Development of a Fuzzy Logic Model for Predicting Faecal Coliforms in a Freshwater Lake

Shibu. K[1], Dr.S.Ayoob[2]

[1] Associate Professor, Department of Civil Engineering, College of Engineering Trivandrum,
Thiruvananthapuram - 695 016, Kerala, India.
shibukrishnanp@gmail.com

[2]Pro-Vice Chancellor, APJ Abdul Kalam Technological University,
Thiruvananthapuram - 695 016, Kerala, India.
ayoobtkm@gmail.com

Abstract: ***Faecal contamination plays a major role in the transmission of diarrhoeal diseases. Thus it is highly desirable to model the water quality in terms of faecal contamination. In this study, an inductive model using fuzzy logic is developed to predict the faecal coliform concentration by taking Sasthamkotta lake as the case study which a freshwater wetland and a Ramsar site in the Southern part of India. Five input water quality parameters namely pH, temperature, dissolved oxygen, biochemical oxygen demand and turbidity during the period 2013 and 2017 was used to develop the said model. A statistical regression model was also developed using the same input parameters. Thereafter, the performance of the model developed using fuzzy logic and regression model was compared in terms of percentage error. It was found that the percentage error obtained in the fuzzy model was 2.8125 % while the same for the conventional regression model was 12.5 %. The said finding implies that the fuzzy model developed was superior compared to that of the conventional regression model in predicting faecal coliform concentration in Sasthamkotta lake.***

Keywords: Freshwater lake, Faecal Coliform, Modeling, Fuzzy logic, Regression.

1. INTRODUCTION

Biological contamination is one of the leading causes of impairments of surface water quality and the same comes from a variety of point and non-point sources. *Faecal coliform bacteria* are used as a proxy for detecting the presence of pathogenic bacteria in environmental water samples. The presence of faecal coliform bacteria in surface water indicates faecal contamination and hence it is highly desirable to model the water quality by studying the presence of faecal coliform in water samples.

Faecal coliform modeling continues to be a challenge due to the numerous sources of bacteria and the corresponding magnitudes of contribution from each source. The various methods available for faecal coliform determination and the associated environmental effects add to the challenges in accurately modeling faecal coliform concentrations in surface waters. For applications where a detailed analysis of the source of faecal contamination and its associated transport into the receiving water bodies is sought, a detailed process-based deductive model still remains the choice [6]. For the development and calibration of such a model the researcher requires sufficient data so that the same will be effective in predicting the presence of faecal coliform in surface water [2]. Such models can be extremely useful in applications where large data sets are available and quick assessment of water quality is sought. The same can minimize any adverse health impacts associated with faecal contamination as well.

Data driven models based on computational intelligence are becoming more and more popular due to a number of water quality modeling applications [10]. Inductive models also

helps in quick and effective management decisions needed to facilitate reliable and safe operation of such systems. If sufficient data is available, using deductive models are very time consuming and ineffective to make quick decisions. Inductive models range from simple linear regression models to more complex nonlinear models based on Artificial Neural Network (ANN) [3], [12]. Today, inductive models developed using principles of evolution and biology is becoming more popular. [11].

Fuzzy logic is such an evolutionary computation method that has found numerous applications in the development and application of inductive models to real world engineering processes [4]. Fuzzy logic provides a system to model uncertainty, the way in which humans arrive at decision making [8]. Experts in the field of environmental engineering often face an uphill task to interpret vast and complex environmental data that is easily understandable by the decision makers so as to arrive at a sustainable conservation measure. The researcher in such cases often fails to make precise statements about inputs and outcomes. In such cases the fuzzy model can adjust the reasoning process and the same can be applied to develop a series of environmental indices which reduces the complexity of the problem [1]. Nowadays fuzzy rule based systems are very common in solving complex environmental problems [5].

The objective of this study is to develop inductive models based on both fuzzy logic and the conventional statistical regression model to predict the faecal coliform concentration in a fresh water wetland named Sathamkotta lake. Thereafter the fuzzy based model was compared with that of the conventional statistical regression model in terms of percentage error.

2. METHODOLOGY

2.1 Study Area and Data Collection

Sasthamkotta lake lies between latitude $9^0$11'N to $9^0$41'N and longitude $76^0$36'30" E to $76^0$40'E and the same spreads out in three Panchayats namely, Sasthamkotta, Mynagapally and Western Kallada. The said lake is a designated wetland of international importance under the Ramsar Convention since November 2002. Figure 1 represents the map of the study area. This freshwater lake caters to the drinking needs of more than 10 lakh people in Kollam district (in the state of Kerala). Sasthamkotta Lake is subjected to serious environmental degradation including faecal coliform concentration. The water spread area of the lake has shrunk and it was observed that the south western tip of Sasthamkotta lake has turned into lush grassland during the field visit. The pollution problems and the resultant environmental degradation warrants an immediate time bound eco-restoration measures.

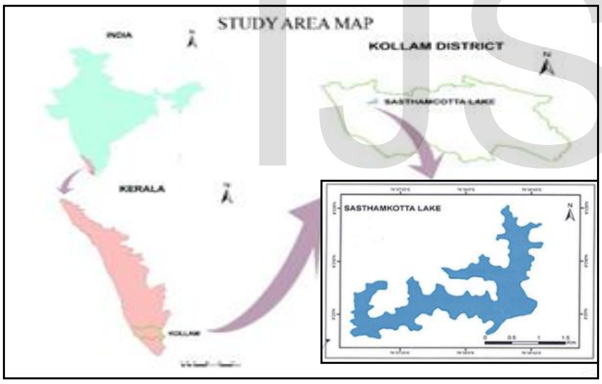

Figure 1: Map of the study area

2.2 Development of Fuzzy Logic Model

In this study, the fuzzy logic formalism was used for developing a model for predicting faecal coliform concentration in Sasthamkotta lake. In fuzzy logic, the mapping is formulated from a given input to the corresponding output and this process is known as fuzzy inference [9], [7]. The process of fuzzy inference was carried out in five steps namely, fuzzification of the input variables, application of the fuzzy operator (AND or OR) in the antecedent, implication from the antecedent to the consequent, aggregation and defuzzification.

The process of fuzzification of the input variables is carried out and the degree of membership to which they belong is determined. In fuzzy logic toolbox software, the input is always a crisp numerical value limited to the universe of discourse of the input variable (in this case the interval is between 0 and 10) and the output is a fuzzy degree of membership in the qualifying linguistic set (always the interval is between 0 and 1).

Once the inputs are fuzzified, the degree to which each component of the antecedent is satisfied for each rule is known. If the antecedent of a given rule has more than one component, the fuzzy operator is applied to obtain one number that represents the result of the antecedent for that rule[8]. This number is then applied to the output function which is the second step in the fuzzy inference process.

Before applying the implication method, the rule's weightage should be determined. Every rule has a *weight* (a value between 0 and 1), which is applied to the number given by the antecedent. From time to time if the researcher wants to weight one rule relative to the others the same can be done by changing its weight value to something other than 1. After assigning proper weightage to each rule, the third step namely the implication method is implemented. The input for the implication process is a single number given by the antecedent, and the output is a fuzzy set represented by the membership function.

The role of implication both in the mathematical and applied aspects of fuzzy set theory is central for developing a successful model. Two built-in methods are supported in the implication process and they are the same functions that are used by the AND method: *min* (minimum), which truncates the output fuzzy set, and *prod* (product), which scales the output fuzzy set.

Since the decisions are based on testing all rules in a Fuzzy Inference System (FIS), then the rules must be combined in some manner in order to make a decision. The fourth step aggregation is the process by which the fuzzy sets that represent the outputs of each rule are combined into a single fuzzy set wherein fuzzy rule base and reasoning mechanism are applied .

The input for the fifth and the final step namely the defuzzification process is the aggregated fuzzy set and the output is a precise quantity. The most commonly used defuzzification method is the centroid calculation, which is adopted in this study.

Statistical regression modeling was carried out for empirical modeling in which the data pertaining to a dependent variable (in this case study it is the faecal coliform concentration) were regressed against those of independent variables namely pH, turbidity, temperature, Biochemical Oxygen Demand (BOD) and Dissolved Oxygen (DO) respectively. The data collected during the period 2013 to 2017 were used to develop the statistical model and the same collected during 2018 was used for the model's validation. The statistical tool namely Statistical Program for Social Science (SPSS) was used for developing the inductive model.

3. RESULTS AND DISCUSSIONS

Five water quality parameters namely pH, turbidity, temperature, BOD and DO were selected as input parameters for the fuzzy inference system to predict faecal coliform concentration. Trapezoidal Membership functions 'low', 'medium' and 'high' were created for each input variable. The membership functions for the input variables namely pH, turbidity, temperature, BOD and DO and the output

ISSN 2229-5518

variable namely faecal coliform concentration are shown in Fig.2, Fig.3, Fig.4, Fig.5, Fig.6 and Fig.7 respectively.

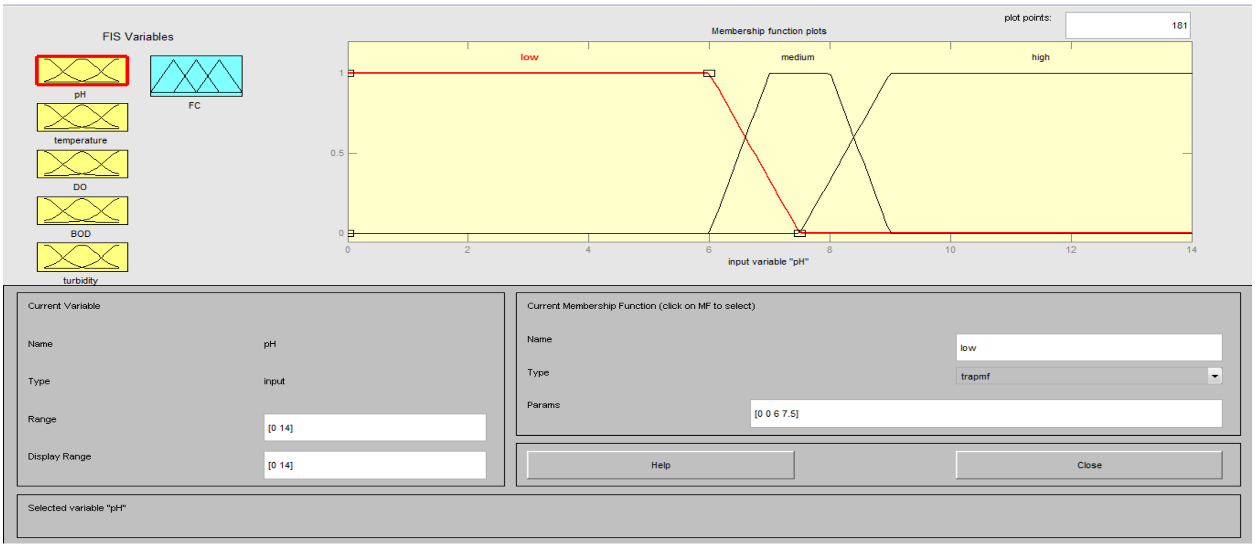

Figure 2: Membership function for the parameter pH

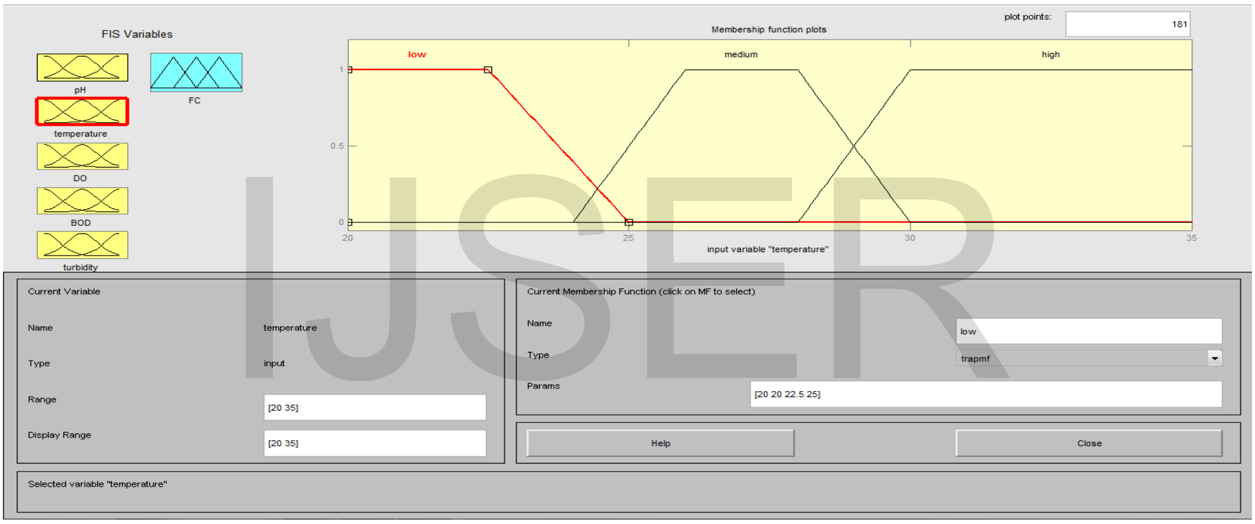

Figure 3: Membership function for the parameter temperature

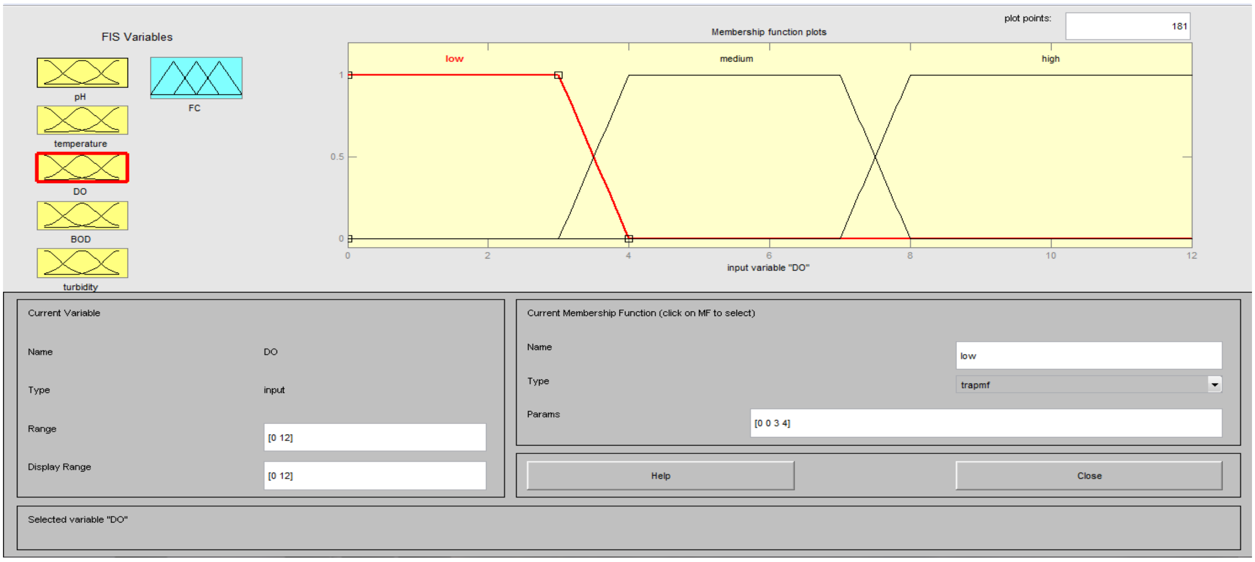

Figure 4: Membership function for the parameter Dissolved Oxygen

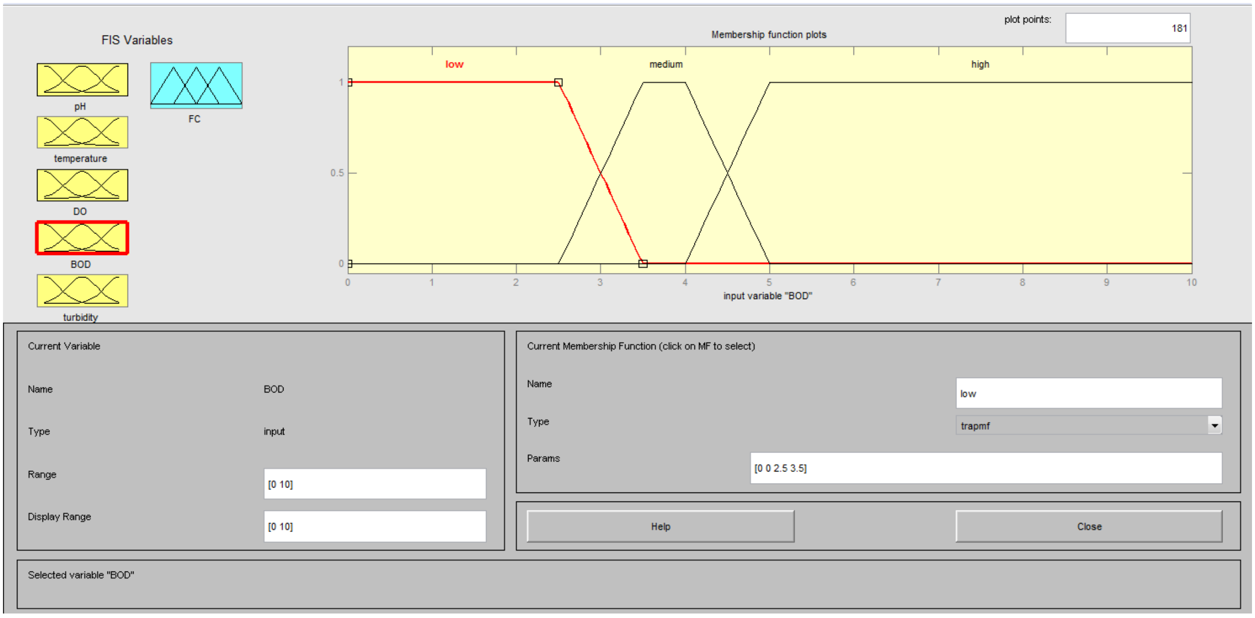

Figure 5: Membership function for the parameter Biochemical Oxygen Demand

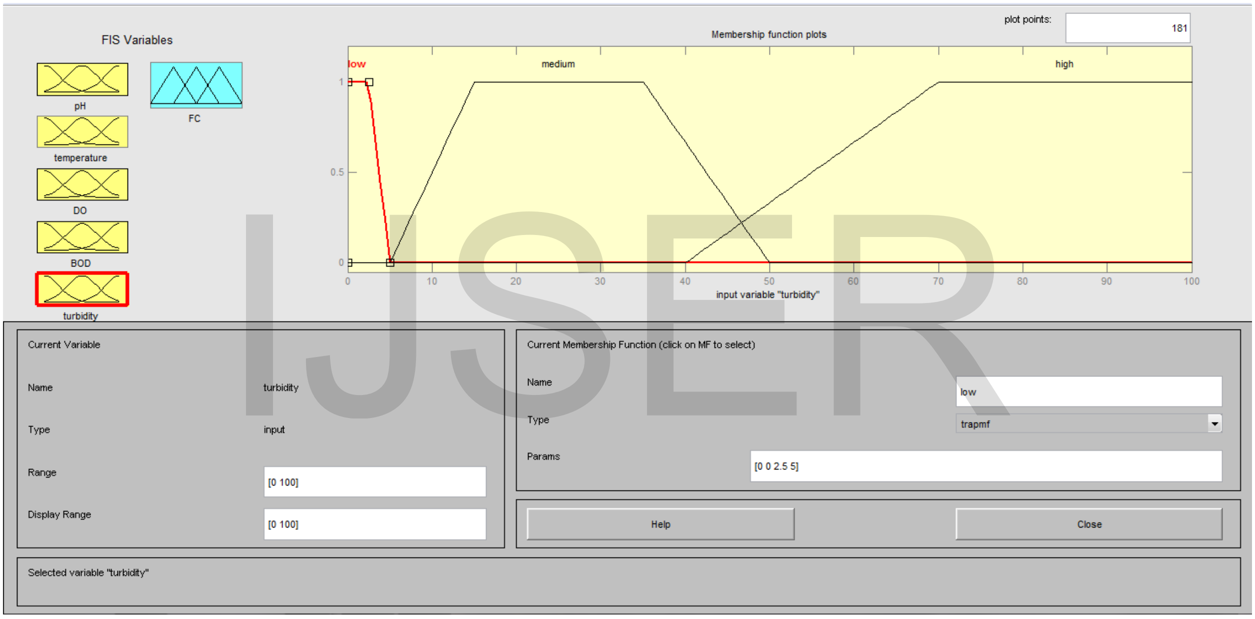

Figure 6: Membership function for the parameter turbidity

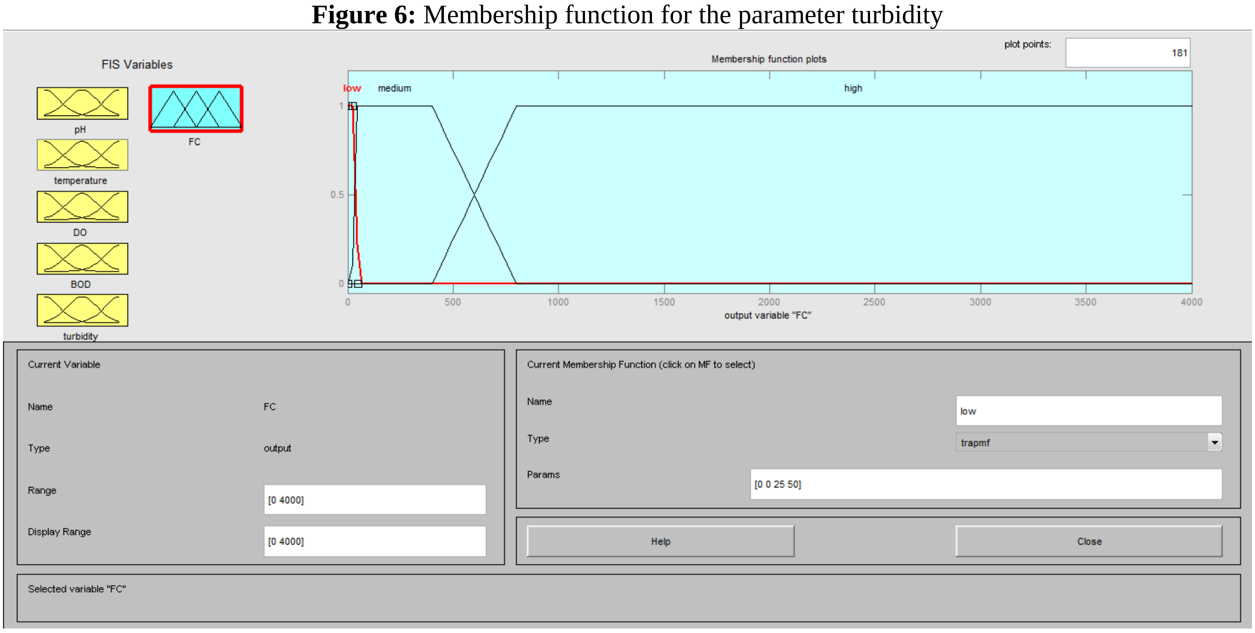

Figure 7: Membership function for the parameter faecal coliforms

ISSN 2229-5518

For formulating the if-then rules using the said parameters, the weight of each parameter was calculated and the same was done to find out the effect of each input parameter in the output while forming the rules. A weight set was established for this purpose and it was developed as A= {a_1, a_2, a_3, a_4, a_5}, which interprets the weights of each of the five indices that sum up the value as 1. In determining the weight of each index, a pair comparison method was adopted regarding the importance of each index and through this comparison, a Judgment Matrix was built up. Specifically, two indices namely u_i, and u_j was taken out each time, and a_{ij} was taken as the comparison of the influence that u_i and u_j each exerts on the overall evaluation. In this way, all the comparison results were got as expressed in A = $(a_{ij})_{5\times5}$ which was called the Judgment Matrix. If the ratio between u_i and u_j was a_{ij}, then the ratio of reversed sequence of the two indexes was found asa_{ji} = $1/a_{ij}$. With the application of Judgment Matrix A, the weight of each index was formulated as equation (1).

$$a_i = \frac{1}{5}\sum_{j=1}^{5}\frac{a_{ij}}{\sum_{k=1}^{5} a_{kj}} \qquad i= 1,2\ldots5 \qquad (1)$$

The weightage of each of the input parameter calculated is given in Table 1.

Table 1: Weightage of each of the input parameters calculated

Input Parameter	pH	Temperature	Turbidity	DO	BOD
Weightage	0.15	0.62	0.03	0.18	0.02

Thereafter the fuzzy rule base was formulated for the model development. Five inputs were chosen and they were connected by if-then rules to give the output. The IF parts were connected using AND operator. For example, IF pH is low and temperature is low and DO is low and BOD is low and turbidity is low, THEN FC is low. A total of 134 fuzzy logic rules were formulated and the same are shown in Fig.8.

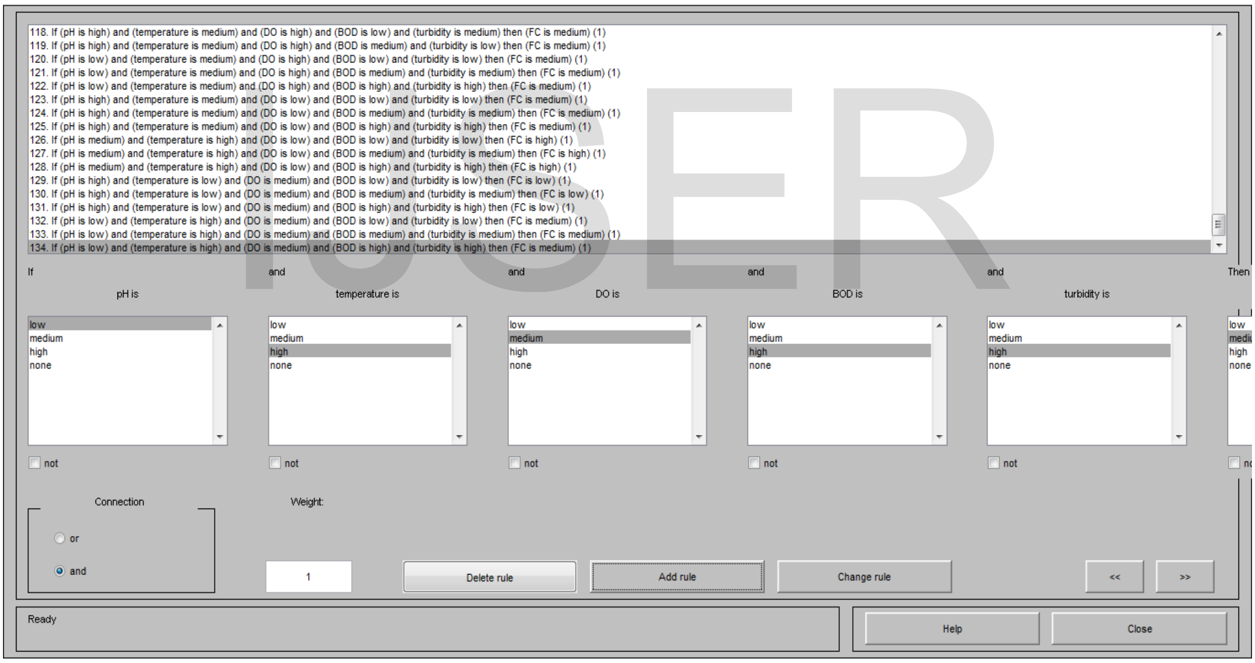

Figure 8: Fuzzy logic rules formulated

In fuzzy logic analysis, the inductive model was developed using the above said five input parameters to predict the faecal coliform concentration in the lake. The model was validated using the actual data for the year 2018. The percentage error between the field value and the predicted value of the faecal coliform concentration was calculated as 2.8125 %. The output from fuzzy logic analysis is shown in Fig.9.

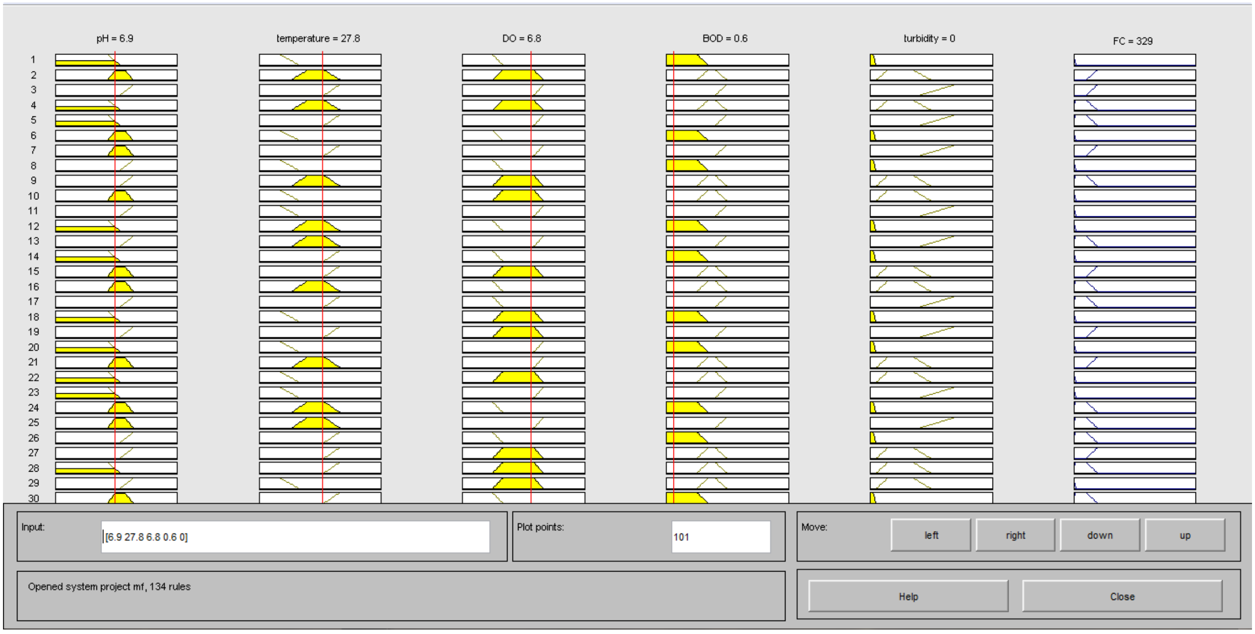

Figure 9: Output from the fuzzy logic analysis

Thereafter the statistical regression model was developed using SPSS. The above said five input parameters were used as independent variables and the parameter faecal coliform concentration was used as the dependant variable and the general regression equation can be written as equation (2).

$$Y = C1 + C2X1 + C3X2 + C4X3 + C5X4 + C6X5 \quad (2)$$

Where Y is the dependant variable, C1, C2, C3, C4, C5 and C6 are constants and X1, X2, X3, X4 and X5 are independent variables (i.e, pH, temperature, turbidity, DO and BOD respectively).

The values of the constants obtained after the regression were: C1= 4380.250, C2 = - 298.270, C3 =- 69. 058, C4= 0, C5 = 0 and C6 = - 186.227 respectively. The correlation coefficient obtained for the model development stage was 1. The faecal coliform concentration was predicted using the developed regression equation. Thereafter the percentage error between the field value and predicted value of the faecal coliform concentration using the said equation was calculated as 12.5 %. The fuzzy based model was developed using five input parameters namely pH, temperature, D.O, BOD and turbidity, which seemed to be the most important factors influencing the presence of faecal coliforms in a surface water body. For generating rule base in fuzzy logic, weights of these parameters were determined using judgement matrix method. It was found that the parameter temperature had the maximum weight and thus proved that temperature was the most influencing factor in faecal coliform prediction. The output generated in fuzzy logic was compared with the actual values and the error was found to be very small. The regression model developed using the same parameters generated an output of higher percentage error. Thus, the superiority of fuzzy model over conventional regression for faecal coliform prediction was proved.

4. CONCLUSION

This study aimed at predicting faecal coliform concentration in surface water using inductive models. The fuzzy rule-based modeling proved to be a reliable and flexible tool to predict faecal coliform concentration. On comparing the observed value and predicted value of faecal coliforms in the lake, it was found that the percentage error obtained in fuzzy model was 2.8125 % while it was 12.5 % in the regression model. Thus, the result obtained from fuzzy model was superior to that of the SPSS model. Hence it can be concluded that fuzzy logic modeling can be used to predict the faecal coliform concentration in surface waters. The fuzzy model developed here would help in quick decision making for pollution abatement programs carried out for Sasthamkotta Lake.

References

[1] Abdullah.M, Sadia.W, Raman.B and Ijaz.M "Development of New Water Quality Model Using Fuzzy Logic System for Malaysia", *Open Environmental Sciences*,2, pp. 101-106, 2008.

[2] Ahmad.S.M, Mohammed.B.T and Soorayya.M, "Intelligent Computational Modeling and Prediction of Coliform Growth in Tropical Lakes based on Hybrid Self Organizing Maps (SOM) and Fuzzy Logic Approaches", *Electronic Journal of Computer Science & Information Technology*, VOL. 1 No. 1, May 2009.

[3] Aslan. M, "Modeling the water quality of lake Eymir using artificial neural networks (ANN) and adaptive neuro fuzzy inference system (ANFIS)", 2008.

[4] Babaei.S, Hassani.A.H, Torabian.A andKarbassi.A.R, "Water quality index development using fuzzy logic: A case study of the Karoon River of Iran", *African Journal of Biotechnology* Vol. 10(50), pp. 10125-10133, 5 September, 2011.

[5] Bouharati.S, Benmahammed.K, Harzallah.D and El-Assaf.Y.M "Application of artificial neuro-fuzzy logic inference system for predicting the microbiological pollution in fresh water", *Journal of Applied Sciences* 8(2), pp. 309-315, 2008.

[6] Brion, G.M. and Lingireddy, S., "Artificial neural network modelling: a summary ofsuccessful applications relative to microbial water quality", *Water Science and Technology*Vol 47 No.3, pp. 235-240, 2003.

[7] Diane.M.L.M and David.P.A., "Comparing artificial neural networks and regression models for predicting faecal coliform concentrations", *Hydrological Sciences Journal*, 52 (4), pp.713-731, 2007.

[8] Mohammed.H, Saad.S.S, M.Y, Haitam.A.A, Aini.H and Ahmed .E, Application of artificial intelligence (AI) techniques in water quality index prediction : A case study in tropical region, Malaysia, *Neural Comput and Applic*, Vol 28 (1), pp.893-905, 2016.

[9] Nasr.S, Rezari.M and Barmaki.D.M,"Analysis of Groundwater Quality using Mamdani Fuzzy Inference System (MFIS) in Yazd Province, Iran", *International Journal of Computer*

ISSN 2229-5518

Applications (0975 – 8887) Volume 59– No.7, December 2012.

[10] Raman.B.V, Reinier.B and Mohan.S, "Fuzzy Logic Water Quality Index and Importance of Water Quality Parameters", *Air, Soil and Water Research* 2009:2, pp.51-59, 2009.

[11] Tufail.M, Ormsbee.L and Teegavarapu.R, "Artificial Intelligence (AI) - based Inductive Models for Prediction and Classification of Faecal Coliform in Surface Waters", *ASCE, Journal of Environmental Engineering*, pp. 789-799, 2006.

[12] Yikui, M.O and Xiangrong, Q.I.A.O, "A Study on the Reliability Evaluation of Urban Transit System", *2nd International Conference on Power Electronics and Intelligent Transportation System*, IEEE, pp. 299-301, 2009.

Authors Profile


Author-1

Shibu K obtained his B.Tech degree in Civil Engineering from University of Kerala, India and thereafter obtained his M.Tech degree in Civil Engineering (Environmental Engineering) from IIT Madras, India. Currently, he is working as Associate Professor and Head of Environmental Engineering Division at the Department of Civil Engineering, College of Engineering Trivandrum, Thiruvananthapuram-695 016, India. His specialisations include Environmental Impact Assessment, Environmental Management, Sustainable Development, Solid Waste Management, Water Treatment, Air Quality Management and Wastewater Treatment.

Author-2


Dr.S.Ayoob received his B.Tech. degree from University of Kerala, India and thereafter his M.Tech. and Ph.D. degree from IIT Kharagpur India. He is currently the Pro-Vice Chancellor of APJ Abdul Kalam Technological University, Thiruvananthapuram - 695 016 Kerala, India. His research interests are in Wastewater Management, Environmental Chemistry, Water supply systems, Solid waste Management, Industrial water pollution control, Environmental Impact Assessment, Sanitary Microbiology, Water Resources Planning and Environmental hydraulics.