MODWT-ANN hybrid models for daily precipitation estimates with time-delayed entries in Amazon region

Evanice Pinheiro Gomes [ORCID] · Claudio José Cavalcante Blanco [ORCID] ·
Patrícia da Silva Holanda [ORCID] · José Francisco de Oliveira Júnior [ORCID]

Abstract Hydrological analyses based on precipitation records in the Amazon are essential due to their importance in climate regulation and regional and global atmospheric circulation. However, there are limitations related to data series with short periods and many gaps and failures at the daily scale. Thus, a hybrid model was developed based on an artificial neural network (ANN) and adaptive neuro-fuzzy inference system (ANFIS) coupled with the maximum overlap discrete wavelet (MODWT) method to obtain precipitation estimates. Six rainfall gauge stations located in different biomes within the studied region were adopted, and satellite data (CMORPH) were used. The interval of data that was have used is 1998–2016. The precipitation data were evaluated by seasonal (wet and dry) periods. The results obtained demonstrated the good capacity of the MODWT-ANFIS model to simulate the daily precipitation. In this case, data entries lagged by 4 days and 5 days performed better, with *Nash* values close to 1.0 and mean square errors (*MSE*) below 0.1.

Keywords Artificial intelligence · Time-delayed inputs · CMORPH · Amazon

E. P. Gomes
Graduate Program of Civil Engineering, Federal University of Pará-PPGEC/ITEC, UFPA, Av. Augusto Corrêa, 01, Belém 66075–110, Brazil
e-mail: gomesevanice@ufpa.br

C. J. C. Blanco (✉)
School of Environmental and Sanitary Engineering, Universidade Federal do Pará – FAESA/ITEC/UFPA, Av. Augusto Corrêa, 01, Belém 66075–110, Brazil
e-mail: blanco@ufpa.br

P. da Silva Holanda
Graduate Program in Natural Resource Engineering in the Amazon of the Federal University of Pará–PRODERNA/ITEC/UFPA, Av. Augusto Corrêa 01, Belém 66075–110, Brazil
e-mail: pholanda@ufpa.br

J. F. de Oliveira Júnior
Applied Meteorology and Environment Laboratory (LAMMA), Institute of Atmospheric Sciences (ICAT), Federal University of Alagoas – LAMMA/ICAT/UFAL, Av. Lourival Melo Mota, S/N, Maceió, Alagoas 57072-970, Brazil

Introduction

Precipitation estimates are essential for the management of water resources, as well as for creating sustainability strategies for these resources for extremely varied applications, such as agriculture, industry, water supply, energy production (hydroelectricity) and waterway transport, especially in extreme weather conditions. However, according to Michot et al. (2019), practical and accurate forecasts may encounter barriers related to the quality of the data (gaps and failures), size of the historical series and availability of the number of rainfall stations. Thus, the use of effective methods for estimating precipitation

is essential. Artificial intelligence (AI) methods are potentially useful approaches to simulate precipitation (Fahimi et al., 2017; Nourani et al., 2014). According to Sulaiman et al. (2018), this utility is due to the remarkable flexibility of AI methods in modelling highly nonlinear systems and stochastic patterns, and these methods do not require prior knowledge of the behaviour of measurement processes.

According to Shoaib et al. (2016), AI methods, for example, artificial neural networks, (ANNs) are able to establish a relationship between historical inputs (precipitation, streamflow, water levels, etc.) and the desired outputs. Such work is carried out through a nonlinear function composed of several factors that are adjusted to the observed data, allowing its prediction, as adopted by Jiménez and Collischonn (2015), Santos et al. (2016), Nourani et al. (2017), Shoaib et al. (2018), Honorato et al. (2018) and Mendonça et al. (2021). ANNs are widely used methods in predicting hydrological variables; however, a single ANN model may not be able to deal with the nonstationary behaviour of time series if the input is not pre-processed (Cannas et al., 2006; Hu et al., 2018; Islam & Sivakumar, 2002). In this sense, wavelet transformation (WT) is a pre-processing methodology capable of filtering and correcting the information contained in the time series of input data (Zeri et al., 2018). According to Nourani et al. (2014), this correction in the inputs considerably favours the efficiency of ANN models in predicting hydrological variables. He et al. (2015) combined feed forward backpropagation WT and ANN to forecast monthly rainfall precipitation in the Australian territory, concluding that the combined model performed better in forecasting when compared to other models. Partal et al. (2015) obtained good results by combining with three types of ANNs (feedback propagation, radial basis function and generalized regression neural network) for daily precipitation prediction.

It is possible to develop these ANN-based prediction models and combine pre-processing tools using a number of variables, such as temperature, radiation and humidity, as inputs. However, few stations are equipped with resources to measure these variables, especially in developing countries, due to economic and technical reasons (Altunkaynak & Nigussie, 2015). Therefore, it is advisable to develop a model that can simulate daily precipitation based on previous records of its historical series. Furthermore, the

amount of minimal input data, which function as a memory in ANN models and act on network learning, is still a matter of concern and needs to be investigated, as demonstrated by Shoaib et al. (2016). Hu et al. (2018) inserted an ANN into LSTM models to simulate the rainfall-runoff process based on flood events from 1971 to 2013 in the Chinese Fen River basin, obtaining satisfactory results with the use of LSTM. Salman et al. (2018) built an LSTM model with an ANN to predict meteorological variables at the Hang Nadim airport in Indonesia, demonstrating that several input layers with different time delays improve the prediction of observed variables. Hammad et al. (2021) developed a new wavelet-coupled multiple order time delay (WMTLNN) ANN model for rainfall prediction in Indus basins, Pakistan. They found that the different inputs with time delays and wavelet pre-processing improved the precipitation forecast in the evaluated basins.

Thus, daily precipitation was estimated through a hybrid model based on a new concept of introduction of several layers of time delay and pre-processed by maximum overlap discrete wavelet (MODWT) via neural networks adaptive neuro-fuzzy inference system (ANFIS). The ANFIS network has been combined with other techniques and has stood out among neural networks due to its good performance in predicting hydrological variables, especially when compared with other models (Choubin et al., 2016; Ahmadlou et al., 2019; Pham et al. 2020; Ebrahimi-Khusfi et al., 2021). The MODWT-ANFIS model was applied to the Amazon basin, which depends on precipitation to sustain its economic activities, in addition to influencing regional and global atmospheric circulation. Precipitation data observed by the National Water Agency (ANA) and Satellite of the Morphing Climate Prediction Center (CMORPH) were adopted. In this case, the models can be applied even in the absence of monitoring by specific precipitation stations.

Material and methods

Study area and database

The Amazon area is approximately 5,015,067.75 km^2, corresponding to approximately 58.9% of the Brazilian territory (IBGE, 2010). The region has an extensive

and dense hydrographic network formed by the largest river in the world, the Amazon, with a length of 6,400 km, of which approximately 3,220 km is within Brazil. Including discharges from its various tributaries, the Amazon River is responsible for 60% of Brazil's water availability and approximately 20% of the flow of all freshwater in the world (Davidson et al., 2012). According to data from Mapbiomas (2016), the Amazon has three characteristic biomes: (i) the Amazon biome (AB), which is the most representative, occupying 83.86% of the region; (ii) the Cerrado biome (CB), located to the east (E) and southeast (SE), corresponding to 14.32%; and (iii) the Pantanal biome (PB), located to the southwest (SW), representing only 1.82% of the total area (Fig. 1). In addition, in these biomes, there are transition areas: the Amazon-Cerrado (EAC) Ecotone is the largest in length, approximately 6,240 km, extending from SE to SW of the region, and the Amazon-Pantanal (EAP) and Amazon Ecotones Pantanal-Cerrado (EAPC) (Fig. 1).

In the context of regional circulation, the forest plays an important role as a source of moisture generation for other regions of Brazil (midwest, southeast and south) and for the South American (SA) continent (Ciemer et al., 2018; Silveira et al., 2017). The Amazon deforested area is 15.19% of the total area, concentrated on the southern and eastern edges of the region, known as the "arc of deforestation" (Fig. 1). This deforestation process is mainly caused by the replacement of forest cover by livestock, agricultural and agro-industrial activities (Lima et al., 2019; Vale et al., 2019).

The temporal series of six rainfall stations (Table 1 and Fig. 1) monitored by the ANA (available at http://www.snirh.gov.br) were used. Daily precipitation data for the CMORPH product were obtained for each location of the rainfall stations. The choice of stations prioritized series with minimal gaps (average of 0.1% of the total observed data), and the period observed was 19 years (1998–2016). Precipitation
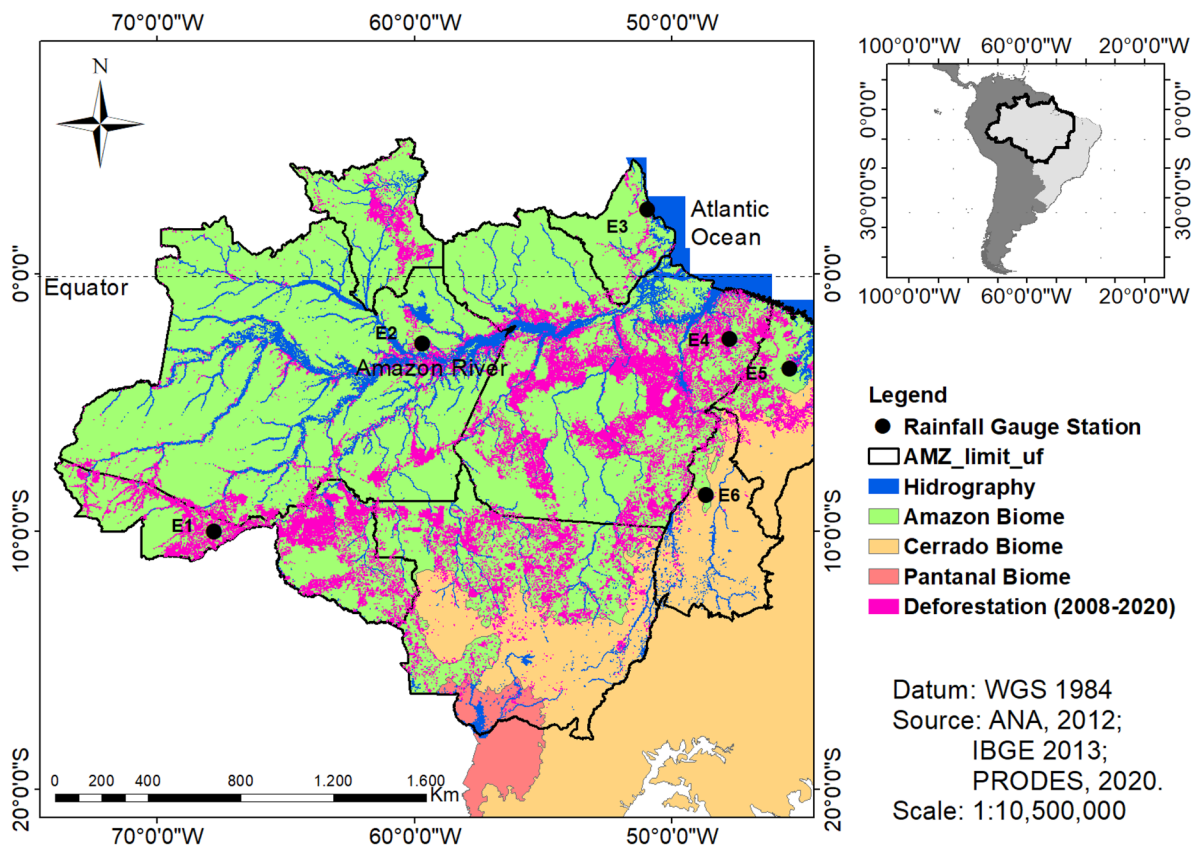

Fig. 1 Amazon and rainfall gauge station locations

Table 1 Data from ANA rainfall stations and average daily rainfall

ID	Rainfall gauge station	Latitude	Longitude	Gaps (%)	ADP ANA	ADP CMORPH	Period
E1	Rio Branco	−9.98	−67.80	0.1	5.4	4.2	1998–2016
E2	Rio Preto da Eva	−2.70	−59.70	0.0	6.1	5.8	
E3	Calçoene	2.50	−50.95	0.0	12.4	8.2	
E4	Badajós	−2.51	−47.77	0.0	6.3	5.5	
E5	Pindaré Mirim	−3.66	−45.44	0.0	5.2	4.0	
E6	Itaporã do Tocantins	−8.57	−48.69	0.0	5.1	4.3	

ID, identification of the rainfall gauge stations; *ADP*, average daily precipitation

from stations stored by ANA is punctual and recorded every 24 h. The information produced by CMORPH has a spatial resolution of 8 km (at the Equator) and is recorded every 30 min. These differences motivated the use of two databases, in addition to the possibility of replacing data, in the absence of punctual monitoring, which is common in some places in the Amazon.

Maximum overlap discrete wavelet transform

For Daubechies (1992), the central idea of WT is the decomposition of the signal at different time scales as a set of basic functions (mother wavelet), revealing information from the original data, such as trends, disintegration points and discontinuities, which the raw signal does not expose (Holdefer & Severo, 2015; Zeri et al., 2018). The WT is divided into two types: continuous wavelet transform (CWT) and discrete wavelet transform (DWT) (Addison et al., 2001; Daubechies, 1992); however, as hydrometeorological data are usually recorded at discrete time intervals, the DWT is preferentially adopted in the hydrological decomposition of time series (Mehr et al., 2014; Ramana et al., 2013). Among the existing TWs, the maximum overlap discrete wavelet transform (MODWT) has stood out in the use of time series decompositions. This is due to its potential to consider boundary conditions (BC) that involve data decomposition, thus avoiding errors that may be introduced throughout the development of the proposed forecasting model. Bašta (2014), Quilty et al. (2016) and Du et al. (2017) demonstrated how BCs influence the decomposition of time series and how they can produce incorrect predictions if not properly treated.

The MODWT definition is derived from the DWT definition, where $(h_{j,k})$ is the DWT filter and $(g_{j,k})$ is the scale filter, with $k = 1\ldots$, representing the filter length (L), with j levels of decomposition. The MODWT wavelet filter $(\hat{h}_{j,k})$ and the MODWT scale filter $(\hat{g}_{j,k})$ are defined as $\tilde{h}_{j,k} = {}^{h_{j,k}}\!/{}_{2^{j/2}}$ and $\tilde{g}_{j,k} = {}^{g_{j,k}}\!/{}_{2^{j/2}}$. Thus, the j-level MODWT wavelet coefficients are defined as the time series convolution (X_t), and the MODWT filters are obtained by Eqs. (1) and (2).

$$\tilde{W}_{j,t} = \sum_{k=0}^{k_j-1} \tilde{h}_{j,k} X_{t-k\,modN} \tag{1}$$

$$\tilde{V}_{j,t} = \sum_{k=0}^{k_j-1} \tilde{g}_{j,k} X_{t-k\,modN} \tag{2}$$

where $\tilde{W}_{j,t}$ is the wavelet coefficient; $\tilde{V}_{j,t}$ is the scale coefficient; $modN$ is the operation module when treating the historical series as periodic, with periods equal to N; and K_j can be obtained by Eq. (3).

$$K_j = \left(2^j - 1\right)(K-1) + 1 \tag{3}$$

The value of K_j represents the number of wavelet coefficients and scales affected by BC for the decomposition level J and the length level of the wavelet filter K. Thus, using this equation, it is possible to obtain wavelet and scale coefficients that have been "corrected by limits", that is, values that avoid the introduction of additionally uncertainty to the wavelets and scale coefficients due to the problem of "future data" (Bašta, 2014).

MODWT uses a high pass filter $(\tilde{h})$ to calculate its wavelet coefficients and applies an iterative

construction of the time series (X_t), which can be reconstructed using Eq. (4).

$$X_t = \tilde{W}_{j,t} + \tilde{V}_{j,t} \tag{4}$$

In practice, MODWT decomposition is performed on a series of data, for which the type of filter (wavelet), the level of decomposition and the limit, can be periodic or reflective, are selected. If periodic, the resulting wavelet and scale coefficients are calculated without duplicating the original series, treating (X_t) as if it were circular. If it is reflection, a new series is reflected twice the length of the original series. In the present study, the periodic limit was adopted, and three types of wavelet families, Daubechies (db4) of levels 6 and 8 (db4-j6 and db4-j8), less asymmetrical (la14) of levels 4 and 6 (la14-j4 and la14-j6) and coiflet (c6) of levels 4 and 6 (c6-j4 and c6-j6), were selected based on the most common hydrological data series (Maheswaran & Khosa, 2012; Santos et al., 2019) and by carrying out diversified decompositions.

Artificial neural network

ANNs are computational models that imitate the functioning of the human brain, with the aim of analysing a given system and reproducing it. The learning of an ANN occurs through an iterative process applied to synaptic weights (wkn) and bias (bk), called training. According to Haykin (2007), the training of an ANN is performed by an algorithm, which adjusts a matrix of synaptic weights. Thus, the output vector must match a desired target value for each input vector. This process is cyclical for the training sample set until a previously stipulated stopping criterion is reached. After training, it is expected that the ANN will be able to generalize information, obtaining coherent outputs with input vectors not used in the training set. It is also expected that the minimum error found in training will be similar to the error in simulation in an entirely different set.

The main architectures of artificial neural networks can be divided into single layer feedforward networks, multilayer feedforward networks, recurrent networks and reticulated networks. The difference between them is related to the arrangement of their neurons, their way of interconnection and the constitution of their layers, as mentioned above. In this study, the ANFIS network was used.

ANFIS network

ANFIS is a neural network that combines the fuzzy inference system (SIF) with an ANN. ANFIS is considered a fuzzy inference system organized in the form of an adaptive network capable of mapping input and output data based on the knowledge of an expert. The adaptive network is a multilayer network with a feedforward architecture arranged by nodes interconnected by unidirectional connections and supervised learning (Jang, 1993). A neuro-fuzzy network is usually made up of three layers. The first layer (fuzzification) represents the fuzzy rules, that is, the terms that precede the rule. The second layer (intermediate) represents the fuzzy rules, and the third layer (defuzzification) represents the output variables, that is, the consequent term of the rule. However, there can be several types of FIS in an ANFIS network, which can vary depending on the reasoning and rules applied.

FIS (Takagi & Sugeno, 1985), adopted in this study, represents a system that associates a set of linguistic rules in the antecedent ("if" part) with fuzzy propositions and in the consequent ("then" part) presented by expressions of type $y = f(x)$ from the linguistic variables of the antecedent. With this system and from a dataset for training (input and output pairs), it is possible to make predictions of a given variable using an ANFIS architecture (Fig. 2). This architecture is composed of five layers, which each have specific purposes (Jang, 1993).

In the first layer, the degree of membership of the input entries x and y is calculated, according to the type of membership function (MF) chosen in these nodes ($A1$, $A2$, $B1$ and $B2$). In the second layer, neurons perform the t-norm operation as the algebraic product (neuron $\prod$) (Eq. 5), considering the MF (μ) and the linguistic terms (Ai, Bi).

$$w_i = \mu_{Ai}(x)\mu_{Bi}(y), i = 1, 2 \ldots \tag{5}$$

In the third layer, the membership functions are normalized (Eq. 6) through the weights (w) of the N neurons.

$$\overline{w}_i = \frac{w_i}{w_1 + w_2}, i = 1, 2 \ldots \tag{6}$$

In the fourth layer, the outputs of neurons are calculated by the product between the normalized firing levels and the value of the consequent rules. Its
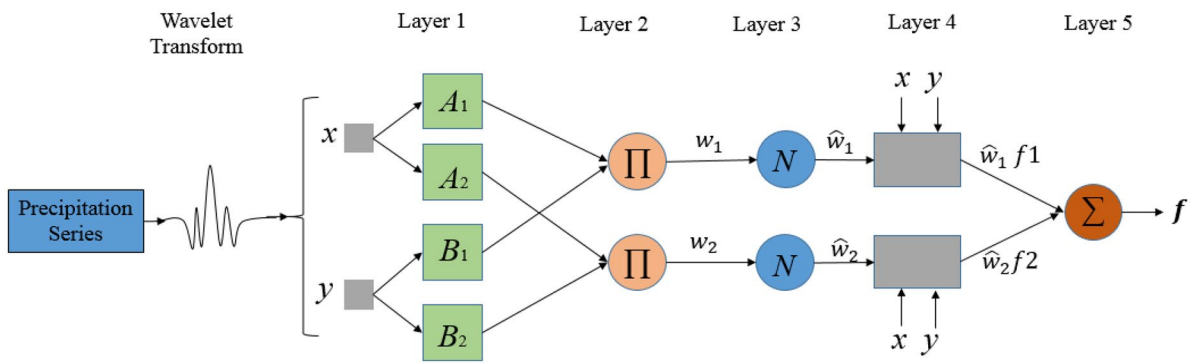
Fig. 2 MODWT and ANN hybrid model with ANFIS network architecture

parameters correspond to the coefficients of the affine expressions and the neuron activation function, which form the fourth layer (Eq. 7), where p_i, q_i and r_i are the parameters associated with the consequents of the rules.

$$z_{4,1} = \overline{w}_i f_i = [\overline{w}_i (p_i + q_i y + r_i)] \tag{7}$$

In the fifth layer, the system output is calculated, which together with the nodes of the third and fourth layers promote the defuzzification or sum total of all input signals (Eq. 8).

$$f = \frac{\sum_i w_i f_i}{\sum_i w_i} \tag{8}$$

For the application of ANFIS in precipitation forecasting, two NMFs (number of membership functions) were adopted as initial parameters for each input variable, and the membership function (MF) type was chosen for the best performance of the network, ranging from triangular, trapezoidal, Gaussian and sinusoidal.

Time-lagged neural network

In problems involving the prediction of time series, neural networks are used as a good artifice, especially in the input layer, where the incorporation of a memory at the input of the network allows the strengthening of the learning of the behaviour of time series, which can be intuitively attached to the other layers of the network, improving the results. Thus, the combination of entries based on antecedent times is suggested in this work. Four combinations were adopted, considering the precipitation of 2, 3, 4 and 5 days before (t-2, t-3, t-4 and t-5) to forecast the current day. In forecasting hydrological variables, the optimal time interval of this delay is not well defined. However, Shoaib et al. (2018), Kim et al. (2020) and Hammad et al. (2021) consider that up to five delays is an acceptable number and this value is adopted in this study. Furthermore, incorporating other climatic variables (air temperature, wind, solar radiation, etc.) in precipitation forecasts can generate errors due to the uncertainty of the real influences that such variables can exert on precipitation.

Seasonality assessment

The daily precipitation data from the rainfall gauge stations were organized in two ways: (1) rainy period, which is formed by 3444 daily precipitations in the months of November–April of 1998–2016, divided into 2584 values for calibration (01/1998–02/2012) and 860 values for validation (02/2012–12/2016); (2) dry period, formed by 3496 precipitations from May–October 1998–2016, divided into 2624 values for calibration (05/1998–06/2012) and 872 and 872 for validation (06/2012–10/2016). This division aims to assess the influence of seasonality on the model's response. In network processing, data were standardized (Eq. 9) and divided for calibration (75%) and validation (25%).

$$P_{pad} = \frac{P_i - P_{min}}{P_{max} - P_{min}} \tag{9}$$

where P_{pad} is the standardized precipitation, P_i is the precipitation to be standardized and P_{min} and P_{max} are the smallest and largest values, respectively, observed in the precipitation series. Standardization implies scaling the samples to the dynamic range of activation functions of hidden layers, typically represented by the logistic function or hyperbolic tangent, to avoid saturation of neurons, as adopted by Nourani et al. (2017).

Performance criteria

Model performance was assessed using statistical parameters, which are used to quantify the agreement between observed and estimated data. In this study, we used two classic criteria, the mean square error (*MSE*, mm) and the Nash–Sutcliffe coefficient (*Nash*), represented by Eqs. (10) and (11), respectively.

$$MSE = \frac{1}{n} \sum_{i=1}^{n} (\overline{X} - Y_{obs})^2 \tag{10}$$

$$Nash = 1 - \frac{\sum (Y_{obs} - Y_{est})^2}{\sum (Y_{obs} - \overline{X})^2} \tag{11}$$

where n is the number of samples, Y_{obs} is the observed precipitation, Y_{est} is the estimated precipitation and $\overline{X}$ is the average of the observed precipitation. The best performing models are those with low *MSE* and *Nash* values close to 1 (Chai & Draxler, 2014; Nash & Sutcliffe, 1970).

The methodology adopted in this study consists of the following steps (Fig. 3):

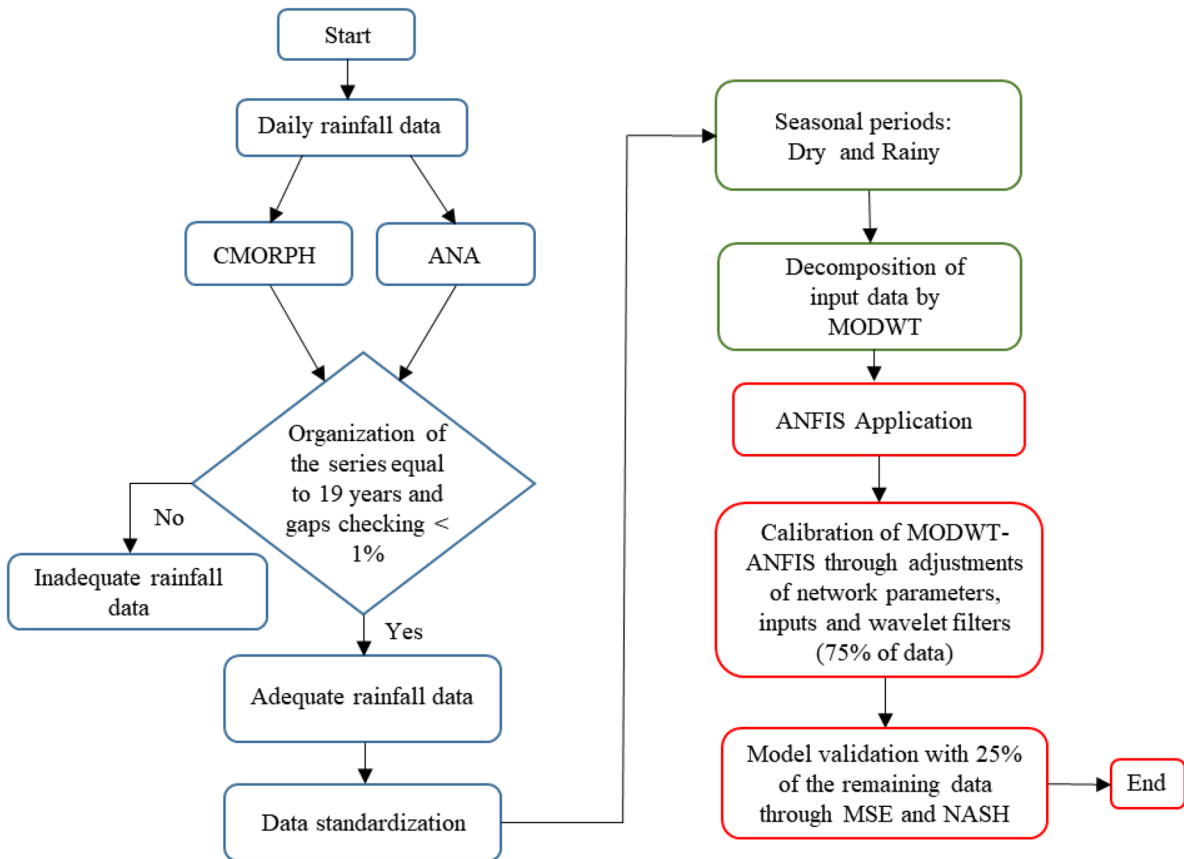– The collection, organization and standardization of precipitation data;

Fig. 3 Methodology flowchart

– The decomposition of the historical series by MODWT with wavelet filters;
– A model calibration performed through MODWT-IA training and adjustments of network parameters, input type and wavelet filters (75% of the historical series); and
– The validation of the model through the adoption of the optimal parameters obtained in the calibration (25% of the historical series) with performance criteria.

Results and discussion

Using the MODWT, the maximum level of decomposition was found to be eight ($J_{máx} = 8$), and the lengths (L) of the wavelet filters were 4 for db4, 14 for la14 and 6 for c6. Thus, using a maximum level of decomposition equal to 8, a K equal to 4 and Eq. (3), K_j is equal to 766 coefficients affected by the limit of

j (this practice was also adopted, for $j = 4$ and 6, and for $L = 6$ and 14). Therefore, the first 766 records of input data from the stations are removed after decomposition with wavelet db4-j8. Then, the training of the ANFIS network was carried out at each station through the method of successive approximations in the dry and rainy periods with data from ANA and CMORPH. Tests were also carried out to assess the optimal parameters. After the simulations at each station, with different filters and levels adopted, the best parameters were defined in relation to the lagged inputs regarding the number of membership functions (NMF), type of membership function (MF) and number of epochs. The MFN of 2 for each entry and the generalized bell MF (gbellmf) were the ones with the lowest errors for training, testing, validation and the FIS of the network (0.01570, 0.01656, 0.01601, 0.01542), with entries delayed by 4 days (Table 2). The selected output function was a constant, and the training method was a hybrid.

Table 2 MODWT-ANFIS model calibration parameters

NMF	Filter	MF	Epochs	Training error	Test error	Check error	FIS error
2 delays							
2 2	db4-j6	trimp	30	0.01886	0.01962	0.02031	0.0189
3 3	db4-j8	gbellmf	30	0.01879	0.02161	0.02053	0.0189
4 4	db4-j8	gbellmf	30	0.01840	0.01957	0.01986	0.0181
5 5	c6-j6	gbellmf	30	0.01868	0.03141	0.02137	0.0187
6 6	db4-j8	gaussmf	30	0.01846	0.02705	0.02483	0.0184
7 7	db4-j6	trapmf	30	0.01848	0.02402	0.02112	0.0185
8 8	la14-j6	trapmf	30	0.01861	0.024151	0.020959	0.0186
3 delays							
2 2 2	db4-j8	gbellmf	30	0.01881	0.08131	0.08131	0.01881
3 3 3	db4-j8	trimp	30	0.01828	0.02051	0.02449	0.01828
4 4 4	db4-j8	gbellmf	30	0.01705	0.01961	0.02034	0.01720
5 5 5	c6-j6	gbellmf	30	0.01763	0.03991	0.11176	0.01763
6 6 6	c6-j4	gaussmf	30	0.01741	0.05599	0.19521	0.01741
7 7 7	db4-j8	gbellmf	30	0.01790	0.15160	0.25428	0.01790
4 delays							
2 2 2 2	**db4-j8**	**gbellmf**	**30**	**0.01570**	**0.01656**	**0.01601**	**0.01542**
3 3 3 3	db4-j6	gauss2mf	30	0.01780	0.02365	0.02612	0.01780
4 4 4 4	la14-j6	gaussmf	30	0.01733	0.43885	0.12759	0.01743
5 5 5 5	db4-j8	gaussmf	30	0.01648	0.12007	0.19133	0.01648
6 6 6 6	c6-j6	gaussmf	30	0.01847	0.18319	0.17287	0.01847
5 delays							
2 2 2 2 2	la14-j4	gaussmf	30	0.028261	0.098502	0.20054	0.028264
3 3 3 3 3	db4-j6	trimp	30	0.027264	0.036274	0.12092	0.027193
4 4 4 4 4	db4-j8	gaussmf	30	0.056044	0.22155	0.081788	0.056044

Through simulations with the ANFIS network, it was found that the increase in the number of membership functions and input lag resulted in an increase in computational time and effort without resulting in gain for the network, as the errors (*MSE*) did not have undergone so much change. Therefore, in this case, increasing the number of inlets and MFN is not advisable for this type of precipitation forecast. This fact may be related to the great effort that the ANFIS network performs with each MF and each input variable, requiring greater computational effort. In this way, the entries with 5 days of delay were made only with 4 NMF to expedite the training and make the training more efficient. Regarding the number of epochs, values from 2 to 100 epochs were adopted. However, the value of 30 epochs presented the lowest *MSE* because from this value, the errors were without a significant reduction. For the wavelet filter, db4-j8 was the most adjusted for the series with four delays in the ANFIS networks (Table 2). Table 3 presents the optimized parameters of the ANFIS network.

The Daubechies (db4) wavelet was able to decompose the seasonality element of the time series more efficiently, and its results for levels (*j*) 6 and 8 and length (*L*) 4 presented small errors and *Nash* values close to the ideal. According to Maheswaran and Khosa (2012), the good db4 performance is due to the broader support in seasonal temporal series and the ability to smooth the signal and good location of time and frequency. This process is necessary for precipitation series that present temporal intercurrence. The less asymmetric wavelet (la14) and the coiflet (c6) combined with the ANN also presented good results with small errors and high *Nash*. However, its performance against db4 was not extensively different.

Table 3 ANFIS parameters after training

ANFIS parameter	Values
Membership functions (MF)	gbellmf
Output function (defuzzification)	constante
Number of nodes	55
Number of fuzzy rules	16
Number of linear parameters	16
Number of nonlinear parameters	24
Total parameters	40
Number of training pairs	1308–1736
Number of validation pairs	656–870

This shows that increasing the length (*L*) of the filter (6 and 14), for this case, did not bring significant improvements and that the db4 filter with a length (*L*) of 4 is sufficient for good signal decomposition.

The best filter, according to Zhang et al. (2015), should be the one with the most similar decomposition to the characteristics of the studied series. However, when choosing a filter, other parameters are also associated with the filter. Thus, according to the tests performed, the factors that most influenced the simulations were the level of decomposition and the length of the wavelet. The fit of the best model with level 8 and length 4 has a smoother adjustment and considers the boundary conditions. It provided a moderate and permissible fit for the decomposition of the precipitation data. The longer length (6 and 14) did not show higher quality and could remove a much larger number of wavelet coefficients adjusted by BC, compromising the amount of input data in the model simulation with ANN.

To avoid errors and circumvent BC, it is necessary to choose an adequate wavelet and sufficient input data for training and forecasting (Du et al., 2017; Quilty et al., 2016; Ramírez-Hernández et al., 2016). In the selection of the precipitation series, this question was adopted by testing three wavelet filters and three levels of decomposition, removing the values that interfere in the coefficients affected by the limit of *j* and by the adjusted division of the number of data used in the calibration and validation. Thus, it was possible to filter the data series, leave them free of uncertainties related to BC and even adjust adequate numbers of input data for the training and validation of neural networks.

In the validation of the MODWT-ANFIS model, tests were performed with 25% of the temporal series, corresponding to the seasonal period (rainy and dry) from 2012 to 2016. In this case, the model presented a *Nash* value close to 1 and an *MSE* value less than 0.1 (Fig. 4).

The effectiveness of the ANFIS model in daily precipitation simulations can be explained by the ability to incorporate fuzzy rules to assist in simulations, being sensitive to learning datasets and able to learn much more during the training period and improve simulations in the testing phase (Seera et al., 2012; Roy & Singh, 2020). Choubin et al. (2016), for example, found that the ANFIS model combined with other techniques can be sufficiently
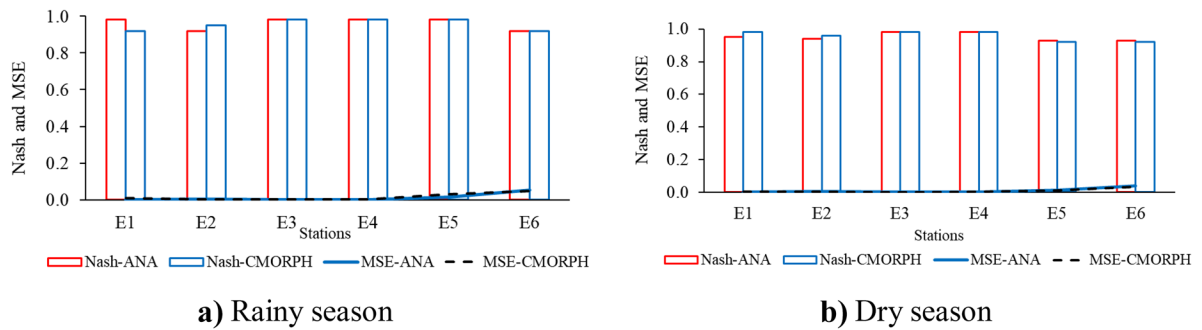
Fig. 4 Performance of the MODWT-ANFIS model in the seasons rainy (**a**) and dry (**b**)

satisfactory in simulating precipitation. The small numbers of modelled data entries with small time delays proved to be effective, as demonstrated by the resulting *Nash* values close to 1.0. This small number of entries can be considered a great advantage of the model, as it allows overcoming the problem of drier periods (Costa et al., 2015; Suhaila et al., 2011), which require more information from previous days to simulate future days. Furthermore, according to Nerantzaki and Papalexiou (2019), the estimation of precipitation events is still a challenge in the literature and requires specific methods for its modelling. The model was also able to satisfactorily simulate the precipitation of stations E1, E2, E3 and E4 (Fig. 1), with high precipitation located in the Amazon biome, and the precipitation of stations E5 and E6 (Fig. 1), with low precipitation located in the transition region and in the Cerrado biome. In other words, the model had no problems reproducing the precipitation resulting from the Amazon's climate variability. However, other models have shown problems with this reproduction (Detzel & Mine, 2011; Liu et al., 2011; Ng et al., 2017; Wilks, 1999).

Conclusion

The MODWT-ANFIS model was calibrated, trained and validated, and it satisfactorily simulated the daily precipitation in the Amazon, considering seasonality and the region's biomes. The small number of data entries input into the model with small time delays proved to be effective and was considered a great advantage of the model. This method can overcome the problems associated with dry periods,

which require more information from previous days to simulate future days. The pre-processing of data performed by MODWT was essential to remove noise from the original time series and correct the boundary conditions that could harm the model's simulations. This stage in the development of the models, together with the time-lagged inputs, configures one of the advantages of hybrid models, such as the analysed model.

The results generated may help future work to better understand the daily precipitation modelling and its behaviour in the Amazon region, which has been suffering from fires and deforestation, impacting the region's hydrological cycle and affecting various activities, such as human supply, sanitation, agribusiness, water supply, hydroelectric production and waterway transport. This hydrological imbalance affects other regions of the country (midwest, southeast and south), which depend on evapotranspiration (ET) from the Amazon to produce rain, which is also important for the water uses mentioned above. Finally, the global climate is sensitive to changes in the Amazon hydrological cycle.

Data availability Data will be made available on reasonable request.

Declarations

Conflict of interest The authors declare no competing interests.

References

Addison, P. S., Murray, K. B., & Watson, J. N. (2001). Wavelet transform analysis of open channel wake flows. *Journal of Engineering Mechanics, 127*(1), 58–70. https://doi.org/10.1061/(ASCE)0733-9399(2001)127:1(58)

Altunkaynak, A., & Nigussie, T. A. (2015). Prediction of daily rainfall by a hybrid wavelet-season-neuro technique. *Journal of Hydrology, 529*, 287–301. https://doi.org/10.1016/j.jhydrol.2015.07.046

Ahmadlou, M., Karimi, M., Alizadeh, S., Shirzadi, A., Parvinnejhad, D., Shahabi, H., & Panahi, M. (2019). Flood susceptibility assessment using integration of adaptive network-based fuzzy inference system (ANFIS) and biogeography-based optimization (BBO) and BAT algorithms (BA). *Geocarto International, 34*(11), 1252–1272. https://doi.org/10.1080/10106049.2018.1474276

Bašta, M. (2014). Additive decomposition and boundary conditions in wavelet-based forecasting approaches. *Acta Oeconomica Pragensia, 22*(12), 48–70. https://doi.org/10.18267/j.aop.431

Cannas, B., Fanni, A., See, L., & Sias, G. (2006). Data preprocessing for river flow forecasting using neural networks: Wavelet transforms and data partitioning. *Physics and Chemistry of the Earth Parts a/b/c, 31*(18), 1164–1171. https://doi.org/10.1016/j.pce.2006.03.020

Chai, T., & Draxler, R. R. (2014). Root mean square error (RMSE) or mean absolute error (MAE)? – Arguments against avoiding RMSE in the literature. *Geoscientific Model Development, 7*(3), 1247–1250. https://doi.org/10.5194/gmd-7-1247-2014

Choubin, B., Khalighi-Sigaroodi, S., Malekian. A., Kişi, Ö. (2016). Multiple linear regression, multi-layer perceptron network and adaptive neuro-fuzzy inference system for forecasting precipitation based on large-scale climate signals. *Hydrological Sciences Journal 61*(6), 1001–1009. https://doi.org/10.1080/02626667.2014.966721

Ciemer, C., Boers, N., Barbosa, H. M., Kurths, J., & Rammig, A. (2018). Temporal evolution of the spatial covariability of rainfall in South America. *Climate Dynamics, 51*(1–2), 371–382. https://doi.org/10.1007/s00382-017-3929-x

Costa, V., Fernandes, W., & Naghettini, M. (2015). A Bayesian model for stochastic generation of daily precipitation using an upper-bounded distribution function. *Stochastic Environmental Research and Risk Assessment, 29*(2), 563–576. https://doi.org/10.1007/s00477-014-0880-9

Daubechies, I. (1992). Ten lectures on wavelet. *Society for Industrial and Applied Mathematics, Philadelphia.* https://doi.org/10.1137/1.9781611970104

Davidson, E. A., de Araújo, A. C., Artaxo, P., Balch, J. K., Brown, I. F., Bustamante, M. M. C., & Wofsy, S. C. (2012).

The Amazon basin in transition. *Nature, 481*, 321–328. https://doi.org/10.1038/nature10717

Detzel, D. H. M., & Mine, M. R. M. (2011). Generation of daily synthetic precipitation series: Analyses and application in La Plata river basin. *The Open Hydrology Journal, 5*, 69–77. https://doi.org/10.2174/1874378101105010069

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. *Journal of Hydrology., 552*, 44–51. https://doi.org/10.1016/j.jhydrol.2017.06.019

Ebrahimi-Khusfi, Z., Taghizadeh-Mehrjardi, R., & Nafarzadegan, A. R. (2021). Accuracy, uncertainty, and interpretability assessments of ANFIS models to predict dust concentration in semi-arid regions. *Environmental Science and Pollution Research, 28*(6), 6796–6810. https://doi.org/10.1007/s11356-020-10957-z

Fahimi, F., Yaseen, Z. M., & El-shafie, A. (2017). Application of soft computing based hybrid models in hydrological variables modeling: A comprehensive review. *Theoretical and Applied Climatology, 128*, 875–903. https://doi.org/10.1007/s00704-016-1735-8

Hammad, M., Shoaib, M., Salahudin, H., Baig, M. A. I., Khan, M. M., Ullah, M. K. (2021). Rainfall forecasting in upper Indus basin using various artificial intelligence techniques. *Stochastic Environmental Research and Risk Assessment*, 1-23.https://doi.org/10.1007/s00477-021-02013-0

Haykin, S. (2007). Redes neurais: Princípios e prática. Bookman publishing company.

He, X., Guan, H., & Qin, J. (2015). A hybrid wavelet neural network model with mutual information and particle swarm optimization for forecasting monthly rainfall. *Journal of Hydrology, 527*, 88–100. https://doi.org/10.1016/j.jhydrol.2015.04.047

Holdefer, A. E., & Severo, D. L. (2015). Análise por ondaletas sobre níveis de rios submetidos à influência de maré. *Revista Brasileira De Recursos Hídricos, 20*(1), 192–201. https://doi.org/10.21168/rbrh.v20n1.p192-201

Honorato, A. G. S. M., Silva, G. B. L., & Guimarães Santos, C. A. (2018). Monthly streamflow forecasting using neuro-wavelet techniques and input analysis. *Hydrological Sciences Journal, 63*, 2060–2075. https://doi.org/10.1080/02626667.2018.1552788

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*(11), 1543. https://doi.org/10.3390/w10111543

IBGE. (2010). Instituto Brasileiro de Geografia e Estatística. http://www.ibge.gov.br/home/. Acessed 20 Feb 2021

Islam, M. N., & Sivakumar, B. (2002). Characterization and prediction of runoff dynamics: A nonlinear dynamical view. *Advancer in Water Resources, 25*(2), 179–190. https://doi.org/10.1016/S0309-1708(01)00053-7

Jang, J. S. (1993). ANFIS: Adaptive-network-based fuzzy inference system. *IEEE Transactions on Systems, Man, and Cybernetics, 23*(3), 665–685. https://doi.org/10.1109/21.256541

Jiménez, K. Q., & Collischonn, W. (2015). Método de combinação de dados de precipitação estimados por satélite e medidos em pluviômetros para a modelagem hidrológica. *Revista Brasileira De Recursos Hídricos, 20*(1), 202–217.

Kim, S., Alizamir, M., Kim, N. W., & Kisi, O. (2020). Bayesian model averaging: A unique model enhancing forecasting accuracy for daily streamflow based on different antecedent time series. *Sustainability, 12*(22), 9720. https://doi.org/10.3390/su12229720

Lima, M., da Silva Junior, C. A., Rausch, L., Gibbs, H. K., & Johann, J. A. (2019). Demystifying sustainable soy in Brazil. *Land Use Policy, 82*, 349–352. https://doi.org/10.1016/j.landusepol.2018.12.016

Liu, Y., Zhang, W., Shao, Y., & Zhang, K. (2011). A comparison of four precipitation distribution models used in daily stochastic models. *Advances in Atmospheric Sciences, 28*, 809–820. https://doi.org/10.1007/s00376-010-9180-6

Maheswaran, R., & Khosa, R. (2012). Comparative study of different wavelets for hydrologic forecasting. *Computers & Geosciences, 46*, 284–295. https://doi.org/10.1016/j.cageo.2011.12.01

Mapbiomas. (2016). Mapa de Limite dos Biomas 1:1.000.000. https://mapbiomas.org/pages/database/reference_maps. Acessed 20 Feb 2021

Mehr, A. D., Kahya, E., Bagheri, F., & Deliktas, E. (2014). Successive-station monthly streamflow prediction using neuro-wavelet technique. *Earth Science Informatics, 7*, 217–229. https://doi.org/10.1007/s12145-013-0141-3

Mendonça, L.M., de Souza, I. G., de Sousa, J. V., Blanco, C. J. C. (2021). Modelagem chuva-vazão via redes neurais artificiais para simulação de vazões de uma bacia hidrográfica da Amazônia. *Revista de Gestão de Água da América Latina 18*(2021), https://doi.org/10.21168/rega.v18e2

Michot, V., Arvor, D., Ronchail, J., Corpetti, T., Jegou, N., Lucio, P. S., & Dubreuil, V. (2019). Validation and reconstruction of rain gauge–based daily time series for the entire Amazon basin. *Theoretical and Applied Climatology, 138*, 759–775. https://doi.org/10.1007/s00704-019-02832-w

Nash, J. E., & Sutcliffe, J. V. (1970). River flow forecasting through conceptual models part I—A discussion of principles. *Journal of Hydrology, 10*, 282–290. https://doi.org/10.1016/0022-1694(70)90255-6

Nerantzaki, S. D., & Papalexiou, S. M. (2019). Tails of extremes: Advancing a graphical method and harnessing big data to assess precipitation extremes. *Advances in Water Resources, 134*, 103448. https://doi.org/10.1016/j.advwatres.2019.103448

Ng, J. L., Aziz, S. A., Huang, Y. F., Wayayok, A., & Rowshon, M. K. (2017). Generation of a stochastic precipitation model for the tropical climate. *Theoretical and Applied Climatology, 133*, 489–509. https://doi.org/10.1007/s00704-017-2202-x

Nourani, V., Baghanam, A. H., Adamowski, J., & Kisi, O. (2014). Applications of hybrid wavelet–artificial intelligence models in hydrology: A review. *Journal of Hydrology, 514*, 358–377. https://doi.org/10.1016/j.jhydrol.2014.03.057

Nourani, V., Andalib, G., & Sadikoglu, F. (2017). Multi-station streamflow forecasting using wavelet denoising and artificial intelligence models. *Procedia Computer Science, 120*, 617–624. https://doi.org/10.1016/j.procs.2017.11.287

Partal, T., Cigizoglu, H. K., & Kahya, E. (2015). Daily precipitation predictions using three different wavelet neural network algorithms by meteorological data. *Stochastic Environmental Research and Risk Assessment, 29*(5), 1317–1329. https://doi.org/10.1007/s00477-015-1061-1

Pham, B. T., Le, L. M., Le, T. T., Bui, K. T. T., Le, V. M., Ly, H. B., & Prakash, I. (2020). Development of advanced artificial intelligence models for daily rainfall prediction. *Atmospheric Research, 237*, 104845. https://doi.org/10.1016/j.atmosres.2020.104845

Quilty, J., Adamowski, J., Khalil, B., & Rathinasamy, M. (2016). Bootstrap rank-ordered conditional mutual information (broCMI): A nonlinear input variable selection method for water resources modeling. *Water Resources Research, 52*(3), 2299–2326. https://doi.org/10.1002/2015WR016959

Ramana, R. V., Krishna, B., Kumar, S. R., & Pandey, N. G. (2013). Monthly rainfall prediction using wavelet neural network analysis. *Water Resources Management, 27*, 3697–3711. https://doi.org/10.1007/s11269-013-0374-4

Ramírez-Hernández, J., Infante-Prieto, S. O., Villa-Angulo, R., & Hallack-Alegría, M. (2016). La influencia del efecto de borde en el pronóstico de precipitaciones utilizando DWT diádica, MODWT, ANN y ANFIS. *Tecnología y Ciencias Del Agua, 7*(3), 93–113.

Roy, B., & Singh, M. P. (2020). An empirical-based rainfall-runoff modelling using optimization technique. *International Journal of River Basin Management, 18*(1), 49–67. https://doi.org/10.1080/15715124.2019.1680557

Salman, A. G., Heryadi, Y., Abdurahman, E., & Suparta, W. (2018). Single layer & multi-layer long short-term memory (LSTM) model with intermediate variables for weather forecasting. *Procedia Computer Science, 135*, 89–98. https://doi.org/10.1016/j.procs.2018.08.153

Santos, C. A., Freire, P. K., Silva, R. M. D., & Akrami, S. A. (2019). Hybrid wavelet neural network approach for daily inflow forecasting using tropical rainfall measuring mission data. *Journal of Hydrologic Engineering, 24*(2), 04018062. https://doi.org/10.1061/(ASCE)HE.1943-5584.0001725

Santos, T. S., Mendes, D., & Torres, R. R. (2016). Artificial neural networks and multiple linear regression model using principal components to estimate rainfall over South America. *Nonlinear Processes in Geophysics, 23*(1), 13–20. https://doi.org/10.5194/npg-23-13-2016

Seera, M., Lim, C. P., Ishak, D., & Singh, H. (2012). Fault detection and diagnosis of induction motors using motor current signature analysis and a hybrid FMM–CART model. *IEEE Transactions on Neural Networks and Learning Systems, 23*(1), 97–108. https://doi.org/10.1109/tnnls.2011.2178443

Shoaib, M., Shamseldin, A. Y., Melville, B. W., & Khan, M. M. (2016). A comparison between wavelet based static and dynamic neural network approaches for runoff prediction. *Journal of Hydrology, 535*, 211–225. https://doi.org/10.1016/j.jhydrol.2016.01.076

Shoaib, M., Shamseldin, A. Y., Khan, S., Khan, M. M., Khan, Z. M., Sultan, T., & Melville, B. W. (2018). A comparative study of various hybrid wavelet feedforward neural network models for runoff forecasting. *Water Resources Management, 32*(12), 83–103. https://doi.org/10.1007/s11269-017-1796-1

Silveira, L. G. T. D., Correia, F. W. S., Chou, S. C., Lyra, A., Gomes, W. B., Vergasta, L., & Silva, P. R. T. (2017). Reciclagem de precipitação e desflorestamento na Amazônia: Um estudo de modelagem numérica. *Revista Brasileira De Meteorologia, 32*(3), 417–432. https://doi.org/10.1590/0102-77863230009

Suhaila, J., Ching-Yee, K., Fadhilah, Y., & Hui-Mean, F. (2011). Introducing the mixed distribution in fitting rainfall data. *Open Journal of Modern Hydrology, 1*(2), 11–22. https://doi.org/10.4236/ojmh.2011.12002

Sulaiman, S. O., Shiri, J., Shiralizadeh, H., Kisi, O., & Yaseen, Z. M. (2018). Precipitation pattern modeling using cross-station perception: Regional investigation. *Environmental Earth Sciences, 77*(19), 709. https://doi.org/10.1007/s12665-018-7898-0

Takagi, T., & Sugeno, M. (1985). Fuzzy identification of systems and its applications to modeling and control. *IEEE Transactions on Systems, Man, and Cybernetics, 15*(1), 116–132. https://doi.org/10.1109/TSMC.1985.6313399

Vale, P., Gibbs, H., Vale, R., Christie, M., Florence, E., Munger, J., & Sabaini, D. (2019). The expansion of intensive beef farming to the Brazilian Amazon. *Global Environmental Change, 57*, 101922. https://doi.org/10.1016/j.gloenvcha.2019.05.006

Wilks, D. S. (1999). Interannual variability and extreme-value characteristics of several stochastic daily precipitation models. *Agricultural and Forest Meteorology, 93*(3), 153–169. https://doi.org/10.1016/S0168-1923(98)00125-7

Zhang, X., Peng, Y., Zhang, C., & Wang, B. (2015). Are hybrid models integrated with data preprocessing techniques suitable for monthly streamflow forecasting? Some experiment evidences. *Journal of Hydrology, 530*, 137–152. https://doi.org/10.1016/j.jhydrol.2015.09.047

Zeri, M., Cunha-Zeri, G., Gois, G., Lyra, G. B., & Oliveira-Júnior, J. F. (2018). Exposure assessment of rainfall to interannual variability using the wavelet transform. *International Journal of Climatology, 39*(1), 568–578. https://doi.org/10.1002/joc.5812