International Journal of Advanced Technology in Civil Engineering

Volume 1 | Issue 4 Article 6

October 2012

ARTIFICIAL NEURAL NETWORK METHOD FOR ESTIMATION OF MISSING DATA

HARSHANAND K. GHUGE
Government College of Engineering, Aurangabad, India, harshanand_ghuge@yahoo.co.in

D.G. REGULWAR
Government College of Engineering, Aurangabad, India, regulwar@gmail.com

ARTIFICIAL NEURAL NETWORK METHOD FOR ESTIMATION OF MISSING DATA

HARSHANAND K. GHUGE, D.G. REGULWAR

Department of Civil Engineering ,Government College of Engineering, Aurangabad, India
Email: harshanand_ghuge@yahoo.co.in, regulwar@gmail.com

Abstract—The availability of precipitation data plays important role for analysis of various systems required for design of water resources systems. The perfect measurements are not available always. The scientist/hydrologists come across the problem of missing data due to a variety of reasons. There may be various reasons of unavailability of data. Measurement of hydrologic variables (e.g. rainfall, stream flows, etc.) is prone to various instrumental/systematic, manual and random errors. In the current study, missing rainfall data is evaluated by using Artificial Neural Network Method. Historical precipitation data from 6 rain-gauge stations in the Maharashtra State, India, are used to train and test the ANN method and derive conclusions from the improvements in result given by ANN. Results suggest that ANN model can be work for estimation of missing data.

***Keywords** - Missing data, Artificial Neural network, Pravara Sub Basin, Maharashtra.*

I. INTRODUCTION

Analysis and designing of the water resources project is essential for obtaining the desired results for society and other fields. The availability of precipitation data and its length plays vital role for hydrologic analysis and design of water resources systems. Mainly the hydrologic data is required for information about Precipitation and Evapotranspiration, Analysis for causes in variation in rainfall Statistical analysis like-Depth-Area-Duration Curve & Intensity-Duration-Frequency Curve, Design of Water Resources System and Hydrologic simulation models, etc.

The scientist/hydrologists come across the problem of missing data due to a variety of reasons. Measurement of hydrologic variables (e.g. rainfall, stream flows, etc.) is prone to various instrumental/systematic, manual and random errors. Instrumental errors in rain gage measurements can be of various types: raindrop splash from outside the rain gauge, water loss during measurement, adhesion loss on the surface of the gage, raindrop splash from the collector. Sometimes, complete data may be loss if rainfall gage malfunctions for a specific time period. Errors in recording of rainfall data are possible due to tree growth, instrumentation problems or techniques used in measuring the rainfall amounts. These errors are critical as they affect the continuity of rainfall data and ultimately influence the results of hydrologic models that use rainfall as input. Therefore, evaluation of missing data is important tasks for design of hydrological models.

A variety of techniques has been used in estimation of missing data. Ramesh S.V. Teegavarapua, V. Chandramouli [1], has studied the Inverse Distance Weighting Method (IDWM), ANN Concepts, Stochastic Interpolation Technique, Kriging are developed and tested. They found that Coefficient of correlation weighing method (CCWM), artificial neural network estimation method (ANNEM) and kriging estimation method (KEM) for estimation of missing precipitation data are superior to other methods.

D.Ruelland, S.Ardoin-Bardin, G. Billen, E.Servat [2], has examines the sensitivity of a hydrological model to several methods of spatial interpolation of rainfall data. They interpolate gridded rainfall data with various methods: Thiessen polygons, inverse distance weighted (IDW) method, thin smooth plate splines (spline), and ordinary kriging and found that interpolations using the IDW and kriging methods are the efficient, particularly when a monthly time step is used.

Jaap Brand, Stef van Buuren, Erik M. van Mulligen1, Teun Timmers, Edzard Gelsema [3], has studied about Model based method and data driven method and found that, data driven imputation method has several advantages compared to model based imputation methods.

P. Coulibaly, N.D.Evora [4], has studied six different types of artificial neural networks namely the multilayer perceptron (MLP) network , the time-lagged feed forward network (TLFN)), the generalized radial basis function (RBF) network, the recurrent neural network (RNN) ,the time delay recurrent neural network (TDRNN), the counter propagation fuzzy-neural network (CFNN) along with different optimization. The experiment results suggest that the

MLP, the TLFN and the CFNN can provide the most accurate estimates of the missing precipitation values. However, overall, the MLP appears the most effective at infilling missing daily precipitation values. The objective of present study is to estimate the missing data by using method and to study merits and demerits.

II. STUDY AREA AND DATA USED

In the catchments of Pravara river in Godavari Basin 6 Raingauge stations namely, Parner, Akole, Newasa, Rahuri, Junnar and Nashik of District Ahmednagar , Pune and Nashik of Maharshtra State, were selected for the missing data analysis in the present study (Fig. 1). The origin of the Pravara River is in the Akole Tahsil of Ahmednagar District, in the range of Kalsubai and Baleshwar. The river has total length of 208 Km in Maharashtra. It has three tributaries namely, Adula, Mahalungi, Mula. It joins Godavari river at Pravara sangam near village toka. The catchment area of the basin in Maharashtra is 6537 Sq.Km.

Topographical maps on scales 1:250000, 47E, 47I, 48M and were utilized for present study pattern. The pattern of rainfall data and location of neighboring stations with reference to the stations, where missing rainfall were required to be estimated, could be studied from these maps.

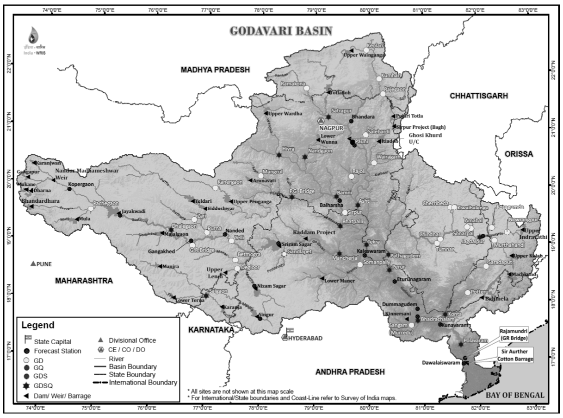

Figure 1. Godavari River Catchments

The rainfall data is obtained for the 6 raingauge stations for the period of 1901 to 1981. The station Akole is used for the study of estimation of missing data.

The rainfall data for Akole station for the month of July from period 1901 to 1981 is considered for study purpose. The data from year 1972 to 1981 is assumed to be missing data.

III. METHODS FOR MISSING RAINFALL PREDICTION

There are many methods for finding out the missing data information, some of the methods are traditional and some are modern. The conventional methods are simply the methods which can be applied for estimation of any missing data although the errors found in this due to covariance. Hence there produced the need to develop the modern techniques to overcome the errors. Some techniques to deal with the missing data are as follows:

A) Deleting the Missing Data: If the number of data available is more and the data is related with each other with less variance, then these methods can be used.

- a. List Wise Deletion
- b. Pair-Wise deletion

B) Traditional Weighting Methods:

- a. Arithmetic Weighting Methods
- b. Normal Ratio Method
- c. Inverse Distance Method

These were the traditional methods, giving less accurate results. The Modern Techniques arises are as follows:

C) Data Driven Method:

- a. Regression and Time Series Analysis
- b. Artificial Intelligence
- c. Stochastic Interpolation Techniques

D) Modern Techniques:

- a. Maximum Likelihood
- b. Multiple Imputations

In the present study, data driven ANN model have been used for estimation of missing data.

Artificial Neural Network

Artificial neural networks (ANN) have been applied extensively in the past decade in the field of hydrology for estimation and prediction of hydrological variables. The architecture of ANN used in the current study is shown in Fig.2 and Fig.3. The architecture is a feed-forward network; chosen for the analysis, in the present study, wherein the input data (rainfall at surrounding stations) are fed into nodes and will pass to the hidden nodes after getting multiplied by weight. The hidden layer neurons are selected using trial and error procedure. The output neurons of the ANN provides the missing value at the stations other than the station of interest.

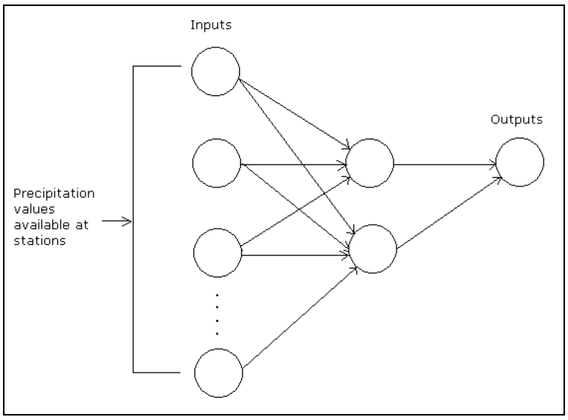

Figure 2. ANN Model for Missing data analysis

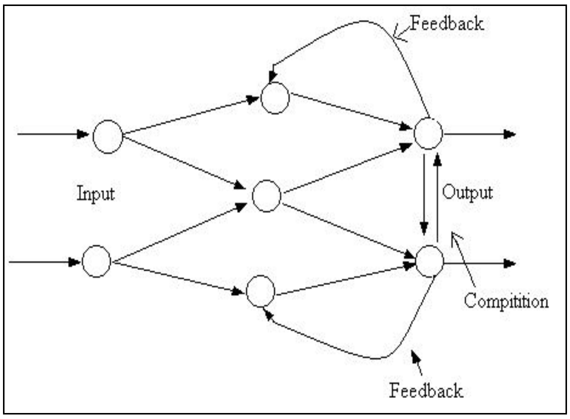

Figure 3. Simple network with feedback and competition

The rainfall data available at station is provided to ANN program. 85% data available is provided to program for training purpose and remaining 15 % are used for determining the performance of weights estimated by the program.

The program is prepared in “C++” for determination of weights to be used for estimation of missing rainfall data. The weights generated by the program are then distributed according to the class of precipitation, viz, 0-30, 31-60, 61-90, 91-120, 121-150, 151-180, 181-210,etc using Excel. The weights from the class are used for the precipitation of missing data with the appropriate class from Excel.

IV. APPLICATION OF ESTIMATION METHODS

The Artificial Neural Network (ANN) is used to estimate the missing rainfall values at Station “Akole”. The monthly rainfall data is used for this study. The rainfall data at all six stations mentioned earlier is available for the period of 1901-1981. The rainfall data at Akole station for the period of 1901-1971 is used for training purpose. The data from 1972-1981 is assumed to be missing for the purpose of testing. It is assumed that the rainfall at Akole station is missing from the period of 1971-1981 for the month of July. The performance of the predicted value by all method is then compared by standard error measurement methods.

The performance of the methods are compared using commonly used error measures, root mean squared error (RMSE), mean relative error (MRE), mean absolute error (MAE) and coefficient of determination (R^2). The error measures, RMSE, MRE and MAE are given by equations (1)-(3), respectively,

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\phi_p - \phi_i)^2} \quad (1)$$

$$\text{MRE} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{(\phi_p - \phi_i)}{\phi_i}\right| \quad (2)$$

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}\left|\phi_p - \phi_i\right| \quad (3)$$

Where,

n = total number of observations,

ϕ_p = predicted value

ϕ_i = actual value of observation

V. RESULTS AND ANALYSIS

As discussed earlier, ANN is applied for 80 years of data (100 % of data).

The precipitation at Akole station is estimated by ANN method. Table (I) shows the actual value of precipitation at Akole station and predicted value of precipitation by ANN method.

TABLE I. COMPARISION OF ACTUAL VALUE AND PREDICTED VALUE OF PRECIPITATION IN MM.

Year	*Actual value of precipitation at Akole Station in July (mm)*	*Predicted Value of precipitation at Akole station in July (mm)*
1972-73	59.50	64.18
1973-74	79.10	80.46
1974-75	35.00	50.03
1975-76	111.00	53.41
1976-77	50.20	78.90
1977-78	181.30	119.54
1978-79	91.00	72.09
1979-80	23.00	29.20
1980-81	185.20	79.14

Table II shows the values of RMSE, MRE, MAE and R^2 are calculated as per Table I.

TABLE II. THE ERRORS CALCULATED AS PER TABLE I.

Sr.No.	*Error type*	*Error by ANN Method*
1	RMSE	46.96160055
2	MRE	0.334054218
3	MAE	6.400728037
4	R2	0.1888748

From the above Table I and Table II, it can be observed that ANN algorithm gives good results.

The graph of actual precipitation value and predicted value by ANN method is shown in Fig.6.

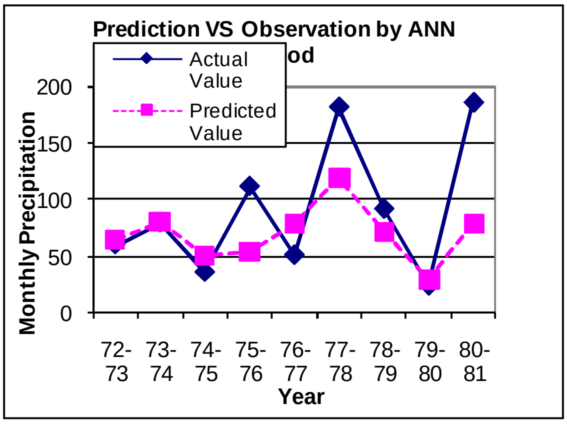

Figure 4. Comparison of actual precipitation and predicted precipitation at
Akole station for July month in mm.

In the Figure 4, it can be observed clearly, that for the year 1972-73, 1973-74, 1974-75, 1976-77, 1978-79 and 1979-80, the predicted value of precipitation is very much closer to the actual value of precipitation.

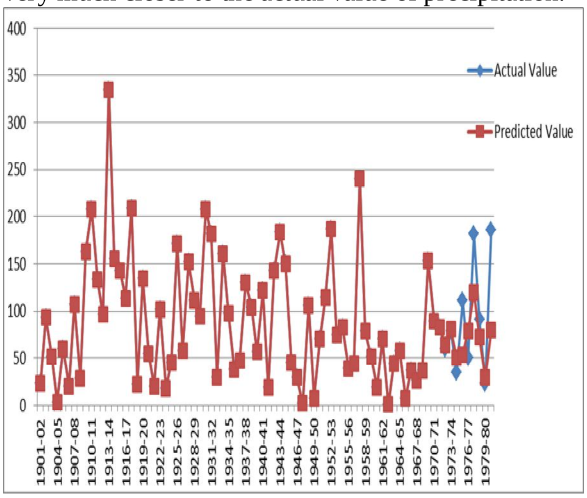

Figure 5. Comparison of actual precipitation and predicted precipitation at

Akole station for July month for 1971-81 in mm.

Figure 5, shows the actual value and predicted value of precipitation from the year 1901-1981 for the month of July month.

VI. CONCLUSIONS

The ANN method is explored in the study and the performance of proposed ANN model is studied and it is observed that the values predicted by the program are reliable to use.

Due to uncertainty of rainfall, the values of RMSE are found towards higher side in ANN. Also the variation in the rainfall is increase year by year for the period of 1972-81 causing increase in the error.

VII. ACKNOWLEDGEMENT

The authors would like to express their sincere thanks to Command Area Development Authority, Nashik (CADA), Maharashtra State, India for providing necessary data for analysis.

REFERENCES

[1] Ramesh S.V. Teegavarapua, V. Chandramouli, “Improved weighting methods, deterministic and stochastic data-driven models for estimation of missing precipitation records”, Journal of Hydrology, Elsevier B.V., 2005, 312 (2005) 191–206.

[2] Ruelland, S.Ardoin-Bardin, G. Billen, E.Servat, “Sensitivity of a lumped and semi-distributed hydrological model to several methods of rainfall interpolation on a large basin in West Africa”, Journal of Hydrology, Elsevier B.V., 2008, 361, 96-117.

[3] Jaap Brand, Stef van Buuren, Erik M. van Mulligen1, Teun Timmers, Edzard Gelsema, “Multiple Imputation as a Missing Data Machine”, AMIA Inc., 1994, 303-306.

[4] P. Coulibaly, N.D.Evora, “Comparision of neural network methods for infilling missing daily weather record”, Journal of Hydrology, Elsevier, 2007, 341, 27-41.

[5] Jan Adamowski, Christina Karapataki, “Comparison of Multivariate Regression and Artificial Neural Networks for Peak Urban Water-Demand Forecasting: Evaluation of Different ANN learning Algorithms”, Journal of Hydrologic Engineering, Vol. 15, No. 10,Pg. 729-743.

◈◈◈