JOURNAL OF CURRENT RESEARCH IN SCIENCE (ISSN 2322-5009)
CODEN (USA): JCRSDJ
2014, Vol. 2, No. 3, pp: 340-345
Available at www.jcrs010.com

ORIGINAL ARTICLE

COMPARING THE RESULTS OF AN EMPIRICAL EQUATION AND ARTIFICIAL INTELLIGENCE TO CALCULATE EVAPOTRANSPIRATION AND COMPARE THEM WITH THE ACTUAL RESULTS (CASE STUDY: *KAVAR* AND *DOROUDZAN* WEATHER STATIONS)

Mahmoud Tabatabaee[1], Mosayeb Mirshekari[2], Zeinab Sheikh Alipour*[2], Reza Tahmasebizade[2]
1- Assistant Professor, Faculty of Water and Soil, University of Zabol, Zabol, Iran
2- M.Sc. Student in Irrigation and Drainage Department, University of Zabol, Zabol, Iran

ABSTRACT: Evapotranspiration is one of the most fundamental parameters in the calculation of crop water requirement. The parameter is calculated to avoid imposing water stress to the plant in different stages of its growth. In this paper, meteorological data including relative humidity, minimum, maximum and mean values of temperature, sunshine hours and evaporation from pan evaporation measurements from *Kavar* weather stations in the *Maharlu* Basin and *Doroudzan* station in the *Kor* River Basin from 1991 to 2011 were used. Calculation has been done by using a combination of experimental methods including FAO-Penman-Monteith, Blaney Criddle and Hargreaves and intelligent techniques including ANFIS and GEP. The results showed that Hargreaves method with $R^2 = 1$ and RMSE = 0.024 in the *Kavar* weather station and with $R^2 = 1$ and RMSE = 0.011 at *Doroudzan* weather station is the best results at estimating evapotranspiration. According to the statistical indicators and analyzing results obtained from this study it observed that combined methods of artificial intelligence methods have better results than empirical methods. In general, for all the experimental methods using ANFIS better performs in estimating evapotranspiration than using empirical relationships and GEP method.
KEYWORDS: ANFIS, *Doroudzan*, Empirical Equation, Evapotranspiration, GEP, *Kavar*.

INTRODUCTION

Evapotranspiration shows water loss from the surface that is down by two separate processes (Evaporation from the soil surface and transpiration by plants). Accurate determination of water requirements of plants is the first and most important step to optimize water use in irrigation. This is usually done by calculating the potential evapotranspiration and then multiplication result by the crop coefficient (Alizadeh, 2008). Shuttleworth, (1993) recommends *Hargreaves* method for areas where some data are not available and *Penman–FAO* method is not possible to use. *FAO* has recommended that parameters provided by Allen *et al.*, (1998) in *Penman-Monteith* equation be used for the calculation of reference evapotranspiration.

Keskin *et al.*, (2004) used data from the weather station near the *West Lake Egirdir Turkey* in a study, and compared the results of evaporation from evaporation pan and the estimated value of the fuzzy model, obtained R2=0.85 and RMSE=6.1.

Also Kisi, (2006) after estimating evaporation of pan evaporation by using ANN and ANFIS and by employing experimental results of *Stephen Stewart*, concluded that ANFIS estimates better results. Rosenberry *et al.*, (2007) in a small mountain lake in *Eastern North America* by comparing 15 methods of calculating evapotranspiration during six agricultural season concluded that in calculating evaporation, methods base on air temperature measuring or base on temperature and solar radiation have better results than the methods require a large number of variables.

Xu and Singh, (2001) compared seven equations based on temperature for determining evaporation in two lake station in *North West Ontario, Canada*; and concluded that by calibration fix parameters for these regions, in all months *Modified Blaney-Criddle Method* with minimal error, and then *Hargreaves* and *Thornthwaite* methods have best results in estimation of evaporation.

The purpose of this paper is to compare results of estimating evapotranspiration with *Blaney-Criddle* (BC), *Hargreaves-Samani* (HS) and artificial intelligence techniques including ANFIS and GEP, and finally compare their results with real data to determine best method for estimating evapotranspiration using standard *FAO-Penman-Monteith* (FPM) method in *Kavar* and *Doroudzan.*

MATERIALS AND METHODS

2.1. The study area

Corresponding Author: Zeinab Sheikh Alipour, M.Sc. Student in Irrigation and Drainage Department, University of Zabol, Zabol, Iran. **E-mail:** z_sh_ir@yahoo.com

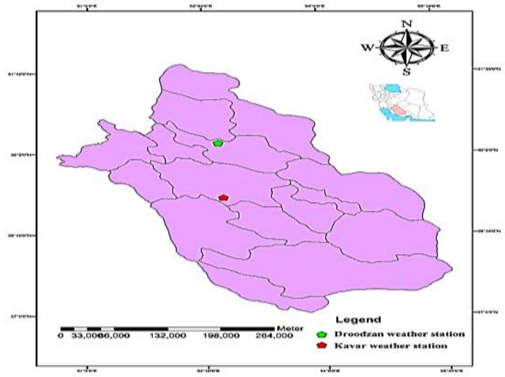

Figure 1: Map of study area in south of Iran

Kavar city in *Fars Province* of *Iran*, is located at a distance of 45 km south-east of *Shiraz* (provincial capital of *Fars*). *Kavar* weather station is located on *Maharlu Lake* basin in *Shiraz* plain. This station is 1639 m above sea level with 480 mm average annual rainfall, have mediterranean climate. Also *Doroudzan* area in *Fars Province* of *Iran* is located at *Marvdasht* plain in *Kor* river basin and at a distance of 100 kilometers from *Shiraz* (Figure 1). This area is 1625 m above sea level with 470 mm average annual rainfall, have the mediterranean climate too. In this study, the data of 21 years from 1991 to 2011 in mentioned weather stations are used.

2.2. FAO Penman-Monteith (FPM)

In this study FPM equation which is recommended as the standard equation to calculate evapotranspiration by *FAO* is used and the equation is:

$$ET_0 = \frac{0.408\Delta(R_n - G) + \gamma[\frac{890}{T + 273}]U_2(e_a - e_d)}{\Delta + \gamma(1 + 0.34U_2)} \quad (1)$$

Where is Slope of the vapor pressure curve, R_n is net radiation at the surface of vegetation cover, G is heat flux into the soil, is moisture coefficient, e_a and e_d are saturation vapor pressure and actual vapor pressure, respectively. T and U_2 is air temperature and wind speed at a height of 2 m above ground surface, respectively. Heat flux into soil is 0 (G = 0) with regard to standard weather station, which has full coverage (Alizadeh, 2008).

2.3. Hargreaves-Samani method (HS)

Hargreaves-Samani method can be used for calculating evapotranspiration in one or several-day periods, and weekly or ten-day periods, though Weekly and ten-day periods have the best results:

$$ET_0 = 0.0023R_a\,(T + 17.8)\sqrt{(T_{max} - T_{min})} \quad (2)$$

Where, T is mean temperature, T_{max} is maximum temperature, T_{min} is minimum temperature and R_a is solar radiation which depends on latitude and month of year.

2.4. Blaney-Criddle Method (BC)

In Blaney-Criddle Method potential evapotranspiration is calculated from the following experimental formula:

$$ET_0 = a + b[p(0.46T + 8.13)] \quad (3)$$

Where, p is coefficient of day duration, a and b are climatic factors depending on the relative humidity, wind speed and ratio of actual sunshine hours to maximum possible sunshine hours. Values of p (coefficient of day duration) and maximum possible sunshine hours, according to the station's location (latitude of 29^oN) obtained from tables for different months (Alizadeh, 2008).

2.5. Artificial Intelligence Methods

2.5.1. Adaptive Neuro–Fuzzy Inference System (ANFIS)

Neuro-Fuzzy networks, by modeling the operation of human brain system, processing experimental data and without regard to physics of the problem, hidden rule among project data. The important point in Fuzzy logic is possibility of communicating between input space with output space; and the primary mechanism for doing this, is a list of If-Then statements which are called rule. In training process, these rules are evaluated and determined parallely. Jang, (1993) for the first time, by considering abilities of Fuzzy theory and neural network, offered ANFIS.

ANFIS is a multi-layer network, consisting of nodes and arcs which connect nodes. As a simple example, a fuzzy inference system with two inputs x and y and one output z is assumed. First order sugeno fuzzy model is a typical rule that with two if - then rules is paired fuzzy and can be expressed as:

If x is A_1 and y is B_1 THEN $F_1=p_1x+q_1y+r_1$ (4)
If x is A_2 and y is B_2 THEN $F_2=p_2x+q_2y+r_2$ (5)

The corresponding ANFIS structure is shown in figure 2. Nodes in the same layer have similar functions. Output of the i_{th} node in layer 1 is specified with $O_{1,i}$.

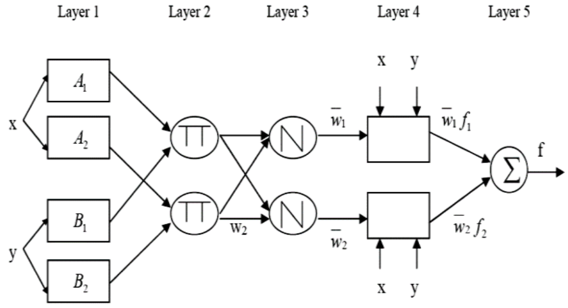

Figure 2: Diagram of an ANFIS with triangular membership function

Layer 1: Every i node in this layer, is a node matches with node function:

$$O_{1,i} = \mu_{A_i}(x) \quad \text{for} \quad i = 1,2 \tag{6}$$

$$O_{1,i} = \mu_{B_i - 2}(y) \quad \text{for} \quad i = 3,4 \tag{7}$$

Where, x (or y) is the input of node i, and A_i (or B_i) is a linguistic variable (such as low or high) compromising with this node.

In other words, $O_{1,i}$ is membership degree of a fuzzy set A = (A1, A2, B1 or B2) and defines the rating allocated to the input x (or y) that assigns to quantifier A. Membership functions for A and B usually are described and generalizations by bell functions. For example:

$$\mu_A(x) = \frac{1}{1 + \left|\frac{x - c_i}{a_i}\right|^{2b_i}} \tag{8}$$

That {ai,bi,ci} is set of parameters. When parameters values change, bell function will change accordingly, so different types of membership functions are displayed in linguistic variable Ai. In fact, every continuous and recognizable piecewise function such as triangular membership functions is qualified candidates for node functions in this layer. Parameters in this layer are called assumptive parameters. The output of these layers is hypothetical values.

Layer 2: This layer consists of nodes labeled with Π, which divides input signals and sends as the product out. For example:

$$O_{2,i} = w_i = \mu_{A_i}(x)\mu_{B_i}(y) \quad \text{for} \quad i = 1,2 \tag{9}$$

Each output node is represents the excitation intensity of a rule.

Layer 3: In this layer, nodes that are identified with label N, calculate the excitation intensity ratio of the i^{th} rule, per sum of excitation intensities of all rules:

$$O_{3,i} = \overline{w}_i = \frac{w_i}{w_1 + w_2} \tag{10}$$

The outputs of this layer are called normalized excitation intensities.

Layer 4: The nodes in this layer are corresponding node functions:

$$O_{4,i} = \overline{w}_i f_i = w_i(p_i x + q_i y + r_i) \tag{11}$$

Where, is output of third layer and {pi,qi,ri} is set of parameters. Parameters in this layer are called Inference parameters.

Layer 5: The single node in this layer is a fixed node called Σ, which computes all outputs as sum of all input signals:

$$O_{5,i} = \sum_i \overline{w}_i f_i = \frac{\sum_i w_i f_i}{\sum_i w_i} \tag{12}$$

Therefore, an adaptive network that functionally is equivalent to a first order fuzzy inference system is created.

2.5.2. Gene Expression Programming (GEP)

Genetic programming is a model similar to genetic algorithm (Goldberg, 1989) with this difference that uses tree model in its result searching system. The first time Koza, (1994), presented genetic programming based on *Darwin's* theory. This method provides a technique for automatic programming to solve an issue like a computer program and it is capable of optimizing the model structure and its parameters. In genetic programming, tree structure is used, but genetic algorithm is based on a binary system. This is fundamental structural difference between these two systems. Use of this method is progressing rapidly in the issues that common methods of computational mathematics are not able to solve them or a clear relations between variables of problem is not available (Ferreira, 2001; Ferreira, 2004).

This method can easily be accountable for the following issues (Banzhaf *et al.*, 1998):

1) The relation between variables of problem is not well understood or validity of current understanding is suspected.

2) Finding size and shape of final solution is very difficult, and forms a major part of problem.

3) Common methods of mathematical analysis are not able to deliver analytical solution methods.
4) Approximate solution is acceptable.
5) Any partial improvement in operation will be measured on a regular basis and this improvement is valuable.
6) Number of data to be analyzed, classified and summarized by computer, is too many.
Gene expression programming was coined in 1999 by Ferreira, (2001). Gene expression programming uses a simple selection of elites and keeps best people of a generation for next generation. Gene expression programming unlike genetic algorithms and genetic programming has several genetic operators to proliferation individuals with modification (Lopez *et al.*, 2004). Tree expression is deduced by a process called translation of chromosomes, and its tree representation will determine its fit in environment and its corresponding chromosomes. While the chromosomes are selected by genetic operators and activities. (Wilson and Banzhaf, 2008).

2.6. Measurement of accuracy models

To evaluate estimation precision of each of models, calibrations have been done by using root mean squared errors (RMSE), mean bias error (MBE), and coefficient of correlation (R^2). Equations of RMSE, MBE and R^2 are represented below:

$$RMSE = \sqrt{\frac{1}{t}\sum_{i=1}^{t}(P_i - O_i)^2} \quad (13)$$

$$MBE = \frac{\sum_{i=1}^{t}(P_i - O_i)}{t} \quad (14)$$

$$R^2 = 1 - \frac{\sum_{i=1}^{t}(P_i - O_i)^2}{\sum_{i=1}^{t}(O_i - \bar{O}_i)} \quad (15)$$

In the above equations, P_i and O_i are observed and estimated values at time t, and i is number of data. RMSE value represents root mean square error between observed and predicted values, and how lower it is, estimated model have more accurately. MBE indicates either computational error is positive or negative. Whether positive or negative, MBE indicates that predicted values are higher or lower than observed values, respectively. R^2 value always varies between zero and one. Whatever it is closer to 1 indicate a better match between estimated and measured values.

RESULTS AND DISCUSSION

In this study, standard *FPM* was used as reference method. Evapotranspiration calculations at *Doroudzan* and *Kavar* stations in *Fars* province were conducted based on *HS* and *BC* models. Then models were modeled by employing parameters used in their equations, with ANFIS and GEP models. In this study data from 1991 to 2011 were used which included 179 data that 104 of them were used for training step and other 75 were used to test step.
The results obtained from each of relations and simulator model of relation, with ANFIS and GEP models were compared with results of *FPM* method. The statistical parameters R^2, RMSE, MBE were used to assess accuracy of models. Precision measurement of models, training and testing models are represented in Tables 1 to 4. As can be seen clearly in Tables 1 and 2, using ANFIS in all models, increases accuracy of models.

Table 1: Statistical indicators of training and testing steps of empirical–ANFIS evapotranspiration of *Kavar* region

	training			Testing		
Model	R^2	RMSE	MBE	R^2	RMSE	MBE
FPM-ANFIS	0.9977	4.36	0	0.87	0.7	-0.47
HS-ANFIS	1	0.02	42.77	1	0.55	0.1
BC-ANFIS	0.93	42.765	0	0.93	47.91	-25.02
HS-FPM	0.97	47.35	-43.61	0.98	35	-33
BC-FPM	0.97	126.62	105.87	0.98	131.38	109.38

Table 2: Statistical characteristics of training and testing steps empirical-ANFIS evapotranspiration of *Doroudzan* region

	training			Testing		
Model	R^2	RMSE	MBE	R^2	RMSE	MBE
FPM-ANFIS	0.997	3.87	0.008	0.86	12.57	1.49
HS-ANFIS	1	0.011	0.00	0.99	0.77	0.078
BC-ANFIS	0.912	40.74	0.00	0.94	38..76	18.39
HS-FPM	0.94	30.54	-23.35	0.95	26.78	-19
BC-FPM	0.96	96.25	76.28	0.96	87.78	68.68

Table 3: Statistical indicators of training and testing steps empirical-GEP evapotranspiration of *Kavar* region

	Training			Testing		
Model	R^2	RMSE	MBE	R^2	RMSE	MBE
FPM-GEP	0.97	13.26	10.62	0.95	19.28	15.14
HS-GEP	0.98	8.23	6.75	0.98	13.19	9.75
BC-GEP	0.92	44.66	29.64	0.93	46.42	37.24
HS-FPM	0.97	47.35	-43.61	0.98	35	-33.01
BC-FPM	0.97	126.62	105.87	0.98	131.38	109.94

Table 4: Statistical indicators of training and testing steps empirical-GEP evapotranspiration of *Doroudzan* region

	training			Testing		
Model	R^2	RMSE	MBE	R^2	RMSE	MBE
FPM-GEP	0.98	8.66	0.23	0.96	23.61	17.00
HS-GEP	0.98	3.38	0.90	0.99	4.12	-1.08
BC-GEP	0.91	40.81	0.27	0.93	40.45	18.28
HS-FPM	0.94	30.54	-23.35	0.95	26.78	-18.68
BC-FPM	0.96	96.25	76.28	0.96	87.78	68.68

According to results of *Kavar* region in Table 1, R^2 of *HS* method which modeled using ANFIS has highest value in both training and test step. Also lowest RMSE value obtained in this method. According to values of statistical indicators, first *HS* method which modeled using ANFIS and then *FPM* method modeled using ANFIS have best operation among other methods.

According to results for *Doroudzan* region in Table 2, it was noticed that R^2 for modeled HS method with ANFIS in both training and testing steps has reached highest value. Also lowest RMSE obtained in this method. According to values of statistical indicators, first *HS* method modeled using ANFIS has best performance among methods, and then *FPM* modeled with ANFIS method have best operation among other methods.

As can be seen clearly from Tables 3 and 4, using GEP in all models increases accuracy of models. According to results of *Kavar* region in Table 3, it observed that R^2 for modeled *HS* with GEP method, in both training and testing steps has highest value. Also lowest RMSE obtained in this method. According to values of statistical indicators, first *HS* modeled with GEP method has best operation among methods, and then *FPM* modeled with GEP method.

According to results of *Doroudzan* region in Table 4, it was observed that R^2 for *HS* modeled with GEP method and then *FPM* modeled with GEP method, in both training and testing steps have highest value. Also lowest RMSE obtained in *HS* which modeled with GEP method. According to values of statistical indicators, first *HS* modeled with GEP method has best performance among methods, and then FPM modeled with GEP method is in second step. Correlation between used models for best model are shown in Figures 3 to 6:

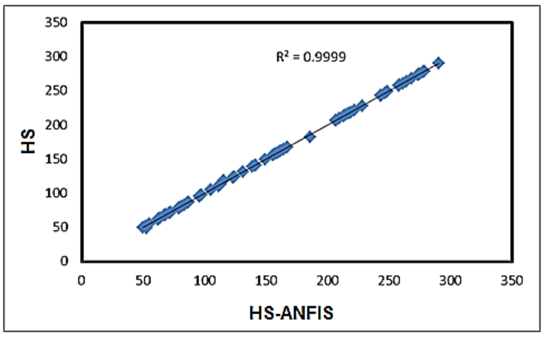

Figure 3: Distribution of *HS* modeled with ANFIS in *Doroudzan* station

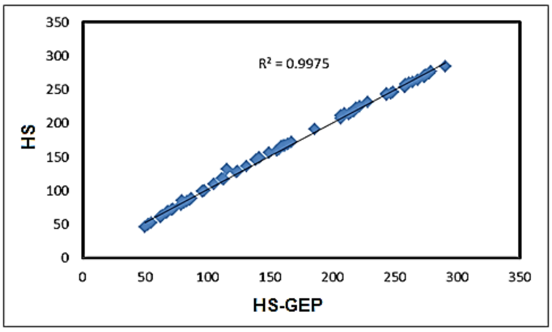

Figure 4: Distribution of *HS* modeled with GEP in *Doroudzan* station

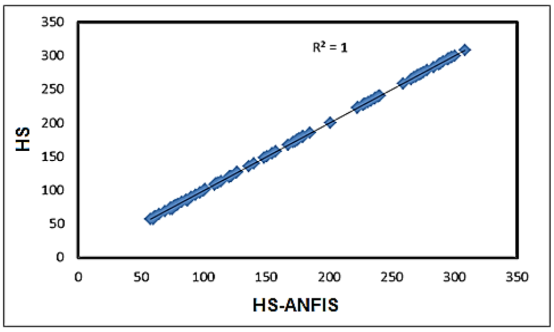

Figure 5: Distribution of HS modeled with ANFIS in *Kavar* station

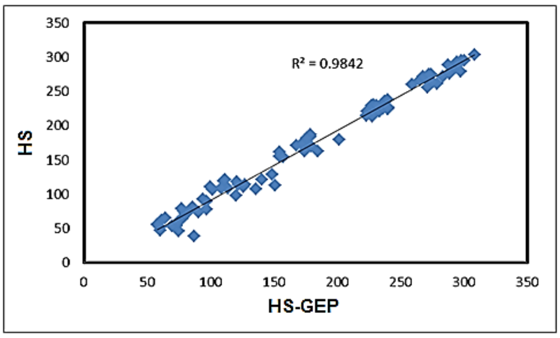

Figure 6: Distribution *HS* modeled with GEP in *Kavar* station

CONCLUSION

The overall results of this study, as some research has been done in the world shows to model evapotranspiration powerful, accurate with high ability method is ANFIS and also is more precise than empirical models. According to the obtained results, it is clearly seen that combined methods of artificial intelligence methods have better results than empirical methods. Also ANFIS have better performance than GEP method in evapotranspiration estimating using empirical equations. It is recommended for future studies, empirical models widely used in other studies to estimate evapotranspiration. Also performance of other intelligent models such as artificial neural networks, genetic algorithms and M5 can be tested.

REFERENCES

Alizadeh A. Principles of Applied Hydrology. 24th edition, Razavi Publications, 2008; pp:870.

Allen RG, Pereira LS, Raes D, Smith M. Crop Evapotranspiration Guidelines for Computing Crop Water Requirements. FAO Irrigation and Drainage Paper 1998;56:300.

Banzhaf W, Nordin P, Keller RE, Francone FD. Genetic Programming. Morgan Kaufmann, San Francisco, CA 1998.

Ferreira C. Gene Expression Programming and the Evolution of Computer Programs. Recent Developments in Biologically Inspired Computing 2004; pp:82-103.

Ferreira C. Gene expression programming: a new adaptive algorithm for solving problems. Complex Systems 2001; 13(2): 87-129.

Goldberg DE. Genetic algorithms in search, optimization, and machine learning. Addison–Wesley, Reading, Mass 1989.

Jang JR. ANFIS: Adaptive Network-Based Fuzzy Inference System. Proc IEEE Conf Trans Syst Man Cybern 1993;23:665-685.

Keskin M, Terzi O, Taylan D. Fuzzy logic model approaches to daily pan evaporation estimation in western Turkey. Journal of Hydrological Sciences 2004;49(6):1001-6.

Kisi O. Daily pan evaporation modelling using a neuro-fuzzy computing technique. Journal of Hydrology 2006;329:636 646.

Koza JR. Genetic programming as a means for programming computers by natural selection. Statistics and Computing 1994;4: 87-112.

Lopez EG, Poli R, Coello Coello CA. Reusing Code in Genetic Programming. Genetic Programming, 7th European Conference 2004;3003:359-368.

Rosenberry DO, Winter TC, Buso DC, Likens GE. Comparison of 15 evaporation methods applied to a small mountain lake in the northeastern USA. Journal of Hydrology 2007;340:149–166.

Shuttleworth WJ. Evaporation. In: Maidment DR (Ed.), Handbook of Hydrology. McGraw Hill, New York 1993.

Wilson GC, Banzhaf W. A Comparison of Cartesian Genetic Programming and Linear Genetic Programming. Proceedings of the 11th European Conference on Genetic Programming 2008;4971:182-193.

Xu CY, Singh VP. Evaluation and generalization of temperature-based methods for calculating evaporation. Hydrol Process 2001; 15: 305–319.