

Estimation of Daily Reference Evapotranspiration Using Support Vector Machines and Artificial Neural Networks in Greenhouse

S.S. Eslamian, J. Abedi-Koupai, M.J. Amiri and S.A. Gohari
Department of Water Engineering, Isfahan University of Technology, Isfahan, Iran

Abstract: In the present study, the meteorological variables including air temperature, solar radiation, wind speed and relative humidity were considered daily. The R^2 of ANNs and SVMs models were obtained 0.92 and 0.96, respectively; whereas the efficiency of ANNs and SVMs models were 0.83 and 0.91, respectively. Both ANNs and SVMs approaches work well for the data set used in greenhouse condition, but the SVMs model works better in comparison with the ANNs model.

Key words: Meteorological variables, ANNs model, SVMs model, efficiency

INTRODUCTION

Nowadays, greenhouse cultivation plays a significant role on fresh vegetable production in Iran. Greenhouse farming, also known as protected cultivation, is one of the farming systems widely used to provide and maintain a controlled environment suitable for optimum crop production leading to maximum profits. This includes creating an environment suitable for working efficiency as well as for better crop growth (Donatelli *et al.*, 2003). The main advantage with the greenhouse farming is that the production can be obtained throughout the year, which is not possible in the open field farming due to heavy rainfall and wind, especially in tropical regions (Allen *et al.*, 1998). In addition, greenhouse technology can contribute to solve the global issues such as the shortage of artificial energy, water, environmental pollution and instability of ecological system in various ways. Irrigation system is one of the most important components affecting the yield and quality of agricultural products from greenhouse farming system. However, more research is needed on irrigation management to establish the appropriate method to be used for estimating crop evapotranspiration (ET), to avoid the excess or deficit water application, soil salinity and groundwater contamination. ET_0 refers to the water removed from a unit ground area completely covered with a reference crop, healthy and unstressed and with ample water supply (Walter *et al.*, 2002). The ET_0 is used to quantify evaporative demand within a region and to estimate crop ET when the ET_0 is multiplied by a crop coefficient (K_c) factor to account for differences between the grass and crop ET (Allen *et al.*, 1998; Schuch and Burger, 1997). Operational software tools in the domains specified above require the estimate of ET_0 and procedures to do that have been repeatedly implemented in software applications. This is one of the reasons why there is an increasing demand for modular approaches in model development (Jones *et al.*, 2001). Such a modular approach leads to the concept of encapsulating the solution of a modeling problem in a discrete, replaceable and interchangeable unit (Donatelli *et al.*, 2004; Troya and Vallecillo, 2001). Reference evapotranspiration can be determined by lysimeter but this method is very expensive and not easy to use. Therefore ET_0 could be estimated by some equipments such as pan, reduced pan and atmometer, but the method based on class A pan evaporation is the most common due to its simplicity and relatively low cost. Many researchers have attempted to estimate the evaporation through the indirect methods using the climatic variables, but some of these techniques require the data which can

not be easily obtained (Rosenberry *et al.*, 2007). The evaporation process is strongly nonlinear in nature; some researchers should emphasize the estimation of relatively accurate evaporation in the research greenhouse using modeling techniques (Bruton *et al.*, 2000; Lindsey and Farnsworth, 1997; Xu and Singh, 1998). Sudheer *et al.* (2002) investigated the prediction of class A Pan Evapotranspiration (PE) using the neural networks model. They used the neural networks model for the evaporation process using proper combinations of the observed climate variables such as temperature, relative humidity, sunshine duration and wind speed for the neural networks model. Kisi (2006) used proper combinations of the observed climatic variables such as air temperature, solar radiation, wind speed, pressure and relative humidity for the neuro-fuzzy model to estimate the daily PE. The aim of this study is estimation of daily pan evaporation using support vector machines and artificial neural networks in greenhouse condition. As the pan evapotranspiration data can be used in the greenhouse and this method is introduced as a suitable method in the greenhouse and as the artificial neural networks and SVMs are used out of the greenhouse (Bruton *et al.*, 2000), the aim of this study was to evaluate if these models can be used inside the greenhouse and if it is possible to estimate Reference evapotranspiration by using climate data inside the greenhouse.

MATERIALS AND METHODS

Research Site and Operating Dates

The experiment was carried out from 23 September 2007 to 23 September 2008 in Isfahan University of Technology, Iran. The geometric characteristics of the glasshouse are as follows: eaves height 3.5 m, ridge height 5 m, total width 10 m and length 20 m. The local altitude is 1624.4 m with latitude 32° 42' N, longitude 51° 28' E, mean annual precipitation of 134 mm, mean annual temperature of 17°C and mean relative humidity of 38%. The glasshouse was built at North-South orientation, covered with a polyethylene film of 0.15 mm thickness. It was naturally ventilated with a single continuous roof vent and lateral windows were kept open during daytime. For measuring microclimate inside the glasshouse, air temperature was measured by Hobo pendant temperature light sensors (2 No.) and relative humidity was measured by RH-sensor. Incoming solar radiation was measured by luxmeter which is placed at the center of glasshouse 0.15 m above the floor. The handy vane anemometer was used to measure daily wind speed at 2 m above ground level and a class A evaporation pan were placed in the southern part of the greenhouse. The class A pan was constructed of galvanized iron sheet, 1.21 m in diameter and 0.25 m in depth. The water depth was maintained between 0.18-0.22 m; the water level was thus 0.07-0.03 m below the rim in order to avoid a great variation in water volume. Water level was measured daily using a hook gauge with resolution of 0.01 mm, which was placed on a still well. Evaporation readings of class A pan were taken everyday between 7:00 and 8:00 a.m. Reading the class A pan installed inside the greenhouse was not influenced by rain directly, since glass cover protected the equipments from rain. Reference evapotranspiration is calculated by the following equation (Abedi-Koupai and Asadkazemi, 2006):

$$ET_0 = K_p E_{pan} \quad (1)$$

Where:

ET_0 = Reference evapotranspiration (mm)

K_p = Pan coefficient

E_{pan} = Pan evaporation (mm)

The K_p values were assumed to be unity as recommended by Fernandes *et al.* (2003) for greenhouse condition.

Support Vector Machines

A support vector machine uses a linear model to separate the sample data through some nonlinear mapping from the input vectors into the high-dimensional feature space (Eslamian *et al.*, 2008; Tay and Coa, 2001). The linear model constructed in the new space can represent a nonlinear decision boundary in the original space. SVM aims at finding a special kind of linear model, the so-called optimal separating hyperplanes. The training points that are closer to the optimal separating hyperplane are called support vectors, which determine the decision boundaries. In general cases where the data is not linearly separated, SVM uses the nonlinear machines to find a hyperplane that minimizes the number of errors on the training set. Consider a training set $D = \{x_i, y_i\}_{i=1}^N$ with input vectors $X_i = \{X_i^1, \dots, X_i^N\} \in R^n$ and target labels $y_i \in \{-1, +1\}$.

SVM binary classifier satisfies the following conditions:

$$y_i (w^T \phi(x_i) + b) \geq 1 \quad i = 1, \dots, N \tag{2}$$

where, w represents the weighting vector and b is the bias.

The nonlinear function $\Phi(0): R^n \rightarrow R^{nk}$ maps the input vectors into a high-dimensional feature space. From Eq. 2, it can be seen that it is possible for multiple solutions to separate training data points. From a generalization perspective, it is the best to choose two bounding hyperplanes at opposite sides of a separating hyperplane $w^T \Phi(X) + b = 0$ with largest margin $2/(|w|^2)$. However, most of the classification problems are linearly non-separable cases. Therefore, it is general to introduce slack variables ξ_i to permit misclassification. Thus the optimization problem becomes as follow:

$$\text{Min}_{w, b, \xi} \left(\frac{1}{2} w^T w + C \sum \xi_i \right) \tag{3}$$

$$\text{s.t.} \begin{cases} y_i (w^T \phi(x_i) + b) - 1 \leq \xi_i \\ \xi_i \geq 0 \end{cases} \quad i = 1, \dots, N \tag{4}$$

where, C is the penalty parameter of the error term. The solution of the primal problem is obtained after constructing the Lagrangian. Then, the primal problem can be converted into the following Q_p -problem.

$$\text{Max}_\alpha \left(e^T \alpha - \frac{1}{2} \alpha^T Q \alpha \right) \tag{5}$$

$$\text{s.t.} \begin{cases} 0 \leq \alpha_i \leq C, i=1, \dots, N \\ \sum_{i=1}^N \alpha_i \xi_i = 0 \end{cases} \tag{6}$$

where, α_i is Lagrange multipliers, $Q_{ij} = y_i y_j \Phi(X_i)^T \Phi(X_j)$. Due to a large amount of computation, inner product is replaced with kernel function which satisfies Mercer's condition, $K(x_i, x_j) = \Phi(X_i)^T \Phi(X_j)$. Finally, we get a nonlinear decision function in primal space for linearly non-separable case.

$$y(x) = \text{sgn} \left(\sum_{i=1}^N \alpha_i y_i k(x, x_i) + b \right) \tag{7}$$

Four common kernel function types of SVMs are given as follows:

- Linear kernel: $k(x_i, x_j) = x_i^T x_j$
- Polynomial kernel: $k(x_i, x_j) = (Y x_i^T x_j + r)^d$

- Radial basis kernel: $k(x_i, x_j) = \exp(-Y\|x_i - x_j\|^2)$
- Sigmoid kernel: $k(x_i, x_j) = \tanh(Yx_i^T x_j + r)$

where, $d, r \in \mathbb{N}$ and $Y \in \mathbb{R}^+$ are constants (Eslamian *et al.*, 2008).

Modeling for SVM

Model selection and parameter search play a crucial role in the performance of SVMs. However, there is no general guidance for selection of SVM kernel function and parameters so far. In general, the Radial Basis Function (RBF) is suggested for SVMs. The RBF kernel nonlinearly maps the samples into the high-dimensional space, so it can handle nonlinear problem. Furthermore, the linear kernel is a special case of the RBF. The sigmoid kernel behaves like the RBF for certain parameter; however, it is not valid under some parameters. The second reason is the number of hyperparameters which influences the complexity of model selection. The polynomial has more parameters than the RBF kernel. Finally, the RBF function has less numerical difficulties. While RBF kernel values are $0 < K_{ij} = 1$, polynomial kernel value may go to infinity or zero when the degree is large. In addition, polynomial kernel takes a longer time in the training stage and is reported to produce worse results than the RBF kernel in the previous studies (Sudheer *et al.*, 2002). The linear kernel SVM has no parameters to tune except for C . For the nonlinear SVM, there are additional parameters, the kernel parameters c to tune. Improper selection of the penalty parameter C and kernel parameters can cause overfitting or underfitting problems. Currently, some kinds of parameter search approach are employed such as cross validation via parallel grid-search, heuristics search and inference of model parameters within the Bayesian evidence framework (Sudheer *et al.*, 2002). For median-sized problems, cross-validation might be the most reliable way for model parameter selection. In v -fold cross-validation, the training set is first divided into v subsets. In the i th ($i = 1, 2, \dots, v$) iteration, the i th set (validation set) is used to estimate the performance of the classifier trained on the remaining ($v-1$) sets (training set). The performance is generally evaluated by cost, e.g., classification accuracy or Mean Square Error (MSE). The final performance of classifier is evaluated by mean costs of v folds subsets. In grid-search process, pairs of (C, c) are tried and the one with the best cross-validation accuracy is picked up. In this study, it is preferred a grid-search on (C, c) using 10- fold cross-validation for the following reasons. Firstly, the cross-validation procedure can prevent the overfitting problem. Secondly, computational time to find good parameters by grid-search is not much more than that by the other methods. Furthermore, the grid-search can be easily parallelized because each (C, c) is independent. While other methods are iterative process, which might be difficult for parallelization. We use LIBSVM software to conduct SVMs experiment. The overall procedure of modeling SVM is shown in Fig. 1.

The adequacy of the ANNs and SVMs evapotranspiration models were evaluated by estimating the coefficient of determination (R), defined based on the evapotranspiration estimation errors as:

$$R^2 = \frac{E_0 - E}{E_0} \quad (8)$$

Where:

$$E_0 = \sum_{i=1}^n (E_{i(\text{pan})} - E_{i(\text{mean})})^2 \quad (9)$$

$$E = \sum_{i=1}^n (E_{i(\text{pan})} - E_{i(\text{simulated})})^2 \quad (10)$$

where, $E_{i(\text{pan})}$ and $E_{i(\text{simulated})}$ are daily pan evaporation measurement and ANNs model evaporation, E_{mean} .

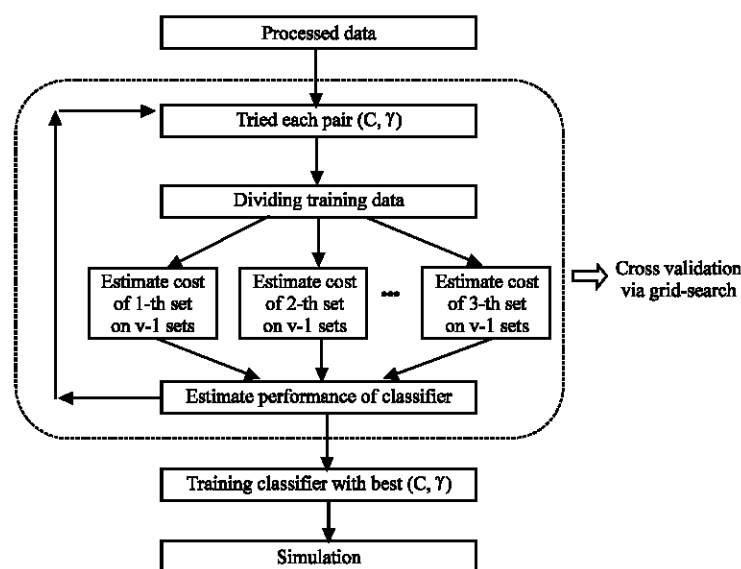


Fig. 1: Overall procedure of modeling SVM

Artificial Neural Networks

Artificial neural networks are a type of parallel computer structure, within which a number of processing units are linked together so that the computer's memory is distributed and information is passed in a parallel manner. A large number of ANN architectures and algorithms have been developed so far, multilayer feedforward networks (Sudheer *et al.*, 2002), self-organizing feature maps Hopfield networks, counterpropagation networks, radial basis function networks and recurrent ANNs (Sudheer *et al.*, 2002). Of these networks, the most commonly used are feedforward networks and radial basis function networks. Multi-layer feedforward networks have been found to perform best when used in hydrological applications and as such they are by far the most commonly used (Sudheer *et al.*, 2002). The attempt to choose between different methods and defining which is superior, is likely to fail as in most cases the choice should be application oriented. It is preferable for every new application to test different types of ANNs rather than use a pre-selected one.

Feed-Forward Propagation Neural Networks (FFNN)

The most commonly used ANN is the three-layer feed-forward ANN. In feedforward neural networks architecture, there are layers and nodes at each layer. Each node at input and inner layers receives input values, processes and passes to the next layer. This process is conducted by weights. Weight is the connection strength between two nodes. The numbers of neurons in the input layer and the output layer are determined by the numbers of input and output parameters, respectively. In the present feed-forward artificial neural networks are used. The model is shown in Fig. 2. In the Fig. 1, i, j, k denote nodes input layer, hidden layer and output layer, respectively. W is the weight of the nodes. Subscripts specify the connections between the nodes. For example, W_{ij} is the weight between nodes i and j . The term feed-forward means that a node connection only exists from a node in the input layer to other nodes in the hidden layer or from a node in the hidden layer to nodes in the output layer; and the nodes within a layer are not interconnected to each other.

To evaluate the performance of these models in daily ET_0 estimates, between the predicted and measured evapotranspiration using the class A pan method values, several performance criteria were

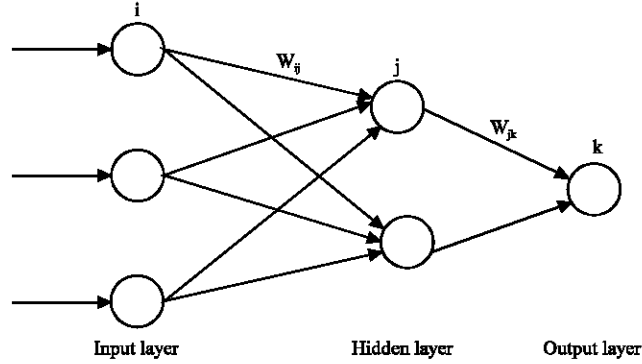


Fig. 2: Feed-forward artificial neural networks with one layer

used including regression analysis, agreement index (D), Mean Absolute Error (MAE), maximum absolute error (MAXE) and efficiency (EF). These criteria are defined as:

$$D = 1 - \frac{\sum_{i=1}^n (O_i - E_i)^2}{\sum_{i=1}^n (|E_i - \bar{O}| + |O_i - \bar{O}|)^2} \quad (11)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |O_i - E_i| \quad (12)$$

$$MAXE = \text{MAX}(|O_i - E_i|) \quad (13)$$

$$EF = \frac{\sum_{i=1}^n (O_i - \bar{O})^2 - \sum_{i=1}^n (O_i - E_i)^2}{\sum_{i=1}^n (O_i - \bar{O})^2} \quad (14)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (O_i - E_i)^2}{n}} \quad (15)$$

Where:

- O_i = Measured value
- E_i = Predicted value
- \bar{O} = mean observed values

RESULTS AND DISCUSSION

Artificial Neural Network

In this study, ANNs model was performed with neuro solution software. Sixty percent of the total data was randomized for as training data, 20% of the total data was randomized as testing performance and 20% was selected for cross validation performance. ANNs evaporation model with four input variables (air temperature, relative humidity, solar radiation and wind speed) are considered. For ANNs model, the number of hidden layers considered after trial and cross validation is two layers and number of hidden neurons is obtained five neurons and the used functions for hidden and output layers are log sigmoid.

The Training Performance

In neuro solution software, 60% of the total data was randomized for training data. This software does not need for standardized input layer and training data was used ordinary in this performance. Table 1 shows a statistical analysis of the PE for training performance. According to Table 1, the best network with 2000 epoch has $MSE = 0.0067$.

The Test Performance

The testing performance applied a cross-validation method in order to overcome the over fitting data. The cross-validate method is not to train all of the training data until MSE was reached to the minimum amount, but is to cross-validate with the testing data at the end of each performance. The correlation coefficient and MSE values were used to judge the performance of ANNs for data. Actual and predicted values of efficiency were also plotted. Table 2 shows a statistical analysis of the reference evapotranspiration for the testing performance of ANNs. Table 2 shows that for cross validation, the values of D, EF and r-square (R^2) were 0.93, 0.83 and 0.920, respectively.

Sensitivity of the Pan Evaporation to Meteorological Variables

From Fig. 3, it is clear that increasing in wind speed, air temperature and solar radiation are significant at 12.566, 7.577 and 5.968 level, while decreasing relative humidity is significant at the 2.251 level. Wind speed and air temperature are the most sensitive variables.

Support Vector Machines

The D, EF and r-square (R^2) values were used to judge the performance of SVMs for the data set. One advantage of using SVMs is the use of a quadratic optimization, which provides a global minimum in comparison with the local minima with back propagation neural network due to the use of non-linear optimization. Both ANNs and SVMs were applied for calculating the D, EF and r-square (R^2) using cross-validation and a percentage split method for the input data set comprising different attributes. Table 3 shows a statistical analysis of the reference evapotranspiration for the testing performance of SVMs. Figure 4 and 5 were show the actual and predicted values of reference evapotranspiration by ANNs and SVMs, that shows fitting of the measured and predicted values of reference evapotranspiration by ANNs and SVMs.

Table 1: Statistical analysis of the pan evaporation for the training performance

Best Networks	Training	Cross validation
Epoch No.	2000	2000
Minimum MSE	0.00675413	0.006126789
Final MSE	0.00675413	0.006126789

Table 2: Statistical analysis of the reference evapotranspiration for testing performance

Test performance	ET
D	0.93
MAE	0.65
MAXE	2.75
RMSE	0.86
EF	0.83
R^2	0.92

Table 3: Statistical analysis of the reference evapotranspiration for testing performance

Test performance	ET
D	0.97
MAE	0.55
MAXE	1.95
RMSE	0.76
EF	0.91
R^2	0.96

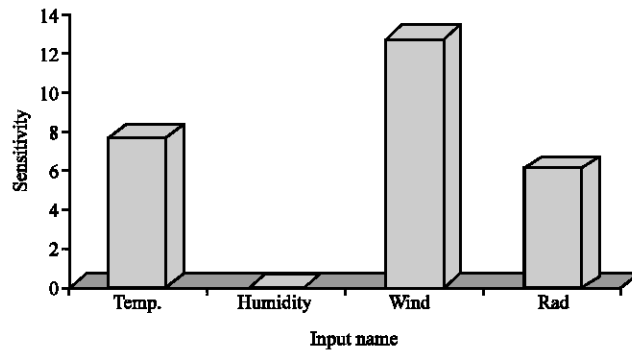


Fig. 3: The sensitivity of the reference evapotranspiration to the four meteorological variables

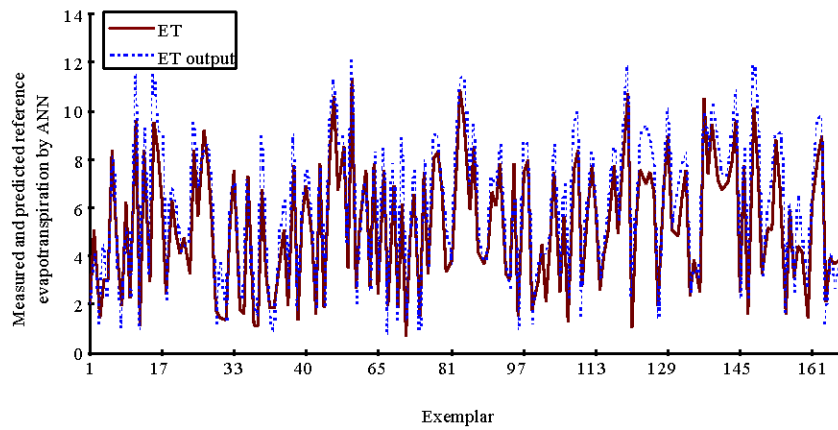


Fig. 4: Predicted reference evapotranspiration using ANNs

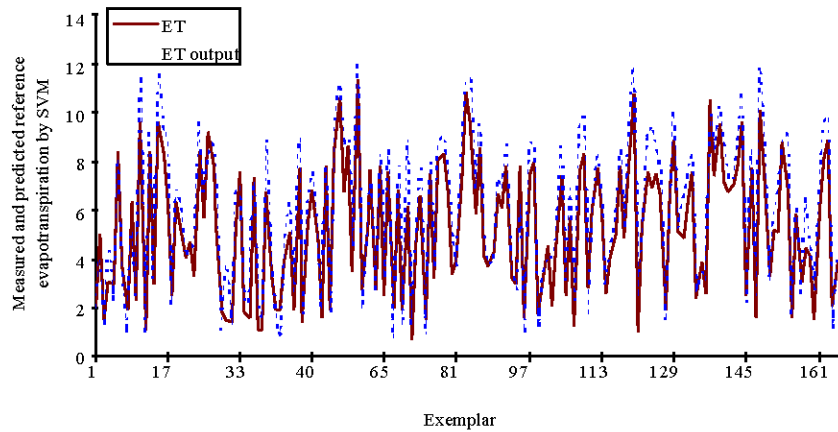


Fig. 5: Predicted pan evaporation using SVM

CONCLUSION

Comparison of the R^2 and efficiency values also suggests an improved performance by both ANNs and SVMs. A possible reason of the better performance by both ANNs and SVMs may be related to that they have a larger number of user-defined parameters. Based on this study, both ANNs and SVMs approaches work equally well for the data set used. The computation cost involved with SVM is significantly smaller than the ANNs algorithm. Based on the values of the efficiency, the SVMs approaches work better than the ANNs model.

REFERENCES

- Abedi-Koupai, J. and J. Asadkazemi, 2006. Effects of a hydrophilic polymer on the field performance of an ornamental plant under reduced irrigation regimes. *Iran. Polym. J.*, 15: 715-725.
- Allen, R.G., L.S. Pereira, D. Raes and M. Smith, 1998. *Crop Evapotranspiration: Guidelines for Computing Crop Water Requirements*. 1st Edn., Irr. and Drain. Paper 56. UN-FAO, Rome, Italy.
- Bruton, J.M., R.W. McClendon and G. Hoogenboom, 2000. Estimating daily pan evaporation with artificial neural networks. *Trans. ASAE*, 43: 491-496.
- Donatelli, M., J. Bolte, F. van Evert and E. Wang, 2003. Modelling cropping systems: Science, software and applications. *Eur. J. Agron.*, 18: 193-195.
- Donatelli, M., A. Omicini, G. Fila and C. Monti, 2004. Targeting usability and replaceability of simulation models for agricultural systems. *Proceedings of the 8th European Society for Agronomy Congress*, July 11–15, Copenhagen, Denmark, pp: 237-238.
- Eslamian, S.S., S.A. Gohari, M. Biabanaki and R. Malekian, 2008. Estimation of monthly pan evaporation using artificial neural networks and support vector machines. *J. Applied Sci.*, 8: 3497-3502.
- Fernandes, C., J.E. Cora and E. Araujo, 2003. Reference evapotranspiration estimation inside greenhouse. *Sci. Agric.*, 60: 591-594.
- Jones, J.W., B.A. Keating and C.H. Porter, 2001. Approaches to modular model development. *Agric. Syst.*, 70: 421-443.
- Kisi, O., 2006. Daily pan evaporation modelling using a neuro-fuzzy computing technique. *J. Hydrol.*, 329: 636-646.
- Lindsey, S.D. and R.K. Farnsworth, 1997. Sources of solar radiation estimates and their effect on daily potential evaporation for use in streamflow modeling. *J. Hydrol.*, 201: 348-366.
- Rosenberry, D.O., T.C. Winter, D.C. Buso and G.E. Likens, 2007. Comparison of 15 evaporation methods applied to a small mountain lake in the Northeastern USA. *J. Hydrol.*, 340: 149-166.
- Schuch, U.K. and D.W. Burger, 1997. Water use and crop coefficients of woody ornamentals in containers. *J. Am. Soc. Hort. Sci.*, 122: 727-734.
- Sudheer, K.P., A.K. Gosain, D.M. Rangan and S.M. Saheb, 2002. Modeling evaporation using an artificial neural network algorithm. *Hydrol. Process.*, 16: 3189-3202.
- Tay, F.E.H. and L. Cao, 2001. Application of support vector machines in financial time series forecasting. *Omega*, 29: 309-317.
- Troya, J.M. and A. Vallecillo, 2001. Controllers: Reusable wrappers to adapt software components. *Inform. Software Technol.*, 43: 189-202.
- Walter, I.A., R.G. Allen, R. Elliott, D. Itenfisu and P. Brown *et al.*, 2002. The ASCE standardized reference evapotranspiration equation. *Proceeding of the Rep. Task Com. On Standardized Reference Evapotranspiration*, July 9, American Society of Civil Engineers, Reston, VA, USA, pp: 1-70.
- Xu, C.Y. and V.P. Singh, 1998. Dependence of evaporation on meteorological variables at different time-scales and intercomparison of estimation methods. *Hydrol. Process.*, 12: 429-442.