

## Evaporation estimation from climatic factors

PANKAJ KUMAR, DEVENDRA KUMAR and RAJDEV PANWAR

*G. B. Pant University of Agriculture & Technology, Pantnagar – 246 123, Uttarakhand, India*

*(Received 22 April 2013, Accepted 12 May 2015)*

**e mail : pankaj591@gmail.com**

**सार** – इस अध्ययन में पंतनगर, उधम सिंह नगर (भारत) जो कि उपआर्द्र और उपउष्णकटिबंधीय जलवायविक क्षेत्र के अंतर्गत आता है, मासिक विभव वाष्पोत्सर्जन का आकलन करने के लिए दो मॉडलों, स्थानीय समाश्रयण (LLR) और कृत्रिम तंत्रिका संजाल (ANN) को क्षमता का मूल्यांकन किया गया। सापेक्षिक आर्द्रता, सौर विकिरण तापमान, पवन गति और वाष्पोत्सर्जन के प्रेक्षणों का उपयोग इन मॉडलों को ट्रेन्ड और टैस्ट करने के लिए किया गया। LLR मॉडल और ANN मॉडल से प्राप्त हुए आकलनों की परस्पर तुलना की गई। इनसे प्राप्त हुए परिणामों से पता चला है कि इन मॉडलों को जिस उद्देश्य के लिए तैयार किया गया था उसे पूरा करने में ये सक्षम हैं। ट्रेनिंग की अवधि में ANN मॉडल के लिए सह-संबंध गुणांक 0.9311 है और टैस्टिंग अवधि 0.9236 है। ट्रेनिंग अवधि का रूटमीन स्केयर एरर 1.070 है और टैस्टिंग अवधि में यह 0.9863 है। LLR मॉडल के मामले में ट्रेनिंग की अवधि के लिए सह-संबंध गुणांक 0.9746 है और टैस्टिंग अवधि 0.9273 है। ट्रेनिंग अवधि के रूटमीन स्केयर एरर का मान 0.6121 है और टैस्टिंग अवधि का मान 1.5301 है।

**ABSTRACT.** This study assessed the ability of two models, Local Linear Regression (LLR) and Artificial Neural Network (ANN) to estimate monthly potential evaporation from Pantnagar, US Nagar (India) which falls under sub-humid and subtropical climatic zone. Observations of relative humidity, solar radiation, temperature, wind speed and evaporation have been used to train and test the developed models. A comparison was made between the estimates provided by the LLR model and ANN model. Results shown that the models were able to well learn the events they were trained to recognize. For ANN model the correlation coefficient for training period is 0.9311 and for testing period is 0.9236 and the value of root mean square error for training period is 1.070 and for testing period it is 0.9863. In case of LLR model the correlation coefficient for training period is 0.9746 and for testing period is 0.9273 and value of root mean square error for training period is 0.6121 and for testing period it is 1.5301.

**Key words** – Evaporation, Estimation, Local linear regression, Artificial neural network.

### 1. Introduction

Evaporation refers to water losses from the surface of a water body to the atmosphere. Evaporation occurs when the number of moving molecules that break from the water surface and escape into the air as vapour is larger than the number that re-enters the water surface from the air and become entrapped in the liquid. Evaporation increases with high wind speed, high temperatures and low humidity. A sizable quantity of water is lost every year by evaporation from storage reservoirs and evaporation of water from large water bodies influences the hydrological cycle. Among the hydrological cycle, evaporation is perhaps the most difficult to estimate due to complex interactions among the components of land-plant-atmosphere system (Singh and Xu, 1997).

The most common and important factors affecting evaporation are solar radiation, temperature, relative humidity, vapour pressure deficit, atmospheric pressure, and wind. Evaporation losses should be considered in the design of various water resources and irrigation systems. In areas with little rainfall, evaporation losses can represent a significant part of the water budget for a lake or reservoir and may contribute significantly to the lowering of the water surface elevation. Therefore, accurate estimation of evaporation loss from the water body is of primary importance for monitoring and allocation of water resources, at farm scales as well as at regional scales. The rate of evaporation depends on a number of meteorological factors such as solar radiation, air temperature, relative humidity, wind speed and to some extent atmospheric pressure. Other factors are

related to the nature of the evaporating surface and the quality of water. Various studies have been conducted to determine which of these factors have the dominant effect on evaporation. Radiation is by far the most important single factor affecting evaporation. In addition to solar radiation, (Chow *et al.*, 1988) claimed that the mechanism of transporting the vapour from the water surface has also a great effect. Vapour pressure deficit, temperature, barometric pressure, humidity and wind speed were emphasized by Singh (1992) as the controlling factors.

Gupta (1992) pointed out that relative humidity, wind velocity and temperature of water and atmosphere are the climatic factors evaporation awfully depends on. In summary, it has been agreed that solar radiation, wind speed, relative humidity, and air temperature have attained special consideration as the most influencing factors by most researchers.

A large number of experimental formulae exist for evaporation estimation. There are direct and indirect methods available for estimating potential evaporation from free water surfaces. Because evaporation is an incidental, nonlinear, complex, and unsteady process, it is difficult to derive an accurate formula to represent all the physical processes involved. As a result, there are new trends in using data mining techniques such as artificial neural networks techniques to estimate evaporation. As a few examples of many such studies, for evaporation modeling are by Sudheer (2002). Kisi (2006), Keskin (2006), Moghaddamia *et al.* (2009) and Kisi (2012).

The main objectives of this study were first to investigate the potential of using LLR and ANN models to predict evaporation as affected by climatic factors. Second, is to evaluate the performance of LLR and ANN models in estimating average monthly evaporation in Pantnagar.

## 2. Materials and methods

### 2.1. Local linear regression (LLR)

LLR technique is a widely studied nonparametric regression method which has been widely used in many low dimensional forecasting and smoothing problems. To make a estimation for a given query point in input space local linear regression (LLR) first finds the  $k$  nearest neighbors of the query point from the given data set and then builds a linear model using these  $k$  data points. Finally the model is applied to the query point thus producing an estimated output. Consequently local linear regression using the  $k$  nearest neighbors (in the training data) of the query point can be accomplished quickly as by Penrose (1955). Thus local linear regression is a very

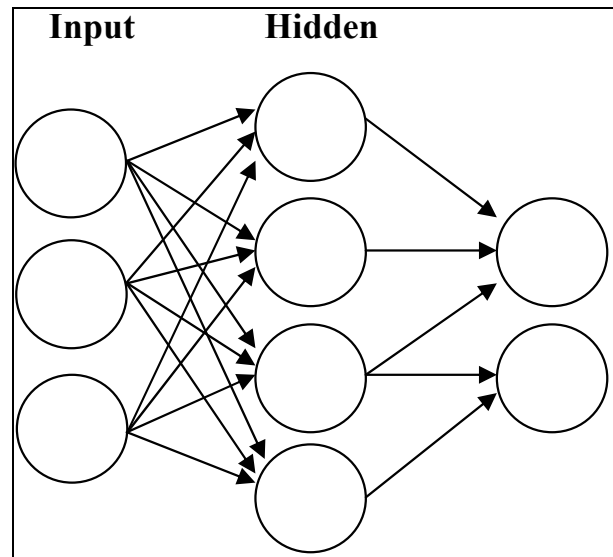


Fig. 1. Architecture of multilayer feed forward neural network (www.wikipedia.org)

fast and capable predictive tool. LLR is most effective in regions of the input space with a high density of data points. For few and far data points in the vicinity of the query point LLR model is not very effective if the underlying function to model is strictly non-linear. LLR model produces accurate predictions in the regions of high data density in input space, but it is predisposed to yield devious results for non-linear functions in regions of low data density. LLR does not generalize fine outcome but is a good interpolative model for large amounts of data. The LLR procedure requires only three data points to obtain an initial prediction and then uses all newly updated data as they becomes available to make further estimation.

### 2.2. Artificial neural networks (ANN)

ANN was first introduced as a mathematical aid by McCulloch and Pitts (1943). They were inspired by the neural structure of the brain. Fig. 1 is a general architecture of a Feed Forward ANN, with one hidden layer. Most ANNs have three layers or more: an input layer, which is used to present data to the network; an output layer, which is used to produce an appropriate response to the given input; and one or more intermediate layers, which are used to act as a collection of feature detectors. The ability of a neural network to process information is obtained through a learning process, which is the adaptation of link weights so that the network can produce an approximate output. In general, the learning process of an ANN will reward a correct response of the system to an input by increasing the strength of the current matrix of nodal weights.

**TABLE 1**  
**Statistical analysis of the total monthly weather data**

S. No.	Data	Maximum	Minimum	Correlation coefficient with evaporation
1.	Air temperature (°C)	32.35	10.45	0.7625
2.	Relative humidity (%)	89	38.5	- 0.640
3.	Wind velocity (m/s)	14.2	0.7	0.6612
4.	Sunshine hours (hour)	10.5	3	0.4931
5.	Evaporation (mm)	13.1	1.1	1.00

There are several features in ANN that distinguish it from the empirical models. First, neural networks have flexible non-linear function mapping capability which can approximate any continuous measurable function with arbitrarily desired accuracy, whereas most of the commonly used empirical models do not have this property. Second, being non-parametric and data-driven, neural networks impose few prior assumptions on the underlying process from which data are generated. Because of these properties, neural networks are less susceptible to model misspecification than most parametric nonlinear methods.

There are a wide variety of algorithms available for training a network and adjusting its weights. In this study, an adaptive technique momentum Levenberg-Marquardt based on the generalized delta rule was adopted.

Let  $x_i$  ( $i = 1, 2, \dots, n$ ) are inputs and  $w_i$  ( $i = 1, 2, \dots, n$ ) are respective weights. The net input to the node can be expressed as :

$$\text{net} = \sum_{i=1}^n x_i w_i \quad (1)$$

The net input is then passed through an activation function  $f(\cdot)$  and the output of the node is computed as

$$y = f(\text{net})$$

Sigmoid function is the most commonly used nonlinear activation function which is given by

$$y = f(\text{net}) = \frac{1}{1 + e^{-\text{net}}} \quad (2)$$

Throughout all ANN simulations, the adaptive learning rates were used for increasing the convergence

velocity. For each epoch, if the performance decreases toward the goal, then the learning rate is increased by the factor of learning increment. If the performance increases, the learning rate is adjusted by the factor of learning decrement.

### 2.3. Study area

The weekly evaporation data for the year 1990 to 2009 (236 months) approximately 19 years and 8 months were collected from Meteorological Observatory, G. B. Pant University of Agriculture and Technology, Pantnagar, District Udham Singh Nagar, India. Pantnagar 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. Generally, monsoon starts in the last of June and continues up to September. The mean annual rainfall is 1364 mm of which 80-90 per cent occurs during June to September. May to June is the hottest months and December and January the coldest. The mean relative humidity remains almost 80-90 per cent from mid-June to February end.

As far as the significance of individual meteorological parameters is concerned, the study revealed that the highest value of correlation coefficient and least value of root mean square error were obtained for evaporation with air temperature, followed by using wind speed and relative humidity (Table 1). While the lowest correlation coefficient was obtained with sunshine hours, which mean bright sunshine hours alone does not appear to influence the evaporation significantly. The effect of air temperature, wind speed and sunshine hours was found to be positive; whereas a negative correlation exists between evaporation and relative humidity (that is evaporation decreases with increase in relative humidity). It is a natural fact that the climatic/meteorological factors

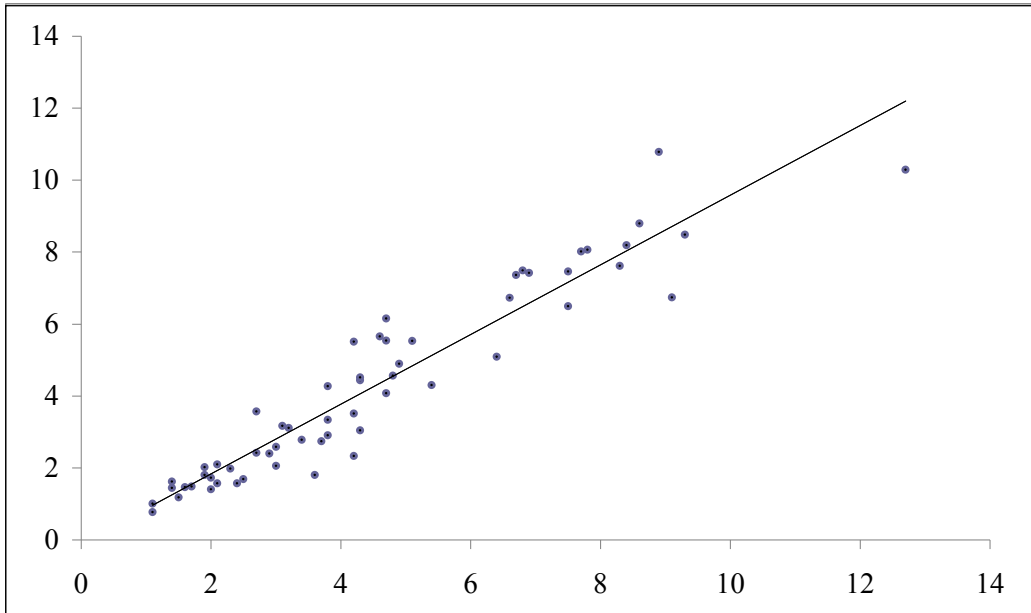


Fig. 2. Observed and predicted monthly evaporation for ANN model in testing period

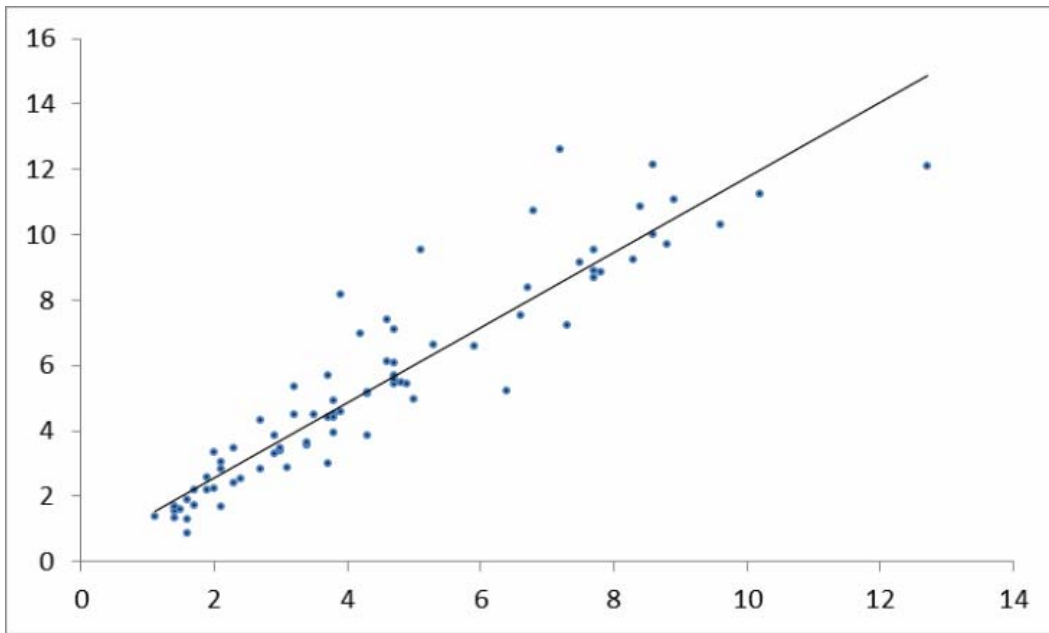


Fig. 3. Observed and predicted monthly evaporation for LLR model in testing period

in general act in concert. Therefore, it is pertinent to take into account the combined influence of all the meteorological parameter on evaporation. By various trials it was suggested that a combination of temperature,

wind speed, sunshine hour and humidity provides a maximum value of correlation coefficient with minimum values of root mean square error in comparison to other inputs combinations.

TABLE 2

Performance evolution of model on training and testing period

	Models	Correlation coefficient	Root mean square error
Training	ANN	0.9311	1.070
	LLR	0.9746	0.6121
Testing	ANN	0.9236	0.9863
	LLR	0.9273	1.5303

The input combinations used in this application to estimate evaporation for Pantnagar station were Air temperature (°C), relative humidity (%), wind velocity (m/s) and sunshine hours (hour) of a month  $t$  and evaporation (mm) of that month  $t$  was considered as output of the models.

Different LLR architectures were tried using these inputs and the appropriate numbers of nearest neighbors. Various LLR models were tested for highest correlation coefficient and lowest root mean square error statistics. 157 data sets were used for training and 79 months data were used for testing for ANN models.

Different Feed Forward ANN architectures were tried using these inputs and the appropriate model structures were determined for each input combination. Then, the ANN models were tested and the results were compared by means of correlation and coefficient root mean square error statistics. 157 data sets were used for training and 79 months data were used for testing for ANN models.

For best fit ANN model in present study multilayer perceptron with one hidden layer and with a sigmoid activation function was used as it works well for this data set. Other user-defined parameters used were momentum learning rate and step size = 0.1, momentum = 0.700, hidden layer nodes = 4 and iterations = 1000. These values were obtained after a number of trials by using different combination of these parameters carried out on data set.

### 3. Results and discussion

The correlation coefficient and root mean square error values of developed model in the training period as well as in testing period are given in Table 2. It can be seen from the table that for ANN model the correlation coefficient for training period is 0.9311 and for testing period is 0.9236. The value of root mean square error for training period is 1.070 and for testing period it is 0.9863. For LLR model the correlation coefficient for training

period is 0.9746 and for testing period is 0.9273 and value of root mean square error for training period is 0.6121 and for testing period it is 1.5301 respectively. It is clear from Table 2 that the higher values of correlation coefficients and lower values of root mean square error suggests the applicability of LLR model for evaporation estimation over the ANN model.

### 4. Conclusion

The present study discusses the application and usefulness of LLR and ANN based modelling approaches in estimation of evaporation losses over a region. The results are quite encouraging and suggest the usefulness of both modeling technique in estimation of the evaporation. In case of LLR model the values of correlation coefficient are better than ANN model for both training and testing periods. Root mean square error value for training period is better for LLR model and for testing period root mean square values of ANN model is found slightly greater than LLR model. Due to having less complexity and less computational efforts to execute LLR model is found appropriate over Feed Forward ANN. This study also concludes that a combination of mean air temperature, wind speed, sunshine hour and mean relative humidity provides better performance in predicting the evaporation losses.

### References

- Chow, V., Maidment, D. and Mays, L., 1988, "Applied Hydrology" N. Y.: McGraw-Hill Pub.
- Gupta, B., 1992, "Engineering Hydrology", Jain, India: N.C.
- Keskin, M. E. and Terzi, O., 2006, "Artificial neural network models of daily pan evaporation", ASCE, *J. Hydrol. Eng.*, **11**, 1, 65-70.
- Kisi, O., 2006, "Daily pan evaporation modelling using a neuro-fuzzy computing technique", *Journal of Hydrology*, **329**, 636- 646.
- Kisi, O., Ali-Baba, A. P. and Shiri, J., 2012, "Generalized neuro-fuzzy models for estimating daily evaporation values from weather data", ASCE, *J. Irrig. Drain. Eng.*, **138**, 4, 349-362.
- McCulloch, W. S. and Pitts, W., 1943, "A logic calculus of the ideas immanent in nervous activity", *Bulletin of Mathematical and Biophysics*, **15**, 115-133.
- Moghaddamia, A., GhafariGousheh, M., Piri, J., Amin, S. and Han, D., 2009, "Evaporatin estimation using artificial neural networks and adaptive neuro-fuzzy inference system techniques", *Adv. Water Resour.*, **32**, 88-97.
- Penrose, R., 1955, "A generalized inverse for matrices", *Proceedings of the Cambridge Philosophical Society*, **51**, 406-413.
- Singh, V., 1992, "Elementary Hydrology", NJ, U.S.A.: Prentice Hall Inc.

Singh, V. P. and Xu, C. Y., 1997, "Evaluation and generalization of 13 mass transfer equations for determining free water evaporation", *Hydrological Processes*, **11**, 311-323.

Sudheer, K. P., Gosain, A. K., Mohan, R. D. and Saheb, S. M., 2002, "Modelling evaporation using an artificial neural network algorithm", *Hydrological Processes*, **16**, 3189-3202.

