Quality of Data Set In Modeling Work: A Case Study in Urban Area for Different Inputs Using Fuzzy Approach

Surendra H. J, Paresh Chandra Deka

Abstract— **All water resources data are ambiguity in nature containing imprecise information, Noise, etc. when these imprecise, noisy data used for modeling, the outcome result may have more error, resulting in discarding the performance of the model. Many researchers may think about Hybrid approaches to improve the accuracy of the model when the single approach fails to get the desired result. This indicates that, quality of the data set is an important approach for any modeling work. If the quality data is not available then it is necessary to make data more precise, noise free to improve the accuracy. In this research work an attempt is made to realize the importance of the quality data for modeling work. For this purpose Fuzzy logic approaches is chosen, since it is capable to handle the imprecise, noisy type of data. Later two different data set such as raw data and normalized data were employed to show the performance of the model for different input scenarios in an urban area. Results revealed that it is necessary to adopt the procedure to improve the statistical property of the data, either by Hybrid approaches or by any processing techniques. Performance of the model is evaluated using indices such as mean absolute error (MAE) and prediction error (PE).**

Index Terms— **Climatic Variables, Fuzzy logic, Normalized data, raw data, Quality Data, Water consumption**

I. INTRODUCTION

The dynamics aspects of water resources management are very difficult to understand since it is associated with various parameters which control the situation in an urban area. Estimating and forecasting of water demand in an urban area is very important as the population is mainly dependent on public water supplies. Hence it is very important to develop the effective forecasting model. However effective water forecasting model requires quality type of data. Growth of areas in urban region will increases stress on water, hence water demand forecasting will highlights the importance for more effective planning and design. Climatic variation could be the one of the determinants of water consumption. The connection between climatic variables and water consumption were important in the region of semi-arid and arid these climatic variables are non-stationary, time varying in nature. Hence it is necessary to normalize the data to improve the accuracy. Since soft computing fuzzy logic method is a data driven model, quality of the data reflects the accuracy of the model. Careful selection of input parameter should be done properly before applying model to particular area. Approach of soft computing technique such as fuzzy logic exploit tolerance for uncertainty, robustness, and human like thinking to achieve the true reality due to its guiding principles. In this research work Mamdani fuzzy inference system is used to develop model which is trained based on climatic data to a certain period and corresponding prediction model were developed for the same period. Present modeling is more focus on water demand forecasting from the past records of climatic variables. Due to variation of climatic variables, challenges on water demand forecasting will be more. Much Fuzzy logic system focus on raw data, we argue that this approach may produce a reasonable analysis and prediction. It is not optimal for non-stationary time series which will impair the result hence it is necessary to normalize the data to improve the accuracy.

Manuscript received March 18, 2015.

Surendra H J , Research scholar , Applied Mechanics and Hydraulics department, NITK Surathkal, Karnataka India.

Dr.Paresh Chandra Deka (Associate professor), Department of Applied Mechanics and Hydraulics, NITK Surathkal, Karnataka India.

II. RELATED LITERATURE REVIEW

There are different approaches to water demand forecasting including statistical or mathematical techniques. Aijun et al., (1996) used a rough set approach for water demand prediction to analyze a set of training data and generate decision rules and it was found to be useful for incomplete and deterministic information. Durga Rao (2005) used multicriteria spatial decision explanatory variables for water demand forecasting. Hongwei, et al., (2009) used system dynamic approach for water demand forecasting based on sustainable utilization strategy of the water resources. Herrera et al., (2010) developed predictive models for forecasting hourly water demand using ANN, projection pursuit regression (PPR), multivariate adaptive regression splines (MARS), random forest and support vector regression (SVR).They also used Monte Carlo simulation designed to estimate predictive performance of model obtained on data set and found that support vector regression model is most accurate one followed by MARS, PPR.

Although Conventional time series modeling methods have served the scientific community for a long time and they provide reasonable accuracy, but suffer from the assumption of stationery and linearity (Kermani & Teshnehlab., 2008). Many new methodologies are developed for modeling the data but current trend seems to be model the data rather than physical process. For modeling the data, artificial intelligence techniques (AI) such as fuzzy logic (FL), artificial neural network (ANN) and adaptive neuro fuzzy inference system(ANFIS) are probably the most attractive techniques among the researchers, which is capable of handling imprecise, fuzzy, noise and probabilistic information to solve complex problem in an efficient manner. Altunkaynak et al.,

(2005) used fuzzy logic approach for water consumption prediction of the Istanbul city, using Takagi Sugeno method for time series data by considering only one lag as input for the analysis. Kermani & Teshnehlab., (2008) used normalized data for water consumption prediction using ANFIS method and also further, auto regressive model is employed for the analysis and they found that ANFIS model is better than autoregressive model. Yurdusev & Firat., (2009) used ANFIS method to forecast monthly water consumption modeling and they have adopted cross correlation method for selection of the input variables. Sen & Altunkaynak., (2009) used Mamdani inference system for modeling of drinking water prediction using different fuzzy sets and rules in the analysis. Also, there were many reports of using ANN in forecasting water demand (Babel & Shinde., 2011, Jain et al 2001, Firat et al 2009 and 2010)

III. Fuzzy Logic

Fuzzy logic is a mathematical tool which provides a simple way of approaching problem to obtain definite conclusion based on imprecise, noisy type data set. Mapping from given input to output will be done based on membership function and rules criteria. Once it is mapped then finally defuzzification is carried out to convert linguistic variables into crisp variables, which is an exact opposition of fuzzification process. Mamdani method was built using fuzzy set theory and employed widely in various field. The linguistic variables used for the analysis are Very low, low, medium and high. Centroid defuzzification method is used to convert linguistic term into crisp output. Figure 1 shows the structure of Mamdani fuzzy inference system. Table 1 shows the rules criteria adopted in the analysis. Figure 2 shows the structure of input and output combination given in fuzzy logic.

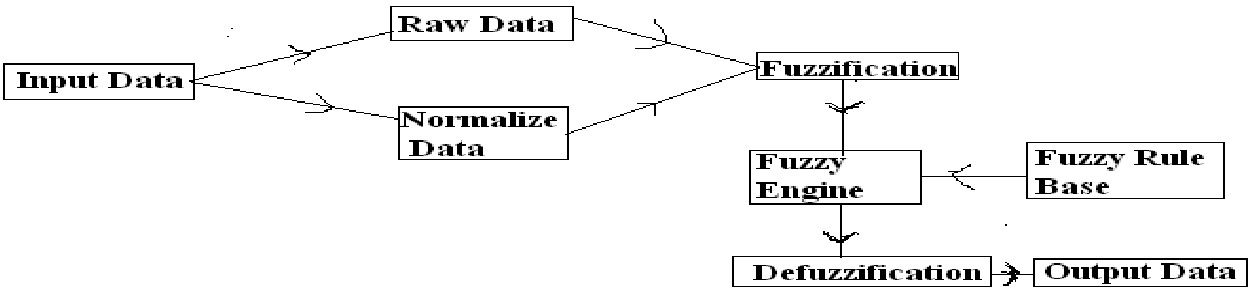


Fig 1: Structure of Mamdani fuzzy Inference System used for analysis

Table 1: Different rules used for analysis

Rule. No	Rule types
R1	If RF is very low, Tmax is very low, Tmin is very low then WD is very low
R2	If RF is low, Tmax is low, Tmin is low then WD is low
R3	If RF is medium, Tmax is medium, Tmin is medium then WD is medium
R4	If RF is high, Tmax is high, Tmin is high then WD is high

Where: - RF; Rainfall, Tmax: Maximum Temperature, Tmin: minimum Temperature, RH: Relative Humidity, WD: Water demand

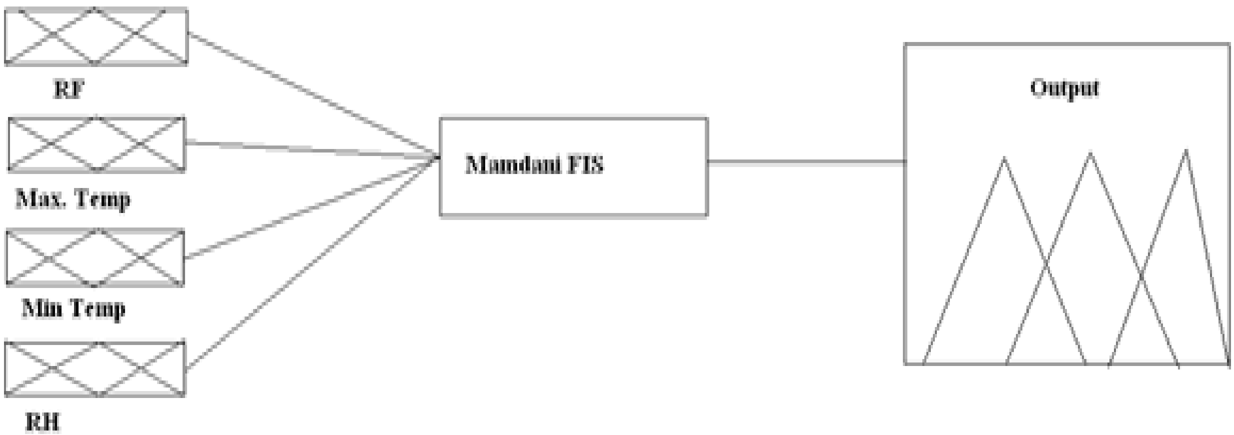


Fig 2: Structure of Input and Output combination used for analysis

IV. STUDY AREA

Due to expansion of the city, increased population and other factors Yelahanka city experience the difficulty to meet the present demand of water. Hence the present city covers the fourth ward of Yelahanka city (Latitude $13^0$06'30",

Longitude $77^0$34'15") which is a sub-urban of Bangalore in the state of Karnataka. City experiences a total rainfall of 1140mm from May to September. Winter from November to February with temperature ranges from 14^0 to $24^0$and summer season starts from March to May with a temperature 20^0 to 35^0. The Climatic variables such as Rainfall, Maximum and Minimum Temperature, Relative Humidity used for this research work were collected on monthly basis from Karnataka state disaster management cell; Bangalore. Water consumption records were collected from BWSSB for a period of ten years. Structure of the model used in the analysis is as shown in the table 2. The selection of input and output variables is done based on correlation coefficient. The correlation coefficient of all variables was shown in the table 3. Location of the Study area is shown in the figure 3. The Monthly variations of Rainfall, Maximum Temperature, Minimum Temperature, Relative Humidity and water consumption were shown in the figure 4, 5, 6 and 7. Figure shows that variables used for the analysis are time varying, non-stationary for this type of data fuzzy logic method is suitable. But to improve the accuracy of the model all the data were normalized using the equation,

$$Xnormalized = \frac{XRaw - Xmaximum}{XMaximum - Xminimum}$$

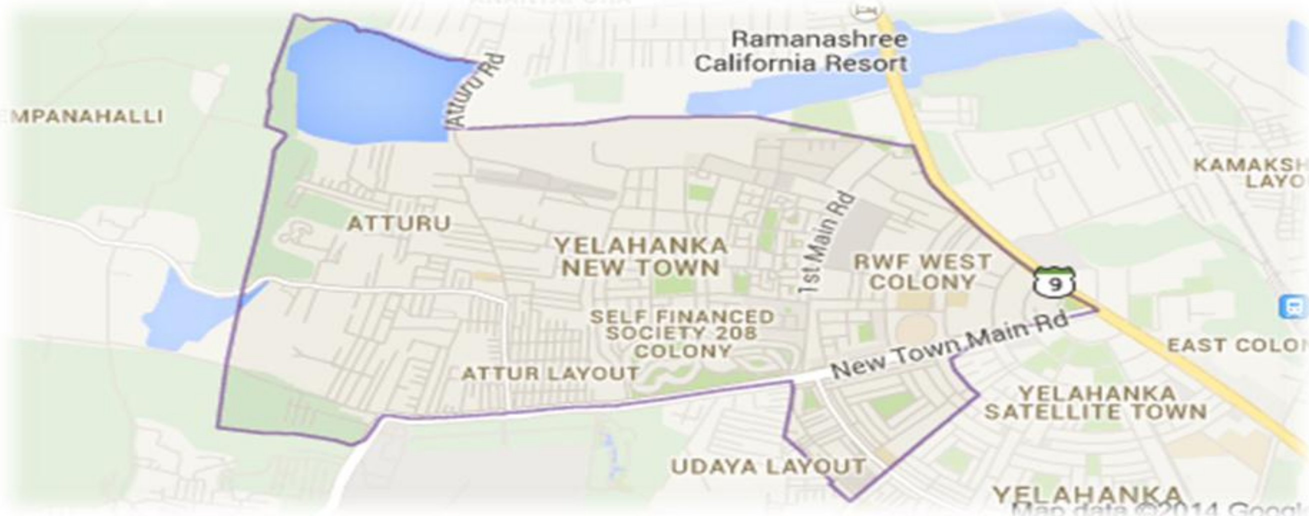


Fig 3: Location of study area (Source: Google map)

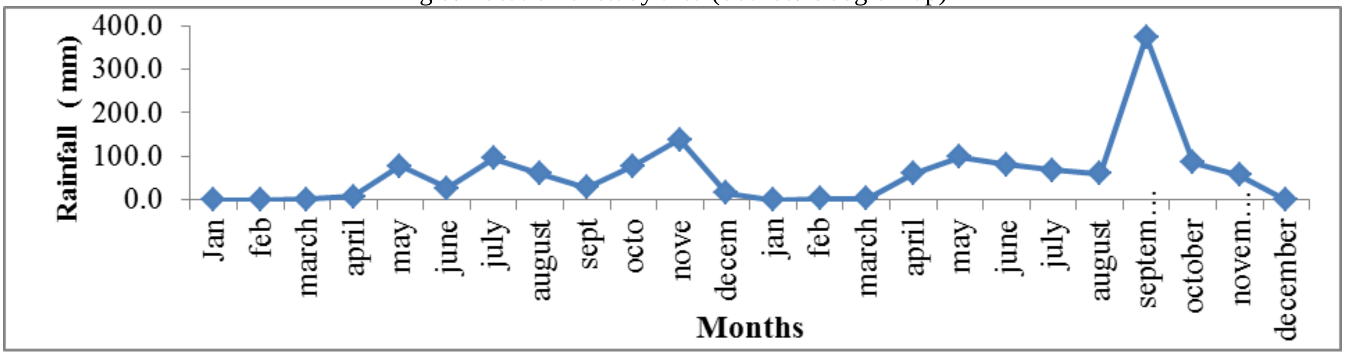


Fig 4: Variation of Monthly Rainfall

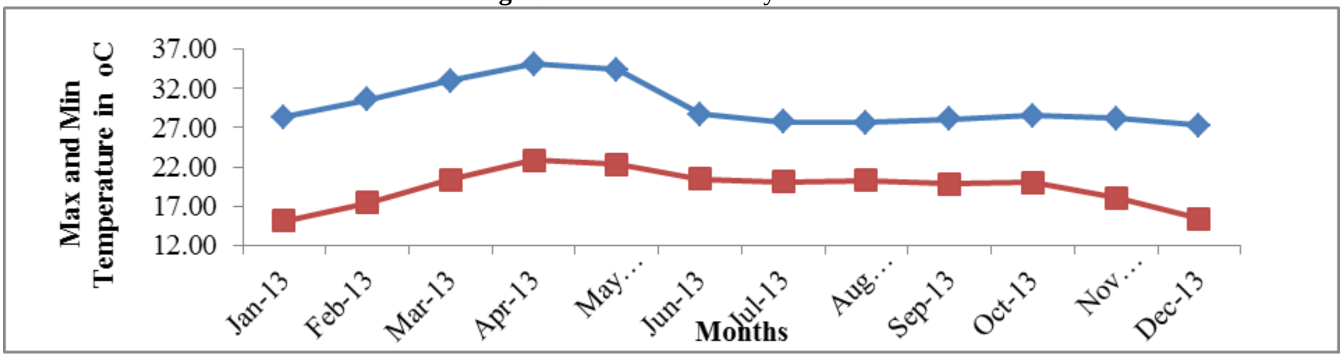


Fig 5: Monthly Maximum and Minimum temperature variation

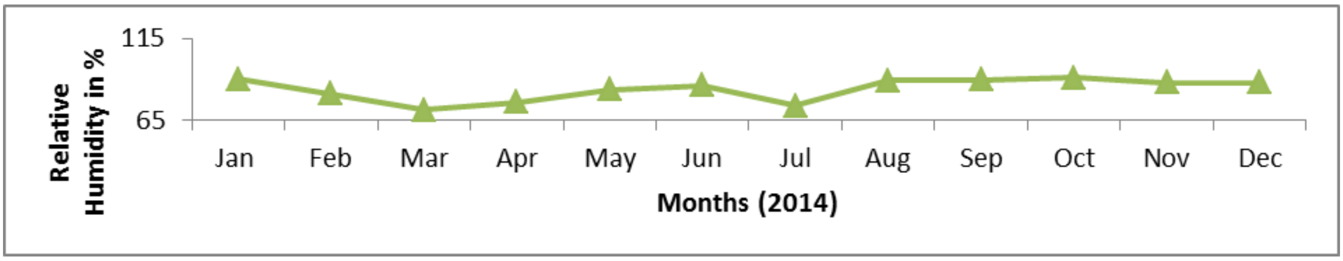


Fig 6: Monthly variation of Relative Humidity

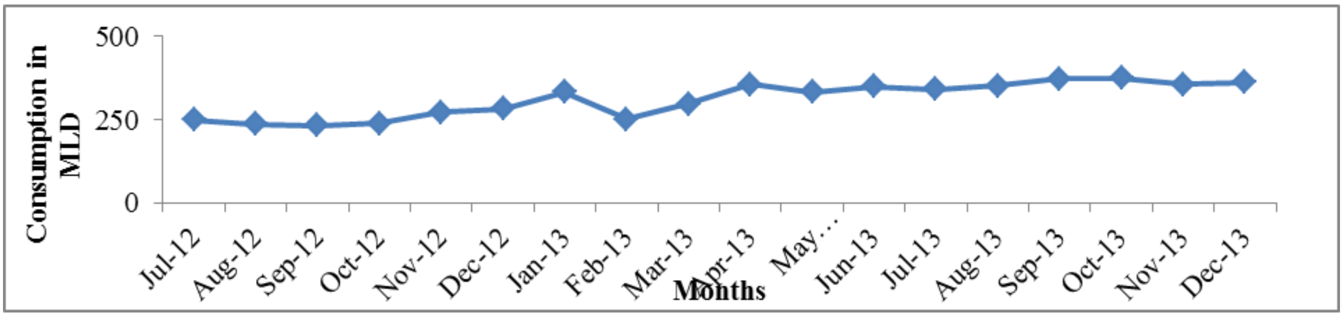


Fig 7: Monthly water consumption variation

Table 2: Model structure

Type	Variable type
Input Variables	Rainfall (RF)
	Maximum Temperature (T max)
	Maximum Temperature (T max)
	Relative Humidity (RH)
Output Variable	Water Demand (WD)

Table 3: Correlation Coefficients of all the variables

CC	Rainfall	Max-Temp	Min-Temp	Relative Humidity	Water Consumption
Rainfall in mm	1.00	0.09	0.05	0.34	0.16
Max-Temp	0.09	1.00	0.42	0.09	0.24
Min-Temp	0.05	0.42	1.00	0.16	0.66
Relative Humidity	0.34	0.09	0.16	1.00	0.04
Water Consumption	0.16	0.24	0.66	0.04	1.00

V. Performance Evaluation

Mean Absolute Error (MAE) and Prediction Error were used to evaluate the model performance. Based on the performance evaluation indices, best model is selected for different input combination and different data sets.

a) Mean Absolute error (MAE): It is defined as the ratio between the differences of observed and predicted values and total number of observation. Smaller the MAE value better will the model result. it is given by the equation :

$$\text{MAE} = \frac{\text{observed values} - \text{predicted values}}{\text{Number of test observation}}$$

b) Prediction Error (P.E): It is defined as the ratio between the differences of observed and predicted values and observed values. If the value is close to zero then the model is treated as best one. prediction Error can be calculated using the equation,

$$\text{P.E} = \frac{\text{predicted values} - \text{observed values}}{\text{observed values}}$$

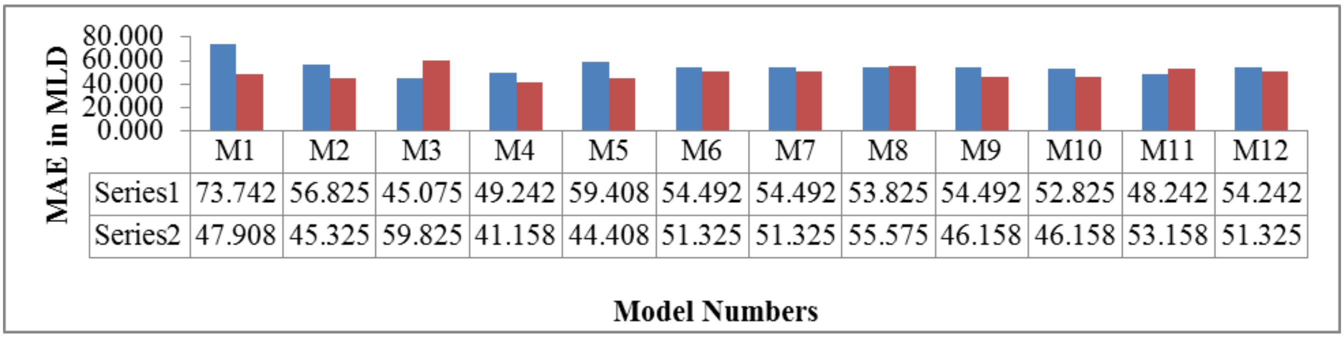
VI. RESULTS AND DISCUSSIONS

The Results of all Fuzzy models for different input combinations and for different data set were represented in the table 4. From the results it is clear that majority of the model performed better (less error) for normalized data compared to original (raw) data. As we can observe Model 1, Model 8 and Model 11 gives less results compared to other models for normalized data since minimum temperature is one of the inputs, which is having less variation. Hence after subjecting to normalization all the Minimum temperature data lies very close. Due to closer values fuzzy rules fails to trigger the relationship for different linguistic terms. Hence performance is low even for normalized data.

Also the model is tested for various inputs combination. Initially Rainfall, Maximum temperature, Minimum Temperature, Relative Humidity is used as separately as input to find out Output. Later number of inputs is increased by trying several combinations as shown in the table 4. For all the combinations of inputs and outputs fuzzy logic is capable of predicting the water consumption using four different rules criteria, four linguistic term and triangular membership function. As we observed since the data is non-linear in nature, all model performance is better for normalized data. From the results table we can reveal that Rainfall, Maximum temperature and Relative Humidity are the most important parameters which influence the water consumption. Error comparison of the entire model is shown in the figure 8. Results show that quality of a data set is very important for model performance. If it is more non-linearity in nature then preprocessing techniques or hybrid technique is necessary to adopt. This research work does not highlight the technique but reveals the importance of quality data for research work.

Table 4: The results of all membership functions and rules criteria

Model No	Inputs	Output	Evaluation Type	Fuzzy Logic Method	
				Raw Data	Normalized Data
M1	RF	WD	PE	0.37	0.24
			MAE	73.74	47.91
M2	Tmax	WD	PE	0.28	0.22
			MAE	56.82	45.32
M3	Tmin	WD	PE	0.22	0.30
			MAE	45.07	59.82
M4	RH	WD	PE	0.24	0.20
			MAE	49.24	41.16
M5	RF, Tmax	WD	PE	0.29	0.22
			MAE	59.41	44.41
M6	RF, Tmax, Tmin	WD	PE	0.27	0.25
			MAE	54.49	51.32
M7	RF, Tmax, Tmin, RH	WD	PE	0.27	0.25
			MAE	54.49	51.32
M8	Tmax, Tmin	WD	PE	0.27	0.28
			MAE	53.82	55.57
M9	RF, RH	WD	PE	0.27	0.23
			MAE	54.49	46.16
M10	Tmax, RH	WD	PE	0.26	0.23
			MAE	52.82	46.16
M11	Tmin, RH	WD	PE	0.24	0.26
			MAE	48.24	53.16
M12	RF, Tmin	WD	PE	0.27	0.25
			MAE	54.24	51.32



	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12
Series1	73.742	56.825	45.075	49.242	59.408	54.492	54.492	53.825	54.492	52.825	48.242	54.242
Series2	47.908	45.325	59.825	41.158	44.408	51.325	51.325	55.575	46.158	46.158	53.158	51.325

Fig 8: Comparative analysis of all the models results.

VII. Conclusions

In this research work an attempt is made to forecast the water demand based on climatic variables using original (raw) and normalized data for various inputs combinations and data set. From the performance indices it is clear that normalized data results were better compared to original data. Also climatic variables such as Rainfall, Maximum Temperature and Relative Humidity influence the water Demand. Hence results reveal that it is necessary to use the quality data for better model performances.

References

[1] Altunkaynak, A., Ozger, M., Cakmakci, M., (2005). Water consumption prediction of Istanbul city by using Fuzzy logic approach. Water resources management 19:641-654.

[2] Aijun,AN., Shan, N., Chan, C., Cercone, N., Ziarko, W., (1996). Discovering rules for water demand prediction: AN enhanced Rough-set approach.PII:SO952-1976:00059-0.

[3] Babel, M., Shinde, R., (2011).Identifying prominent explanatory variables for water demand prediction using artificial neural network: A case of Bangkok. Water resource management 25:1653-1676.

[4] Durga Rao, K.H.V., (2005).Multicriteria spatial decision analysis for forecasting urban water requirement: A case study of Dehradun city, India. Landscape and Urban planning 71:163-174.

[5] Firat, M., Yurdusev, M., Turan, M., (2009). Evaluation of Artificial Neural Network Techniques for Municipal Water Consumption Modeling. Water Resources Manage 23:617-632.

[6] Firat, M., Tarun, M., Yurdusev, M., (2010). Comparative analysis of Neural network technique for predicting water consumption time series. Journal of hydrology 384: 46-51.

[7] Gato, S., Jayasuriya, N., Roberts, peter., (2007). Temperature and rainfall thresholds for base urban water demand modeling. Journal of hydrology 337: 364-376.

[8] Herrera, M., Torgo, L., Izquierdo, J., Garcia, R., (2010). Predictive models for forecasting hourly urban water demand. Journal of hydrology 387: 141-150.

[9] Hongwei, Z ., Xuehua , Z., Bao, Z., (2009). System Dynamic Approach to Urban Water Demand Forecasting. : A Case Study of Tianjin. Tianjin University and Springer-Verlag 15:070-074.

[10] Jain, A., Varshney, K., Joshi, U., (2001). Short Term Water Demand Forecasting Modeling at IIT Kanpur Using Artificial Neural Networks. Water Resources Management 15: 299-321.

[11] Kermani, Z., Teshnehlab, M., (2008). Using adaptive Neuro fuzzy inference system for hydrological time series prediction. Applied soft computing 8:928-936.

[12] Kim, H., Hwang, S., Shin, H., (2001). A Neuro-Genetic approach for daily water demand forecasting. KSCE Journal of civil engineering, volume. No 5, PP 281-288.

[13] Lertpalangsunti, N., Chan, W., Mason, R., Tonhwachwuthikul, P., (1999). A toolset for construction of hybrid intelligent forecasting systems; Application for water demand prediction. Artificial intelligence in Engineering 13: 21-42.

[14] Lee, s., Tong, L.,(2011).Forecasting time series using a methodology based on Autoregressive integrated moving average and Genetic programming. Knowledge Based system 24:66-72.

[15] Mohamed, M., Mualla, A., (2010). Water demand forecasting in Ummal-Quwain using the constant rate model. Desalination 259: 161-168.

[16] Nasseri, M., Moeini, A., Tabesh, M., (2011). Forecasting monthly urban water demand using extended Kalman filter and Genetic programming. Expert system with applications38:7387-7395.

[17] Qi, C., Chang, B., (2011). System dynamic modeling for municipal water demand estimation in an urban region under uncertain economic impact. Journal of environmental management 92:1628-1641.

[18] Sen, Z., Altunkaynak, A.,(2009). Fuzzy system modeling of drinking water consumption prediction. Expert systems with applications 36; 11745-11752.

[19] Wang, C., Chou, W., Chang, T., Qiu, L., (2009). A Comparison of performance of several artificial intelligence methods for forecasting monthly discharge time series. Journal of hydrology 374: 294-306.

[20] Yurdusev, A., Firat, M., (2009). Adaptive Neuro Fuzzy Inference system approach for municipal water consumption modeling. Journal of hydrology 265: 225-234.

[21] Zhou, S.L., McMahon, T.A., Walton, A., Lewis, J., (2000). Forecasting daily urban water demand: A case study of Melbourne. Journal of hydrology 236:153-164.

[22] Zhou, S.L., McMahon, T.A., Walton, A., Lewis, J., (2000). Forecasting operational demand for an urban water supply zone. Journal of hydrology 259:189-202.

Surendra H J , Research scholar , Applied Mechanics and Hydraulics department, NITK Surathkal, Karnataka India. B.E from VTU 2009, M.Tech from NITK Surathkal 2009. Presently Research scholar in the Department of Applied Mechanics and Hydraulics since from 2012, published many papers in International Journal as well as in Conferences.

Dr.Paresh Chandra Deka (Associate professor), Department of Applied Mechanics and Hydraulics, NITK Surathkal, Karnataka India. Published several Books, papers in International and National Journals as well as in Conferences. Guided Three Ph.D students and guiding another six Research Scholars. Area of Interest is soft computing and Hydrologic Modeling