Journal of Civil Mechanical Engineering

Volume No. 12

Issue No. 1

January - April 2024

ENRICHED PUBLICATIONS PVT. LTD

S-9, IInd FLOOR, MLU POCKET,
MANISH ABHINAV PLAZA-II, ABOVE FEDERAL BANK,
PLOT NO-5, SECTOR-5, DWARKA, NEW DELHI, INDIA-110075,
PHONE: - + (91)-(11)-47026006

Journal of Civil Mechanical Engineering

Aims and Scope

The journal of civil and mechanical engineering publishes original papers within the broad field of civil mechanical engineering it publishes both theoretical and experimental papers which explore or exploit new ideas and techniques in the following areas: structural engineering (structures, machines and mechanical systems), mechanics of materials (elasticity, plasticity, fatigue, fracture mechanics), materials science (metals, composites, ceramics, plastics, wood, concrete, etc., their structures and properties, methods of evaluation) Theoretical papers, practice-oriented papers including case studies, state-of-the-art reviews are all welcomed and encouraged for the advance of science and technology in civil engineering. , All papers are subject to a referee procedure.

Journal of Civil Mechanical Engineering

Journal of Civil Mechanical Engineering

(Volume No. 12, Issue No. 1, January - April 2024)

Contents

Performance Evaluation Of Extreme Learning Machine For Modeling Evapotranspiration In Arid Regions

Amit Prakash Patil[1] and Paresh Chandra Deka[2]
[12]Department of Applied Mechanics and Hydraulics,
National Institute of Technology Karnataka, Mangalore 572025, India.
Email**:** amitpatil.nitk@gmail.com, pareshdeka@yahoo.com

A B S T R A C T

Managing irrigation systems in arid and semiarid climates is a difficult task owing to the limited availability of water resources and overexploitation of the existing ones. As evapotranspiration plays a vital role in irrigation management, accurate estimation of evapotranspiration becomes evident. Recently, artificial neural network (ANN) is being widely used to model the process of evapotranspiration. However, ANN faces issues like local minima, slow learning and tuning of meta-parameters. To overcome these problems, an improved single layered feed forward neural network algorithm called extreme learning machine (ELM) was used for modeling the process of evapotranspiration. The study was carried out for Jodhpur and Pali weather stations situated in the arid regions of Thar Desert, India. Gamma test was employed to determine inputs that are able to proficiently model the process under limited data availability condition. Primary objective of the study was to evaluate the performance of ELM in modeling the process of evapotranspiration. Further, the results of the ELM model were compared to the well-established ANN and Least-square support vector machine (LS-SVM) models. The ELM model exhibited promising results (testing RMSE of 0.73 mm and 0.72 mm for Jodhpur and Pali respectively) when compared to the ANN model (testing RMSE of 0.76 mm and 0.73 mm for Jodhpur and Pali respectively). The performance of ELM model was at par with the LS-SVM model. Based on the result it can be concluded that ELM is more expeditious tool to estimate evapotranspiration.

Keywords: Evapotranspiration; Limited data; Extreme Learning Machine; Arid region; least square support vector machine.

1. INTRODUCTION

Scarcity of water and growing need for food supplies emphasize on developing improved methods for crop-water estimation. As evapotranspiration (ET) plays a vital role in determining crop-water requirement, accurate estimation of evapotranspiration becomes evident. Besides, precise estimation of evapotranspiration is of great importance to many disciplines like hydrological studies, agriculture, meteorology and drainage studies. Considering the significance of accuracy in estimating evapotranspiration, hydrologists have mainly focused on developing more reliable and accurate methods for estimating evapotranspiration.

Evapotranspiration is a very complex process which depends on the interaction of various atmospheric, plant and soil parameters. Generally, lysimeters are used for direct measurement of ET, but high

operating costs and need for accuracy in measurements has limited the use of lysimeters (López-Urrea et al., 2006). Over the years, hydrologists have developed numerous physical, empirical and semi-empirical equations that used meteorological variables to estimate reference crop evapotranspiration (ETo). The Food and Agricultural Organization of United Nations (FAO) has accepted the FAO Penman-Monteith (FAO-56PM) as the standard equation to estimate ETo (Allen et al., 1998). The FAO-56PM is a physical based equation which offers best results in estimating evapotranspiration of living grass reference crop. Application of the FAO-56PM equation requires various meteorological variables like air temperature, solar radiation, humidity and wind speed. In developing countries like India, the network of weather stations capable of measuring all these parameters is sparse. However, small weather stations capable of measuring only few parameters like air temperature are ample in number. Therefore, there is an immediate need to develop a proficient system capable of estimating ETo using the limited climatic data available.

Recently, application of machine learning (ML) techniques (e.g., artificial neural network, adaptive neuro-fuzzy inference system and support vector machines) in modelling hydrological processes like evapotranspiration have received much attention from the researchers (Aksoy et al., 2007; Partal and Kişi, 2007; Rahimi Khoob, 2007). Machine learning algorithms provide explanation of an externally driven processes without a need of complex physical models. Ease of experimentation, simple yet fast in the training and testing phases and low computational burden are some advantages of using ML techniques. Kumar et al., (2002) used ANN for the estimation of ETo. Sudheer et al., (2003) employed radial-basis function (RBF) type ANN for computing the daily values of ET for rice crop. Zanetti et al., (2007) simplified the input variables used for the ANN and estimated ETo as a function of extra-terrestrial solar radiation, air temperature and sunshine hours. Aytek et al., (2008) proposed explicit neural network formulation for estimating reference evapotranspiration. The results were compared to five conventional ETo equations and a linear regression model. Chauhan and Shrivastava, (2008) and Rahimikhoob, (2010) used only maximum and minimum temperature datasets to estimate ETo. Kumar et al., (2008) compared the performance of different ANN models for arid and humid climates. Landeras et al., (2008) used various input combinations of meteorological variables for estimating evapotranspiration using ANN models. The performance of ANN models were compared to locally calibrated ETo models. Kumar et al., (2009) proposed generalized artificial neural network model for estimating ETo. Kisi and Cimen, (2009) studied the potential of support vector machine (SVM) in modelling ETo for central California. Kisi, (2009) compared the performance of two different ANN models. (Martí, González-Altozano, & Gasque, 2010) proposed ANN models with exogenous inputs. The models performed better than the existing temperature based models, which considered only local temperature data. Tabari et al., (2012) evaluated the performance of SVM, adaptive neuro-fuzzy

inference system, regression and climate based models for modelling evapotranspiration. Kisi, (2013) employed least square support vector machines (LS-SVM) to estimate daily ETo values. Wen et al., (2015) employed SVM to model ETo with limited climatic data in arid regions.

Generally, back-propagation (BP) algorithms are used to train ANN models. The BP algorithm needs iterative tuning to obtain optimal model parameters and may face issues like trapping in local minima, and are time consuming in learning. Additionally, building of an ANN model involves specification of several parameters like transfer function, learning rate, number of hidden layers and number of nodes in the hidden layer. Furthermore, the commonly used ANN applications treat these parameters as user defined functions. A non-expert user generally uses a trial and error method to set these parameters. Trial and error method does not always result into an optimum setting and may lead to low prediction accuracies despite using a good algorithm. Therefore, there is an immediate need to address these problems. In this study an improved single layered feed forward neural network (SLFN) algorithm called extreme learning machines (ELM) is adopted to estimate weekly ETo. ELM algorithm does not need prior tuning of meta-parameters like input weights and hidden layer biases (Guang et al., 2012). This distinguishes it from the traditional neural network methodology. The ELM algorithm has exhibited promising results in some recent studies (Huang et al., 2011; Wang et al., 2011). Şahin et al., (2014) compared the performance of ELM to ANN model for estimation of solar radiation. The comparison showed that the ELM model gave better estimation than ANN model. Acharya et al., (2014) used ELM for estimating northeast monsoon rainfall over south peninsular India. Deo and Şahin, (2015) employed ELM for prediction of effective drought index. Based on the results they concluded that ELM was an expeditious tool for prediction of drought.

This paper attempts to model the process of evapotranspiration in arid regions of India under limited data scenario. The primary objective of the study is to evaluate the capabilities of ELM to model the process of evapotranspiration. Further, the results of the ELM model are compared to the well-established ANN and LS-SVM models. The secondary objective of the study is associated with input selection. Input parameters play a vital role in the performance of any ML based model. This study makes an attempt to study the effectiveness of using Gamma test for selection of most influential inputs that can be used to estimate evapotranspiration under limited data scenario.

2 BASICS OF ELM, ANN AND LS-SVM

2.1 Extreme learning Machines (ELM)

The algorithms like BP use some rules to adjust the weights based on the given batch of training examples. On the other hand, weights are chosen randomly in ELM. Tanmura, (1997) and Huanga, (2003) found that SLFNs with randomly adopted input weights can efficiently learn distinct training examples with minimum error. On choosing the input weights and hidden layer biases the SLFN can be considered as a linear system and the output weights analytically determined by simple generalized inverse operation of the hidden layer output matrices. This simplified approach makes ELM work faster than the feed forward algorithm. The basic theory of ELM can be given as follows;

For M arbitrary distinct inputs (x_i, y_i) with $x_i \in \mathbb{R}^d$ and $y_i \in \mathbb{R}$; a standard SLFN with N hidden nodes and activation function f can be modeled as the following sum

$$\sum_{i=1}^{\mathcal{N}} \beta_i \mathrm{f}(w_i x_j + b_i), j \in \{1,2,3,\dots,M\}, \quad (1)$$

where w_i are the input weights to the i^{th} neuron in the hidden layer, b_i the biases and β_i are the output weights.

In the case where the SLFN would perfectly approximate the data, the relation is

$$\sum_{i=1}^{\mathcal{N}} \beta_i \mathrm{f}(w_i x_j + b_i) = y_j, j \in \{1,2,3,\dots,M\}, \quad (2)$$

which can compactly be written as,

$$\boldsymbol{H}\beta = \boldsymbol{Y} \quad (3)$$

where, $\boldsymbol{H}$ the hidden layer output matrix is defined as,

$$\boldsymbol{H} = \begin{pmatrix} \mathrm{f}(w_1 x_1 + b_1) & \cdots & \mathrm{f}(w_n x_1 + b_N) \\ \vdots & \ddots & \vdots \\ \mathrm{f}(w_1 x_M + b_1) & \cdots & \mathrm{f}(w_N x_M + b_N) \end{pmatrix} \quad (4)$$

where $\beta = (\beta_1 \dots \beta_N)^T$ and $Y = (y_1 \dots y_M)^T$.

Considering the randomly initialized first layer of the ELM and the training inputs, the hidden layer hidden layer output matrix $\boldsymbol{H}$ can be computed. Given $\boldsymbol{H}$ and the target outputs, output weight β can be solved by finding the least square solution to the linear system defined by Eq. 4. This solution is given by $\beta = \boldsymbol{H}^{\dagger}\boldsymbol{Y}$, where $\boldsymbol{H}^{\dagger}$ is the Moore-Penrose generalized inverse of the matrix $\boldsymbol{H}$. More details on the ELM algorithm can be found in the original paper (Huang et al., 2006).

2.2 Artificial neural network (ANN)

ANN an information processing system, stimulates the ability of a human brain to sort out patterns and learn from trial and error. It has the ability to extract relationships that exist within the data. Typically, ANN architecture consists of a series of processing elements called neurons. Neurons are arranged in layers, namely; input layer, output layer and one or more hidden layer. Each layer is fully connected to the next layer by interconnection weights. In the training process of most commonly BP algorithm, the estimated outputs are first compared to the known outputs, then the errors occurred are back propagated to obtain appropriate weight adjustments necessary for minimizing the errors.

2.3 Least-square support vector machine (LS-SVM)

Support vector machines (SVM) are a robust and efficient algorithm for classification and regression. SVMs based on structural risk minimization hypothesis for minimizing empirical risk and confidence interval of learning machine are able to achieve good generalization capabilities. Basically SVMs maps the original data set from the input space to a high dimensional feature resulting into a simpler feature space. SVM comes with an advantage of using kernel trick to minimize both model complexities and prediction errors simultaneously. Least square support vector machine (LS-SVM) provides fast implementation of the traditional SVM.

2.4 Gamma test

Selection of proper inputs plays a vital role in improving the efficiency of any ML model. It is expected that one should select explicit inputs for extracting an accurate model out of the available database. A non-linear modeling and analysis technique called gamma test (GT) (Stefánsson and Jones, 1997) can be used to evaluate the efficiency of different input combinations in modeling the output function. GT allows us to measure the extent to which a smooth relationship can be established between input and output without relying on information about a specific machine-learning model (Tsui et al., 2002). Gamma test scores give an indication of the unaccountable variance that exists between input and output datasets. It is a very useful statistic for comparing performances of different input variables in modelling the desired output. As the objective of this paper is to evaluate the performance of ML models for estimating ETo under limited data scenario, gamma test was used to determine the input variable that is capable of modeling the process of evapotranspiration more efficiently.

3 MATERIALS AND METHODS

3.1 Climate data

In this study, weekly climatic data from Jodhpur (26°28'N latitude and 73°02'E longitude) and Pali (25°77'N latitude and 73°33' E longitude) weather stations were used. The data were obtained from the Central Arid Zone Research Institute. Both the stations are located in The Thar Desert which is classified as Arid Region (BW) according to the Koppen climate classification. The weekly weather data used in this study were maximum air temperature (Tmax), minimum air temperature (Tmin), maximum relative humidity (RHmin), minimum relative humidity (RHmin), solar radiation (Rs) and wind speed measured at 2 meters height above ground (U2). The data samples, covers forty (1970-2010)

records of aforementioned climatic parameters. All the models were trained using first thirty five years (1970 to 2005) data while the remaining five years (2006 to 2010) data were used for testing. Computation for all the data necessary to calculate weekly ETo was done according to the procedure prescribed in Chapter 3 of FAO-56.

3.2 Model development

In this study, the input variables were selected based on the Gamma test scores of different input combinations. Table 1 shows Gamma test results for Jodhpur weather station.

Table 1 Gamma scores for Jodhpur station

Input variables	Gamma Scores		
	Entire set	Training set	Testing set
T	0.056	0.056	0.050
RH	0.197	0.192	0.193
U_2	0.125	0.132	0.106
n/N	0.082	0.095	0.087

GT scores for each input was calculated separately and the climatic parameter that provided the least GT score was assumed as the most competent input for estimating ETo under limited data scenario. The variables tested were namely maximum and minimum temperature (T), maximum and minimum relative humidity (RH), wind speed at 2 meter elevation (U2), the relative sunshine duration (n/N). From the Gamma test results (Table 1) it was found that temperature (GT score=0. 056) explained most amount of variance in the evapotranspiration process of an arid region. Whereas, relative humidity (GT score=0.197) explained the least amount of variance in the output function. It was also observed that inspite of testing different lengths of datasets; the GT scores for all the three datasets were quite similar. After analyzing the gamma test results, it was decided to use maximum temperature (Tmax), minimum temperature (Tmin) and extraterrestrial radiation (Ra) as inputs to model the process of evapotranspiration under limited data scenario.

A three layered architecture was adopted for ELM model development. The first layer (input layer) used meteorological parameters as inputs. The output layer had one neuron representing the estimated weekly ETo. For the hidden layer a maximum of 200 neurons were tested for each model. For determining the optimum number of neurons in the hidden layer, initially 10 neurons were tested and subsequently the number of neurons was gradually increased to 200 by an interval of ten. Radial basis activation function was employed for all the ELM models tested.

ANN has been widely used for modelling the process of ET. In this study a feed forward backpropagation network with Levenberg-Marquet (LM) algorithm for weight optimization was used. Different ANN architectures for modelling daily evapotranspiration were tested by varying the number of neurons in the hidden layer. A trial and error procedure was adopted to find the optimum number neurons in the hidden layer. The performance of the various sigmoid activation functions (hyperbolic-tangent sigmoid and log-sigmoid) in the hidden layer with linear activation function at the output node was also evaluated. As data normalization provides for initial weight allocation according to distribution and not the magnitude of data, the inputs and outputs were normalized using min-max normalization.

LS-SVM uses equality optimization constrains instead of inequalities constrains used in the traditional SVM. Equality optimization results in a direct least square solution by avoiding quadratic programming. Choice of kernel functions and hyper-parameters are some critical issues needed to be addressed before the application of LS-SVM. Radial basis function (RBF) a more compactly supported kernel function is able to reduce the computational complexity of the training process and provides a good performance. Hence, RBF kernel function was employed in this study. Different techniques for tuning of the hyper-parameters related to the regularization constant are available in the literature. In this study the regularization parameter gamma (γ) and kernel function parameter (σ^2) were obtained by grid search technique based on leave-one-out cross validation.

3.3 Model performance criteria

In this study root mean square error (RMSE) was used to evaluate the performance of models. RMSE represents the standard deviation of differences between actual values and values predicted by the model. In order to test the robustness of the developed model it is also necessary to test the model using some other performance evaluation indicators like Nash-Sutcliffe model efficiency coefficient (NSE) and threshold statistics (TS). In the present study TS for absolute relative error levels of 5, 10 and 15 percent (TS5, TS10 and TS15) were used to measure the effectiveness of the models regarding their ability to accurately predict data from the calibrated model. The Nash-Sutcliffe coefficient was used to access efficiency of the models. NSE for a model can range from - to 1. An efficiency of 1 (NSE=1) corresponds to a perfect match between estimate and observations. Scatter plots were also used to evaluate the accuracies of the models.

4. RESULTS AND DISCUSSION

This study focused on comparing the performance of ELM models with ANN, LS-SVM and Hargreaves empirical model for estimating weekly evapotranspiration. The ELM model was compared to other models in respect of RMSE and R statistics for Jodhpur Pali station in Table 2 and Table 3 respectively. In Table 2 and Table 3 the model parameters for ANN and ELM model represent the number of neurons in the input, hidden and output layers, whereas, for the LS-SVM model the model parameters represent γ an ^{2} values.

Table 2 Performance of ANN, LS-SVM and ELM models for Jodhpur station

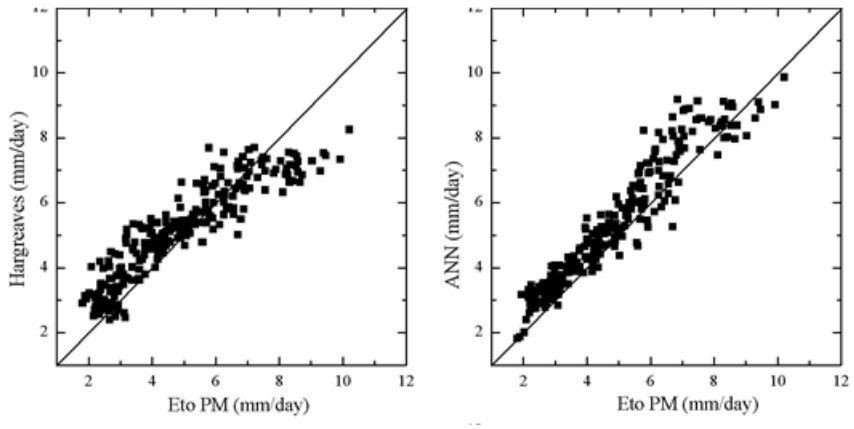
Model	Model parameters	RMSE (mm/day)	NSE	Ts5	Ts10	Ts15
Hargreaves	--	0.90	0.79	20.8	37.7	53.8
ANN1	3, 3, 1	0.76	0.85	21.9	40.4	57.7
LS-SVM1	154, 1.7	0.76	0.85	20.4	40.0	57.3
ELM1	3, 20, 1	0.73	0.86	21.9	42.7	59.2

From the results it was observed that all the artificial intelligence based model performed better than the conventionally used Hargreaves equation. It was also seen that the performance of ELM model was better than the ANN and LS-SVM model for both Jodhpur and Pali station. Conversely, the threshold statistics show a more or less similar performance of all the models. From table 2 it can be observed that the ELM model with twenty neurons in the hidden layer was the best model to estimate evapotranspiration at Jodhpur station. The Hargreaves equation (RMSE= 0.9) displayed a NSE of 0.79 while the ML based models showed relatively higher NSE values. Ts statistic clearly indicates the superiority of ELM model for estimating evapotranspiration at Jodhpur station.

Table 3 Performance of ANN, LS-SVM and ELM models for Pali station

Model	Model parameters	RMSE (mm/day)	NSE	Ts5	Ts10	Ts15
Hargreaves	--	0.99	0.81	23.8	41.9	61.5
ANN1	3, 5, 1	0.73	0.89	24.23	48.85	68.85
LS-SVM1	188, 1.6	0.72	0.90	21.54	51.92	70.00
ELM1	3, 30, 1	0.72	0.90	30.38	52.31	72.31

All the model performed better at Pali station compared to Jodhpur station. At the Pali station performance of ELM and LS-SVM were similar but the Ts statistics show ELM model performing slightly better than the LS-SVM model. The Ts statistics show that for ELM model 72.31 percent of the estimated ETo values are less than 15 percent relative error, displaying the efficiency of ELM model. Figure 1 show scatter plots for all the models at Jodhpur station.



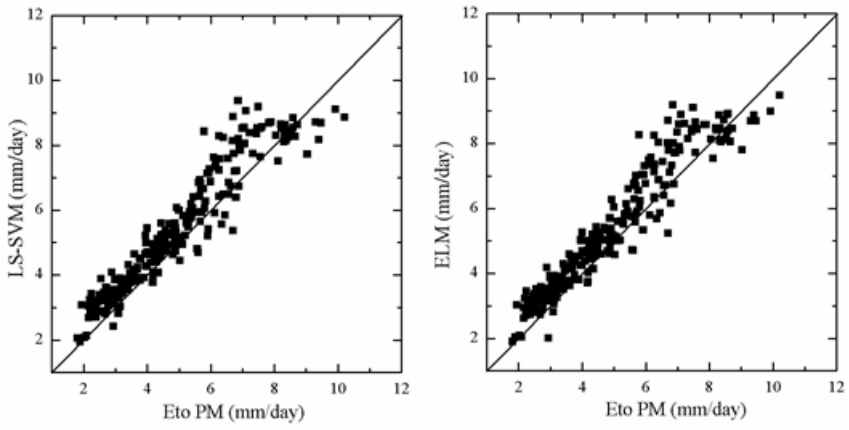


Figure 1 Scatter plot of Hargreaves, ANN, LS-SVM and ELM models in testing phase for Jodhpur.

The scatter plot shows that the Hargreaves equation overestimated the lower ETo values but overestimated the higher ones. However, all the ML models overestimated the ETo values. The amount of overestimation was small for lower values while significant overestimation was observed for the higher ETo values. It is also evident from the scatter plot that the performance LS-SVM and ELM was better than the ANN model. Similar results were also observed for the Pali station.

4 . CONCLUSION

In this study an attempt was made to evaluate the performance of Extreme Learning Machine to model the process of evapotranspiration. The performance of the ELM model was further compared to the conventional Hargreaves equation and commonly used ML models like ANN and LS-SVM. From the results it can be concluded that the ELM model performed better than the Hargreaves and ANN model. The ELM model with thirty neurons in the hidden layer was found to be the best model to estimate weekly evapotranspiration. The performance of ELM and LS-SVM model was alike, but considering the fact that ELM requires less human intervention for selecting optimum model parameters, ELM can be considered as a practical option to model evapotranspiration. Additionally the simplicity in the application of ELM algorithm makes it much faster than the ANN and SVM models.

REFERENCES

Acharya, N., Shrivastava, N. A., Panigrahi, B. K., & Mohanty, U. C. (2014). Development of an artificial neural network based multi-model ensemble to estimate the northeast monsoon rainfall over south peninsular India: an application of extreme learning machine. Climate Dynamics, 43(5-6), 1303–1310.

Aksoy, H., Guven, A., Aytek, A., Yuce, M. I., & Unal, N. E. (2007). Discussion of "Generalized regression neural networks for evapotranspiration modelling." Hydrological Sciences Journal, 52(4), 825–831.

Allen, R., Pereira, L., Raes, D., & Smith, M. (1998). Crop evapotranspiration-Guidelines for computing crop water requirements-FAO Irrigation and drainage paper 56. Food and Agriculture Organization of the United Nations, Rome.

Aytek, A., Guven, A., Yuce, M. I., & Aksoy, H. (2008). An explicit neural network formulation for evapotranspiration. Hydrological Sciences Journal, 53(4), 893–904.

Chauhan, S., & Shrivastava, R. K. (2008). Performance Evaluation of Reference Evapotranspiration Estimation Using Climate Based Methods and Artificial Neural Networks. Water Resources Management, 23(5), 825–837.

Deo, R. C., & Şahin, M. (2015). Application of the extreme learning machine algorithm for the prediction of monthly Effective Drought Index in eastern Australia. Atmospheric Research, 153, 512–525.

Guang-Bin Huang, Hongming Zhou, Xiaojian Ding, & Rui Zhang. (2012). Extreme Learning Machine for Regression and Multiclass Classification. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), 42(2), 513–529.

Huang, G.-B., Wang, D. H., & Lan, Y. (2011). Extreme learning machines: a survey. International Journal of Machine Learning and Cybernetics, 2(2), 107–122.

Huang, G.-B., Zhu, Q.-Y., & Siew, C.-K. (2006). Extreme learning machine: Theory and applications. Neurocomputing, 70(1-3), 489–501.

Kisi, O. (2013). Least squares support vector machine for modeling daily reference evapotranspiration. Irrigation Science, 31(4), 611–619.

Kişi, Ö. (2009). Modeling monthly evaporation using two different neural computing techniques. Irrigation Science, 27(5), 417–430.

Kişi, O., & ÇIMEN, M. (2009). Evapotranspiration modelling using support vector machines Hydrological Sciences Journal, 54(5), 918–928.

Kumar, M., Bandyopadhyay, A., Raghuwanshi, N. S., & Singh, R. (2008). Comparative study of conventional and artificial neural network-based ETo estimation models. Irrigation Science, 26(6), 531–545.

Kumar, M., Raghuwanshi, N. S., & Singh, R. (2009). Development and Validation of GANN Model for Evapotranspiration Estimation. Journal of Hydrologic Engineering, 14(2), 131–140.

Kumar, M., Raghuwanshi, N. S., Singh, R., Wallender, W. W., & Pruitt, W. O. (2002). Estimating Evapotranspiration using Artificial Neural Network, (August), 224–233.

Landeras, G., Ortiz-Barredo, A., & López, J. J. (2008). Comparison of artificial neural network models and empirical and semi-empirical equations for daily reference evapotranspiration estimation in the Basque Country (Northern Spain). Agricultural Water Management, 95(5), 553–565.

López-Urrea, R., Martín de Santa Olalla, F., Fabeiro, C., & Moratalla, a. (2006). Testing evapotranspiration equations using lysimeter observations in a semiarid climate. Agricultural Water Management, 85(1-2), 15–26.

Martí, P., González-Altozano, P., & Gasque, M. (2010). Reference evapotranspiration estimation without local climatic data. Irrigation Science, 29(6), 479–495.

Partal, T., & Kişi, Ö. (2007). Wavelet and neuro-fuzzy conjunction model for precipitation forecasting. Journal of Hydrology, 342(1-2), 199–212.

Rahimi Khoob, A. (2007). Comparative study of Hargreaves's and artificial neural network's methodologies in estimating reference evapotranspiration in a semiarid environment. Irrigation Science, 26(3), 253–259.

Rahimikhoob, A. (2010). Estimation of evapotranspiration based on only air temperature data using artificial neural networks for a subtropical climate in Iran. Theoretical and Applied Climatology, 101(1-2), 83–91.

Şahin, M., Kaya, Y., Uyar, M., & Yildirim, S. (2014). Application of extreme learning machine for estimating solar radiation from satellite data. International Journal of Energy Research, 38(2), 205–212.

Stefánsson, A., Končar, N., & Jones, A. J. (1997). Computing & Applications A Note on the G a m m a Test. Neural Computing & Applications, 5(3), 131–133.

Sudheer, K. P., Gosain, A. K., & Ramasastri, K. S. (2003). Estimating Actual Evapotranspiration from Limited Climatic Data Using Neural Computing Technique. Journal of Irrigation and Drainage Engineering, 129(3), 214–218.

Tabari, H., Kisi, O., Ezani, A., & Hosseinzadeh Talaee, P. (2012). SVM, ANFIS, regression and climate based models for reference evapotranspiration modeling using limited climatic data in a semi-arid highland environment. Journal of Hydrology, 444-445, 78–89.

Tsui, a. P. M., Jones, a. J., & Guedes de Oliveira, A. (2002). The Construction of Smooth Models using Irregular Embeddings Determined by a Gamma Test Analysis. Neural Computing & Applications, 10(4), 318–329.

Wang, Y., Cao, F., & Yuan, Y. (2011). A study on effectiveness of extreme learning machine. Neurocomputing, 74(16), 2483–2490.

Wen, X., Si, J., He, Z., Wu, J., Shao, H., & Yu, H. (2015). Support-Vector-Machine-Based Models for Modeling Daily Reference Evapotranspiration With Limited Climatic Data in Extreme Arid Regions. Water Resources Management, 29(9), 3195–3209.

Zanetti, S. S., Sousa, E. F., Oliveira, V. P., Almeida, F. T., & Bernardo, S. (2007). Estimating Evapotranspiration Using Artificial Neural Network and Minimum Climatological Data. Journal of Irrigation and Drainage Engineering, 133(2), 83–89.

SWAT Model Assessment of Hydrological Impacts of HRU-scale Invasion by Parthenium Hysterophorus Weed

Soham Adla[1], Shivam Tripathi[2]
[1]Department of Civil Engineering, Indian Institute of Technology Kanpur, Kanpur – 208016, India
[2]Department of Civil Engineering, Indian Institute of Technology Kanpur, Kanpur – 208016, India
Email**: sohamadla@gmail.com**
Telephone/Mobile No.: +91 9559753681

A B S T R A C T

Parthenium hysterophorus is a notorious invasive alien species with an increasing presence in India attributed to its high growth rate due to its ability to compete against many native species and adaptability to varied environmental conditions. Despite extensive studies on its impacts on human health, livestock, agricultural productivity, and infestation mitigation techniques, there have been few studies highlighting its hydrological impact on components of the water balance at any scale. This study used the physically based SWAT (Soil and Water Assessment Tool) model to simulate Parthenium invasions at the scale of a hydrologic response unit (HRU) – the smallest user defined homogenous unit of computation within the SWAT model. The primarily agricultural basin of the Punpun river in Bihar was chosen as the study area, fallow land and kharif crop HRUs were infested with Parthenium, and results were analysed across soil type and slope classes. SWAT modeled higher (but statistically insignificant) levels of ET losses for Parthenium compared to both the native LUs, subject to soil moisture availability which was a limiting factor for ET during the dry season. Average monthly soil moisture levels were correspondingly and consistently slightly lower for Parthenium contradicting heuristic experience of local water managers. The study highlights the need for localized measurement of crop parameters to resolve such contradictions.

Keywords: Parthenium Hysterophorus; SWAT; HRU; water balance; invasive alien species

1. INTRODUCTION

Parthenium Hysterophorus is an invasive alien plant species in India. It has invaded about 2 million hectares of agricultural land, mainly in Punjab, Haryana and Uttar Pradesh (Dwivedi et al. 2009) because of its high germination capacity, ability to withstand and grow in diverse climatic and soil conditions, and allelopathic ability to restrict the growth of native plant species. Its spread may result in impacts on human health on exposure (Patel 2011), decrease in crop productivity and harmful impacts on livestock and biodiversity (Kumari 2014).

1.1 Motivation

The increasing spread of Parthenium Hysterophorus weed in the fertile crop-producing plains of North India, particularly in the Ganga basin, is critical not only from the perspective of its impacts on human and animal health and inhibitory effects on crop productivity, but also in terms of potentially significant impacts on the water resources of the infested catchment, by influencing the water balance components.

A study by Adla and Tripathi (2014) used the Soil and Water Assessment Tool (SWAT) model (Arnold et al. 1998) to investigate the basin scale hydrological effects of an invasion by Parthenium Hysterophorus. The land use of the primarily agricultural Punpun river basin (Bihar) was modified under the framework of spatially random scenarios of incremental Parthenium cover, representative of increasing extents of a hypothetical invasion. Results indicated that the while the presence of Parthenium did not alter the annual water balance significantly, an increasing land cover reduced evapotranspiration (ET) losses which subsequently led to higher soil moisture buildup before the onset of the monsoon. The study, however, had a few limitations.

The plant parameters of Parthenium used in the SWAT model were derived from a combination of existing literature (Pandey et. al 2003) and in-situ measurements which implied that the parameter estimates were not internally consistent, and did not follow the standards for SWAT model plant parameter estimation. It was assumed that the kharif crop was rain-fed, and hence irrigation inputs were excluded from the model setup. On request, the USDA-ARS (US Department of Agriculture – Agricultural Research Service) has developed an official set of Parthenium plant parameters now included within the SWAT model by default. In this study, irrigation has been incorporated realistically as a management practice during the kharif cropping season. However, the major conceptual limitation of the previous study was that a model had been constructed to simulate a river basin scale invasion by Parthenium. Since regular agricultural management practices by farmers (applying weedicides, manual uprooting etc.) preclude any fully basin scale invasions of agricultural land by most invasive species, the study of Adla and Tripathi (2014) was not representative of actual conditions on the ground, but served to provide an overall perspective of the hydrological impacts of such an extensive land use change at the intra-annual time scale. Also, such an approach implied challenges in isolating the particular processes through which this land use change may have causally impacted water balance variables. The framework of the land use change deployed in the present study can potentially circumvent this limitation and isolate the drivers of changes in the water balance, by simulating Parthenium invasions at the scale of a hydrologic response unit (HRU) – the smallest user defined homogenous unit of computation within the SWAT model. As invasion at the scale of HRU is representative of the scale of an actual invasion, the use of the LU-split option in SWAT can establish a common ground for comparing the native and invaded land covers and their hydrological effects.

1.2 Objectives

The objective of the study was to simulate invasions by Parthenium Hysterophorus at management-appropriate scales of HRUs, the basic computational units in the SWAT model. Each HRU had a unique combination of native land cover, soil type and slope category. A subsequent partial invasion of each such HRU would then help to detect the marginal effect of a Parthenium invasion. The study aimed at fulfilling the following objectives:

1. To construct a representative hydrological model of the Punpun river basin by using the SWAT model.
2. To simulate invasions by Parthenium Hysterophorus at the HRU-scale
3. To quantify and explain the impacts of invasion by Parthenium Hysterophorus on the water balance components at annual and intra-annual levels.

2. MATERIALS AND METHODS

2.1 The SWAT Model

The Soil and Water Assessment Tool (SWAT) Model (Arnold et al. 1998) is a semi-distributed river basin model which simulates water, sediment, nutrient and point-source pollution yields at a daily time step (Gassman et al. 2007). It was developed through by USDA to assist water resource managers in assessing impacts of land-use management on water and diffuse pollution for large ungauged catchments with different soil types, land uses and management practices (Arnold and Fohrer 2005).

The SWAT model framework divides the catchment into multiple sub-basins, each of which is further subdivided into hydrologic response units (HRUs). An HRU is the smallest unit of computation in the SWAT model. It does not have any spatial reference (the sub-basin is the smallest unit with spatial meaning) and has a unique combination of land use, soil and slope class characteristics, which can be modified by user inputs. The SWAT model computes the daily water balance for each HRU in the model according to Equation 1:

$$\Delta(Storage) = \Delta(Snow) + \Delta(SW) + \Delta(Sh.Aq) + \Delta(Dp.Aq) + \Delta(Surf\_daylag) + \Delta(Lat\_daylag) \quad (1)$$

where Storage is the sum of all storage terms, Snow is the amount of water stored as snow, SW is the amount of water stored in the soil profile on a given day, Sh.Aq is the depth of water in the soil aquifer, Dp.Aq is the depth of water in the deep aquifer, Surf_daylag is the amount of surface runoff lagged over a day, Lat_daylag is the amount of lagged lateral flow, and Δ represents the change in each of the terms over a daily time step. The water losses are computed according to Equation 2:

$$Waterloss = Pcp_{day} + Irr_{day} - Surf_{day} - Lat_{day} - ET_{day} - GWQ_{day} - Rvap_{day} - Rchrg_{day} - Seep_{day} - Tloss \quad (2)$$

where Waterloss is the net movement of water out of the HRU, Pcpday is the precipitation, Irrday is the irrigation water application, Surfday is the surface runoff loading to the main channel, Latday is the lateral flow, GWQday is the groundwater contribution to streamflow, Rvapday is the amount of water moving from shallow aquifer to the soil profile or absorbed by plant roots in the shallow aquifer, Rchrgday is the amount of water recharging both aquifers, Seepday is the seepage leaving the bottom of the soil profile and Tloss are the transmission losses in surface runoff, all computed at the scale of an HRU for a particular day.

2.2 *Study Area: Punpun River Basin*

The study area (Figure 1) chosen was the Punpun river basin, in southern Bihar (India). The modeled area of the basin was 5495.51 km^2 and the outlet was chosen as the Central Water Commission gauging

site at Sripalpur (25°30′6″N, 85°6′8 ″E) in Patna district. The choice of the area was in part due to the fact that its land use is majorly agricultural (~74%) which is convenient for land use representation in the model.

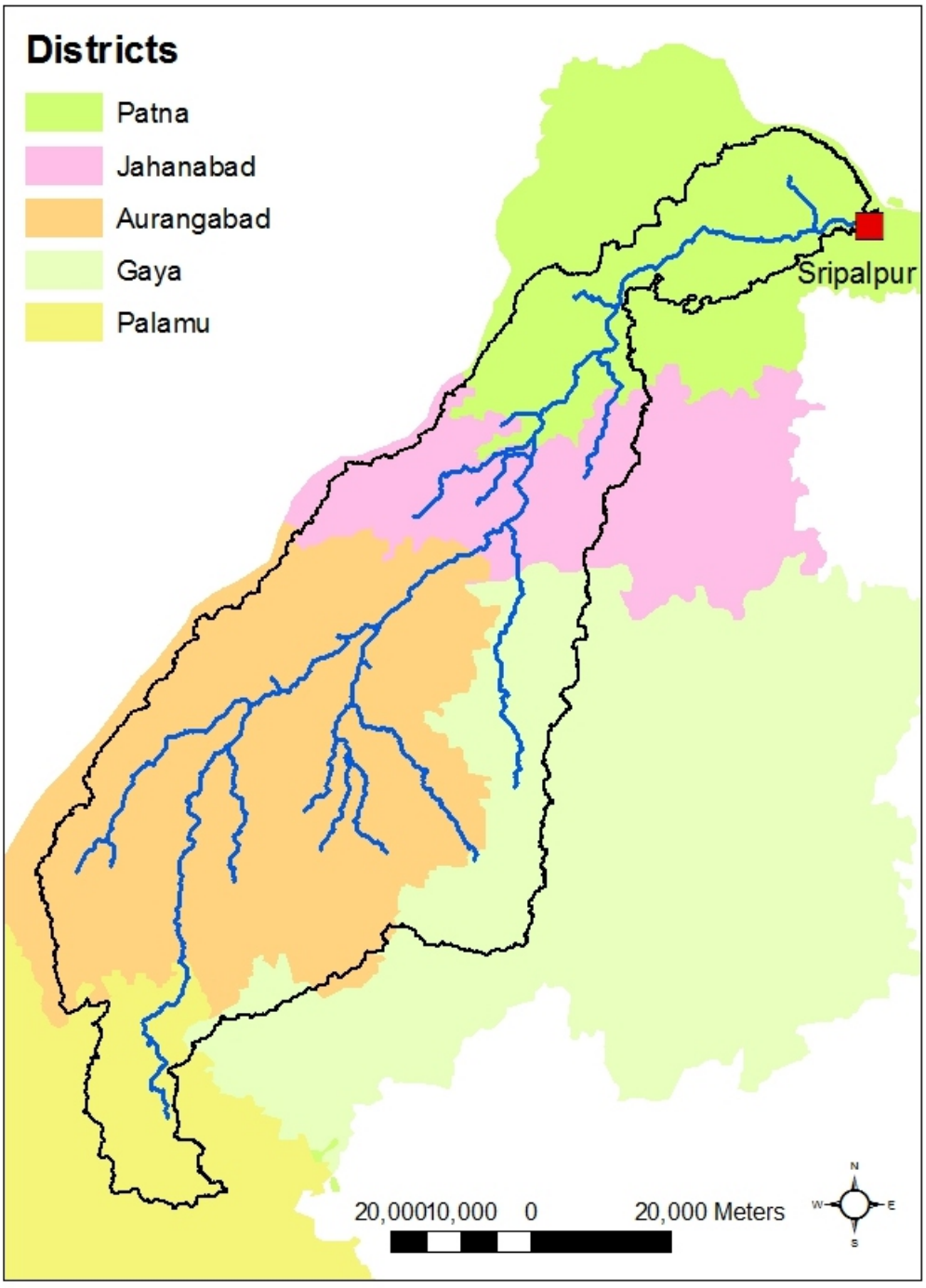


Figure 1: District map of the Punpun river basin

2.3 Model Setup

The input datasets required to run the SWAT model include topographical data, land use land cover (LULC) data, soil data, weather data (daily precipitation, temperature, RH, solar radiation and wind speed)

and observed discharge data (output) for model calibration. Following the SWAT modeling guidelines proposed by Abbaspour et al. (2015), various combinations of input datasets were tested in multiple model runs without calibration and the combination with the best output variable simulation was chosen for calibration and further analyses. Topographic data were available from the Shuttle Radar Topography Mission (SRTM) in the form of a Digital Elevation Model (DEM) raster file. The LULC raster file was obtained from the National remote Sensing Centre (NRSC), Indian Space Research Organization (ISRO). The soil data were based on the raster dataset prepared by the NRSC and National Bureau of Soil Survey (NBSS). Most (81%) of the area had Hydrologic Group C/D soils. Daily precipitation data were derived from the Asian Precipitation Highly Resolved Observational Data Integration Towards Evaluation (APHRODITE) of Water Resources database, with a 0.5° × 0.5° resolution. The daily maximum and minimum temperature and wind speed data were extracted from the Princeton University weather dataset. The daily solar radiation and relative humidity data were derived from the National Centers for Environmental Prediction (NCEP) Climate Forecast System Reanalysis (CFSR) dataset. Daily observed discharge data (m^3/s) were available for the CWC gauge station at Sripalpur from 1959.

Table 1 LULC data for Punpun basin (NRSC, ISRO)

Sl. No.	LULC class	Area (km^2)	Fraction of total area (%)
1	*Kharif* Surface Irrigation	1472.74	26.80%
2	*Kharif* Conjunctive Irrigation	1423.25	25.90%
3	Current fallow	767.90	13.97%
4	Double/Triple Crop Surface Irrigation	555.33	10.11%
5	Double/Triple Crop Conjunctive Irrigation	300.91	5.48%
6	Forest-Deciduous	227.98	4.15%
7	Others (range grasses, mixed forests, *zaid* cropping)	746.81	13.59%

2.4 One-at-a-time Sensitivity Analysis and Calibration/Validation

According to the calibration protocol laid out by Abbaspour et al. (2015), the default model was used to calibrate the outlet discharge. A set of parameters expected to significantly influence the simulated discharge towards the observations was chosen out of the 26 SWAT hydrological parameters (van Griensven et al. 2006). One-at-a-time (OAT) sensitivity analysis was carried out with each of those parameters using the Latin Hypercube (LH) sampling technique. Subsequently 4 parameters (Table 2) of significance were identified, initial ranges were assigned to them and 300 simulations were run using the SWAT-CUP Sequential Uncertainty FItting Algorithm (SUFI2). The study used the Nash-Sutcliffe (NS) efficiency criterion in the objective function.

Table 2 Model parameters used for calibration after LH-OAT sensitivity analysis

Sl. No.	Model Parameter	Definition	Process
1	CN2	SCS runoff curve number for moisture condition II	Runoff
2	ESCO	Soil evaporation compensation factor	Evaporation
3	REVAPMN	Threshold depth of water in the shallow aquifer required for return flow to occur (mm)	Groundwater
4	SOL_AWC	Available water capacity of the soil layer (mm/mm soil)	Soil

2.5 Simulation of invasion by Parthenium Hysterophorus weed at HRU-scale

The 'Land Use Split' feature of the ArcSWAT GIS interface of SWAT was instrumental in 'splitting' the chosen LUs so that a part (50%) of the chosen HRU could be 'invaded' by *Parthenium* and then its outputs could be compared conveniently to the original LU. The choice of LUs for *Parthenium* invasion was based on a tradeoff between simulating invasions, which were most representative of actual conditions on the field, and increasing the complexity of the model by adding more details regarding phenology and weed management operations. Two criteria were used to select the LUs to be 'invaded': to choose the LU with the maximum ease of invasion for the weed, and the LU with the highest impact potential on agricultural productivity. These criteria led to two categories of LU changes or 'invasions' (Table 3) – fallow land during the dry season (January-May) and *kharif* crop LUs during the monsoon. *Kharif* crop LU management operations were incorporated from Kaur et al. (2003).Weed management was not incorporated as it would preclude the quantification of the full extent of hydrological impacts of an invasion.

Table 3 HRU-scale *Parthenium Hysterophorus* invasion using the LU-Split feature

Sl. No.	Original LU	Distribution of new LU	Growing season, Additional LU management inputs
1	Fallow Land	Fallow Land: 50% *Parthenium*: 50%	January – March No additional management input for both LUs
2	*Kharif* season conjunctive use | surface irrigation	*Kharif* season conjunctive use: 50% *Parthenium*: 50%	25th June – 25th September *Kharif*: 2 irrigations (120 mm each, 30% irrigation efficiency, auto-fertilization) *Parthenium*: No management inputs

3. RESULTS

3.1 Calibration/Validation of Outlet Discharge

The default model (without *Parthenium* 'invasions') was run for 8 years (1990-1997) with a two year warm-up period. Observed and simulated streamflows at the outlet are given in Figure 2.

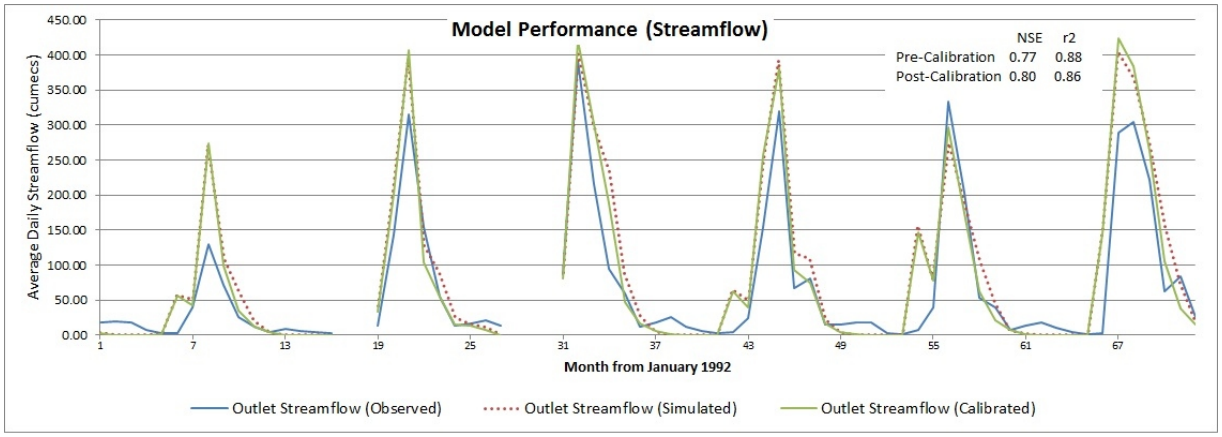


Figure 2: Observed, simulated and calibrated streamflows at basin outlet (Sripalpur): 1992-1997

It was observed that the peaks flows in the monsoonal wet season were consistently over-estimated by SWAT, whereas the baseflows during the dry season (January-May) were highly underestimated. According to Abbaspour et al. (2015), the over-estimation of the peak flows could be resolved by decreasing the curve number (CN2), increasing the available soil water capacity (SOL_AWC) and increasing the soil evaporation compensation factor (ESCO). Also the low base flows could be resolved by decreasing the minimum threshold for groundwater flow to occur (GWQMN), decreasing the groundwater 'revap' coefficient GW_REVAP ('revap' is the movement of water from shallow aquifer to soil profile through evaporation or root uptake by deep rooted trees) and increasing the minimum threshold of shallow aquifer for 'revap' to take place (REVAPMN). Since GWQMN=0, and GW_REVAP=0.02 were by default the minimum values that are allowed by SWAT, after LH-OAT sensitivity analysis, the other 4 parameters were chosen for calibration. Model calibration led to slightly better estimates of middle flows, but peak flows and low flows were not significantly improved. Simulated streamflow after calibration is given in Figure 2.

3.2 HRU-scale Water Balance

The average annual HRU-scale water balances were examined by equating (1) and (2) using the variables given in Table 4. SWAT computes the HRU water balance at the daily scale, so there were a few minor discrepancies when the values were aggregated to the annual level. This was resolved by the assumption that at an annual scale, the changes in the storage terms (Equation 1) were negligible. Equation 2 was used to compute an annual 'water loss/flow' balance.

Table 4 HRU-scale Annual Water Balance (units in mm)

PCP	IRR	ET	RCHRG	SEEP	SURQ	LATQ	GWQ	H_2O Losses|Flows
936.98	144.42	586.03	243.29	240.65	277.18	2.52	231.00	-1.35%

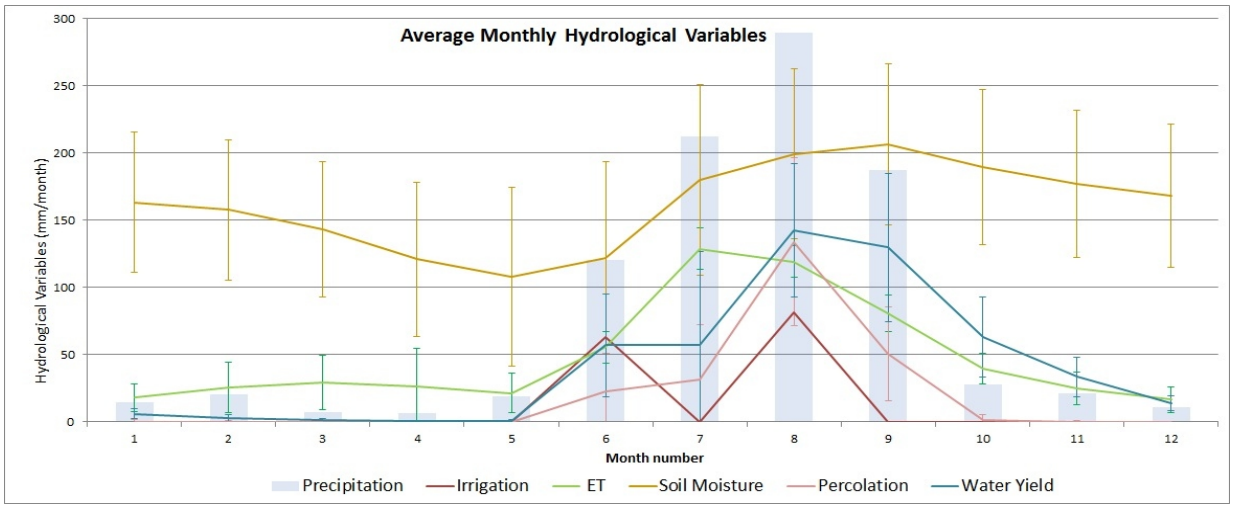


Figure 3: Mean monthly hydrological variables with standard deviation for relevant HRUs

A subsequent analysis was conducted on the LUs of relevance (fallow land, *kharif* crop, and *Parthenium*) by varying the soil type and slope class, independently, to understand how this variation influenced the water balance. The influence of the slope class variation on hydrological output variables while keeping other HRU components constant was negligible. This was because even despite the classification, most of the HRUs had slopes around 1% and computations did not yield large differences at this level. However, the classification of HRUs according to soil type, keeping other components constant, did yield significant differences in hydrological variables. The soil types were categorized according to Soil Hydrologic Group. The soil groups ranged from Group A to Group D. The single Group A soil had a high infiltration rate with lowest runoff potential, whereas the infiltration rates decreased simultaneously with increasing runoff potential moving towards Group D. For the sake of clarity, only the extreme soil groups A and D were chosen to compare relevant hydrological flow variables. The comparison is illustrated in Figure 4. The other groups were intermediate in their hydrological response and had variables between those of the extreme soil groups.

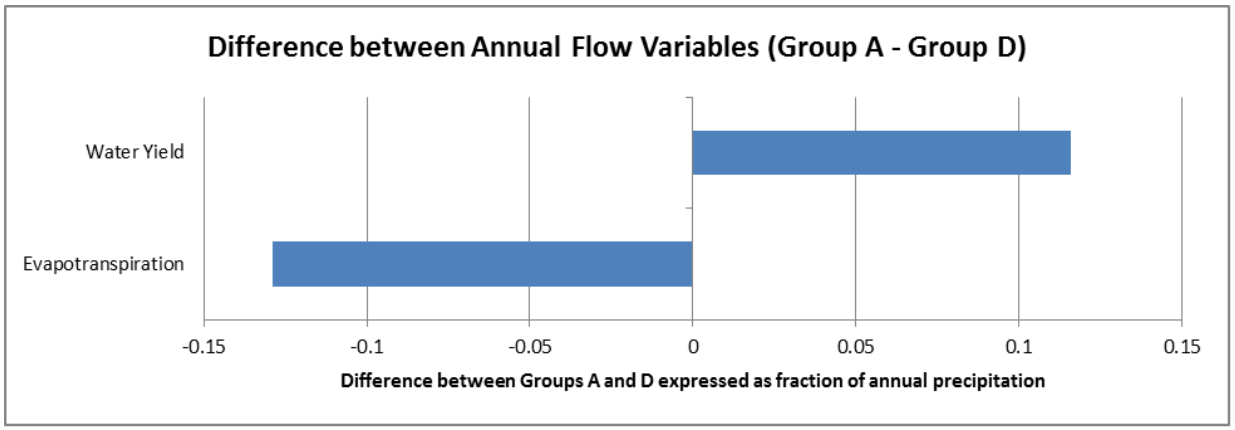


Figure 4: Difference of annual flow variables between Group A and Group D soils

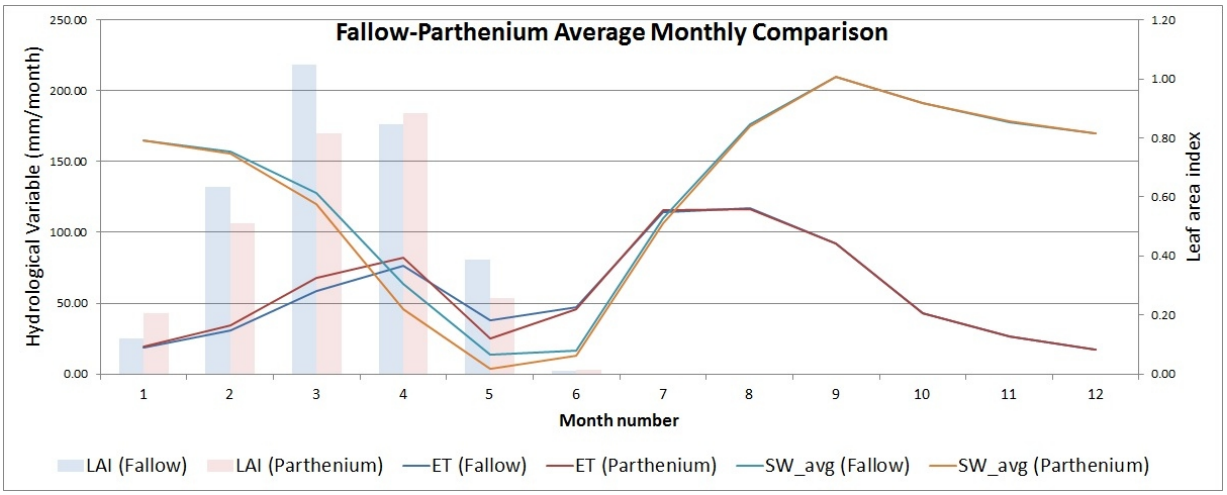
The lower infiltration potential of Group D as compared to Group A led to a significantly higher amount of soil moisture storage in Group D which was contrasted slightly by higher aquifer storage in Group A due to higher infiltration potential. Within the water flows, Group D had significantly higher ET losses contrasted by higher surface runoff, lateral flow and base flow (Water Yield) in Group A.

3.4 HRU-scale analysis of Water Balance after Parthenium Invasion

The average monthly hydrological variables of *Parthenium* LU were compared to the 2 invaded LUs. Subsequently, hydrological responses of HRUs to the simulated invasions were analysed while varying the soil hydrologic groups. The *Parthenium* invasion is essentially a change in the SWAT crop parameters, and therefore only ET and soil moisture components of the water balance get affected. The chosen SWAT method for computing ET was the Penman-Monteith equation (Monteith J.L. 1965) which is a function of temperature, relative humidity, wind speed, solar radiation and crop parameters related to plant height, leaf area index (LAI), and stomatal conductance.

ET losses and monthly soil moisture (SW_avg) simulations averaged over all soil types and slope classes for *Parthenium* invasions on fallow land and *kharif* LU are given in Figures 5(a) and 5(b), respectively. The LAI of the crop/weed/fallow land during the growing periods, which are different for the two scenarios, are also indicated. The significant results inferred from the graphs are:

1. In the Fallow-*Parthenium* scenario, the ET losses are slightly higher for the weed during the first 4 months of the year, accompanied by slightly lower average soil moisture levels. Soil moisture levels are consistently lower for the weed LU. However, despite soil moisture levels decreasing for both the native and invasive LU during the dry season, the difference in ET losses leads to a relative buildup of soil moisture in the Fallow LU compared to the weed LU, and there is a short pre-monsoon period of slightly higher ET losses of Fallow LU as compared to *Parthenium* LU.
2. For the *Kharif-Parthenium* scenario, ET losses of *Parthenium* are consistently slightly higher than for the *Kharif* crop LU. This is accompanied by slightly lower average soil moisture levels for the weed throughout the growing season.
3. The observed differences between ET and soil moisture are statistically insignificant at α=0.05 for both the full year (*p-value* = 0.98 and 0.89 for Fallow-*Parthenium* ET and soil moisture, respectively; and *p-value* = 0.91 and 0.78 for *Kharif-Parthenium* ET and soil moisture, respectively) and also for only the growing season with non-zero LAI values (*p-value* = 0.95 and 0.86for Fallow-*Parthenium* ET and soil moisture; and *p-value* = 0.84 and 0.72 *Kharif-Parthenium* ET and soil moisture, respectively).



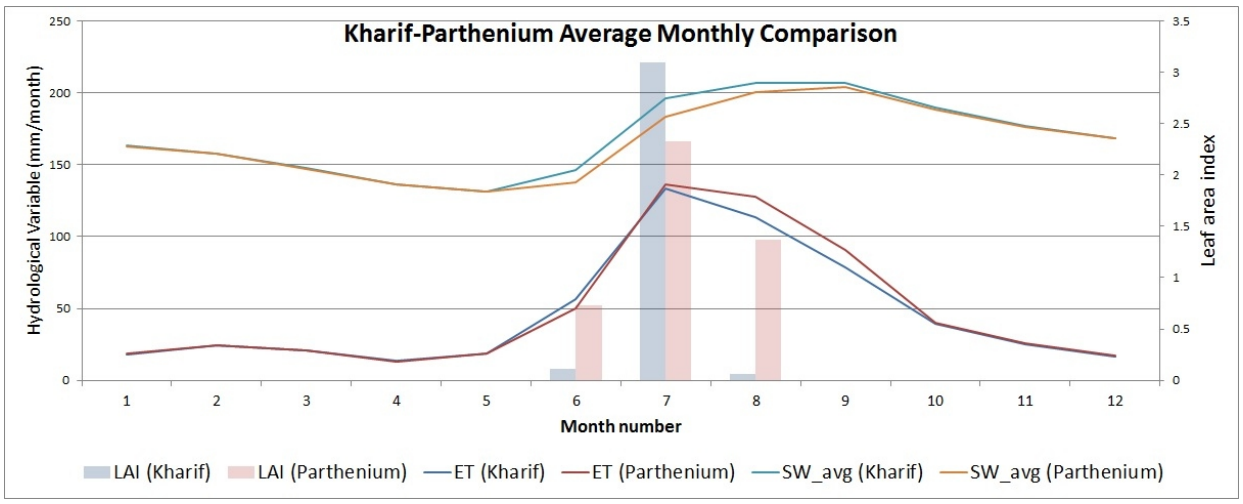


Figure 5: Comparison of variables for (a) Fallow-*Parthenium* and (b) *Kharif-Parthenium* simulations

These results can be explained by analyzing the SWAT crop parameters for the relevant LU types (Table 5). Higher stomatal conductance, maximum canopy height and maximum LAI (favourable leaf area development curve) lead to higher ET losses in the Penman-Monteith potential evapotranspiration (PET) calculations. This coupled with available soil moisture (which may be a limiting factor for ET in dry conditions as seen in the dry months of the Fallow-*Parthenium* scenario) can explain the modeled results satisfactorily.

Table 5 SWAT Crop Parameters

	Maximum LAI	Maximum stomatal conductance (s/m)	Maximum canopy height (m)
Fallow Land	2.5	0.005	1
Kharif season crop	4	0.006	0.9
Parthenium	3	0.007	1.7 (Pandey et al. 2003)

The same analysis when conducted by varying the soil type conserves the patterns of ET and soil moisture, but changes the amount of average monthly soil moisture. The average monthly soil moisture increases from Group A to Group D, as expected, however none of the changes were statistically significant

4. DISCUSSION AND CONCLUSION

The SWAT model simulation of *Parthenium* invasions on fallow land and native *kharif* crops using the SWAT crop parameters results in the inference that *Parthenium hysterophorus* weed has slightly higher but statistically insignificant ET losses as compared to both the native LUs. Additionally, the other objective of this HRU-scale analysis is that these ET losses follow the same trends across hydrologic soil types, which differ only in the amount of average soil water buildup in the soil layers. However, this is in conflict to the heuristic knowledge of local experts who maintain that *Parthenium hysterophorus* is a weed known for its less water uptake, and hence lesser ET losses. This discrepancy may be due to non-localized measurements of SWAT crop parameters. One example of this is the crop parameter 'Maximum Canopy Height' for *Parthenium* which was modified from 1m (SWAT estimate) to 1.7m (Pandey et al. 2003) as it was simply visible that unregulated *Parthenium* plants often grew even taller than 2m on fields and fallow land. Such regional differences may exist in other crop parameters of the relevant LUs thus requiring calibration of parameters for the local conditions of the study area.

REFERENCES

Abbaspour, K. C. (2007). SWAT calibration and uncertainty programs. A user manual, EAWAG: Swiss Federal Institute of Aquatic Science and Technology, 103.

Adla S. & Tripathi S. The Effect of Parthenium Hysterophorus weed on Basin Hydrology (2014). In Tiwari H.L., Suresh S., Jaiswal R.K. Hydraulics, Water Resources, Coastal and Environmental Engineering. Paper presented at 19th International Conference on Hydraulics, Water Resources and Environmental Engineering: HYDRO 2014 International, Maulana Azad National Institute of Technology, Bhopal (pp 531-541). New Delhi.

Arnold, J. G. & Fohrer, N. (2005) SWAT2000: Current Capabilities and Research Opportunities in Applied Watershed Modelling. Hydrol. Processes 19(3), 563–572.

Arnold, J. G., R. Srinivasan, R. S. Muttiah, and J. R. Williams. (1998). Large-area hydrologic modeling and assessment: Part I. Model development. J. American Water Resour. Assoc. 34(1): 73-89.

Dwivedi, P., Vivekanand, V., Ganguly, R., and Singh, R. P. (2009). Parthenium sp. as a plant biomass for the production of alkalitolerant xylanase from mutant Penicillium oxalicum SAUE-3.510 in submerged fermentation. Biomass Energy, 33, 581-588.

Gassman, P. W., Reyes, M. R., Green, C. H. & Arnold, J. G. (2007) The Soil and Water Assessment Tool: Historical Development, Applications, and Future Research Directions. Trans. Am. Soc. Agric. Biol. Engrs 50(4), 1211–1250.

K.C. Abbaspour, E. Rouholahnejad, S. Vaghefi, R. Srinivasan, H. Yang, B. Kløve (2015). A continental-scale hydrology and water quality model for Europe: Calibration and uncertainty of a high-resolution large-scale SWAT model. Journal of Hydrology, 524, 733-752.

Kaur, R., Srinivasan, R., Mishra, K., Dutta, D., Prasad, D., & Bansal, G. (2003). Assessment of a SWAT model for soil and water management in India. Land Use and Water Resources Research, 3, 1-7.

Monteith, J. L. (1965). "Evaporation and Environment." Symposia of the Society for Experimental Biology, 19.

Pandey, D. K., Palni, L. M. S., and Joshi, S. C. (2003). "Growth, Reproduction and Photosynthesis of Ragweed Parthenium (Parthenium Hysterophorus)." Weed Science, 51(2), 191-201.

Patel, S. (2011). Harmful and beneficial aspects of Parthenium Hysterophorus: an update. 3 Biotech, 1(1), 1–9. doi:10.1007/s13205-011-0007-7

van Griensven, A., Meixner, T., Grunwald, S., Bishop, T., Diluzio, M., & Srinivasan, R. (2006). A global sensitivity analysis tool for the parameters of multi-variable catchment models. Journal of hydrology, 324(1)

Modeling Transport Of A Non-Aqueous Phase liquid's (NAPL) Vapor In Variably Saturated Zone

Shachi[1], Brijesh Kumar Yadav[2]
[1]Research Scholar, Department of Hydrology, IIT Roorkee, Roorkee-247667 INDIA
[2]Assisstant Professor, Department of Hydrology, IIT Roorkee, Roorkee-247667 INDIA
Email: [1]shachi2809@gmail.com, [2]brijeshy@gmail.com
Telephone No.: +91 1332 284755/+91 9455688191

ABSTRACT

One of the common groundwater contaminants are non-aqueous phase liquids (NAPL), belonging to the category of volatile organic compounds. When these pollutants come in contact with air and/or water they form separate phases covering large area with long term source of subsurface contamination. Numerical experiments are used here for finding out the effect of NAPL density and characteristics of porous media on the transport of NAPL vapor in the subsurface. The governing equations used for this purpose are the mass conservation equations for three phases viz air, water and oil along with the supporting soil water retention curves. The integrated volume finite difference method is used for discretizing the physical domain spatially and Euler differencing scheme for the temporal domain integrated with Newton-Raphson iteration technique. The results shows the transport of NAPL vapor in subsurface are density dependent. The density effects become less prominent when the hydraulic conductivity was lowered. In case of heterogeneity, the fast movement of NAPL vapor observed in the upper surface of the porous media where sand was considered above silt. The comparative observation of the concentration profile curves for dense and light NAPLs shows the higher concentration of dense NAPL throughout the domain. The results of this study may assist in predicting transport of various NAPL in stratified subsurface and for applying subsequent remediation measures in field.

Keywords: Groundwater Contamination Modelling, Non-aqueous phase liquids, Numerical modeling, subsurface.

1. INTRODUCTION

In the 21st century, the most threat natural resources are groundwater, which are important to fulfil the needs of drinking water, irrigation, and other demands throughout the globe. Due to intense industrial activities, the hazardous chemicals like, volatile organic compounds are originated and get into the groundwater during infiltration. The non-aqueous phase liquids (NAPLs) belonging to the category of volatile organic compounds are classified as dense (DNAPLs) and light (LNAPLs) (EPA 2015). The density of DNAPLs is more than water and LNAPLs have density lesser than water. When NAPLs come in contact with water and/or air forms a separate immiscible phase and differ in physical properties to water. Therefore, physical interfacesare formed between them and prevent them from further mixing which reduce the NAPLs solubility. It migrates downward because of gravitational forces through the subsurface and a part of NAPL is held in soil pores by capillary forces as residue (Yadav and

Hassanizadeh 2011). The migration continues until it encounters a physical barrier or the force exerted by water becomes equal to that exerted by NAPL near the water table (C. J. Newell et. al. 2015). The different soil hydraulic variables affect migration patterns of NAPL. The soil hydraulics parameters to heterogeneous domain having variable values of hydraulic conductivity become more realistic to field conditions. The distribution of grain size affects the migration and entrapment therefore the dissolution process is greatly affected by the effective permeability (Miller et al., 1998).In the heterogeneous system consisting of two phase the larger diameter pores are occupied by gas while the smaller diameter pores are occupied by water (Brusseau 1991).

Previously, the gas phase transport and evaporation of NAPL from the subsurface source was investigated by very few researchers (Falta et al. 1989). Most of the previous investigations considered the aqueous phase and/or dissolution, for example Nambi and Powers(2000)investigated the NAPL dissolution in heterogeneous system. They suggested that effective permeability variation affects the amount of water flowing through source zone of NAPL and the extent of dilution around source zone. The advective-dispersive transport of DNAPL vapor in unsaturated zone was investigated by Mendoza and Frind, 1990. Similarly, Thompson et. al, 1997, investigated factors affecting gas and aqueous phase plume in subsurface and found that plume are sensitive to the vertical position of contaminant source. Furthermore, Jang and Aral (2006) saw the potential of groundwater contamination in the saturated zone has increased because of density driven advection of gas phase contaminant in the unsaturated zone.The objective of this study was to model the effect of NAPL density and characteristics of porous media on the transport of NAPL vapor in the subsurface.

2. SIMULATION EXPERIMENT

The migration of NAPLvapor in subsurface becomes very complex in nature due to multiphase characteristics; and for understanding that various numerical models have been developed. In this simulation experiment the fate and transport mechanisms in porous media was consideredfor multiple phases. Therefore, the mass conservation equation uses solid matrix, aqueous, NAPL and gas as the fluid phases for expressing the conserved quantities in the form of transport and storage. To fulfil the objectives of this study, simulation experiments were performed, with and without considering the effect of density, impact of variation in hydraulic conductivity on the movement of contaminant in heterogeneous subsurface. The numerical method uses integrated volume finite difference method for discretizing the computational domain spatially and implicit scheme was used for the temporal discretization of the solution domain.The simulation experiments were performed using sub-surface transport over multiple phases (STOMP) simulator with water, oil and air operational mode. The

STOMP simulator was written in FORTRAN 77 language using American National Standards Institute (ANSI) standards developed by Pacific Northwest National Laboratory (PNNL) (White and Oostrom, 2000).

The equations that are obtained after discretizing the computational domain spatially and temporally results into non-linear algebraic equations. The resulting equations provide the relationship for the primary variables between the neighboring grid points. The secondary variables of the governing equation depend on the primary unknowns because of which nonlinearity arises.For removing the nonlinearity from the governing equations Newton-Raphson iteration technique is used for a multivariable system. If the starting values obtained for the primary variables are close to the solution, quadratic convergence is displayed by the Newton-Raphson technique (White et. al, 1995). This technique results into a Jacobian matrix which is comprised of the partial differential equations. The matrix is solved to obtain the solution of the linear system of equation by the simulator.

2.1 Governing Equations

The NAPL vapor migration in subsurface was simulated using mass conservation equations for multiple phase transport as:

$$\frac{\partial}{\partial t}\left[\sum_{\gamma=l,g,i}\left(n_D\omega_\gamma^c\rho_\gamma s_\gamma\right)\right] = -\sum_{\gamma=l,g}\left(\nabla F_\gamma^c + \nabla J_\gamma^c\right) - \nabla F_l^s + m^c forc = w,\ a,\ o \qquad (1)$$

$$F_\gamma^c = \frac{-\omega_\gamma^c\rho_\gamma k_{r\gamma}K}{\mu_\gamma}\left(\nabla_{P_\gamma} + \rho_\gamma g z_g\right) forc = w,\ a,\ o \qquad (1a)$$

$$J_\gamma^c = -\tau_\gamma n_D \rho_\gamma S_\gamma \frac{M^c}{M_\gamma} D_\gamma^w \nabla \chi_\gamma^w \qquad (1b)$$

$$F_l^s = D_l^s \nabla s_l \qquad (1c)$$

When component (c) is water and air; phase $\gamma = l, g$ and for component oil, phase$\gamma = l, g, n$.Equations (1a) represent advective flux of water, air and oil in phase γ, kg/m²s; and equation (1b) represent diffusive-dispersive flux of component water, air and oil in γ phase, kg/m²s; equation (1c) represent advective flux of salt in liquid or aqueous phase, kg/m²s. where, n_D is diffusive porosity; $\omega_\gamma^c$is mass fraction of component in γ phase; ρ_γ is phase density for phase γ; s_γ saturation of phase γ; $F_\gamma^c$is advective flux of component in phase γ; $J_\gamma^c$is pressure dispersion flux of component for phase γ; $F_l^s$is advective flux of salt in aqueous phase; m^c is mass source rate of component ; $k_{r\gamma}$is fluid relative permeability of phase γ; K is intrinsic permeability; g is acceleration due to gravity; $P_\gamma$is pressure of phase γ; $z_g$is unit gravitational direction vector; $\tau_\gamma$is phase Tortuosity for phase γ; $M^c$is molecular weight of

component in kg/kg-mol; M_γ is molecular weight of phase γ; D_γ^c is diffusion coefficient of component for phase γ, m²/s; χ_γ^c is mole fraction of component in phase γ; μ_γ is kinematic viscosity of phase γ, Pa-s; D_l^s is salt diffusion coefficient for aqueous phase; s_l is salt concentration in aqueous phase.

2.2 Study Domain

The physical domain used in the simulation process was 30 m in dimension in longitudinal direction and 10 m in dimension in lateral direction (i.e. X and Z direction). At the bottom of the domain the water table was placed and firstly a uniform porous medium comprising of sand is taken.

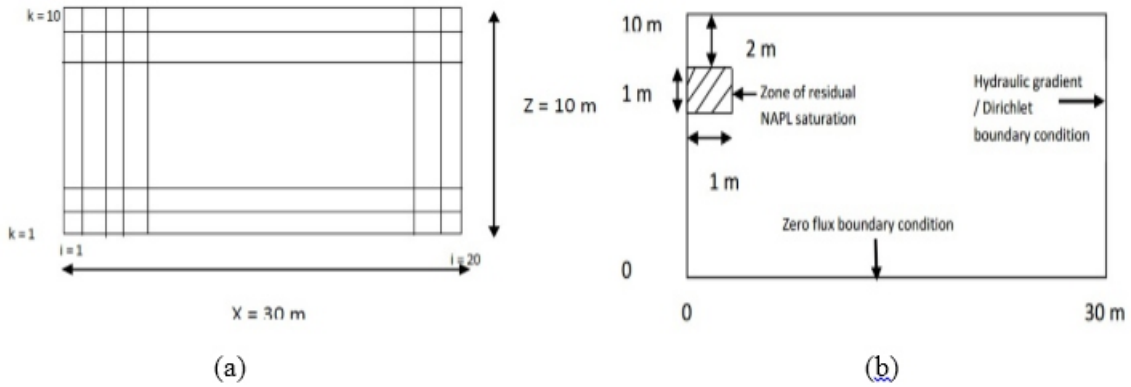


Fig. 1: Study domain (a): Physical domain comprising 30 m in X-dir and 10 m in Z-dir (b): location of source point of NAPL in the physical domain

To incorporate heterogeneity in the defined domain both sand and silt were taken as the porous medium. The 1 m3 region asa zone of residual NAPL saturation of 0.2 was positioned near the left hand side boundary at a distance of 2 m from top shown in figure 1b. The vapor pressure of 12,000 Pa was taken for NAPL. The carbon tetrachloride having molecular weight of 153.82 g/molwas taken as representative of DNAPLand toluene having molecular weight of 92.14 g/mol was taken as representative of LNAPL. The comparative physical properties of both representative NAPLs listed in table 1.The total time period of simulation process was 100 days.

Table1 Fluid and Porous medium properties used in the simulation for both representative

Parameter	DNAPL Value	LNAPL Value
alpha a	2.5	1.25
n	2.0	2
Irreducible Water Saturation S_m	0.1	0.1
Porosity	0.4	0.4
Density	1623 Kg/m^3	867 Kg/m^3
Viscosity	0.97 e-3 Pa-s	0.59 e-3 Pa-s

2.3 Boundary Conditions

In simulation experiments, for study domain showed in figure 1b different boundary conditions were applied. The pressure at the right hand side boundary is described by for gaseous and aqueous phases hydraulic gradient boundary condition was used and for NAPL phase Dirichlet boundary condition was used. At the bottom boundary the pressure for all the phases was described by zero flux boundary condition shown in figure 1b.

3. RESULTS AND DISCUSSION

The spatial and temporal relative concentration was represented using concentration profile curve for the different observation points within the domain. The simulated concentration profile curves were presented in fig. 2 to 7 which represents the different effects of density on the transport of NAPL. In fig 2(a) and 3(a) the effect of density was considered and in fig 2(b) and 3(b) the density effect was not considered for the study domain having DNAPL and LNAPL respectively. The figures (2a-3a) shows that advection of gaseous phase was responsible for the migration of both representatives NAPL near the source location and phase partitioning produces the highly concentrated region at the water table.

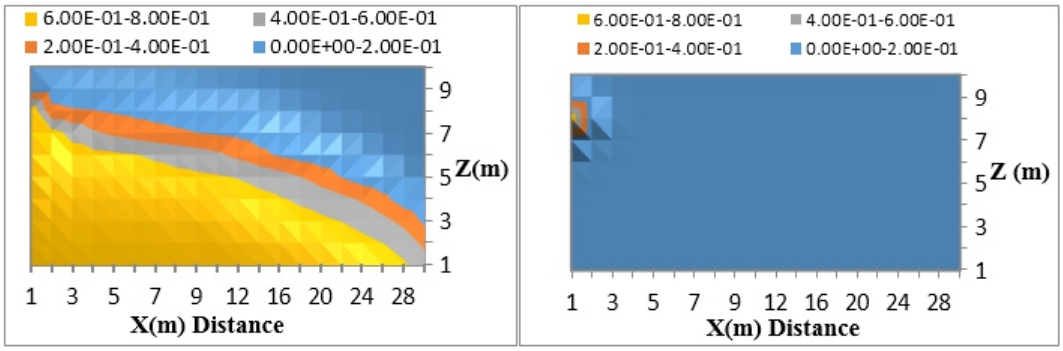


Fig. 2(a) Fig. 2(b)

Fig. 2:DNAPL gas concentration (100 days) (a) when density effect is considered (b) when density effect is not considered.

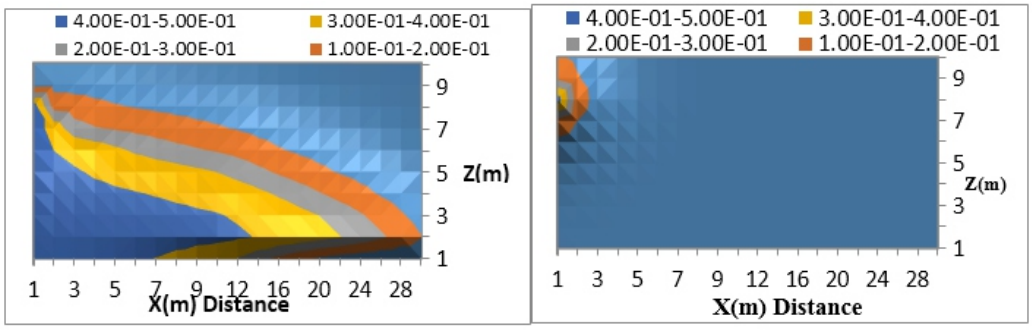


Fig. 3(a) Fig. 3(b)

Fig.3 LNAPL gas concentration (100 days) (a) when density effect is considered (b) when density effect is not considered.

NAPL vapor concentration profile curve of both representatives NAPL (dense and light) for 5 days and 50 days in fig 4 (a, b) and 5(a, b). The comparative profile of both NAPLs showed the DNAPL having fast movement than LNAPL towards the water table whereas the vapor concentrations are more in LNAPL profile. Furthermore, the movement was increasing as time advances for both domains (fig 5).

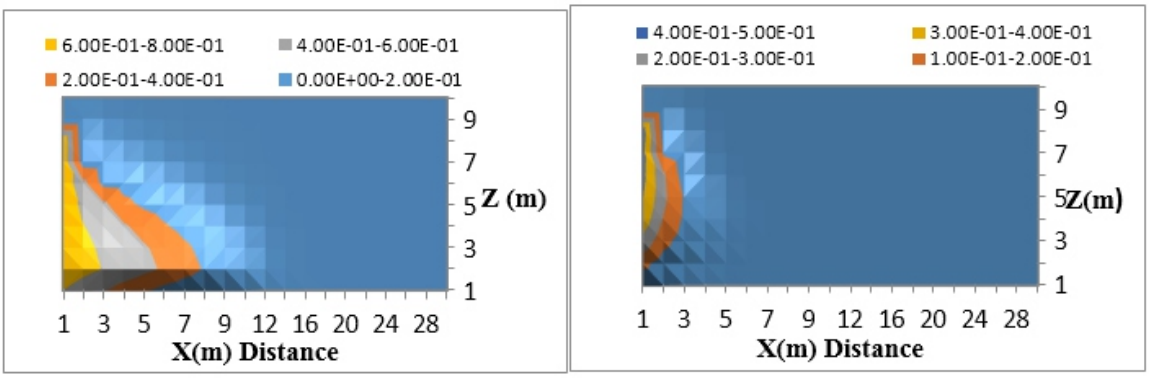


Fig. 4(a) Fig. 4(b)

Fig. 4: NAPL concentration in gas phase after 5 days (a) DNAPL profile (b) LNAPL profile

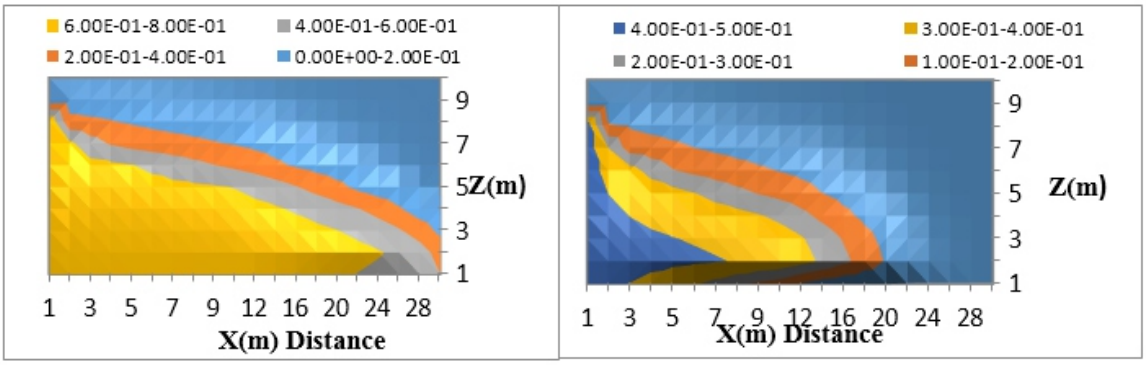


Fig. 5(a) Fig. 5(b)

Fig. 5:NAPL concentration in gas phase after 50 days (a) DNAPL profile (b) LNAPL profile

In the next case study, the effect of hydraulic conductivity was incorporated in the domain. The results indicateNAPL vapor movement becomes comparatively slow by reducing the hydraulic conductivity and also NAPL vapor were not able to cover the entire domain as shown in figure 6.

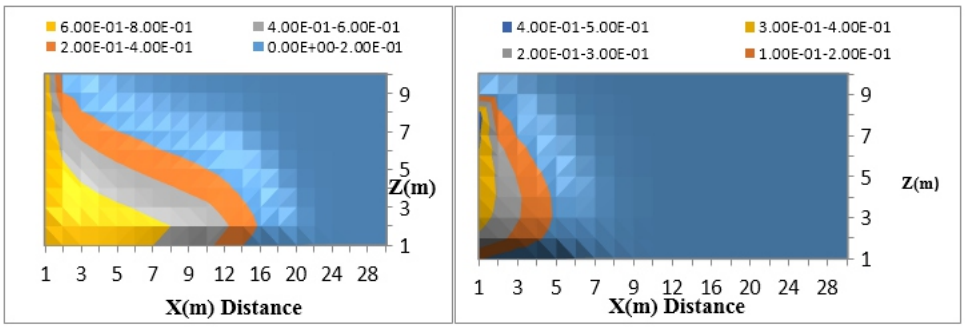


Fig. 6(a) Fig. 6(b)

Fig. 6:NAPL concentration when hydraulic conductivity is reduced 10 times (a) DNAPL profile (b) LNAPL profile

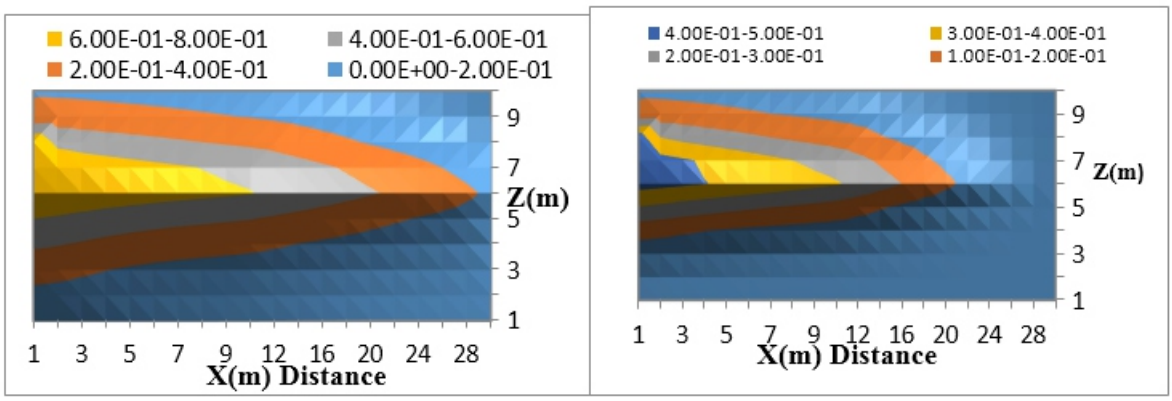


Fig. 7(a) Fig. 7(b)

Fig. 7: NAPL vapor behavior in layered porous media consisting sand and silt (a) DNAPL profile (b) LNAPL profile

To create more realistic domain to field, the heterogeneity was considered for the domain using sand and silt material. The layered porous media was taken consisting of sand having hydraulic conductivity 100.0 m/day in the upper half and silt having hydraulic conductivity 1.0 m/day in the bottom half domain. The result indicates that the movement was dominated in domain having sand than silt. Similarly, for same simulation time period the DNAPL propagates more distance as compared to LNAPL as shown in figure 7.

4. CONCLUSION

In this study, the simulation experiments were conducted for transport of NAPL vapor in the subsurface under different soil hydraulic conditions. The simulation experiments were performed with and without considering the effect of density, impact of hydraulic conductivity and the subsurface heterogeneity. In the investigation, the soil domain having different boundary conditions and the zone of residual NAPL saturation was identified. First, the impact of density was considered for both domain having LNAPL and DNAPL respectively. Subsequently, the effect of hydraulic conductivity and heterogeneity was considered in both the domains. The result shows that the concentration of DNAPL and LNAPL was decreasing in the direction away from the source zone of residual saturation.The DNAPL plume propagates and covers a larger surface area as compared to LNAPL plume. In heterogeneous domain, the movement was dominated in sandy domain than silty zone. The study investigation may help in determining extent of subsurface contamination due to light and dense NAPL and can assist in applying remediation technologies in field.

ACKNOWLEDGMENTS

The authors are thankful to the Department of Science and Technology (DST), India for funding this research under the scheme of Ramanujan fellowship.

REFERENCES

Brusseau, L. (1991). Transport of Organic Chemicals by Gas Advection in Structured or Heterogeneous Porous Media: Development of a Model and Application to Column Experiments, water resources research, vol. 27, no. 12, pages 3189-3199

Charles, J. Newell., Steven, D. Acree., Randall, R. Ross., and Scott G. Huling (2015). Groundwater Issue: light Nonaqueous phase liquids. EPA on www.epa.gov dated 15/08/2015

Falta, R. W., Javandel, I., Pruess, K., and Witherspoon, P. A., (1989), Density driven flow of gas in the unsaturated zone due to the evaporation of volatile organic compounds. Water Resources Research, Vol. 25, No. 10: 2159-2169.

Jang, W., & Aral, M. M. (2007). Density-driven transport of volatile organic compounds and its impact on contaminated groundwater plume evolution, 353–374. http://doi.org/10.1007/s11242-006-9029-8

Mendoza, C. A. &Frind, E. O., Advective-dispersive transport of dense organic vapors in the unsaturated zone 1. Model development. Water Resour. Res., 26 , 379- 87.

Miller, C.T., Christakos, G., Imhoff, P.T., McBride, J.F., Pedit, J.A., Trangenstein, J.A., 1998. Multiphase flow and transport modeling in heterogeneous porous media: challenges and approaches. Adv. Water Resour. 21 (2), 77–120.

Nambi, I. M., & Powers, S. E. (2000). NAPL dissolution in heterogeneous systems : an experimental investigation in a simple heterogeneous system, Journal of Contaminant Hydrology 44, p 161–184

White &Oostrom (2000). Subsurface Transport Over Multiple Phases Version 2.0 Theory Guide.

White, M. D., Oostrom, M., &Imhard, J. (1995). Modeling fluid flow and transport in variably saturated porous media with the STOMP simulator .1 .Nonvolatile three-phase model description, 18(6), 353–364.

Yadav, B. K., &Hassanizadeh, S. M. (2011). An overview of biodegradation of LNAPLs in coastal (semi)-arid environment. Water, Air, and Soil Pollution. doi:10.1007/s11270-011-0749-1.

Derivation of Intensity Duration Frequency Curve Based on Multivariate Empirical Mode Decomposition and Scaling Theory

Adarsh S[1*], Sreelakshmy S[2] Pavithra CP[3]
[1]Assistant Professor, Department of Civil Engineering,
TKM College of Engineering Kollm-691005
[2]Post Graduate Student, Department of Applied Mechanics and Hydraulics,
NIT Suratkal-575025
[3]Formerly Graduate Student, Department of Civil Engineering,
TKM College of Engineering Kollm-691005
Email**:** adarsh_lce@yahoo.co.in*, lechusree93@gmail.com,
pavithracp93@gmail.com
Mobile No.: +91 9446915388*

ABSTRACT

Intensity-Duration-Frequency (IDF) relationship of rainfall is one of the most commonly adopted tools in the design of urban storm water drainage systems. The information on the time scale invariance property of rainfall is a useful mean for estimation of IDF relationships of shorter durations from the data in coarse time resolution though the disaggregation operation in time domain. This paper proposes a novel method for deriving the IDF relationships based on the scaling property, in which the Multivariate Empirical Mode Decomposition (MEMD) method is used for determination of scaling exponent by considering rainfall intensity series for different durations simultaneously. The logarithmic plot between Probability Weighted Moments (PWM) of Intrinsic Mode Functions (IMFs) obtained from MEMD and the duration gives the scaling exponent, and finally the IDF relationships are derived based on Generalized Extreme Value (GEV) formulations involving scaling exponents. To validate the correctness of the proposed method, first the method is applied for hourly rainfall data from Bangalore station, India, and the results are compared with those reported in published literature. The difference in estimation of intensity values by the two methods is quantified in terms of Root Mean Square Error (RMSE) statistics, which is found to be increasing with increase in return period and the derivation by the classical frequency factor method is found to be an over estimation. Then the method is applied for the derivation of IDF relationships for different sub-daily durations from the daily rainfall data of Kollam urban area in the state of Kerala. The encouraging results showed that the proposed method is a robust alternative for derivation of regional IDF curves.

Keywords: IDF, Scaling, MEMD, Rainfall

1. INTRODUCTION

Urban flooding is a disaster of major concern among the hydrologists nowadays. The accumulation of high intense rainfall over short duration can be prominent in urban areas of coastal regions, either due to coupled effect of rainfall with tidal actions or due to insufficient/improper drainage facilities may eventually lead to flooding in such areas. In this context the design of urban drainage systems is of considerable practical significance. For design of such systems and the evaluation of flood risk associated with it, the intensity-duration-frequency (IDF) relationships are one of the key tools still in practice. The frequency of occurrences of a storm of a given intensity and duration can be estimated statistically based on the IDF curves prepared for a specific region. Eventhough, such short duration

rainfall (sub daily or sub hourly) information necessary for the design of such systems, the national hydrological services provide the data mainly in daily or monthly time step. The traditional methods of IDF curve generation may fail in such circumstances and the scale invariance theory can be used to solve this problem. Also the output of spatial downscaling studies of rainfall are primarily available in monthly time resolutions and in such cases this concept can be used in the derivation of IDFs under a changing climate scenario (Ali and Mishra 2014; Afrin et al. 2015; Herath et al. 2015; Chandrarupa et al. 2015). The scale invariance theory has gotten much attention in hydrology in the past and some of them addressed development of simple and multiscaling models for regional precipitation estimation (Waymire and Gupta 1981; Waymire et al. 1984; Veneziano et al. 1996). Stemming from the scale invariance theory, alternative approaches were also evolved to construct IDF relationships, based on the fractal properties of rainfall (which implies scaling invariance). In these studies scaling formulas were proposed to extend the IDF relationships from daily time scale to shorter time intervals based on scaling properties of rainfall. For example, Gupta and Waymire (1990) studied the concepts of simple and multiple scaling to characterize the probabilistic structure of the precipitation process, Koutsoyiannis and Foulfoula-Georgiu (1993) used a scaling model to predict storm hyetographs. Menabde et al. (1999) showed that based on the empirically observed scaling properties of rainfall and some general assumption about the cumulative distribution function for the annual maxima of mean rainfall intensity, it is possible to derive simple IDF relationships. Yu et al. (2004) developed regional rainfall intensity formulae based on the scaling property of rainfall. Nhat et al. (2008) developed the IDF curves for sub hourly and hourly durations for different stations in Asia-Pacific region, assuming Extreme Value 1(EV1) distribution. Bara et al. (2009) developed sub daily IDF curves for Slovak region. Afrin et al. (2015) developed the a regional IDF relationship for Dhaka city for present as well as future climatic scenarios based on simple scaling. Ali and Mishra (2014) performed an extensive study for developing sub-daily IDF curves for major urban cities in India under a changing climate scenario. In all of the above studies, the scaling exponent was obtained by the classical procedures.

Recently Kuo et al. (2013) proposed a new method for deriving regional intensity-duration-frequency (IDF) based on the scaling property of precipitation. The representative scale exponents for different durations are obtained by ensemble empirical mode decomposition (EEMD) method. The method was applied to derive IDF curve for Edmanton, Canada. The results show that quantiles derived from generalized extreme value (GEV) probability distribution with parameters derived by the probability-weighted moment (PWM) are more accurate than those derived from the EVI distribution with parameters derived by the method of moment (MOM), It was further observed the underestimation of rainfall intensity by the EV1-MOM method for high return period (greater than 25 years) and short

duration (less than 1 h). However, the EEMD method decomposes each of the rainfall intensity series independently. As the EEMD method perform the decomposition adaptively (considering the complexity of the data), the number of modes obtained by the decomposition may vary. This may introduce errors in the estimation of scaling exponents and offer difficulties in its computation. Hence it is believed that the decomposition by the identification of common time scales in different intensity series is a better alternative and the common scales present in different rainfall intensity can be captured by the Multivariate EMD (MEMD) method. Here the decomposition of all intensity series is performed in single step. This enables to maintain the same number of modes for all intensity series and hence simplify the process considerably. Therefore this study proposes a new method for developing the IDF curves based on MEMD to find the representative scaling exponent. In short, this paper performs (i) the scaling property based estimation of IDF curves of Bangalore city and compare the results by frequency factor method; (ii) apply the method for deriving hourly IDF curves for the coastal city of Kollam in the state of Kerala India, from daily rainfall data.

2. MULTIVARIATE EMPIRICAL MODE DECOMPOSITION

Huang et al. (1998) proposed a purely data adaptive decomposition procedure namely Empirical Mode Decomposition to analyse the non-linear and non-stationary time series. Multivariate EMD proposed by Rehman and Mandic (2010) is an extension of the traditional EMD, which decomposes multiple time series simultaneously after identifying the common scales inherent in different time series of concern. In this method, multiple envelops are produced by taking projections of multiple inputs along different directions in an n-dimensional space. A brief description of the MEMD algorithm, is presented below (after Rehman and Mandic 2010; Hu and Si 2013).

Assuming $V(t)=\{v_1(t),v_2(t)....v_n(t)\}$ being the n spatial datasets as a function of time t and $X^{\theta_k}=\{x_1{}^k,x_2{}^k,....,x_n{}^k\}$ denoting the direction vector along the direction given by angles $\theta_k=\{\theta_1{}^k,\theta_2{}^k,...\theta_{n-1}{}^k\}$ in a direction set X (k=1,2.3,….K, K is the total number of direction). Then IMFs of n time series datasets can be obtained by the following algorithm

1. Generate a suitable set of direction vectors X
2. Calculate the projection $p^{\theta_k}(t)$ of the datasets $V(t)$ along the direction vector X^{θ_k} for all k
3. Find spatial instants $s_i{}^{\theta_k}$ corresponding to the maxima of projection for all k
4. Interpolate $t_i{}^{\theta_k},V(t_i{}^{\theta_k})$ to obtain multivariate envelop curves $e^{\theta_k(t)}$ for all k
5. The mean of envelope curves ($M(t)$) is calculated by $M(t)=\frac{1}{K}\sum_{k=1}^{K}e^{\theta_k}(t)$
6. Extract the 'detail' $D(t)$ using $D(t)$ = $V(t)$-$M(t)$. If $D(t)$ fulfills the stoppage criterion for a multivariate IMF, apply the above procedure to $V(t)$ the above procedure to $V(t)$-$D(t)$, otherwise apply it to $D(t)$

The stopping criteria proposed by Huang et al. (2003) can be used for MEMD with the difference that the condition for equality of number of zero crossings and extrema cannot be imposed, as the extrema cannot be defined for multivariate signals (Hu and Si 2013).

3. SCALING THEORY AND ESTIMATION OF SCALING EXPONENT BASED ON MEMD

The derivation of IDF curves for sub-daily durations used in this study is based on the concept of scale invariance theory. Scaling implies that the statistical properties of the process observed at various scales are governed by the same relationship (Olsson 1998; Olsson and Berndtsson 1998). Let X(t) and X(λt) denotes the observations (time series) at two distinct time scales t and λt, where λ is the scale resolution factor (a positive quantity). If X(t) is scaling, then there exists a function $f(\lambda)$ such that

$$X(t) \overset{dist}{=} f(\lambda)\, X(\lambda t) \qquad (1)$$

where $f(\lambda) = \lambda^{-\beta}$ $\overset{dist}{=}$ denote the equality of the probability distribution of the two random variables X(t) and X(λt) and β is the scaling exponent, also known as the Mandelbrot–Kahane–Peyriere (MKP) Function (Mandelbrot et al. 1974).

In other words, $X(t) \overset{dist}{=} \lambda^{-\beta} X(\lambda t)$ (2)

This is known as strict sense simple scaling (SSSS) and it states "that equality in the probability distribution of the rainfall depth observed at two different time scales holds" (Gupta and Waymire 1990). This implies that the quantiles and raw moments of any order are also scale invariant i.e., $E[X(t)^q] = \lambda^{-\beta q} E[X(\lambda t)^q]$ (referred to as wide sense simple scaling, WSSS), where q is the order of the moment. In simple scaling, the slope β can be obtained from the slope of the linear regression relation between log transformed values of $E[(X(\lambda t)^q]$ and the scale resolution factor λ for different orders of moment.

The annual maximum rainfall intensity Id (defined by the maximum value of moving average of width d of the continuous rainfall process) is the key random variable of concern in deriving the rainfall IDF relationships. According to the scaling theory, a random variable Id obey the simple scaling properties if it obeys the following (Gupta and Waymire 1990) :

$$I_d = \left(\frac{D}{d}\right)^{-\beta} I_D \qquad (3)$$

where D is the aggregated time duration, which can be related to the duration d by defining the scaling ratio $\lambda_d = \frac{D}{d}$

It is to be noted that λ_d can be substituted as 1, 24 and 720 respectively if we use the hourly data, daily data and monthly data in an exercise of deriving IDF curve for sub daily durations.

Hence by scaling theory,

$$I_d = \lambda_d^{-\beta} I(\lambda_d d) \quad (4)$$

and $$E[I(d)^q] = \lambda_d^{-\beta q} E[I(\lambda_d d)^q] \quad (5)$$

The assumption of WSSS enable us to derive the IDF relationships for sub daily durations from daily or monthly time series data. The scaling exponent β can be computed based on the scaling properties of the probability weighted moments (PWMs) of rainfall intensity time series for various durations based on multivariate empirical mode decomposition (MEMD). The PWM (Greenwood et al. 1979) is defined as

$$M^{i,j,q} = E[I_d{}^i F_{I_d}{}^j (1 - F_{I_d})^s] = \int_0^1 I_d{}^i F_{I_d}{}^j (1 - F_{I_d})^s dF_{I_d} \quad (6)$$

Where F_{I_d} is the cumulative distribution function (CDF) of I_d. Adopting $i=1$, $j=q$ and $s=0$, (which are linear in I_d) is generally sufficient for parameter estimation (Hosking 1986). Thus the left hand side of the Equation (6) becomes $M^{1,q,0}$, which is generally denoted as M^q.

If two rainfall intensity time series of duration d and $\lambda_d d$ (I_d and I_D) have similar scaling property, their q order PWMs will have the following relationship :

$$M_d{}^q = \lambda_d^{-\beta} M_D{}^q \quad (7)$$

The scaling exponent can be obtained from the log-linear relationship of PWM for different moment orders as explained earlier. Also in simple scaling, the exponent will be independent of moment order q. The scaling exponent can be derived by the MEMD approach by adopting the following steps :

1. Prepare annual rainfall intensity series for different larger resolution durations (i.e., 1 day, 2 day etc.)
2. Decompose all of the rainfall intensity series simultaneously using the MEMD method
3. For each IMF, find the PWM values for different durations, for different moment orders
4. Fit a log linear relationship between the PWM and the duration and estimate the slope (β), for each IMF, for different moment orders
5. Discard such IMFs which give low value of coefficient of determination (R^2) statistic (say below 0.8) in the fit or which do not show simple scaling (considerably different β for different moment orders)
6. Form a new series by the addition of all of the remaining IMFs and repeat steps 2 to 5 till all of the IMFs pass the criteria mentioned in step 5.
7. Compute the average value of the scaling exponent β as the average of slopes for different IMFs and for different moment orders
8. Use the scaling exponent computed in step 7 for the derivation of IDF curves for different return periods, using the extreme value based expression illustrated in the next section.

The overall procedure is given in the form of a flowchart, in Figure 1.

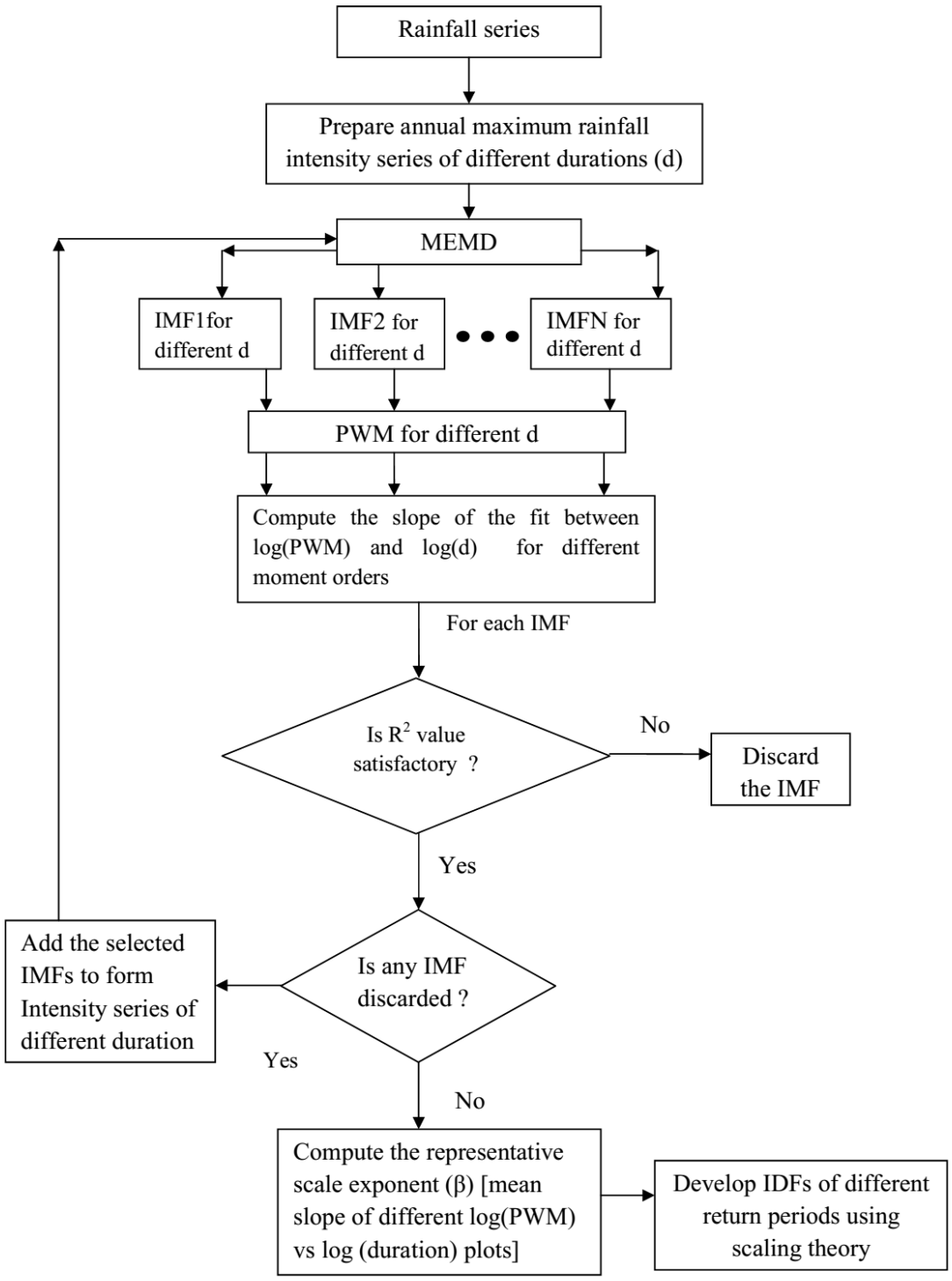


Figure 1 Flowchart of the proposed MEMD-GEV-PWM coupled approach for preparation of IDF curves

3.1 Derivation of IDF based on scaling theory

The general form of IDF relationships can be

$$i_d = \frac{a(T)}{(d+\theta)^{\eta}} \tag{8}$$

i_d is the intensity for duration d, T is the return period , θ is a positive constant and $0< \eta <1$.
The cumulative distribution function (CDF) of the random variable I_d is given by

$$P(I_d \le i) = F_d(i) = 1 - \frac{1}{T(i)} \tag{9}$$

If the GEV distribution is adopted, the scaling IDF can be derived as (Kuo et al. 2013)

$$i_{d,T} = \left[\mu_D + \frac{\sigma_D}{k}\left(1-y^k\right)\right]\left(\frac{D}{d}\right)^{-\beta} \quad \text{where } y = -\ln(1-1/T) \tag{10}$$

$$i_{d,T} = \left[\mu + \frac{\sigma}{k}\left(1-y^k\right)\right]\left(\frac{1}{d}\right)^{-\beta} \tag{11}$$

where μ and σ are the location parameter and scale parameter respectively given by $\mu = \mu_D (D)^{-\beta}$ and $\sigma = \sigma_D (D)^{-\beta}$

4. DATA DETAILS

This study demonstrates the application of the proposed method by considering two different datasets. The dataset from Bangalore city in India is used for the example validation. The study area is located at 12.59oN, 77.57oE, and is situated at an altitude of 920 m above mean sea level. The average annual rainfall of the city is 974 mm, summer temperature ranges from 18 to 38oC and the winter temperature ranges from 12 to 25oC. Bangalore receives rainfall from both southwest (June to September) and northeast monsoons (October to December). The convective heat transfer is the predominant factor of summer monsoon rainfall of the city. The hourly maximum rainfall and intensity series for durations 1 hr, 2 hr, 6 hr, 12 hr and 24 hr for the period 1960-2003 are collected from literature (Mujumdar 2008) and used for preparation of IDF curve by the GEV-PWM implementation based on scaling theory and MEMD. After validation, the proposed methodology is used for developing the IDF frequency curve for sub-daily durations based on the daily time series data obtained from Kollam city station in state of Kerala for the period 1993-2012. The daily data for the period 1993-2012 has been collected from India Meteorological Department, Trivandrum (http://www.imdtvm.gov.in/). For selection of appropriate probability distribution, the Akaikai Information Criteria (AIC) and Bayesian Information Criteria (BIC) can be used and lower the value of the criteria, best will be the distribution for the given time series (Ghosh 2010). The annual maximum rainfall intensity series of Bangalore station are evaluated with AIC and BIC and the results are presented in Table 1/

Table 1 Selection of best probability distribution based on AIC and BIC

Duration	GEV		Log-Normal		Gamma		EV1	
	AIC	BIC	AIC	BIC	AIC	BIC	AIC	BIC
1 hr	279.11	283.60	279.05	282.04	283.45	286.44	330.01	333.00
2 hr	263.10	267.59	262.45	265.44	265.43	268.42	309.65	312.64
6 hr	205.78	210.27	213.03	216.02	221.91	224.91	281.21	284.21
12 hr	160.85	165.34	168.55	171.54	179.14	182.14	244.82	247.81
24 hr	118.94	123.43	127.66	130.66	137.62	140.61	200.69	203.68

From Table 1, it is clear that the GEV distribution is best probability distribution to represent the rainfall intensity series. Hence MEMD-GEV-PWM coupled approach is followed to derive the IDF relationships of Bangalore city.

The annual maximum rainfall intensity series for different durations (d =1, 2, 6, 12, 24 hrs) collected from literature is first decomposed by MEMD method. The decomposition resulted in a total of six modes ranging from high to low frequency and the results of decomposition are presented in Figure 2. Adopting the procedure discussed in section 3, PWM values are computed. The linear fitting of PWM and durations are made for each IMF, for different moment orders. The results of fitting are shown in Table 2. It shows that the R^2 statistics of the fitting of IMF4 is the least followed by that of IMF3. Thus these two modes are excluded and a new time series is formed by considering all the remaining modes. The decomposition of new series also resulted in six modes and the fitting statistics gave R^2 values of all modes consistently greater than 0.85 and for brevity, the results are not presented here. The mean representative scaling exponent (β) is found to be -0.7175. This exponent is used in the scaling formula to obtain the IDF for different durations and return period.

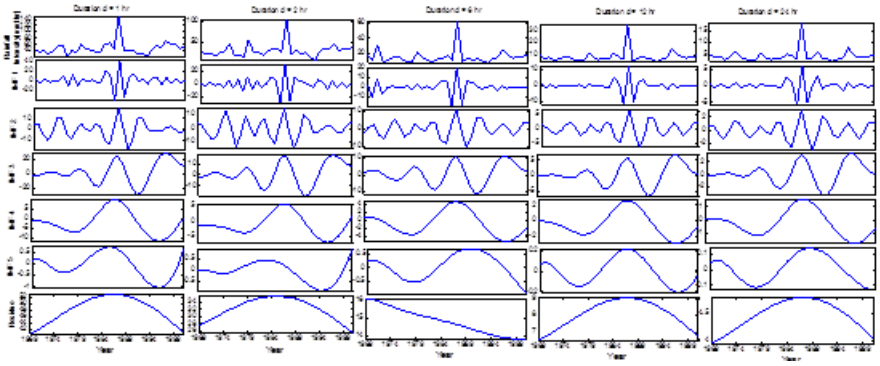


Figure 2 Decomposition of rainfall intensity time series of different durations

Table 2 Scale exponent (β) and R^2 of IMFs of original intensity series for different moment orders. The italic figures indicate that the IMF components 4 and 5 are discarded while forming the new intensity series

IMFs	Scale exponent (β) and R^2 Statistics	Moment order			
		1	2	3	4
IMF1	β	-0.6895	-0.6637	-0.659	-0.6534
	R^2	0.9378	0.939	0.9389	0.9399
IMF2	β	-0.5234	-0.5321	-0.5338	-0.5346
	R^2	0.8845	0.8838	0.8789	0.8747
IMF3	β	-0.6886	-0.6791	-0.6768	-0.6757
	R^2	0.9634	0.9662	0.9668	0.9673
IMF4	*β*	*-0.4723*	*-0.5088*	*-0.5131*	*-0.5134*
	R^2	*0.8134*	*0.8994*	*0.9128*	*0.918*
IMF5	*β*	*-0.4163*	*-0.4769*	*-0.4919*	*-0.498*
	R^2	*0.5635*	*0.6688*	*0.6986*	*0.7129*
Residue	β	-0.7495	-0.7494	-0.7478	-0.7449
	R^2	0.9861	0.9785	0.971	0.9628

The IDF curves are also prepared by the classical frequency factor method. The IDF obtained by both methods are presented in Figure 3. The comparison of IDF curves by both methods confirmed that the results by frequency factor method shows an over estimation over the present method. To quantify the Difference, the root mean square (RMSE) statistics is used and the values of RMSE for IDF curve of different return period are presented in Figure 4. It showed for larger return period (5 years and 100 years), the difference is larger. Similar results were reported by Kuo et al. (2013) in an exercise of preparation of IDF curve for Edmanton city, Canada.

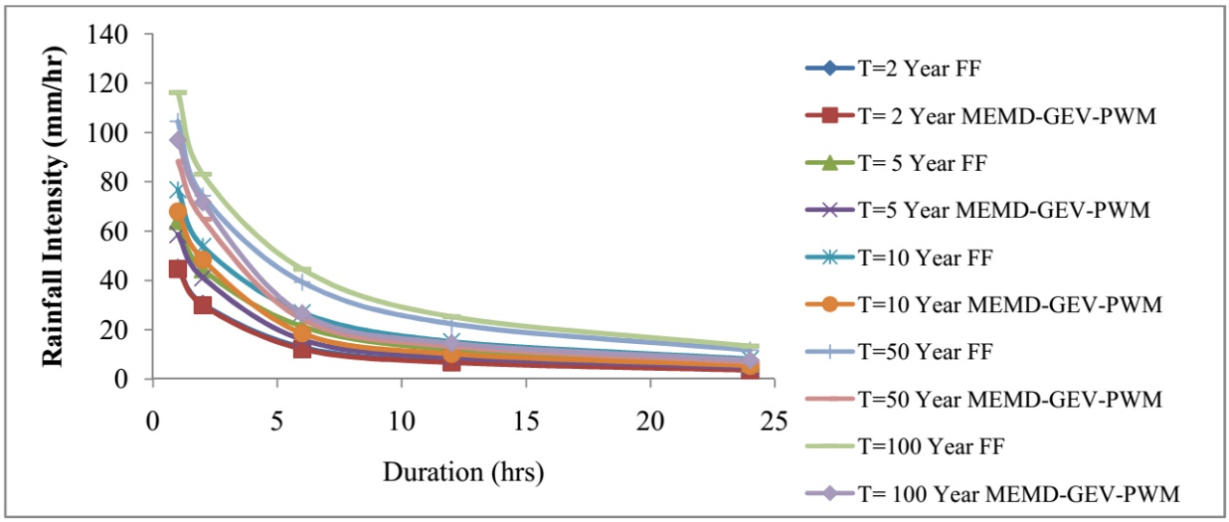


Figure 3 IDF curves for Bangalore city by MEMD-GEV-PWM method and frequency factor method

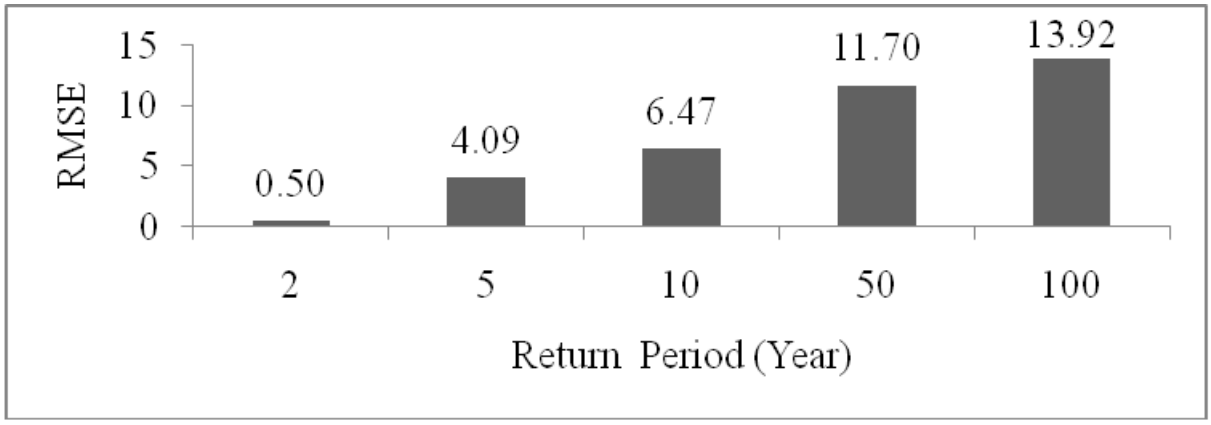


Figure 4 RMSE statistics of the IDF curves of Bangalore city for different return periods, derived by the frequency factor method and proposed MEMD-GEV-PWM method

After understanding the capability of the proposed method, the IDF hourly IDF curves of the city of Kollam are prepared form the daily time series. The annual maximum rainfall intensity are estimated and here also GEV distribution is found to be the best among the four candidate probability distributions. Subsequently, the proposed procedure is implemented. Here the decomposition resulted in five modes and the IMF3 is eliminated in the first cycle of operation. After the second cycle, it is found that the scaling exponent is -0.5752 , it can be adopted to derive the IDF relationships. It is to be noted that in the scaling IDF formula, the scaling factor to be considered as 24 in deriving the hourly IDF curve from daily time series. The resulting IDF curves for Kollam city are presented in Figure 5. The IDF curve can be used for the design of urban storm water drains in the city of Kollam. The proposed methodology being a general one, it can be applied for deriving IDF relationships for shorter durations (like sub-hourly, sub-hourly) from data of larger temporal resolutions (monthly). It can be extended for regional IDFs and for deriving similar relationships for a future rainfall projected by downscaling methods under changing climate scenario.

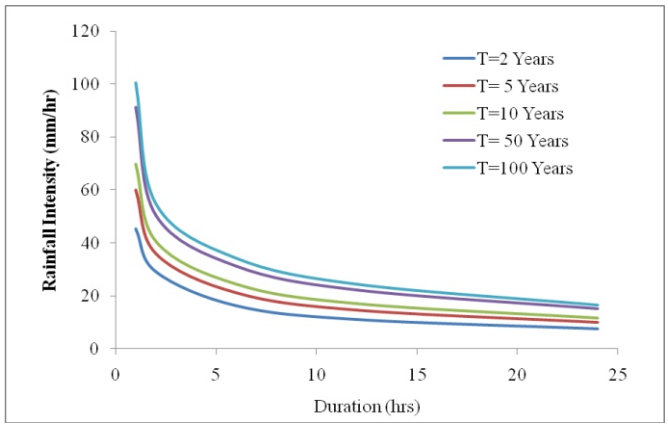


Figure 5 IDF curve of Kollam city by MEMD-GEV-PWM method

5. CONCLUSIONS

The IDF curves for sub-daily durations have lot of practical significance in the design of urban drainage systems. The derivation of hourly IDF curves from daily/monthly data can be facilitated by the scaling theory. A novel method for deriving the scaling theory based IDF curve preparation is proposed in this paper, in which the representative scaling exponent in the scaling expressions is obtained by Multivariate Empirical Mode Decomposition (MEMD) method. The MEMD approach is capable to decompose the time series of different durations simultaneously after identifying the common scales present in different series, which makes the uniqueness in number of decomposition modes which finally makes the procedure simpler when compared with the decomposition based on the more popular EEMD. The comparison of IDF curves prepared based on the proposed MEMD-GEV-PWM coupled approach is compared with results by classical frequency factor showed an over estimation of intensity values particularly for larger return period. Finally, the proposed method is applied for developing IDF curves for the coastal city of Kollam in the state of Kerala. The hourly IDF curves derived are helpful in the design of urban drainage systems of Kollam city.

REFERENCES

Ali, H., and Mishra, V. (2014). Development of sub-daily Intensity Duration Frequency (IDF) curves for major urban areas in India. (Paper ID Number: 14778), AGU Fall meeting 2014, San Francisco, US

Chandra Rupa, Sha U, and Mujumdar PP. (2015). Model and parameter uncertainty in IDF relationships under climate change. Advances in Water Resources, 79 (2015) 127–139

Bara, M., Kohnova, S., Gaal, L., Szolgay, J, and Hlavcov, K.. (2009). Estimation of IDF curves of extreme rainfall by simple scaling in Slovakia. Contributions to Geophysics and Geodesy, 39(3), 187–206

Ghosh, S. (2010). Modelling bivariate rainfall distribution and generating bivariate correlated rainfall data in neighbouring meteorological subdivisions using copula. Hydrological Processes, 24(24), 3558-3567

Greenwood, J. A., Landwehr, J. M., Matalas, N. C., and Wallis, J. R. (1979). Probability weighted moments: Definition and relation to parameters of several distributions expressible in inverse form. Water Resources Research, 15(5), 1049–1054

Gupta, V.K., and Waymire, E. (1990). Multiscaling properties of spatial rainfall and river flow distributions. Journal of Geophysical Research, 95, 1999-2009

Herath, H. M. S. M., Sarukkalige, P. R., and Nguyen, V. T. V. (2015) Downscaling approach to develop future sub-daily IDF relations for Canberra Airport Region, Australia. Proceedings of IAHS, 369, 147-155

Hosking, J. R. M. (1986). The theory of probability weighted moments. Research Report RC 12210, IBM Research Division, Yorktown Heights, NY

Hu, W. and Si, B.C. (2013). Soil water prediction based on its scale-specific control using multivariate empirical mode decomposition. Geoderma 193,180-188.

Huang, N.E., Shen Z, Long SR, Wu MC, Shih HH, Zheng Q, Yen NC, Tung CC and Liu HH (1998) The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis. Proceedings of Royal Society London, Series A, 454, 903–995

Huang, N.E., Wu M.C.L., Long, S.R., Shen, S.S.P., Qu, W., Gloersen, P., and Fan, K.L. (2003) A confidence limit for the empirical mode decomposition and Hilbert spectral analysis. Proceedings of the Royal Society Series A 459:2317–2345.

Koutsoyiannis D., and Foufoula-Georgiu E., (1993) A scaling model of storm hyetograph. Water Resour. Res., 29, 7, 2345–2361

Kuo, C., Gan, T., and Chan, S. (2013). Regional Intensity-Duration-Frequency Curves Derived from Ensemble Empirical Mode Decomposition and Scaling Property. *Journal of Hydrologic Engineering* 18(1), 66–74

Mandelbrot, B.B. (1974). Intermittent turbulence in self-similar cascades: divergence of high moments and dimension of the carrier. *Journal of Fluid Mechanics* 62:331–58.

Menabde, M.., Seed, A., and Pegram C.G.S. (1999). A simple scaling model for extreme rainfall. *Water Resources Research*, 35(1) 335-339

Mujumdar P.P. (2008). NPTEL lecture notes on stochastic hydrology, IISc Bangalore, L29.

Nhat L.M., Tachikawa, Y., Sayama, T. and Takara, K. (2008). Estimation of Sub-hourly and Hourly IDF Curves Using Scaling Properties of Rainfall at Gauged Site in Asian Pacific Region. *Annuals of Disaster Prevention Research Institute*, Kyoto University, No. 51 B, 2008

Olsson, J., and Berndtsson, R. (1998) Temporal rainfall disaggregation based on scaling properties. *Water Science Technology* 37(11), 73–9

Olsson J. (1998) Evaluation of a scaling cascade model for temporal rainfall disaggregation. *Hydrology and Earth System Science* 2, 19–30

Rehman, N., and Mandic, D.P. (2010) Multivariate empirical mode decomposition**.** *Proceedings of the Royal Society Series A*, 466(2117), 1291-1302.

Waymire E. C., and Gupta V. K., (1981) The mathematical structure of rainfall representation, 1, A review of stochastic rainfall models, 2, A review of the point processes theory, 3, Some applications of the point process theory to rainfall processes. *Water Resources Research*, 17(5), 1261–1294.

Waymire, E.C., Gupta V. K.., and Rodriguez-Iturbe I. (1984) A spectral theory of rainfall intensity at the meso-β scale. *Water Resources Research*, 20(10), 1453–146539

Veneziano D., Bras R. L, and Niemann J. D., (1996) Nonlinearity and self-similarity of rainfall in time and a stochastic model. *Journal of Geophysical Research*, 101, D21, 371–392.

Yu P.S., Yang T.C., and Lin C.-S. (2004) Regional rainfall intensity formulas based on scaling property of rainfall. *Journal of Hydrology*, 295, 1-4, 108–123

Stochastic Modelling for Inflow Prediction into Ukai Reservoir, India

Priyank J. Sharma[1], P. L. Patel[2], V. Jothiprakash[3]

[1]Research Scholar, Centre of Excellence on 'Water Resources and Flood Management', Department of Civil Engineering, Sardar Vallabhbhai National Institute of Technology Surat, Surat – 395007, India.

[2]Professor, Department of Civil Engineering, Sardar Vallabhbhai National Institute of Technology Surat, Surat – 395007, India.

[3]Professor, Department of Civil Engineering, Indian Institute of Technology Bombay, Mumbai – 400076, India.

Email:pjs230688@gmail.com, plpatel@ced.svnit.ac.in, vprakash@iitb.ac.in

ABSTRACT

The present study deals with investigating the ability of Auto Regressive Moving Average (ARMA) models in forecasting the monthly inflows into Ukai reservoir in the Tapi basin. The monthly data of Ukai reservoir for a period of 39 years (1975 – 2014) are used for the development of stochastic model. The orders of the candidate models has been decided from the autocorrelation function (ACF) and partial autocorrelation function (PACF) plots with all possible differencing schemes. The periodicities have been observed in the original data series from the autocorrelation function, and effectively removed by standardization rather than differencing of the time series. The parameters of different candidate models are estimated using Maximum Likelihood (ML) procedure, and the best modelhas been selected based onAkaike Information Criterion (AIC) and Bayesian Information Criterion (BIC). The selected model is validated by performing diagnostic check of the residuals, i.e. residuals should exhibit normal distribution and uncorrelated. The forecasted inflows using selectedARMA model are compared with the observed series, and it is found that the uncertainties in parameter estimations are minimal. The model performance is further assessed by employing different performance criteria. The model performed well for low flows as well as high flows (R = 0.99 and 0.79 respectively), while for moderate flows requires further investigation (R = 0.58). The long-term forecasting of inflows would be helpful in assessment of existing reservoir operating policies, and operation of reservoir on real-time basis.

Keywords: Stochastic modelling, parameter estimation, validation, forecasting, Tapi basin

1. INTRODUCTION

Time series modelling is one of the popular techniques used in stochastic modeling of hydrologic and climatic variables such as rainfall, runoff, temperature, etc. In practice, the hydrologists often face problems such as limited amount of data length as well as imperfect knowledge about future data. Thus, stochastic time series models provides help in data generation as well as data forecasting of hydrologic and climatic variables. Forecast of inflows into a reservoir form a vital component in planning, management and operation of available water resources. The inflow forecasting models are useful in diverse water resources application such as flood control, drought forecasting, optimal operation of reservoirs, and operation of hydropower plants (Yeh, 1985).Large number of time series models, in

time domain as well as frequency domain,were used in the past for forecasting hydrological time series (Yevjewich, 1963; Box and Jenkins, 1976). These time series models capture the statistical features of the historical data series and generate synthetic sequence of the same. If sufficiently longer length of the observed data are available, then, it is assumed that these data inherently capture the catchment characteristics (viz. topography, land use/ land cover, etc.), and thus, simple regression models would give better results (Magar and Jothiprakash, 2011).

Mujumdar and Kumar (1990) presented few case studies to investigate the best model for streamflow forecasting from the proposed candidate models of the Auto-Regressive Moving Average (ARMA) family. Their results reflected that the selected model was parsimonious and validated by testing the residuals. Mohan and Vedula (1995) employed seasonal Auto-Regressive Integrated Moving Average (ARIMA) model to forecast the monthly inflows into Bhadra reservoir, Karnataka, India. The forecasted values when compared with actual values revealed the adequacy of ARIMA model for long term inflow forecasting. Mishra and Desai (2005) employed ARIMA and SARIMA models for drought forecasting in Kansabati River basin, West Bengal, India which would be helpful for sustainable management of available water resources. Kote and Jothiprakash (2009) compared the performance of stochastic and Artificial Neural Network (ANN) models for inflow forecasting in Pawana reservoir, Maharashtra, India. The results of their study pointed that seasonal model gave improved performance compared to the monthly model since the zero inflow periods are not considered and only monsoon period flows are modelled. Mondal and Chowdhury (2013) developed deseasonalized ARMA model for generation of ten-daily flows of Brahmaputra River in Bangladesh for risk-based evaluation of proposed water resource project on the River.The present study investigates the suitability of a stochastic time series model of ARMA or ARIMA family to represent the monthly streamflows into Ukai reservoir, Gujarat, India. The forecasts from the developed stochastic model will be helpful in planning and management of water resources i.e. evaluating the existing reservoir policies as well as for drought management by assessing the command area requirements.

2. MATERIALS AND METHODS

2.1 Study area

The area selected for present study is Ukai reservoir in Tapi River basin, India. The basin lies between east longitudes of 72º 38' to 78º 17' and north latitudes of 20º 5' to 22º 3' and extends over an area of 65,145 km2. The location of study area is shown in Figure 1. The Ukai dam has a gross storage capacity of 7414.29 × 106 m3 with a water spread area of 60,096 hectares at full reservoir level (FRL) of 105.15

m. Ukai reservoir is a multipurpose project with the prime objective of drinking and industrial water supply, irrigation, hydropower generation as well as partial flood control.

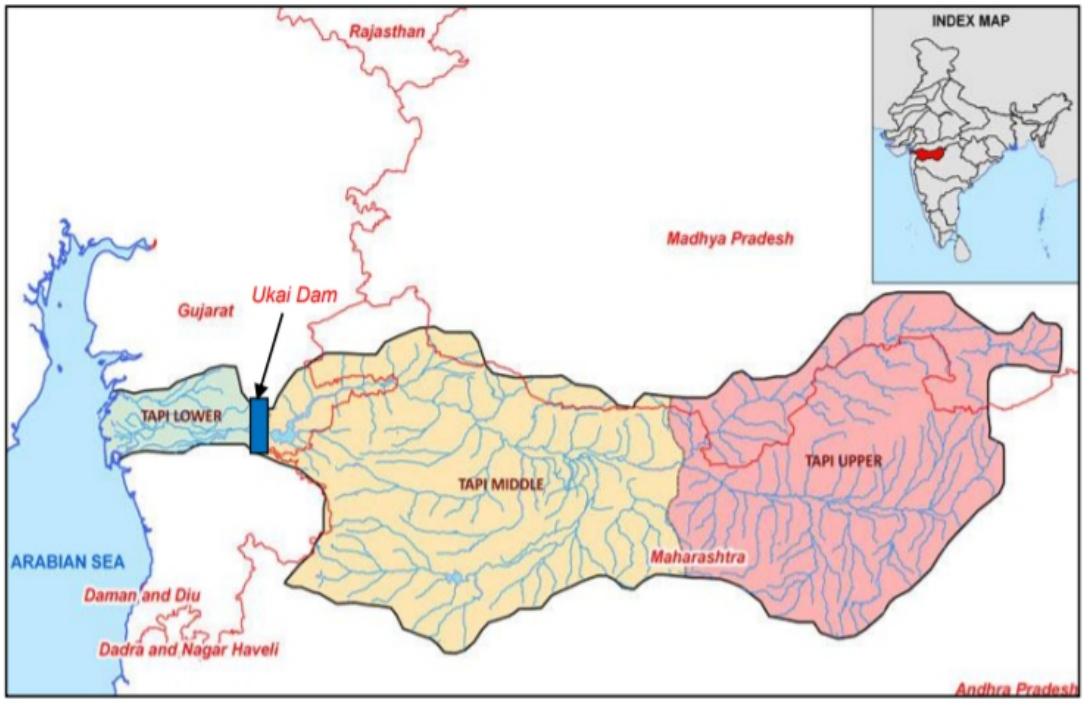


Figure 1 Location map of Tapi Basin (Source: iomenvis.nic.in)

The observed monthly inflow time series at Ukai reservoir for a period of 39 water years (June 1975 – May 2014) is shown in Figure 2.To determine the persistence structure of the time series, sample autocorrelation (ACF) and sample partial autocorrelation (PACF) are estimated from the original series and shown in Figure 3. It is evident from the ACF plot that the time series exhibits periodicity with little exponential decay, and PACF plot also shows significant correlation structure. Hence, before modelling the series, the series must be transformed by either standardization or differencing or normalization, and then, identify all possible models of the Box and Jenkins family. Further, a part of time series is used for training or calibration of the model and other part for testing or validation of the calibrated model. Generally, the calibration part of data length should be preferably long compared to validation part in order to determine the governing pattern of the time series in a better way. In present study, the entire data series was divided into two parts: a calibration set consisting of first 27 years (70% of data length), i.e. from June 1975 – May 2002, and a validation set consisting of remaining 12 years (30% of data length), i.e. from June 2002 – May 2014.

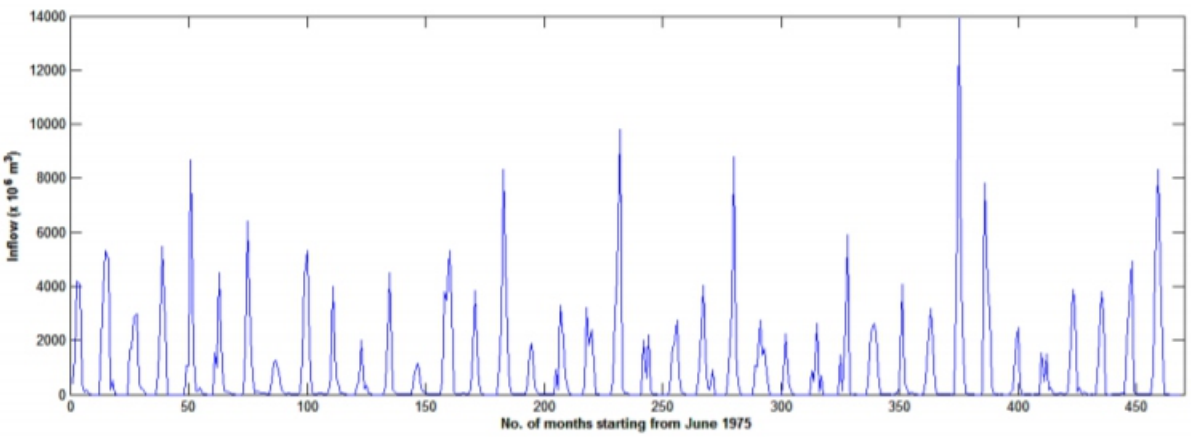


Figure 2 Time series of observed monthly inflows from June 1975 – May 2014

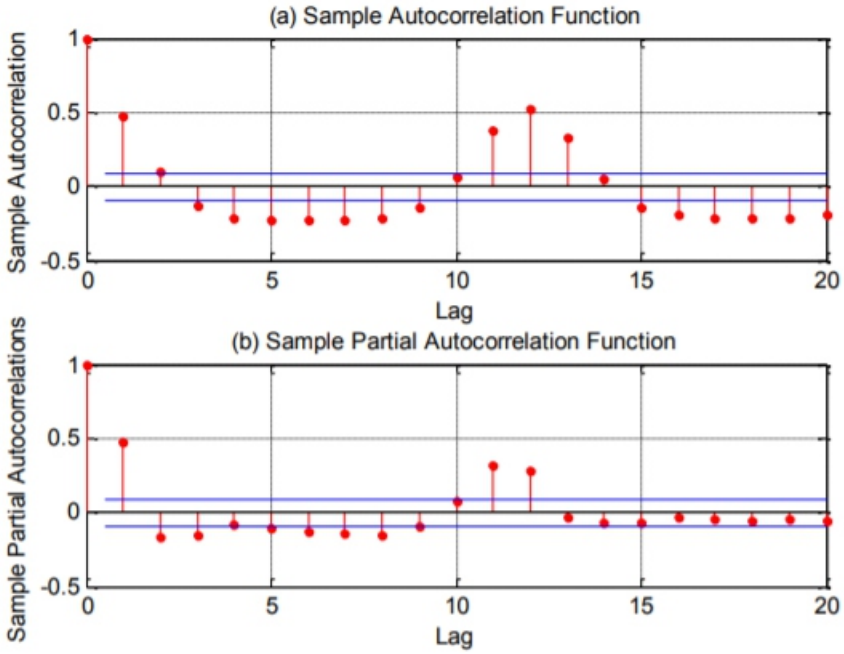


Figure 3 (a) Sample autocorrelation function (ACF), (b) Sample partial autocorrelation function (PACF)for original monthly time series

2.2 Stochastic models

The Autoregressive (AR) models can be efficiently coupled with moving average (MA) models to form a more generalized model of Box and Jenkins family i.e. autoregressive moving average (ARMA) models. In ARMA models, the current value of the time series is expressed as linear aggregates of p previous values and weighted sum of q previous deviations (original value minus fitted value of previous data) plus a random parameter (Mishra and Desai, 2005). However, ARMA models deals with stationary data only. If the data is non-stationary, it can be transformed into stationary by employing commonly used logarithm transformation or standardization procedure, and then they could be

modelled using ARMA. Another class of models called ARIMA models can be extended to non-stationary data also, which involves differencing of time series to attain stationarity. Moreover, ARMA (p,q) is the general case of ARIMA (p,d,q) where the differencing term (d) is zero. The general formulation of non-seasonal ARIMA (p,d,q) is represented as:

$$Y_t = c + \sum_{i=1}^{p} \phi_i Y_{t-i} - \sum_{j=1}^{q} \theta_j \varepsilon_{t-j} + \varepsilon_t \quad \ldots (1)$$

where, ε_t = random component, c = constant, $\phi_1, \phi_2, \ldots, \phi_p$ = autoregressive model parameters, $\theta_1, \theta_2, \ldots, \theta_q$ = moving average model parameters, and Y_t is resulted by d^{th} difference of the original series X_t through the preprocessing operation.

$$Y_t = X_t - X_{t-1} = (1 - B)X_t \quad \ldots (2)$$

where, B = backshift operator and its effect is to shift the argument one step behind.

If a given time series is non-stationary, i.e., it has periodicity associated with it, then it can be converted to a stationary time series by standardizing the same. Mathematically, it is expressed as:

$$Z_t = \frac{(X_t - \bar{X}_i)}{S_i} \quad \ldots (3)$$

where, X_t = original time series of time period t, $\bar{X}_i$ = mean of particular month or season i, and S_i = standard deviation of particular month or season i.

2.3 Methodology

The detailed methodology adopted in the present study is shown in Figure 4. The model is developed using monthly inflow data of 39 water years from June 1975 to May 2014. The model parameters arecalibrated using 70% of data length, i.e. from June 1975 to May 2002, while validation is done for the remaining 30% of data length. While selecting the calibration period, it is ensured that all type of data, i.e. low and peaks are captured well for model building. The best model is selected using AIC and BIC criteria. Further, the model performance for different inflow class is assessed using statistical indices like root mean square error (RMSE), mean absolute error (MAE) and coefficient of correlation ®.

3. RESULTS AND DISCUSSION

3.1 Model identification

The statistics of the original and transformed (standardized) monthly time series are shown in Table 1. Standardization ensures the removal of inherent periodicities in the process. The standardized monthly series is also shown in Figure 5, and corresponding ACF and PACF plots are shown in Figure 6. From Figure 6, it is clearly evident that the ACF is significant upto two lags and then decays down fairly quickly, while the PACF has only one significant lag. Thus, several combinations of autoregressive (p)

and moving average (q) terms are formed to investigate the best fit ARMA(p,q) model for the given time series.

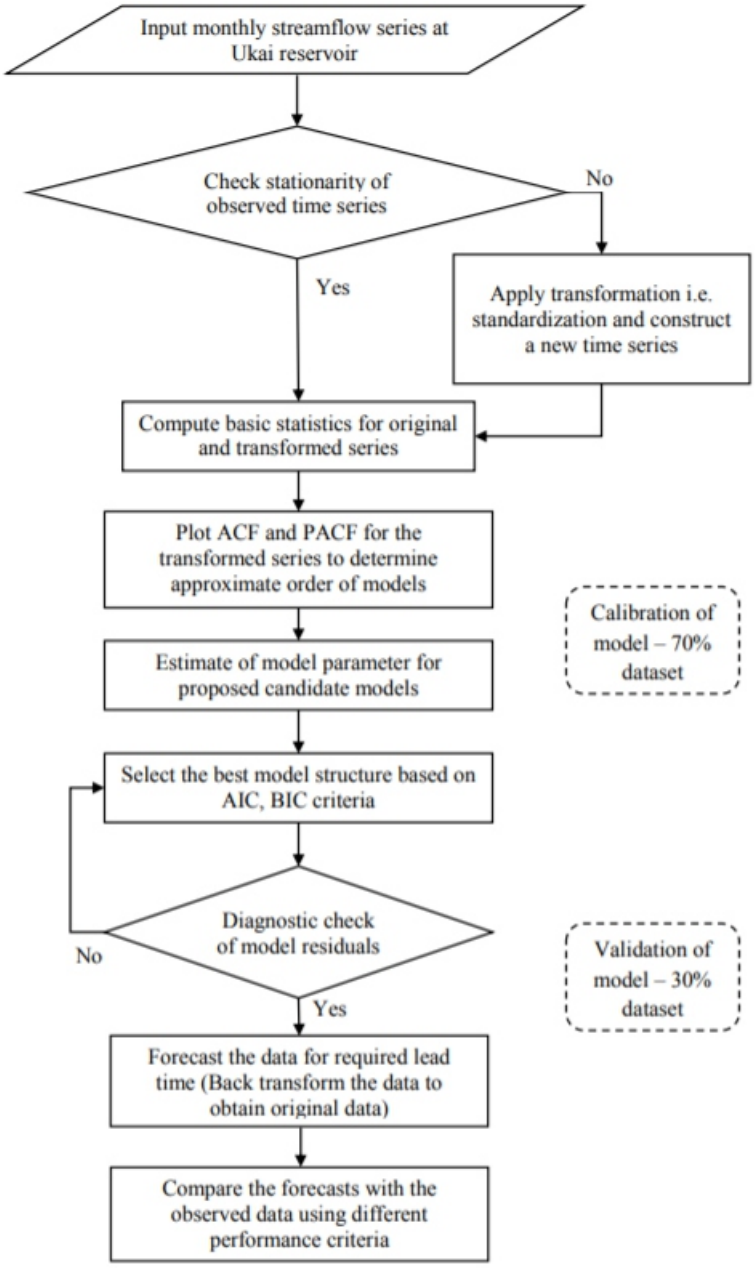


Figure 4Methodology for stochastic model development

Table 1 Statistical properties of observed and transformed monthly data series

Statistical properties	Entire dataset (Jun 1975 - May 2014)	Entire dataset [transformed] (Jun 1975 - May 2014)	Calibration dataset [transformed] (Jun 1975 - May 2002)	Validation dataset [transformed] (Jun 2002- May 2014)
Mean (10^6 m^3)	809.46	0.15	0.16	0.12
Standard Deviation (10^6 m^3)	1651.22	0.94	0.92	0.97
Skewness	3.28	2.17	2.22	2.11
Kurtosis	14.46	6.44	7.12	5.42
Coefficient of Variation	2.04	6.42	5.79	8.37

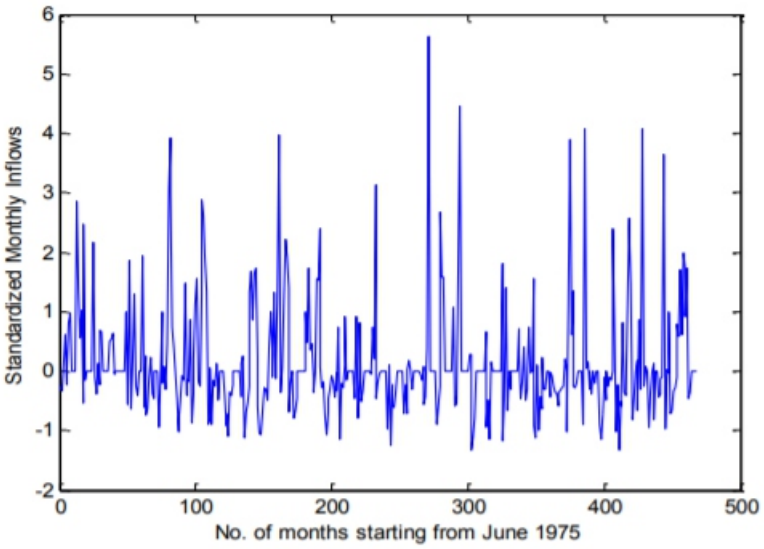


Figure 5 Standardized monthly time series

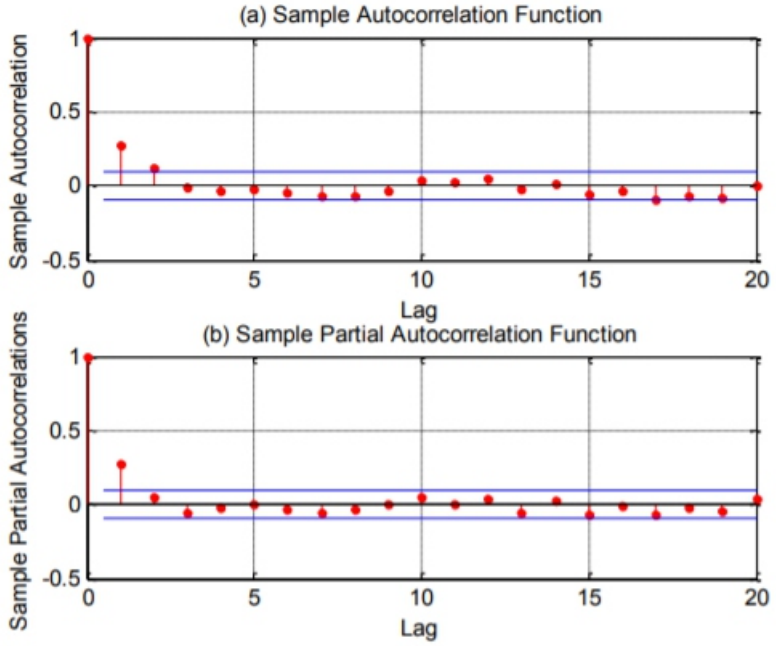


Figure 6 (a) Sample autocorrelation function (ACF), (b) Sample partial autocorrelation function (PACF) for standardized monthly time series

3.2Parameter Estimation

The model parameters for the proposed candidate models were estimated using the maximum likelihood objective function, and the best model was selected based on Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC). The AIC and BIC values for the candidate models are listed in Table 2.It is seen that ARMA (1,1) model has the least BIC value, while ARMA(3,2) has the least AIC value. The model that results in minimum value of AIC and BIC is selected as parsimonious model. However, the difference of AIC and BIC values for ARMA(1,1) model being very small and

considering the principle of parsimony, ARMA(1,1) model is tentatively selected and its parameters are described in Table 3. The model shall be validated before employing it for forecasting.

Table 2 Selection of best model using AIC and BIC

Candidate model structure	AIC	BIC
ARMA(1,1)	**829.14**	**841.59**
ARMA(1,2)	829.40	845.99
ARMA(1,3)	830.58	851.32
ARMA(2,1)	830.78	847.37
ARMA(2,2)	830.50	851.24
ARMA(2,3)	830.57	855.46
ARMA(3,1)	831.14	851.88
ARMA(3,2)	828.87	853.76
ARMA(3,3)	830.61	859.65

Table 3 Parameter estimates of selected ARMA(1,1) model

Parameter	Constant	ϕ_1	θ_1	Variance
Estimated value	0.0973	0.3856	–0.0442	0.7428

3.3 Diagnostic check

The selected model is diagnostically checked to examine whether the basic assumptions being used in building the model are valid for the selected model: (i) The residual series is normally distributed; (ii) No significant periodicity is present in the residual series; and (iii) The residual series is uncorrelated. Figure 7 shows the standardized residual series and a QQ plot of sample data versus standard normal. It is clearly evident that the residual is normally distributed as most of the data points fall on the standard normal line. Further, the ACF and PACF plots do not exhibit any periodic or correlation structure, see Figure 7 (c) and (d). Hence, it can be inferred that the selected ARMA(1,) model can be used for inflow forecasting.

3.4 Inflow forecasting

$$\hat{X}_{t+1} = \sum_{j=1}^{p} \phi_j X_{t-j} + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j} + c \qquad \ldots (7)$$

The one step ahead forecast error is given by:

$$e_{t+1} = X_{t+1} - \hat{X}_{t+1} \qquad \ldots (8)$$

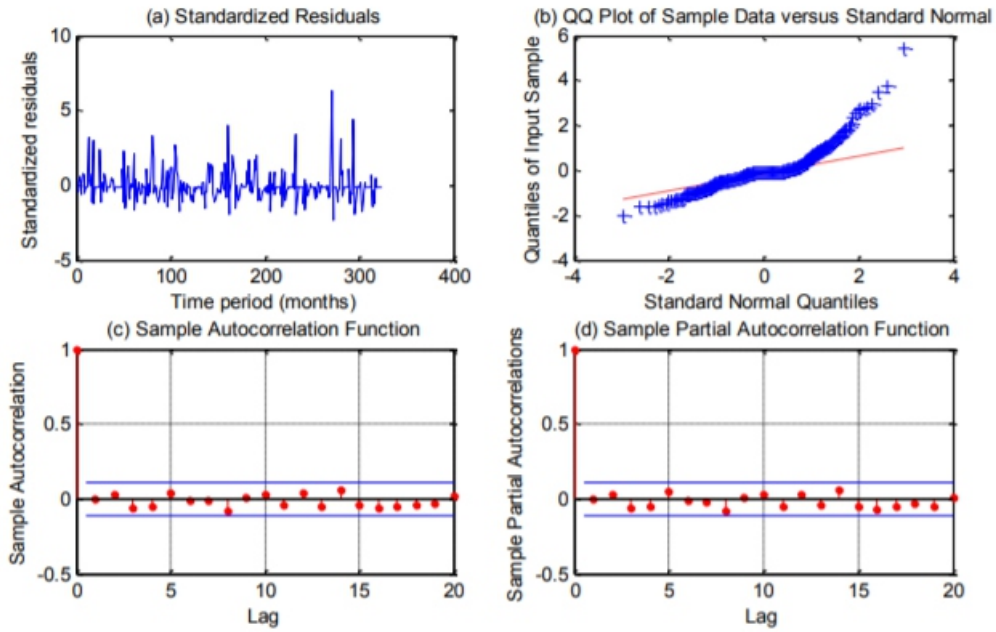


Figure 7 Diagnostic check of residuals of fitted ARMA (1,1) model

The error in the model is computed as the difference between observed and forecasted value at that time step. This error term is incorporated in the next time step as the random shock. Hence, the error is continuously updated in the model. The calibrated ARMA(1,1) model is selected for forecasting inflows for a period of 12 years (i.e. 30% of data set). Figure 8 shows a comparison between the observed and forecasted inflows during the period.

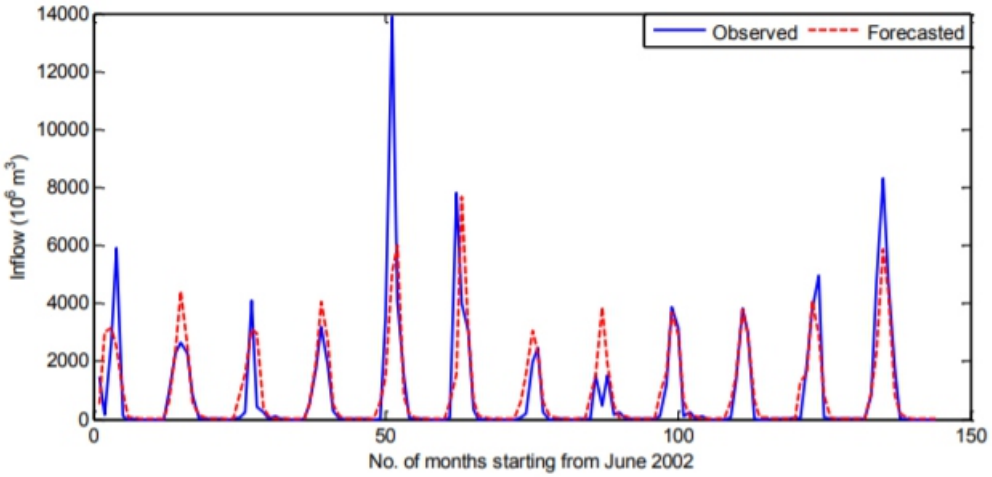


Figure 8 Comparison between observed and forecasted inflows

The performance of the model can be evaluated using different performance criteria such as root mean square error (RMSE), mean absolute error (MAE), coefficient of correlation (R) between the observed

and forecasted inflows. The RMSE and MAE are a good measures of indicating goodness-of-fit at moderate and high output values, while the R value quantifies the efficiency of a model in capturing the complex, dynamic and non-limear nature of the physical processes being modelled, and its value equal to unity shows perfection (Kote and Jothiprakash, 2009). The mathematical expression for different perofrmance criteria are listed in Appendix A.

The observed and forecasted series is classified in different inflow categories, viz., low, moderate and high. The summary of computed statistics is shown in Table 4. It is reflected from the results that coefficient of correlation is much better for low and high values compared to moderate values. However, the model is not able to simulate certain peak flows. Further, the detailed data analysis revealed that about 97.9 % of inflow volume is received during monsoon months, i.e., June to October. In other months, the flows are very low, and most of the times are zero. The zero values add to the skewness of the model, and hence, sometimes result in improper estimates. Hence, the seasonality effect should be taken in future to improve the accuracy of forecasts particularly for moderate flows.

Table 4 Performance criteria for ARMA (1,1) model

Inflow class (10^6 m^3)	Performance criteria		
	RMSE (10^6 m^3)	MAE (10^6 m^3)	R
Low inflow 0 - 30	139.89	50.33	0.99
Moderate inflow 31 - 3000	948.98	611.22	0.58
High inflow > 3000	3051.74	2007.75	0.79

4. CONCLUSION

The present study focussed on forecasting inflows into reservoir by employing time series model of Box and Jenkins family. The following are the specific findings of the present study:

(a) The observed inflow time series into Ukai reservoir indicates the existence of periodicity with little exponential decay. The original time series, in turn, is standardized for development of stochastic model for future inflow prediction into reservoir.

(b) The series of stochastic models were tried using AIC and BIC criteria for prediction of inflows into Ukai reservoir. Finally, ARMA(1,1) model (parameters: c = 0.0973, 1 = 0.3856, 1 = 0.0442, variance = 0.7428) has been proposed for prediction of inflow into the reservoir.

(c) The performance of the proposed model isvalidated using independent data set for period of 12 years (June 2002 – May 2014) using standard statistical performance indices like RMSE, MAE and R. The model performed satisfactory for low and high flows; however its performance is required to be

investigated for moderate flows.

ACKNOWLEDGMENTS

Authors wish to acknowledge the partial funds and necessary support received from Centre of Excellence (CoE) on "Water Resources and Flood Management", TEQIP-II, Ministry of Human Resources Development (MHRD), Government of India for conducting the study reported in the paper.

Appendix A

Sr. No.	Performance criteria	Equation
1.	Root Mean Square Error (RMSE)	$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(obs_i - calc_i)^2}{N}}$
2.	Akaike Information Criterion (AIC)	$AIC = Nln(RMSE) + 2(p+q)$
3.	Bayesian Information Criterion (AIC)	$BIC = Nln(RMSE) + (p+q)ln(N)$
4.	Mean Absolute Error (MAE)	$MAE = \frac{1}{N}\sum_{i=1}^{N}\lvert obs_i - calc_i\rvert$
5.	Coefficient of Correlation (R)	$R = \frac{\sum_{i=1}^{N}(obs_i - avgobs_i)(calc_i - avgcalc_i)}{\sqrt{\sum_{i=1}^{N}(obs_i - avgobs_i)^2}\sqrt{\sum_{i=1}^{N}(calc_i - avgcalc_i)^2}}$

where, N = length of time series data, p and q = the orders of the model, obs_i = observed inflow, $calc_i$ = calculated/forecasted inflow, $avgobs_i$ = average observed inflow, $avgcalc_i$ = average calculated/forecasted inflow.

REFERENCES

Box, G. E. P., and Jenkins, G. M. (1976). Time Series Analysis: Forecasting and Control, revised ed. Holden-Day, San Francisco, California, USA.

Kote, A. S., and Jothiprakash, V. (2009). Stochastic and artificial neural network models for reservoir inflow prediction. Journal of the Institution of Engineers (India): Civil Engineering Division, 90 (Nov), 25-33.

Magar, R. B., andJothiprakash, V. (2011). Intermittent reservoir daily-inflow prediction using lumped and distributed data multi-linear regression models. Journal of earth system science, 120(6), 1067-1084.

Mishra, A. K., and Desai, V. R. (2005). Drought forecasting using stochastic models. Stochastic Environmental Research and Risk Assessment, 19(5), 326-339.

Mohan, S., andVedula, S. (1995). Multiplicative seasonal arima model for longterm forecasting of inflows. Water resources management, 9(2), 115-126.

Mondal, M. S., and Chowdhury, J. U. (2013). Generation of 10-day flow of the Brahmaputra River using a time series model. Hydrology Research, 44(6), 1071-1083.

Mujumdar, P. P., and Kumar, D. N. (1990). Stochastic models of streamflow: some case studies. Hydrological Sciences Journal, 35(4), pp. 395-410.

Yeh, W. W. G. (1985). Reservoir management and operations models: A state-of-the-art review. Water Resources Research, 21(12), 1797-1818.

Yevjevich, V. M. (1963). Fluctuations of Wet and Dry Years, Part I-Research Data Assembly and Mathematical Models. Colorado State University Hydrology Paper 1, 55 pp.

Instructions for Authors

Essentials for Publishing in this Journal

1 Submitted articles should not have been previously published or be currently under consideration for publication elsewhere.

2 Conference papers may only be submitted if the paper has been completely re-written (taken to mean more than 50%) and the author has cleared any necessary permission with the copyright owner if it has been previously copyrighted.

3 All our articles are refereed through a double-blind process.

4 All authors must declare they have read and agreed to the content of the submitted article and must sign a declaration correspond to the originality of the article.

Submission Process

All articles for this journal must be submitted using our online submissions system. http://enrichedpub.com/ . Please use the Submit Your Article link in the Author Service area.

-

Manuscript Guidelines

The instructions to authors about the article preparation for publication in the Manuscripts are submitted online, through the e-Ur (Electronic editing) system, developed by **Enriched Publications Pvt. Ltd**. The article should contain the abstract with keywords, introduction, body, conclusion, references and the summary in English language (without heading and subheading enumeration). The article length should not exceed 16 pages of A4 paper format.

Title

The title should be informative. It is in both Journal's and author's best interest to use terms suitable. For indexing and word search. If there are no such terms in the title, the author is strongly advised to add a subtitle. The title should be given in English as well. The titles precede the abstract and the summary in an appropriate language.

Letterhead Title

The letterhead title is given at a top of each page for easier identification of article copies in an Electronic form in particular. It contains the author's surname and first name initial .article title, journal title and collation (year, volume, and issue, first and last page). The journal and article titles can be given in a shortened form.

Author's Name

Full name(s) of author(s) should be used. It is advisable to give the middle initial. Names are given in their original form.

Contact Details

The postal address or the e-mail address of the author (usually of the first one if there are more Authors) is given in the footnote at the bottom of the first page.

Type of Articles

Classification of articles is a duty of the editorial staff and is of special importance. Referees and the members of the editorial staff, or section editors, can propose a category, but the editor-in-chief has the sole responsibility for their classification. Journal articles are classified as follows:

Scientific articles:

1. Original scientific paper (giving the previously unpublished results of the author's own research based on management methods).

2. Survey paper (giving an original, detailed and critical view of a research problem or an area to which the author has made a contribution visible through his self-citation);

3. Short or preliminary communication (original management paper of full format but of a smaller extent or of a preliminary character);

4. Scientific critique or forum (discussion on a particular scientific topic, based exclusively on management argumentation) and commentaries. Exceptionally, in particular areas, a scientific paper in the Journal can be in a form of a monograph or a critical edition of scientific data (historical, archival, lexicographic, bibliographic, data survey, etc.) which were unknown or hardly accessible for scientific research.

Professional articles:

1. Professional paper (contribution offering experience useful for improvement of professional practice but not necessarily based on scientific methods);
2. Informative contribution (editorial, commentary, etc.);
3. Review (of a book, software, case study, scientific event, etc.)

Language

The article should be in English. The grammar and style of the article should be of good quality. The systematized text should be without abbreviations (except standard ones). All measurements must be in SI units. The sequence of formulae is denoted in Arabic numerals in parentheses on the right-hand side.

Abstract and Summary

An abstract is a concise informative presentation of the article content for fast and accurate Evaluation of its relevance. It is both in the Editorial Office's and the author's best interest for an abstract to contain terms often used for indexing and article search. The abstract describes the purpose of the study and the methods, outlines the findings and state the conclusions. A 100- to 250-Word abstract should be placed between the title and the keywords with the body text to follow. Besides an abstract are advised to have a summary in English, at the end of the article, after the Reference list. The summary should be structured and long up to 1/10 of the article length (it is more extensive than the abstract).

Keywords

Keywords are terms or phrases showing adequately the article content for indexing and search purposes. They should be allocated heaving in mind widely accepted international sources (index, dictionary or thesaurus), such as the Web of Science keyword list for science in general. The higher their usage frequency is the better. Up to 10 keywords immediately follow the abstract and the summary, in respective languages.

Acknowledgements

The name and the number of the project or programmed within which the article was realized is given in a separate note at the bottom of the first page together with the name of the institution which financially supported the project or programmed.

Tables and Illustrations

All the captions should be in the original language as well as in English, together with the texts in illustrations if possible. Tables are typed in the same style as the text and are denoted by numerals at the top. Photographs and drawings, placed appropriately in the text, should be clear, precise and suitable for reproduction. Drawings should be created in Word or Corel.

Citation in the Text

Citation in the text must be uniform. When citing references in the text, use the reference number set in square brackets from the Reference list at the end of the article.

Footnotes

Footnotes are given at the bottom of the page with the text they refer to. They can contain less relevant details, additional explanations or used sources (e.g. scientific material, manuals). They cannot replace the cited literature.
The article should be accompanied with a cover letter with the information about the author(s): surname, middle initial, first name, and citizen personal number, rank, title, e-mail address, and affiliation address, home address including municipality, phone number in the office and at home (or a mobile phone number). The cover letter should state the type of the article and tell which illustrations are original and which are not.

Address of the Editorial Office:

Enriched Publications Pvt. Ltd.
S-9,IInd FLOOR, MLU POCKET,
MANISH ABHINAV PLAZA-II, ABOVE FEDERAL BANK,
PLOT NO-5, SECTOR -5, DWARKA, NEW DELHI, INDIA-110075,
PHONE: - + (91)-(11)-45525005

Notes: