e-ISSN 2320 –7876 www.ijfans.org
Vol.11,S Iss 1, 2022
Research paper UGC CARE Listed (Group -I) Journal

Estimation of Pan Evaporation Under Limiting Data Condition: Jorhat (Assam) Northeast India

Pankaj Kumar Pandey[1], Vanita Pandey[2] Madhusudhan Mishra[3]

[1,2]Department of Agricultural Engineering, North Eastern Regional Institute of Science & Technology, Nirjuli, Arunachal Pradesh

[3]Department of Electronics and Communication Engineering, North Eastern Regional Institute of Science & Technology, Nirjuli, Arunachal Pradesh

*Corresponding Author: pkpnerist@gmail.com (P.K. Pandey)

Abstract

Evaporation is a crucial factor in numerous hydrological and water resource systems initiatives. This research seeks empirical relations for computing monthly pan evaporation based on climatic temperature, relative humidity, and wind speed data. This study used 11 years of meteorological data (e.g., from January 2004 to December 2014). A parameter selection approach is utilized with multiple linear regression to determine the optimal model forms. Using the power and exponential functions, the correlations of evaporation with temperature and relative humidity different trendlines were fitted to identify the hidden pattern in meteorological data. The suggested evaporation estimating models with the optimal climatic parameters combinations were demonstrated to yield results that fit well with the observed data.

Keywords: Pan evaporation, multiple linear regression, modelling

Introduction

Accurate evaporation estimation is critical for monitoring, surveying, and managing water resources, as well as constructing irrigation and drainage systems and scheduling. Some commonly used methods for estimating evaporation include mass transfer and water balance (Gundalia and Dholakia,2013). The most precise direct method of measurement of evaporation is a Class-A pan evaporimeter. For measurement of evaporation using the direct method requires the installation of an evaporimeter at the location, and that is not always practical to install an evaporimeter at every location where a reservoir and irrigation project is proposed or already in existence. Furthermore, constructing or maintaining costly meteorological sensors is not always viable in remote locations. Pan performance is influenced by instrumental restrictions and operational issues such as human errors, instrumentation errors, water turbidity, and other maintenance issues that might influence evaporation measurement accuracy. Hydrologists and meteorologists are particularly interested in estimating evaporation for isolated rural locations. Literature suggests numerous attempts have been made to accurately quantify pan evaporation (Ep), pan coefficient, and their agreement with meteorological parameters under different climates. Evaporation is a highly nonlinear process. To understand nonlinearity better, Ep estimates could be investigated using models that handle the process's inherent nonlinearity.

Many academics have also utilized models based on meteorological data to estimate evaporation. The Penman equation (1956) is a standard method for evaporation estimation. The Penman (1956) model was developed based on an experiment in lakes and developed boundary limits for open water evaporation. To successfully apply the Penman model, the function of environmental, meteorological conditions and heat transfer within the water body necessitates temperature profile observations. As a result, if one or more of the Penman equation's parameters are unavailable from meteorological weather station readings, the Penman equation (1956) cannot be used (Kumar et al. 2013). As a result, several simplified analytical and empirical equations have been devised. For example, Fennessey and Vogel (1996) used regression methods to build models for regional monthly average evaporation in the United States as a function of readily available factors such as temperature and the longitude and elevation of the site. Linacre (1973), Hargreaves and Samani (1982), and Tabari et al. (2009) assessed evaporation in the semi-arid region of Iran using both artificial neural network approaches and multivariate nonlinear regression. Temperature, relative humidity, wind speed, solar radiation, and precipitation are the meteorological variables used in the procedures (Bruton, 2002).

In estimating pan evaporation using statistical procedures (such as multiple linear regression and ANOVA), a well-established relationship between evaporation and other climatological factors is produced (Piri et al. 2010). Multiple linear regression is an essential method for determining the dependent parameter using known input parameters, often known as predictors (Kisi 2013). It is commonly implemented for forecasting and predicting. ANOVA is a statistical test that compares the means of many groups and extends the t-test to include more than two groups. These are useful for comparing groups or parameters' statistical significance (testing). This research aims to develop empirical relationships suitable for calculating daily and monthly evaporation in the northeastern region of India using readily available meteorological data.

2.0 Dataset and Methodology

2.1 Geographical Detail of Study area

Daily mean meteorological parameters (2004 to 2014) were collected from Tea Research Association, Jorhat, Assam (Longitude 94^0 12' E, and Latitude 26^0 47' N). The study area is under a humid climate.

2.2 MLR (Multiple Linear Regression):

The quantitative relationship between a dependent parameter and different correlated parameters is often described using regression analysis (Tabari et al., 2009). The functions in multiple linear regression are restricted to a linear equation, i.e., a straight line in the form

Y=$\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_{k,i}$+$e_i$ Eq. (1)

Where, Y = dependent variable, $\beta_0 - \beta_k$= equation parameters for the linear relation, X_1, X_k= independent variables for this system and e_i = random error

The least squares criterion with the minimal sum of squares of error terms (S) is the most used approach for estimating the values of _0, _1, and _k. $\beta_1, \ldots$ and β_k to minimize.

$$S=\sum_{i=1}^{n}\left(yi_{observed} - \beta_0 - \beta_1 X_{1,i} - \cdots - \beta_k X_{k,i}\right)^2 \quad \text{Eq. (2)}$$

$$=\sum_{i=1}^{n}(yi_{observed} - yi_{calculated})^2$$

$$=\sum_{i=1}^{n} e_i^2$$

As a result, $\beta_0, \beta_1, \ldots$ and β_k must satisfy

$$\frac{\partial S}{\partial b_j} = 2\sum_{i=1}^{n} e_i \frac{\partial e_i}{\partial b_j} = 0,$$

Eq.(3)

And since $e_i = y_{i\,observed} - y_{i\,calculated}$, the above equation becomes

$$\frac{\partial S}{\partial b_j} = -2\sum_{i=1}^{n} e_i \frac{\partial e_{i\,calculated}}{\partial b_j} = 0, \quad \text{Eq. (4)}$$

j=0, 1… k

Initially, a general multiple regression model expressing evaporation can be assumed regarding those climatological parameters.

$$E_i = \beta_1 T_i + \beta_2 H_i + \beta_3 u_i \quad \text{Eq. (5)}$$

Where, *E* is pan evaporation (mm day^{-1}); *T* is the temperature (◦C); *u* is wind speed (m s^{-1}); *H* is relative humidity (%). And is in the linearized form

$$E_i = \beta_1 T_i^{*} + \beta_2 H_i^{*} + \beta_3 u_i \quad \text{Eq. (6)}$$

.

The correlation coefficients are the strength of an association between two variables, which can be measured by examining the correlation coefficients before and after a particular statistical transformation is carried out on data (Almedeij, 2012).

2.3 Evaluation of Developed Models

The evaluation of the developed models was carried out by using the following statistical indices:

e-ISSN 2320 –7876 www.ijfans.org
Vol.11,S Iss 1, 2022
Research paper UGC CARE Listed (Group -I) Journal

2.3.1 RmSE (Root Mean Square Error):
The RMSE can be written in their usual meaning as reported by statistical literature as

$$\text{RmSE}=\sqrt{\frac{1}{n}\sum_{i=1}^{n}(yi_{calculated}-yi_{observed})^2} \qquad \text{Eq. (7)}$$

2.3.2 MaPE (Mean Absolute Percentage Error):
The MaPE is a commonly used index to know the deviation of fitting from mean values. Mathematically it can be written as:

$$\text{MaPE}=\frac{1}{n}\sum_{i=1}^{n}\left|\frac{yi_{calculated}-yi_{observed}}{yi_{observed}}\times 100\right| \qquad \text{Eq. (8)}$$

2.3.3 NsE (Nash Sutcliffe efficiency):

Nash–Sutcliffe efficiency indicates a degree of agreement between modelled and observed values. It can be written as

$$\text{NsE}=1-\frac{\sum_{i=1}^{n}(yi_{observed}-yi_{calculated})^2}{\sum_{i=1}^{n}(yi_{observed}-\overline{yi}_{observed})^2}$$

Eq.(9)

2.3.4 R^2_a (Adjusted Coefficient of Determination):
It can be given as

$$R^2_a=1-\frac{n-1}{n-p}\times(1-R^2) \qquad \text{Eq.}$$

(10)
Where, n is the sample size; *p* is number of explanatory variables.

3.0 Results and Discussion

3.1Selection and Transformation of Input variables

The applicability of the various collected meteorological variables for fitting this regression model was investigated. Based on analytical and statistical indices, all feasible combinations were discussed to locate suitable input. Tmax, RHmax, and wind speed were chosen as appropriate variables, as indicated in Table 1. Fitting regression revealed that the relationships between Ep and Tavg, Ep and RHmax, and Ep and U are not linear, as illustrated in Figs. 1–3. The selected variables were then translated into a linear form utilizing the created relation below (Table 1). The independent parameters, namely average temperature, relative humidity, and wind speed, can be transformed to linearize these variables.

Table 1: Fitting accuracy measure of different input parameters.

Variables	Ra²	MAPE (%)	RMSE (mm)	NSE
Tmax	0.7892	0.91318	0.183	0.776
Tmin	0.8432	0.7452	0.163	0.822
Tavg	0.827	0.810	0.169	0.809
RHmax	0.573	10.057	0.253	0.573
RHmin	0.380	16.09	0.333	0.258
RH avg	0.152	18.045	0.391	0.151
Rainfall	0.824	8.059	0.162	0.824
Wind Speed	0.760	13.703	0.189	0.760

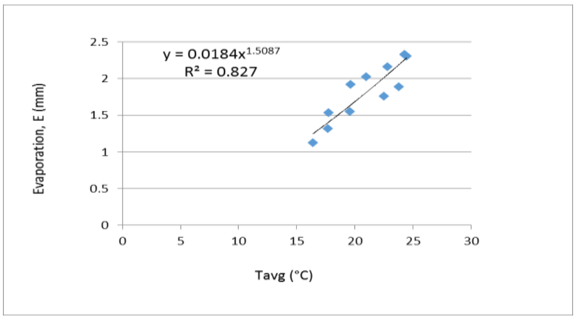
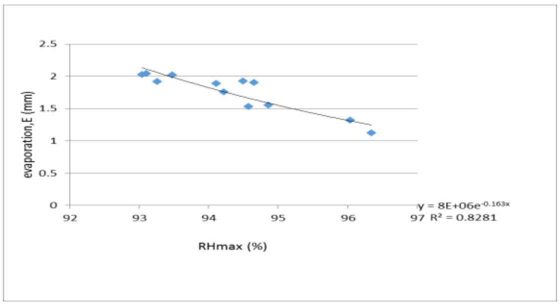


Fig.1 Evaporation Vs Temperature (in power function).



e-ISSN 2320 –7876 www.ijfans.org
Vol.11,S Iss 1, 2022
 UGC CARE Listed (Group -I) Journal

Fig.2 Evaporation Vs RHmax (in exponential function)

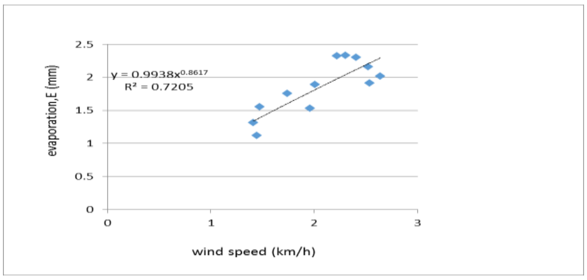


Fig. 3 Evaporation Vs wind speed (in power function)

A linear relationship between the independent variable is frequently used to create multiple linear regression. As a result, all selected values were translated into a linear form. The transformed equations that have been constructed are as follows:

$$E=\beta T^{\alpha} = 0.0184\ T_{avg}^{\ 1.5087} \qquad \text{Eq. (11)}$$

$$E= \beta e^{\alpha H} =29\ e^{-0.032RH_{max}} \qquad \text{Eq. (12)}$$

$$E=0.9938u^{0.8617} \qquad \text{Eq. (13)}$$

Tavg = average temperature, RHmax = maximum relative humidity, and u = wind speed. The transformed relation in Fig.4 to Fg.6 clearly shows that by adopting transformed variables, the relationship between pan evaporation and Tavg, RH, and U may have represented a linear form.

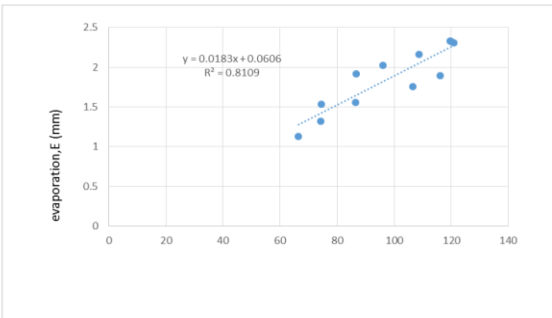


Fig: 4 Evaporation Vs Temperature (average) in a linear function.

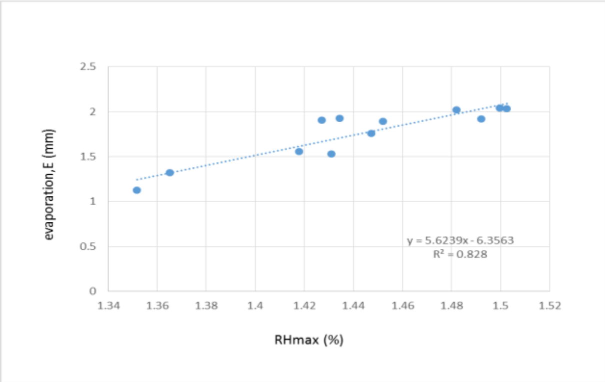


Fig. 5 Evaporation Vs RHmax (in linear form)

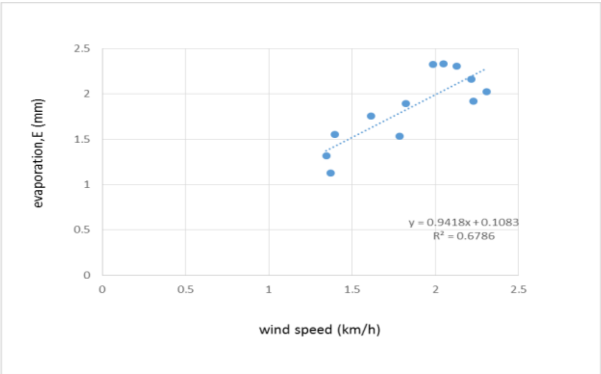


Fig.6 Evaporation Vs wind speed (in linear form)

3.2. Development of one variable model:

The dependent value (evaporation) and a single variable value are required to create a model with a single parameter. Here, the variables considered are average temperature, maximum RH, and U. As illustrated in Table 3, the evaporation value is compared with each specified variable to provide the values Ra^2, MaPE, RmSE, and NsE. For both Tables 4 and 5, comparisons between evaporation and RHmax were calculated using regression analysis (ANOVA). RHmax is the best variable, as depicted in Table 2. Analysis of Table 2 displays the fitting precision of regression models for a single parameter using monthly average temperature data for Jorhat, Assam, from 2004 to 2014. In addition, Table 3 shows the ANOVA results for the identified models of a single parameter using monthly average climatic data for Jorhat, Assam.

$$E=\beta T^{\alpha} = 0.0184\, T_{avg}^{1.5087} \quad \text{Eq. (14)}$$

$$E= \beta e^{\alpha H} =29\, e^{-0.032RHmax} \quad \text{Eq. (15)}$$

$$E=0.9938u^{0.8617} \quad \text{Eq. (16)}$$

Table 2 Fitting accuracy measure of single input parameter developed a model for Jorhat, Assam.

Input parameter	Fitting Statistics			
	Ra^2	MaPE (%)	RmSE (mm)	NsE
Single input parameter model				
Tavg	0.8109	8.100	0.0488	0.8095
RHmax	0.828	7.7930	0.0378	0.8117
U	0.6786	13.7025	0.2208	0.6760

Table 3 ANOVA analysis of single input parameter developed model for Jorhat, Assam,

Developed Eq.		DF	ss	ms	F	F Significance
	Regression	1	1	0.8	58.5	true
1	Residual	10	0	0		
	Total	11	1			

3.3 Development of two variable models:

Average temperature and wind speed are the best input parameters for Model 2. The accuracy result of model 2 based on Ra^2, MaPE, RmSE, and NsE equals 0.864, 11.667,0.2322,0.8575 (for temperature and relative humidity) 0.947, 3.57, 0.28, 0.9553 (for temperature and relative humidity) (for temperature and wind speed). Table 4 depicts the relative humidity and wind speed-based model performance indices. The analysis results confirm that the developed model based on RHmax and wind speed input parameters may be ranked the poorest. The ANOVA analysis and regression results of two variable models are shown in Tables 6 and 7. The F-test value indicates a statistically significant linear trend between evaporation and modified input parameter (Almedeij 2012). Table 5 shows the fitting accuracy of regression models of two input variables using monthly average meteorological data for Jorhat, Assam.

Model 2, $Ep = 0.09T_{avg} + 0.41u - 0.88$ Eq. (17)

$Ep = 0.10T_{avg} - 0.13\, RHmax + 11.81$ Eq. (18)

Ep = 0.55 u- 0.09 RHmax +9.6 Eq. (19)

Table 4: Fitting accuracy measure of double input parameter developed model for Jorhat, Assam

Parameters	Fitting Statistics			
	R_a^2	MaPE(%)	RmSE(mm)	NsE
Double input parameters model				
T_{avg}, RH_{max}	0.864	11.669	0.232	0.857
T_{avg}, U	0.947	3.560	0.283	0.955
RH_{max}, U	0.625	13.702	0.445	0.619

Table 5: ANOVA analysis of single input parameter developed model for Jorhat, Assam

Develope d Eq.		**DF**	**ss**	**ms**	**F**	**F Significance**
	Regression	2	2	0.9	98.	true
2	Residual	9	0	0		
	Total	11	2			

3.4 Development of three variable models:

The input of Model 3 is Tavg, RH, and U. From Table 8, the accuracy result of model 3 is based on R_a^2= 0.94, MaPE= 10.21 % RmSE= 0.193 mm and NsE=0.9495. Model three had the best performance indices if wind speed data is available and could be used to estimate pan evaporation with higher accuracy. Table 6 depict the fitting accuracy of regression models of three variable using monthly average meteorological data for Jorhat, Assam, from 2004 to 2014 (calibration range). Table 7 shows that the ANOVA analysis of the F test statistic showed a significant linear trend between the evaporation and selected input parameters.

Table 6: Fitting accuracy measure of three input parameter developed model for Jorhat, Assam

Variable	Fitting Statistics			
	R_a^2	MaPE(%)	RmSE(mm)	NsE
Three variable model				
T_{avg}, RH_{max}, U	0.94	10.2103	0.193	0.9495

Table 7: ANOVA analysis of three input parameters developed model for Jorhat, Assam

Developed Eq.		DF	ss	ms	F	F Significance
	Regression	3	2	0.6	58.5	true
3	Residual	0	0	0		
	Total	11	2			

3.5 Developed Models analysis:

Based on the findings, Model 1 (temperature) is the best one-variable model for assessing statistics. Model 2 (temperature and relative humidity) had the best attributes of all two-variable model combinations based on all evaluation criteria. The third model considers input parameters, such as Tavg, RH, and U. The form of the three following models are as follows:

Model 1, Ep=$29e^{-0.032RH_{max}}$ and Ep = 0.0184 $T_{avg}^{1.5087}$ Eq. (20)

Model 2, Ep= $0.09T_{avg} + 0.41u - 0.88$ and Ep= $0.10T_{avg} - 0.13\ RHmax + 11.81$ Eq.(21)

Model 3, Ep= $0.09T_{avg} + 0.42u + 0.01RH_{max} - 1.68$ Eq. (22)

The results of the RmSE and MaPE tests (Table 2) reveal a difference between Model 2 and Model 1 of 3.57% and 8.1%, respectively. Therefore, it could be concluded that Model 2 is the best-developed Eq. to estimate Ep accurately in the study region.

The results analysis found that developed Eq.1(model1) has the poorest performance among the three identified models (e.g., model1, model2, and model3); nevertheless, it has the advantage of simplicity in that it can estimate evaporation from a single input parameter, e.g. temperature. This approach is practical for estimating evaporation when the site's meteorological input variables are insufficient. According to our findings developed, Eq. (22) (model3) has the advantage of including all three variables investigated in this study; nevertheless, the high scatter in the wind speed data can compromise this model's accuracy. Model 2 reduces relative humidity using the enhanced performance of the two parameters, temperature and wind speed. Tables 3 and Table4 present ANOVA and analysis results for the three developed models. The F-test value indicates an excellent linear association between evaporation and different identified input parameters for all three models. As expected, the value for the wind speed variable in Model 3 is significant, showing that the zero-slope null hypothesis is correct. Model 2 is hence the most efficient evaporation model available.

4.0 Conclusion

Evaporation is a crucial factor in numerous hydrological and water resource systems initiatives. This study was undertaken at Jorhat, Assam, to develop empirical relations for estimating monthly pan evaporation as a function of Tavg, RH, and U. The data utilized for modelling are daily observations of considerable continuous coverage from January 2004 to December 2014. The optimal model forms can be determined using multiple linear regression approaches in conjunction with a variable selection procedure. Using the power and exponential functions, the correlations of evaporation with Tavg and RH are modified to linear form the existing nonlinear patterns of the data. The findings generated by the developed models with the optimal variable combinations correspond reasonably well to the observed values.

Multiple regression analysis is an effective method for predicting the value of an unknown variable based on the known values of two or more predictor variables (Gundalia et al., 2013). It is commonly employed for forecasting and predicting. ANOVA gives a statistical test to determine whether the means of many groups are equal and extends the t-test to include more than two groups (Hanson, 1987). These are useful for comparing (testing) the statistical significance of three or more means (groups or variables). Therefore, multiple linear regression was planned to estimate pan evaporation in limited data conditions. In the present study, empirical relations were derived for modelling pan evaporation applicable to Jorhat, Assam.

The study's primary findings are as follows: The modifications used for temperature, relative humidity, and wind speed variables improved the correlation results. The three models used here produced generally acceptable outcomes with decent accuracy. Two input parameter models were deemed the best based on the evaluation criteria used. If wind speed data is provided, three variable models perform well.

References

Abtew, W., (2001). Evaporation estimation for Lake Okeechobee in South Florida. J. Irri. Drain. Eng., ASCE. 127 (3), 140 –147.

Almedeij, J., (2012). Modeling Pan Evaporation for Kuwait by Multiple Linear Regression. The Scientific World Journal, doi:10.1100/2012/574742.

Bruton, J. M., McClendon, R. W., Hoogenboom, G., (2000). Estimating daily pan evaporation with artificial neural Networks. American Society of Agricultural Engineers, VOL. 43(2): 491-496.

Fekih, M., Bourabaa, A., Mohamed, S., (2013). Evaluation of two methods for estimation of evaporation from Dams water in arid and semi-arid areas in Algeria. International Journal of Application or Innovation in Engineering & Management, Volume 2, Issue 1.

Fennessey, N. M., & Vogel, R. M. (1996). Regional models of potential evaporation and reference evapotranspiration for the northeast USA. *Journal of Hydrology*, *184*(3-4), 337-354.

Goven, A., Kisi, O., (2001). Daily pan evaporation modeling using linear genetic programming technique. Irrig Sci, 29:135–145.

Gundalia, M.J., Dholakia, M.B., (2013). Estimation of Pan Evaporation Using Mean Air Temperature and Radiation for Monsoon Season in Junagadh Region. Int. Journal of Engineering Research and Applications, Vol. 3(6):64-70.

Gundalia, M.J., Dholakia, M.B., (2013). Modeling pan evaporation using mean air temperature and mean pan evaporation relationship in middle south Saurashtra Region. International Journal of Water Resources and Environmental Engineering, Vol. 5(11):622-629.

Hanson, C.L., (1987). Prediction of class A Pan evaporation in southwest Idaho. J. Irri. Drain. Eng., ASCE. 115 (3), 140 –147.

Hargreaves, G. H., & Samani, Z. A. (1982). Estimating potential evapotranspiration. Journal of the irrigation and Drainage Division, 108(3), 225-230.

Kişi, O., Tombul, M., (2013). Modelling monthly pan evaporations using fuzzy genetic approach. Journal of Hydrology, Vol. 477: 203–212.

Kumar, P., Rasul, G., Kumar, D., (2013). Evaporation Estimation from Climatic Factors. Pakistan Journal of Meteorology, Vol. 9, Issue 18.

Linacre, E. T. (1973). simpler empirical expression for actual evapotranspiration rates--a discussion. *Agricultural meteorology*.

Matsui, H., (2003). Estimation of Mean Monthly Pan Evaporation Based on Hargreaves Equation. Journal of japan society of hydrology and water resources, Vol. 16: 589-594.

Piri, J., Amin, S., Moghaddamnia, A., Keshavarz, A., Han, D., Remesan, R., (2010). Daily Pan Evaporation Modelling in a Hot and Dry Climate. Journal of hydrologic engineering, Vol. 14(8):803-811.

Piri, J., Ansari, h., (2012). Daily Pan Evaporation Modeling with ANFIS and NNARX. Iran Agricultural Research, Vol. 31.

Tabari, H., Safar, M., Sabziparvar, A., (2009). Estimation of daily pan evaporation using artificial neural network and multivariate nonlinear regression. Journal of irrigation science, Vol. 28:399-406.

Penman, H. L. (1956). Estimating evaporation. *Eos, Transactions American Geophysical Union*, *37*(1), 43-50.