The Korean Journal of Applied Statistics (2009)
22(3), 515–530

A Statistical Approach to Examine the Impact of Various Meteorological Parameters on Pan Evaporation

Swati Pandey[1] · Manoj Kumar[2] · Soubhik Chakraborty[3] · N.C. Mahanti[4]

[1]Department of Applied Mathematics, Birla Institute of Technology;
[2]Department of Applied Mathematics, Birla Institute of Technology;
[3]Department of Applied Mathematics, Birla Institute of Technology;
[4]Department of Applied Mathematics, Birla Institute of Technology

(Received November 2008; accepted December 2008)

Abstract

Evaporation from surface water bodies is influenced by a number of meteorological parameters. The rate of evaporation is primarily controlled by incoming solar radiation, air and water temperature and wind speed and relative humidity. In the present study, influence of weekly meteorological variables such as air temperature, relative humidity, bright sunshine hours, wind speed, wind velocity, rainfall on rate of evaporation has been examined using 35 years(1971–2005) of meteorological data. Statistical analysis was carried out employing linear regression models. The developed regression models were tested for goodness of fit, multicollinearity along with normality test and constant variance test. These regression models were subsequently validated using the observed and predicted parameter estimates with the meteorological data of the year 2005. Further these models were checked with time order sequence of residual plots to identify the trend of the scatter plot and then new standardized regression models were developed using standardized equations. The highest significant positive correlation was observed between pan evaporation and maximum air temperature. Mean air temperature and wind velocity have highly significant influence on pan evaporation whereas minimum air temperature, relative humidity and wind direction have no such significant influence.

Keywords: Pan evaporation, meteorological predictors, ANOVA test, regression analysis.

1. Introduction

One of the most important factors in agriculture is water availability. Water is provided to the crops naturally through precipitation and subsurface moisture, but when these supplies prove to be inadequate for crop use, growers must resort to irrigation. Water availability is also a critical variable for virtually every other economic activity, including industry, the energy sector, and public use. In recent years, water availability has become an issue even in the relatively wet state

[2]Corresponding author: Department of Applied Mathematics, Birla Institute of Technology, Mesra, Ranchi -835215, India. E-mail: msinha_09@rediffmail.com

of Louisiana as periods of prolonged drought have stressed both agriculture and nonagricultural sectors in Jharkhand (India). To schedule irrigation properly, a grower must know the environmental demand for surface water. For the grower, this surface water loss occurs primarily through evapotranspiration(ET). ET is simply the amount of water returned to the atmosphere through evaporation (moisture loss from the soil, standing water, *etc.*) and transpiration (biological use and release of water by vegetation) (Hansen *et al.*, 1980). Evaporation is an important component of the hydrologic cycle. Evaporation along with other meteorological parameters influences the microclimate of a region. The rate of evaporation is primarily controlled by incoming solar radiation, air and water temperature, wind speed and relative humidity. Also the amount of water lost by evaporation from any water body varies with its exposed surface area.

The ET rate is a function of factors such as temperature, solar radiation, humidity, wind, and characteristics of the specific vegetation that is transpiring, which may vary significantly between vegetation types (Allen *et al.*, 1998). If the demand for water(ET) exceeds the availability to the plant through precipitation or stored in the soil, then transpiration may stop resulting in crop loss. Therefore, reliable estimates of ET, along with knowledge of precipitation totals and soil moisture storage capacity, can provide estimates of water need via irrigation. Several methods are available to measure evapotranspiration directly. For instance, a lysimeter is used to measure ET by routinely measuring the change in soil moisture of known volume of soil that is covered with vegetation (Watson and Burnett, 1995). Lysimetry can be expensive both economically and in time investments to install, check, and maintain the equipment. Evaporation pans measure the loss of a known quantity of water through evaporation, but they do not measure transpiration, and therefore they must be adjusted using coefficients to represent ET (Dingman, 1994; Allen *et al.*, 1998; Barnett *et al.*, 1998). ET can also be measured by determining the flux of moisture from the vegetative surface to the atmosphere by using highly-sensitive sensors that detect the change in meteorological variables between the surface and a fixed level above the surface. The determination of ET through these flux-related methods, while highly accurate, can be difficult and are generally used only in research settings (Allen *et al.*, 1998; Geiger *et al.*, 2003). In summary, the measurement of ET can be difficult, requiring methods that are either labor or financially-intensive or that are indirect proxies of evapotranspiration.

To simplify the process of determining ET, models have been developed to estimate ET for use in environments that lack direct ET measurements. Many of these have been derived empirically through field experiments; others have been derived through theoretical approaches. A major complication in modeling ET is the requirement for meteorological data that may not be easily available (*e.g.* solar radiation). This restriction at times prohibits use of more accurate models, and necessitates use of models that have less demanding data requirements. A modification of the ET concept is reference evapotranspiration(ETo) that provides a standard crop (a short, clipped grass) with an unlimited water supply so that a user can calculate maximum evaporative demand from that surface for a given day. This value, adjusted for a particular crop, is the consumptive use (or demand), and deficit represents that component of the consumptive use that goes unfilled, either by precipitation or by soil-moisture use, during the given time period. This deficit value is the amount of water that must be supplied through irrigation to meet the water demand of the crop (Dingman, 1994; Allen *et al.*, 1998).

The climate in the Jharkhand state ranges from dry semi humid to semi arid types and is influenced by geographic location and physical features of the region (Kumar *et al.*, 2007). Many parts of the state experience agricultural drought like conditions during monsoon season due to deficit rainfall and excessive evaporation. Estimation of evaporation and its dependency on other meteorological parameters is therefore of prime importance in characterization of such agricultural drought. Estimation of evaporation from free water surfaces, from the ground and of evapotranspiration from vegetation-covered surfaces are also of great importance in hydrological modeling, hydro meteorological and agricultural studies.

A number of studies conducted world wide has indicated a declining trends in pan evaporation rate typically between 2 and 4mm a-2 for United States, former Soviet union and parts of Asia from 1950 to 1990 (Paterson *et al.*, 1995; Golubev *et al.*, 2001) and about 3mm a-2 for China from 1955 to 2000 (Liu *et al.*, 2004). In India reported decrease appears to be much larger at about 12mm a-2 for 1961–1992 (Chattopadhyay and Hulme, 1997). The vapour pressure deficit, wind speed, net radiation together with temperature determines the evaporative demands of the atmosphere (Rosenberg *et al.*, 1989).

Effect of meteorological parameters on pan evaporation has been studied by several workers in different parts of the world (Singh *et al.*, 1981; Khan, 1992; Khanikar and Nath, 1998; Sahu *et al.*, 1994; Hordofa *et al.*, 2004).

Hordofa *et al.* (2004) studied the influence of meteorological variables on monthly pan evaporation under sub-humid climatic conditions in Ethiopia. Based on twenty years(1981–2000) of meteorological data from three meteorological stations located at different agro climatic regions of Ethiopia, he opined that maximum temperature and relative humidity were highest correlated whereas minimum and mean air temperatures were least correlated wih pan evaporation in all the three regions in Ethiopia.

Although the inter relationship between dependent and independent meteorological variables using regression model have been established by earlier workers (Ali *et al.*, 2006) but no significant work has been done to check these models with time order sequence of residual plots and to develop new standardized regression models using standardized equations.

In the present study various meteorological parameters *i.e.*, rainfall, relative humidity, soil temperature, wind velocity, wind direction, dry bulb temperature, wet bulb temperature, maximum temperature, minimum temperature are used to evaluate their influence on pan evaporation using regression model based on 35 years of weather data recorded at Ranchi, the capital of Jharkhand state.

Statistical analysis was carried out employing linear regression models using Sigma plot (*ver.* 10.0) statistical software. Using the correlation matrix and by selecting parameters which exhibit significant correlations between them, the regression models have been developed using one response variable and one or more predictors. The developed regression models were tested for goodness of fit, multicollinearity along with normality test and constant variance test. Regression models were subsequently validated using the observed and predicted parameter estimates with the meteorological data of the year 2005. Further these models were checked with time order sequence of residual plots to identify the trend of the scatter plot and then new standardized regression models were developed using standardized equations.

Study area

Ranchi district of Jharkhand state has been selected for the present study. It is located between 22°34′N to 23°43′N latitude and 84°57′E to 85°54′E longitude and covering a total area of 7698sq.km. Topographically Ranchi district is plateau region with elevation ranging from 150 m to 600 m above mean sea level. The region is characterized by dry semi humid to semi arid climate with annual rainfall of about 1370mm received mainly during southwest monsoon season(June to September).

Materials and methodology

Evaporation rates can be measured by calculating the change of depth of water in a container over a given time period. Many U.S. weather stations determine evaporation rates using a U.S. Class-A pan, which is a stainless steel pan, 25.4 centimeters high and 1.2 meters in diameter. The pan, which is normally installed on a wooden platform set on the ground in a grassy open area, is filled with water to within 6.35 centimeters of the top and left exposed. The pan evaporation rate is simply the amount of water that evaporates from the pan in a given period of time. This rate is measured by manual readings or using an evaporation gauge. Water is added to the pan to bring it back to its original level each day.

In order to carry out the study, daily meteorological data pertaining to pan evaporation(PE), rainfall(Rf), relative humidity(RH), soil temperature(ST), wind speed(WS), wind direction(WD), dry bulb temperature(DBT), wet bulb temperature(WBT), maximum temperature(MaxT), minimum temperature(MinT), bright sunshine hours(BSS) were collected from meteorological observatory at Birsa Agricultural University, Ranchi for a period of 35 years(1971–2005).

A year was divided into 52 standard meteorological weeks(MSW). The normal ratio averaging method was employed to fill up the gaps in the data set. For this purpose mean value of meteorological parameters for the same week of the preceding year and the following were used. Thereafter the data series for various meteorological parameters for corresponding weeks were added together and average value for the respective weeks were found. In this way a 52 weeks data series for each meteorological parameters were generated.

In the present work, correlation matrix among the variables was evaluated for selection of significantly correlated meteorological parameters using Sigma plot (*ver.* 10.0). Using significant correlations between the two select parameter regression models were developed. These models were simulated with the observed data of year 2005 to calculate the predicted values using prediction equation. The difference of observed and predicted values were then plotted to check the time order sequence of the related residuals.

Further regression models, correlation coefficient, coefficient of determination, standard error, ANOVA table, F test, normality test, constant variance test and normality test of residuals were carried out using various parameters. A brief description of various statistical tools applied to the present work is given in the appendix.

2. Results and Discussions

Influence of meteorological variables on pan evaporation

The results of correlation and regression equations that defines the relationship between pan evaporation and meteorological parameters located in the dry semi humid to humid semi arid type climatic region of Jharkhand are given in Figure 2.1.

Maximum air temperature has shown consistently the highest positive correlation, 0.943 (Figure 2.1a) with pan evaporation among all other meteorological parameters. This indicates that evaporation is directly and primarily related to the air temperature and it increases as the air temperature

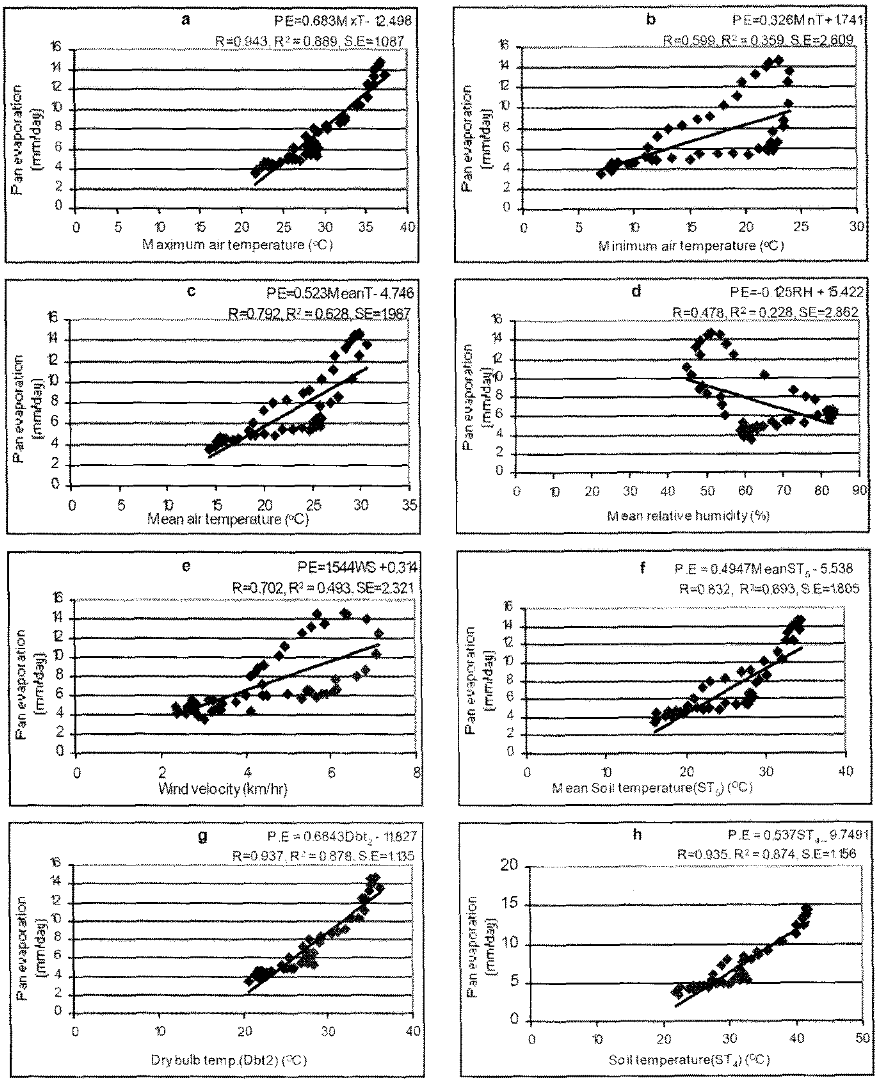

Figure 2.1. Scatter plot of pan evaporation with various meteorological parameters

increases. The warmer the air, the stronger the temperature gradient and therefore higher the rate of evaporation. Due to this the evaporation is competitively more in summer season than in winter season. Mean air temperature has shown significant influence on pan evaporation with correlation coefficient 0.792 (Figure 2.1c) whereas minimum air temperature exhibit a non-significant effect on pan evaporation ($R = 0.599$) as shown in figure1b.

Among the dry bulb temperatures DBT2 (Figure 2.1g) has shown highly significant correlation coefficient ($R = 0.937$) in comparison to DBT1 ($R = 0.660$). Wet bulb temperatures have non-significant correlation.

Wind velocity has also significant influence on pan evaporation ($R = 0.702$, Figure 2.1e) whereas wind direction has no significant influence on pan evaporation as shown by low correlation coefficient

Table 2.1. Regression models for pan evaporation(PE) using predictors

Eq No.	Prediction Equation	R^2	MSE	β estimates and their S.E. (respectively)	NT Normality test (overall)	CV Constant variance test	NT Normality test (residuals)
1	$0.683*\text{MaxT} - 12.498$	0.889	1.087	$\beta_0 = -12.498$, 1.004 $\beta_1 = 0.683$, 0.034	Failed $P = 0.010$	Passed $P = 0.662$	K.S. dist. = 0.106 $P = 0.147$ Passed
2	$-0.165*\text{WBT}_1 + 0.818*\text{MaxT} - 13.612$	0.931	0.864	$\beta_0 = -13.612$, 0.823 $\beta_1 = -0.165$, 0.030 $\beta_2 = 0.818$, 0.036	Passed $P = 0.583$	Passed $P = 0.085$	K.S. dist. = 0.071 $P > 0.200$ Passed
3	$2.516*\text{DBT}_1 - 2.239*\text{WBT}_1 - 0.163*\text{MaxT} + 1.474$	0.964	0.633	$\beta_0 = 1.474$, 2.370 $\beta_1 = 2.516$, 0.382 $\beta_2 = -2.239$, 0.316 $\beta_3 = -0.163$, 0.151	Passed $P = 0.096$	Passed $P = 0.188$	
3*a*	$-0.187*\text{DBT}_1 + 0.882^b\text{MaxT} - 14.659$	0.926	0.896	$\beta_0 = -14.659$, 0.935 $\beta_1 = -0.187$, 0.037 $\beta_2 = 0.882$, 0.049	Passed $P = 0.745$	Passed $P = 0.083$	K.S. dist. = 0.068 $P > 0.200$ Passed
4	$2.110*\text{DBT}_1 - 1.908*\text{WBT}_1 - 1.057$	0.963	0.634	$\beta_0 = -1.057$, 0.313 $\beta_1 = 2.11$, 0.067 $\beta_2 = -1.908$, 0.072	Passed $P = 0.271$	Passed $P = 0.067$	
4*a*	$0.368*\text{DBT}_1 + 0.209$	0.437	2.445	$\beta_0 = 0.209$, 1.193 $\beta_1 = 0.368$, 0.059	Passed $P = 0.131$	Failed $P \leq 0.001$	K.S. dist. = 0.100 $P > 0.200$ Passed
5	$2.875*\text{DBT}_1 - 2.532*\text{WBT}_1 - 0.311*\text{DBT}_2 + 3.474$	0.966	0.610	$\beta_0 = 3.474$, 2.079 $\beta_1 = 2.875$, 0.353 $\beta_2 = -2.532$, 0.292 $\beta_3 = -0.311$, 0.141	Passed $P = 0.081$	Passed $P = 0.403$	
5a	$-0.181*\text{DBT}_1 + 0.879*\text{DBT}_2 - 13.770$	0.913	0.968	$\beta_0 = -13.77$, 0.979 $\beta_1 = -0.181$, 0.040 $\beta_2 = 0.879$, 0.053	Passed $P = 0.801$	Passed $P = 0.133$	K.S. dist. = 0.086 $P > 0.200$ Passed
6	$2.710*\text{DBT}_1 - 2.397*\text{WBT}_1 - 0.553*\text{DBT}_2 + 0.306*\text{MaxT} + 2.260$	0.967	0.607	$\beta_0 = 2.260$, 2.301 $\beta_1 = 2.710$, 0.377 $\beta_2 = -2.397$, 0.311 $\beta_3 = -0.553$, 0.245 $\beta_4 = 0.306$, 0.254	Passed $P = 0.290$	Passed $P = 0.216$	
7	$0.495*\text{Mean ST}_5 - 5.538$	0.693	1.805	$\beta_0 = -5.538$, 1.237 $\beta_1 = 0.495$, 0.046	Failed $P = 0.037$	Failed $P \leq 0.001$	K.S. dist. = 0.173 $P < 0.001$ Failed
8	$0.501*\text{Mean ST}_5 - 0.130*\text{Mean RH} + 2.697$	0.939	0.814	$\beta_0 = 2.679$, 0.809 $\beta_1 = 0.501$, 0.021 $\beta_2 = -0.130$, 0.009	Passed $P = 0.206$	Failed $P = 0.032$	K.S. dist. = 0.072 $P > 0.200$ Passed
9	$0.399*\text{Mean T} + 0.145*\text{Mean ST}_5 - 0.142*\text{Mean RH} + 3.547$	0.943	0.795	$\beta_0 = 3.547$, 0.919 $\beta_1 = 0.399$, 0.220 $\beta_2 = 0.145$, 0.197 $\beta_3 = -0.142$, 0.011	Passed $P = 0.167$	Passed $P = 0.094$	
10	$0.561*\text{Mean T} - 0.147*\text{Mean RH} + 3.930$	0.942	0.791	$\beta_0 = 3.930$, 0.754 $\beta_1 = 0.561$, 0.022 $\beta_2 = -0.147$, 0.009	Passed $P = 0.236$	Passed $P = 0.089$	K.S. dist. = 0.107 $P = 0.141$ Passed
11	$0.431*\text{Mean T} - 0.152*\text{Mean RH} + 0.507*\text{WS} + 4.886$	0.956	0.699	$\beta_0 = 4.886$, 0.711 $\beta_1 = 0.431$, 0.039 $\beta_2 = -0.152$, 0.008 $\beta_3 = 0.507$, 0.132	Passed $P = 0.086$	Passed $P = 0.015$	K.S. dist. = 0.084 $P > 0.200$ Passed

(0.407). Higher wind velocity indicates faster circulation of cold wind from higher atmospheric levels to near surface. Wind parameter helps in transportation of moving moisture laden winds to upper atmospheric levels, which enhances the evaporation.

Soil temperature recorded in the morning at 7am (ST1, ST2, ST3) and afternoon at 2pm (ST4, ST5, ST6) at different depths of 5cm, 15cm and 30cm from ground surface helps in understanding the subsurface temperature gradient, which also controls the rate of evaporation. Among the six

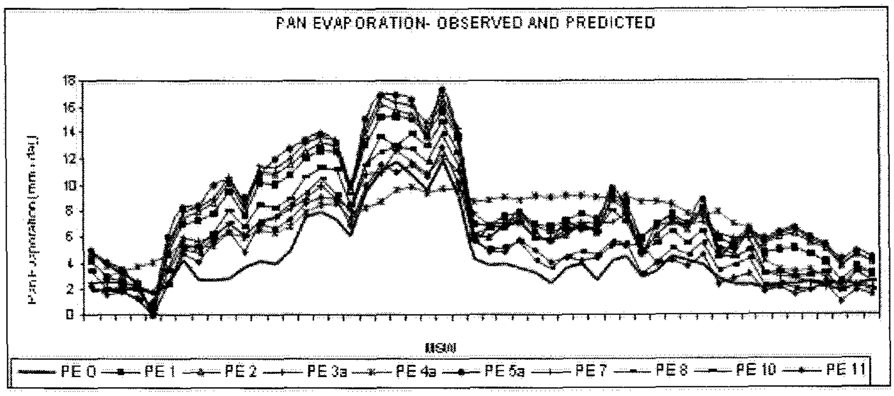

Figure 2.2. Plot showing relationship between pan evaporation observed and predicted models

observations, soil temperature recorded at 5cm depth at 2pm (ST4) exhibit highest correlation coefficient (0.935) with the pan evaporation (Figure 2.1h). This indicates that upper level soil temperature recorded in the afternoon exert the highest control on evaporation due to higher temperature in the upper soil zone, which permit higher proportion of water movement from deeper soil zone towards the surface, and its subsequent evaporation to atmosphere.

Mean soil temperature (Mean ST5), which is the mean of ST1 and ST4 shows significant correlation ($R = 0.832$) with pan evaporation (Figure 2.1f).

Relative humidity has shown low correlation [($R = 0.478$), Figure 2.1d] whereas bright sunshine hours and rainfall were found least correlated with pan evaporation. It has been observed that there is no regular trend of relative humidity.

Statistical analysis

The regression models for predicting pan evaporation using dry bulb temperatures (DBT1, DBT2); wet bulb temperatures(WBT1, WBT2); maximum air temperature(MaxT); minimum air temperature(MinT); mean air temperature(Mean T); soil temperatures (ST1 and ST4); wind speed(WS); mean relative humidity (Mean RH), mean soil temperature [Mean ST5 = (ST1 + ST4)/2] were developed.

The regression models in Table 2.1 have been developed using one response variable and one or more predictors, but in some of the models *i.e.* (Equation 3, 4, 5, 6 & 9) multi-collinearity was observed. In Equation 3, 4, 5 removing one predictor have solved the problem of multi-collinearity. So the new models that were derived after removing multicollinearity were represented by Equation 3a, 4a, 5a and their regression models are tabulated above. However this problem could not be solved in Equation 6 and 9. The residual plots for the models, which do not exhibit the problem of multi-collinearity in Table 2.1, are shown later in Figure 2.3.

The predicted equations derived from regression models were tested for accuracy using the year 2005 meteorological data assorted on weekly basis. The plot showing relationship between pan evaporation observed for the year 2005 and the various predicted models of Table 2.1 are shown in Figure 2.2, where PE-O represents the observed values of pan evaporation for the year 2005 and PE1, PE2, PE11 are the predicted values of pan evaporation calculated from the models 1,211 respectively.

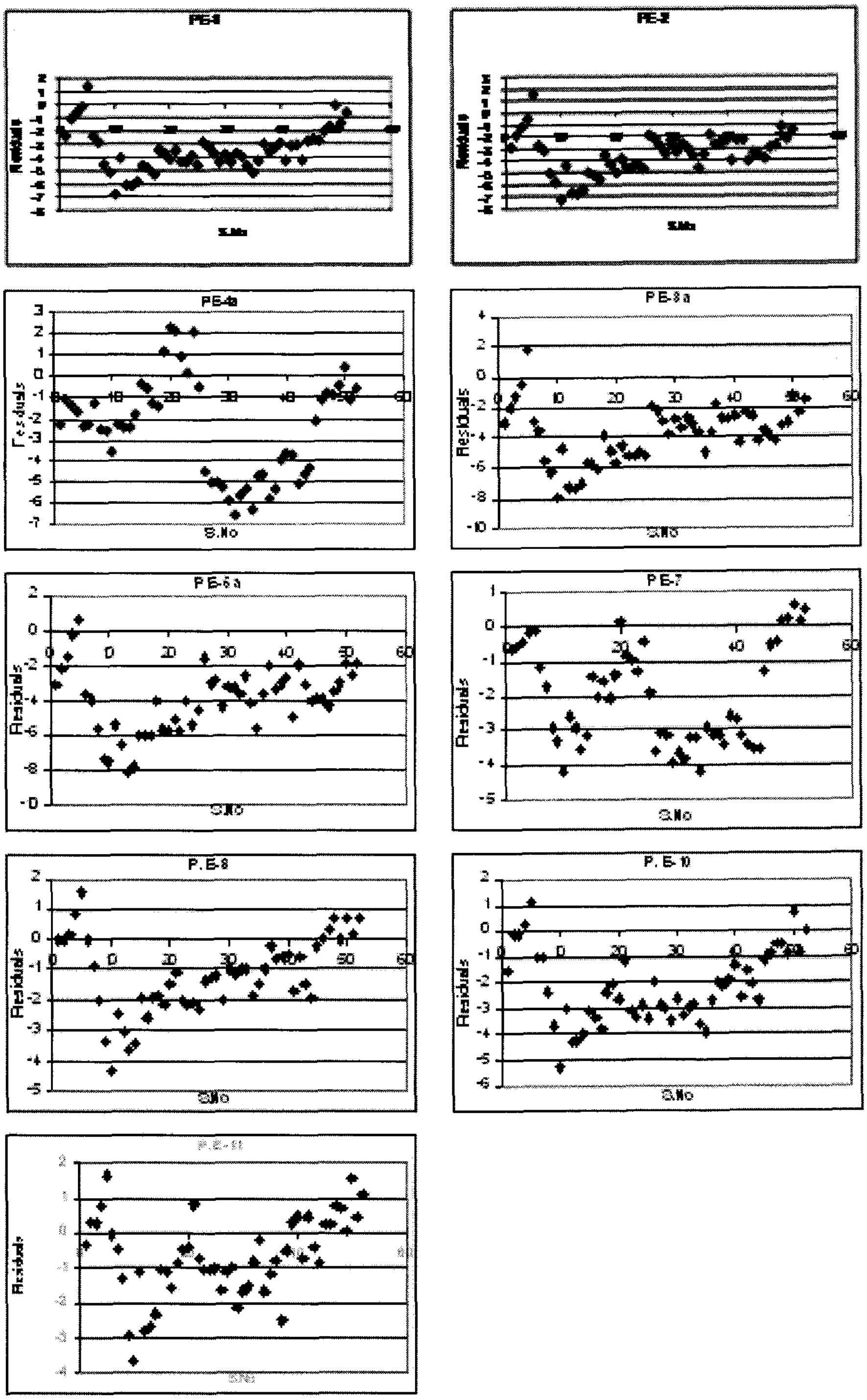

Figure 2.3. The residual plots of various regression models (Table 2.1) in time order sequence

ANOVA table for various models (1–11) represented in Table 2.1 at 5% level of significance are given below:

1. PE = 0.683*MaxT − 12.498

Analysis of Variance:

SV	DF	SS	F	Table F	Remark
Regression	1	471.538	399.276	4.03	significant
Residual	50	59.049			
Total	51	530.587			

2. PE = −0.165*WBT1 + 0.818*MaxT − 13.612

SV	DF	SS	F	Table F	Remark
Regression	2	494.038	331.172	3.18	significant
Residual	49	36.549			
Total	51	530.587			

3. PE = 2.516*DBT1 − 2.239*WBT1 − 0.163*MaxT + 1.474

SV	DF	SS	F	Table F	Remark
Regression	3	511.380	426.009	2.79	significant
Residual	48	19.206			
Total	51	530.587			

3a. PE = −0.187*DBT1 + 0.882*MaxT − 14.659

SV	DF	SS	F	Table F	Remark
Regression	2	491.271	306.139	3.18	significant
Residual	49	39.316			
Total	51	530.587			

4. PE = 2.110*DBT1 − 1.908*WBT1 − 1.057

SV	DF	SS	F	Table F	Remark
Regression	2	510.915	636.331	3.18	significant
Residual	49	19.671			
Total	51	530.587			

4a. PE = 0.368*DBT1 + 0.209

SV	DF	SS	F	Table F	Remark
Regression	1	231.656	38.747	4.03	significant
Residual	50	298.931			
Total	51	530.587			

5. PE = 2.875*DBT1 − 2.532*WBT1 − 0.311*DBT2 + 3.474

SV	DF	SS	F	Table F	Remark
Regression	3	512.722	459.211	2.79	significant
Residual	48	17.864			
Total	51	530.587			

5a. PE = −0.181∗DBT1 + 0.879∗DBT2 − 13.770

SV	DF	SS	F	Table F	Remark
Regression	2	484.665	258.578	3.18	significant
Residual	49	45.922			
Total	51	530.587			

6. PE = 2.710∗DBT1 − 2.397∗WBT1 − 0.553∗DBT2 + 0.306∗MaxT + 2.260

SV	DF	SS	F	Table F	Remark
Regression	4	513.258	348.021	2.56	significant
Residual	47	17.329			
Total	51	530.587			

7. PE = 0.495∗Mean ST5 − 5.538

SV	DF	SS	F	Table F	Remark
Regression	1	367.647	112.817	4.03	significant
Residual	50	162.940			
Total	51	530.587			

8. PE = 0.501∗Mean ST5 − 0.130∗Mean RH + 2.697

SV	DF	SS	F	Table F	Remark
Regression	2	498.159	376.372	3.18	significant
Residual	49	32.428			
Total	51	530.587			

9. PE = 0.399∗Mean T + 0.145∗Mean ST5 − 0.142∗Mean RH + 3.547

SV	DF	SS	F	Table F	Remark
Regression	3	500.246	263.803	2.79	significant
Residual	48	30.341			
Total	51	530.587			

10. PE = 0.561∗Mean T − 0.147∗Mean RH + 3.930

SV	DF	SS	F	Table F	Remark
Regression	2	499.902	399.149	3.18	significant
Residual	49	30.684			
Total	51	530.587			

11. PE = 0.561∗Mean T − 0.147∗Mean RH + 3.930

SV	DF	SS	F	Table F	Remark
Regression	3	507.140	346.075	2.79	significant
Residual	48	23.446			
Total	51	530.587			

Table 2.2. Regression models for standardized pan evaporation and predictors

Eq. No.	Prediction Equation	MSE	β estimates and their S.E. (respectively)	NT Normality test (overall)	CV Constant variance test
S1	0.943*MaxT − 0.000143	0.337	$\beta_0 = -0.000143$, 0.0467 $\beta_1 = 0.943$, 0.0472	Failed $P = 0.010$	Passed $P = 0.563$
S2	−0.278*WBT$_1$ + 1.129*MaxT − 0.000182	0.268	$\beta_0 = -0.000181$, 0.0371 $\beta_1 = -0.278$, 0.0506 $\beta_2 = 1.129$, 0.0506	Passed $P = 0.583$	Passed $P = 0.102$
S3a	−0.336*DBT$_1$ + 1.217*MaxT − 0.000162	0.278	$\beta_0 = -0.000162$, 0.0385 $\beta_1 = -0.336$, 0.0677 $\beta_2 = 1.217$, 0.0677	Passed $P = 0.745$	Passed $P = 0.114$
S5a	−0.326*DBT$_1$ + 1.204*DBT$_2$ + 0.0000522	0.300	$\beta_0 = 0.0000522$, 0.0416 $\beta_1 = -0.326$, 0.0733 $\beta_2 = 1.204$, 0.0733	Passed $P = 0.801$	Failed $P = 0.008$
S7	0.833*Mean ST$_5$ − 0.0000629	0.560	$\beta_0 = -0.0000629$, 0.0776 $\beta_1 = 0.833$, 0.0784	Failed $P = 0.037$	Passed $P = 0.155$
S8	0.843*Mean ST$_5$ − 0.496*MeanRH − 0.0000671	0.252	$\beta_0 = -0.0000671$, 0.0350 $\beta_1 = 0.843$, 0.0353 $\beta_2 = -0.496$, 0.0353	Passed $P = 0.206$	Passed $P = 0.236$
S10	0.850*Mean T − 0.564*Mean ST − 0.0000917	0.245	$\beta_0 = -0.0000917$, 0.0340 $\beta_1 = 0.850$, 0.0345 $\beta_2 = -0.564$, 0.0345	Passed $P = 0.236$	Failed $P = 0.027$

Residual analysis

The residuals are defined as $e_i = Y_i - \hat{Y}_i$, $i = 1, 2, 3, \ldots, n$ where Y_i is an observed value and $\hat{Y}_i$ is the corresponding fitted / predicted value obtained by use of the regression model. Thus residuals e_i are the differences between what is actually observed, and what is predicted by the regression equation. The residual of each regression model could be examined to see if they provide any indication that the model is adequate or not (Draper and Smith, 1981). The residuals have been plotted in the time sequence plot (Figure 2.3).

The i^{th} standardized residual eis is defined as $e_{is} = e_i/s$, where s is the standard deviation of residuals. The standardized residuals e_{is} have zero mean and unit standard deviation. The residuals should be distributed approximately as independent, normal deviates for a large sample (Chatterjee and Price, 1977). The plots of residuals in the time order exhibit that most of the equations were showing large number of negative residuals points in the scatter plots. This is because mean of the error is not zero and the variance is too large. However plots 4a and 11 are showing both positive and negative residuals.

The equations, which are showing large number of negative residuals points in the scatter plots, were taken up for the standardization of response variable and predictors.

Normality test of the residuals were carried out by the method of Kolmogorov-Smirnov test. This test was done only for those models, which does not show multi-collinearity or those, which were derived after removing multi-collinearity. A test that passes the normality test indicates that the data matches the pattern expected if the data was drawn from a population with a normal distribution.

Standardization of the response variable and the predictors (Garrett, 1947)

We have opted for standardization the Equations (1, 2, 3a, 5a, 7, 8, 10) to ensure both negative and positive residuals in the scatter plot.

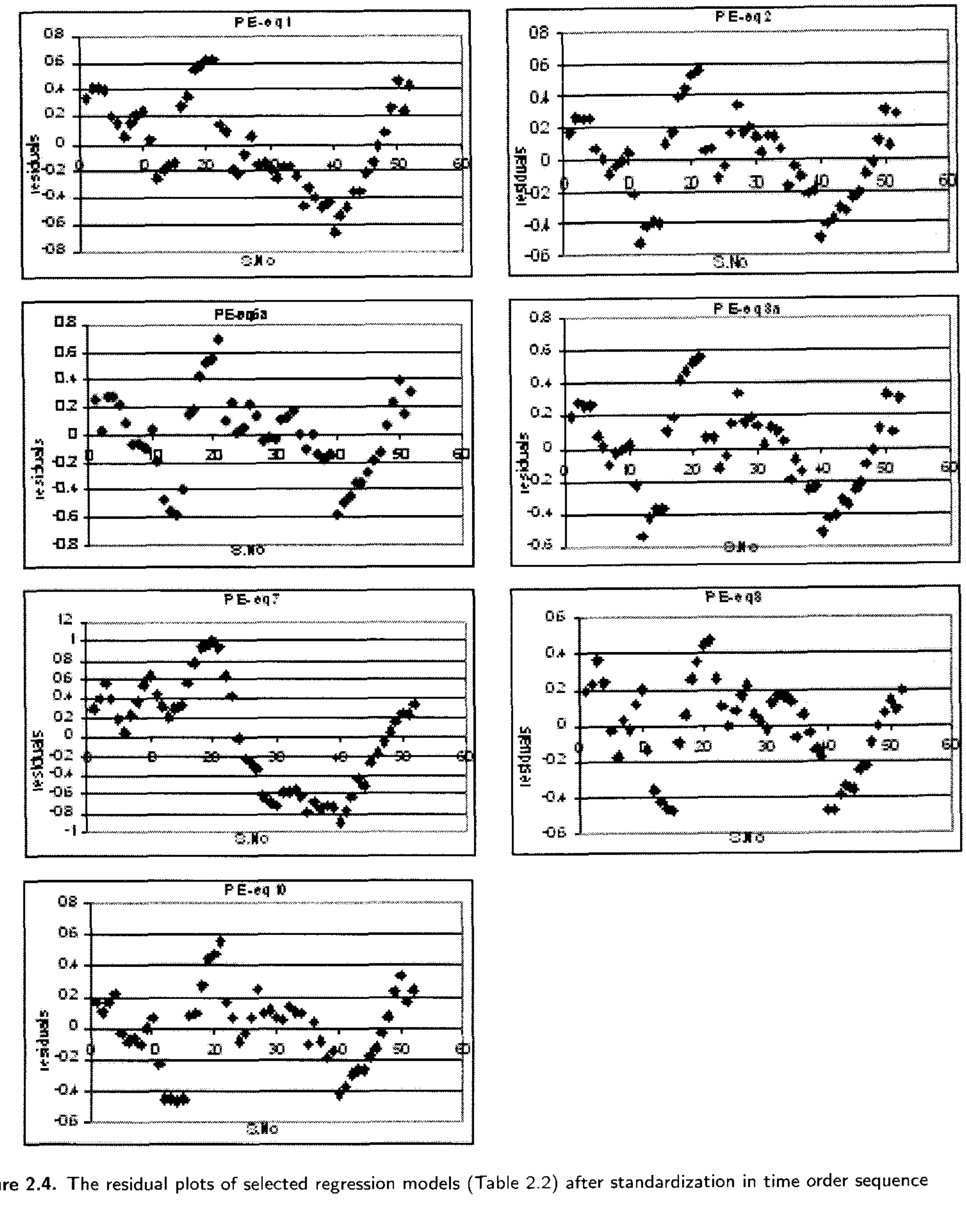

Figure 2.4. The residual plots of selected regression models (Table 2.2) after standardization in time order sequence

In performing the regression analysis various assumptions about the errors need to be considered. The errors are independent, have zero means, a constant variance, and they follow a normal distribution. (Draper and Smith, 1981). Therefore in the present study, if the fitted/standardized model is correct then the residuals should fulfill these assumptions. The new Equations (1, 2, 3a, 5a, 7, 8, 10) after standardization are expressed as (S1, S2, S3a, S5a, S7, S8, S10) and the regression models are expressed in Table 2.2. The residual plots of the said regression models (Table 2.2) after standardization in time order sequence are shown later in Figure 2.4.

ANOVA table for the various standardized models represented in Table 2 at 5% level of significance are given below:

S1. PE = 0.943*MaxT − 0.000143

SV	DF	SS	F	Table F	Remark
Regression	1	45.337	399.276	4.03	significant
Residual	50	5.677			
Total	51	51.015			

S2. PE = −0.278*WBT1 + 1.129*MaxT − 0.000182

SV	DF	SS	F	Table F	Remark
Regression	2	47.501	331.172	3.18	significant
Residual	49	3.514			
Total	51	51.015			

S3a. PE = −0.336*DBT1 + 1.217*MaxT − 0.000162

SV	DF	SS	F	Table F	Remark
Regression	2	47.235	306.139	3.18	significant
Residual	49	3.780			
Total	51	51.015			

S5a. PE = −0.326*DBT1 + 1.204*DBT2 + 0.0000522

SV	DF	SS	F	Table F	Remark
Regression	2	46.600	258.578	3.18	significant
Residual	49	4.415			
Total	51	51.015			

S7. PE = 0.833*Mean ST5 − 0.0000629

SV	DF	SS	F	Table F	Remark
Regression	1	35.349	12.817	4.03	significant
Residual	50	15.666			
Total	51	51.015			

S8. PE = 0.843*Mean ST5 − 0.496*Mean RH − 0.0000671

SV	DF	SS	F	Table F	Remark
Regression	2	47.897	376.372	3.18	significant
Residual	49	3.118			
Total	51	51.015			

S10. PE = 0.850*Mean T − 0.564*Mean RH − 0.0000917

SV	DF	SS	F	Table F	Remark
Regression	2	48.065	399.149	3.18	significant
Residual	49	2.950			
Total	51	51.015			

3. Conclusion and Future Work

Comparative evaluation of regression models

A comparative evaluation of regression models (*i.e.* pre and post standardization) as presented in Table 2.1 and 2.2 revealed the following characteristics about the pre and post standardized data sets:

(i) The normality test results were identical for all the pre and post standardization models as it passed for all the models (except for 1, 7).

(ii) The constant variance test passed for all models (except for 4a, 7, 8 in Table 2.1), (except for 5a, 10 in Table 2.2).

(iii) The mean squares of residuals in the standardized models were also found to be reduced from that before standardization.

(iv) The β estimates and their standard error(S.E) of the standardized models were also reduced from pre-standardized models.

Transformation of the variables

In regression analysis a convenient starting point is that the model describing the data is linear in the variables. The necessity for transforming the data arises because the original variable, or the model in terms of the original variable, violates one or more of the standard assumptions. The most commonly violated assumptions are those concerning the linearity of the model and the constancy of the error variance. When the error variance is not constant over all the observations, the error is said to be heteroscedastic. Heteroscedasticity can be removed by means of a suitable transformation. The transformation is not only to stabilize the variance, but also have the effect of making the distribution of the transformed variable closer to the normal distribution (Chatterjee and Price, 1977).

It is noted that models in Table 2.1 which failed the test *i.e.*, models 1, 7 in case of Normality test and models 4a, 7, 8 in case of Constant variance test can be transformed further using suitable transformations in order to pass the test. Therefore in the present investigation an attempt has been made to develop transformed model for select failed models.

(a) The model 1 – $\mathrm{PE} = 0.683{*}\mathrm{MaxT} - 12.498$ was transformed into

$$\mathrm{PE} = -31.890 + 7.299{*}\sqrt{\mathrm{MaxT}}\ (R = 0.931,\ R^2 = 0.867,\ \mathrm{S.E} = 1.188)$$

Normality test = Passed ($P = 0.086$), Constant variance test = Passed

(b) The model 7 – $\mathrm{PE} = 0.495{*}\mathrm{Mean\ ST5} - 5.538$ was transformed into

$$1/\mathrm{PE} = 0.631 - 0.0931{*}\sqrt{\mathrm{Mean\ ST5}}\ (R = 0.897,\ R^2 = 0.805,\ \mathrm{S.E} = 0.025)$$

Normality test = Passed ($P = 0.0655$), Constant variance test = Passed ($P = 0.956$)

It is to remark that the maximum air temperature primarily influence pan evaporation besides mean air temperature and wind velocity which also have significant influence on pan evaporation. In contrast minimum air temperature, relative humidity and wind direction has non-significant influence. Interestingly rainfall and bright sunshine hours exhibit least correlation with pan evaporation.

The regression models developed in table 1using one response variable and one or more predictors have problems of multicollinearity in few models, which was solved by removing one predictor. In the case of models 6 and 9, we are looking for suitable functional relationships of the predictors to be placed as coefficients (z_i's) in the model so that problem may be solved.

To examine whether the model is adequate or not, the residuals of each regression model were plotted in the time sequence plot. The plots of residuals in the time order exhibit that most of the equations were showing negative residuals points in the scatter plots. Therefore these equations were taken up for the standardization of response variable and predictors. It was found that in the standardized models the mean squares of residuals were reduced from those before standardization as well as the estimates and their standard errors(S.E) were also reduced from pre- standardized models.

Further investigations are ongoing for some other suitable transformation to stabilize the error variance for those models, which failed the constant variance test after trying some selected transformations proposed by Draper and Smith (1981). Use of weighted least squares, forward inclusion and backward elimination selection techniques are also under consideration. Although sensitivity analysis (detection and removal of outliers or influential observations) in linear regression is reserved for the future, at a cursory glance the authors felt that the present data sets have not been influenced by extremely large or extremely small observations.

As a final comment, although multiple regression analysis which we have used is a powerful statistical tool, it may not be a bad idea to compare the present results with those obtained by alternative methods such as principal components, factor analysis and partial least squares based on latent variables.

Acknowledgements

Authors are thankful to Prof. A Wadood, Birsa Agricultural University Ranchi, India for helping in providing meteorological data and useful scientific discussions during preparation of the manuscript. We also thank an anonymous referee for several suggestions for improving the manuscript.

References

Ali, M. S., Sattar, A. and Chaturvedi, O. P. (2006). Correlation of meteorological parameters with the incidence and activity of shoot and trunk borer in shisham stands, *Journal of Agrometeorology*, **8**, 296–299.

Allen, R. G., Pereira, L. S., Raes, D. and Smith, M. (1998). *Crop Evapotranspiration-Guidelines for computing crop water requirements,* Food and Agriculture Organization, Rome.

Barnett, N., Madramootoo, C. A. and Perrone, J. (1998). Performance of some evapotranspiration equations at a site in Quebec, *Canadian Agricultural Engineering*, **40**, 89–95.

Chatterjee, S. and Price, B. (1977). *In Regression Analysis by Example,* Wiley, New York.

Chattopadhyay, N. and Hulme, M. (1997). Evaporation and potential evapotranspiration in India under conditions of recent and future climate change, *Agricultural and Forest Meteorology*, **87**, 55–73.

Dingman, S. L. (1994). *Upper Saddle River,* Prentice Hall, New Jersey.

Draper, N. R. and Smith, H. (1981). *Applied Regression Analysis,* Wiley, New York, 86–99.

Garrat, H. E. (1947). *In: Statistics in Psychology and Education*, Paragon International Publishers 1982, 492pp.

Geiger, R., Aron, R. H. and Todhunter, P. (2003). *The Climate Near the Ground*, Rowman & Littlefield, New York.

Golubev, V. S., Larimore, J. H., Groisman, P. Y., Speranskaya, N. A., Zhuravin, S. A., Menne, M. J., Peterson, T. C. and Malone, R. W. (2001). Evaporation changes over the contiguous United States and the former USSR: A reassessment, *Geophysical Research Letters*, **28**, 2665–2668.

Hansen, V. E., Israelsen, O. W. and Stringham, G. E. (1980). *Irrigation: Principles and Practices,* 4th Edition, John Wiley & Sons, New York.

Hordofa T., Sharma A., Singh, R. and Dashora, P. K. (2004). Influence of meteorological variables on monthly pan evaporation under sub-humid climatic conditions of Ethiopia, *Indian Journal of Soil Conservation*, **32**, 1–4.

Khan, M. A. (1992). Evaporation of water from free water surface - Climatic influence, *Indian Journal of Soil Conservation*, **20**, 22–27.

Kumar, M., Pandey, S. and Mahanti, N. C. (2007). Water resource potentials of Jharkhand State, India, In *Proceeding International Conference on Water and Flood Management(ICWFM-2007)*, Dhaka, Bangladesh, 525–530.

Liu, B., Xu M., Henderson, M., and Gong, W. (2004). A spatial analysis of pan evaporation trends in China, 1955–2000, *Journal of Geophysical Research*, **109**, D15102 doi: 10.1029/2004JD004511.

Paterson, T. C., Golubev, V. S. and Groisman, P. Y. (1995). Evaporation losing its strength, *Nature*, **377**, 687–688.

Rosenberg, N. J., McKenney, M. S. and Martin, P. (1989). Evapotranspiration in a greenhouse-warmed world: A review and a simulation, *Agricultural and Forest Meteorology*, **47**, 303–320.

Sahu, D. D., Pandya, H. R. and Dixit, S. K. (1994). Annual and seasonal affects of meteorological and associated parameters on pan evaporation, *Gujarat Agricultural University Research Journal*, **19**, 82–86.

Singh, R. V., Chauhan, H. S. and Ali, A. B. (1981). Pan evaporation as related to meteorological parameters, *J. Agri. Engg.* **18**, 48–53.

Watson, I. and Burnett, A. D. (1995). *Hydrology: An Environmental Approach,* CRC Press, Florida.