Turkish Journal of Engineering & Environmental Sciences

http://journals.tubitak.gov.tr/engineering/

Research Article

Turkish J Eng Env Sci
(2013) 37: 211 – 219
© TÜBİTAK
doi:10.3906/muh-1212-5

M5 model trees and neural network based modelling of ET_0 in Ankara, Turkey

Mohammad Taghi SATTARI,[1,*] **Mahesh PAL,**[2] **Kadri YÜREKLİ,**[3] **Ali ÜNLÜKARA**[4]

[1]Department of Water Engineering, Faculty of Agriculture, Tabriz University, Tabriz, Iran
[2]Department of Civil Engineering, National Institute of Technology, Kurukshetra, Haryana, India
[3]Department of Farm Structures and Irrigation, Faculty of Agriculture, Gaziosmanpaşa University, Tokat, Turkey
[4]Department of Farm Structures and Irrigation, Seyrani Faculty of Agriculture, Erciyes University, Kayseri, Turkey

Received: 07.12.2012 • **Accepted:** 17.06.2013 • **Published Online:** 04.07.2013 • **Printed:** 29.07.2013

Abstract: This paper investigates the potential of back propagation neural network and M5 model tree based regression approaches to model monthly reference evapotranspiration using climatic data of an area around Ankara, Turkey. Input parameters include monthly total sunshine hours, air temperature, relative humidity, wind speed, rainfall, and monthly time index, whereas the reference evapotranspiration calculated by FAO–56 Penman–Monteith was used as an output for both approaches. Mean square error, correlation coefficient, and several other statistics were considered to compare the performance of both modeling approaches. The results suggest a better performance by the neural network approach with this dataset, but M5 model trees, being analogous to piecewise linear functions, provide a simple linear relation for prediction of evapotranspiration for the data ranges used in this study. Different scenario analysis with neural networks suggests that rainfall data does not have any influence in predicting evapotranspiration.

Key words: Evapotranspiration, M5 model tree, ANN, Penman–Monteith, Ankara

1. Introduction

Evapotranspiration (ET) is the combined process of plant transpiration and soil evaporation. Plant transpiration is the loss of water from the plant through tiny pores in the leaves known as stomates. The water enters the plants through the roots in liquid form and leaves the plants through the stomates in gaseous form. Soil evaporation is the direct evaporation of water from the surface of the soil into the atmosphere. The evapotranspiration rate from a reference surface is called reference evapotranspiration (ETo). A number of methods have been proposed to model reference evapotranspiration (Allen et al., 1998). Most of these models are highly complex and depend on meteorological data such as temperature, solar radiation, wind speed, and humidity (Kumar et al., 2002). Reference evapotranspiration is estimated by a physically based equation (e.g., FAO-56 Penman–Monteith equation) or empirical relationships between meteorological variables (Hargreaves–Samani and Blaney–Criddle relations). Studies have compared the results by these models using datasets collected under specific climatic and agronomic conditions (Nandagiri and Kovoor, 2006) and found that these models are valid only under specific climatic and agronomic conditions and cannot perform well under conditions different from those under which they were originally developed. For this reason, these models require local calibration to produce reliable estimates of ETo. The use of multiple regressions between meteorological variables and evapotranspiration is one such approach used for local calibration of models.

*Correspondence: mtsattar@gmail.com

In recent years, feed forward artificial neural networks (ANNs) have been extensively used in modeling nonlinear and nonstationary time series data in hydrology (Tayfur, 2012) and found to perform well in comparison to other statistical models. On the other hand, several studies reported the use of neural network techniques in modeling ETo. Rahimikhoob (2009) applied an ANN in estimating pan evaporation (E_{Pan}) as a function of air temperature data in the Safiabad Agricultural Research Center (SARC) located on Khuzestan plain in the southwest of Iran. Shirsath and Singh (2010) investigated the application of ANN, statistical regression, and climate based models, viz.: Penman, Priestley–Taylor and Stephens and Stewart, for estimation of daily pan evaporation. Piri et al. (2009) applied an ANN model to estimate evaporation in hot and dry regions. Awchi (2008) used radial basis function neural networks to predict evapotranspiration for Mosul meteorological station in the north of Iraq and found it performed well with their dataset. Zanetti et al. (2007) tested an ANN to estimate ETo as a function of the maximum and minimum air temperatures in the Campos dos Goytacazes county, State of Rio de Janeiro, whereas Keskin and Terzi (2006) applied an ANN method for developing a model to estimate daily pan evaporation in Eğirdir Lake, Turkey. Kumar et al. (2002) investigated the usefulness of ANNs to estimate daily grass crop ETo in comparison to the Penman–Monteith method. These studies suggest an improved performance of ANNs in predicting daily pan evaporation in comparison to other models.

Within the last decade, several studies have reported the use of an M5 model tree, a decision tree based regression approach, for water resource applications (Khan and See, 2006; Siek and Solomatine, 2007; Stravs and Brilly, 2007; Londhe and Dixit 2011). Pal and Deswal (2009) used an M5 model tree to model daily ETo using climatic data of the Davis station maintained by California irrigation Management Information System (CIMIS) and found it performed well in comparison to empirical relations. In other research, Sattari et al. (2013) compared the capabilities of an M5 model tree and support vector machine in predicting daily stream flows in the River Sohu, located within the municipal borders of Ankara, Turkey. They showed that the M5 model tree works well up to 7-day ahead forecasting in comparison of SVM. Keeping in view the potential of the M5 model tree based regression approach, the present study explored its capabilities in predicting the ETo using climatic data for the Ankara area (Turkey) in comparison to the ANN approach.

2. Artificial neural networks

The feed forward ANN is the most widely used neural network in water resource engineering. Its design consists of 1 input layer, at least 1 hidden layer, and 1 output layer. Connections are directed and allowed only in the forward direction, e.g., from input to hidden, or from hidden layer to a subsequent hidden or output layer. Each layer is made up of nonlinear processing units called neurons; the connections between neurons in successive layers carry associated weights. Nonlinear processing is performed by applying an activation function to the summed inputs to a unit. The backpropagation method, a gradient-descent algorithm that minimizes the error between the output of the training input/output pairs and the actual network outputs, is used to adjust the connecting weights (Bishop, 1995). Therefore, a set of input/output pairs is repeatedly presented to the network and the error is propagated from the output back to the input layer. The weights on the backward path through the network are updated according to an update rule and a learning rate. ANNs are not solely specified by the characteristics of their processing units and the selected training or learning rule. A neural network based modeling approach requires setting up several user-defined parameters like learning rate, momentum, optimal number of nodes in the hidden layer and the number of hidden layers, so as to have a less complex network with a better generalization capability. Further, training a neural network requires a number of iterations and a large number of training iterations may force ANN to over train, which may affect the predicting capabilities of the model.

3. M5 model tree

Model trees generalize the concepts of regression trees, which have constant values at their leaves (Witten & Frank, 2005). Therefore, they are analogous to piece-wise linear functions (and hence nonlinear). The M5 model tree (Quinlan, 1992) is a binary decision tree having linear regression functions at the terminal (leaf) nodes, which can predict continuous numerical attributes. Tree-based models are constructed by a divide-and-conquer method. A model tree generation requires 2 different stages. The first stage involves using a splitting criterion to create a decision tree. The splitting criterion for the M5 model tree algorithm is based on treating the standard deviation of the class values that reach a node as a measure of the error at that node and calculating the expected reduction in this error as a result of testing each attribute at that node. The formula to compute the standard deviation reduction (SDR) is

$$SDR = sd\ (T)\ -\ \sum \frac{|T_i|}{|T|}\ sd\ \ (T_i)\,, \tag{1}$$

where T represents a set of examples that reaches the node, T_i represents the subset of examples that have the ith outcome of the potential set, and sd represents the standard deviation. Due to the splitting process, the data in child nodes have less standard deviation as compared to parent node and are thus more pure. After examining all the possible splits, M5 chooses the one that maximizes the expected error reduction. This division often produces a large tree-like structure that may cause overfitting. To counter the problem of overfitting, the tree must be pruned back, for example by replacing a subtree with a leaf. Thus, the second stage in the design of the model tree involves pruning the overgrown tree and replacing the subtrees with linear regression functions. This technique of generating the model tree splits the parameter space into areas (subspaces) and builds in each of them a linear regression model. For further details of the M5 model tree, readers are referred to Quinlan (1992).

4. Study area and data

Ankara, the capital of Turkey, is an important commercial and industrial city and serves as the marketing center for the surrounding agricultural area. The metropolitan city area of Ankara province is located between $39°50'$ and $40°00'$ north latitudes and $32°35'$ and $33°00'$ east longitudes. It is at an altitude of 800–850 m in the center of Anatolia on the eastern edge of the great, high Anatolian plateau. Ankara plain is formed by Çubuk stream and its branches. This plain is surrounded to the north by Mount Çiçek, which forms the southern parts of Mount Mire, to the east by the western parts of Mount İdris, and to the south by mounts Çaldağıand Elmadağ (Figure 1). The plain is open from the western side and is connected to Mürted plain, which is formed by Ova stream. Its climate characteristics include a harsh dry continental climate with cold snowy winters and hot dry summers, and rainfall occurs mostly during spring and autumn. The mean temperature varies from 10 to 13 °C and average monthly precipitation is between 11 and 55 mm.

The FAO–56 Penman–Monteith (PM) relationship was used to calculate the monthly reference evapotranspiration value (ETo) using meteorological data from Ankara meteorological station for the time period between Jan 1975 to Dec 2006 (a total of 384 instances). Ten meteorological parameters, namely monthly total sunshine hours (SunH); monthly mean, min, and max air temperature (AvT, MinT, and MaxT); monthly mean, min, and max relative humidity (AvH, MinH, and MaxH); monthly average wind speed (Wind2); rain, and additional monthly time index (MTI), were considered inputs. In order to train both the ANN and M5 model tree approaches, 75% of instances (i.e. Jan 1975 to Dec 1998) were used, whereas the remaining 25% of instances (Jan 1999 to Dec 2006) were used for testing the models. Table 1 provides the statistical properties of the various meteorological parameters used in this study.

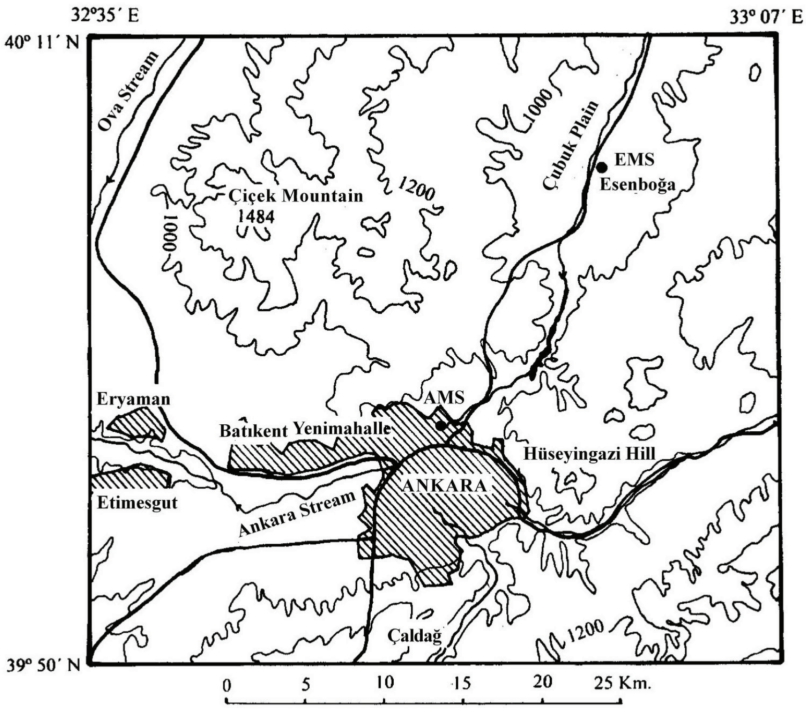

Figure 1. Study area, Ankara, Turkey.

Table 1. Statistical properties of all data.

Statistics	MinT (°C)	MaxT (°C)	SunH (h)	Wind2 (m/s)	MinH (%)	MaxH (%)	Rain (mm)	ETo(PM) (mm)
Mean	–0.16	25.02	210.28	1.43	27.15	92.34	33.48	106.75
Standard error	0.42	0.44	5.07	0.02	0.55	0.28	1.34	2.92
Standard deviation	8.19	8.65	99.35	0.31	10.81	5.49	26.33	57.28
Kurtosis	–0.86	–1.08	–1.25	–0.11	–0.23	2.97	0.30	–1.17
Skew	–0.18	–0.31	0.02	0.05	0.86	–1.88	0.90	0.22
Max	15.00	40.80	399.20	2.40	59.00	100.00	122.40	238.99
Min	–21.50	4.40	31.80	0.70	9.00	71.00	0.00	21.23

Neural network training can be made more efficient if certain preprocessing steps are performed on the network inputs and targets. It is often useful, before training, to scale the inputs and targets so that they always fall within a specified range. In the present study, the input and output data have been scaled to make it bounded in the intervals –1 and +1, which is preferable when a tan-sigmoid activation function is used in the network. The standardization equation used for scaling the dataset is represented by

$$Z = \frac{2 \times (X_i - X_{\min})}{(X_{\max} - X_{\min})} - 1, \tag{2}$$

where Z is standardized input values lying in the range of [–1, +1], and min X and max X are minimum and maximum input values, respectively. After simulation, all the output values are destandardized by multiplying by the respective standardization factor to get actual ETo values. This step helps the neural network training to be more efficient (Awchi, 2008).

5. Results

The aim of the present study was to explore the potential of the ANN and M5 model tree for the prediction of the monthly ETo of the study area. Several neural networks models were created utilizing various input combinations (Table 2), whereas estimated monthly ETo calculated by using the PM method was considered the output for all ANN models. The performance evaluation process included 2 statistical tests, i.e. correlation coefficient (R) and the mean square error (MSE), for the test dataset. Various user-defined parameters of the ANN were determined by trial and error using MSE as the main performance criterion. A large number of trials were carried out using different combinations of user-defined parameters to find the optimal values of these parameters. With this dataset, the neural network with 1 hidden layer having 4 nodes was found to work well.

Table 2. ANN model results for different scenarios.

Scenarios	Inputs	MSE	R
1	SunH, AvT, MinT, MaxT, AvH, MinH, MaxH, Wind2, Rain, MTI	0.025	0.981
2	SunH, AvT, MinT, MaxT, AvH, MinH, MaxH, Wind2, MTI	0.017	0.982
3	SunH, AvT, MinT, MaxT, MinH, MaxH, Wind2, MTI	0.011	0.991
4	MinT, MaxT, MaxH, MinH, Wind, SunH, MTI	0.002	0.997
5	MinT, MaxT, SunH, Wind, AvH, Rain, MTI	0.003	0.996
6	AvT, MaxH, MinH, Wind, SunH, MTI	0.006	0.990
7	MinT, MaxT, MaxH, MinH, MTI	0.012	0.988
8	MinT, MaxT, Wind, MTI	0.005	0.992
9	MinT, MaxT, SunH, MTI	0.011	0.991
10	MinT, MaxT, Rain, MTI	0.013	0.987
11	MinT, MaxT, AvH, MTI	0.013	0.987
12	MinT, MaxT, MTI	0.015	0.986
13	AvT, MTI	0.013	0.984

In Table 2 different scenarios are presented as an input for ANN. Scenarios 1, 2, 10, and 12 checked the effect of rainfall data in monthly ET0 prediction. The comparison between these scenarios showed that elimination of rainfall data had no effect on monthly ET0.

Results from Table 2 suggest that a combination of 7 input parameters (MinT, MaxT, MaxH, MinH, Wind, SunH, MTI; MSE = 0.002, R = 0.997) performs well in comparison to other combinations using the ANN approach. The observed and estimated values of reference evapotranspiration by using 7 input parameters are plotted in Figure 2. This figure depicts almost perfect agreement between the actual and predicted values of reference evapotranspiration. A plot between actual and predicted values of reference evapotranspiration with time is provided in Figures 3. A comparison of the results (Figure 3) suggests than the ANN approach works well in predicting reference evapotranspiration with this dataset.

In order to assess the usefulness of the M5 model tree in predicting reference evapotranspiration, the same training and test dataset (with 7 input parameters) performing well with ANN was considered. Several statistical parameters were used to compare the performance of the M5 model tree and ANN approach. Table 3 provides values of these statistical parameters for the M5, ANN, and PM methods with the test dataset. The results from Table 3 suggest that the M5 model tree can effectively be used to predict reference evapotranspiration for the study area. A plot between actual and predicted values of ETo by the M5 model tree (Figure 4) suggests that most of the predicted values lie on the line of perfect agreement. A high value of R (0.995) and smaller value of MSE (4.533) with the M5 model tree also confirm that this approach also works well in predicting reference evapotranspiration with this dataset. Figures 5 shows a plot of measured versus predicted values of reference evapotranspiration obtained using the M5 model tree with the test dataset. Apart from providing accuracy comparable to that of the ANN approach, a major advantage of the M5 model tree approach is the availability of simple linear relations in predicting reference evapotranspiration (Figure 6).

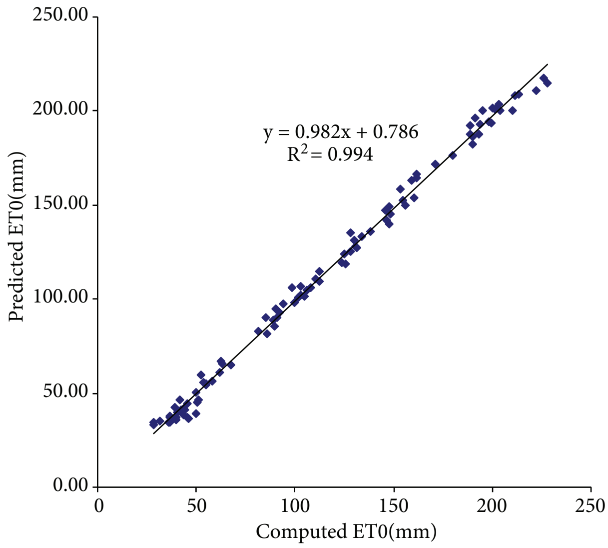

Figure 2. Scatter plot provided by ANN approach.

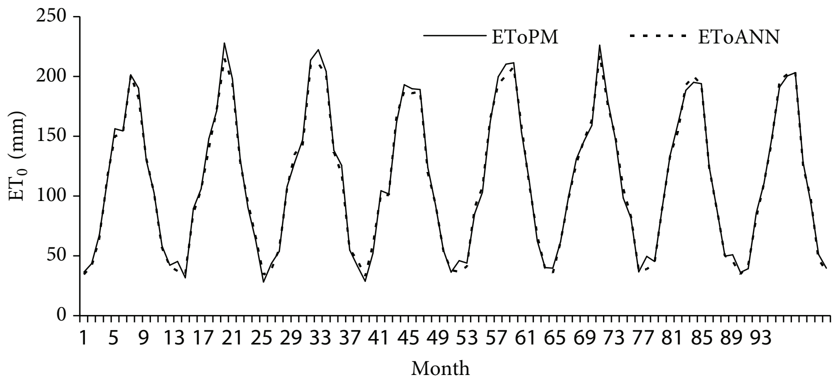

Figure 3. Time series plot using ANN modeling approach.

Table 3. Statistical properties for observed (PM) and modeled (ANN, M5) for test data.

For test period	Observed ETo (PM)	ANN	M5
Mean	114.2789	113.0781	114.1729
Minimum	28.1	33.01	26.7
Maximum	227.82	217.79	222.8
Variance	3,608.88	3,502.31	3,433.42
Standard deviation	60.0739	59.1803	58.5954
Skewness	0.2323	0.189	0.1899
Kurtosis	1.7611	1.7079	1.7369

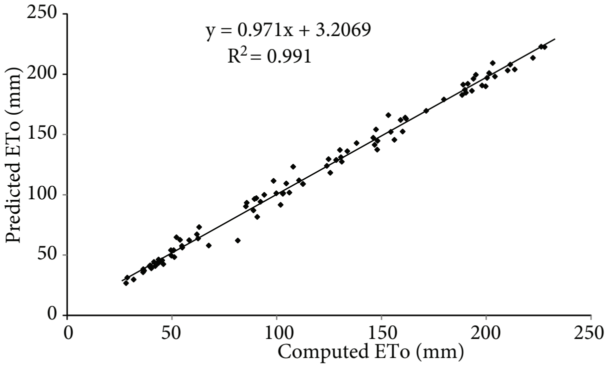

Figure 4. Scatter plot by M5 model tree approach.

In order to compare the performance of both M5 and ANN models on the test dataset, global performance metrics as suggested by Dawson et al. (2007) was also used (Table 4). Results from this table suggest that the ANN model consistently outperformed the M5 tree models in terms of different parameters of evaluation metrics for this dataset except for RVE and ME but the availability of simple linear relations provided by the M5 model tree is a major advantage for field conditions.

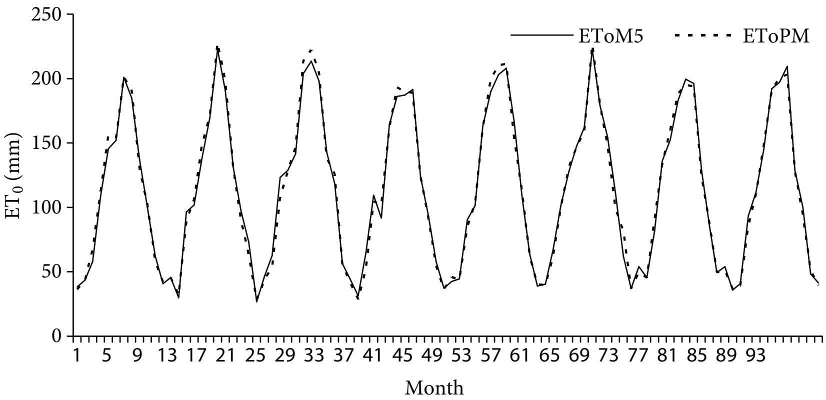

Figure 5. Time series plot of M5 model tree results.

```
MaxT <=  21.65 :
|  MaxT <= 16 : LM1
|  MaxT > 16 : LM2
MaxT > 21.65 : LM3
LM1:   ETo =    -0.2122 * MTI + 0.0116 * MinT + 1.7809 * MaxT + 0.0221 * SunH+ 16.903 *
Wind2 - 0.3899 * MinH- 0.0247 * MaxH+ 9.6565
LM2:   ETo =    -0.4649 * MTI - 0.2572 * MinT + 3.2309 * MaxT + 0.0221 * SunH + 28.146 *
Wind2 - 0.6663 * MinH+ 0.6744 * MaxH - 84.0023
LM 3:   ETo  =  -3.5752 * MTI + 1.1924 * MinT + 3.9626 * MaxT + 0.1658 * SunH + 37.9179 *
Wind2 + 0.0257 * MinH- 0.0158 * MaxH - 61.6546
```

Figure 6. Linear models provided by M5 model tree approach.

Table 4. Comparison of results in test period.

Statistic	ANN	M5
Absolute maximum error (AME)	12.84	19.46
Peak difference (PDIFF)	10.03	5.02
Mean absolute error (MAE)	3.5928	4.4009
Mean error (ME)	1.2007	0.1059
Root mean squared error (RMSE)	4.5225	5.8127
Fourth root mean quadrupled error (R4MS4E)	5.9803	8.1064
Number of sign changes (of residuals) (NSC)	44	50
Relative absolute error (RAE)	0.0687	0.0842
Percent error in peak (PEP)	4.4026	2.2035
Mean absolute relative error (MARE)	0.0424	0.0459
Median absolute percentage error (MdAPE)	2.9365	3.1671
Mean squared relative error (MSRE)	0.0039	0.0042
Relative volume error (**RVE**)	0.0105	0.0009
Coefficient of determination (RSqr)	0.9949	0.991
Index of agreement (IoAd)	0.9986	0.9976
Coefficient of efficiency (CE)	0.9943	0.9906
Coefficient of persistence index (PI)	0.9844	0.9742

6. Conclusion

This study compared the performance of the M5 model tree and ANN approaches in predicting monthly reference evapotranspiration using meteorological data from Ankara weather station (Turkey). The results presented are quite encouraging and suggest that both ANN and M5 model tree approach works well in predicting reference evapotranspiration. Further, it can be concluded from this study that in comparison to the M5 model tree the ANN approach works well with this dataset but the M5 model tree approach provides simple linear relations, which can be used to predict the reference evapotranspiration. This study also suggests that rainfall is not an important parameter in prediction of reference evapotranspiration with this dataset.

References

Allen, R.G., Pereira, L.S., Raes, D. and Smith, M.,"Crop Evapotranspiration, Guidelines for Computing Crop Water Requirements". In FAO Irrigation and Drainage, Paper no. 56, Food and Agriculture Organization of the United Nations: Rome. 1998.

Awchi, T.A.,"Application of Radial Basis Function Neural Networks for Reference Evapotranspiration Prediction", Al-Rafidain Engineering, 16, 117–129, 2008.

Bishop, C.M., Neural Networks for Pattern Recognition. Oxford: Clarendon Press. 1995.

Keskin, M.E. and Terzi, O.,"Artificial Neural Network Models of Daily Pan Evaporation", Journal of Hydrologic Engineering, 11, 65–70, 2006.

Khan, A.S. and See, L., Rainfall-runoff modeling using data driven and statistical methods. Institute of Electrical and Electronics Engineers (IEEE) 2006.

Kumar, M., Raghuwanshi, N.S., Singh, R., Wallender, W.W. and Pruitt, W.O.,"Estimating Evapotranspiration Using Artificial Neural Network". ASCE Journal of Irrigation and Drainage Engineering 128, 224–233, 2002.

Londhe, S.N and Dixit, P.R.,"Forecasting Stream Flow Using Model Trees", International Journal of Earth Sciences and Engineering, 4, 282–285, 2011.

Nandagiri, L. and Kovoor, G.M.,“Performance Evaluation of Reference Evapotranspiration Equations across a Range of Indian Climates”, ASCE Journal of Irrigation and Drainage Engineering, 132, 238–249, 2006.

Pal, M. and Deswal, S.,“M5 Model Tree Based Modelling of Reference Evapotranspiration”, Hydrological Processes, 23, 1437–1443, 2009.

Piri, J., Amin, S., Moghaddamnia, A., Keshavarz, A., Han, D. and Remesan, R.,“Daily Pan Evaporation Modeling in a Hot and Dry Climate”, Journal of Hydrologic Engineering, 14, 803–811, 2009.

Quinlan, J.R., Learning with continuous classes. In: Proc. AI’92 (Fifth Australian Joint Conf. on Artificial Intelligence) (ed. by A. Adams & L. Sterling), 343–348. World Scientific, Singapore. 1992.

Rahimikhoob, A.,“Estimating Daily Pan Evaporation Using Artificial Neural Network in a Semi-Arid Environment”, Theoretical and Applied Climatology, 98, 101–105, 2009.

Sattari, M.T., Pal, M., Apaydin, H. and Ozturk, F.,“M5 Model Tree Application in Daily River Flow Forecasting in Sohu Stream, Turkey”, Water Resources, 40, 233–242, 2013.

Shirsath, P.B. and Singh, A.K., “A Comparative Study of Daily Pan Evaporation Estimation Using ANN, Regression and Climate Based Models”. Water Resources Management, 24, 1571–1581, 2010.

Siek, M. and Solomatine D.P.,“Tree-Like Machine Learning Models in Hydrologic Forecasting: Optimality and Expert Knowledge”. Geophysical Research Abstracts, Vol. 9, 2007.

Stravs, L. and Brilly, M.,“Development of a Low Flow Forecasting Model Using the M5 Machine Learning Method”, Hydrological Sciences, 52, 466–477, 2007.

Tayfur, G., Soft Computing Methods in Water Resources Engineering, ANN, FL, GA. WIT Press, 267 pages, Southampton, UK. 2012.

Witten, I.H. and Frank, E., Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations. Morgan Kaufmann: San Francisco. 2005.

Zanetti, S.S., Sousa, E.F., Oliveira, V.P.S., Almeida, F.T. and Bernardo, S.,“Estimating Evapotranspiration Using Artificial Neural Network and Minimum Climatological Data”, Journal of Irrigation and Drainage Engineering, 133, 83–89, 2007.