Estimation of Potential Evapotranspiration by Multiple Linear Regression Method

Sriram A. V.[1], Rashmi C. N.[2]

[1](Assistant Professor, Department of Civil Engineering, UVCE, Bangalore University)
[2](Assistant Professor, Department of Civil Engineering, PESIT, Bangalore)

ABSTRACT: *Prediction of evapotranspiration is important for study, design and management of irrigation systems. In this study, the potential of Multiple Linear Regression model using least squares is investigated in modeling of mean monthly evapotranspiration (PET) obtained using the standard FAO-56 Penman–Monteith equation. Various combinations of daily climatic data, namely solar radiation, air temperature, relative humidity and wind speed, are used as inputs to the model, so as to evaluate the degree of effect of each of these variables on PM estimated PET. Residual Analysis, Co-efficient of Determination and Residual Error is used as comparison criteria for evaluation of the model performance. The Multiple Linear Regression model gave Residual errors of 0.83mm/day and Co-efficient of Determination of 0.995 for the meteorological station considered. Based on the residual analysis results it was found that the Multiple Linear Regression model followed a linear trend. The input variables of the model were adequate and this model could be successfully employed in estimating the monthly Potential Evapotranspiration in the study area.*
Keywords: Method of Least squares, Multiple Linear Regression, Potential Evapotranspiration.

I. Introduction

Evapotranspiration (ET) is the sum of volume of water used by vegetation, evaporated from the soil, and intercepted precipitation [1]. ET plays an important role in our environment at global, regional, and local scales. In many areas where water resources are scarce, the calculation of this loss becomes imperative in the planning and management of irrigation practices [2].Accurate quantification of evapotranspiration is crucial for better management and allocation of water resources.

Since there are many difficulties for the direct measurement of PET, it is usually estimated using the standardized, physically based equation of FAO Penman-Monteith Equation[3] which is widely accepted to provide the best estimates of PET over various climatic types [4,5]. This equation, nevertheless, is demanding in terms of input data, which are not measured in the majority of the weather stations. To overcome this deficiency, a number of empirical equations, less demanding in terms of input data, has been proposed for the estimation of PET. Yet, due to the fact that these are developed for specific regions and climatic types it is possible that their estimates will be deemed by source of bias deriving from site-specific conditions. It is important therefore, besides the general evaluation of the empirical equation prior to their use, to detect and identify the specific parameters that lead in partial failure of the empirical equations even when they overall perform well.

We often use regression analysis to model the relationship between dependent (response) and independent (explanatory) variables. In traditional regression analysis, residuals are assumed to be due to random errors. Thus, statistical techniques are applied to perform estimation and inference in regression analysis. [6] and [7] employed regression methods to develop models for regional monthly average evaporation in USA as a function of readily available variables such as temperature and site's longitude and elevation. The empirical models were shown to be an improvement over other temperature-based models such as that of Linacre and that of Hargreaves and Samani.

The objectives of this present study are to generate PET values using FAO 56 Penman-Monteith equation and to develop a multiple linear regression model for the same. The linearity is checked by residual analysis and Comparison for the accuracy of MLR model is done with PET measured by FAO 56 Penman-Monteith method in the humid city of Chennai in South India.

II. Methodology

2.1 Introduction

The monthly potential evapotranspiration values are calculated using the daily values for Meenambakkam; Chennai meteorological station for a data of 5 years (2007-2011) by FAO 56 Penman-Monteith equation and this was compared with a multiple linear regression model for Evapotranspiration developed using Least-squares method, whose linearity is ascertained by residual Analysis.

2.2 Penman-Monteith equation

This equation is used to predict evapotranspiration from standard climatological records of sunshine, temperature, humidity and wind speed. This approach includes all parameters that govern energy exchange and corresponding latent heat flux. The Penman-Monteith form of the combination equation is given as

$$ET_o = \left[\frac{0.408\, \Delta\, (R_n - G) + \gamma \left(\dfrac{900}{T + 273} \right) u_2\ (VPD)}{\Delta + \gamma\ (1 + 0.34\, u_2)} \right] \tag{1}$$

Where:

ET_o	= daily reference ET [mm/d], for longer periods 900 becomes 37
T	= mean air temperature at 2m high [°C]
VPD	= vapor pressure deficit [kPa]
u_2	= wind speed at 2 m high [m/s] = 2 m/s
R_n	= net radiation at the crop surface [MJ m^{-2} d^{-1}]
Δ	= slope vapour pressure curve [kPa °C^{-1}]
γ	= psychometric constant [kPa °C^{-1}]
G	= soil heat flux density [MJ m^{-2} d^{-1}]

2.3 Multiple Linear Regression Methodology

The meteorological data of the study area can be examined for the suitability of fitting a multiple linear regression model. The relation between potential evapotranspiration and the other climatological parameters of temperature, relative humidity, and wind speed on a monthly basis can be obtained. For a multiple linear regression model, the dependent variable is assumed to be a function of several independent variables. The general form of the equation is written as

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2, + \ldots + \beta_p X_p + \varepsilon \tag{2}$$

Where, β_0, β_1... β_p are fitting constants; p represents the pth observations of each of the variables x1, x2, x3, ... xp respectively; ε is a random error term representing the remaining effects on of variables not explicitly included in the model. The most common procedure for estimating the values of β, and ε is to employ the least squares criterion with the minimum sum of squares of error terms.

The least squares function is

$$L = \sum_{i=1}^{n} \epsilon_i^2 = \sum_{i=1}^{n} \left(y_i - \beta_0 - \sum_{j=1}^{k} \beta_j x_{ij} \right)^2 \tag{3}$$

We want to minimize L with respect to β0, β1, and, Xk. The least squares estimates of β0, ε1, p, Xk must satisfy

$$\frac{\partial L}{\partial \beta} = 0 \tag{4}$$

Simplifying the equation we obtain the Least squares Normal Equations as

$$n\hat{\beta}_0 + \hat{\beta}_1 \sum_{i=1}^{n} x_{i1} + \hat{\beta}_2 \sum_{i=1}^{n} x_{i2} + \cdots + \hat{\beta}_k \sum_{i=1}^{n} x_{ik} = \sum_{i=1}^{n} y_i$$

$$\hat{\beta}_0 \sum_{i=1}^{n} x_{i1} + \hat{\beta}_1 \sum_{i=1}^{n} x_{i1}^2 + \hat{\beta}_2 \sum_{i=1}^{n} x_{i1} x_{i2} + \cdots + \hat{\beta}_k \sum_{i=1}^{n} x_{i1} x_{ik} = \sum_{i=1}^{n} x_{i1} y_i$$

$$\vdots \qquad \vdots \qquad \vdots \qquad \vdots \qquad \vdots$$

$$\hat{\beta}_0 \sum_{i=1}^{n} x_{ik} + \hat{\beta}_1 \sum_{i=1}^{n} x_{ik} x_{i1} + \hat{\beta}_2 \sum_{i=1}^{n} x_{ik} x_{i2} + \cdots + \hat{\beta}_k \sum_{i=1}^{n} x_{ik}^2 = \sum_{i=1}^{n} x_{ik} y_i \tag{5}$$

This model is a system of *n* equations that can be expressed in matrix notation as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \tag{6}$$

Where,

$$
\mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \quad
\mathbf{X} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1k} \\ 1 & x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nk} \end{bmatrix} \quad
\boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} \quad \text{and} \quad
\boldsymbol{\epsilon} = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix} \tag{7}
$$

2.4 Residual Analysis

The residuals from the multiple regression model, defined by ε, play an important role in judging model adequacy. The standard regression model assumes that the residuals, or ε's, are independently, identically distributed as normal with $\mu = 0$ and variance σ^2.

2.5 Checks to be done on a multiple Linear Regression Model.
(a) The association is not linear: - This check is done by looking at the scatter plot of X and Y. If it doesn't look like a straight line, then a linear regression should not be run.
(b) Error terms do not have constant variance: - This can be observed in the residual plots. We can detect this by plotting the residuals against the predictor variable. The residual plot should have near constant variance and look like a horizontal band of points.
(c) The error terms must be independent: - we can perform a scatter plot of residuals against time.
(d) Outliers: - (scatter plot of Y and X, examining the numerical value of the residuals, plotting residuals against the predictor).
(e) Error terms (residuals) are not normally distributed: - This is checked by the normal-normal plot.
(f) Wrong structural model (mis-specified model):- we can also use residuals to check whether an additional variable should be added to a regression equation.

III.   Study Area And Data Used
3.1 Study Area: The study area is Meenambakkam, Chennai, Tamilnadu State lying in tropical climates of India having sub humid climatic conditions. In the present study, a multiple linear regression model for evapotranspiration is developed for the study area. The general details of the meteorological station are summarized in Table 1.

Table 1: General details of the study area

Location	Climate Type	Latitude	Longitude	Altitude in m
Chennai	Humid	13^oN	80^o11'E	15

3.2 Data used: The data was obtained from the weather station located at Meenambakkam airport for a span of 5 years (2007-2011). Fig. 1 gives the Average annual values of mean temperature, Minimum Humidity, Maximum Humidity and wind speed data of the meteorological station. It is seen that as mean temperature increases, relative humidity decreases and vice-versa for the data obtained for the meteorological station.
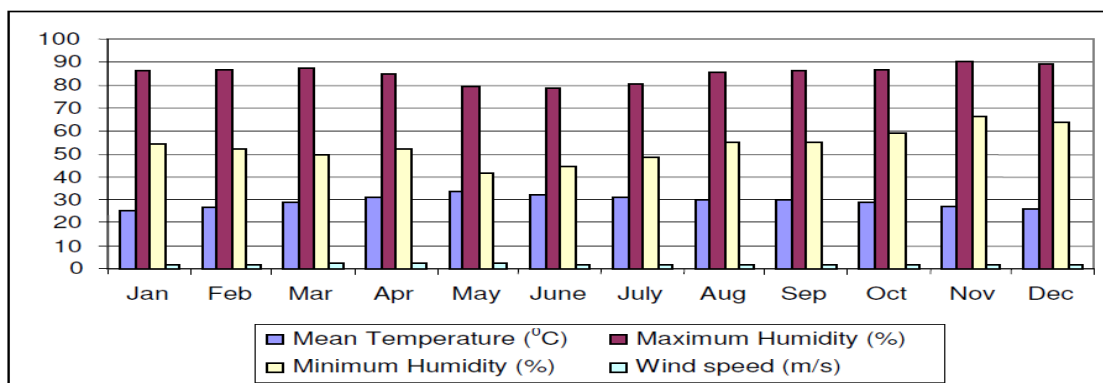

Figure 1: Average Annual values of Mean Temperature, Maximum Humidity, Minimum Humidity and Wind speed for Chennai (2007 - 2011)

IV.   Results and Discussions

Table 2: PET values in mm/day at Chennai Meteorological station by FAO 56 PM Method.

Month	Chennai				
	2007	2008	2009	2010	2011
January	3.27	3.09	3.24	3.24	3.41
February	3.95	3.69	4.06	3.94	3.73
March	4.51	4.23	4.72	5.20	4.78
April	5.31	4.63	5.41	6.02	4.97
May	6.94	**7.14**	6.46	6.05	6.01
June	5.06	5.63	6.54	4.92	5.74
July	4.90	5.12	6.45	4.17	4.82
August	4.21	4.39	4.87	4.48	4.20
September	4.22	4.58	4.52	3.85	4.45
October	3.52	3.52	4.35	3.75	3.50
November	3.03	3.14	2.66	2.89	2.93
December	2.69	3.04	2.73	**2.58**	2.97

Table 2 shows the PET values obtained for a period of 5 years. It can be seen that high values are obtained in the months of April, May and June which may be due to the fact that these months have higher temperature, lesser humidity and high wind speeds compared to other months. Amongst all the years of data considered the highest value was obtained in May 2008 and the lowest in December 2010 which has highest and lowest temperature, RH and wind speed in the dataset and November 2010 has recorded low temperature and wind speed and high relative humidity.

4.1 Multiple Linear Regression Model:
Multiple linear regression techniques can be used to model the evapotranspiration data for the study region in terms of the local climatological parameters of Relative Humidity(X_1), Temperature(X_2), Wind speed(X_3) and Solar Radiation(X_4). The results obtained for a multiple linear regression model run on Microsoft excel software are shown below. The Multiple Linear regression model obtained is

$Y = 1.787 - (0.0645X_1) + (0.1353X_2) + (0.7176X_3) + (0.1846X_4)$

Table 3: Model of the Multiple Linear Regression for the Meteorological station

Station	Model	R^2	$e = Y_i - \hat{Y}_i$
Chennai	$Y = 1.787 - (0.0645X_1) + (0.1353X_2) + (0.7176X_3) + (0.1846X_4)$	0.9954	0.83

Table 4: Summary statistics of Multiple Regression model Residual Analysis

Regression Statistics	
Multiple R	0.996661
R Square	0.993332
Adjusted R Square	0.992847
Standard Error	0.097589
Observations	60

ANOVA

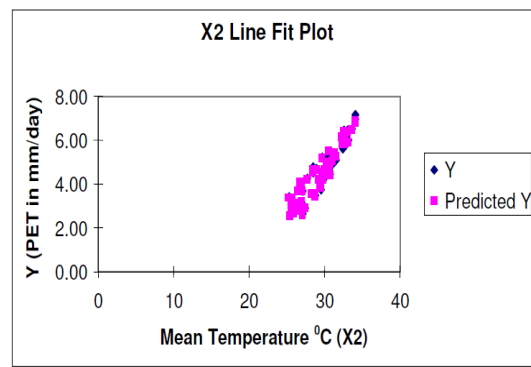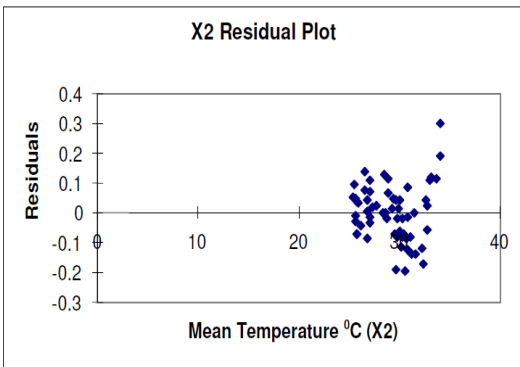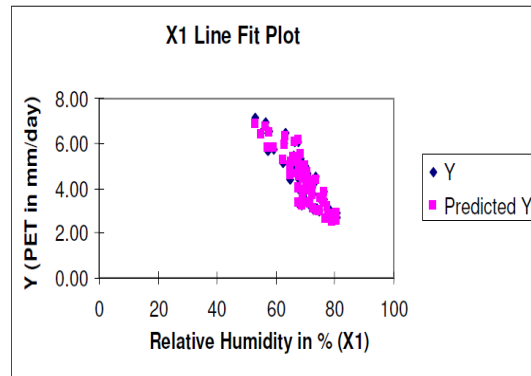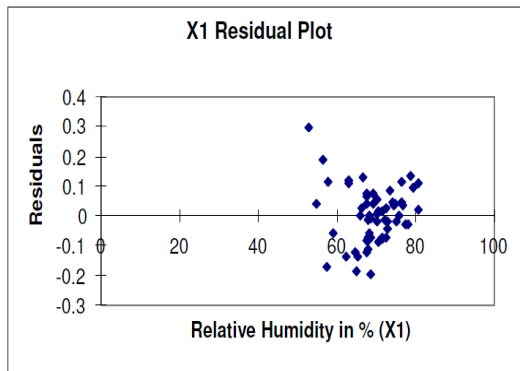|	df	SS	MS	F	Significance F
Regression	4	78.0324307	19.50811	2048.393	4.08882E-59
Residual	55	0.52379882	0.009524		
Total	59	78.5562295			
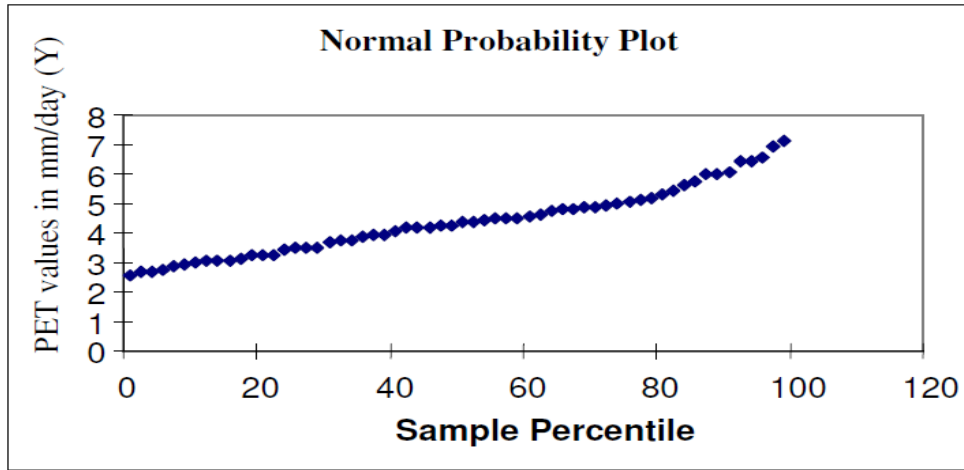
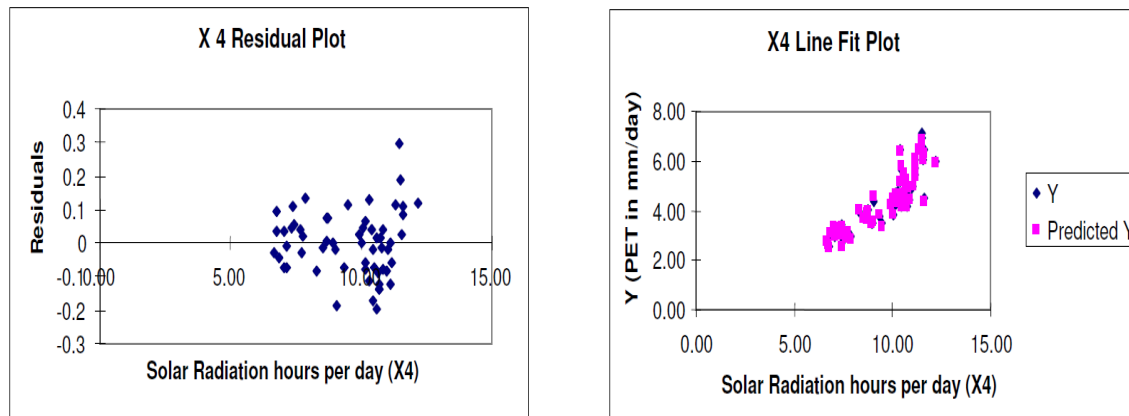Figure 2: Normal probability plot for Chennai Meteorological station
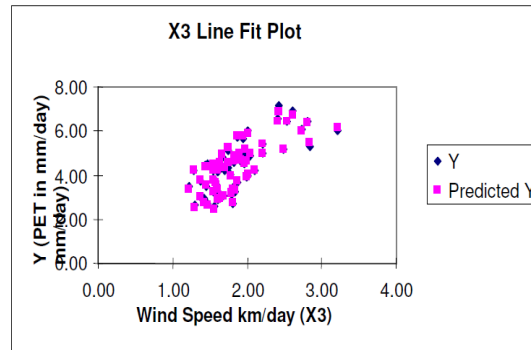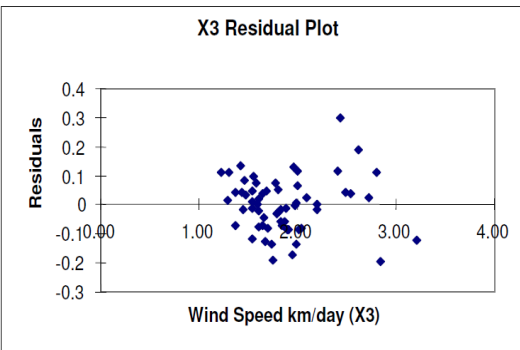
Figure 3: Residual and Line of Fit plots for Variables X1(Relative Humidity), X2(Mean Temperature), X3(Wind speed) and X4(Solar Radiation) for the study Area.

It can be seen from figures 2 and 3 and Table 4 that for Chennai Meteorological station,

- **The association is linear**: - The scatter plot of X and Y is a straight line for both the Meteorological stations (Figure 2)
- **Error terms have constant variance**: - This can be observed by plotting the residuals against the predictor variable. The residual plot does not have a systematic pattern and the plot looks like a horizontal band of points (Figure 3)
- **Error terms (residuals) are normally distributed**: - Since the normal plot of residuals is a straight line for the station and the residuals are normally distributed (Figure 2)

V. Conclusions

From this study, it has been found that

- The PET values are obtained by FAO-56 full data Penman-Monteith equation and it was found that highest PET value was obtained in May 2008 (7.14) and the lowest value in December 2010 (2.58) for Chennai. Higher values of PET values indicate higher temperature, lower Humidity and higher wind speed as observed from the dataset.
- The multiple linear regression equation developed for the meteorological stations is found to be a good fit ($R^2 = 0.995$, e = 0.83) for Chennai meteorological station.
- A Residual Analysis performed shows that the association between the variables is Linear, the error terms have a constant variance, and residuals are normally distributed.

REFERENCES
[1]   Singh, V.P. (1988). Hydrologic system rainfall-runoff modeling. Vol. 1, Prentice Hall, ISBN: 0134480511, NJ, USA.
[2]   Kisi, O., (2009). Evapotranspiration modelling from climatic data using a neural computing technique Hydrological process, Volume 21, Issue 14, pp 1925-1934.
[3]   Allen, R.G, Pereira, L.S., Raes, D., and Smith, M., 1998: Crop evapotranspiration - Guidelines for computing crop water requirements - *FAO Irrigation and drainage paper 56*, FAO, Rome.
[4]   Droogers, P., and Allen, R.G., 2002. Estimating reference evapotranspiration under inaccurate data conditions. Irrig. Syst.16, 33-45.
[5]   Gavilan, P., Lorite, I. J., Tornero, S., and Berengena J. 2006. Regional calibration of Hargreaves equation for estimating reference ET in a semiarid environment, Agr. Water Manage. 81, 257–281.
[6]   Cahoon, J.E, Costello,T.A, Ferguson, J A., 1991. Estimating pan evaporation using limited meteorological observations. *Agricultural and Forest Meteorology*. 1991; 55(3-4):181–190.
[7]   Fennessey, N.M, Vogel, R.M., 1996. Regional models of potential evaporation and reference evapotranspiration for the northeast USA. *Journal of Hydrology*. 1996; 184(3-4):337–354.