# Evaporation Estimation Using Adaptive Neuro-Fuzzy Inference System and Linear Regression

**Dr. Ali H. Al-Aboodi**
Engineering College, University of Basrah /Basrah
Email: alialaboodi90@gmail.com

**Received on: 5/12/2013 & Accepted on: 7/8/2014**

## ABSTRACT

Evaporation is important for water planning, management and hydrological practices, and it plays an influential role in the management and development of water resources. This study demonstrates the application of two different models, adaptive neuro-fuzzy inference system (ANFIS), and linear regression (LR) models for estimating monthly pan evaporation in Basrah City, south of Iraq. In the first part of this study, the ANFIS model is used twice, in the first one, the temperature is used as input data only, and in the second one, the temperature and relative humidity are used as input data for predicting the evaporation. A verification test is added to check the model correctness by matching the calculated evaporation with the once observed in Basrah city for the period (1980-2009). In the second part of the study, the results obtained by ANFIS models are compared with results of linear regression model. The comparison reveals that the ANFIS models give better accuracy in estimating monthly pan evaporation than the linear regression model. The accuracy is improved about 5% in correlation coefficient (R) and determination coefficient ($R^2$). The results proved that monthly pan evaporation could be successfully estimated through the use of ANFIS models.

**Keywords**: Evaporation, ANFIS, linear regression, Basrah, Iraq.

## تخمين التبخر باستخدام نظام الاستدلال العصبي الضبابي المكيف والانحدار الخطي

**الخلاصة**

تخمين التبخر مهم في عملية التخطيط المائي, والإدارة والتطبيقات الهيدرولوجية، كذلك يلعب دورا مؤثرا في إدارة وتطوير الموارد المائية. في هذه الدراسة اجري تطبيق نموذجين مختلفين، هما نموذج الاستدلال العصبي الضبابي المكيف (ANFIS) ونموذج الانحدار الخطي (LR) لتخمين التبخر الشهري في مدينة البصرة، جنوب العراق. في الجزء الأول من هذه الدراسة، تم استخدام نموذج ANFIS مرتين، في الاولى استخدمت درجة الحرارة فقط كمعلومات ادخال وفي النموذج الثاني استخدمت درجة الحرارة والرطوبة النسبية كمعلومات ادخال لتخمين التبخر وقد اجري التحقق من النموذجين بمقارنت نتائجهما مع القياسات الحقلية للتبخر الشهري لمدينة البصرة المسجلة للفترة (1980-2009). أما في الجزء الثاني، تمت مقارنة نتائج نموذج (ANFIS) مع نموذج الانحدار الخطي. بينت النتائج، إن نموذج (ANFIS) يعطي نتائج أكثر دقة من نموذج الانحدار الخطي، حيث طرأ تحسن في النتائج حوالي 5% في معامل الارتباط (R) ومعامل التحديد ($R^2$)، وقد أثبتت النتائج إمكانية تخمين التبخر الشهري بنجاح باستخدام نظام الاستدلال العصبي الضبابي المكيف.

## INTRODUCTION

Evaporation is the process of a liquid becoming vaporized. In other words, a change in phase in the atmosphere occurs when substances change from a liquid to a gaseous, or vapor form. Evaporation is the primary pathway that water moves from the liquid state back into the water cycle as atmospheric water vapor. Estimation of evaporation is important for water planning, management and hydrological practices [1].

An evaporation pan is used to hold water during observations for the determination of the quantity of evaporation at a given location. Such pans are of varying sizes and shapes, the most commonly used being circular or square [2] The best known of the pans are the "Class A" evaporation pan and the "Sunken Colorado Pan" [3]. In Europe, India and South Africa, a Symon's Pan (or sometimes Symon's Tank) is used ( e.g., the most commonly used is the US Weather Bureau Class A, which is 4 ft in diameter and 10 in deep and is mounted on a timber grill about 6 in above the soil surface) [4]. Often the evaporation pans are automated with water level sensors and a small weather station is located nearby.

In the recent years, the computing techniques have been successfully used in modeling pan evaporation. Terzi and Erol Keskin (2005) used Gene Expression Programming (GEP) for modeling evaporation as a function of air temperature, solar radiation, and relative humidity [1]. Moghaddamnia et al. (2009) explored evaporation estimation method based on artificial neural networks (ANN) and adaptive neuro-fuzzy inference system (ANFIS) technique [5]. Shirsath, and Singh (2010) presented application of artificial neural networks (ANN), statistical regression and climate based model for estimating of daily pan evaporation [6]. Kumar et al (2012) developed artificial neural networks (ANN) and adaptive neuro-fuzzy inference system (ANFIS) models were developed to forecast monthly potential evaporation in Pantagar, U.S. Nagar (India) based on four explanatory climatic factors (relative humidity, solar radiation, temperature, and wind speed) [7]. Kisi (2013) proposes the application evolutionary neural networks (ENN) for modeling monthly pan evaporations [8].

The first objective in this study attempted to evaluate the application of adaptive neuro-fuzzy inference system (ANFIS) and linear regression (LR) model for monthly pan evaporation estimation using different climate input of Basrah City, such as temperature and relative humidity. While, the second objective of this study aimed of comparing (ANFIS) approach with (LR) for monthly pan evaporation estimation.

### Study Area and Data Set

Basrah City is located on the Shatt al-Arab river in southern Iraq. It is located between longitude line ($47^o\ 30^' – 48^o\ 30^'$) and latitude line ($30^o\ 00^' – 30^o\ 30^'$) as shown in figure (1). The great rivers Euphrates and Tigris flow through Iraq before converging into Shatt al-Arab, shortly before reaching Basrah. The huge river connects the city with the Arabian Gulf and gives it a coastal feel. The summer heat is extreme, and in winter frost is not unknown. Nevertheless the climate is considered healthy and agreeable. Rainfall usually begins in October and continuous till May. The climate information used in this study is obtained from the meteorological recording station in Basrah City for the period (1980-2009).

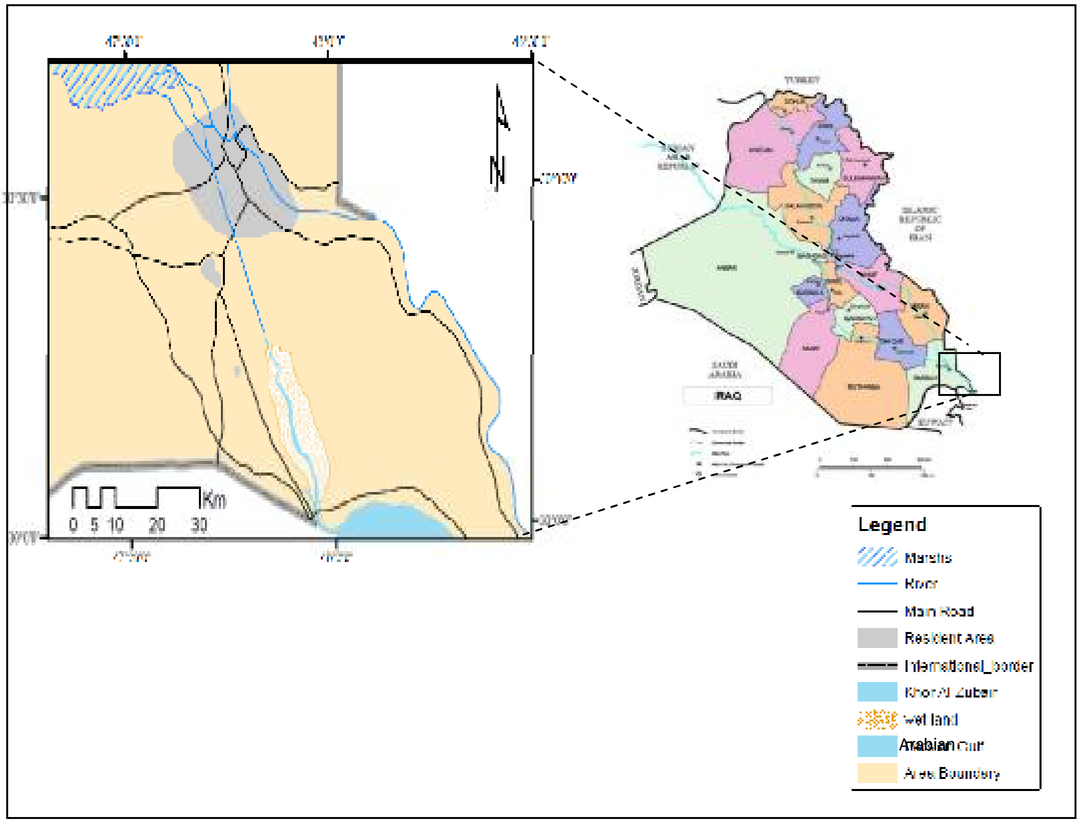


**Figure (1) Location of study area in reference to map of Iraq.**

## Adaptive Neuro-Fuzzy Inference System (Anfis)

Adaptive Neuro-Fuzzy Inference System (ANFIS) is an artificial intelligence technique that has been successfully used for mapping input-output relationship based on available data sets [9]. ANFIS uses a Neural Network (NN) learning algorithm for constructing a set of fuzzy if-then rules with appropriate membership functions (MFs) from special input-output pairs. It is based on the first order Sugeno-fuzzy inference system proposed by Jang (1993) [10]. Interpreting an if-then rule involves distinct parts: first evaluating the antecedent (which involves fuzzifying the input and applying any necessary fuzzy operators) and second applying the results to the consequent (known as implication). If the ANFIS structure is consisted of five layers as shown in figure (2), it can be described as a multi-layered neural network. Each node output in first layer is fuzzified by membership grade of a fuzzy set corresponding to each input, each node output in the second layer represents the firing strength of a rule, which performs fuzzy, AND operation, each node in this layer, labeled Π, is a stable node which multiplies incoming signals and sends the product out, the third layer normalizes the membership functions (MFs), the fourth layer executes the consequent part of the fuzzy rules, and finally the last layer computes the output of the fuzzy system by summing up the outputs of the fourth layer.

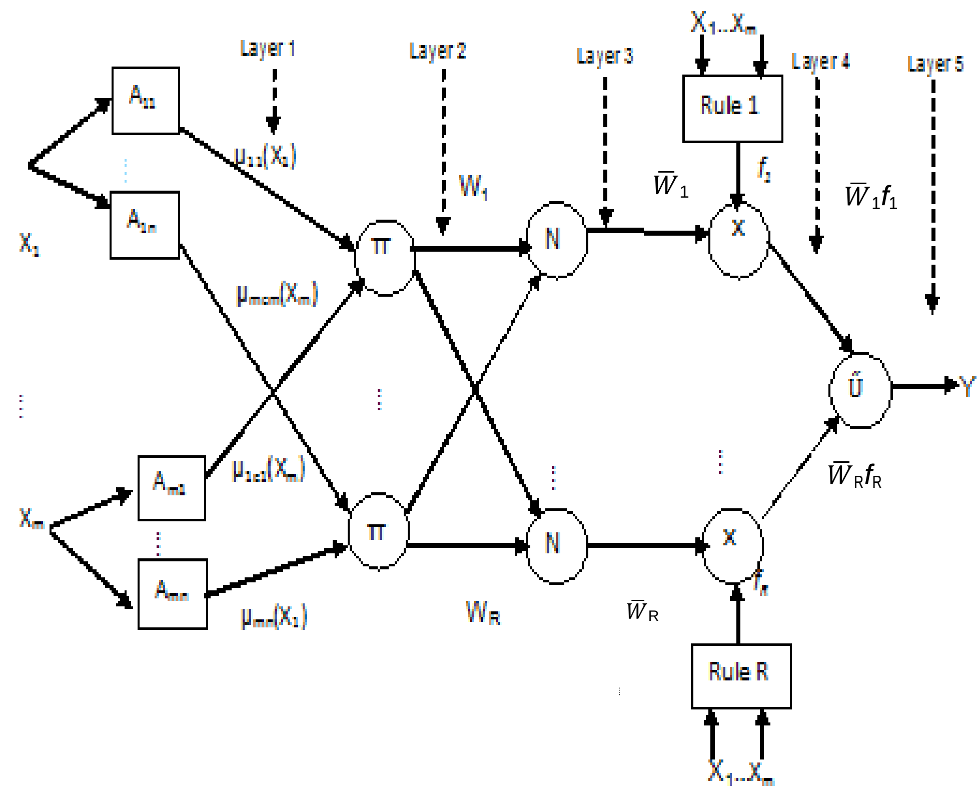


**Figure (2) Basic structure of the ANFIS.**

For constructing a fuzzy system, the linguistic variables should be provided in addition to the numerical variables. Then, the system requires If/Then fuzzy rules to qualify simple relationships between the fuzzy variables [11]. A typical rule set with two fuzzy If/Then rules in a first-order Sugeno system, as follow

*Rule 1 : If x is* $A_1$ *and y is* $B_1$*, then* $F_1 = p_1x + q_1y + r_1$ …(1)

*Rule 2 : If x is* $A_2$ *and y is* $B_2$*, then* $F_2 = p_2x + q_2y + r_2$ … (2)

Where:

x,y : input, and output variables respectively.

A,B : linguistic terms of the precondition part with MF.

p,q,r : consequent parameters.

For a zero-order Sugeno model, the output level $F$ is a constant (p=q =0). The output level $F_i$ of each rule is weighted by the firing strength $w_i$ of that rule. For example, for an AND rule with Input 1 = x and Input 2 = y, the firing strength is

$$wi = And\ Method(F1(x), F2(y)) \quad \dots (3)$$

Where:
$F_1, F_2$ (.) are the membership functions for inputs 1 and 2.
The final output of the system is the weighted average of all rule outputs, computed as

$$final\ output = \frac{\sum_{i=1}^{N} w_{i.F_i}}{\sum_{i=1}^{N} w_i} \quad \dots (4)$$

Where:
N : number of rules.

## LINEAR REGRESSION

Regression analysis uses more sophisticated equations to analyze larger sets of data and translates them into coordinates on a line or curve. Linear regression is a statistical technique used to observe trends, determine correlation, and predict future observations. The underlying principle of this technique is called the least-squared, which is the process of minimizing the distance between the predicted value of an observation and the actual value of that observation. Finding the best linear fit between two paired variables is very useful in many science applications [12]. The prediction of the dependent variable ($\hat{Y}$) is given by the following equation.

$$\hat{Y} = a + bX \quad \dots (5)$$

Where:
$X$ : independent variable.

The least square method defines the estimate of these parameters as the values which minimize the sum of the squares between the measurements and the model (i.e., the predicted values). This amounts to minimizing the expression [12]:

$$\mathcal{E} = \sum_i (Y_i - \hat{Y}_i)^2 = \sum_i [Y_i - (a + bX_i)]^2 \quad \dots (6)$$

Where:
$\mathcal{E}$: error (which is the quantity to be minimized).
$Y_i$, $X_i$: observations are used to find a function relating the value of the dependent variable ($Y$) to the values of an independent variable ($X$).

## Methodology

In this study, ANFIS and LR models are used to predict monthly pan evaporation in Basrah City. The monthly pan evaporation for the period (1980-2009) is used here. For the year 1980, the highest monthly pan evaporation was 663.1 mm and the lowest value 40.8 mm was recorded. The ANFIS model is used twice, in the first one, the temperature is used as input data only, and in the second one, the temperature and relative humidity are used as input data for predicting the evaporation. The Gaussian member function is adopted in this study since it is the most popular form as illustrated in figure (3). A membership function (MF) is a curve that defines how each point in the input space is mapped to a membership value (or degree of membership) between 0 and 1. The number of MF is selected equal to (4) according to trial and error procedure based on root mean squared error (RMSE). The total number of observations is 192, these observations are divided into two statistically parts, 70%

(135) for training, and 30% (57) for checking. The idea behind using a checking data set for model validation is that after a certain point in the training, the model begins overfitting the training data set. In principle, the model error for the checking data set tends to decrease as the training takes place up to the point that overfitting begins, and then the model error for the checking data suddenly increases [13]. Overfitting is accounted for by testing the fuzzy inference system (FIS) trained on the training data against the checking data, and choosing the membership function parameters to be those associated with the minimum checking error if these errors indicate model overfitting [14]. In the present model, the number of training epoch (40), error tolerance (0), initial step size (0.01), step size decrease rate (0.9), and step size increase rate (1.1). While, the methods of this model are, andMethod (prod), orMethod (max), defuzzMethod (weighted average), impMethod (prod), and aggMethod (max).

For each model (ANFIS & LR), root mean squared error (RMSE) (Eq. 7), mean absolute error (MAE) (Eq. 8), determination coefficient ($R^2$) (Eq. 9), and correlation coefficient (R) (Eq. 10) were used as evaluation criteria.

$$RMSE = \left(\frac{\sum_{j=1}^{n}(Y_j - \hat{Y}_j)^2}{n}\right)^{1/2} \quad \ldots(7)$$

$$MAE = \frac{\sum_{j=1}^{n}|Y_j - \hat{Y}_j|}{n} \quad \ldots(8)$$

$$R^2 = 1 - \frac{\sum_{j=1}^{n}(Y_j - \hat{Y}_j)^2}{\sum_{j=1}^{n}(Y_j - \bar{Y})^2} \quad \ldots(9)$$

$$R = \frac{\sum_{j=1}^{n}\left[(Y_j - \bar{Y})(\hat{Y}_j - \bar{\hat{Y}})\right]}{\left[\sum_{j=1}^{n}(Y_j - \bar{Y})^2 \sum_{j=1}^{n}(\hat{Y}_j - \bar{\hat{Y}})^2\right]^{1/2}} \quad \ldots(10)$$

Where:

$Y$ *and* $\hat{Y}$ are the observed and estimated values respectively and $n$ is the number of observations. $\bar{Y}$ *and* $\bar{\hat{Y}}$ are mean of observed and estimated values.

The RMSE shows the goodness of fit relevant to high values whereas the MAE measures a more balanced perspective of goodness of fit at moderate value. The $R^2$ shows the degree to which two variables are linearly related [15]. R is a measure of how well trends in the predicted value follow trends in past actual value [16].

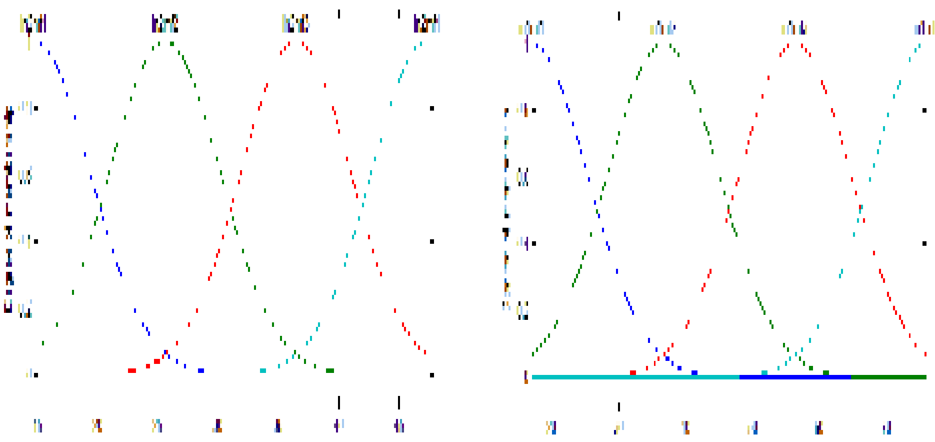


**Figure (3) The shape of membership function of temperature (T) and relative humidity (RH).**

**Results and Discussion**

Two models are used for estimating monthly pan evaporation in Basrah City. The first model is adopted by using ANFIS models, while the last model is adopted by using linear regression LR model. These models are described as follow:

*Model No.1:*

*a, E = f (T)……. ANFIS*

*b, E = f (T, RH)………ANFIS*

*Model No.2:*

*E = f (T)……… LR*

Based on the idea, that the measuring of temperature is very easy, therefore, this variable is adopted as an input for estimating the amount of evaporation by using ANFIS and LR models. The second ANFIS model (b) is developed by introducing another input variable (relative humidity) in order to study the effect of this variable on the results. Table (1) gives the tabulation of performance of the different models. The performances indicators throughout the study are root mean squared error (RMSE), mean absolute error (MAE), determination coefficient ($R^2$), and correlation coefficient (R) over the checking period.

**Table (1) The results of models computed over the checking period**

| Model No. | Description | RMSE | MAE | $R^2$ | R |
|---|---|---|---|---|---|
| 1 | a, ANFIS | 64.126 | 48.402 | 0.905 | 0.986 |
| | **b, ANFIS** | **63.597** | **49.16** | **0.929** | **0.987** |
| 2 | LR | 64.42 | 51.7 | 0.879 | 0.938 |

Obviously, estimating the monthly pan evaporation accuracy is significantly improved when using the ANFIS models compared with the LR model. It can be depicted that accuracy improvement about 5% in R and $R^2$. The ability of the ANFIS model to accurately forecast monthly pan evaporation might be attributed to the method by which the ANFIS model recognizes the input variables. The use of fuzzy membership function to represent the different classes of temperature and relative humidity data allows considering the input data as a range rather than as a crisp input. Figures 4 to 6 present the details of the observed and calculated monthly pan evaporation and their corresponding scatter plots for the developed models.

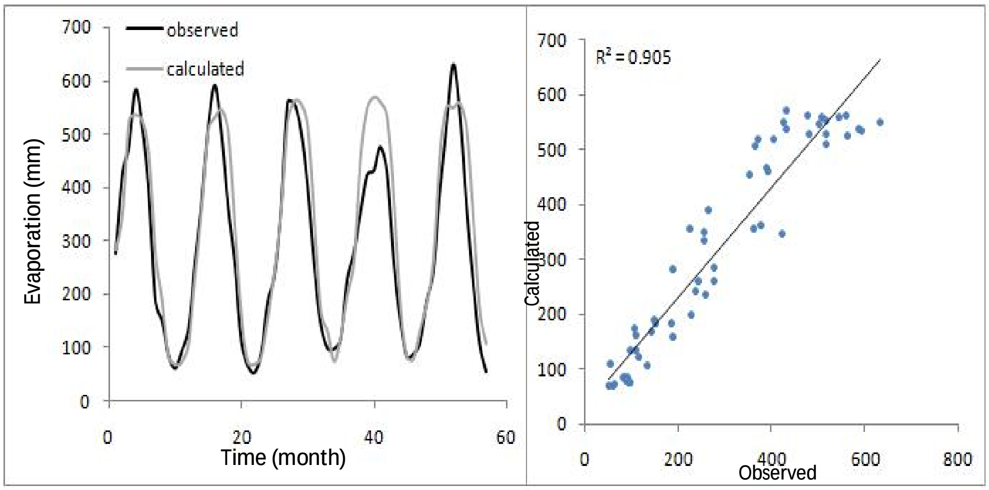


**Figure (4) Comparative plots of observed and calculated monthly pan evaporation and their corresponding scatter plot for model No. 1 (a) over the checking period.**

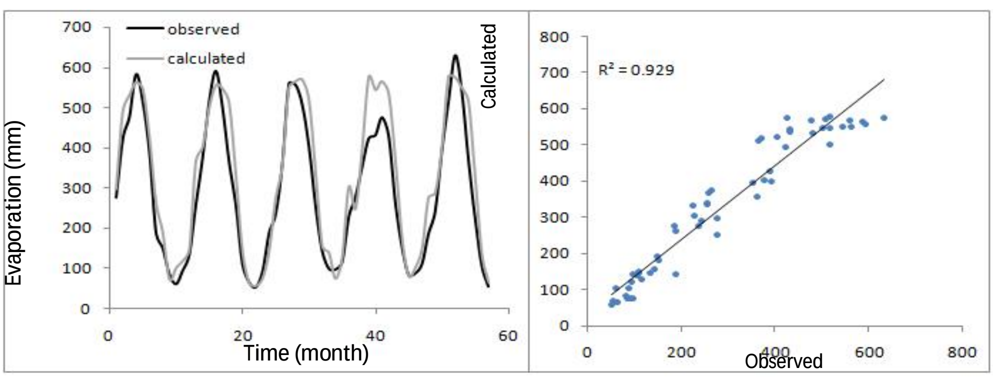


**Figure (5) Comparative plots of observed and calculated monthly pan evaporation and their corresponding scatter plot for model No. 1 (b) over the checking period.**

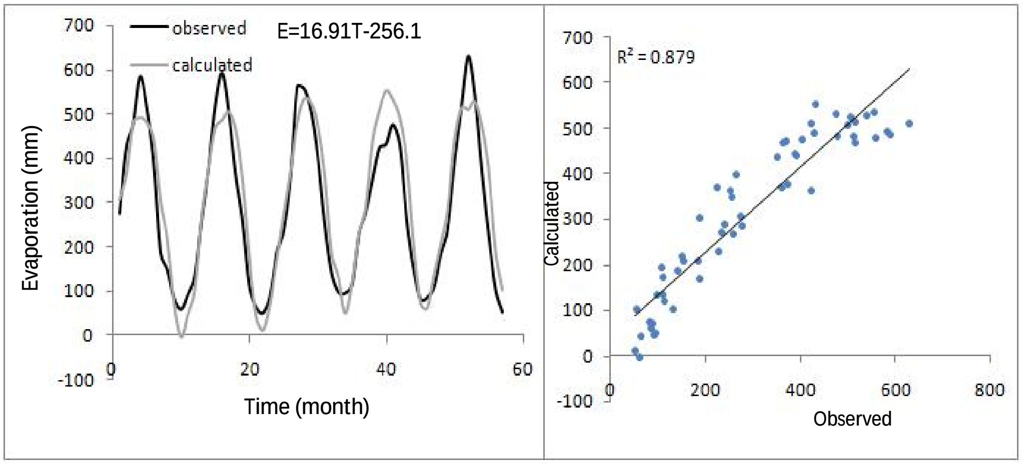


**Figure (6) Comparative plots of observed and calculated monthly pan evaporation and their corresponding scatter plot for model No. 2.**

## CONCLUSIONS

In the present study, monthly pan evaporation was estimated by using two different models (ANFIS & LR). The ANFIS model is used twice, in the first one, the temperature is used as input data only, and in the second one, the temperature and relative humidity are used as input data for predicting the evaporation. A verification test is added to check the model correctness by matching the calculated evaporation with the once observed. The comparison reveals that the ANFIS models give better accuracy in estimation of monthly pan evaporation than the linear regression model; the accuracy is improved about 5% in R and $R^2$. Based on the idea, that the measuring of temperature is very easy and low cost, it could be used as input data for estimation pan evaporation. This model could provide a check on doubtful measurements.

## REFERENCES

[1]. Terzi, O., Erol Keskin, M., "Evaporation estimation using gene expression programming". J. of applied science 5(3), 508-512, 2005.

[2]. Karamouz M., Nazif S., Falahi M., Hydrology and Hyroclimatology Principles and Application, Taylor and Francis Group, LLc, 2013.

[3]. Sethuraman G., Naidu S., International Encyclopedia of Agricultural Science and Technology, Kishan Mittal for Mittal Publication, 2008.

[4]. Farnsworth R.K., Thompson E.S., "Mean monthly, seasonal, and annual evaporation for the United States". NOAA Technical Report NWS 34, Washington, DC;. 82p, 1982.

[5]. Moghaddamnia, A., Gousheh, M. G., Piri, J., Amin., S., Han, D., "Evaporation estimation using artificial neural networks and adaptive neuro-fuzzy inference system techniques". Advance in Water Resources 32: 88-97, 2009.

[6]. Shirsath, P.B., Singh, A.K., "A comparative study of daily pan evaporation estimation using ANN, regression and climate based model". Water Resou. Manage 24:1571-1581, 2010.

[7]. Kumar, P., D. Kumar, Jaipaul , Tiwari A. K., "Evaporation estimation using artificial neural networks and adaptive neuro-fuzzy inference system techniques. Pakistan Journal of Meteorology. Vol. 8, Issue 16, 2012.

[8]. Kisi, O. "Evolutionary neural networks for monthly pan evaporation modeling". J. of Hydrology. 498:36-45, 2013.

[9]. Jang, J.S.R., Sun, C.T.. Neuro-Fuzzy and Soft Computing: A computational Approach to Learning and Machine Intelligence. Prentice Hall, Englewood Cliffs, New jersey. 1997.

[10]. Jang, J.S.R. "ANFIS: adaptive-network-based fuzzy inference system". IEEE Trans Syst Man Cybern 23 (3):665-685, 1993.

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

[12]. Harper H.L. "The method of least squares and some alternatives". Part I, II, II, IV, V, VI. International Satistical Review, 42, 147–174; 42, 235–264; 43, 1–44; 43, 125–190; 43, 269–272; 44, 113–159, (1974–1976).

[13]. Paraskevi T. Z, "An Adaptive Neuro-fuzzy Inference System for Robot Handling Fabrics with Curved Edges towards Sewing". J. Intell Robot Syst, DOI 10.1007, 2009.

[14]. Wafaa El-Shahat, Alaa H. I. H. "Permeability and porosity prediction from wireline logs using neuro-fuzzy technique. Ozean Journal of Applied Sciences 3(1), 2010.

[15]. Karunanithi, N., Grenney, W.T., Whitley, D., Bovee, K., "Neural networks for river flow prediction". J Comput Civil Eng ASCE 8(2):201-220, 1994.

[16]. Dinesh, C.S., Jangid, A., "Discharge modeling using adaptive neuro-fuzzy inference system". International journal of advanced science and technology. Vol. 31: 99-114, 2011.