrol.2012.01.011

Knoben WJ, Freer JE, Woods RA (2019) Inherent benchmark or not? comparing Nash–Sutcliffe and Kling–Gupta efficiency scores. Hydrol Earth Syst Sci 23:4323–4331. https://doi.org/10.5194/hess-23-4323-2019

Kotu V, Deshpande B (2019) Time series forecasting. In: Kotu V, Deshpande B (Eds.), Data Science. Second ed. Morgan Kaufmann, pp. 395–445. https://doi.org/10.1016/B978-0-12-814761-0.00012-5

Koutroumanidis T, Ioannou K, Arabatzis G (2009) Predicting fuelwood prices in Greece with the use of SARIMA models, artificial neural networks and a hybrid SARIMA–ANN model. Energy Policy 37(9):3627–3634. https://doi.org/10.1016/j.enpol.2009.04.024

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(11):6005–6022. https://doi.org/10.5194/hess-22-6005-2018

Krasnopolsky VM, Lin Y (2012) A neural network nonlinear multimodel ensemble to improve precipitation forecasts over continental US. Adv Meteorol 2012. https://doi.org/10.1155/2012/649450

Kumar D, Singh A, Samui P, Jha RK (2019) Forecasting monthly precipitation using sequential modelling. Hydrol Sci J 64(6):690–700. https://doi.org/10.1080/02626667.2019.1595624

Lee J, Kim CG, Lee JE, Kim NW, Kim H (2018) Application of artificial neural networks to rainfall forecasting in the Geum river basin. Korea Water 10:1448. https://doi.org/10.3390/w10101448

Li P, Lai ES (2004) Short-range quantitative precipitation forecasting in Hong Kong. J Hydrol 288:189–209. https://doi.org/10.1016/j.jhydrol.2003.11.034

Li Y, Yang R, Yang C, Yu M, Hu F, Jiang Y (2017) Leveraging LSTM for rapid intensifications prediction of tropical cyclones. ISPRS Ann Photogramm Remote Sens Spatial Inf Sci 4(2):101–105. https://doi.org/10.5194/isprs-annals-IV-4-W2-101-2017

Li P, Zhang J, Krebs P (2022) Prediction of flow based on a CNN-LSTM combined deep learning approach. Water 14:993. https://doi.org/10.3390/w14060993

Liu YP, Hou D, Bao JP, Qi Y (2017) Multi-step ahead time series forecasting for different data patterns based on LSTM recurrent neural network, 14th Web Information Systems and Applications Conference (WISA), pp. 305–310, Liuzhou, China, 11–12 November 2017. https://doi.org/10.1109/WISA.2017.25

Luk KC, Ball JE, Sharma A (2001) An application of artificial neural networks for rainfall forecasting. Math Comput Model 33:683–693. https://doi.org/10.1016/S0895-7177(00)00272-7

Mahmud I, Bari SH, Rahman M (2017) Monthly rainfall forecast of Bangladesh using autoregressive integrated moving average method. Environ Eng Res 22(2):162–168. https://doi.org/10.4491/eer.2016.075

Majumdar SJ, Sun J, Golding B, Joe P, Dudhia J, Caumont O, Chandra Gouda K, Steinle P, Vincendon B, Wang J, Yussouf N (2021) Multiscale forecasting of high-impact weather: Current status and future challenges. Bull Am Meteorol 102:E635–E659. https://doi.org/10.1175/BAMS-D-20-0111.1

Moazenzadeh R, Mohammadi B, Shamshirband S, Chau KW (2018) Coupling a firefly algorithm with support vector regression to predict evaporation in northern Iran. ng Appl Comput Fluid Mech 12(1):584–597. https://doi.org/10.1080/19942060.2018.1482476

Moeeni H, Bonakdari H (2017) Forecasting monthly inflow with extreme seasonal variation using the hybrid SARIMA-ANN model. Stoch Envl Res Risk A 31(8):1997–2010. https://doi.org/10.1007/s00477-016-1273-z

Moradi H, Avand MT, Janizadeh S (2019) Landslide susceptibility survey using modeling methods. Spatial modeling in GIS and R for earth and environmental sciences. Elsevier, pp 259–275. https://doi.org/10.1016/B978-0-12-815226-3.00011-9

Narejo S, Jawaid MM, Talpur S, Baloch R, Pasero EGA (2021) Multi-step rainfall forecasting using deep learning approach. Peer J Comput Sci 7:e514. https://doi.org/10.7717/peerj-cs.514

Nastos P, Moustris K, Larissi I, Paliatsos A (2013) Rain intensity forecast using artificial neural networks in Athens, Greece. Atmos Res 119:153–160. https://doi.org/10.1016/j.atmosres.2011.07.020

Nourani V, Davanlou Tajbakhsh A, Molajou A, Gokcekus H (2019) Hybrid wavelet-M5 model tree for rainfall-runoff modeling. J Hydrol Eng 24:04019012. https://doi.org/10.1061/(ASCE)HE.1943-5584.0001777
Nourani V, Kisi Ö, Komasi M (2011) Two hybrid artificial intelligence approaches for modeling rainfall–runoff process. J Hydrol 402(1–2):41–59. https://doi.org/10.1016/j.jhydrol.2011.03.002
Osuch M, Romanowicz RJ, Booij MJ (2015) The influence of parametric uncertainty on the relationships between HBV model parameters and climatic characteristics. Hydrol Sci J 60:1299–1316. https://doi.org/10.1080/02626667.2014.967694
Panchal G, Ganatra A, Kosta Y, Panchal D (2011) Behaviour analysis of multilayer perceptrons with multiple hidden neurons and hidden layers. Int J Comput Theor Eng 3:332–337. https://doi.org/10.7763/IJCTE.2011.V3.328
Popescu MC, Balas VE, Perescu-Popescu L, Mastorakis N (2009) Multilayer perceptron and neural networks. WSEAS Trans Circuits Syst 8:579–588. https://doi.org/dl.acm.org/doi/abs/10.5555/1639537.1639542
Rayner N, Parker D, Horton E, Folland CK, Alexander LV, Rowell DP, Kent EC, Kaplan A (2003) Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. J Geophys Res Atmospheres 108:4407. https://doi.org/10.1029/2002JD002670
Ren X, Li X, Ren K, Song J, Xu Z, Deng K, Wang X (2021) Deep learning-based weather prediction: a survey. Big Data Res 23:100178. https://doi.org/10.1016/j.bdr.2020.100178
Rhee K, Shin HC (2018) Electromyogram-based hand gesture recognition robust to various arm postures. Int J Distrib Sens Netw 14(7):1550147718790751. https://doi.org/10.1177/1550147718790751
Ridwan WM, Sapitang M, Aziz A, Kushiar KF, Ahmed AN, El-Shafie A (2021) Rainfall forecasting model using machine learning methods: Case study Terengganu, Malaysia. Ain Shams Eng J 12:1651–1663. https://doi.org/10.1016/j.asej.2020.09.011
Rivera V (2009) Interrelación Canal Albear, Presa Ejército Rebelde y parámetros de control de la Cuenca Almendares-Vento. Master's thesis. Universidad Tecnológica de La Habana José A. Echeverría. La Habana, Cuba
Sadeghi M, Nguyen P, Hsu K, Sorooshian S (2020) Improving near real-time precipitation estimation using a U-Net convolutional neural network and geographical information. Environ Model Softw 134:104856. https://doi.org/10.1016/j.envsoft.2020.104856
Sakib S, Ahmed N, Kabir A, Ahmed H (2018) An overview of Convolutional Neural Network: Its architecture and applications. https://doi.org/10.20944/preprints201811.0546.v4. Preprint 2018:2018110546
Samad A, Gautam V, Jain P, Sarkar K et al (2020) An approach for rainfall prediction using Long Short Term Memory Neural Network. In: IEEE 5th International Conference on Computing Communication and Automation (ICCCA), IEEE, pp. 190–195, Greater Noida, India, 30–31 October 2020. https://doi.org/10.1109/ICCCA49541.2020.9250809
Shabani S, Samadianfard S, Sattari MT, Mosavi A, Shamshirband S, Kmet T, Várkonyi-Kóczy AR (2020) Modeling pan evaporation using Gaussian process regression K-nearest neighbors random forest and support vector machines; comparative analysis. Atmosphere 11(1):66. https://doi.org/10.3390/atmos11010066
Shahid M, Rahman KU, Haider S, Gabriel HF, Khan AK, Pham QB, Mohammadi B, Linh NTT, Anh DT (2021) Assessing the potential and hydrological usefulness of the CHIRPS precipitation dataset over a complex topography in Pakistan. Hydrol Sci J 66:1664–1684. https://doi.org/10.1080/02626667.2021.1957476
Sivapalan M, Yaeger MA, Harman CJ, Xu X, Troch PA (2011) Functional model of water balance variability at the catchment scale: 1. Evidence of hydrologic similarity and space-time symmetry. Water Resour Res 47. https://doi.org/10.1029/2010wr009568
Solomatine DP, Ostfeld A (2008) Data-driven modelling: some past experiences and new approaches. J Hydroinformatics 10:3–22. https://doi.org/10.2166/hydro.2008.015
Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R (2014) Dropout: a simple way to prevent neural networks from overfitting. J Mach Learn Res 15:1929–1958. https://jmlr.org/papers/volume15/srivastava14a/srivastava14a.pdf
Sumi SM, Zaman MF, Hirose H (2012) A rainfall forecasting method using machine learning models and its application to the Fukuoka city case. Int J Appl Math Comput Sci 22:841–854. https://doi.org/10.2478/v10006-012-0062-1
Sun D, Wu J, Huang H, Wang R, Liang F, Xinhua H (2021) Prediction of short-time rainfall based on deep learning. Math Probl Eng 2021. https://doi.org/10.1155/2021/6664413
Sun Y, Wang X, Tang X (2015) Deeply learned face representations are sparse, selective, and robust. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2892–2900, Boston, MA, USA, 07–12 June 2015. https://doi.org/10.1109/CVPR.2015.7298907
Teschl R, Randeu WL, Teschl F (2007) Improving weather radar estimates of rainfall using feed-forward neural networks. Neural Netw 20:519–527. https://doi.org/10.1016/j.neunet.2007.04.005

Thiemig V, Rojas R, Zambrano-Bigiarini M, De Roo A (2013) Hydrological evaluation of satellite-based rainfall estimates over the Volta and Baro-Akobo Basin. J Hydrol 499:324–338. https://doi.org/10.1016/j.jhydrol.2013.07.012

Towner J, Cloke HL, Zsoter E, Flamig Z, Hoch JM, Bazo J et al (2019) Assessing the performance of global hydrological models for capturing peak river flows in the Amazon basin. Hydrol Earth Syst Sci 23(7):3057–3080. https://doi.org/10.5194/hess-23-3057-2019

Tseng FM, Yu HC, Tzeng GH (2002) Combining neural network model with seasonal time series ARIMA model. Technol Forecast Soc Change 69:71–87. https://doi.org/10.1016/S0040-1625(00)00113-X

Valcarce-Ortega RM (2007) Geofísica de pozos y diagnosis matemática en el estudio de la vulnerabilidad de acuíferos. Universidad Tecnológica de La Habana José A. Echeverría

Vathsala H, Koolagudi SG (2017) Prediction model for peninsular Indian summer monsoon rainfall using data mining and statistical approaches. Comput and Geosci 98:55–63. https://doi.org/10.1016/j.cageo.2016.10.003

Velasco LCP, Serquiña RP, Zamad MSAA, Juanico BF, Lomocso JC (2019) Week-ahead rainfall forecasting using multilayer perceptron neural network. Procedia Comput Sci 161:386–397. https://doi.org/10.1016/j.procs.2019.11.137

Venkatesh R, Balasubramanian C, Kaliappan M (2021) Rainfall prediction using generative adversarial networks with convolution neural network. Soft Comput 25:4725–4738. https://doi.org/10.1007/s00500-020-05480-9

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

Wang L, Kisi O, Hu B, Bilal M, Zounemat-Kermani M, Li H (2017a) Evaporation modelling using different machine learning techniques. Int J Climatol 37:1076–1092. https://doi.org/10.1002/joc.5064

Wang L, Niu Z, Kisi O, Li CA, Yu D (2017b) Pan evaporation modeling using four different heuristic approaches. Comput Electron Agric 140:203–213. https://doi.org/10.1016/j.compag.2017.05.036

Watson GL, Telesca D, Reid CE, Pfister GG, Jerrett M (2019) Machine learning models accurately predict ozone exposure during wildfire events. Environ Pollut 254:112792. https://doi.org/10.1016/j.envpol.2019.06.088

Weathington BL, Cunningham CJ, Pittenger DJ (2012) Appendix B: Statistical Tables in Understanding business research. John Wiley & Sons, pp 435–483. https://doi.org/10.1002/9781118342978.app2

Wongsathan R, Seedadan I (2016) A hybrid ARIMA and neural networks model for PM-10 pollution estimation: The case of Chiang Mai city moat area. Procedia Comput Sci 86:273–276. https://doi.org/10.1016/j.procs.2016.05.057

Wu C, Chau KW, Fan C (2010) Prediction of rainfall time series using modular artificial neural networks coupled with data-preprocessing techniques. J Hydrol 389:146–167. https://doi.org/10.1016/j.jhydrol.2010.05.040

Xiaojian G, Quan Z (2009) A traffic flow forecasting model based on BP neural network. In: 2nd International Conference on Power Electronics and Intelligent Transportation System (PEITS), IEEE. pp. 311–314. Shenzhen, China, 19–20 December 2009. https://doi.org/10.1109/PEITS.2009.5406865

Xu X, Xuezhen Z, Erfu D, Wei S (2014) Research of trend variability of precipitation intensity and their contribution to precipitation in China from 1961 to 2010. Geogr Res 33:1335–1347

Xu W, Jiang Y, Zhang X, Li Y, Zhang R, Fu G (2020) Using Long Short-Term Memory Networks for river flow prediction. Hydrol Res 51:1358–1376. https://doi.org/10.2166/nh.2020.026

Yang Q, Lee CY, Tippett MK (2020) A long short-term memory model for global rapid intensification prediction. Wea Forecast 35(4):1203–1220. https://doi.org/10.1175/WAF-D-19-0199.1

Yuan S, Wang C, Mu B, Zhou F, Duan W (2021) Typhoon intensity forecasting based on LSTM using the rolling forecast method. Algorithms 14(3):83. https://doi.org/10.3390/a14030083

Zhang GP (2003) Time series forecasting using a hybrid ARIMA and neural network model. Neurocomputing 50:159–175. https://doi.org/10.1016/S0925-2312(01)00702-0

Zhang P, Cao W, Li W (2021) Surface and high-altitude combined rainfall forecasting using convolutional neural network. Peer-to-Peer Netw Appl 14:1765–1777. https://doi.org/10.1007/s12083-020-00938-x

Zhao Q, Liu Y, Yao W, Yao Y (2021) Hourly rainfall forecast model using supervised learning algorithm. IEEE Trans Geosci Remote Sens 60:4100509. https://doi.org/10.1109/TGRS.2021.3054582

Authors and Affiliations

Albenis Pérez-Alarcón[1,2] · Daniel Garcia-Cortes[3] · José C. Fernández-Alvarez[1,2] · Yoel Martínez-González[3]

✉ Albenis Pérez-Alarcón
albenisp@instec.cu

[1] Departamento de Meteorología, Instituto Superior de Tecnologías y Ciencias Aplicadas, Universidad de La Habana, 10400 La Habana, Cuba

[2] Centro de Investigación Mariña, Environmental Physics Laboratory (EPhysLab), Universidade de Vigo, Campus As Lagoas s/n, 32004 Ourense, Spain

[3] Departamento de Medio Ambiente, Instituto Superior de Tecnologías y Ciencias Aplicadas, Universidad de La Habana, 10400 La Habana, Cuba