

Reservoir inflow forecasting with a modified coactive neuro-fuzzy inference system: a case study for a semi-arid region

Mohammed Falah Allawi¹ · Othman Jaafar¹ · Firdaus Mohamad Hamzah¹ ·
Nuruol Syuhadaa Mohd² · Ravinesh C. Deo³ · Ahmed El-Shafie²

Received: 27 July 2017 / Accepted: 6 October 2017 / Published online: 14 October 2017
© Springer-Verlag GmbH Austria 2017

Abstract Existing forecast models applied for reservoir inflow forecasting encounter several drawbacks, due to the difficulty of the underlying mathematical procedures being to cope with and to mimic the naturalization and stochasticity of the inflow data patterns. In this study, appropriate adjustments to the conventional coactive neuro-fuzzy inference system (CANFIS) method are proposed to improve the mathematical procedure, thus enabling a better detection of the high nonlinearity patterns found in the reservoir inflow training data. This modification includes the updating of the back propagation algorithm, leading to a consequent update of the membership rules and the induction of the centre-weighted set rather than the global weighted set used in feature extraction. The modification also aids in constructing an integrated model that is able to not only detect the nonlinearity in the training data but also the wide range of features within the training data records used to simulate the forecasting model. To demonstrate the model's efficacy, the proposed CANFIS method has been applied to forecast monthly inflow data at Aswan High Dam (AHD), located in southern Egypt. Comparative analyses of the forecasting skill of the modified CANFIS and the conventional ANFIS model are carried out with statistical

score indicators to assess the reliability of the developed method. The statistical metrics support the better performance of the developed CANFIS model, which significantly outperforms the ANFIS model to attain a low relative error value (23%), mean absolute error (1.4 BCM month⁻¹), root mean square error (1.14 BCM month⁻¹), and a relative large coefficient of determination (0.94). The present study ascertains the better utility of the modified CANFIS model in respect to the traditional ANFIS model applied in reservoir inflow forecasting for a semi-arid region.

1 Introduction

1.1 Background

Dam reservoirs are significant components of water resource management systems. The main purpose of reservoirs is to provide and regulate sufficient irrigation water, drought and flood control, and hydropower generation and to perform other hydrological functions for day to day living. Therefore, efficient reservoir operations by means of deducing a suitable schedule of water release policy is important for strategic water resources management. An accurate and reliable reservoir inflow forecast is a vital reference for decision-makers in reducing the impacts of water surpluses or water deficits. Furthermore, determining the most proper model for forecasting future water inflows can play an essential role in making appropriate decisions for reservoir management and for providing effective and successful reservoir policies. In this context, developing model for reservoir inflow forecasting remains of particular interest in the operational areas of hydrology and water resources planning (Coulbaly et al. 2000; Lohani et al. 2012).

✉ Mohammed Falah Allawi
mohmmd.falah@gmail.com

¹ Civil and Structural Engineering Department, Faculty of Engineering and Built Environment, Universiti Kebangsaan Malaysia, Bangi, Malaysia

² Civil Engineering Department, Faculty of Engineering, University of Malaya, 50603 Kuala Lumpur, Malaysia

³ School of Agricultural, Computational and Environmental Sciences, Institute of Agriculture and the Environment, University of Southern Queensland, Springfield, QLD 4300, Australia

Reservoir inflow patterns entail highly complex processes to be described using simple predictive models because of the nonlinearity, spatial distribution, and time varying characteristics of the data (Bai et al. 2015; Valizadeh et al. 2017). There are two primary methods for inflow forecasting that have been examined in previous studies: mechanistic or “physically-based” models and system theoretic (or data-driven) models. Developing mechanistic models for encapsulating the hydrological processes is a devious task for decision-makers, yet they require high-level erudite mathematical procedures and significant amount of calibration (or initial/boundary condition) data to secure an acceptable level of modeling accuracy. Consequently, the hydrologists have paid a lot of attention to utilize the rapidly evolving data-driven models based on system theoretic principles rather than using the highly complex mechanistic models as a suitable alternative to model the complex hydrological process. This is due to the fact that data-driven models have the aptitude to adequately mimic the input-output dynamics of water systems without requiring an insightful understanding of the fundamental physical processes of the system. Recently, most conceptual reservoir inflow forecasting models are embracing data-driven models for reservoir inflow forecasting presumably due to practical reasons with physical models, such as the calibration issues and data unavailability (El-Shafie and Noureldin 2011). Furthermore, physically-based models also tend to disregard the time-varying, and stochastic characteristic of the inflow system, so data-driven models which attempt to combine the representative nonlinear relationships into the reservoir inflow values (as time series and climatology parameters) are becoming popular tools for water resource managers.

Data-driven (or system theoretic) methods forecast the inflow patterns by direct mapping of relationship between input variables and a target (output) set-based time series procedure (El-Shafie et al. 2009; Keshtegar et al. 2016). This mapping between the variables is achieved without the need for elaborate consideration of the internal framework of the physical processes. One of the most conventional method under this approach is the auto regressive moving average (ARMA) which is a linear approach proposed by Box and Jenkins (1970). The ARMA has commonly been utilized for inflow forecasting due to its straightforward procedure and effortless development process (Valipour et al. 2013). Although, such approaches have successfully revealed satisfactory results for reservoir inflow forecasting, these models were unsuccessful in several aspects of their implementations, mainly due to their inability to address nonlinear and dynamical behavior of the inflow values. Since these models do not always enable highly accurate values (Valipour et al. 2013), alternative nonlinear models are highly warranted.

1.2 Problem statement

Soft computing approaches such as the artificial neural networks (ANN), fuzzy inference set (FIS), and adaptive neuro-fuzzy inference systems (ANFIS) have become increasingly popular modeling tool for inflow forecasting. Over the recent three decades, ARMA models have been slowly replaced by soft computing models that show greater ability to detect the nonlinearity dynamics present in the reservoir inflow patterns. This popularity is evidenced in recent research works that have dealt with the implementation of soft computing approaches (Smith and Eli 1995; Whigham and Crapper 2001; Jothiprakash and Magar 2009, 2012; Ju et al. 2009; El-Shafie et al. 2009; Lin and Wu 2011; Guo et al. 2011; Kisi et al. 2012; Danandeh Mehr et al. 2013; Chen et al. 2013; Deo and Şahin 2016; Allawi and El-Shafie 2016). However, there can be a difficulty in the associated mathematical procedure in the ANN and the ANFIS model to detect the highly stochastic patterns and wide range of the inflow attributes present in the data, which warrants a need to enhance soft-computing procedures.

In this study, a consideration is given to the variation of the ANFIS model with an aim to improve its performance. The ANFIS model is a good candidate because it has been proven to possess the highest ability among all of the soft-computing approaches to detect the nonlinearity patterns in the input data. In fact, the ANFIS method, which integrated the ANN with FIS algorithm, was first developed as multiple inputs—single output procedure where the merits of the two standalone (ANN & FIS) models were combined for improved performance. The study of Jang et al. (1997) (Jang et al. 1997) proposed further enhancements to the ANFIS procedure, where they aimed to consider applications with multiple inputs—multiple output patterns and also introduced a method namely, the coactive neuro-fuzzy inference system (CANFIS). Such new procedure is likely to permit enhancements in the conventional ANFIS model since it encompasses the back propagation procedure to allow a re-evaluation of the rule of the membership function used in the prediction algorithm. With this improvement, the CANFIS method is likely to better detect the nonlinearities and the stochastic patterns at several local points in the predictor dataset. This is particularly useful for long historical input data used in a predictive model, as used in our case study.

1.3 Objectives

This research paper investigates the performance of a reinigorated CANFIS procedure to construct a reliable forecasting model for reservoir inflow forecasting. A total of 130 years monthly basis inflow data measured at Aswan High Dam (AHD) has been used to examine the predictive

model. The different training approaches with different data splitting criterion have been considered to examine the ability of the model to detect the nonlinearity pattern between the input and the output. The present study discusses the influence of the four different training approaches on the reservoir inflow forecasting accuracy. Finally, a comprehensive comparison analysis between the improved model (CANFIS) and the conventional model (ANFIS) is carried out via several statistical indexes to examine the reliability of the inflow forecasting accuracy.

2 Case study

The main water resource infrastructure in Egypt is the Aswan High Dam (AHD). The AHD was constructed to provide Egypt a long-term protection against floods and drought. The Nile River flows from south to north, supplying water to AHD; it has a length of approximately 6850 km. Almost 10% of all African countries depend on the Nile catchment basin. The reservoir of the dam (i.e., Lake Nasser) supplies water for different purposes such as irrigation, industrial use, navigation, and energy production. This research has utilized 130 years monthly inflow data (1870 to 2000), recorded from the upstream gauge (i.e., Dongola gauge station). The data were collected from the Nile Water Authority (NWA), Aswan High Dam Authority (AHDA), and Ministry of Water Resources and Irrigation, Egypt (El-Shafie and Noureldin 2011). Figure 1 shows the location of Lake Nasser which was created by the AHD. The water storage capacity of this lake is about 160 billion m³ (BCM).

By a preliminary analysis of the data that were collected, it is observable that the historical inflow data during the study period has a high monthly fluctuation pattern. The natural inflow of the AHD is presented in Fig. 2. It is evident that the average values of the inflow are different for the entire 130 years for both the yearly and the monthly time scales. Figure 2a–c shows the different extreme events that occurred in three different months, which show the dynamical behavior of the water inflow in the AHD. For example, it is noticeable that the maximum record in August, February, and June were (29.1 BCM month⁻¹), (6 BCM month⁻¹), and (5.16 BCM month⁻¹), respectively. Conversely, the low inflow values observed in August, February, and June were found to be (6.5 BCM month⁻¹), (1.15 BCM month⁻¹), and (0.9 BCM month⁻¹) respectively. Such a wide range of data recorded in the historical inflow period reveals the challenge in accomplishing accurate forecasting of such data, and the need for optimized models to attain reliable forecasting accuracy.

3 Modeling methodology

3.1 Adaptive neuro-fuzzy inference system (ANFIS) method

ANFIS model has a proven capability for highly nonlinear feature mapping and this model is seen to outperform the ordinary linear approaches when utilizing nonlinear time series data (Jang et al. 1997). Developed as a multi-layer feed forward network, this model adopts a neural network learning algorithm coupled with a fuzzy-logic tool to map the input-output space in such a way that the mean square error between forecasted and observed data in training period are minimized.

In an ANFIS model, a popular rule set with two fuzzy IF-THENs is defined in the following way (Eqs. 1 and 2):

$$\text{Rule 1 : if } u \text{ is } u_1 \text{ and } n \text{ is } n_1 \text{ then } f_1 = p_1u + q_1n + r_1 \quad (1)$$

$$\text{Rule 2 : if } u \text{ is } u_2 \text{ and } n \text{ is } n_2 \text{ then } f_2 = p_2u + q_2n + r_2 \quad (2)$$

where u and n are the membership functions related to input variable(s), and the outputs are the p , q , and r parameters. The architecture of the ANFIS method comprises of five primary layers, which are the fuzzifications layer, firing strength layer, normalized firing strength layer, consequent parameters layer, and the overall output layer, (Lohani et al. 2007). Further theoretical details of the ANFIS model can be found in (Jang et al. 1997).

3.2 Co-active neuro-fuzzy inference system method

The conventional form of the CANIFIS model is an extension of the original ANFIS model (Jang et al. 1997). The underlying concept of the CANFIS model can be extended to consider any number of input and output pairs, as shown in Fig. 3a (Shoab et al. 2016). The essential component of the CANFIS is generally similar to the components of the ANFIS model where a fuzzy neuron that represents a parameter called the membership function (MF) is used to construct the modeling framework. There are several types of MFs that can be utilized, including the triangular, trapezoidal, sigmoidal, Gaussian, z-shape, and s-shape functions, pi, the general bell and the Gaussian equations. CANFIS structure includes the normalizing axon that aims to normalize the output variables in the range of 0–1. The architectural network of CANFIS model has the combined axon applied the target (or output) of the MFs to the target for the neuronal network (Alecsandru and Ishak 2004), which is still similar to the common ANFIS procedure.

CANFIS model is improved such that, in order to generate multiple outputs, the model aims to maintain the same antecedent behavior of the ANFIS procedure, with the restructure that allows the modeling procedure to use fuzzy rules that are constructed with shared membership values to express

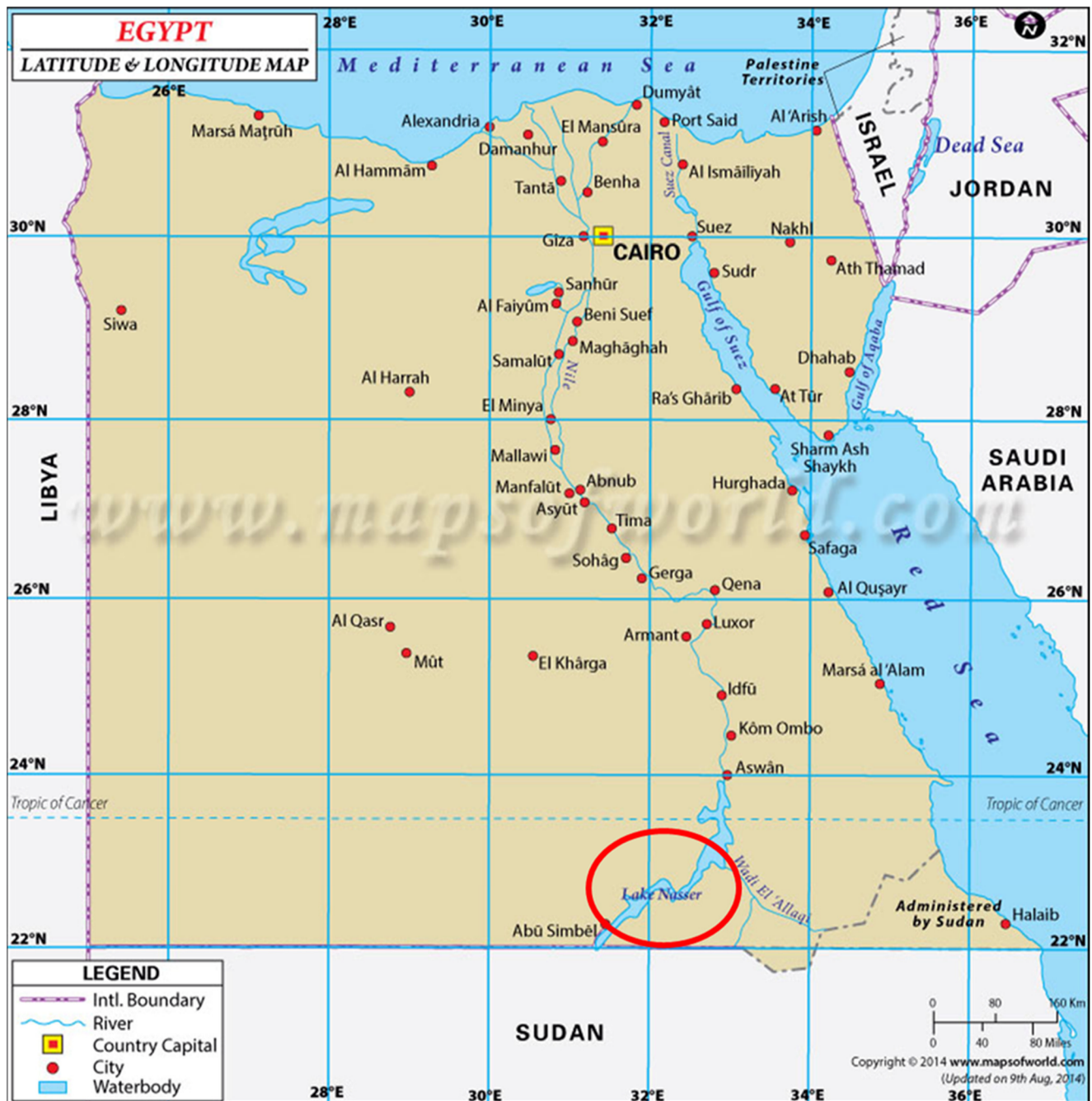


Fig. 1 Geographic location of the Aswan High Dam (AHD) and the Lake Nasser region in Egypt highlighted by the red circle (source: https://www.mapsofworld.com/lat_long/egypt-lat-long.html)

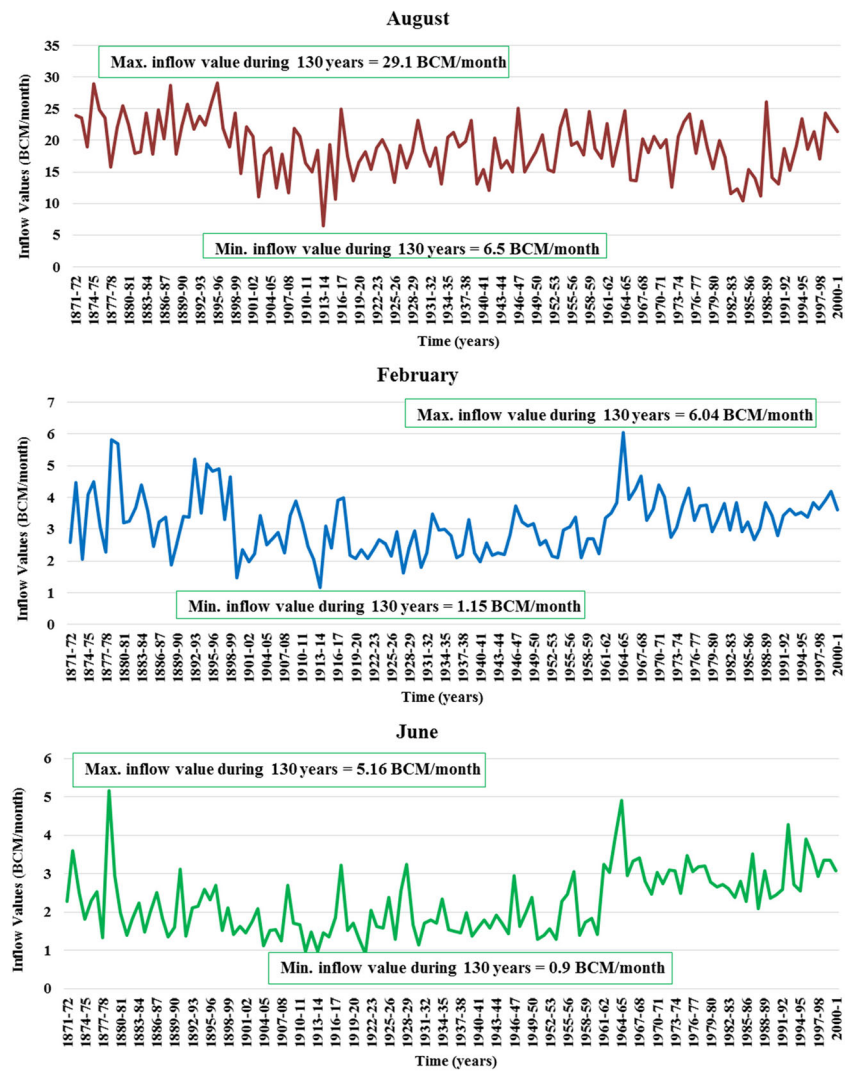
possible correlations with the output. Furthermore, when the consequent parts are correlated (i.e., when two neural consequents, neural rule1 and neural rule 3 are fused into one NN_1 and neural rule 2 and neural rules 4 are fused into another NN_2), a typical modular network is constructed in similar fashion. The outputs of the two NNs are mediated and trained by an integrating unit (typically a gating network). The training is based on an independent training scheme whereby the antecedent parts and the consequent parts are trained individually and then positioned together. Another training approach

used in the CANFIS model is to train the antecedent and the consequent parts concurrently.

3.3 The modified CANFIS model

One of the substantial tasks that needs to be performed in the CANFIS-based modeling is the predefined selection of the optimal internal parameters for the CANFIS architecture (including the number and shape of the MFs) (Tabari et al. 2012; Lohani et al. 2012). The number of MFs selected before the

Fig. 2 Historical naturalized inflow trends at the AHD for months of August, February, and June for years between 1870 and 2000



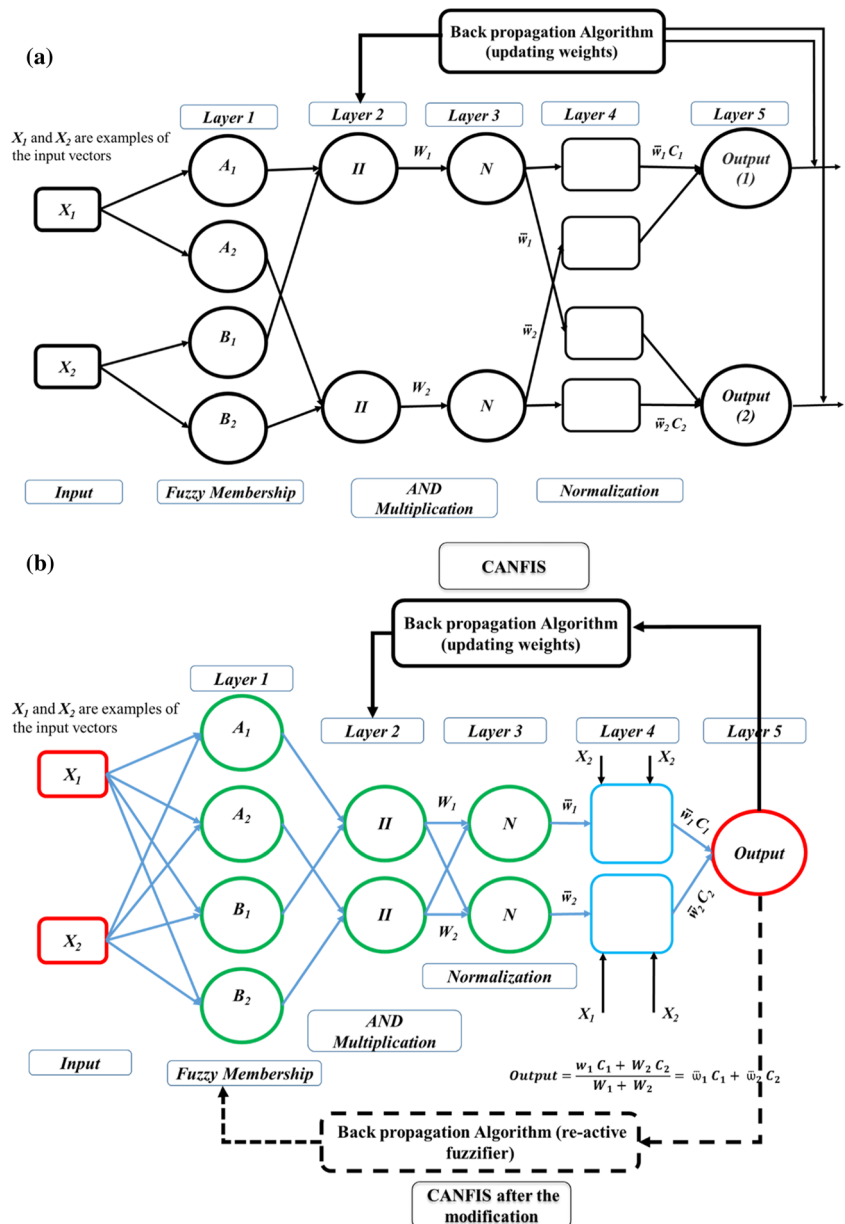
training phase contributes to the definition of every mapping between input and output variables, hence providing a desirable and an improved predictive model (El-Shafie et al. 2007). Deducing the proper values of the CANFIS membership functions is a thus significant process for sufficiently mapping the variables (Lohani et al. 2012).

A noteworthy point is that the CANFIS procedure permits the modeler to introduce a further step to expand the back propagation phase to commence at the fuzzification layer. In respect to ANFIS, a major change in the CANFIS procedure is to begin the common nodes of the input set that are passed to the last layer of output. However, the errors are modified by a back propagation phase toward the weights and biases within the fuzzification layer, which accords to the membership rule. In this context, the fuzzy axons and their MFs could be moderated using the back-propagation framework through the system exercises and the adjustment of the weight and biases for the particular patterns, leading to an overall improvement in the model's accuracy.

As presented in Fig. 3b, “the dotted line connections,” show an implementation of the modified CANFIS model procedure by means of modifying the available barebones back propagation algorithm. In accordance with this, the CANFIS back propagation procedure stems from the pattern-dependant weights between the consequent layer and the fuzzy association layer, namely, the membership rule. The modification proposed for the back propagation procedure to the membership rule layer is known as the “fuzzification rule.” In this context, the membership values correspond to those of dynamically changeable weights that depend on the input pattern. In fact, the current back propagation procedure before modification (Fig. 3b, “solid line connections”) attempts to deduce one specific set of weights that are common to all training patterns. In other word, the weights are utilized in a global fashion in the CANFIS model.

In this study, we propose to introduce modifications in the CANFIS model such that it allows the back-propagation algorithm to adjust the weights in a much local manner (i.e.,

Fig. 3 The architecture of the CANFIS model; **a** original structure with multiple inputs—multiple outputs; **b** modified structure with new back propagation algorithm procedure and multiple inputs—single output



allowing a better feature extraction from the training inflow dataset). With this modification that aims to improve the overall accuracy, the CANFIS model structure has a new mechanism whereby it is able to generate a center-weighted response to small receptive fields, thus, localizing the primary input excitation. In this sense, the proposed modified CANFIS model in this study can be functionally equivalent to the classification of the input pattern to different categories of features with different set of weights.

Every node of layer 1 has the MFs degree for a fuzzy group (i.e., R_1, R_2, T_1, T_2) and this specifies the grade of the presented input that is associated with one of fuzzy sets. The second step (layer) receives the results of each target (output) which has arrived from the previous step. A third

step includes two major sections, known as the upper and the lower section. The first component (upper section) applies the MFs to all input variables, while the second section (lower section) is the demonstration of the employment network. The latter calculates the total firing strengths for all outputs (Aytek 2009; Yan et al. 2010).

Layer 4 generates a definitive output of the network, which provides the final shape for the output variables by computing the normalized weights of the output that arrives from the upper and lower components in the third layer. Two models are fundamentally utilized as the training procedure, the Tsukamoto and the Sugeno model. For the MFs, the three parameters belonged for the bell type, whereas two parameters for correction belonged for Gaussian

function type (Chang and Chang 2006). The bell-shaped function is defined as below:

$$\mu_{Ai} = \frac{1}{1 + \left| \frac{x+c_i}{a_i} \right|^{2b_i}} \quad (3)$$

$$\mu_{Ai} = \frac{1}{1 + \left| \frac{x-c_i}{a_i} \right|^{2b_i}} \quad (4)$$

where a , b , and c values are the parameters of the bell-shaped function that change the shape of the membership function. The generalized bell function is more adaptable compared to the Gaussian function. Accordingly, the bell function has been utilized for the training model in this research paper.

The networks are training with an error correction learning, which implies that a target response of a system should be recognized. This correction works through a system response at a processing element at iteration n , $y_i(n)$. In addition, a target response $d_i(n)$ of a certain input pattern and the immediate error $e_i(n)$ is expressed as:

$$e_i = d_i(n) - y_i(n) \quad (5)$$

Utilizing the concept of the gradient descent learning algorithm, the weights in the CANFIS model is acclimated to using the correction of the current weights for the terms, which is symmetrical of a current input variables as well as the errors which are encountered in the weights (Memarian et al. 2013).

The back-propagation algorithm works to update the weight coefficients for every input pattern in the CANFIS model. With the modified CANFIS model, this algorithm modifies the membership function in addition to updating the rules of the modular neural network. It should be noted that the original CANFIS model generally minimizes the error from the updating of the weights for the neural network without changing fuzzy axons. However, the processing technique used by the CANFIS model after introducing the modification, as this study, aims to facilitate the reactive fuzzifier process. This improvement provides a new membership function value where the re-adjustment of the fuzzy axons by itself is performed, thus creating new rules to not only reduce the training error but also to create local weight sets for each input pattern. In respect of the significant improvement introduced in this study, the new CANFIS model is considered to be superior to the original CANFIS and the predecessor, i.e., ANFIS model. With the modifications introduced for the CANFIS model in this paper, the new adaptive learning procedure for the MFs is likely to lead to a greater precision within a fixed (and reasonable) amount of the computation time.

3.4 Model evaluation

In this study, several statistical measures have been utilized to examine the performance accuracy of the proposed method for reservoir inflow forecasting. These indicators aim to examine the effect of the training approach and the testing performance of the models. The statistical measures used for model evaluation are: the mean absolute error (MAE), correlation coefficient (R^2), root-mean-square error (RMSE), and percentage of relative error (%RE). The coefficient of determination (R^2) (Eq. 6) is a popular indicator that compares the covariance of the observed and forecasted data in the testing phase (Gillberg and Wahlström 2008). The resulting error between the forecasted and the observed inflow data is computed through the RMSE and MAE (Eqs. 7 and 8, respectively). The relative error is the difference between forecasted and actual values over the actual inflow, as shown in Eq. (9).

$$R^2 = \frac{\sum_{t=1}^n \left[\left((I_o) - \overline{(I_o)} \right) \left((I_f) - \overline{(I_f)} \right) \right]}{\sqrt{\sum_{t=1}^n \left((I_o) - \overline{(I_o)} \right)^2 \sum_{t=1}^n \left((I_f) - \overline{(I_f)} \right)^2}} \quad (6)$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{t=1}^n \left((I_f) - (I_o) \right)^2} \quad (7)$$

$$MAE = \frac{1}{N} \sum_{t=1}^n |I_f(t) - I_o(t)| \quad (8)$$

$$\%RE = \frac{I_f(t) - I_o(t)}{I_o(t)} * 100 \quad (9)$$

The forecasted values are represented in the above equations by I_f , while the observed inflow values are given the symbol I_o , and the parameter n represents the number of data points in the test phase. For further analysis, two different visual diagnostic evaluation measures, the Taylor diagram and the boxplot of error distributions, have also been considered to examine the actual and forecasted inflow data.

3.5 Data and model development

Table 1 shows the statistical parameters of the monthly inflow data for the case study site, (i.e., the AHD). Here, the mean (X_{mean}), maximum (X_{max}), and minimum (X_{min}) values over the 130 years of operation have been enumerated. In addition, this table includes the standard deviation (S_x), skewness (C_{sx}), variation coefficient (C_v), and the median values for the monthly inflow data. In terms of the dynamical changes, a relatively high standard deviation is evident for the September data, while for the March records, the standard deviations are relatively low. The inflow values for the month of May reveal a high variation coefficient, whereas a low variation coefficient is observed for the records obtained

Table 1 Statistical analysis of the monthly inflow data at the AHD for the study period between 1870 and 2000

Month	X_{mean} (BCM month ⁻¹)	S_x (BCM month ⁻¹)	C_v (S_x / X_{mean})	C_{sx}	X_{max} (BCM month ⁻¹)	X_{min} (BCM month ⁻¹)	Median (BCM month ⁻¹)
January	4.30	1.003	23.30	0.718	7.7	1.72	4.2
February	3.15	0.90	28.51	0.654	6.04	1.15	3.12
March	2.75	0.86	31.44	0.816	5.81	1.07	2.69
April	2.52	1.05	41.61	0.428	5.25	0.95	2.5
May	2.33	1.01	43.54	0.449	4.71	0.8	1.96
June	2.24	0.82	36.59	0.810	5.16	0.9	2.1
July	5.24	1.76	33.66	0.758	11.02	1.74	5.03
August	18.96	4.32	22.82	-0.074	29.1	6.5	18.86
September	21.50	5.053	23.49	-0.196	32.79	7.31	21.45
October	14.51	4.23	29.16	0.168	27.4	5.96	14.8
November	8.11	2.24	27.65	0.564	14.4	4.12	8
December	5.54	1.36	24.55	1.198	11	2.83	5.22

X_{mean} mean value, S_x standard deviation, C_v coefficient of variation, C_{sx} skewness, X_{min} minimum value, X_{max} maximum value)

during the month of August. Another remarkable feature in Table 1 is that the high and low skewness indicators correspond to the months of December and September, respectively, and the minimum and maximum monthly inflow values during the 130 year period were recorded for the months of May and September, respectively. In terms of the median values, the data show that of the second half of the year median values were relatively high compared to the other months over the 130 year period. This is attributable to the fact that reservoir of the present dam generally receives a large water volume beginning from the month of August until the end of the year.

In order to construct the models, the reservoir inflow data from 1870 to 2000 were separated into two distinct sets: the training and validation set (or testing). This study has investigated the ability of the proposed models by the inclusion of four different training procedures. That is, we have utilized four distinct training approaches to develop a robust predictive model based on the modified CANFIS and the ANFIS modeling techniques.

The first procedure (denoted as “training approach #1”) aimed to establish the model structure based on the partitioned data with 75% of all available values was allocated to training of the models and the remaining 25% of data used to verify (or test) the trained models. In the second training procedure (“training approach #2”), a new model was built after changing the percentage of the inflow data, to about 80% of the inflow values being used during the training phase, and the remainder 20% of data was used to validate (or test) the predictive models. In the third training procedure (“training approach #3”), the monthly inflow data for the period from 1930 to 1997 (i.e., 15% of total data) are used for testing the performance of the models and using the remainder 85% of the

all data for training the models. Finally, the fourth training procedure (“training approach #4”) allocated 90% of total data to the training of the models and 10% of the data used for the model testing phase. The utilization of four distinct data partitioning, and the subsequent modeling procedures, enabled a robust extraction of the data attributes necessary to attain the most accurate predictive model. The distribution of inflow data during training and testing period with four different training procedures are shown in Fig. 4.

The present study has developed the most optimal architecture of the modified CANFIS model comprised of the multiple inputs—single output system where statistically significant lagged combinations of historical inflow data, I_a , as per earlier studies (e.g., Chiew et al. 1998; Deo and Şahin 2016), were used to construct the most accurate predictive models.

The present models utilized five sets of input data series: $I_a(t-1)$, $I_a(t-2)$... $I_a(t-5)$, and considered the role of statistical memory in constructing forecasted value of the inflow (I_f). The architecture of the inflow forecasting models using the ANFIS and modified CANFIS techniques are expressed as:

$$Model\ 1 : (I_f = I_a(t-1)), \tag{10}$$

$$Model\ 2 : (I_f = I_a(t-1), I_a(t-2)), \tag{11}$$

$$Model\ 3 : (I_f = I_a(t-1), I_a(t-2), I_a(t-3)),$$

$$Model\ 4 : (I_f = I_a(t-1), I_a(t-2), I_a(t-3), I_a(t-4)), \tag{12}$$

$$Model\ 5 : (I_f = I_a(t-1), I_a(t-2), I_a(t-3), I_a(t-4), I_a(t-5)), \tag{13}$$

where I_f = the forecasted inflow representing the output and I_a = actual reservoir inflow representing the input data.

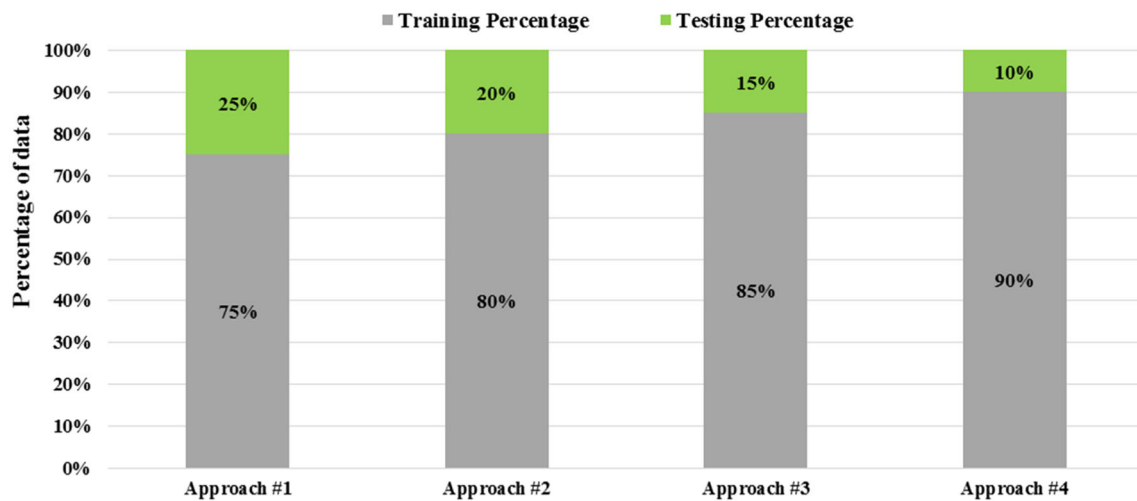


Fig. 4 Data splitting during training and testing phases in the four different training approaches

4 Result and discussion

The modified CANFIS and the comparative ANFIS models for reservoir inflow forecasting were developed according to the procedure discussed in the previous sections. A comprehensive analysis of the results and a comparison between their performances were performed to examine the functionality of the proposed techniques. The evaluation of the models' performance and the influence of the training approach on the forecasted results have been summarized in this section. To yield a deeper and more conclusive analysis of the present modeling approach, the evaluation of the methods' reliability was carried out in the model training and testing phases with respect several statistical indices.

The statistical indicators of model accuracy including the MAE, RMSE, R^2 , and the maximum percentage of relative error (%RE) for both methods using training approach #1 are shown in Table 2. It can be seen that model II is considerably better than other models for both of the proposed methods. Importantly, the results clearly show that the modified CANFIS model is more effective than the ANFIS model when both methods are subjected to the testing dataset. This is evident in the R^2 obtained by the modified CANFIS model, which is approximately 0.78 (a relatively high utility value compared to the ANFIS model). Furthermore, the MAE and RMSE values generated by the modified CANFIS model are less than those using the ANFIS model. On the other hand, the maximum percentage of the relative error of model II using the modified CANFIS is less than those yielded by the ANFIS model.

Table 3 enumerates the statistical analysis of the results obtained by proposed artificial-intelligence models using training approach #2. It is remarkable to note that the performance of the ANFIS and the model modified CANFIS using two input variables (i.e., $I_f = I_a(t-1), I_a(t-2)$) used in Model II) is better

compared than using two input variables in training approach # 1, suggesting that the utilization of 80% of data for training leads to an optimization of the feature extraction in reservoir inflow forecasting. Notably, in the calibration (or training) period, the worst performance is also attained using model I where only $I_f = I_a(t-1)$ is used as the model input. In concurrence with the results of training approach # 2, the results of the model II obtained by the modified CANFIS model is more accurate than those provided by the ANFIS model.

Proceeding to the assessment of both models with training approach # 3, Table 4 demonstrates the values of RMSE, MAE, R^2 , and the maximum percentage of relative error (%RE) attained using the ANFIS and modified CANFIS. Note that here, the data partitioning used 85% (training) and 15% (testing) ratios. Further improvement in the model accuracy for the training phase is evident for the ANFIS model in the case of model II with a smaller value of error and a larger coefficient of determination compared to the ANFIS model in training approach # 2 for 1-month-ahead inflow forecasting. We also note that the predictive ability of the modified CANFIS model using two antecedent values of reservoir inflow data as the model input is superior to the ANFIS method with same input variables (i.e., with relative error of 23 vs. 31%). It should be highlighted that the relative error indicator in the testing phase is considered to be one of the important statistical indicators using to examine the proposed models. It can be seen from Table 4 that the maximum error percentage using the modified CANFIS method with two previous inflow records, $I_f = I_a(t-1), I_a(t-2)$ is considerably less than those obtained by other ANFIS model. This shows that the modified CANFIS technique is able to yield accurate forecasts of reservoir inflow data.

Further examination of the modified CANFIS vs. the ANFIS model for different input combinations (Eqs. 10, 11, 12, and 13) is presented Table 5 where the predictive models

Table 2 Performance indicators (MAE, RMSE, R^2 , and maximum %RE) values for the training approach #1 using the ANFIS and the modified CANFIS methods during training and testing periods. The optimal model has been italicized

CANFIS																
Model	Training				Testing				Training				Testing			
	MAE (BCM month ⁻¹)	RMSE (BCM month ⁻¹)	R^2	%RE	MAE (BCM month ⁻¹)	RMSE (BCM month ⁻¹)	R^2	%RE	MAE (BCM month ⁻¹)	RMSE (BCM month ⁻¹)	R^2	%RE	MAE (BCM month ⁻¹)	RMSE (BCM month ⁻¹)	R^2	%RE
M1	3.02	4.11	0.47	42	3.2	4.37	0.52	42	2.07	4.15	0.46	40	3.29	4.41	0.57	37
M2	1.34	3.11	0.72	41	1.95	3.31	0.68	40	<i>1.57</i>	<i>2.53</i>	<i>0.81</i>	<i>40</i>	<i>1.92</i>	2.93	<i>0.78</i>	<i>35</i>
M3	2.03	3.24	0.65	46	2.25	3.78	0.60	43	2.23	2.82	0.73	43	2.4	3.28	0.70	39
M4	2.06	3.41	0.63	45	2.31	3.84	0.58	47	1.82	3.24	0.61	44	3.01	3.79	0.63	40
M5	1.96	3.32	0.71	48	2.28	3.63	0.62	48	2.47	3.73	0.52	47	2.88	4.09	0.50	42

Table 3 Performance indicators (MAE, RMSE, R^2 , and maximum %RE) values for the training approach #2 using the ANFIS and the modified CANFIS methods during training and testing periods. The optimal model has been italicized

CANFIS																
Model	Training				Testing				Training				Testing			
	MAE (BCM month ⁻¹)	RMSE (BCM month ⁻¹)	R^2	%RE	MAE (BCM month ⁻¹)	RMSE (BCM month ⁻¹)	R^2	%RE	MAE (BCM month ⁻¹)	RMSE (BCM month ⁻¹)	R^2	%RE	MAE (BCM month ⁻¹)	RMSE (BCM month ⁻¹)	R^2	%RE
M1	2.93	3.98	0.63	41	3.1	4.18	0.65	38	2.91	3.97	0.61	37	3.18	4.21	0.68	34
M2	1.34	2.97	0.73	37	1.86	3.11	0.71	34	<i>1.24</i>	<i>2.34</i>	<i>0.84</i>	<i>35</i>	<i>1.73</i>	2.63	<i>0.80</i>	29
M3	1.97	3.27	0.64	43	2.14	3.54	0.62	37	1.31	2.34	0.76	38	1.85	2.67	0.77	33
M4	2.03	3.28	0.62	42	2.19	3.58	0.61	38	2.42	3.56	0.60	39	2.63	3.75	0.57	35
M5	1.98	3.23	0.63	46	2.15	3.4	0.64	41	1.94	3.22	0.65	42	2.01	3.34	0.63	35

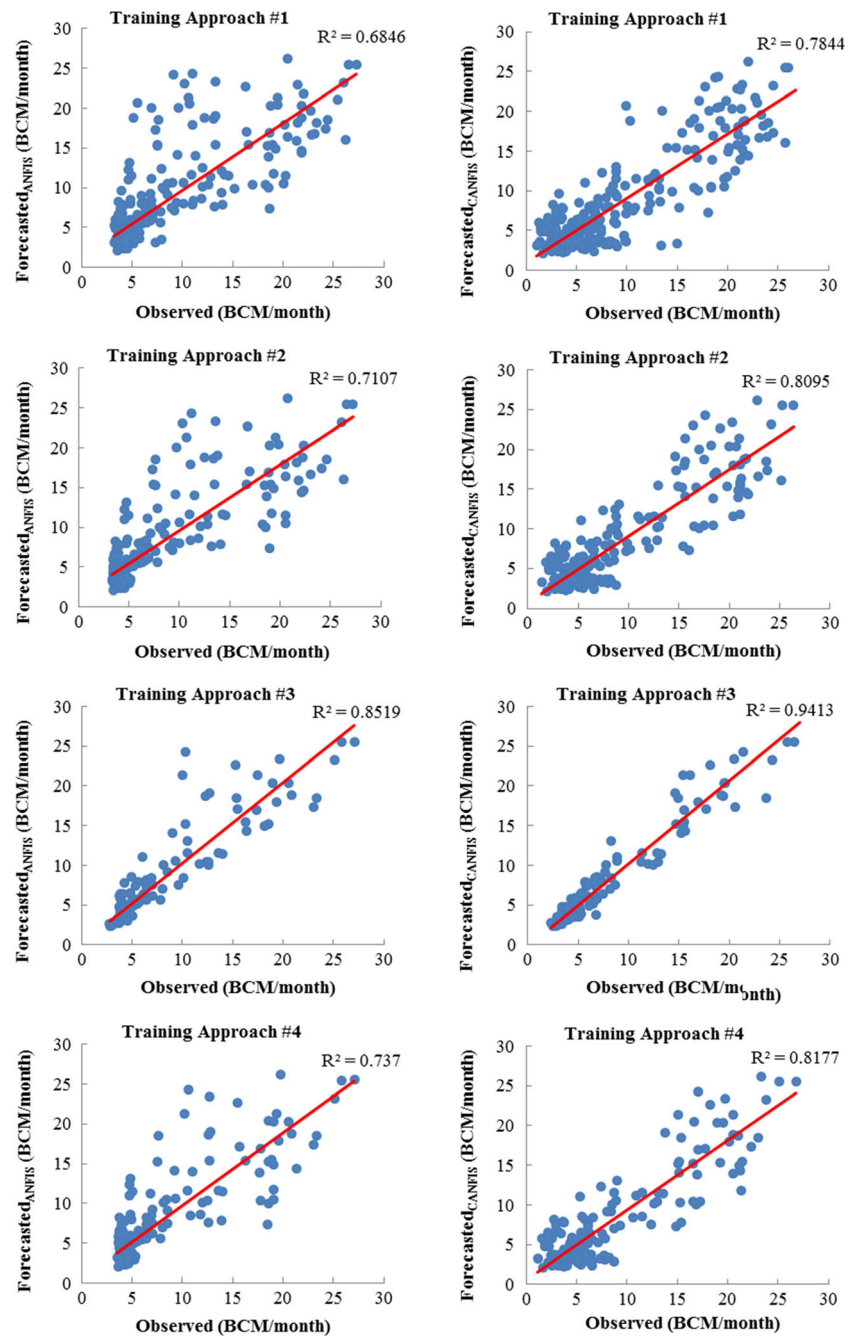
Table 4 Performance indicators (MAE, RMSE, R^2 , and maximum %RE) values for the training approach #3 using the ANFIS and the modified CANFIS methods during training and testing periods. The optimal model has been italicized

CANFIS												
Model	Training					Testing						
	MAE (BCM month ⁻¹)	RMSE (BCM month ⁻¹)	R^2	%RE	MAE (BCM month ⁻¹)	RMSE (BCM month ⁻¹)	R^2	%RE	MAE (BCM month ⁻¹)	RMSE (BCM month ⁻¹)	R^2	%RE
M1	2.98	4.03	0.62	38	3.09	4.23	0.68	35	3.02	3.82	0.68	34
M2	1.52	2.55	0.75	32	1.81	2.94	0.85	31	1.23	1.12	0.93	26
M3	1.92	3.08	0.68	41	2.05	3.36	0.66	33	1.54	2.07	0.84	37
M4	1.68	2.83	0.71	35	2	3.21	0.68	35	1.61	2.11	0.72	32
M5	1.82	3.01	0.70	34	2.09	3.27	0.67	37	1.64	2.51	0.70	30

Table 5 Performance indicators (MAE, RMSE, R^2 , and maximum %RE) values for the training approach #4 using the ANFIS and the modified CANFIS methods during training and testing periods. The optimal model has been italicized

CANFIS												
Model	Training					Testing						
	MAE (BCM month ⁻¹)	RMSE (BCM month ⁻¹)	R^2	%RE	MAE (BCM month ⁻¹)	RMSE (BCM month ⁻¹)	R^2	%RE	MAE (BCM month ⁻¹)	RMSE (BCM month ⁻¹)	R^2	%RE
M1	3.12	4.34	0.51	44	3.25	4.46	0.65	42	3.21	4.05	0.63	43
M2	1.75	2.96	0.77	42	1.84	3.07	0.73	40	1.52	1.98	0.85	40
M3	1.85	2.98	0.71	48	2.03	3.51	0.68	45	1.66	2.25	0.79	47
M4	1.95	3.02	0.68	54	2.08	3.55	0.67	47	2.37	3.08	0.74	48
M5	2.01	3.02	0.72	57	2.03	3.36	0.70	54	1.88	3.02	0.73	51

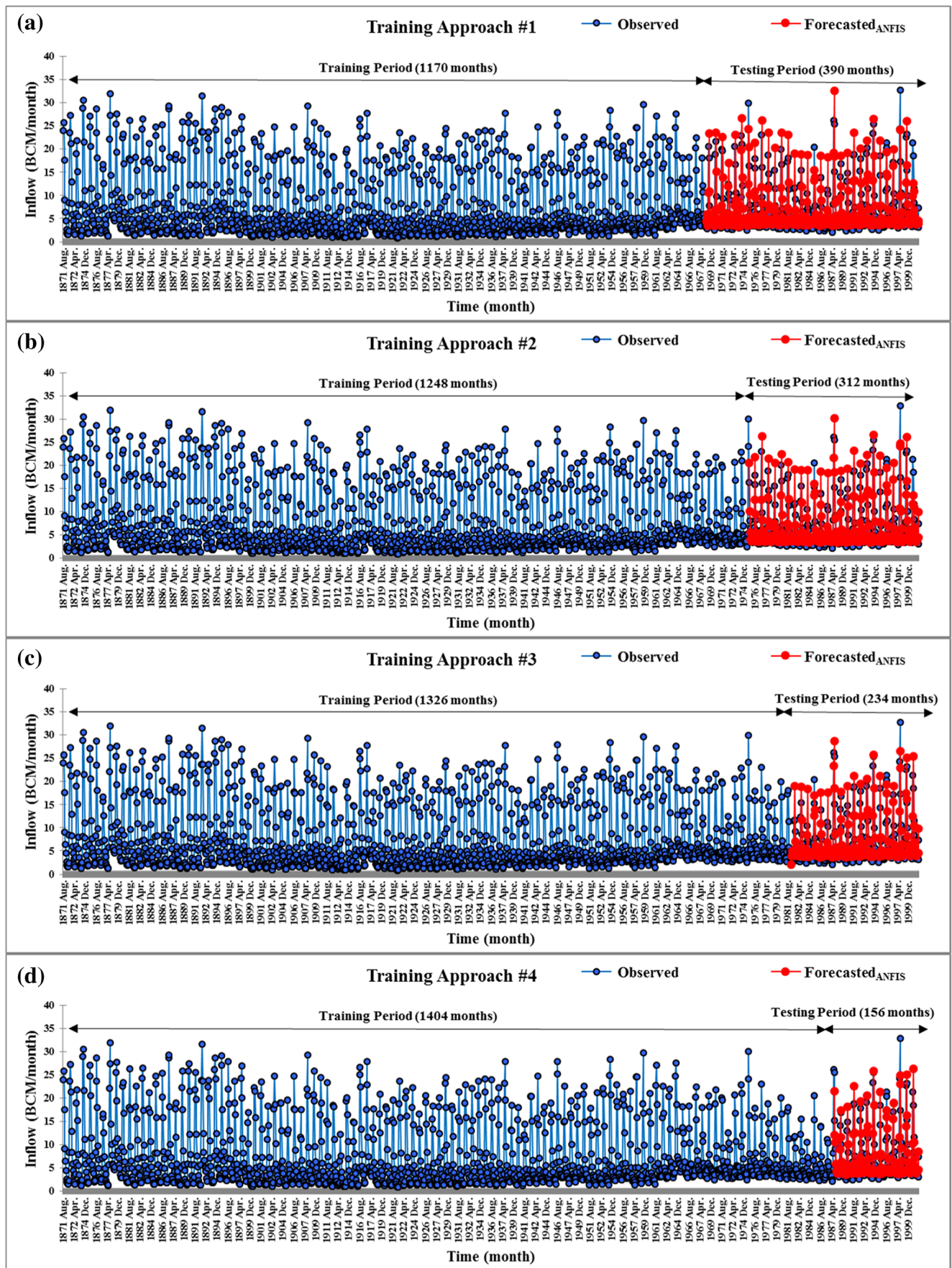
Fig. 5 Scatter plot of model II with four training approaches using ANFIS (left) and modified CANFIS model (right)

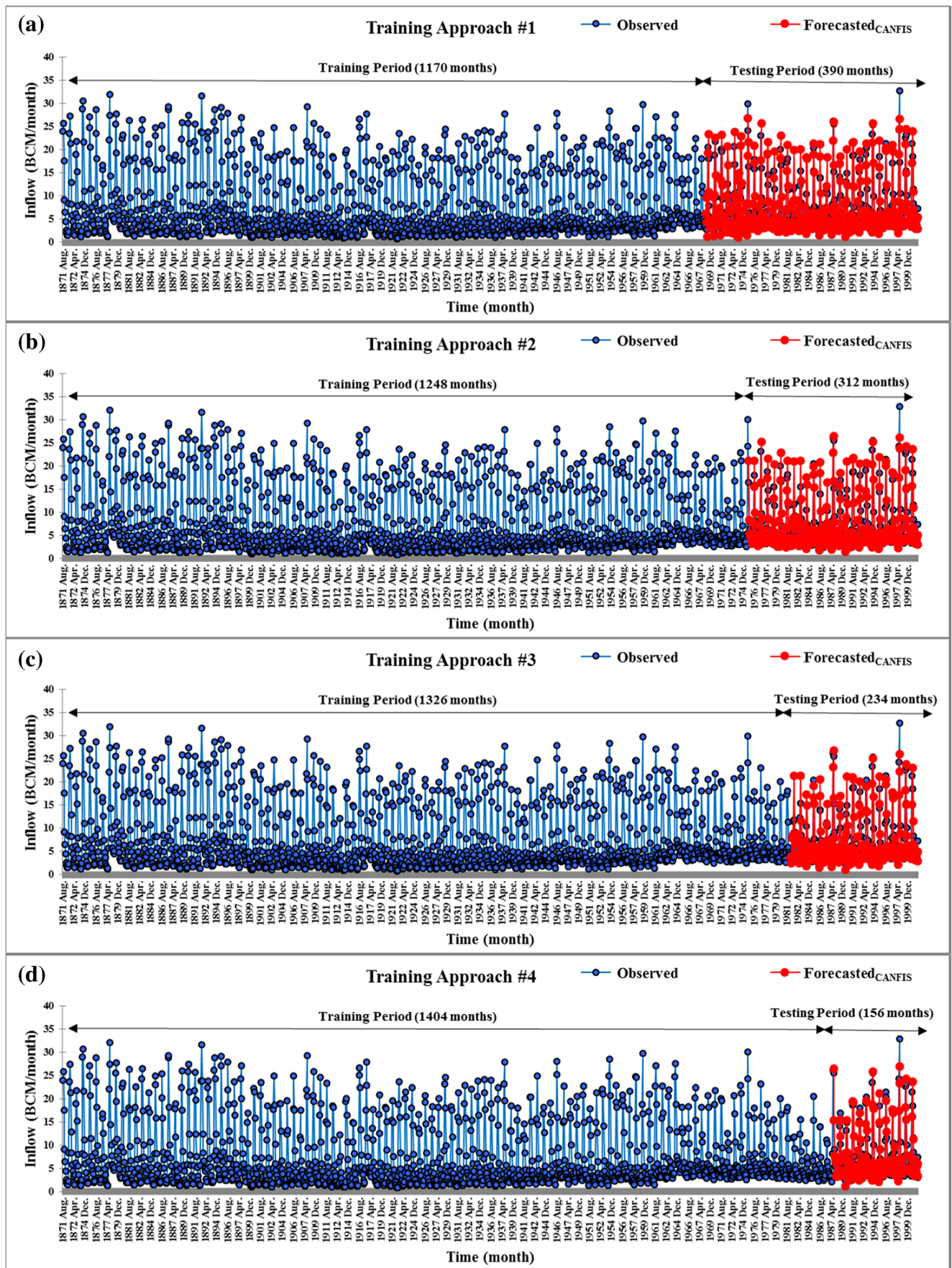


are trained with 90% and tested with 10% of all lagged combination of reservoir inflow data. Evidently, this table shows that in the calibration phase, the modified CANFIS-model II has a higher accuracy compared to the other models. Furthermore, it is noticeable that model II with inputs defined by $I_f = I_a(t-1), I_a(t-2)$ and constructed with the modified CANFIS algorithm generated a minimum root-mean-square error (i.e., $2.51 \text{ BCM month}^{-1}$), mean absolute error (i.e., $1.8 \text{ BCM month}^{-1}$), and correlation coefficient of approximately 0.81. Conversely, the results of the best model attained by the

ANFIS method for training approach #4 were found to yield $\text{RMSE} = 3.07 \text{ BCM month}^{-1}$, $1.84 \text{ BCM month}^{-1}$, and $R^2 = 0.73$. It can be further seen that the minimum relative error was attained using the modified CANFIS model (with train/test scenario denoted as model II) when trained with 90% and tested with 10% of the lagged combinations of inflow

Fig. 6 Observed versus forecasted inflow pattern using the ANFIS method-Model II for testing period; **a** training approach #1; **b** training approach #2; **c** training approach #3; **d** training approach #4





◀ **Fig. 7** Observed versus forecasted inflow pattern using the modified CANFIS method-Model II for testing period; a) Training approach #1; b) Training approach #2; c) Training approach #3; d) Training approach #4

data. There is no doubt that the performance of model II utilizing the modified CANFIS approach is more accurate than those provided by ANFIS approach (Tables 2, 3, 4, and 5).

For an optimized data-driven predictive model, it is necessary to identify the best data-partitioning approach undertaken to construct and evaluate the final model as there is “no rule of thumb” for data division, and in fact, this optimization is likely to depend on the problem under investigation.

The partitioning of the predictor data in this study followed the notion that many researchers have used different data divisions between testing and model training sets that can vary with the problem. Kurup and Dudani (2002) used 63% of data used for training, (Boadu 1997) used 80% of the data used for training, (Pal 2006) used 69% of data used for training, and (Samui and Dixon 2012) used 70% of data for training and 30% for testing purposes to attain optimum results. In this study, the ability of the objective model, modified CANFIS vs. the comparative model ANFIS for reservoir inflow forecasting, was established by a trial of four different training approaches, as listed in Tables 2, 3, 4, and 5.

It is noticeable that in the model verification (i.e., testing) period, the results are quite variable among the four training approaches. Obviously, the worst results are attained by both models that utilized training approach #1, where the error values between the forecasted inflow and the observed inflow data are considerably higher (Table 2) than those found in the other training approaches (Tables 3, 4, and 5). The performance criteria of these models register a notable improvement in the results of models that utilize training approach #2 and #3, but overall, models constructed with training approach 3 (where 85% data are applied for training and 15% for testing remains superior). It implies that the training approach #3 is more suitable for both the ANFIS and the modified CANFIS models used in the current problem of reservoir inflow forecasting which is presumably due to the fact that both model construction methods were able to extract optimal information during the training phase, and consequently, to learn the pattern embedded in the inflow data. Another remarkable observation is that, although the number of data points used in the training phase are greater in models with training approach #4 (i.e., 90% train vs. 10% test) compared to training approach #3 (i.e., 85% train vs. 15% test), the present models provide poor accuracy using training approach #4. This indicates that an increase in training dataset length might lead to counterproductive

results, at least in the present problem under investigation, so a proper selection of the optimal train: test-ratioed dataset is necessary to attain the most accurate simulation model.

The results of the scatter plot diagram for model II utilizing the ANFIS and the modified CANFIS methods with different training approaches are presented in Fig. 5. Based on the values of the correlation between the observed and the forecasted data, the training approach 3 seems to provide the best results with both artificial intelligence methods. Furthermore, the level of correlation between the actual monthly reservoir inflow and the forecasted inflow values generated by the modified CANFIS method is higher those by the ANFIS model. It is remarkable that the higher correlation utilizing the ANFIS attained a value of (i.e., $R^2 = 0.85$) with training approach #3, while the modified CANFIS method attained an R^2 value of 0.94. According to the correlation indicators, there is no doubt that the efficiency of the modified CANFIS model applied for reservoir inflow forecasting is considerably better than the ANFIS model under all training approaches.

From a practical viewpoint, it is useful to check the patterns and level of agreement between the forecasted and the observed inflow values in the testing phase in order to generate a detailed comparison between the present results. Fig. 6 presents the hydrograph of the best model (i.e., Model II) for each training approach using the ANFIS technique. In this figure, there appear to be many peak values of the inflow data shown as a quasi-periodic pattern during the testing phase. On the other hand, the patterns of inflow data forecasted by model II utilizing the modified CANFIS method with four different training approaches is also presented in Fig. 7. It can be seen that both proposed techniques have a high ability to forecast monthly inflow data, but the optimal forecast is attained by using training approach #3. However, it is apparent from Fig. 6c that the ANFIS model is not able to forecast the low and the high inflow values as accurately as the modified CANFIS model. It can be further observed that the ANFIS model is able to detect only the pattern of medium inflow values most of the time, to reflect a lesser degree of the inflow forecasting. However, in accordance with Fig. 7c, we notice that the modified CANFIS method is able to forecast the data values that are more closely matched to the actual low and high inflow values in the testing phase. Based on Figs. 6 and 7, it is evident that the use of training approach #3 for model development and the application of the modified CANFIS model (in contrast to the ANFIS model) led to a significant improvement in the model performance.

Further statistical analysis is performed using the Taylor diagram that shows the statistical patterns (and their relative location from a reference point) of the best models using the ANFIS and the modified CANFIS algorithms under different training approach (see Fig. 8). Note that the Taylor diagram is a model inter-comparison tool where the reference (i.e.,

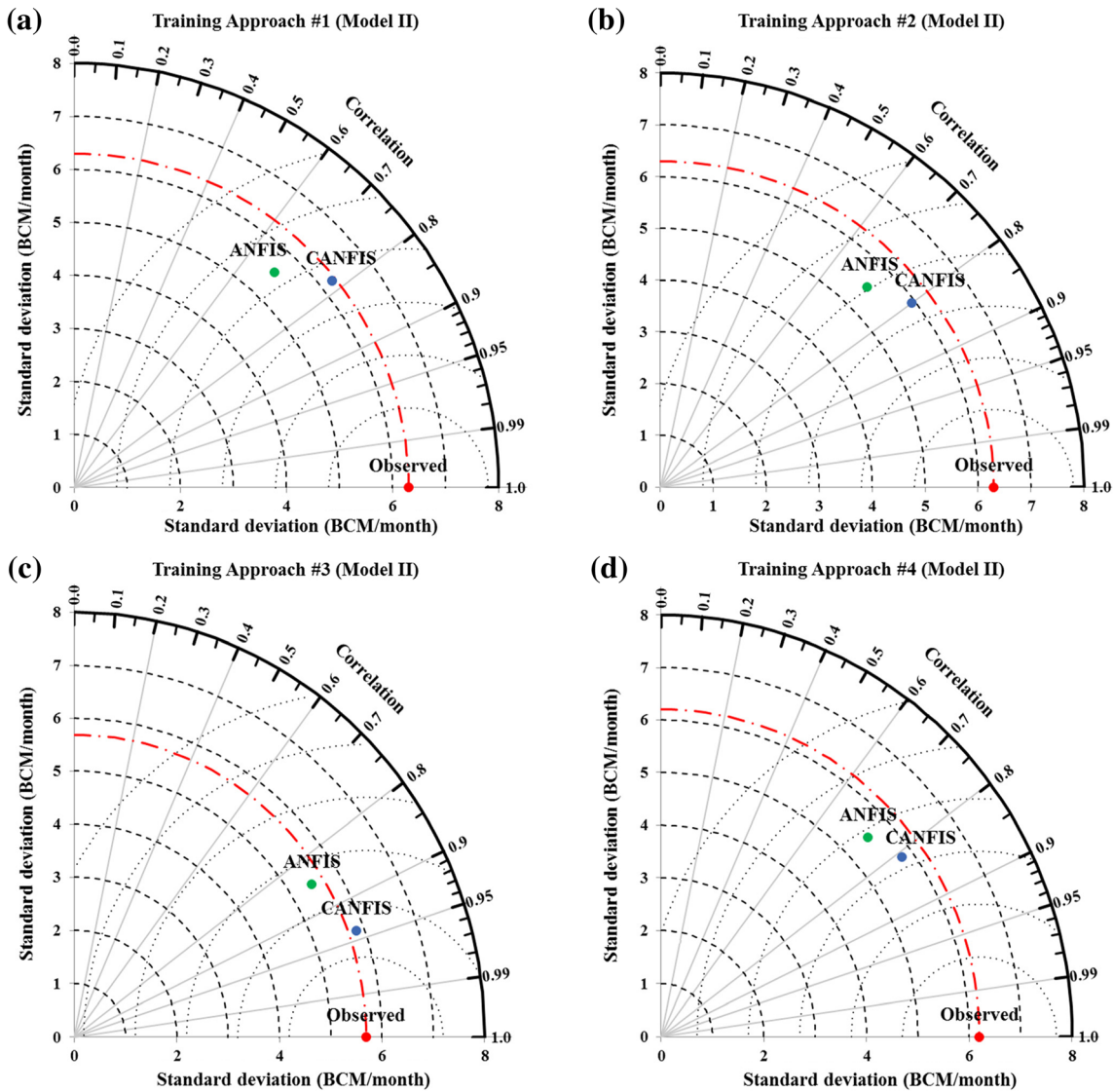


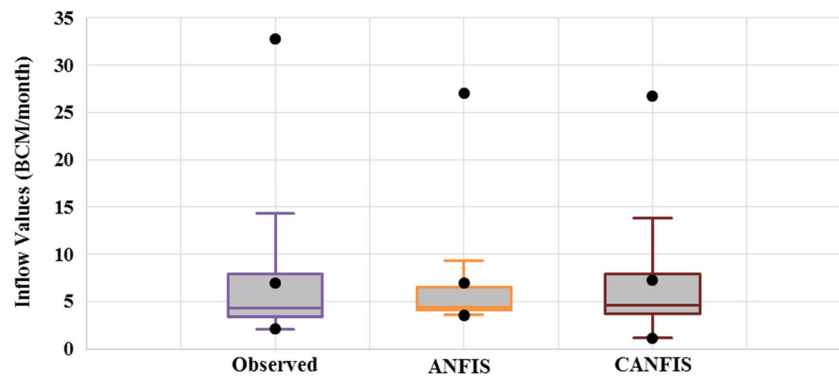
Fig. 8 Taylor diagram for the best results using the ANFIS and the modified CANFIS methods under four training approaches compared to the observed data; **a** training approach #1 (model II); **b** training approach #2 (model II); **c** training approach #3 (model II); **d** training approach #4 (model II)

observed) data are used to benchmark the location of predictive model data based on the root-mean-square difference (RMSD) and the triangle inequality comparisons (Taylor 2001). It is observed that the modified CANFIS model performance is better than the ANFIS model for all training approach since it is located closer to the reference line (RMSD) and observed dataset. In terms of comparing the models, the position of the best model in the diagram using the modified CANFIS technique is attained by training approach #3.

Figure 9 shows a boxplot of the distribution of the forecasted reservoir inflow values for model II by two methods under training approach #3 compared to the actual inflow values in the testing phase. The Whisker-based indicators display the extremity values of the forecasted and the actual inflow properties including their respective quartiles. Note that the lower-end of

the boxplot shows the lower-quartile, I_{25} (representing the 25th percentile data), while the upper-end of the boxplot, I_{75} , represents the 75th percentile value. On the other hand, the second quartile is the median of the reservoir inflow value represented by I_{50} (i.e., 50th percentile). Whiskers are stretched outwards from the top-quartile to the bottom-quartile and the lower and upper quartile (I_{25} , I_{75}) extends the smallest and largest outlier values, respectively. These results indicate that the distribution of the forecasted inflow values by the modified CANFIS model are relatively close to actual inflow data in the testing phase. However, the low inflow values tend to be slightly underestimated when compared with the observed values. In contrast, the ANFIS model provided a different distribution of the forecasted inflow data compared with the modified CANFIS model, which in fact showed a reduction in the accuracy for reservoir inflow forecasts.

Fig. 9 Boxplots of the observed inflow and forecasted values using the ANFIS method-model II and the modified CANFIS method-model II under training approach #3



5 Conclusion

Forecasting monthly reservoir inflow data is a vital requirement for water resources planning and management and everyday decision-tasks implemented in real hydrological practices. In literature, there are several models used to forecast the reservoir inflow, but in this study, the modifications to the original CANFIS method which is extension of ANFIS method were made to purposely enhance its mathematical procedures. This included two major stages: introduction of the fuzzy axons and modular networks, resulting in the construction of the modified CANFIS method. To evaluate their versatility, both the ANFIS and the modified CANFIS methods have been applied to forecast the monthly reservoir inflow for the Aswan High Dam, located in Egypt. The forecasting model has been structured by utilizing different scenario for the input patterns detection that included 1 to 5 previous (or antecedent) statistically significant inflow records used to construct the five different models using each of the two algorithms. In addition, four different training approaches have been examined considering the different splitting of the input-target data for the entire 130 years of monthly records of natural inflow.

In order to examine the performance of the modified CANFIS versus the conventional ANFIS model, four statistical indicators were used to evaluate the performance of each training approach with the various input structures (i.e., five models). These statistical indicators were the %RE, MAE, RMSE, and R^2 values attained by both methods. For further analysis, two different visual data evaluation methods namely, the Taylor diagram and boxplot were considered to examine the correspondence between the actual and forecasted inflow from both methods. Generally, the modified CANFIS model outperformed the ANFIS model for all modeling scenarios (i.e., different lagged input combinations of antecedent inflow data) and different training approaches (different ratios for the partitioning of data into train/test sets). The present results revealed that the forecasted results attained by the modified CANFIS model that utilized input structure for Model II (with $I_f = I_a(t-1), I_a(t-2)$; with training approach #3) are the most

accurate case compared to the other input combinations and training-testing approaches. The results also showed that the modified CANFIS model for the case of model II with training approach #3 attained the minimum %RE (23), MAE (1.4 BCM month⁻¹), RMSE (1.14 BCM month⁻¹), and the highest value of $R^2 = 0.94$. Clearly, the present study supports the preferable use of modified CANFIS model with two antecedent lagged input combinations and training approach #3 for optimal forecasting of reservoir inflow data.

In spite of the good performance of the modified CANFIS model demonstrated in this study, the study does contain potential drawbacks that could address in follow-up works. Since the reservoir data is expected to rely on other climate and environmental inputs (e.g., temperature, humidity, winds, etc.), further investigation is necessary to incorporate these factors, with proper model input structure be carried out into the CANFIS model, to improve the forecasting ability. In fact, this step is an essential for understanding the most influential variable on the accuracy of the model output and for designing real-life expert systems. Further improvement could be introduced in terms of the training approach by considering a set of 12 different models, i.e., “one model for each month” to encapsulate the high variability of the inflow patterns for each month that are expected to be distinct due to different climatic and environmental patterns. In fact, train the CANFIS model with the high variability of the historical inflow pattern is likely to experience difficulty if the modelers fail to mimic the patterns with other variables/factors and to provide a better mapping between the input (i.e., predictor) and the output (i.e., target data). Finally, there is also room to introduce further improvements by the use of ‘add-in’ optimizer algorithms (e.g., Firefly Algorithm; Particle Swarm Optimisation; Genetic Algorithm and Multiverse Optimisation) (e.g., (Wu and Chau 2006; Sedki and Ouazar 2010; Raheli et al. 2017; Ghorbani et al. 2017) to increase the learning ability of the modified CANFIS model. In this context, the back-propagation procedure within the modified CANFIS method could overcome some of the weakness in the fine-tuning of its internal parameters if such methods are used to construct hybrid CANFIS models.

Taken together, future enhancements made to the proposed CANFIS model could focus on the three different phases of the modeling procedure. First, the utilization of a large pool of environmental variables (e.g., temperature, humidity, solar radiation and wind data) and the implementation of the model's input selection through a suitable procedure for identifying data pattern considering the most influence variables is important to achieve a better forecasting accuracy. Second, for the training approaches, modelers must investigate more efficient training methods that provide the model the ability to mimic all the possible inflow patterns. Third, within the CANFIS method itself, modelers must utilize more effective optimization algorithms to search for optimal CANFIS parameters. These approaches are likely to lead to a more robust CANFIS model that is responsive to the attributes within the input data, and hence, yield more accurate performance to increase its ability to be implemented in real-time reservoir inflow simulation systems.

Acknowledgements This work was supported by a research grant from the Ministry of Higher Education, FRGS/1/2016/STG06/UKM/02/1, and a grant from the University of Malaya's BKP Grant BK023-2015. The authors thank the Nile Water Authority (NWA) and Aswan High Dam Authority (AHDA), Ministry of Water Resources and Irrigation, Egypt for providing the natural inflow data sets used in this study. Dr. R C Deo also acknowledges the University of Southern Queensland s-ADOSP research grant (Sept – Nov 2017). The authors would like to thank the research facilities support received from DIP-2015-012 project funded by University Kebangsaan Malaysia for the sixth author. The authors thank all reviewers and the Editor-In-Chief for their insightful comments that has improved the quality of the final manuscript.

Compliance with ethical standards

Conflicts of interest The authors declare that they have no conflict of interest.

References

- Alecsandru C, Ishak S (2004) Hybrid model-based and memory-based traffic prediction system. *Transp Res Rec J Transp Res Board* 1879: 59–70. <https://doi.org/10.3141/1879-08>
- Allawi MF, El-Shafie A (2016) Utilizing RBF-NN and ANFIS methods for multi-lead ahead prediction model of evaporation from reservoir. *Water Resour Manag* 30:4773–4788. <https://doi.org/10.1007/s11269-016-1452-1>
- Aytek A (2009) Co-active neurofuzzy inference system for evapotranspiration modeling. *Soft Comput* 13:691–700. <https://doi.org/10.1007/s00500-008-0342-8>
- Bai Y, Wang P, Xie J et al (2015) Additive model for monthly reservoir inflow forecast. *J Hydrol Eng* 20:4014079. [https://doi.org/10.1061/\(ASCE\)HE.1943-5584.0001101](https://doi.org/10.1061/(ASCE)HE.1943-5584.0001101)
- Boadu FK (1997) Rock properties and seismic attenuation: neural network analysis. *Pure Appl Geophys* 149:507–524. <https://doi.org/10.1007/s000240050038>
- Box GEP, Jenkins GM (1970) *Time series analysis: forecasting and control*, Rev. ed. Holden-Day, San Francisco
- Chang F-J, Chang Y-T (2006) Adaptive neuro-fuzzy inference system for prediction of water level in reservoir. *Adv Water Resour* 29:1–10. <https://doi.org/10.1016/j.advwatres.2005.04.015>
- Chen C-H, Yao T-K, Kuo C-M, Chen C-Y (2013) RETRACTED: evolutionary design of constructive multilayer feedforward neural network. *J Vib Control* 19:2413–2420. <https://doi.org/10.1177/1077546312456726>
- Chiew FHS, Piechota TC, Dracup JA, McMahon TA (1998) El Nino/southern oscillation and Australian rainfall, streamflow and drought: links and potential for forecasting. *J Hydrol* 204:138–149. [https://doi.org/10.1016/S0022-1694\(97\)00121-2](https://doi.org/10.1016/S0022-1694(97)00121-2)
- Coulibaly P, Ancil F, Bobée B (2000) Daily reservoir inflow forecasting using artificial neural networks with stopped training approach. *J Hydrol* 230:244–257. [https://doi.org/10.1016/S0022-1694\(00\)00214-6](https://doi.org/10.1016/S0022-1694(00)00214-6)
- Danandeh Mehr A, Kahya E, Olyaeie E (2013) Streamflow prediction using linear genetic programming in comparison with a neuro-wavelet technique. *J Hydrol* 505:240–249. <https://doi.org/10.1016/j.jhydrol.2013.10.003>
- Deo RC, Şahin M (2016) An extreme learning machine model for the simulation of monthly mean streamflow water level in eastern Queensland. *Environ Monit Assess* 188:90. <https://doi.org/10.1007/s10661-016-5094-9>
- El-Shafie A, Noureldin A (2011) Generalized versus non-generalized neural network model for multi-lead inflow forecasting at Aswan high dam. *Hydrol Earth Syst Sci* 15:841–858. <https://doi.org/10.5194/hess-15-841-2011>
- El-Shafie A, Taha MR, Noureldin A (2007) A neuro-fuzzy model for inflow forecasting of the Nile river at Aswan high dam. *Water Resour Manag* 21:533–556. <https://doi.org/10.1007/s11269-006-9027-1>
- El-Shafie A, Abdin AE, Noureldin A, Taha MR (2009) Enhancing inflow forecasting model at Aswan high dam utilizing radial basis neural network and upstream monitoring stations measurements. *Water Resour Manag* 23:2289–2315. <https://doi.org/10.1007/s11269-008-9382-1>
- Ghorbani MA, Deo RC, Yaseen ZM, et al (2017) Pan evaporation prediction using a hybrid multilayer perceptron-firefly algorithm (MLP-FFA) model: case study in North Iran. *Theor Appl Climatol* 1–13. <https://doi.org/10.1007/s00704-017-2244-0>
- Gillberg C, Wahlström J (2008) Chromosome abnormalities in infantile autism and other childhood psychoses: a population study of 66 cases. *Dev Med Child Neurol* 27:293–304. <https://doi.org/10.1111/j.1469-8749.1985.tb04539.x>
- Guo Z, Wu J, Lu H, Wang J (2011) A case study on a hybrid wind speed forecasting method using BP neural network. *Knowl-Based Syst* 24: 1048–1056. <https://doi.org/10.1016/j.knosys.2011.04.019>
- Jang J-SR, Sun C-T, Mizutani E (1997) *Neuro-fuzzy and soft computing, a computational approach to learning and machine intelligence*. Prentice Hall, New Jersey
- Jothiprakash V, Magar R (2009) Soft computing tools in rainfall-runoff modeling. *ISH J Hydraul Eng* 15:84–96. <https://doi.org/10.1080/09715010.2009.10514970>
- Jothiprakash V, Magar RB (2012) Multi-time-step ahead daily and hourly intermittent reservoir inflow prediction by artificial intelligent techniques using lumped and distributed data. *J Hydrol* 450–451:293–307. <https://doi.org/10.1016/j.jhydrol.2012.04.045>
- Ju Q, Yu Z, Hao Z et al (2009) Division-based rainfall-runoff simulations with BP neural networks and Xinanjiang model. *Neurocomputing* 72:2873–2883. <https://doi.org/10.1016/j.neucom.2008.12.032>
- Keshtegar B, Allawi MF, Afan HA, El-Shafie A (2016) Optimized River stream-flow forecasting model utilizing high-order response surface method. *Water Resour Manag* 30:3899–3914. <https://doi.org/10.1007/s11269-016-1397-4>
- Kisi O, Shiri J, Nikoofar B (2012) Forecasting daily lake levels using artificial intelligence approaches. *Comput Geosci* 41:169–180. <https://doi.org/10.1016/j.cageo.2011.08.027>

- Kurup PU, Dudani NK (2002) Neural networks for profiling stress history of clays from PCPT data. *J Geotech Geoenviron Eng* 128:569–579. [https://doi.org/10.1061/\(ASCE\)1090-0241\(2002\)128:7\(569\)](https://doi.org/10.1061/(ASCE)1090-0241(2002)128:7(569))
- Lin G-F, Wu M-C (2011) An RBF network with a two-step learning algorithm for developing a reservoir inflow forecasting model. *J Hydrol* 405:439–450. <https://doi.org/10.1016/j.jhydrol.2011.05.042>
- Lohani AK, Goel NK, Bhatia KKS (2007) Deriving stage–discharge–sediment concentration relationships using fuzzy logic. *Hydrol Sci J* 52:793–807. <https://doi.org/10.1623/hysj.52.4.793>
- Lohani AK, Kumar R, Singh RD (2012) Hydrological time series modeling: a comparison between adaptive neuro-fuzzy, neural network and autoregressive techniques. *J Hydrol* 442–443:23–35. <https://doi.org/10.1016/j.jhydrol.2012.03.031>
- Memarian H, Balasundram SK, Tajbakhsh M (2013) An expert integrative approach for sediment load simulation in a tropical watershed. *J Integr Environ Sci* 10:161–178. <https://doi.org/10.1080/1943815X.2013.852591>
- Pal M (2006) Support vector machines-based modelling of seismic liquefaction potential. *Int J Numer Anal Methods Geomech* 30:983–996. <https://doi.org/10.1002/nag.509>
- Raheli B, Aalami MT, El-Shafie A et al (2017) Uncertainty assessment of the multilayer perceptron (MLP) neural network model with implementation of the novel hybrid MLP-FFA method for prediction of biochemical oxygen demand and dissolved oxygen: a case study of Langat River. *Environ Earth Sci* 76:503. <https://doi.org/10.1007/s12665-017-6842-z>
- Samui P, Dixon B (2012) Application of support vector machine and relevance vector machine to determine evaporative losses in reservoirs. *Hydrol Process* 26:1361–1369. <https://doi.org/10.1002/hyp.8278>
- Sedki A, Ouazar D (2010) Hybrid particle swarm and neural network approach for streamflow forecasting. *Math Model Nat Phenom* 5:132–138. <https://doi.org/10.1051/mmnp/20105722>
- Shoab M, Shamseldin AY, Melville BW, Khan MM (2016) Hybrid wavelet neuro-fuzzy approach for rainfall-runoff modeling. *J Comput Civ Eng* 30:4014125. [https://doi.org/10.1061/\(ASCE\)CP.1943-5487.0000457](https://doi.org/10.1061/(ASCE)CP.1943-5487.0000457)
- Smith J, Eli RN (1995) Neural-network models of rainfall-runoff process. *J Water Resour Plan Manag* 121:499–508. [https://doi.org/10.1061/\(ASCE\)0733-9496\(1995\)121:6\(499\)](https://doi.org/10.1061/(ASCE)0733-9496(1995)121:6(499))
- Tabari H, Hosseinzadeh Talaei P, Abghari H (2012) Utility of coactive neuro-fuzzy inference system for pan evaporation modeling in comparison with multilayer perceptron. *Meteorol Atmos Phys* 116:147–154. <https://doi.org/10.1007/s00703-012-0184-x>
- Taylor KE (2001) Summarizing multiple aspects of model performance in a single diagram. *J Geophys Res Atmos* 106:7183–7192. <https://doi.org/10.1029/2000JD900719>
- Valipour M, Banihabib ME, Behbahani SMR (2013) Comparison of the ARMA, ARIMA, and the autoregressive artificial neural network models in forecasting the monthly inflow of Dez dam reservoir. *J Hydrol* 476:433–441. <https://doi.org/10.1016/j.jhydrol.2012.11.017>
- Valizadeh N, Mirzaei M, Allawi MF et al (2017) Artificial intelligence and geo-statistical models for stream-flow forecasting in ungauged stations: state of the art. *Nat Hazards* 86:1377–1392. <https://doi.org/10.1007/s11069-017-2740-7>
- Whigham PA, Crapper PF (2001) Modelling rainfall-runoff using genetic programming. *Math Comput Model* 33:707–721. [https://doi.org/10.1016/S0895-7177\(00\)00274-0](https://doi.org/10.1016/S0895-7177(00)00274-0)
- Wu CL, Chau KW (2006) A flood forecasting neural network model with genetic algorithm. *Int J Environ Pollut* 28:261. <https://doi.org/10.1504/IJEP.2006.011211>
- Yan H, Zou Z, Wang H (2010) Adaptive neuro fuzzy inference system for classification of water quality status. *J Environ Sci* 22:1891–1896. [https://doi.org/10.1016/S1001-0742\(09\)60335-1](https://doi.org/10.1016/S1001-0742(09)60335-1)