ESTIMATION OF DISCHARGE AND TOTAL WATER LEVEL AT YEDGAON DAM USING DATA DRIVEN TECHNIQUES

Dr. Preeti S. Kulkarni1, Prof. Shreenivas Londhe2, Nikita Sainkar3, Sayali Rote4
1,2 Associate Professor and Professor, Department of Civil engineering, Vishwakarma Institute of Information Technology, Pune, 411048, India
3,4 Student of M.Tech-Water Resources and Environmental Engineering, Department of Civil engineering, Vishwakarma Institute of Information Technology, Pune, 411048, India. Email: sayali_219m0029@viit.ac.in

Abstract:- A reservoir operation planning using Data driven Techniques is gaining its momentum in hydrological area with good prediction and Estimation capabilities. The present work aims at using the 5 years data of Water Level to estimate the discharge and water level at the Yedgaon dam which is like pick up weir having its own yield and storage. It receives water from Dimbhe (though DLBC), Wadaj (through MLBC), Manikdoh (through river) and through Pimpalgaojoge (through river), in the Kukadi project of Maharashtra State, India. 4 different models were developed to estimate the water level using the Data Driven Techniques: M5 Model Tree, Support Vector Regression, Multi Gene Genetic Programming and Random Forest. The Accuracy of the developed models is assessed by the values of coefficient of correlation, coefficient of efficiency, mean absolute error and root mean squared error and comparison is done between actual values and Predicted values. The results indicated that the MGGP model was superior as compared to other techniques with correlation coefficient as 0.86 with an advantage of a single equation to estimate the waterlevel.

Keywords: Yedgaon dam, Kukadi Project, Data Driven techniques, Modelling.

1. Introduction
Among the all other basic needs of human beings, water is considered to be the prime natural resources. The amount of water available on earth is about 1400 million km$^3$, of which only 2.7% is available as freshwater (Sinha, 2005). Hence, only a small fraction of can be utilized and is available for all the purposes. A reservoir operation plan is designed to achieve the maximum benefits from the storage capacity. The flow characteristics of the stream i.e. history of its past performances should be known to plan the reservoir operation. A reservoir operation is described as the amount of water to be released from the storage at any time depending upon the state of the reservoir, level of demands and any information about the likely inflow in the reservoir. Considering the monsoon season of about three to four months in India, there occurs the necessity of creation of large storage and utilization of runoff. The regional variation is also extreme from about 100 mm in Rajasthan to...
11000 mm in Meghalaya of northeast India, causing the drought prone situation at some part and flood condition at the other in same time (Sinha, 2005). In order to overcome this drought flood situation in the country, it is necessary to plan and construct reservoirs and other water storage plants for conservation and utilization of water resources to its maximum benefits.

For the present study, a reservoir operation is to be performed on the Kukadi project of Maharashtra state, India. Regression analysis can be done using data sets at five different stations namely, Dimbhe, Manikdoh, Wadaj, Pimpla ga joge and Yedgaon. The Yedgaon dam is like pick up weir having its own yield and storage. It also receives water from Dimbhe dam (through DLBC), Wadaj dam (through MLBC), Manikdoh dam (through river) and through Pimpla ga joge dam (through river) (Birajdar, 2012).

In recent years, Data Driven Techniques have found increasing applications in Hydrology and Water Resources Engineering oriented towards estimating or predicting values of various parameters like stream flow, stage, discharge, etc. Mujumdar et.al (1997) developed a short- term real-time reservoir operation model for irrigation of multiple crops. The model was applied to a case study of an existing reservoir in India. Babovic et.al (2002) used data driven techniques such as genetic programming (GP) to model runoff on the basis of available hydro meteorological data. Solomatine et al. (2004) investigated flexibility and optimality in M5 model tree by proposing two new algorithms, namely M5flex and M5opt. For experiments six hydrological data sets and five benchmark data sets were used. Mathur et.al (2012) studied support vector machines (SVMs) construct a ground water level forecasting system. Further the proposed SVM-PSO model was employed in estimating the groundwater level of Rentachintala region of Andhra Pradesh in India. Lily et.al (2013) used the monthly groundwater level data of about 20 years from the Hamadan Bahar Plain, west of Iran, based on peizometric height related to hydrologic years. The support vector machine (SVM) was used to predict groundwater level. Sivapragasam et al. (2015) used Genetic Programming (GP) to model spatial variation of groundwater in Arjuna Nadhi sub basin region. Bing et.al (2016) aimed to develop an efficient model for forecasting Lake water level variations, exemplified by the Poyang Lake (China) case study. A random forests (RF) model was first applied and compared with artificial neural networks, support vector regression, and a linear model.

The aim of the present work is develop a model to estimate outflow at Yedgaon dam using outflows at Manikdoh, Wadaj, Dimbhe, Pimpla ga joge dams and analyze the results using various soft computing tools such as M5 Model Tree, Support Vector Regression, Multi Gene Genetic Programming and Random Forest. Accuracy of the models is assessed by the values of coefficient of correlation, coefficient of efficiency and root mean squared error and comparison is done between actual and predicted values of discharge and the total water level. The outline of the paper is as follows. Section 2 follows the salient features of the data used to train and test the Data Driven Technique model. Section 3 provides the brief information about all the soft computing tools. After this model, results along with a discussion of the reliability of the predictions are presented next section. Concluding remarks are given at the end.

2. Study Area and Data
The present study aims to predict daily discharge values and total water level at Yedgaon. The study area consists of Western Ghats of Maharashtra of Sahyadri hill range where five dams of the Kukadi integrated system are situated. Study area also extends up to command area in three districts of Pune, Solapur and Ahmednagar.

Figure 1 given below shows the Google map for the Kukadi complex which gives the general idea of the study area we have considered for the present work.
The total irrigable command area under the entire Kukadi project is 156278 Ha. For this study, Daily Discharge Data and total water level Data was obtained from the Irrigation Department (Kukadi Irrigation Division No.1, Narayangaon) from 1st June 2015 to 31st August 2019 each for Manikdoh, Dimbhe, Wadaj, Pimpalgaojoge and Yedgaon for Kukadi Irrigation Project. The total no. of observations for daily discharge data available for each station is 592 each and that for total water level is 1548 each. Table 1 below shows Statistical parameters of the daily discharge (Model 1) and Table 2 show Statisticalparameters of total water level (Model 2) data for the study sites.

Table 1 Statistical parameters of the daily discharge for the study sites

| Stations     | Range of parameters | Average value | Standard Deviation |
|--------------|--------------------|---------------|--------------------|
| Manikdoh     | 0 to 1250          | 276           | 456.30             |
| Wadaj        | 0 to 364           | 69            | 97.77              |
| Dimbhe       | 0 to 650           | 480           | 192.40             |
| Pimpalgaojoge| 0 to 1450          | 258           | 452.09             |
| Yedgaon      | 138 to 11966       | 1445          | 881.01             |

All values of discharge are in cusecs.

Table 2 Statistical parameters of the total water level data for the study sites.

| Stations     | Range of parameters | Average value | Standard Deviation |
|--------------|--------------------|---------------|--------------------|
| Manikdoh     | 681.75 to 709.99   | 694.95        | 8.42               |
| Wadaj        | 699.69 to 717.53   | 712.57        | 4.43               |
| Dimbhe       | 682.55 to 719.15   | 705.30        | 11.83              |
| Pimpalgaojoge| 673.01 to 686.62   | 682.56        | 3.31               |
| Yedgaon      | 634.22 to 641      | 694.95        | 8.42               |

All values of Total Water Level are in Meters.
3. Methodology
In the current study, four models were developed using each of the techniques viz. MT, MGGP, SVR and RF to predict the discharge and total water level at Yedgaon. Objective of the present study is to correlate upstream and downstream stations using Data driven technique (M5 Model Tree, SVR, MGGP and Random forest). After obtaining the maximum and minimum daily values, two models were formed for predicting discharge and total water level on daily basis by MT, SVR, MGGP and RF. To find out the predicted values, M5 Model tree (MT) tool, Support vector regression (SVR) and Random forest is used by using WEKA (3.9.4) and Multi Gene Genetic Programming (MGGP) tool is used by using MATLAB R2019a. The following table 3 shows the model (M) and their input (I1, I2, I3 and I4) and output parameters.

Figure 2 given below shows the methodology used for this particular study using the various data driven techniques (M5 Model Tree, Support Vector Regression, Multi Gene Genetic Programming and Random Forest).

Table 3 Model formulation

| M  | I1      | I2      | I3      | I4      | Output       |
|----|---------|---------|---------|---------|--------------|
| 1  | Manikdoh| Wadaj   | Dimbhe  | P joge  | Yedgaon      |
| 2  | Manikdoh| Wadaj   | Dimbhe  | P joge  | Yedgaon      |

Fig. 2: Flowchart for Methodology.
4. Results and Discussion

Consolidated summary of model performance is described in Table 4.

| Tool          | $r$  | RMSE   | MAE   |
|---------------|------|--------|-------|
| M5 MT         | 0.46 | 0.02   | 1331.22 | 1367.34 | 495.21 | 1.9021 |
| SVR           | 0.37 | 0.06   | 1431.64 | 1.100   | 514.12 | 0.9036 |
| RF            | 0.36 | 0.09   | 1415.22 | 1242.28 | 345.59 | 0.9944 |
| MGGP          | 0.47 | 0.06   | 1341.02 | 1.4152  | 239.49 | 0.7576 |

The first model (Discharge) has 592 values each for 4 inputs and a single output. All the four data driven techniques (M5 Model Tree, Support Vector Regression, Multi Gene Genetic Programming and Random Forest) estimated output values reasonably well in testing. The values of coefficient of correlation $r$ are 0.46, 0.17, 0.16 and 0.47 for M5 MT, SVR, RF and MGGP respectively. Fig. 3 shows the combined graph for all the 4 soft tools and the actual values of the output obtained analytically.

However, among these techniques, MGGP showed better results compared to other three, i.e. $r = 0.47$, RMSE = 1847.02, MAE = 339.49.

The second model (Total Water Level) has total 1548 values of TWL of Manikdoh, Dimbhe, Wadaj and Pimpalgaonjoge as input with Yedgaon as the output. All the four data driven techniques (MT, SVR, MGGP &RF) estimated the output reasonably well in testing with ($r$=0.80 for MT, $r$=0.84 for SVR, $r$=0.85 for MGGP and $r$=0.78 for RF). Fig. 4 shows the combined graph for the four techniques.

However among all MGGP shows the better result compared to other techniques giving ($r = 0.85$, RMSE = 1.4152 and MAE = 0.7576).

Other error measures MAE and RMSE indicate the similar performance for both the models with MGGP giving the better results among all the techniques. Other error measures i.e. RMSE, MAE indicate a similar performance for both models with MGGP a better in all.

Fig. 3: Combined graph of actual vs. predicted Discharge values using all soft tools
Fig. 4: Combined graph of actual vs. predicted TWL values using all soft tools.

Figure 5 given below shows the scatter plots for model 1 and model 2 given by MGGP as it showed the better results than the other three.

Fig. 5: Scatter plots for MGGP model 1 and 2 for actual vs. predicted values.

The scatter plots shown above for model 1 and model 2 are varying. The values of inputs in model 1 i.e. the outflows at Manikdoh, Dimbhe, Wadaj and Pimpalgaojoge are somewhere either low or zero causing the extreme variation in the predicted output at Yedgaon. Ultimately it gives poor relation in the scatter plot. Hence, it was decided to consider the total water level for the further study.

The consolidated summary of the results is that the values of correlation coefficient ‘r’ between the estimated total water level and observed total water level is between 0.78 (M2, RF) and 0.85 (M2, MGGP). The mean absolute error varies between 0.99m (M2, RF) and 0.7576m (M2, MGGP). The root mean square (RMSE) showed a variation of 1.2425m (M2, RF) to 1.4152m (M2, RF).

5. Conclusions
In the present work, development of a model using Data Driven Techniques like M5 Model Tree, Support Vector Regression, Multi Gene Genetic Programming and Random Forest is done on the Kukadi Project of Maharashtra state, India. Present work aims at estimation of outflow at Yedgaon using outflow at Manikdoh, Wadaj, Dimbhe, and Pimpalgaojoge and analyze the results. For the model formulation, Discharge data and the total water level data has been used. The performance of the model was evaluated using scatter plots and error measures like coefficient of correlation, root mean squared error, mean absolute error. Using the results, the values are compared. The values of
coefficient of correlation are good in case of model 2 as compared to model 1 as they are nearest to one. All data driven techniques showed good results in case of model 2. The results indicated that the Multi Gene Genetic Programming technique was superior as compared to others for both the models. For Model 1, the ‘r’ values varied from 0.1649 to 0.4631 and that for model 2 it varied from 0.7858 to 0.8479 indicating the superiority of Data Driven Techniques in model 2. Hence, it can be concluded that these data driven techniques can be adopted for the estimation of Discharge and Total Water Level.

6. References
1. Babovic V. Keijzer M., —Rainfall Runoff Modelling Basedon Genetic Programming!, Nordic Hydrology, 33 (5), 2002, 33 1-34.
2. Bing Li, Guishan Yang, Rongrong Wan, Xue Dai and Yanhui Zhang, “Comparison of random forests and other statistical methods for the prediction of lake water level: a case study of the Poyang Lake in China”, Hydrology Research, 2016.
3. Birajdar C.A., Siddamal U.V., “Sustainable Development of Irrigation in Kukadi Integrated Project by transfer of Surplus Water From one Reservoir to Other”, WIT Transactions on Ecology and the Environment, Vol 168, © 2012 WIT Press.
4. Lily T., Alireza R., Abbas M., “Prediction the groundwater level of Hamadan-Bahar Plain, west of Iran using support vector machines”, Journal of research in Health Sciences.
5. Mujumdar P.P. And Ramesh T. S. V., “Real- Time Reservoir Operation for Irrigation”, Water Resources Research, Vol. 33, No. 5, Pages 1157– 64, May 1997.
6. Sinha S.K., “Real Time Integrated Operation of Reservoirs”, Government of India, Central Water Commission, Reservoir Operation Directorate, April-2005.
7. Solomatine D. P. & Yunpeng Xue (2004). “M5 Model Trees and Neural Networks: Application to Flood Forecasting in the Upper Reach of the Huai River in China”, Journal of Hydrologic Engineering © Asce / November/December 2004: 491-501.
8. Sudheer Ch., Shashi Mathur, “Groundwater level forecasting using SVM-PSO”, Int. J. Hydrology Science and Technology, Vol. 2, No. 2, 2012.