RESEARCH PAPER | International Journal of Agricultural Engineering | Volume 10 | Issue 2 | October, 2017 | 503-507
e ISSN–0976–7223 ■ Visit us : *www.researchjournal.co.in* ■ **DOI: 10.15740/HAS/IJAE/10.2/503-507**

Evaporation estimation from meteorological parameters using multiple linear regression model

■ **NEERAJ KUMAR, GAURANG JOSHI, ROHIT KUMAR AND PANKAJ KUMAR**

Received : 22.07.2017; **Revised :** 23.08.2017; **Accepted :** 09.09.2017

See end of the Paper for authors' affiliation

Correspondence to :

NEERAJ KUMAR
Department of Soil and Water Conservation Engineering, College of Technology, Govind Ballabh Pant University of Agriculture and Technology, Pantnagar, U.S. NAGAR (UTTARAKHAND), INDIA
Email : neerajkumarsaini.126@gmail. com

■ **ABSTRACT :** Evaporation is one of the main elements affecting water storage and temperature in the hydrological cycle and it plays an important role in evaluation of water availability. Considering the difficulty involved in direct method of evaporation estimation and limitation of empirical methods, an attempt has been made to estimate evaporation by multiple linear regression with the aid of gamma test (GT). The data of meteorological parameters *viz.*, average temperature (T_{avg}), wind speed (W), average relative humidity (Rh_{avg}) and sunshine hours (S) were used as input parameters and evaporation was considered as output parameter. The performance of developed model was evaluated in terms of mean square error (MSE) and correlation co-efficient (r). In developed model, MSE was found to be 1.13 and 0.92 in training and testing phase, respectively. The model demonstrated good values of correlation co-efficient, respectively as 0.91 and 0.95 for training and testing period indicating the suitability of model to estimate the evaporation.

■ **KEY WORDS :** Evaporation, Meteorological parameters, MLR, Gamma test

■ **HOW TO CITE THIS PAPER :** Kumar, Neeraj, Joshi, Gaurang, Kumar, Rohit and Kumar, Pankaj (2017). Evaporation estimation from meteorological parameters using multiple linear regression model. *Internat. J. Agric. Engg.*, **10**(2) : 503-507, **DOI: 10.15740/HAS/IJAE/10.2/503-507.**

Evaporation is the process whereby atoms or particles in a liquid state (or solid state if the substance sublimes) gain sufficient energy and changes to the gaseous state. It is the surface phenomena and a reverse process of condensation. The thermal motion of a molecule of liquid must be sufficient to overcome the surface tension and evaporate, that is, its kinetic energy must exceed the work function of cohesion at the surface.

Evaporation is important part of water cycle which influences the water availability and temperature in the hydrological cycle. Water loss due to evaporation becomes unavoidable during survey, management and designing of water conveyance structure and surface water reservoirs. Thus estimation of water loss due to evaporation is necessity for optimum water requirement. Direct estimation of evaporation is considered as an important element for further improvement in water management. The rate of evaporation depends upon vapour pressure difference and energy available in the medium which further majorly depends upon temperature, relative humidity, sunshine hours and wind speed. Evaporation can be estimated by direct or indirect methods. Evaporation pans (Class A evaporation pan, sunken Colorado pan etc.) are used for the direct estimation of evaporation. However, there are also numbers of empirical methods available like penman for indirect estimation of evaporation but these are relatively more complicated. Many researchers developed the models for evaporation estimation using different techniques (Stewart and Rouse, 1976; Bruin, 1978; Anderson and Jobson, 1982 and Abtew, 2001). To overcome the complexity of empirical formulas, Cahoon *et al.* (1991) and Fennessey and Vogel (1996) applied the regression method to develop model in order to find the monthly average evaporation in USA. Murthy and Gawande(2006) proposed the linear dependency between evaporation and climatic variables by local linear regression method.

The researchers also applied neural network techniques for prediction of evaporation and reported good results (Hoskins and Himmelblau, 1988; Chiu *et al.*, 1995; Cook and Chiu, 1997; Gao *et al.*, 1997). But for high accuracy, these techniques require long time series data for training and testing of model. Hastle and Loader (1993); Wand and Jones (1995) and Fan and Gijbels (1996) explored extensive literatures on theoretical and practical properties of local linear based modeling. With linear regression technique, initial prediction can be performed only with three data points and other data can be used to make future predictions which make it a dynamic process for predicting time series data.

Linear regression is widely used in statistical assessment. The profits of a linear model are due to its simplicity and ease of use, while high model bias becomes its drawback. Nowadays, non-parametric regression is attracting much attention in the background in various fields: striking progress in computers, advancements in studies on neural networks, enhancement of the demand for data mining, and accumulation of diverse data by allowing the improvement of measurement and observation techniques. Linear regression method is one of the reliable statistical modelling techniques which can predict the results even with a small amount of sample data while high data density in the input space moves it towards more accurate prediction.

Cochrane-Orcutt, Kernel, weighted least square (WLS) etc. are different estimators for linear regression but weighted least square is widely used estimator for fitting of linear regression based model. Many researchers worked for modifications and perfections of kernel and local linear estimators to reduce the bias. Abramson (1982); Samiuddin and El-Sayyad (1990) and Jones *et al.* (1995) purposed the idea of taking linear combinations of some local linear estimators as well. Robinson (1983); Yakowitz (1987); Diggle and Hutchinson (1989); Hardle(1990) and Tjostheim (1994) applied kernel regression for solving the problems having time series data. But, Nottingham (1995) and Fan and Gijbels (1996) demonstrated the superiority of local linear regression over kernel regression in many applications due to occurrence of boundary bias in kernel regression.

Chandra *et al.* (1988) applied multiple linear regression (MLR) to investigate the effect of climatic parameters on evaporation and reported that air temperature, relative humidity, wind speed and number of sunshine hours can be well correlated with pan evaporation. Alshaikh (1998) conducted a study using linear, nonlinear single and multiple regression to find the best fit evaporation model for arid region of Saudi Arabia. Shrivastava *et al.* (2000) made a linear relationship between climatic parameters namely wind speed, minimum temperature and maximum relative humidity and evaporation for Sunderbans, West Bengal. The model explained 92.1% variability of evaporation. Senturk and Kocyigit (2010) developed four different models based on MLR technique by changing number of independent variables. Among all, a model involving all independent parameters *viz.*, temperature, wind speed, relative humidity and radiation was found more reliable. Almedeij (2012) developed a MLR based model for estimation of daily and monthly evaporation in Kuwait.

The literature shows that linear regression can easily fit equation between response variable and predictors even with small data. Considering complexity and limitation experienced with empirical methods, this study was undertaken to develop a model in order to estimate weekly evaporation from meteorological data using multiple linear regression (MLR) with the aid of gamma test (GT). The climatic variables sunshine hours, temperature, relative humidity and wind speed were used as independent variables and evaporation was used as dependent variable.

Study area :

The data considered for this study concerns to Meteorological Observatory, G.B. Pant University of Agriculture and Technology, Pantnagar. The selected location falls in sub-humid and subtropical climatic zone and situated in Tarai belt of Shivalik range, of foot hills of Himalayas. Geographically it is located at 29°N latitude and 79.29°E longitude and an altitude of 243.84 m above mean sea level. The mean annual rainfall is 1364 mm of which 80-90% occurs during June to September. The mean relative humidity remains almost 80-90% from mid-June to February end.

METHODOLOGY

The weekly meteorological data (2002 to 2011) used in this study were collected from Meteorological Observatory, G.B. Pant University of Agriculture and Technology, Pantnagar, India. The statistical details of weekly meteorological data are given in Table A. In this study, multiple linear regression with weighted least square estimator was used for fitting of the model. Out of 10 years meteorological data, 7 year data were used for training and remaining 3 year data were used for testing of model. The mean square error (MSE) and correlation co-efficient (r) calculated, respectively from Eqns. (1) and (2) were used to judge the performance of developed model.

$$\text{MSE} = \frac{\sum_{i=1}^{i=n}(E_{i,pred} - E_{i,act})^2}{n} \quad (1)$$

$$\text{Correlation co-efficient (r)} = \frac{n(\sum E_{act} E_{pred}) - (\sum E_{act})(\sum E_{pred})}{\sqrt{[\{n\sum E_{act}^2 - (\sum E_{act})^2\}\{n\sum E_{pred}^2 - (\sum E_{pred})^2\}]}} \quad (2)$$

where,

E_{act} = Actual value of evaporation (mm),

E_{pred} = Predicted value of evaporation (mm),

and n = Number of data sets.

Table A : Statistical details of weekly meteorological data (2002 to 2011)

Parameters	Min. value	Max. value	Mean value	Standard deviation
Sunshine hours (h)	0.5	11.7	6.99	2.36
Wind speed (km/h)	0.6	19.8	5.03	3.07
Temperature (^{o}C)	7	33.4	23.34	6.26
Relative humidity (%)	33	90.5	67.79	12.32
Evaporation (mm)	0.5	17.2	4.40	2.70

Selection of model input parameters using gamma test :

Gamma test was used for selecting the best combination of independent parameters used in the model. It is a modelling tool widely used in non-parametric methods for selection of effective parameters in order to develop a model with less mean square error (MSE). Agalbjorn *et al.* (1997) and Koncar (1997) used gamma test in their study to find the best combination of parameters. The gamma test can be performed easily in win Gamma™ software (Tsui *et al.*, 2002). Based on gamma test result, the input parameters with minimum gamma value can be selected and a model with less MSE can be developed.

RESULTS AND DISCUSSION

The results obtained from the present investigation as well as relevant discussion have been summarized under following heads :

Gamma test :

The gamma test was performed in order to select best combination of input parameters and thereby reducing the possibility of higher mean square error in the model. The gamma values attained with different combination of input parameters are presented in Table 1. It can be observed in Table 1 that the input parameters *viz.*, wind speed (W), sunshine hours (S), average temperature (T_{avg}) and average relative humidity (Rh_{avg}) indicated minimum gamma value. So these parameters were used in the model as predictors.

Training of multiple linear regression model:

Out of total (520), 364 meteorological datasets were used in the training period. The multiple linear regression was used to explore the relation between output and input parameters. The model developed the relation between evaporation (E) and input parameters *viz.*, W, S, T_{avg} and Rh_{avg} as given in Eqn. (3). The developed model showed mean square error and correlation co-efficient as 1.13 and 0.91, respectively. A good agreement can be observed between actual and predicted evaporation as shown in Fig. 1.

$$E_{pred} = 6.5074 - 0.1009S + 0.1585W + 0.2361T_{avg} - 0.1139\ Rh_{avg} \quad (3)$$

Testing of multiple linear regression model :

Testing of any developed model is necessary to

Table 1 : Gamma value for different input parameters

Parameters	W, S, T_{avg}, Rh_{avg}	W, S, T_{max}, Rh_{max}	W, S, T_{min}, Rh_{max}	W, S, T_{max}, Rh_{min}	W, T_{max}, Rh_{min}	W, T_{max}, Rh_{max}
Gamma value	0.48819	0.49749	0.51206	0.54896	0.58066	0.60521

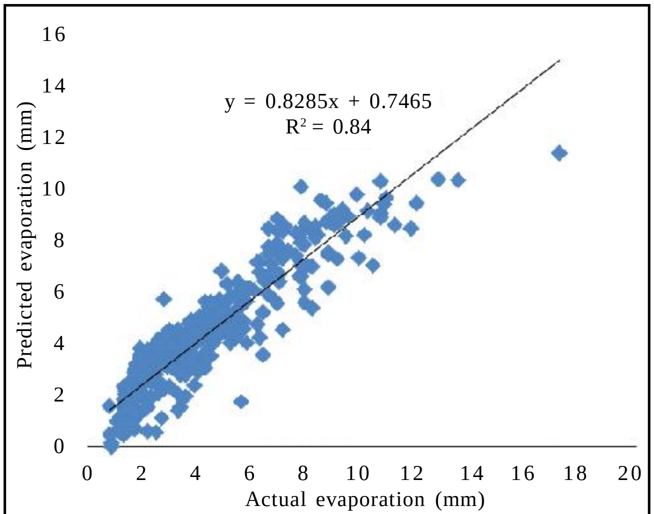

Fig. 1 : Actual vs. predicted evaporation for training period

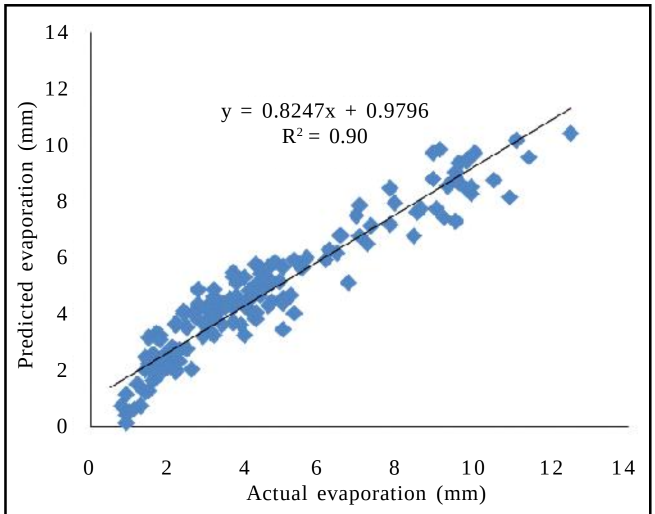

Fig. 2 : Actual vs. predicted evaporation for testing period

Table 2 : ANOVA for multiple linear regression model

Source	DF	SS	MS	F	p-value
Regression	4	3042.88	760.72	704.37	<0.0001
Residual error	515	556.54	1.08		
Total	519	3794.50	7.31		

Table 3 : Mean square error and correlation co-efficient for multiple linear regression model

Mean square error			Correlation co-efficient		
Training	Testing	Overall	Training	Testing	Overall
1.13	0.92	1.07	0.91	0.95	0.92

examine its working accuracy with different data. The developed model was tested with 156 meteorological data sets. The values of mean square error and correlation co-efficient were found to be 0.92 and 0.95, respectively for testing period. A good linear dependency between actual and predicted evaporation can be observed in Fig. 2. The analysis of variance (ANOVA) for multiple linear regression model is presented in Table 2. The overall results of developed multiple linear regression model is shown in Table 3.

From the Table 2 it is clear that p-value is less than significance level of 0.001. It provides strong evidence for alternative hypothesis which states that at least one of the regression co-efficients is not equal to zero which is also supported by Eqn. (3).

Conclusion :

From this study, it can be concluded that gamma test helped in the selection of best combination of independent parameters resulting in lower value of MSE during modeling. The developed model predicted weekly evaporation with good accuracy. A good linear dependency observed between actual and predicted values of evaporation in training as well as in testing phase of model. So, multiple linear regression model can be used to estimate the evaporation from meteorological data easily as compared to other time consuming and complex methods of evaporation estimation.

Authors' affiliations:
GAURANG JOSHI, ROHIT KUMAR AND PANKAJ KUMAR, Department of Soil and Water Conservation Engineering, College of Technology, Govind Ballabh Pant University of Agriculture and Technology, Pantnagar, U.S. NAGAR (UTTARAKHAND), INDIA

REFERENCES

Abramson, I.S. (1982). On bandwidth variation in kernel estimates-a square root law. *Ann. Statist.*, **10** : 1217–1223. MR0673656

Abtew, W. (2001). Evaporation estimation for Lake Okeechobee in South Florida. *Irrigation & Drainage Engg.*,

127 : 140-147.

Agalbjorn, S., Koncar, N. and Jones, A.J. (1997). A note on the gamma test. *Neural Comput. Appl.,* **5**(3) : 131–133. ISSN 0-941-0643.

Almedeij, J. (2012). Modeling pan evaporation for Kuwait by multiple linear regression.*The Scientific World J.,* 2012 (article id 574742), pp. 1-9.

Alshaikh, A. (1998). Analysis of evaporation data as affected by climatic factors in arid regions. Water and land resources development and management for sustainable use. Vol. II-A. The tenth ICID Afro-Asian regional conference on irrigation and drainage, Denpasar, Bali, Indonesia, 19-26 July 1998. A-27, 9 –10.

Anderson, M.E. and Jobson, H.E. (1982). Comparison of techniques for estimating annual lake evaporation using climatological data. *Water Resour. Res.,* **18** : 630-636.

Bruin, H.D. (1978). A simple model for shallow lake evaporation. *Applied Meteorol.,* **17** : 1132-1134.

Cahoon, E.J., Costello, T.A. and Ferguson, J.A. (1991). Estimating pan evaporation using limited meteorological observations. *Agric. & Forestry Meteorol.,* **55** :181-190.

Chandra, A., Shrikhande, V.J. and Kulshreshta, R. (1988). Relationship of pan evaporation with meteorological parameters. *J. Indian Water Reso. Soc.*, **8**(2) : 41- 44.

Chiu, C.C., Cook, D.F. and Pignatiello, J.J. (1995). Radial basis function neural network for Kraft pulping forecasting. *Internat. J. Ind. Eng*., **2** (3) : 209–215.

Cook, D.F. and Chiu, C.C. (1997). Predicting the internal bond strength of particleboard utilizing a radial basis function neural network. *Engg. Appl. AI.,* **10**(2) : 171–177.

Diggle, P.J. and Hutchinson, M.F. (1989). On spline smoothing with auto-correlated errors. *Australian J. Statist.,* **31** :166–182.

Fan, J. and Gijbels, I. (1996). *Local polynomial modeling and its applications*. Chapman & Hall, London. *Great lakes buoy and associated synoptic weather patterns*. J. Hydro Meteorol.*,* **3**(1) : 3-12.

Fennessey, N.M. and Vogel, R.M. (1996). Regional models of potential evaporation and reference evapotranspiration for the northeast USA. *J. Hydrol.,* **184** : 337-354.

Gao, X.M., Gao, X.Z., Tanskanen, J. and Ovaska, S.J. (1997). Power prediction in mobile communications systems using an optimal neural network structure. *IEEE Trans. Neural Networks,* **8**(6) : 1446–1455.

Hardle, W. (1990). *Applied nonparametric regression*. Cambridge University Press, New York.

Hastle, T. and Loader, C. (1993). Local regression: Automatic kernel carpentry (with discussion). *Statist. Sci.,* **8** : 120–143.

Hoskins, J.C. and Himmelblau, D.M. (1988). Artificial neural networks models of knowledge representation in chemical engineering. *Comput. Chemi. Engg.,* **12**(9) : 881–890.

Jones, M.C., Linton, O. and Nielsen, J.P. (1995). A simple bias reduction method for density estimation. *Biometrika*, **82**(2) : 327–338.

Koncar, N. (1997). Optimisation methodologies for direct inverse neurocontrol. Ph.D. Thesis, Department of computing, Imperial College of Science, Technology and Medicine, University of London.

Murthy, S. and Gawande, S. (2006). Effect of meteorological parameters on evaporation in small reservoirs 'Anand Sagar' Shegaon – a case study. *J. Prudushan Nirmulan,* **3**(2) : 52-56.

Nottingham, Q.J. (1995). Model-robust quantile regression. Ph.D. Dissertation, Virginia.

Robinson, P.M. (1983). Nonparametric estimators for time series. *J. Time Series Anal*., **4** :185–207.

Samiuddin, M. and El-Sayyad, G. M. (1990). On nonparametric kernel density estimates. *Biometrika,* **77** : 865–874. MR1086696

Senturk, K. and Kocyigit, F.O. (2010). A case study: Evaporation estimation at Oymapinar Dam. Proc., BALWOIS 2010, Balkan institute for water and environment, Montpellier, France, 1–7.

Shrivastava, S.K., Misra, S.K., Sahu, A.K. and Bose, D. (2000). Correlation between pan evaporation and climatic parameters for Sunderbans- a case study. *IE(I) Journal – AG*. 81, 55-58.

Stewart, R.B. and Rouse, W.R. (1976). A simple method for determining the evaporation from shallow lakes and ponds. *Water Resour. Res.,* **12** : 623-627.

Tjostheim, D. (1994). Nonlinear time series: a selective review. *Scandinavian J. Statist.,* **21** : 97–130.

Tsui, A.P.M., Jones, A.J. and Oliveira, A.G. (2002). The construction of smooth models usingirregular embeddings determined by a gamma test analysis. *Neural Comput Appl.,* **10**(4) : 318–329. doi:10.1007/s005210200004

Wand, M.P. and Jones, M.C. (1995). *Kernel smoothing*, Chapman and Hall, London. MR1319818.

Yakowitz, S. (1987). Nearest-neighbor methods for time series analysis. *J. Time Series Anal*., **8** : 235–247.

10th Year
★★★★★ of Excellence ★★★★★