Regression Models for Predicting Reference Evapotranspiration

N.Manikumari*, G.Vinodhini**
*Department of Civil Engineering,
** Department of Computer Science & Engineering,
Faculty of Engineering & Technology, Annamalai University.

Abstract

The FAO-56 Penman-Monteith method is a universally adopted method for computing Reference crop Evapotranspiration (ET_0) approved worldwide to arrive Crop Evapotranspiration. The main focus of this work is to model the Reference evapotranspiration by three different regression models including the Multi Linear Regression (MLR) and Support Vector Regression (SVR) models. The analysis was carried out based on the data collected in the command area of Veeranam tank system during the period 1987 –2008 in India. Veeranam tank is one of the major irrigation tanks that forms part of Cauvery basin, the largest river basin of Tamilnadu, India. The results were compared with Penman–Monteith (FAO-PM) to select the best model. The results show that all models provided a closer agreement with the calculated values for FAO-PM (R^2 > 0.98). However, the SVR model gives better estimates than the other two models for estimating Reference evapotranspiration (ET_0). The results showed that both SVR models provided better Root Mean Square (RMSE), Mean Absolute Error (MAE) and Correlation coefficient (R^2).

***Keywords**: water; irrigation; regression; evapotranspiration*

1. Introduction

Evapotranspiration (ET) represents the sum of water consumed by the process of transpiration by plants and evaporation from the adjacent irrigated soil. ET continues to be the parameter of foremost importance in irrigated agriculture. The initial step in analyzing water losses from the land surface is to estimate the Reference crop Evapotranspiration (ET_0) [8]. The United Nations Food and Agriculture Organization (FAO) Irrigation and Drainage Paper No. 56 (Allen et al., 1998) presents guidelines for computing crop water requirements. It incorporated advances in research and included an accurate procedure for calculating Reference ET and Crop ET from meteorological data and crop coefficients. Various mathematical methods for estimating the ET_0, such as the Penman-Monteith, Makkink, Priestley-Taylor, Turc, Hargreaves-Samani, Thornthwaite and Blaney-Criddle methods were developed long back. In many locations, the complete dataset of meteorological variables required for PMF56 calculations is not available. Therefore, simple empirical evapotranspiration models, using a lesser number of input variables are used instead [4]. Today, applications of modeling and regression methods have been widely spread in different fields of science and engineering. These methods are applied for increasing the quality of forecasts using the sample data sets from experiments and different investigations. Appropriate use of these methods is very useful to resolve uncertainties and predict results, when obtaining of information is difficult and sometimes impossible. Therefore, the main purpose of this work is to predict the ET_0 value by employing multiple linear regression and support vector regression for conditions in the study location using the Penman-Monteith model as a reference [2,3]. This approach ensures that the evapotranspiration values calculated using regression models will agree better with the values predicted by the Penman-Monteith model and be more reliable. The paper is organized as follows: Section 2 describes the background of the work investigated. The methods and materials used in this research are explained in Section 3. Section 4 presents the results obtained in this analysis.

2. Background

Tabari, H. et al. [7] adopted a method to estimate daily ET_0 using Artificial Neural Network (ANN) and Multivariate Non-Linear Regression (MNLR) methods in the semi-arid region of Iran. Five different ANN and MNLR models comprising various combinations of daily meteorological variables that are relative humidity, air temperature, solar radiation, wind speed and precipitation were developed to evaluate the degree of effect of each of these variables on ET_0.

Cobaner, M. [1] analyzed the efficiency of a Wavelet Regression (WR) model in estimating reference evapotranspiration based on only Class A pan evaporation. The results of the WR model were compared with those of three pan-based equations, namely the FAO-24 pan, Snyder ET_0 and Ghare ET_0 equations and their calibrated versions. Daily Class A pan evaporation data from the Fresno and Bakersfield stations of the United States Environmental Protection Agency in California, USA, were used. The WR model estimates were compared with those of the FAO-56 Penman–Monteith equation. Results showed that the WR model is capable of accurately predicting the ET_0 values as a product of pan evaporation data.

Medeiros, P. V. et al. [5] described two correction procedures for potential evapotranspiration estimates by temperature, making the results more reliable. Initially, the standard FAO Penman-Monteith method was evaluated with a complete climatologic data set for the period between 2002 and 2006. Then temperature-based estimates by Camargo and Jensen-Haise methods have been adjusted by error autocorrelation evaluated in biweekly and monthly periods. In a second adjustment, simple linear regression was applied. The adjusted equations have been validated with climatic data available for the year 2001. Both proposed methodologies showed good agreement with the standard method indicating that the methodology can be used for local potential evapotranspiration estimates.

Silva, H. J. F. et al. [6] estimated the reference evapotranspiration using FAO Penman-Monteith method. Multiple regression analysis was used as a selection process of significant variables for the model fit. Generated values by the proposed evapotranspiration models were compared to observed values for validation. Results indicated that the model with three variables (mean temperature, wind speed and insolation) satisfactorily estimated reference evapotranspiration with great performance for annual data.

However, these studies focus on limited regression methods and use less recorded data. Furthermore, comparison of the regression methods such as support vector regression has not been investigated much. In the present paper, attempts are made to apply multi linear regression (MLR) and support vector regression (SVR) models to estimate reference evapotranspiration values.

3. Methods and Material

The command area of Veeranam tank was taken for the study. Veeranam tank system, one of the major irrigation tanks of Tamilnadu forms part of Cauvery basin, the largest river basin of Tamil nadu. The sub-basin to which it belongs is referred as Udayarpalayam-Cauvery basin, a minor basin of Coleroon basin. Veeranam tank system is the largest tank irrigation system in India in terms of the area irrigated and number of canals, and is the second largest (1465 Mcft) in terms of tank capacity, next to Chembarambakkam tank (3645 Mcft) which is in Chennai (Arumugam and Mohan, 1997). The daily meteorologic and lysimetric data measured at Indian Meteorological Observatory, Annamalainagar of Tamilnadu is taken for the computation of Evapotranspiration for lowland rice. The FAO Penman-Monteith method is the standard method for computing ET_o, as recommended by FAO (FAO, 1977) since it involves many climatological parameters representative of the location. The Penman-Monteith method adopted in this study considers the crop physiological factors also, in addition to the meteorological factors considered in the Modified Penman method. Annamalainagar is located at Chidambaram, Cuddalore District at $11^o25'$ North latitude and $79^o44'$ East longitude. In this study, daily data on maximum temperature, minimum temperature, maximum relative humidity, minimum relative humidity, actual bright sunshine hours and wind speed observed at Indian Meteorological Observatory, Annamalainagar were collected for computing Daily Reference crop ET (ET_0).

Dataset was derived based on the period of the commencement of the release of water from the tank in different seasons from 1987-2008. The tank was operated effectively for irrigation for three of the release groups July, August and September group. Reference evapotranspiration measured during these water release commencement period was alone considered. The detailed description of the dataset used is given in Table 1.

Table 1. Description of dataset

Properties	Value
Number of instances	552
Number of attributes	10 (9 + 1)
Period of analysis	1987-2008
Training instances	386
Test instances	166
Dependent variable	ET_0

The statistical characteristic of attributes for the dataset is provided in Table 2. The statistical properties of data models show that each of the attribute has its own uniqueness in terms of mean and the range.

Table 2. Statistical properties of attributes

Attribute	Description	Data type	Mean	Range
RAIN	Rainfall	real	2.497 +/- 9.094	[0.000 ; 76.000]
RHMX	Maximum Relative Humidity	integer	80.322 +/- 12.880	[8.000 ; 100.000]
RHMN	Minimum Relative Humidity	integer	54.792 +/- 14.709	[29.000 ; 100.000]
WS	Wind Speed	real	7.208 +/- 3.096	[0.000 ; 21.700]
TMAX	Maximum Temperature	real	26.910 +/- 13.075	[0.000 ; 50.700]
TMIN	Minimum Temperature	real	26.917 +/- 4.962	[19.000 ; 37.600]
SSH	Sunshine Hours	real	11.926 +/- 8.268	[0.000 ; 27.200]
EVP	Evaporation	real	5.177 +/- 6.052	[0.000 ; 82.800]
AET	Actual Evapotranspiration	real	0.493 +/- 1.650	[0.000 ; 8.500]
ET_0	Reference Evapotranspiration	real	4.632 +/- 1.460	[1.906 ; 7.982]

The regression methods used in the experimental analysis are multiple linear regression and support vector regression.

a. Multiple Linear Regression (MLR)

Multiple linear regression techniques are used to model the reference evapotranspiration data in terms of the local climatological parameters of temperature, relative humidity, and wind speed. For a multiple linear regression model, the dependent variable y is assumed to be a function of k independent variables $x_1, x_2, x_3, \ldots, x_k$. The model is expressed in the form

$$y_i = b_0 + b_1 x_{1,i} + \cdots + b_k x_{k,i} + e_i\ ,$$

where $b_0, b_1, \ldots$ and b_k are fitting constants; $y_i, x_{1,i}, \ldots, x_{k,i}$ represent the ith observations of each of the variables y, $x_1, \ldots x_k$ respectively; e_i is a random error term representing the remaining effects on y of variables not explicitly included in the model. For simple regression models, ei can be assumed to be an uncorrelated variable with zero mean. The most common procedure for estimating the values of b_0, $b_1, \ldots b_k$ is to employ the least squares criterion with the minimum sum of squares of error terms.

b. Support Vector Regression

Support vector regression (SVR) is a widely used regression technique. It is extended from support vector classification. The problem of regression is that of finding a function which approximates mapping from an input domain to the real numbers on the basis of a training sample. The loss function determines this measure. Each choice of loss function will result in a different overall strategy for performing regression. Support vector regression performs linear regression in the attribute space using ε - insensitive loss function and, at the same time, tries to reduce model complexity by minimizing $\| w \|^2$. This can be described by introducing (non-negative) slack variables ξ_i, ξ_i^* i=1,…n to measure the deviation of training samples outside the ε – insensitive zone. The SVR is formulated as

$$\text{Min} \frac{1}{2} \| w \|^2 + C \sum_{i=1}^{n} (\xi_i + \xi_i^*)$$

Subject to the following conditions,

$$w^T \emptyset(x_i) - y_i \leq \epsilon + \xi_i$$

$$y_i - w^T \emptyset(x_i) - y_i \leq \epsilon + \xi_i$$

$$\xi_i, \xi_i^* \geq 0, i = 1 \ldots\ldots . m$$

Φ maps the data point x into high-dimensional attribute space and linear regression with ε (insensitive loss function) is performed in that attribute space. As a consequence, the dimension of w is that of φ(w) and hence there is no attempt to express the target function in terms of w.

4. Results & Discussion

The results obtained for a multiple linear regression model run on Weka tool is shown below in Fig. 1. For evaluation, 70% of the dataset was used as training set and 30% was used as test set. The coefficients are in fact the weights that are applied to each attribute before adding them together. In Fig. 1 the attribute SSH (Sunshine Hours) has larger influence on predicting the reference evapotranspiration.

ET_0 = (-0.0001 * RAIN) + (0.0018 * RHMX) + (0.005 * RHMN) + (-0.0227 * WS) + (0.0569 * TMAX) + (0.0571 * TMIN) + (0.2352 * SSH) + (0.0017 * EVP) –(0.0045 * AET) -1.4457

Fig.1. MLR equation for ET_0 prediction

A one unit increase in SSH is associated with a 0.235 unit increase in ET_0 value holding other attributes for ET_0 as constant. The multiple regression equation thus obtained can be used to estimate ET_0 as a function of Rain, RHMX, RHMN, WS, TMax, TMin, SSH, EVP and AET. The inclusion of all 9 attributes in the generated multi linear regression equation shows that all of the 9 attributes are independent of each other. Each of the contributing attributes differs in their weight values. The equation generated for Support Vector Regression is shown in Fig. 2.

ET_0 = (0.0107 * (normalized) RAIN) +(0.0413 * (normalized) RHMX) + (0.061 * (normalized) RHMN) - (0.1135 * (normalized) WS) + (0.5317 * (normalized) TMAX) + (0.1969 * (normalized) TMIN) + (1.0317 * (normalized) SSH) + (0.1415 * (normalized) EVP) – (0.0084 * (normalized) AET) - 0.3953

Fig.2. SVR equation for predicting ET_0

In the Fig. 2 the attribute SSH and TMAX have larger influence on predicting the reference evapotranspiration. The support vector regression equation thus obtained does not omit any of the nine attributes used. This indicates that all the nine attributes have a contributing role in predicting ET_0. Each of the contributing attributes differs in their weight values.

The predicted and actual ET_0 values are plotted in Fig.3 and Fig.4 for MLR and SVR models respectively. The plot obtained for MLR is highly scattered than the SVR plot as shown in Fig.3. Here, the solid line represents the condition of perfect agreement, and the points plotted away from the solid line represent discrepancies of ET_0 . It is seen that the higher scatter in reference evapotranspiration value is evident in Fig.3 for MLR model.

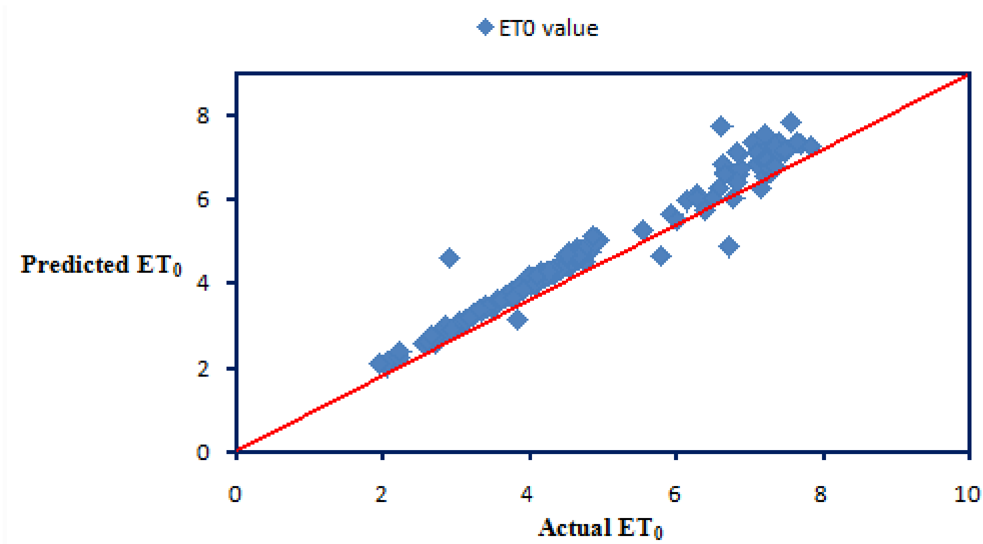

Fig.3. Actual Vs Predicted ET_0 for MLR model

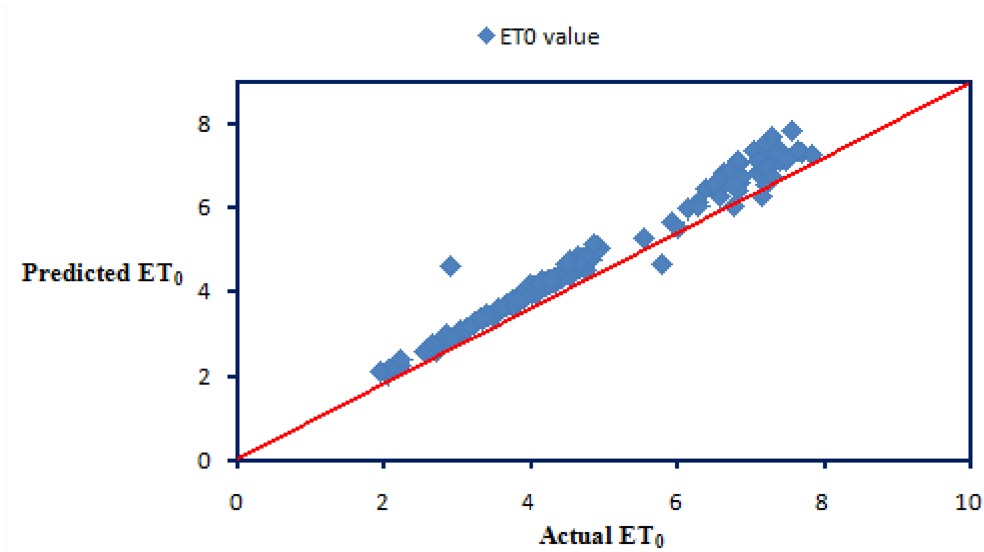

Fig.4. Actual Vs Predicted ET_0 for SVR model

The results obtained from the P–M empirical equation were to be applied in the proposed regression models for prediction of ET_0. The accuracy of the predicted data was compared with the FAO-56 PM ET_o using the RMSE, MAE and the R. The RMSE, MAE and R of a model prediction with respect to the actual value is given as

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

$$MAE = \frac{1}{N}\sum_{i=1}^{N}| ETi_{calculated} - ETi_{predicted} |100$$

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(ETi_{calculated} - ETi_{predicted})^2}{\sum_{i=1}^{N}(ETi_{calculated} - meanET_{predicted})^2}$$

Where $ETi_{calculated}$ represents the ET_o obtained from FAO-56 PM method and $ETi_{predicted}$ represents the model predicted output ET_0. Table 3 and Fig. 5 show the results obtained for regression models evaluated in terms of two error metrics and regression coefficient.

Table 3. Performance of Regression models

Methods	RMSE	MAE	Regression Coefficient
MLR	0.26	0.14	0.98
SVR	0.19	0.10	0.99

The multiple regression statistics reveals that the value of coefficient of determination (R^2) is 0.98 and it is significant at 1% level. The values of RMSE of SVR model have also reduced slightly by 7%. The values of MAE of SVR model have also reduced to 0.10 from 0.14 of MLR model. This may be due to the fact that ET_0 values do not exhibit much of nonlinearity. The study reveals that the support vector regression model proposed may be adopted satisfactorily in the ET_0 estimation at the location selected for the present study.

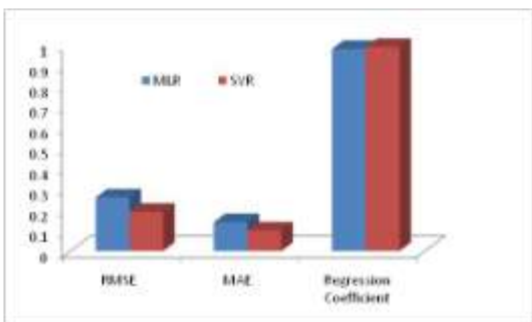

Fig. 5. Comparison of regression models

Table 4 shows the time taken to build the model for the two regression models employed. The values clearly depict that SVR model takes less time compared to MLR model. Thus the SVR is superior in terms of all the performance measures.

Table 4. Time taken by the models.

Models	Time taken to build model (sec)
MLR	1.32
SVR	0.56

5. Conclusion

The present work focused on the development of Multiple Linear Regression (MLR) and Support Vector Regression (SVR) models for predicting Reference Evapotranspiration. The developed regression models of Reference Evapotranspiration were applied on the data collected from Veeranam tank area. The experimental analysis showed that climatic parameters mostly influenced ET_0 estimation at the study location. The linear regression models were proposed in terms of the climatic parameters in the ET_0 estimation. The SVR model proposed showed a marginal improvement over MLR models. The SVR model may therefore be adopted for better estimation of ET_0.

References

[1]. Cobaner, M. (2013). Reference evapotranspiration based on Class A pan evaporation via wavelet regression technique. Irrigation Science, 31(2), 119-134.

[2]. Jadhav, P. B., Kadam, S. A., & Gorantiwar, S. D. (2015). Comparison of methods for estimating reference evapotranspiration for Rahuri region. Journal of Agrometeorology, 17(2), 204.

[3]. Jhajharia, D., Dinpashoh, Y., Kahya, E., Singh, V. P., & Fakheri-Fard, A. (2012). Trends in reference evapotranspiration in the humid region of northeast India. Hydrological Processes, 26(3), 421-435.

[4]. 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.

[5]. Medeiros, P. V., Marcuzzo, F. F. N., Youlton, C., & Wendland, E. (2012). Error Autocorrelation and Linear Regression for Temperature-Based Evapotranspiration Estimates Improvement1. JAWRA Journal of the American Water Resources Association, 48(2), 297-305.

[6]. Silva, H. J. F., Santos, M. S., Cabral Junior, J. B., & Spyrides, M. H. C. (2016). Modeling of reference evapotranspiration by multiple linear regression. Journal of Hyperspectral Remote Sensing, 6(1), 44-58.

[7]. Tabari, H., Marofi, S., & Sabziparvar, A. A. (2010). Estimation of daily pan evaporation using artificial neural network and multivariate non-linear regression. Irrigation Science, 28(5), 399-406.

[8]. Vicente-Serrano, S. M., Van der Schrier, G., Beguería, S., Azorin-Molina, C., & Lopez-Moreno, J. I. (2015). Contribution of precipitation and reference evapotranspiration to drought indices under different climates. Journal of Hydrology, 526, 42-54.