Feedforward Backpropagation, Genetic Algorithm Approaches for Predicting Reference Evapotranspiration
(Perambatan balik Maklum Balas ke Depan, Pendekatan Algoritma Genetik untuk Meramalkan Rujukan Penyejatpeluhan)

SHAFIKA SULTAN ABDULLAH*, M.A. MALEK, NAMIQ SULTAN ABDULLAH & A. MUSTAPHA

ABSTRACT

Water scarcity is a global concern, as the demand for water is increasing tremendously and poor management of water resources will accelerates dramatically the depletion of available water. The precise prediction of evapotranspiration (ET), that consumes almost 100% of the supplied irrigation water, is one of the goals that should be adopted in order to avoid more squandering of water especially in arid and semiarid regions. The capabilities of feedforward backpropagation neural networks (FFBP) in predicting reference evapotranspiration (ET_0) are evaluated in this paper in comparison with the empirical FAO Penman-Monteith (P-M) equation, later a model of FFBP+Genetic Algorithm (GA) is implemented for the same evaluation purpose. The study location is the main station in Iraq, namely Baghdad Station. Records of weather variables from the related meteorological station, including monthly mean records of maximum air temperature (T_{max}), minimum air temperature (T_{min}), sunshine hours (R_n), relative humidity (R_h) and wind speed (U_2), from the related meteorological station are used in the prediction of ET_0 values. The performance of both simulation models were evaluated using statistical coefficients such as the root of mean squared error (RMSE), mean absolute error (MAE) and coefficient of determination (R^2). The results of both models are promising, however the hybrid model shows higher efficiency in predicting ET_0 and could be recommended for modeling of ET_0 in arid and semiarid regions.

Keywords: Evapotranspiration; genetic algorithm; neural networks; Penman-Monteith

ABSTRAK

Kekurangan air adalah satu kebimbangan global, kerana permintaan bekalan air semakin bertambah dan pengurusan sumber air yang lemah akan secara dramatik mempercepatkan pengurangan air sedia ada. Jangkaan tepat untuk penyejatpeluhan (ET), yang menggunakan hampir 100% daripada bekalan air pengairan merupakan salah satu matlamat yang perlu diterima pakai bagi mengelakkan lebih banyak pembaziran air terutamanya di kawasan-kawasan gersang dan separa gersang. Keupayaan rangkaian neural perambatan balik maklum balas ke depan (FFBP) untuk meramalkan rujukan penyejatpeluhan (ET_0) dinilai dalam kertas ini berbanding dengan persamaan empirikal FAO Penman-Monteith (P-M), kemudian model FFBP + genetik algoritma (GA) dijalankan bagi tujuan penilaian yang sama. Lokasi kajian ialah stesen utama di Iraq, iaitu stesen Baghdad. Rekod pemboleh ubah cuaca dari stesen kajicuaca berkaitan, termasuk rekod bulanan purata suhu udara yang maksimum (T_{max}), suhu udara minimum (T_{min}), jam cahaya matahari (R_n), kelembapan relatif (R_h) dan kelajuan angin (U_2) dari stesen kajicuaca berkaitan digunakan dalam jangkaan untuk nilai ET_0. Prestasi kedua-dua model simulasi dianalisis menggunakan pekali statistik seperti punca min ralat kuasa dua (RMSE), min ralat mutlak (MAE) dan pekali penentuan (R^2). Keputusan kedua-dua model adalah menggalakkan. Walau bagaimanapun model hibrid menunjukkan kecekapan yang lebih tinggi dalam meramalkan ET_0 dan boleh disyorkan untuk pemodelan ET_0 di kawasan gersang dan separa gersang.

Kata kunci: Algoritma genetik; Penman-Monteith; penyejatpeluhan; rangkaian neural

INTRODUCTION

Water scarcity is mostly attributed to improper means of water usage in many parts of the world. Iraq, which consists of both arid and semiarid climates, consumes most of the available water in irrigation projects. Hence, an understanding of the accurate prediction of evapotranspiration, which represent the highest percentage of water consumption, would allow for optimization of water use in irrigation projects.

There are several conventional methods for the prediction of ET, these methods are either direct or indirect measurements (Kumar et al. 2010; Landeras et al. 2008), but the Food and Agriculture Organization (FAO) of the United Nations proposed that the combination of energy balance/aerodynamic equations provides the most accurate results for prediction of ET_0, and adopted the FAO Penman-Monteith (P-M) equation as the only standard equation for estimation of ET_0 (Allen et al. 1998).

Artificial Neural Networks (ANNs) are programming techniques that represent a resemblance to the performance of biological brain nerve cells (neurons) and the way they interact (Rumelhart & McClelland 1986; Zurada 1992). The ANNs applications have proved to be a success for the last couple of decades; therefore ANNs are considered now as a common executive means of handling various scientific tasks such as: Bioinformatics (Wang et al. 2008); pattern recognition and forecasting predictions (Basheer & Hajmeer 2000; Warner & Misra 1996). The principal pattern in the performance of ANNs-based models, in general, could be divided into training, validation and testing, respectively (Bishop 1994). The literature cited (Kiranyaz et al. 2009) indicates that the testing performance is considered as the major evaluation procedure for determining the efficiency of the ANNs model in meeting the predefined objectives.

Pursuing the current research status showed a widespread popularity of artificial intelligence (AI) in the scientific research; however, it still attracts researchers in almost all fields and hydrologists are not exceptions, they have investigated the efficiency of ANNs in modeling reservoirs sedimentations, ET, runoff analysis, stream flow and rainfall (Gautam et al. 2000). The application of multiple layers perception (MLP); radial-based function network (RBFN); generalized regression neural network (GRNN) of ANNs provided a reliable estimations of ET_0 (Kisi 2006; Kumar et al. 2002).

The results of adaptive neurofuzzy inference system models (ANFIS) in modeling ET_0, compared by (Kisi & Öztürk) in 2008 to ANNs and other empirical methods emphasize the superiority of ANFIS with inputs of all required weather parameters over the ANNs and empirical methods.

Examining the accuracy of triple ANN techniques, namely, MLP and radial basis neural networks (RBNNs), the generalized regression (GR), in a model of P-M ET_0 indicated that both RBNN and MLP can be successfully used in modeling ET_0 (Kisi 2008).

Landeras et al. (2008) investigated the results of Seven ANNs techniques consisting of different inputs in comparison with results of ten empirical and semi-empirical ET_0 empirical equations, using meteorological data as inputs and calibrated to traditional FAO P-M. The statistical error techniques are used as comparison criteria. ANNs techniques obtained the most efficient results among other calibrated equations. El-Baroudy et al. (2010) compared the evolutionary polynomial regression (EPR) to both ANNs and genetic programming (GP). The EPR model provided a comparable performance to that of ANN and GP.

The main obstacle in the application of FAO P-M equation, as identified by (Tabari & Talaee 2012), is the wide range of meteorological data required for the estimation of ET_0. The authors used a multi-layer neural network (MLNN) for modeling ET_0 in Hamedan in Iran, with different input sets of data. The model with complete set of climatic parameters performed the best among the rest of MLNN models with other incomplete sets.

Khoshhal and Mokarram (2012) evaluated the performance of 10 ANN models, compared with P-M equation, in predicting ET_0 values using meteorological records of Eghlid station in Iran. The model with inputs of minimum air temperature (T_{min}), maximum air temperature (T_{max}), relative humidity (R_h), sunshine duration (R_n) and wind speed (U_2) data produced the nearest ET_0 values to results of P-M equation.

This study investigates the performance of the proposed AI techniques in predicting ET_0; the applied techniques in this study are the feedforward backpropagation neural networks (FFBP) and the FFBP hybridized with genetic algorithm (FFBP+GA), in comparison with values obtained using the empirical P-M Equation with historical records of weather parameters from Baghdad meteorological station in Iraq. The root of mean squared error (RMSE), mean absolute error (MAE) and coefficient of determination (R^2) are the statistical coefficients used to evaluate the performance of FFBP and FFBP+GA models.

STUDY AREA

The study area covers Baghdad District, which represents central part of Iraq as shown in Figure 1. Baghdad Station (code No. 650) is located between latitude 33° 18' 00" N, longitude 44° 24' 00" E and at altitude of 31.7 m above sea level. The weather data consists of 14 years (2000-2013) of monthly means of Tmax, Tmin, Rh, Rn and U_2 of this meteorological station (Iraqi Meteorological Organization and Seismology 2014).

FIGURE 1. Study location in Iraq

APPLIED TECHNIQUES IN PREDICTING REFERENCE EVAPOTRANSPIRATION PENMAN-MONTEITH EMPIRICAL EQUATION

The accurate determination of ET_0 is a fundamental requirement in operating irrigation projects. Where the distribution and adequacy of water supply depend entirely on previous estimates of ET_0.

FAO paper 56, issued by Allen et al. (1998), presented the FAO P-M equation as the globally accepted equation

adopted by the organization for the prediction of ET_0. The weather variables are supposed to be the only elements affecting the ET_0 predictions, thus, ET_0 values could be obtained using spatial weather records of certain locations. The equation in the FAO P-M form shows all weather parameters that influence the evapotranspiration process.

$$ET_0 = \frac{0.408\Delta\left(R_n - G\right) + \gamma\dfrac{900}{T + 273}u_2\left(e_s - e_a\right)}{\Delta + \gamma\left(1 + 0.34u_2\right)}. \qquad (1)$$

ET_0 is the reference evapotranspiration (mm/day), Rn is the net radiation (MJ $m^{-2}day^{-1}$), G is the density of soil heat flux (MJ $m^{-2}day^{-1}$), ϒ is the psychometric constant (kPa °C^{-1}), U_2 is the wind speed at height 2 m (m s^{-1}), (e_s-e_a) = deficit in saturation vapor pressure (kPa), T is the average daily air temperature (°C), and Δ is the slope of vapor pressure curve (kPa °C^{-1}).

The geographical characteristics (altitude and latitude) are essential for the calculations of radiation, daylight hours (N) or for adjustment of some weather elements for the local averages of atmospheric pressure values.

Standard climate data, including Tmax, Tmin, Rh, R and U_2, over a period of 14 years were collected, processed and organized in monthly averages, then applied in P-M equation for estimation of ET_0.

FEEDFORWARD BACKPROPAGATION NEURAL NETWORKS

ANNs are considered an efficient attempts to reproduce some of the abilities and skills of human brain by artificial techniques. The typical neural network process starts with receiving data; learning and adapting; pattern recognition; and then targeting the required function by testing different weights and effects of each of the data elements in a computational model. A conventional ANNs model consists of input layer that contains the input data; the hidden layers, with number of neurons that should satisfy the simulation conditions, would organize the data and recognize patterns through several trial processes to predict the targeted output; and the output layer (El-Baroudy et al. 2010).

The supervised FFBP is investigated in this study for the prediction of ET_0. In the FFBP, the network model has an input layer, one hidden layer with number of neurons that act in parallel as a summing and nonlinear mapping junctions (Zurada 1992) and an output layer, using one hidden layer to represent the nonlinear relationship of ET_0 is sufficient (Kumar et al. 2002). The first 70% of total data are used for training the model; the rest of the data records are used in testing the performance of this model. The backpropagation algorithm is then employed to activate the ten (10) neurons in the hidden layer for weighing and training of information elements to achieve the desired target (Badde et al. 2013). The neurons are connected such as each neuron in a certain neural network layer is connected to neurons of the previous layer; the best number of neurons was determined by using the trial-and-error methods with the application of 3, 5, 9, 10, 15 and 20 neurons; the architecture of the FFBP model is shown
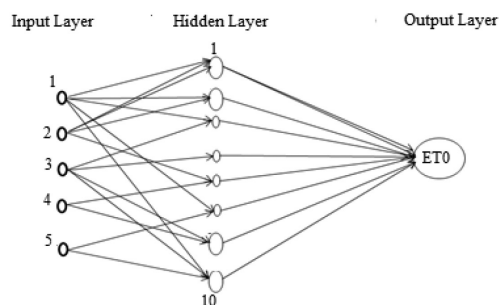

FIGURE 2. The feedforward backpropagation neural networks architecture for prediction of P-M ET_0

in Figure 2.

The backpropagation learning algorithm is a gradient descent technique used to establish arbitrarily chosen weights in the supervised FFBP, these weights are to be continuously adjusted by propagation from the output layer to input layer during the training process until the boundary signal falls below a predetermined threshold (Huang et al. 2006). The function Min and Max (premnmx, postmnmx, tramnmx) is applied to standardize the data within a range of [-1, 1] (Demuth et al. 2009). The MATLAB software is employed for implementation of ANNs model.

$$XN = (X - MinX) / (MaxX - MinX) * 2 - 1. \qquad (2)$$

The FFBP model has also been upgraded to prepare it for making predictions and then to predict ET_0 for any time interval depending on new climatic data that have not been employed previously in training or testing the model.

GENETIC ALGORITHM

The genetic algorithm (GA) is a probabilistic optimization method based on evolutionary theory. There are three major stages in evolution from one generation to another: Selection of strings; crossover of strings; and random mutation of the selected strings to produce the new generation (Azzini 2005). The same records of historical climatic data used in ANNs model are used in this part of study in a hybridized model of FFBP with Binary GA (FFBP+GA). The GA is applied to improve and optimize the FFBP performance by varying the number of neurons and to increase the number of hidden layers for maximum three (3) instead of only one (1) layer initially used in the ANNs method.

The standard performance evaluation functions are the statistical parameters RMSE, MAE and R^2, the range of R^2 varies between 0 and 1 (Draper & Smith 1998). The linear regression equation y = ax + b0 is applied for evaluating the performance of ET_0 predictions, where y is the dependent variable (Model ET_0); x is the independent variable (P-M ET_0); a is the slope; and b is the intercept (Khoshhal & Mokarram 2012).

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}\left(ETi_{calculated} - ETi_{predicted}\right)^2}{N}}. \qquad (3)$$

$$MAE = \frac{1}{N} \sum_{i=1}^{N} \left| ETi_{calculated} - ETi_{predicted} \right| 100. \qquad (4)$$

$$R^2 = 1 - \frac{\sum \left(ET_{predicted} - ET_{calculated} \right)^2}{\sum \left(ET_{calculated} - mean\ ET_{predicted} \right)^2}. \qquad (5)$$

RESULTS AND DISCUSSION

The complicated impact of meteorological parameters and geographical characteristics of the study location, represented by latitude and altitude, on the prediction of ET_0; inspired this study to investigate the capabilities of AI in producing reliable ET_0 values. The mean values of weather parameters for the period 2000-2013 are employed in the empirical P-M equation for the calculation of average monthly ET_0 as shown in Figure 3. The illustrated ET_0 value indicate the inequality of the ET_0 rates at different season of the year. The design of the irrigation systems should accommodate peak values, which starts from mid-May and extends to mid of September, to satisfy the minimum supply of water according to its monthly consumptive use.

The main identified hindrance in the application of P-M equation is the comprehensive weather data required in the estimation process (Droogers & Allen 2002; Shiri et al. 2012as a major component of the hydrological cycle, is of importance for water resources management and
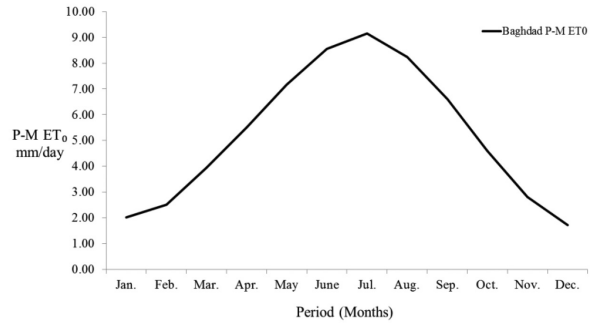

FIGURE 3. Average monthly P-M ET_0 at Baghdad station for the period of 2000-2013

development, as well as for estimating the water budget of irrigation schemes. This study presents a Gene Expression Programming (GEP; Tabari & Talaee 2012), in addition to the complicated nonlinear relation between climate variables and the constraints of location determinants ascribed to latitude and altitude, induced this study to produce correct and dependable values of ET_0 with cost, effort and time saving method.

The values of ET_0 obtained from application of FAO P-M equation are employed in the ANNs model to investigate the efficiency of the network system in predicting ET_0. The correlation between ET_0 values obtained from P-M equation and the FFBP ET_0 predicted values during the training and testing stages are shown in Figure 4.
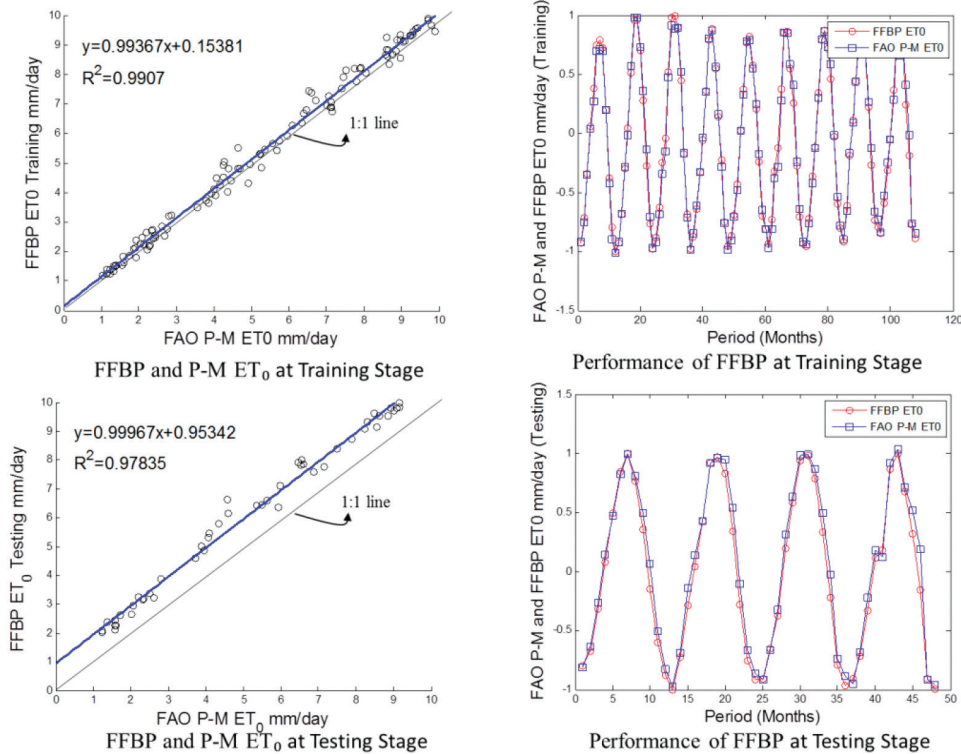

FIGURE 4. The correlation between FAO P-M ET_0 and FFBP ET_0 during training and testing stages for Baghdad station

The scatterplot shows that the results of FFBP model performance during the testing stage are so close to the corresponding results at training stage, the fit line equation and obtained R^2 values indicate a very satisfactory FFBP performance during both training and testing periods.

The statistical evaluation coefficients including RMSE, MAE and R^2 as presented in Table 1, enhanced the results illustrated in the scatterplots and confirm the efficiency of the proposed ANNs model in predicting ET_0.

TABLE 1. The statistical evaluation coefficients of ANNs model during training and testing stage in Baghdad station for the period of 2000-2013

Station	Stage	RMSE (mm/day)	MAE (mm/day)	R^2
Baghdad	Training stage	0.061	0.045	0.991
	Testing stage	0.104	0.074	0.978

The results of the statistical coefficients showed a significant consistency between calculated ET_0 using FAO P-M equation and FFBP predicted ET_0 values for the selected station.

The same values of ET results obtained by application of FAO P-M equation were also applied against a hybrid model (FFBP+GA). In this hybrid model; the population size

is (20), each pop meter is represented by 4 bits, maximum number of iterations is (200) and the mutation rate is 0.15. The principle architecture of the FFBP+GA model consists of one input layer, number of hidden layers (varies from 1-3 layers) and the number of associated neurons will depend on the result of implementing the hybridization model and one output layer.

The hybridization of ANNs with GA enhanced the model performance and maximized the fitness function and R^2 represented by the statistical parameters shown in Figure 5.

The statistical evaluation coefficients results shown in Table 2 confirm the positive effect of hybridization process and enhanced results shown in Figure 5. The evaluation results clarify that the FFBP+GA simulation model has modified the performance of ANNS model, during both training and testing stages, in terms of all included evaluation coefficients.

TABLE 2. The statistical evaluation coefficients of FFBP+GA model during training and testing stage in Baghdad Station for the period of 2000-2013

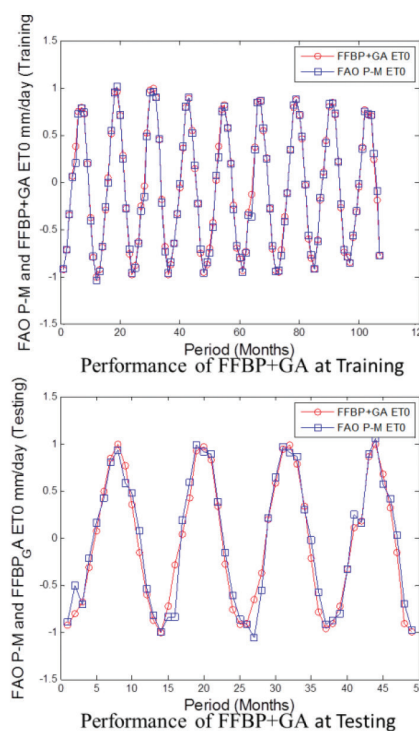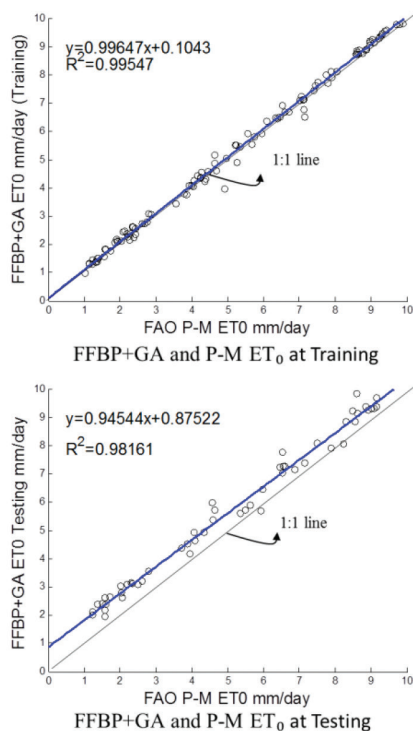| Station	Stage	RMSE (mm/day)	MAE (mm/day)	R^2
Baghdad	Training stage	0.059	0.044	0.995
	Testing stage	0.088	0.069	0.982

FIGURE 5. The correlation between calculated ET_0 using FAO P-M equation and FFBP+GA predicted ET_0 during training and testing stage for Baghdad station

The results of the evaluation coefficients presented in Table 2 confirm that the hybridization of FFBP with GA has upgraded the performance of FFBP model and increased the consistency of statistical results for the selected station.

CONCLUSION

ET_0 is the key parameter in designing and operating any irrigation project. In the P-M equation, both latitude and altitude values are considered properties of the study location and have great effect on the ET_0 values. Normalization of input and output parameters in the FFBP and FFBP+GA models will eliminate the effect of those two parameters. The development of the FFBP and hybrid FFBP+GA intelligent models would provide an effective alternative to the empirical method used in the estimation of ET_0. The proposed AI models have significant performance as compared to the limitations embedded in the traditional empirical FAO P-M equation method. This study shows that both FFBP and FFBP+GA models can be used for the prediction of ET_0 in Iraq, taking into consideration the unstable and sudden changes of weather conditions plus the need of new updated ET_0 predictions from time to time not only for developing projects but even for improving performance of existing irrigation projects. Both AI models could be recommended for application in arid and semiarid regions without restrictions of location specifications, with superiority of FFBP+GA on the FFBP model used, the proposed models will help to save a lot of time and efforts needed for achieving dependable predictions of ET_0.

REFERENCES

Allen, R., Pereira, L., Raes, D. & Smith, M. 1998. *Crop Evapotranspiration - Guidelines for Computing Crop Water Requirements - FAO Irrigation and Drainage Paper 56*. Rome, Italy.

Azzini, A. 2005. A New Genetic Approach for Neural Network Design and Optimization. Ph.D. Thesis, Universita Degli Studi Di Milano, Italy (unpublished) http://www.dti.unimi.it/azzini/wwwmat/TesiAzziniAntonia.pdf.

Badde, D.S., Gupta, A.k. & Patki, V.K. 2013. Cascade and feed forward back propagation artificial neural network models for prediction of compressive strength of ready mix concrete. *IOSR Journal of Mechanical and Civil Engineering (IOSR-JMCE)* 3: 1-6.

Basheer, I.A. & Hajmeer, M. 2000. Artificial neural networks: Fundamentals, computing, design, and application. *Journal of Microbiological Methods* 43(1): 3-31.

Bishop, C.M. 1994. Neural networks and their applications. *Rev. Sci. Instrum.* 65(6): 1803-1832.

Demuth, H., Beale, M. & Hagan, M. 2009. *Neural Network Toolbox ^{TM} 6 User's Guide*. The MathWorks, Inc.

Draper, N.R. & Smith, H. 1998. *Applied Regression Analysis*. New Jersey: John Wiley & Sons, Inc.

Droogers, P. & Allen, R.G. 2002. Estimating reference evapotranspiration under inaccurate data conditions. *Irrigation and Drainage Systems* 16(1): 33-45.

El-Baroudy, I., Elshorbagy, A., Carey, S.K., Giustolisi, O. & Savic, D. 2010. Comparison of three data-driven techniques in modelling the evapotranspiration process. *Journal of Hydroinformatics* 12(4): 365-379.

Gautam, M.R., Watanabe, K. & Saegusa, H. 2000. Runoff analysis in humid forest catchment with artificial neural network. *Journal of Hydrology* 235(1-2): 117-136.

Huang, G., Zhu, Q. & Siew, C. 2006. Extreme learning machine: A new learning scheme of feedforward neural networks. *Neurocomputing* 70: 489-501.

Iraqi Meteorological Organization and Seismology. 2014. *Ministry of Transportation*. Baghdad, Iraq: Ministry of Transportation. http://meteoseism.gov.iq/en/index.php?name=Pages&op=page&pid=74.

Khoshhal, J. & Mokarram, M. 2012. Model for prediction of evapotranspiration using MLP neural network. *International Journal of Environmental Sciences* 3(3): 1000-1009.

Kiranyaz, S., Ince, T., Yildirim, A. & Gabbouj, M. 2009. Evolutionary artificial neural networks by multi-dimensional particle swarm optimization. *Neural Networks : The Official Journal of the International Neural Network Society* 22(10): 1448-1462.

Kisi, Ö. 2008. The potential of different ANN techniques in evapotranspiration modelling. *Hydrological Processes* 22(14): 2449-2460.

Kisi, Ö. & Öztürk, Ö. 2008. Adaptive neurofuzzy computing technique for evapotranspiration estimation. *Journal of Irrigation and Drainage Engineering* 133(4): 368-379.

Kisi, O. 2006. Generalized regression neural networks for evapotranspiration modelling. *Journal of Hydrological Sciences* 51(12): 1092-1105.

Kumar, M., Raghuwanshi, N.S. & Singh, R. 2010. Artificial neural networks approach in evapotranspiration modeling: A review. *Irrigation Science* 29(1): 11-25.

Kumar, M., Raghuwanshi, N.S., Singh, R., Wallender, W.W. & Pruitt, W.O. 2002. Estimating evapotranspiration using artificial neural network. *Journal of Irrigation and Drainage Engineering* 128(4): 224-233.

Landeras, G., Ortiz-Barredo, A. & López, J.J. 2008. Comparison of artificial neural network models and empirical and semi-empirical equations for daily reference evapotranspiration estimation in the Basque country (Northern Spain). *Agricultural Water Management* 95(5): 553-565.

Rumelhart, E. & McClelland, L. 1986. *Parallel Distributed Processing*. Cambridge: The MIT Press.

Shiri, J., Kişi, Ö., Landeras, G., López, J.J., Amir Hossein Nazemi & L.C.P.M. Stuyt. 2012. Daily reference evapotranspiration modeling by using genetic programming approach in the Basque country (Northern Spain). *Journal of Hydrology* 414-415 (1): 302-316.

Tabari, H. & Hosseinzadeh Talaee, P. 2012. Multilayer perceptron for reference evapotranspiration estimation in a semiarid region. *Neural Computing and Applications* 23(2): 341-348.

Wang, G., Zhao, Y. & Wang, D. 2008. A protein secondary structure prediction framework based on the extreme learning machine. *Neurocomputing* 72(1-3): 262-268.

Warner, B. & Misra, M. 1996. Understanding neural networks as statistical tools. *The American Statistician* 50(4): 284-293.

Zurada, J.M. 1992. *Introduction to Artificial Neural Systems*. California: West Publishing Company.

Shafika Sultan Abdullah* & M.A. Malek
Department of Civil Engineering
Universiti Tenaga Nasional, Putrajaya Campus
Jalan IKRAM-UNITEN
43000 Kajang, Selangor Darul Ehsan
Malaysia

Namiq Sultan Abdullah
Department of Electrical and Computer Engineering
Zakho Street 38, 1006 AJ Duhok
Duhok Governorate - Kurdistan Region - Iraq
P.O Box 78

A. Mustapha
Faculty of Computer Science and Information Technology
Universiti Putra Malaysia
43400 Serdang, Selangor Darul Ehsan
Malaysia

Shafika Sultan Abdullah*
Akre Technical Institute, Dohuk Polytechnic University
61 Zakho Road, 1006 Mazi Qr Duhok
Kurdistan-Iraq

*Corresponding author; email: sha_akre@yahoo.com