Open access e-Journal
Earth Science India, eISSN: 0974 – 8350
Vol. 4(I), January, 2011, pp.15-20
http://www.earthscienceindia.info/

Application of Multivariate Adaptive Regression Splines to Evaporation Losses in Reservoirs

Pijush Samui[1] and D.P. Kothari[2]
[1]Centre for Disaster Mitigation and Management
[1,2]VIT University, Vellore-632014, India
Email: pijush.phd@gmail.com; advisor@vit.ac.in

Abstract

This study adopts Multivariate Adaptive Regression Splines (MARS) for prediction of Evaporation Losses (E) in reservoirs. MARS is a technique to estimate general functions of high-dimensional arguments given sparse data. The input variables of MARS model are Mean air temperature, Average wind speed, Sunshine hours, and Mean relative humidity. An equation has been developed for prediction of E based on the developed MARS. The results of MARS are compared with the Artificial Neural Network (ANN) model. This study shows that the developed MARS is a robust model for prediction of E in reservoirs.

Key Words: Evaporation Loss; Prediction; Multivariate Adaptive Regression Splines

Introduction

The Evaporation Loss (E) from reservoirs has an impact on climate. Therefore, the determination of E is an imperative task in earth science. The determination of E is a difficult task due to complex interactions among the components of land-plant-atmosphere system (Singh and Xu, 1997). Researchers use different methods for prediction of E in reservoir (Stewart and Rouse, 1976; Bruin, 1978; Anderson and Jobson, 1982; Abtew, 2001; Murthy and Gawande, 2006). Recently, Deswal and Pal (2008) have successfully employed Artificial Neural Network (ANN) for determination of E in reservoir. ANN has been successfully applied by Indian researches (Kanungo *et al.,* 2006; Das and Basudhar, 2006; Kumar and Samui, 2007). The input parameters of the developed ANN model are air temperature (T), wind speed (WS), sunshine hours (SH) and relative humidity (RH). They have used feed-forward back propagation learning algorithm with one hidden layer, momentum = 0.1, learning rate =0.2, hidden layer nodes = 6 and iterations = 1000. The developed ANN has been criticized for its long training process in obtaining the network's topology, not easy to identify the relative importance of potential input variables (Lee and Chen, 2005).

This study investigates the feasibility of Multivariate Adaptive Regression Splines (MARS) for prediction of E in reservoirs. The dataset has been collected from work of Deswal and Pal (2008). The data are collected from a reservoir in Anand Sagar, Shegaon, Maharashtra.

The data of evaporation loss were collected for one year only. Whereas, the other data for a period of fifteen year (from 1990 to 2004) were obtained from a full climatic station at Manasgaon, about 9 Km from Shegaon, lying under water resources division, Amravati Hydrology Project Maharashtra, India. The dataset contains information about E, T (^{0}C),

WS(m/sec), SH(hrs/day), and RH(%). MARS is a flexible, more accurate, and faster simulation method for both regression and classification problems (Friedman, 1991; Salford Systems, 2001). It is capable of fitting complex, nonlinear relationships between output and input variables. The paper has following aims:

- To examine the feasibility of MARS model for prediction of E in reservoirs.
- To determine an equation for prediction of E based on the developed MARS.
- To make a comparative study between MARS and ANN model developed by Deswal and Pal (2008).

Details of MARS

The MARS model splits the data into several splines on an equivalent interval basis (Friedman, 1991). In every spline, MARS splits the data further into many subgroups. Several knots are created by MARS. These knots can be located between different input variables or different intervals in the same input variable, to separate the subgroups. The data of each subgroup are represented by a basis function (BF). The general form of a MARS predictor is as follows:

$$f(x) = \beta_0 + \sum_{j=1}^{P}\sum_{b=1}^{B}\left[\beta_{jb}(+)Max(0, x_j - H_{bj}) + \beta_{jb}(-)Max(0, H_{bj} - x_j)\right] \quad (1)$$

Where x=input, f(x) =output, P= predictor variables and B=basis function. Max (0,x-H) and Max(0,H-x) are BF and do not have to each be present if their coefficients are 0. The H values are called knots. The MARS algorithm consists of (i) a forward stepwise algorithm to select certain spline basis functions, (ii) a backward stepwise algorithm to delete BFs until the "best" set is found, and (iii) a smoothing method which gives the final MARS approximation a certain degree of continuity. BFs are deleted in the order of least contributions using the generalized cross-validation (GCV) criterion (Craven and Wahba, 1979). The GCV criterion is defined in the following way:

$$GCV = \frac{\frac{1}{N}\sum_{i=1}^{N}[y_i - f(x_i)]^2}{\left[1 - \frac{C(B)}{N}\right]^2} \quad (2)$$

Where N is the number of data and C (B) is a complexity penalty that increases with the number of BF in the model and which is defined as:

$$C(B) = (B+1) + dB \quad (3)$$

Where d is a penalty for each BF included into the model. It can be also regarded as a smoothing parameter. Friedman (1991) provided more details about the selection of the d.
This article adopts the above MARS methodology for prediction of E. Table-1 shows the different statistical parameters of input and output variables.

The data has been divided into two sub-sets; a training dataset, to construct the model, and a testing dataset to estimate the model performance. So, for our study a set of 34 data are

Earth Science India, eISSN: 0974 – 8350
Vol. 4(I), January, 2011, pp.15-20
http://www.earthscienceindia.info/

considered as the training dataset and remaining set of 14 data are considered as the testing dataset. The training and testing dataset have been chosen randomly form the original dataset. The program of MARS has been developed by using MATLAB.

Table- 1: Statistical parameters of input and output variables.

Variable	Minimum	Maximum	Range	Mean	Standard Deviation	Skewness	Kurtosis
Evaporation (E)(mm/day)	2.80	15.30	12.50	5.5667	3.0528	1.4565	4.2828
Mean air temperature (T)(^{0}C)	19.64	35.46	15.82	21.1175	4.4931	0.1473	2.1636
Average wind speed(WS) (m/sec)	2.10	11.10	9.00	5.5375	2.8121	0.5129	1.9797
Sunshine hours(SH) (hrs/day)	2.40	11.20	8.80	7.3333	2.4223	-0.6824	2.2231
Mean relative humidity (RH)(%)	36.40	87.65	51.25	62.6708	15.4015	-0.0538	1.8776

Results and Discussion

Coefficient of correlation(R) has been adopted to assess the performance of the developed MARS model. For good model, the value of R is close to one. The value of R has been determined by using the following formula:

$$R = \frac{\sum_{i=1}^{n}\left(E_{ai} - \overline{E}_a\right)\left(E_{pi} - \overline{E}_p\right)}{\sqrt{\sum_{i=1}^{n}\left(E_{ai} - \overline{E}_a\right)}\sqrt{\sum_{i=1}^{n}\left(E_{pi} - \overline{E}_p\right)}} \quad (4)$$

Where E_{ai} and E_{pi} are the actual and predicted E values, respectively, $\overline{E}_a$ and $\overline{E}_p$ are mean of actual and predicted E values corresponding to n patterns. In this study, the value of n is 34 and 14 for training and testing dataset, respectively. First, the forward stepwise procedure was carried out to select 7 basis functions (BF) to build the MARS model. This was followed by the backward stepwise procedure to remove redundant basis functions. The final model includes 4 basis functions, which are listed in Table 2 together with their corresponding equations. The value of GCV is 0.000015. The final equation for the prediction of E based on MARS model is given below:

$$E = 0.191 + 0.919 * BF1 - 1.131 * BF2 - 8.626 * BF3 + 1.141 * BF4 \quad (5)$$

The performance of training and testing dataset has been determined by using the above equation (5). Fig.1 illustrates the performance of training dataset. The performance of testing has been shown in Fig. 2. It is observed form Fig.1 and 2 that the value of R is close to one. Therefore, the developed MARS model has the capability to predict E in reservoir. Fig. 3 shows the performance of training and testing dataset.

Table- 2: Basis functions and their corresponding equations.

Basis Function	Equation
BF1	$\max(0, WS - 0.143)$
BF2	$\text{BF1} * \max(0, RH - 0.233)$
BF3	$\text{BF1} * \max(0, 0.233 - RH)$
BF4	$\max(0, T - 0.429)$

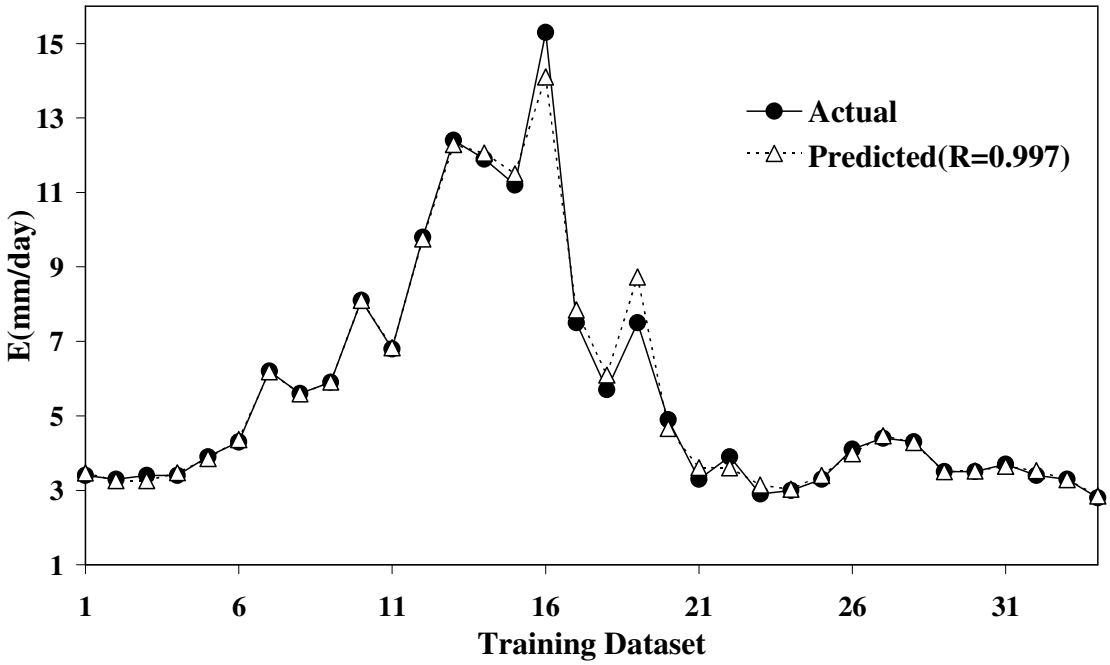

Fig. 1: Performance of training dataset.

A comparative study has been carried out between developed MARS and ANN model developed by Deswal and Pal (2008). Comparison has been done for testing dataset. The value of R and Root Mean Square Error (RMSE) for the ANN model is 0.960 and 0.865 respectively. Whereas, the developed MARS gives R=0.995 and RMSE =0.404. So, the performance of the developed MARS is slightly better than ANN model. MARS explicitly indicates the important inputs and discards the unneeded ones. Whereas ANN still needs to consider these inputs even though they may not produce significant impacts to the final result.

Earth Science India, eISSN: 0974 – 8350
Vol. 4(I), January, 2011, pp.15-20
http://www.earthscienceindia.info/

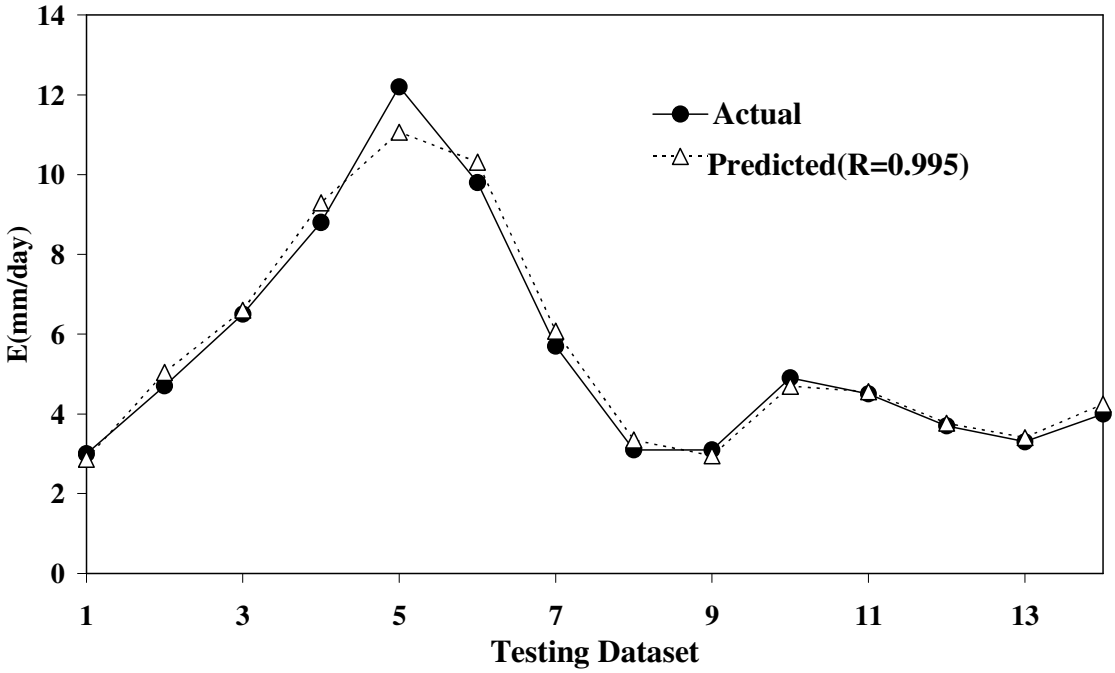

Fig. 2: Performance of testing dataset.

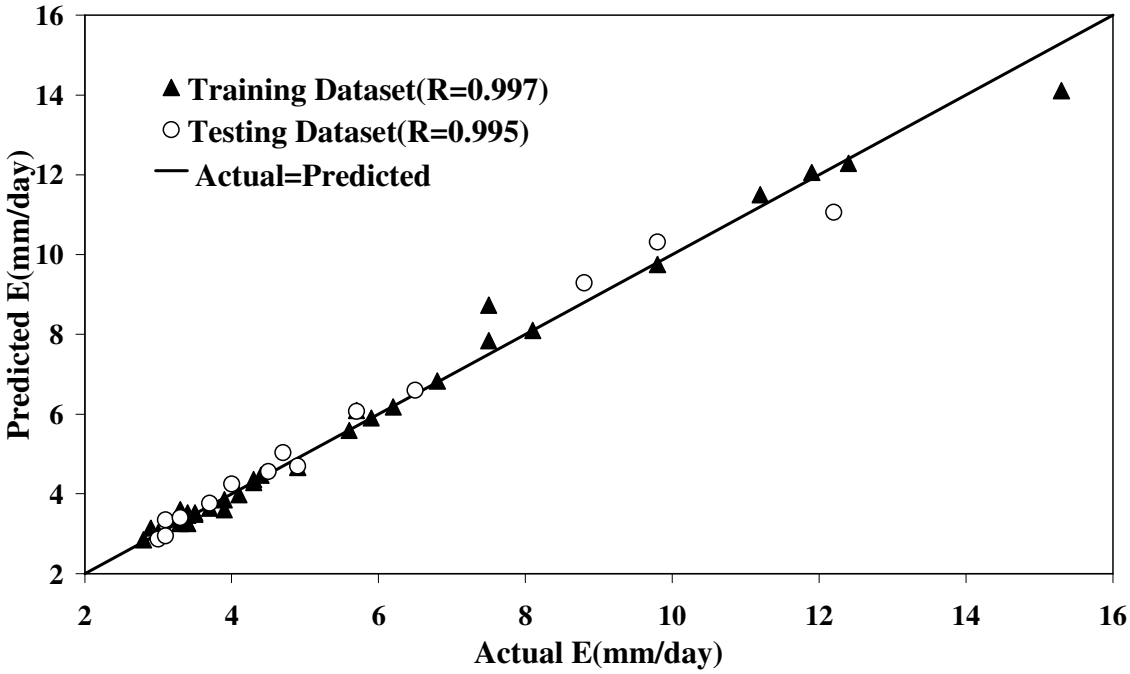

Fig. 3: Actual E versus Predicted E for training and testing dataset.

Conclusion

This study has described MARS for prediction of E. 48 data are used to develop MARS model. MARS uses data not only to calculate the parameter values of some function specified in advance, also the structure of the model is automatically determined. The performance of MARS is encouraging. The performance of MARS is also comparable with ANN. User can use the developed equation for prediction of E in reservoirs. It is concluded that the MARS technique is an effective tool for prediction of E in reservoir.

References

Abtew, W. (2001) Evaporation estimation for Lake Okeechobee in South Florida. Irrigation and Drainage Eng., v.127, pp. 140-147.

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

Bruin, H.A.R.D. (1978) A simple model for shallow lake evaporation. Applied Meterol, v.17, pp.1132-1134.

Craven, P. and Wahba, G. (1979) Smoothing noisy data with spline functions: estimating the correct degree of smoothing by the method of generalized cross-validation. Numer. Math., v. 31, pp.317–403.

Das, S. K. and Basudhar, P. K. (2006) Undrained lateral load capacity of piles in clay using artificial neural network. Computers and Geotechnics, v. 33(8), pp. 454-459.

Deswal, S. and Pal, M. (2008) Artificial Neural Network based Modeling of Evaporation Losses in Reservoirs. Inter. J. Mathematical, Physical and Engineering Sciences, v. 2(4), pp.177-181.

Eaton, E.D. (1958) Control of evaporation losses. United States Government Printing Office, Washington, USA.

Friedman, J. H. (1991) Multivariate adaptive regression splines. Ann Stat, v.19, pp.1–141.

Kanungo, D.P., Arora, M.K., Sarkar, S. and Gupta, R. (2006) A comparative study of conventional, ANN black box, fuzzy and combined neural and fuzzy weighting procedures for landslide susceptibility zonation in Darjeeling Himalayas. Engineering Geol., v. 85, pp. 347-366.

Kumar, B. and Samui, P. (2007) Application of ANN for predicting pore water pressure response in a shake table test. Inter. J. Geotechnical Engineering, v. 2(2), pp. 153-160.

Lee, T.S. and Chen, I.F. (2005) A two-stage hybrid credit scoring model using artificial neural networks and multivariate adaptive regression splines. Expert Systems with Applications, v.28, pp.743–752.

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

Salford Systems (2001) MARSTM User Guide. Salford Systems, San Diego, California, USA.

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

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