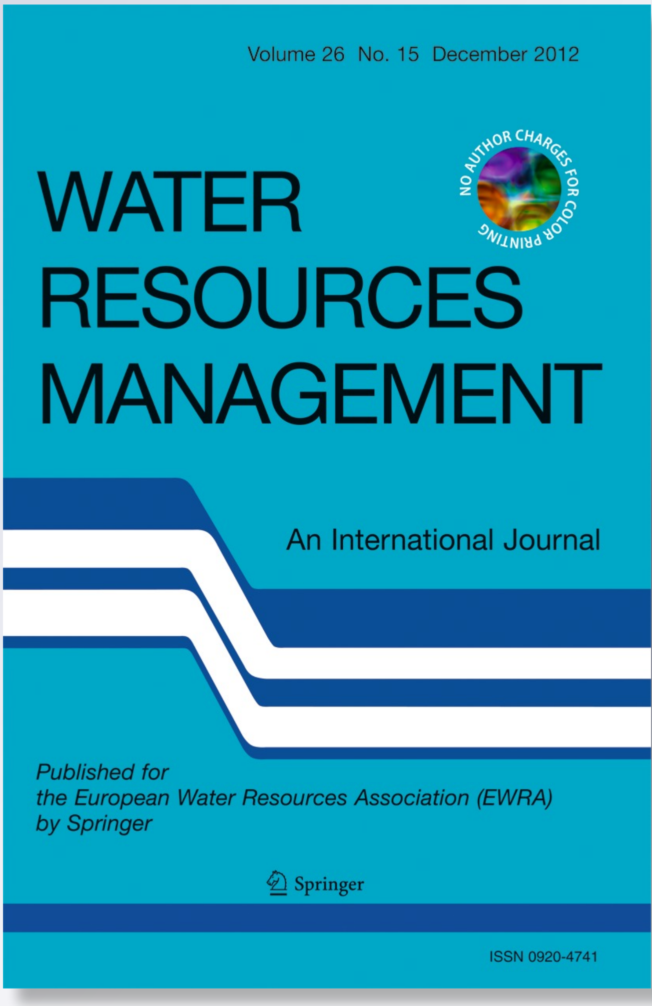
Estimation of Daily Pan Evaporation Using Two Different Adaptive Neuro-Fuzzy Computing Techniques

Hadi Sanikhani, Ozgur Kisi, Mohammad Reza Nikpour & Yagob Dinpashoh

Water Resources Management
An International Journal - Published for the European Water Resources Association (EWRA)

ISSN 0920-4741
Volume 26
Number 15

Water Resour Manage (2012)
26:4347-4365
DOI 10.1007/s11269-012-0148-4



Springer

Water Resour Manage (2012) 26:4347–4365
DOI 10.1007/s11269-012-0148-4

Estimation of Daily Pan Evaporation Using Two Different Adaptive Neuro-Fuzzy Computing Techniques

Hadi Sanikhani · Ozgur Kisi ·
Mohammad Reza Nikpour · Yagob Dinpashoh

Received: 13 March 2012 / Accepted: 17 September 2012 /
Published online: 30 September 2012

Abstract This paper investigates the ability of two different adaptive neuro-fuzzy inference systems (ANFIS) including grid partitioning (GP) and subtractive clustering (SC), in modeling daily pan evaporation (E_{pan}). The daily climatic variables, air temperature, wind speed, solar radiation and relative humidity of two automated weather stations, San Francisco and San Diego, in California State are used for pan evaporation estimation. The results of ANFIS-GP and ANFIS-SC models are compared with multivariate non-linear regression (MNLR), artificial neural network (ANN), Stephens-Stewart (SS) and Penman models. Determination coefficient (R^2), root mean square error (RMSE) and mean absolute relative error (MARE) are used to evaluate the performance of the applied models. Comparison of results indicates that both ANFIS-GP and ANFIS-SC are superior to the MNLR, ANN, SS and Penman in modeling E_{pan}. The results also show that the difference between the performances of ANFIS-GP and ANFIS-SC is not significant in evaporation estimation. It is found that two different ANFIS models could be employed successfully in modeling evaporation from available climatic data.

Keywords Adaptive neuro-fuzzy inference system · Grid partitioning · Subtractive clustering · Evaporation modeling

1 Introduction

Evaporation takes place whenever there is a vapour pressure deficit between a water surface and the overlying atmosphere and sufficient energy is available. The most common and

H. Sanikhani
Young Researcher Club, Tabriz Branch, Islamic Azad University, Tabriz, Iran

O. Kisi (✉)
Architectural and Engineering Faculty, Civil Engineering Department, Canik Basari University, Samsun, Turkey
e-mail: kisi@erciyes.edu.tr

M. R. Nikpour · Y. Dinpashoh
Water Engineering Department, Faculty of Agriculture, University of Tabriz, Tabriz, Iran

important meteorological parameters affecting the rate of evaporation are solar radiation, air temperature, relative humidity, vapor pressure deficit and wind speed. Evaporation losses should be considered in design of various water resources and irrigation systems (McCuen 1998). Evaporation is one of the less understood components of hydrologic cycle (Jackson 1985).

Daily pan evaporation (E_{pan}) is an important parameter in water budgeting estimations and in modeling crop water response to different weather conditions. It has been widely used as an index of lake and reservoir evaporation, potential or reference crop evapotranspiration and irrigation scheduling (Snyder 1993).

Some of researchers used climatic variables to predict E_{pan} values (Reis and Dias 1998; Coulomb et al. 2001; Gaven and Agnew 2004; Rahimikhoob 2009; Trajkovic 2010; Trajkovic and Kolakovic 2010; Sabziparvar et al. 2010). Because evaporation is a nonlinear, stochastic and complex process, it is difficult to derive an accurate formula to represent all the physical processes involved (Moghadamnia et al. 2009). In recent years, application of artificial intelligence techniques, such as Artificial Neural Networks (ANNs), Adaptive Neuro-Fuzzy Inference System (ANFIS) and Genetic Programming (GP) in estimation of hydrological parameters have been widely considered by most of the researchers (Trajkovic et al. 2000; Trajkovic 2009; Khu et al. 2001; Kisi 2006a, b, 2007a, 2009a, b; El-Shafie et al. 2007; Rahimikhoob 2008; Aytek 2009; Chu and Chang 2009; Guldal and Tongal 2010; Traore et al. 2010; Traore and Guven 2011; Goyal and Ojha 2011; Cobaner 2011). Sudheer et al. (2002) used an ANN for estimating E_{pan} and found that the ANN performed better than the other conventional approach. Keskin et al. (2004) used fuzzy approach for modeling daily pan evaporation of western Turkey. Keskin and Terzi (2006) developed multi-layer perceptron (MLP) models to estimate daily E_{pan} and found that the ANN model showed a considerably better performance over the conventional method. Tan et al. (2007) modeled hourly and daily open water evaporation rates by using ANN technique. Kisi (2009a) used three different ANN techniques, namely the MLP, radial basis neural network (RBNN) and generalized regression neural network (GRNN), in daily E_{pan} modeling and found that the MLP and RBNN performed significantly better than the GRNN. Piri et al. (2009) used ANN model for estimating daily E_{pan} in a hot and dry climate. Moghadamnia et al. (2009) explored evaporation estimation methods based on ANN and ANFIS techniques. It has been found that the ANN and ANFIS techniques have much better performances than the empirical formulas. Keskin et al. (2009) used the fuzzy sets and ANFIS for modeling daily E_{pan} and found that ANFIS approach could be employed more successfully in modeling evaporation process than fuzzy sets. Dogan et al. (2010) used ANFIS approach for estimating daily pan evaporations from the reservoir of Yuvacik Dam, Turkey. Tabari et al. (2010) investigated the ability of ANN and multivariate non-linear regression techniques for modeling daily pan evaporation and found that the ANN performed better than the non-linear regression. Guven and Kisi (2011) modeled daily pan evaporations using linear genetic programming and ANN models. All above studies were employed grid partitioning (GP) method in ANFIS modeling. Cobaner (2011) used two different ANFIS models for estimation of evapotraspiration and found that the ANFIS with subtractive clustering (SC) model yields plausible accuracy with fewer amounts of computations as compared to the ANFIS-GP and ANN models.

In this study, MNLR and two different ANFIS methods i.e. ANFIS-GP and ANFIS-SC are applied for estimating E_{pan} by using climatic variables. To the best knowledge of the authors, no study has been carried out for applying the ANFIS-SC approach to estimate daily E_{pan}. Furthermore, in the past studies, no criteria were used for selecting input variables to apply in intelligent approaches. In this research, for selecting input variables to the MNLR model, the stepwise regression method was used. After determination of different suitable combinations as input variables, same inputs are used in artificial intelligent methods.

The paper is set as follows. Section 2 presents a description of the methods used in the study. Section 3 provides the information about the used data, methodological properties and statistical indices. The applicability of the models on evaporation estimation and the results are presented in Section 4. Conclusions are presented in Section 5.

2 Material and Methods

First, MNLR was applied for determining different input combinations including important climatic variables described before. Logarithmic transferred values of all climatic variables and E_{pan} were calculated. Then, the multiple linear regression using stepwise method was used to select important variables at 5 % significance level. For applying linear regression, SPSS software was used. After determining various input combinations, two different ANFIS models were used to model E_{pan}. Two different program codes, including fuzzy logic toolbox, were written in MATLAB for this purpose. Various ANFIS structures were tried using these codes and the appropriate model structures were determined for each input combination. Finally, two different ANFIS estimates were compared with those of the MNLR, ANN, SS and Penman in modeling daily pan evaporation of two stations.

2.1 Multivariate Non-Linear Regression

Some statistical methods, such as regression models, are known as the best tools for investigating any relation between small sample sizes of dependent and independent variables (Razi and Athappilly 2005). The MNLR is a method used to model the non-linear relationship between a dependent variable and one or more independent variables. It is based on least squares. The model is fit such that the sum-of-squares of differences of observed and predicted values is minimized. Estimation of E_{pan} could be considered by models that can address the inherent non-linearities in evaporation process. Tabari et al. (2010) showed that multivariate logarithmic regression is able to estimate pan evaporation at desirable level of accuracy. Therefore, the logarithmic model was applied for estimation of pan evaporation here. First, the transferred logarithmic values of E_{pan} and climatic variables were prepared and then multiple linear regression based on stepwise method was conducted to find important variables which can describe E_{pan}. In this matter, SPSS software version 15.0 was used. After determining the significant input variables, different combinations of them were used to predict E_{pan} using the two different ANFIS models (ANFIS-GP and ANFIS-SC).

2.2 Stepwise Regression Method

Variable selection methods which identify good subset models have been developed. These methods are referred to as stepwise regression methods. The subset models are identified sequentially by adding or deleting, depending on the method (forward selection or backward elimination), the one variable that has the greatest impact on the residual sum of squares. Stepwise selection of variables requires more computing than forward or backward selection but has an advantage in terms of the numbers of potential subset models checked before the model for each subset size is decided. It is reasonable to expect stepwise selection to have a greater chance of choosing the best subsets in the sample data, but selection of the best subset for each subset size is not guaranteed (Rawlings 1988).

2.3 Adaptive Neuro-Fuzzy Inference System (ANFIS)

Jang (1993) presented a learning procedure for the fuzzy inference system (FIS) that uses a NN learning algorithm for constructing a set of fuzzy if-then rules with appropriate membership functions (MFs) from specified input–output pairs. Figure 1 shows a basic structure of ANFIS.

An ANFIS is a network structure consisting of a number of nodes connected through directional links. Each node is characterized by a node function including fixed or adjustable parameters. Training phase of a NN is a process to determine parameter values to sufficiently fit the training data. The basic learning rule is the well-known back-propagation method which seeks to minimize some measure of error between network's outputs and desired outputs (Drake 2000).

Depending on the types of inference operations upon 'if-then rules', most FISs can be classified into three types; Mamdani's system (Mamdani and Assilian 1975), Sugeno's system (Takagi and Sugeno 1985) and Tsukamoto's system (Tsukamoto 1979). Mamdani's system is the most commonly used; meanwhile, Sugeno's system is more compact and computationally efficient. The output of Sugeno's system is crisp and it has a mathematically intractable defuzzification operation. It is by far the most popular candidate for sample-data based fuzzy modeling and it lends itself to the use of adaptive techniques (Takagi and Sugeno 1985).

To build-up a fuzzy system, firstly, the linguistic variables should have been provided in addition to numerical variables. Then, the system requires If/Then fuzzy rules to qualify simple relationships between fuzzy variables. A typical rule set with two fuzzy If/Then rules in first-order Sugeno's system, can be shown as:

$$\textit{Rule } 1: \textit{ If } x \textit{ is } A1 \textit{ and } y \textit{ is } B1; \textit{ then } f_1 = p_1 x + q_1 y + r_1 \tag{1}$$

$$\textit{Rule } 2: \textit{ If } x \textit{ is } A2 \textit{ and } y \textit{ is } B2; \textit{ then } f_2 = p_2 x + q_2 y + r_2 \tag{2}$$

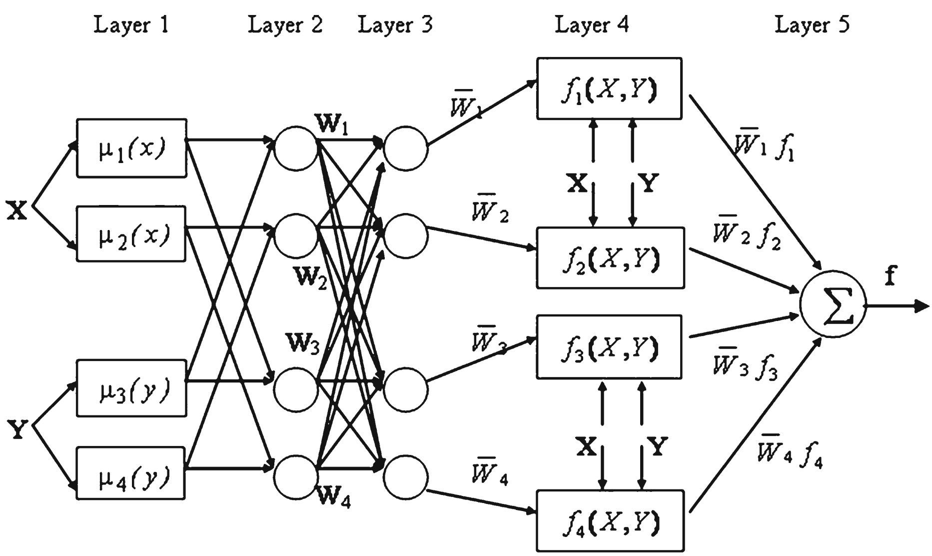


Fig. 1 Basic structure of the ANFIS (see Kisi et al. 2009)

Where x and y refer to inputs and f_2 represents the output variable, respectively. The A and B terms denote the linguistic terms of the precondition part with MF. The 'If' part of the rule 'x is A' is called the premise, while the 'Then' part of the rule is called the consequent. The p, q, r indicate the consequent parameters (Sayed et al. 2003).

According to Fig. 1, ANFIS consists of five layers as follow:

Layer 1 Every node i in this layer is an adaptive node, including MFs generally described by generalized bell functions, e.g.

$$f_{1,i} = \mu_1(X) = \frac{1}{1 + |(X - c_1)/a_1|^{2b_1}} \tag{3}$$

where X is input to the node and a_1, b_1 and c_1 are adaptable variables known as premise parameters. The membership values of the premise part constitute the outputs of this layer.

Layer 2 This layer composes of the nodes which multiply incoming signals and sending the product out. This product represents the firing strength of a rule. For example in Fig. 1

$$f_{2,1} = w_1 = \mu_1(x)\mu_3(y) \tag{4}$$

Layer 3 In this layer, the nodes calculate the ratio of the ith rule's firing strength to the sum of all rules firing strengths

$$f_{3,1} = \overline{w}_1 = \frac{w_1}{w_1 + w_2 + w_3 + w_4} \tag{5}$$

Layer 4 The nodes of this layer are adaptive with node functions

$$f_{4,1} = \overline{w}_1 f_1 = \overline{w}_1(p_1 x + q_1 y + r_1) \tag{6}$$

where $\overline{w}_1$ is the output of Layer 3 and $\{p_i, q_i, r_i\}$ is the parameter set. This layer's parameters are referred to as consequent parameters.

Layer 5 Single fixed node in this layer computes the final output as the summation of all incoming signals

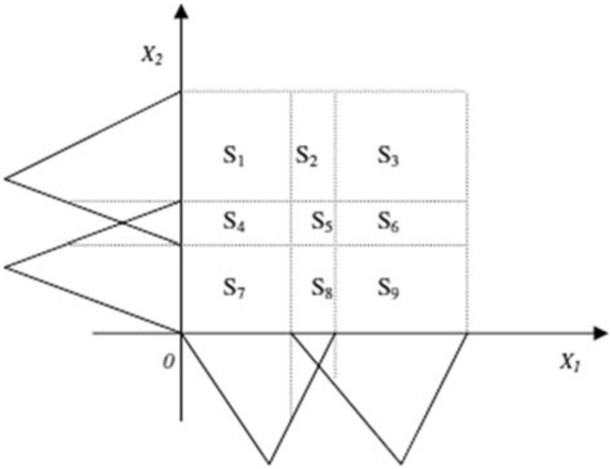
$$f = \sum_{i=1}^{n} \overline{w}_i f_i \tag{7}$$

A detailed description of ANFIS can be found in Jang (1993).

2.3.1 Grid Partitioning

By combining ANFIS and grid partition, ANFIS Grid Partition (ANFIS-GP) model was obtained. Grid partition divides the input space into rectangular subspaces using a number of local fuzzy regions by axis-paralleled partition based on predefined number of MFs and their types in each dimension. Figure 2 shows a schematic description of this model. For calculating fuzzy sets and parameters, least square method according to the partition and MF type is used.

Fig. 2 Grid partition of an input domain with two input variables and two MFs for each input variable (see Wei et al. 2007)



During constructing the fuzzy rules consequent parameters in the linear output MF are set to zero. Therefore, by using ANFIS, parameters are identified and refined. GP and its combination with ANFIS are illustrated by Abonyi et al. (1999) in details. By increasing the number of input variables, the number of fuzzy rules is exponentially increased. For instance, if there are n input variables and m MFs for each input variable for the problem, the total number of fuzzy rules equals m^n (Wei et al. 2007). For application of grid partition, the number of input variable must be small and less than 6 (http://www.cs.nthu.edu.tw/~jang/an.sfaq.htm). In current study, E_{pan} was estimated by using four input variables and therefore applying ANFIS-GP model in this paper is reasonable.

2.3.2 Subtractive Clustering

By combining ANFIS and subtractive clustering, ANFIS subtractive clustering (ANFIS-SC) model was obtained. This model is an extension of mountain clustering method proposed by Yager and Filev (1994) in which each data point (not a grid point) is considered as a center for potential cluster center (Chiu 1994). Using this method, the number of effective "grid points" to be evaluated equals to the number of data points, independent of the dimension of the problem. Another advantage of this method is that it eliminates the need to specify a grid resolution, in which tradeoffs between accuracy and computational complexity must be considered. The subtractive clustering method also extends the criterion of the mountain method for accepting and rejecting cluster centers. The subtractive clustering method works as follows:

By considering a collection of n data points $\{x_1,x_2,\ldots,x_n\}$ in an M dimensional space, without loss of generality and assuming that the data points have been normalized in each dimension so that they are bounded by a unit hypercube. Each data point is considered as a potential cluster center and a measure of the potential of data point X_i defined as:

$$P_i = \sum_{j=1}^{n} e^{-\alpha \left\| X_i - X_j \right\|^2} \tag{8}$$

where $\alpha = 4/r_a^2$, $\left\| X_i - X_j \right\|^2$ indicates the Euclidean distance, and r_a is a positive constant. So, the measure of the potential for a data point is a function of its distances to all other data points. A data point with many neighboring data points will have a high potential value. The constant r_a is the radius defining a neighborhood; data points outside this radius have little influence on the potential. After the potential of every data point has been computed, the data point with the highest potential is selected as the first cluster center. If X_1^* is the location of the first cluster center and $P*_1$ is its potential value, potential of each data point x_i represented by the following formula:

$$P_i \Leftarrow P_i - P_1^* e^{-\beta \left\| X_i - X_1^* \right\|^2} \tag{9}$$

where $\beta = 4/r_b^2$ and r_b is a positive constant. So, an amount of potential from each data point as a function of its distance from the first cluster center is subtracted. The data points near the first cluster center will have greatly reduced potential, and therefore will unlikely be selected as the next cluster center. The constant r_b is effectively the radius defining the neighborhood which will have measurable reductions in potential. To avoid obtaining closely spaced cluster centers, r_b is set somewhat greater than r_a and a good value is r_b equals to 1.25 r_a.

When the potential of all data points has been revised according to Eq. (9), the data point with the highest remaining potential as the second cluster center is selected. Then further reduce in the potential of each data point according to their distance to the second cluster center is done. In general, after the *k*'th cluster center has been obtained, the potential of each data point is given by the formula:

$$P_i \Leftarrow P_i - P_k^* e^{-\beta \left\| X_i - X_k^* \right\|^2} \tag{10}$$

where X_k^* is the location of the *k*'th cluster center and P_k^* is its potential value.

The process of acquiring new cluster center and revising potentials repeats until the remaining potential of all data points falls below some fraction of the potential of the first cluster center P_1^*. Other criteria are available for accepting and rejecting cluster centers that help avoid marginal cluster centers (Chiu 1997).

The influential radius is critical for determining the number of clusters. Selecting a smaller radius results to many smaller clusters in the data space and more rules are required and vice versa. Therefore it is substantial to select proper influential radius for clustering the data space. The number of fuzzy rules and premises fuzzy MF is then determined. Finally, the linear least squares estimate is used to determine the consequent in the output MF, resulting in a valid FIS. Subtractive clustering and ANFIS has been used in different fields of engineering (Wei et al. 2007 and Cobaner 2011).

3 Data Used

The data-driven modeling approaches such as ANFIS are based on high quality data (Wang et al. 2007; 2009). Therefore, the daily climatic data of two automated weather stations, San Francisco Station (latitude 37° 37′ N, longitude 122° 23′ W) and San Diego Station (32° 44′ N, longitude 117° 10′ W) also previously used by other researchers (i.e. Kisi 2009a, b) were used in the study. These stations operated by the US Environmental Protection Agency (US EPA) have high quality climatic data. The locations of the San Francisco and San Diego Station in California are shown in Fig. 3. The weather parameters considered in the study are the air temperature (T),

Fig. 3 The location of the San Francisco and San Diego Stations in California

wind speed (W), solar radiation (SR) and relative humidity (RH). The altitudes of San Francisco and San Diego Station stations are 2 and 4 m, respectively. The measured daily climatic data for these stations were downloaded from the US EPA web server (http://www.epa.gov/ceampubl/tools/metdata/us_met.htm).

The data sample consisted of 4 years (1987–1990) of daily records of T, SR, W, RH and pan evaporation (E_{pan}). For each station, the first 3 years (1987–1989) data were used to train the models and the remaining data were used for testing. The daily statistical parameters of the climatic data are given in Table 1. In the table, the x_{mean}, x_{max}, x_{min}, Sx, Cv and Csx denote the mean, maximum, minimum standard deviation, coefficient of variation and coefficient of skewness, respectively. SR variable seems to be the most effective on E_{pan} (see the correlations between SR and E_{pan} in Table 1). In both stations, the mean RH is more than 60 % in due to their location in a coastal area (See Fig. 1). The RH data of the San Francisco seem to be more effective on E_{pan} than those of the San Diego. The W and RH data of the San Diego station have more skewed distribution than those of the San Francisco.

Table 1 The statistical parameters of data set for the stations

Station	Data set	Unit	X_{mean}	X_{max}	X_{min}	Sx	Cv	Csx	Correlation with E_{pan}
San Francisco	T	°C	13.89	25.5	0.4	3.39	0.24	−0.36	0.66
	W	mile/h	11.21	28.1	1.82	4.35	0.39	0.43	0.61
	SR	Langley	409.89	747.1	57.8	183.72	0.45	0.04	0.90
	RH	%	67.43	99	21	11.91	0.18	−0.68	−0.56
	E	mm	4.34	12.6	0.4	2.32	0.53	0.4	1
San Diego	T	°C	17.26	28.9	7.2	3.34	0.19	−0.19	0.59
	W	mile/h	8.18	24.38	2.15	2.08	0.25	1.23	0.30
	SR	Langley	435.71	743.2	78.4	154.15	0.35	0.05	0.82
	RH	%	63.95	95	10	15.51	0.24	−1.22	−0.33
	E	mm	4.76	12.8	0	1.77	0.37	0.4	1

4 Results and Discussions

4.1 Performance Indices

Three statistical evaluation criteria, coefficient of determination (R^2), the root mean square error (RMSE) and mean absolute relative error (MARE), were used to assess models' performances. The RMSE and MARE are defined as:

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(E_{io} - E_{ie}\right)^2} \tag{11}$$

$$MARE = \frac{1}{N}\sum_{i=1}^{N}\left|\frac{E_{io} - E_{ie}}{E_{io}}\right|100 \tag{12}$$

where E_{io} and E_{ie} denote the observed and estimated evaporation values, respectively and N is the number of data sets.

Coefficient of determination (R^2) indices standardizes the differences between observed and modeled means and variances and is also more sensitive to outliers, so it alone should not be used as a fitness measure. Therefore, it is suitable to quantify the error by using RMSE or MARE measures in addition to the application of coefficient of determination. The combined use of these measures is found to be adequate for evaluating the applied models (Legates and McCabe 1999).

4.2 San Francisco Station

By applying stepwise method, the first significant (at 5 % level) variable entered to the model was the SR. The second variable that entered to the model was RH. The third one was W. Mean T was finally entered to the model. So, by using stepwise method the input combinations at different steps are summarized as follow:

i) SR
ii) SR, RH
iii) SR, RH, W
iv) SR, RH, W, T

The results of stepwise method are shown in Table 2.

The regression equations obtained for each combination are:

$$\begin{aligned}
&Combination(i): \ln\left(E_{pan}\right) = 1.063\ \ln(SR) - 4.97\\
&Combination(ii): \ln\left(E_{pan}\right) = 0.961\ \ln(SR) - 0.806\ \ln(RH) - 0.992\\
&Combination(iii): \ln\left(E_{pan}\right) = 0.749\ \ln(SR) - 1.018\ \ln(RH) + 0.493\ \ln(W) - 0.004\\
&Combination(iv): \ln\left(E_{pan}\right) = 0.614\ \ln(SR) - 1.104\ \ln(RH) + 0.499\ \ln(W) + 0.369\ \ln(T) + 0.131
\end{aligned}$$

The same input combinations are used in two different ANFIS models. The statistical parameters of different models including MNLR, ANFIS-GP and ANFIS-SC for test period are presented in Tables 2, 3 and 4. The performance of ANFIS models is found to be better than the MNLR for all input combinations. In most of the input combinations, the performance of ANFIS-GP model seems to be slightly better than the ANFIS- SC model (input combination (ii) and (iii)). For the input combination (i), the performance of ANFIS-SC model is slightly better than the ANFIS-GP model. All models gave their best estimates for the four model inputs, SR, RH, W and T. This indicates that all these variables are needed for better E_{pan} modeling.

As can be seen from Tables 2, 3 and 4, The RMSE, MARE and R^2 values of the optimal ANFIS-GP model are very close to those of the optimal ANFIS-SC model. Using only the SR as an input (input combination (i)) gave relatively good estimation. The reason may be attributed to the relatively high correlation (0.90) between SR and E_{pan} (see Table 1). Adding temperature to the input combination (input combination iv) significantly increased the performance of ANFIS models. The reductions in the amount of RMSE for the ANFIS-GP and ANFIS-SC are 58 and 56 %, respectively.

For the ANFIS-GP model, different MF types and numbers were considered. For all input combinations, triangular MF(trimf) was found to be the optimal (Table 3). In the case of one variable (Model No: 1), 4 MF yielded the best results. Kisi et al. (2009) demonstrated that two or three MF was sufficient for evaporation estimation. For ANFIS-SC model the performance of the models was evaluated by changing Radii values in the range between 0 and 1 (Table 4). The best value for Radii was obtained by trial and error. For example, the optimal value of Radii was found to be 0.40 for the input combination (iv).

MNLR and ANFIS models were also compared with ANN, SS and Penman models. The conjugate gradient algorithm was used for adjusting the weights of the ANN model because this technique is more powerful and faster than the conventional gradient descent technique (Kisi and Uncuoglu 2005; Kisi 2007b). The sigmoid activation functions were used for the hidden and output nodes. The ANN network training was stopped after 250 epochs following the suggestion of Kisi and Uncuoglu (2005) and Kisi (2007b). For the ANN model, nine hidden nodes were found to be sufficient. Al-Shalan and Salih (1987) compared 23 well-known climatic methods of evaporation estimation and concluded that the Stephens-Stewart model was found to perform the best of all. The model is:

$$E = SR(a + bT) \tag{13}$$

Table 2 The statistical parameters of MNLR for test period-San Francisco

Model No.	Model input	R^2	RMSE	MARE
1	SR	0.810	0.952	19.06
2	SR,RH	0.870	0.796	18.00
3	SR, RH,W	0.909	0.661	12.90
4	SR, RH,W,T	0.932	0.644	10.87

Table 3 Error statistics of optimal model of ANFIS-GP models based on different MF type and different MF number (for test period)-San Francisco

Model input	MF type	MF number	R^2	RMSE	MARE
SR	trimf	4	0.810	0.951	20.34
SR, RH	trimf	2 2	0.880	0.750	17.75
SR, RH,W	trimf	2 3 2	0.951	0.485	10.51
SR, RH,W,T	trimf	3 2 2 3	0.992	0.204	4.42

where E is the daily class A pan evaporation, SR is the daily solar radiation, T is the mean daily air temperature, a and b are the fitting parameters. In this study, the least squares method was used to obtain the values of the parameters, a and b. The Penman (1948) method for estimating the pan evaporation is:

$$E = \frac{\frac{\Delta(Rn-G)}{\lambda} + \gamma_p . E_a}{\Delta + \gamma_p} \tag{14}$$

where, E = pan evaporation (mm/day), λ = latent heat of the evaporation (MJ/Kg), Δ = slope of the saturation vapor pressure versus temperature function (kPa/°C), R_n = net radiation (MJ/m^{2}day), G = Soil heat flux density (MJ/m^{2}day), γ_p = psychometric constant (kPa/°C), and E_a = aerodynamic function (mm/day).

The optimal MNLR, ANFIS-GP and ANFIS-SC models are compared with ANN, SS and Penman models in respect of RMSE, MARE and R^2 statistics in Table 5. It is clear from the table that the ANFIS and MNLR models performed better than the ANN and SS models. Penman model seems to be better than the MNLR model.

Different model's estimates for E_{pan} at San Francisco are presented in Fig. 4 in the form of time series hydrograph and scatter plot. As it can be seen from the hydrographs and scatter plots that the ANFIS estimates are closer to the corresponding observed E_{pan} values than the MNLR, SS, Penman and ANN models. The difference between the ANFIS-GP and ANFIS-SC models cannot be clearly seen from the figures. It can be inferred from the Fig. 4 that the performance of the ANFIS models in estimation of maximum values of E_{pan} is significantly better than the MNLR, SS, Penman and ANN models. The underestimation of the peak values is clearly seen for the SS and Penman models. The estimation of E_{pan} in annual time scale is considered in the study because of its importance in irrigation management. Total annual evaporation was underestimated by different models in test period. ANFIS-GP, ANFIS-SC, MNLR models estimated the observed total evaporation value of 1586.7 mm as 1562.7 mm, 1,564 mm and 1472.9 mm, with underestimations of 1.51, 1.43 and 7.17 while the SS and Penman resulted in 1592.4 mm and 1718.8 mm, with overestimations of 0.36 and 8.33 %, respectively. The SS estimate is found to be the closest estimate to the observed one with the smallest relative error. The estimates of the ANFIS models are also closer to the observed value than those of the MNLR, ANN and Penman models.

Table 4 Error statistics of optimal model of ANFIS-SC models based on different cluster radius (Radii) values (for test period)-San Francisco

Model input	Radii	R^2	RMSE	MARE
SR	0.38	0.811	0.950	20.32
SR, RH	0.86	0.875	0.769	18.44
SR, RH,W	0.79	0.947	0.497	10.59
SR, RH,W,T	0.40	0.990	0.216	4.29

Table 5 Comparison of optimal MNLR, ANFIS-GP, ANFIS-SC, ANN, Penman and SS models in the test period-San Francisco

Model	Model input	R^2	RMSE	MARE
MNLR	SR, RH,W,T	0.932	0.644	10.87
ANFIS-GP	SR, RH,W,T	0.992	0.204	4.42
ANFIS-SC	SR, RH,W,T	0.990	0.216	4.29
ANN	SR, RH,W,T	0.892	0.742	14.9
Penman	SR, RH,W,T	0.956	0.589	14.2
SS	SR,T	0.838	0.883	17.7

The ANFIS-GP, ANFIS-SC, MNLR, SS, Penman and ANN models have 337, 331, 215, 158, 169 and 172 estimates lower than the 10 % relative error in test period, respectively. Furthermore, the ANFIS-GP, ANFIS-SC, MNLR, SS, Penman and ANN have 248, 262, 97, 78, 77 and 69 estimates lower than the 5 % relative error in test period, respectively. It seems that the performance of the ANFIS-GP and ANFIS-SC models are significantly better than the MNLR, SS, Penman and ANN models.

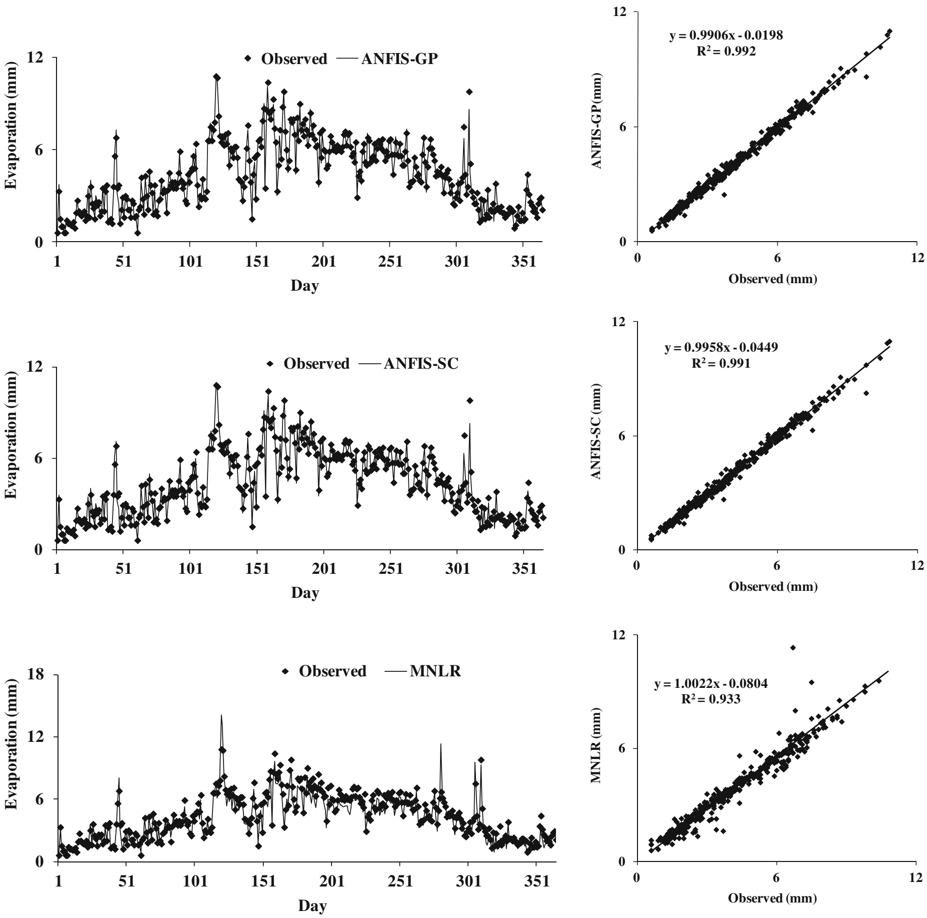


Fig. 4 The observed and estimated E_{pan} of San Francisco Station for test period

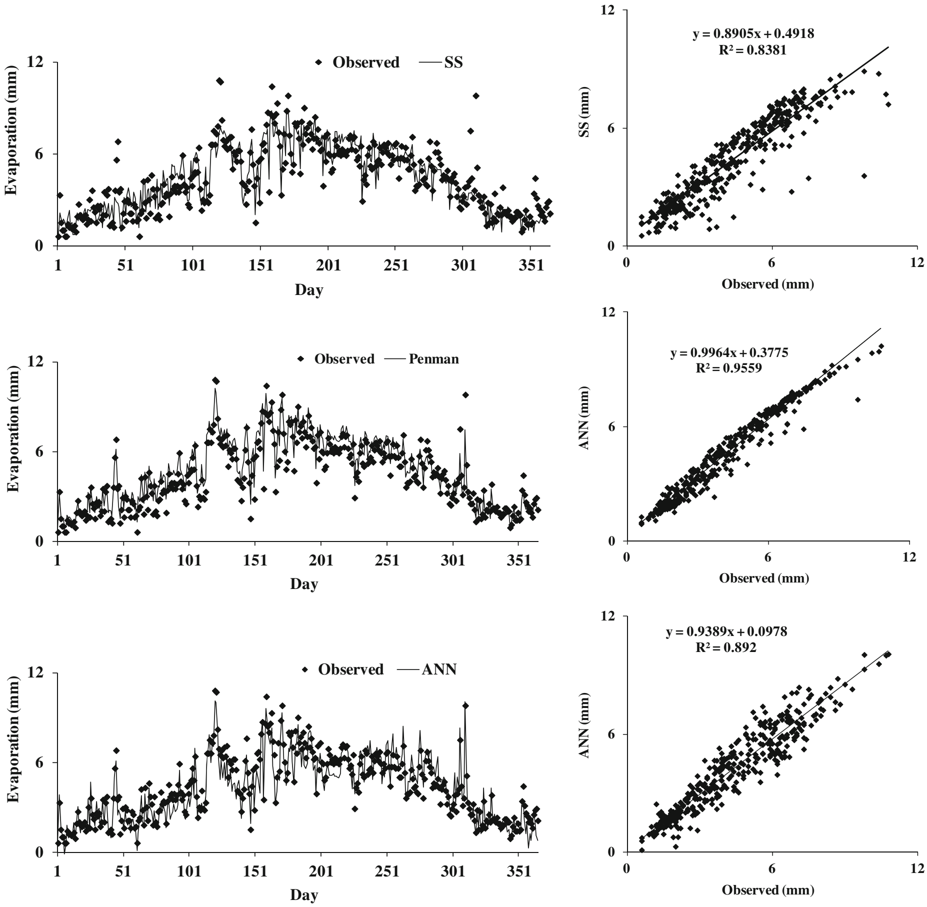


Fig. 4 (continued)

The results were also tested by ANOVA for verifying the robustness (the significance of differences between the observed values and model estimates) of the models. Tests were set at a 95 % significant level. The statistics of the tests are given in Table 6. The SS model gives smaller testing values with higher significance level than the other models. Both ANFIS-GP and ANFIS-SC models show similar test results and are more robust than the MNLR and Penman models. According to the ANOVA analysis, the Penman model is the worst in robustness for estimating E_{pan}.

4.3 San Diego Station

The MNLR equations obtained for San Diego are:

$$Combination(i) : \ln(E_{pan}) = 0.835\ \ln(SR) - 3.533$$
$$Combination(ii) : \ln(E_{pan}) = 0.871\ \ln(SR) - 0.488\ \ln(RH) - 1.739$$
$$Combination(iii) : \ln(E_{pan}) = 0.804\ \ln(SR) - 0.609\ \ln(RH) + 0.474\ \ln(W) - 1.822$$
$$Combination(iv) : \ln(E_{pan}) = 0.68\ \ln(SR) - 0.655\ \ln(RH) + 0.465\ \ln(W) + 0.44\ \ln(T) + 2.107$$

The statistical indices of MNLR, ANFIS-GP and ANFIS-SC models in test period are presented in Tables 7, 8 and 9. It is clear from Table 8 that the ANFIS-GP model whose

Table 6 ANOVA of MNLR, ANFIS-GP, ANFIS-SC, ANN, Penman and SS models in the test period-San Francisco

Method	*F*-Statistic	Resultant Significance Level
MNLR	3.862	0.050
ANFIS-GP	0.167	0.683
ANFIS-SC	0.149	0.699
ANN	1.093	0.296
Penman	4.950	0.026
SS	0.010	0.922

inputs are SR, RH, W and T has the smallest RMSE (0.206), MARE (3.67 %) and the highest R^2 (0.986) values. The best performance of MNLR and ANFIS-SC model is also obtained by applying this input combination. ANFIS-GP results seem to be parallel to those of the ANFIS-SC models and both ANFIS models perform better than the MNLR models. For this station, using only SR as input (input combination (i)) has worse results than those of the San Francisco Station (Tables 7, 8 and 9). This may be related to low correlation (0.82) between SR and E_{pan} in comparison to the San Francisco Station. As can be seen from the Tables 7, 8 and 9, adding RH input (input combination (ii)) increased the models' performances in terms of R^2, RMSE and MARE. Adding T into the input combination (iii) significantly increased the ANFIS models. Accordingly, the RMSE and MARE values of ANFIS-GP, were decreased by 58 % and 54 % respectively but in the case of ANFIS-SC, they decreased by 57 % and 51 %, respectively. For this station also triangular MF yielded the best results in comparison of other MF types for the all input combinations (Table 8).

The optimal MNLR, ANFIS-GP and ANFIS-SC models are compared with ANN, Penman and SS models in Table 10. It is obvious from the table that the ANFIS-GP and ANFIS-SC models performed better than the ANN, SS and Penman models. In contrast to San Francisco, the ANN model showed better accuracy than the MNLR model.

The hydrographs and scatterplots of observed and estimated E_{pan} for San Diego are shown in Fig. 5. As can be seen from Fig. 5, the estimation of E_{pan} using ANFIS-GP and ANFIS-SC models closely follow the corresponding observed values. As it can be seen from the fit line equations (with assuming linear equation as $y=Ax+B$) in the scatter plots that the A and B coefficients for ANFIS-GP and ANFIS-SC models are respectively, closer to the 1 and 0 with a higher R^2 values of 0.986 and 0.985 than those of the MNLR, SS, Penman and ANN models. This confirms the results of performance indices presented in Tables 10. The total evaporation estimates of ANFIS-GP, ANFIS-SC, MNLR, SS, ANN and Penman models were 1.18, 1.07, 2.51, 5.23 and 2.19 % lower and 4.56 % higher than the observed value (1,814 mm) in test period, respectively. ANFIS-GP and ANFIS-SC models have 346

Table 7 The statistical parameters of MNLR in the test period-San Diego

Model No.	Model input	R^2	RMSE	MARE
1	SR	0.688	0.992	14.74
2	SR, RH	0.808	0.768	13.05
3	SR, RH,W	0.881	0.620	9.69
4	SR, RH,W,T	0.904	0.544	7.32

Table 8 The error statistics and properties of ANFIS-GP model in the test period-San Diego

Model input	MF type	MF number	R^2	RMSE	MARE
SR	trimf	4	0.699	0.958	14.36
SR, RH	trimf	3 4	0.843	0.688	11.83
SR, RH,W	trimf	3 4 4	0.923	0.502	7.91
SR, RH,W,T	trimf	4 2 2 2	0.986	0.206	3.67

and 338 estimates lower than the 10 % relative error in test period while the MNLR, SS, Penman and ANN models have 285, 181, 189 and 284 estimates. Furthermore, the ANFIS-GP and ANFIS-SC have 274 and 267 estimates lower than the 5 % relative error in test period while the MNLR, SS, Penman and ANN has 164, 85, 86 and 195 estimates lower than 5 % error. In this matter, ANFIS-GP and ANFIS-SC almost have the same results and superior to the MNLR, SS, Penman and ANN models. The ANOVA test statistics are given in Table 11. As found for the San Francisco station, both ANFIS-GP and ANFIS-SC models show similar test results and are more robust than the ANN, MNLR, Penman and SS model. In contrast to San Francisco station, SS model is the worst in robustness for estimating E_{pan}.

In general, results indicated that the performance of ANFIS-GP and ANFIS-SC in modeling evaporation is better than the MNLR, ANN, SS and Penman models. It seems that the ANFIS models are sufficient approaches in modeling sophisticated phenomena like evaporation that relationship between variables are nonlinear. Comparison of results of ANFIS-GP and ANFIS-SC showed that both models yield similar performances in evaporation modeling and the difference of estimated values of ANFIS models are not significant.

5 Conclusions

The ability of two different ANFIS models, ANFIS-GP and ANFIS-SC, in modeling daily pan evaporation was investigated in this study. Daily air temperature, wind speed, solar radiation and relative humidity data of two stations in California State of US were used for E_{pan} modeling. For selecting the appropriate input combination for ANFIS models, MNLR model by using stepwise method was used. The logarithmic MNLR method was used to select effective meteorological input variables. For both stations, solar radiation was identified as the most important variable by using stepwise method and other variables were the next ranks. Different input combinations were applied by using MNLR, ANFIS-GP and ANFIS-SC models. The best results were generally obtained by using all of the variables as input combination. It was found that using only the solar radiation input gives relatively acceptable estimates for

Table 9 The error statistics and properties of ANFIS-SC model in the test period-San Diego

Model input	Radii	R^2	RMSE	MARE
SR	0.60	0.696	0.964	14.45
SR, RH	0.55	0.845	0.682	11.63
SR, RH,W	0.53	0.922	0.503	7.91
SR, RH,W,T	0.70	0.984	0.215	3.87

Table 10 Comparison of optimal MNLR, ANFIS-GP, ANFIS-SC, ANN, Penman and SS models in the test period-San Diego

Model	Model input	R^2	RMSE	MARE
MNLR	SR, RH,W,T	0.904	0.544	7.32
ANFIS-GP	SR, RH,W,T	0.986	0.206	3.67
ANFIS-SC	SR, RH,W,T	0.984	0.215	3.87
ANN	SR, RH,W,T	0.953	0.387	6.77
Penman	SR, RH,W,T	0.888	0.623	11.2
SS	SR,T	0.715	0.992	15.4

the San Francisco station. Adding air temperature to input combinations significantly increased the models' performances. The results indicated that the impact of air temperature on evaporation is so large. The performances of ANFIS-GP and ANFIS-SC models were compared with the MNLR, ANN, SS and Penman models. The comparison results indicated that the performance of ANFIS-GP and ANFIS-SC

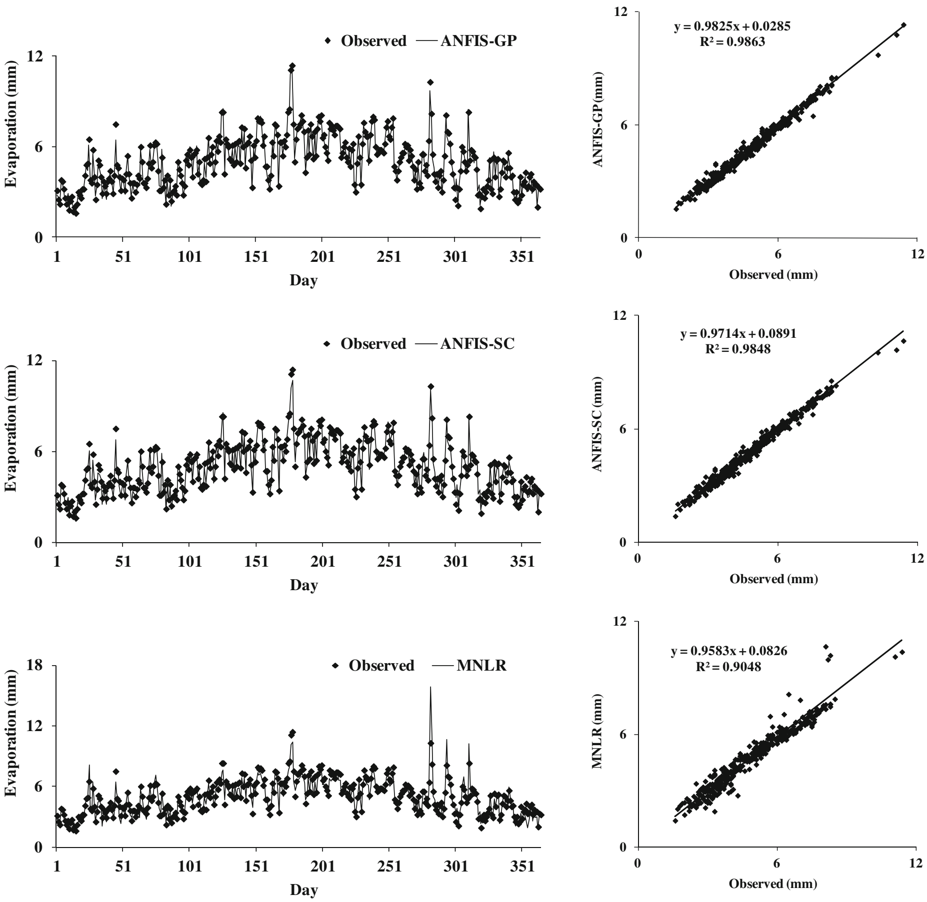


Fig. 5 The observed and estimated E_{pan} of San Diego Station for test period

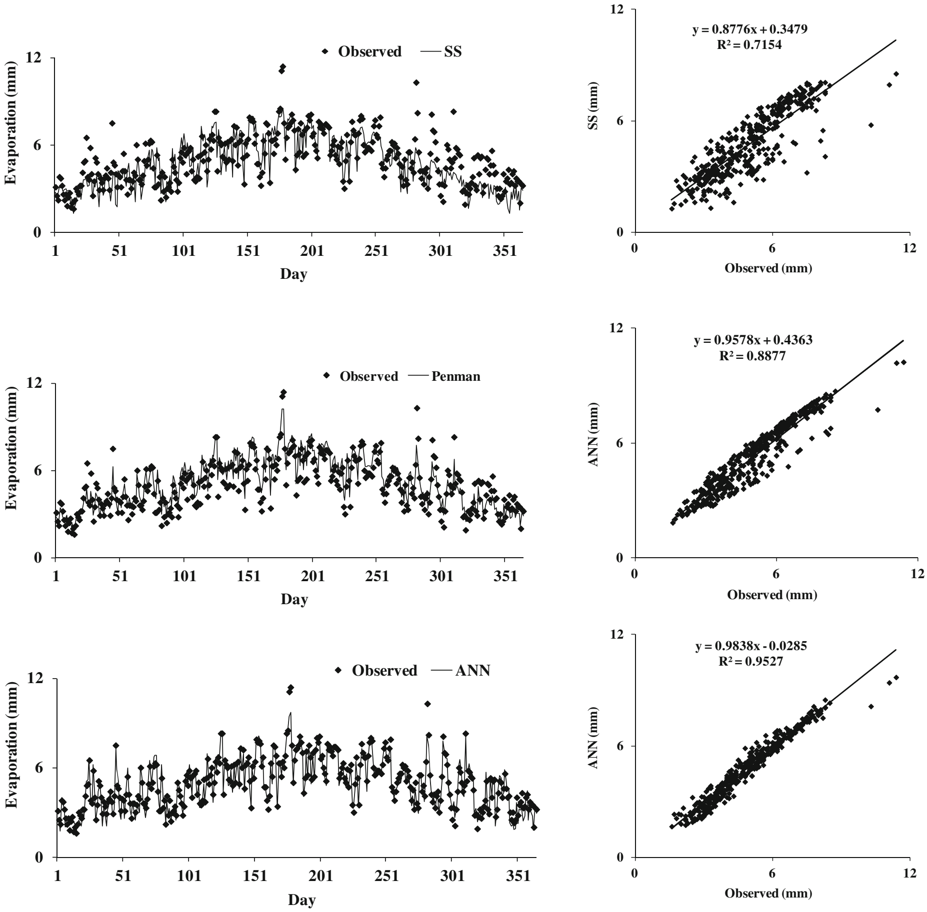


Fig. 5 (continued)

is better than the MNLR, ANN, SS and Penman in modeling evaporation. Results also showed that the ANFIS-GP and ANFIS-SC have similar performances and there is no considerable difference between the results of ANFIS models. However, simple ANFIS-SC models with less computation can be successfully used as an alternative to the more complex ANFIS-GP models in E_{pan} modeling.

Table 11 ANOVA of MNLR, ANFIS-GP, ANFIS-SC, ANN, Penman and SS models in the test period-San Diego

Method	*F*-Statistic	Resultant Significance Level
MNLR	0.984	0.321
ANFIS-GP	0.221	0.639
ANFIS-SC	0.184	0.668
ANN	0.750	0.387
Penman	3.229	0.073
SS	4.161	0.042

References

Abonyi J, Andersen H, Nagy L, Szeifert F (1999) Inverse fuzzy-process-model based direct adaptive control. Math Comput Simul 51:119–132

Al-Shalan A, Salih AMA (1987) Evapotranspiration estimation in extremely arid areas. ASCE J Irrig Drain Eng 113:565–574

Aytek A (2009) Co-active neurofuzzy inference system for evapotranspiration modeling. Soft Comput 13:691–700

Chiu S (1994) Fuzzy model identification based on cluster estimation. J Intell Fuzzy Syst 2(3):267–278

Chiu S (1997) Extracting fuzzy rules from data for function approximation and pattern classification. In: Dubois D, Prade H, Yager R (eds) Fuzzy information engineering: A Guided Tour of Applications. Springer, Berlin, pp 149–162

Chu HJ, Chang LC (2009) Application of optimal control and fuzzy theory for dynamic groundwater remediation design. Water Resour Manag 23(4):647–660

Cobaner M (2011) Evapotranspiration estimation by two different neuro-fuzzy inference systems. J Hydrol 398:292–302

Coulomb CV, Legesse D, Gasse F, Travi Y, Chernet T (2001) Lake evaporation estimates in tropical Africa (Lake Ziway, Ethiopia). J Hydrol 245(1–4):1–18

Dogan E, Gumrukcuoglu M, Sandalci M, Opan M (2010) Modelling of evaporation from the reservoir of Yuvacik dam using adaptive neuro-fuzzy inference systems. Eng Appl Artif Intell 23:961–967

Drake JT (2000) Communications phase synchronization using the adaptive network fuzzy inference system. Dissertation, New Mexico State University

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

Gaven H, Agnew CA (2004) Modelling actual, reference and equilibrium evaporation from temperate wet grassland. Hydrol Process 18(2):229–246

Goyal MK, Ojha CSP (2011) Estimation of scour downstream of a ski-jump bucket using support vector and M5 model tree. Water Resour Manag 25(9):2177–2195

Guldal V, Tongal H (2010) Comparison of recurrent neural network, adaptive neuro-fuzzy inference system and stochastic models in Egirdir Lake level forecasting. Water Resour Manag 24(1):105–128

Guven A, Kisi O (2011) Daily pan evaporation modeling using linear genetic programming technique. Irrig Sci 29(2):135–145

Jackson RD (1985) Evaluating evapotranspiration at local and regional scales. Proc IEEE 73(6):1086–1096

Jang JSR (1993) ANFIS: adaptive-network-based fuzzy inference system. IEEE Trans Syst Man Cybern 23 (3):665–685

Keskin ME, Terzi O (2006) Artificial neural network models of daily pan evaporation. J Hydrol Eng 11(1):65–70

Keskin ME, Terzi O, Taylan D (2004) Fuzzy logic model approaches to daily pan evaporation estimation in Western Turkey. Hydrol Sci J 49(6):1001–1010

Keskin ME, Terzi O, Taylan C (2009) Estimating daily pan evaporation using adaptive neural-based fuzzy inference system. Theor Appl Climatol 98:79–87

Khu ST, Liong SY, Babovic V, Madsen H, Muttil N (2001) Genetic programming and its application in real-time runoff forming. J Am Water Resour Assoc 37(2):439–451

Kisi O (2006a) Evapotranspiration estimation using feed forward neural networks. Nord Hydrol 37(3):247–260

Kisi O (2006b) Daily pan evaporation modeling using a neuro-fuzzy computing technique. J Hydrol 329:636–646

Kisi O (2007a) Evapotranspiration modeling from climate data using a neural computing technique. Hydrol Process 21(6):1925–1934

Kisi O (2007b) Streamflow forecasting using different artificial neural network algorithms. ASCE J Hydrol Eng 12(5):532–539

Kisi O (2009a) Daily pan evaporation modeling using multi-layer perceptrons and radial basis neural networks. Hydrol Process 23:213–223

Kisi O (2009b) Modelling monthly evaporation using two different neural computing techniques. Irrig Sci 27 (5):417–430

Kisi O, Uncuoglu E (2005) Comparison of three backpropagation training algorithms for two case studies. Indian J Eng Mater Sci 12:443–450

Kisi O, Haktanir T, Ardiclioglu M, Ozturk O, Yalcin E, Uludag S (2009) Adaptive neuro-fuzzy computing technique for suspended sediment estimation. Adv Eng Softw 40:438–444

Legates DR, McCabe GJ (1999) Evaluating the use of goodness-of-fit measures in hydrologic and hydro-climatic model validation. Water Resour Res 35(1):233–241

Mamdani EH, Assilian S (1975) An experimental in linguistic synthesis with fuzzy logic controller. Int J Man–Mach Stud 7:1–13

McCuen RH (1998) Hydrologic analysis and design. Prentice Hall, Englewood Cliffs

Moghadamnia A, Gousheh MG, Piri J, Amin S, Han D (2009) Evaporation estimation using artificial neural networks and adaptive neuro-fuzzy inference system techniques. Adv Water Resour 32:88–97

Penman WR (1948) Natural evaporation from open water, bare soil and grass. Proc R Soc Edinb A 193:120–145

Piri J, Amin S, Moghaddamnia A, Keshavarz A, Han D, Remesan R (2009) Daily Pan evaporation modeling in a hot and dry climate. J Hydrol Eng 14:803–812

Rahimikhoob A (2008) Artificial neural network estimation of reference evapotranspiration from pan evaporation in a semi-arid environment. Irrig Sci 27:35–39

Rahimikhoob A (2009) An evaluation of common pan coefficient equations to estimate reference evapotranspiration in a subtropical climate (north of Iran). Irrig Sci 27(4):289–296

Rawlings JO (1988) Applied regression analysis: A research tool. Wadsworth & Brooks/Cole Advanced Books & Software, Pacific Grove, California

Razi MA, Athappilly K (2005) A comparative predictive analysis of neural networks (NNs), nonlinear regression and classification and regression tree (CART) models. Expert Syst Appl 29:65–74

Reis RJ, Dias NL (1998) Multi-season lake evaporation: energy-budget estimates and CRLE model assessment with limited meteorological observations. J Hydrol 208(3–4):135–147

Sabziparvar A-A, Tabari H, Aeini A, Ghafouri M (2010) Evaluation of class A pan coefficient models for estimation of reference crop evapotranspiration in cold semi-arid and warm arid climates. Water Resour Manag 24(5):909–920

Sayed T, Tavakolie A, Razavi A (2003) Comparison of adaptive network based fuzzy inference systems and b-spline neuro-fuzzy mode choice models. J Comput Civ Eng 17(2):123–130

Snyder RL (1993) Equation for evaporation pan to evapotranspiration conversions. J Irrig Drain Eng 118 (6):977–980

Sudheer KP, Gosain AK, Rangan DM, Saheb SM (2002) Modelling evaporation using an artificial neural network algorithm. Hydrol Process 16:3189–3202

Tabari H, Marofi S, Sabziparvar AA (2010) Estimation of daily pan evaporation using artificial neural network and multivariate non-linear regression. Irrig Sci 28(5):399–406

Takagi T, Sugeno M (1985) Fuzzy identification of systems and its applications to modeling and control. IEEE Trans Syst Man Cybern 15:116–132

Tan SBK, Shuy EB, Chua LHC (2007) Modelling hourly and daily open-water evaporation rates in areas with an equatorial climate. Hydrol Process 21(4):486–499

Trajkovic S (2009) Comparison of radial basis function networks and empirical equations for converting from pan evaporation to reference evapotranspiration. Hydrol Process 23(6):874–880

Trajkovic S (2010) Testing hourly reference evapotranspiration approaches using lysimeter measurements in a semiarid climate. Hydrol Res 41(1):38–49

Trajkovic S, Kolakovic S (2010) Comparison of simplified pan-based equations for estimating reference evapotranspiration. J Irrig Drain Eng 136(2):137–140

Trajkovic S, Stankovic M, Todorovic B (2000) Estimation of FAO Blaney-Criddle b Factor by RBF Network. J Irrig Drain Eng 126(4):268–271

Traore S, Guven A (2011) New algebraic formulations of evapotranspiration extracted from gene-expression programming in the tropical seasonally dry regions of West Africa. Irrig Sci. doi:10.1007/s00271-011-0288-y

Traore S, Wang YM, Kerh T (2010) Artificial neural network for modeling reference evapotranspiration complex process in Sudano-Sahelian zone. Agric Water Manag 97:707–714

Tsukamoto Y (1979) An approach to reasoning method. In: Gupta M, Ragade RK, Yager RR (eds) Adv Fuzzy Set Theo Appl, Amsterdam, p 137–149

Wang D, Singh V, Zhu Y (2007) Hybrid fuzzy and optimal modeling for water quality evaluation. Water Resour Res 43(5):W05415

Wang D, Singh V, Zhu Y, Wu J (2009) Stochastic observation error and uncertainty in water quality evaluation. Adv Water Resour 32(10):1526–1534

Wei M, Bai B, Sung AH, Liu Q, Wang J, Cather ME (2007) Predicting injection profiles using ANFIS. Inform Sci 177:4445–4461

Yager RR, Filev DP (1994) Approximate clustering via the mountain method. IEEE Trans Syst Man Cybern 24(8):1279–1284