Analyzing The Prediction of Water Quality Using WQI Values

Rishabh Bakliwal[1], Priyanka Dubey[2]

[1]M. Tech Scholar in Dr. A.P.J Abdul Kalam University, Indore, Madhya Pradesh, India
[2]Assistant Professor, Department of Civil Engineering in Dr. A.P.J Abdul Kalam University, Indore, Madhya Pradesh, India

Abstract

Artificial intelligence methods can remarkably reduce costs for water supply and sanitation systems and help ensure compliance with the quality of drinking and wastewater treatment. Therefore, modelling and predicting water quality to control water pollution has been widely researched. The novelty of the proposed system is presented to develop an efficient operation of monitoring drinking water to ensure a sustainable and friendly green environment. In this work, the adaptive neuro-fuzzy inference system (ANFIS) algorithm was developed to predict the water quality index (WQI). The dataset has eight significant parameters, but seven parameters were considered to show significant values. The proposed methodology was developed based on these statistical parameters. Prediction results demonstrated that the ANFIS model was superior for the prediction of WQI values.

Keywords: *Prediction, Water, Quality, Modelling, Sanitation*.

1. INTRODUCTION

Water is one of the prime elements responsible for life on the earth. The six billion people on earth use nearly 30 percent of the world's total accessible renewal water supply. Yet billions of people are deprived of basic water availability. Among other countries in the world, India is one of the few selected countries endowed with reasonably good land as well as water resources. India is a country with vast geographic, biological and climatic diversity. Average annual precipitation including snowfall is approx. 4000 billion cubic meters (BCM) over the country. The average annual water resources in various river basins are estimated to be 1869 BCM, of which 1086 BCM is utilizable including 690 BCM of surface water and 396 BCM of ground water. The rest of the water is lost by evaporation or flows into the sea and goes unutilised.

Water is one of the key elements of the environment that determines the survival of life and restricts the socio-economic growth of the people. Overseas and inland surface and sub-surface water systems play an incredible role in everyday life activities mainly for drinking, agricultural, industrial, recreational, and other public uses. Our everyday lives depend on the availability and quality of water. Accessibility of suitable water quality for different purposes is becoming difficult due to rapid population growth and expansion of agro-industries. Some industrial, agricultural, and human activities have a serious effect on ecological diversity. In addition, surface water quality depends on natural phenomena; the quality of water in lakes and dams is suffering from incessant degradation due to natural processes resulting from eutrophication and anthropogenic causes.

Water quality modelling is a useful tool for gaining a better understanding of the river system and for simulating conditions that may not be obtained by field monitoring. Environmental models can be highly unreliable due to our limited knowledge of environmental systems, the difficulty of mathematically and physically representing these systems, and limitations to the data used to develop, calibrate and run these models. The extensive range of physical, biochemical and ecological processes within river systems is represented by a wide variety of models: from simpler one-dimensional advection dispersion equation (1D ADE) models to complex eutrophication models. Gaining an understanding of uncertainties within catchment water quality models across different spatial and temporal scales for the evaluation and regulation of water compliance is still required. Water quality modelling assists in the assessment and improvement of water systems by simulating and predicting water quantity and quality conditions. The complexity of water quality models spans from zeroth dimensional models, representing water volumes and concentrations without dispersion to biochemical models which include reaction terms to describe the biochemical processes in river systems.

2. LITERATURE REVIEW

Ping Liu, Jin Wang, Arun Kumar Sangaiah, Yang Xie and Xinchun Yin (2019) This research paper focuses on a water quality prediction model which requires high-quality data. In the process of construction and operation of smart water quality monitoring systems based on Internet of Things (IoT), more and more big data are produced at a high speed, which has made water quality data complicated. Taking advantage of the good performance of long short-term memory (LSTM) deep neural networks in time-series prediction, a drinking-water quality model was designed and established to predict water quality big data with the help of the advanced deep learning (DL) theory in this paper. The drinking-water quality data measured by the automatic water quality monitoring station of Guazhou Water Source of the Yangtze River in Yangzhou were utilized to analyze the water quality parameters in detail, and the prediction model was trained and tested with monitoring data from January 2016 to June 2018. The results of the study indicate that the predicted values of the model and the actual values were in good agreement and accurately revealed the future developing trend of water quality, showing the feasibility and effectiveness of using LSTM deep neural networks to predict the quality of drinking water.

Mosleh Hmoud Al-Adhaileh and Fawaz Waselallah Alsaade (2021) Artificial intelligence methods can remarkably reduce costs for water supply and sanitation systems and help ensure compliance with the quality of drinking and wastewater treatment. Therefore, modelling and predicting water quality to control water pollution has been widely researched. The novelty of the proposed system is presented to develop an efficient operation of monitoring drinking water to ensure a sustainable and friendly green environment. In this work, the adaptive neuro-fuzzy inference system (ANFIS) algorithm was developed to predict the water quality index (WQI). Feed-forward neural network (FFNN) and K-nearest neighbors were applied to classify water quality. The dataset has eight significant parameters, but seven parameters were considered to show significant values. The proposed methodology was developed based on these statistical parameters. Prediction results demonstrated that the ANFIS model was superior for the prediction of WQI values. Nevertheless, the FFNN algorithm achieved the highest accuracy (100%) for water quality classification (WQC). Furthermore, the ANFIS model accurately predicted WQI, and the FFNN model showed superior robustness in classifying the WQC. In addition, the ANFIS model showed accuracy during the testing phase, with a regression coefficient of 96.17% for predicting WQI, and the FFNN model achieved the highest accuracy (100%) for WQC. This proposed method, using advanced artificial intelligence, can aid in water treatment and management.

Sabrina Sorlini, Daniela Palazzini, Joseph M. Sieliechi, Martin B. Ngassoum (2013) Unsafe drinking water is one of the main concerns in developing countries. In order to deal with this problem, a cooperation project was set up by the ACRA Foundation in the Logone valley (Chad-Cameroon). Water supplies were sampled throughout the villages of this area mostly from boreholes, open wells, rivers and lakes as well as some piped waters. The samples were analysed for their physical-chemical and microbiological quality in order to identify the contamination problems and suggest appropriate solutions. Results of the assessment confirmed that in the studied area there are several parameters of health and aesthetic concern. Elevated lead levels were detected both in aquifers and in surface waters, confirming that further investigations of the occurrence of lead contamination in the Logone valley are warranted.

Camacho Suarez, Vivian V. (2020) Maintaining healthy river ecosystems is crucial for sustaining human needs and biodiversity. Therefore, accurately assessing the ecological status of river systems and their response to short and long-term pollution events is paramount. Water quality modelling is a useful tool for gaining a better understanding of the river system and for simulating conditions that may not be obtained by field monitoring. Environmental models can be highly unreliable due to our limited knowledge of environmental systems, the difficulty of mathematically and physically representing these systems, and limitations to the data used to develop, calibrate and run these models. The extensive range of physical, biochemical and ecological processes within river systems is represented by a wide variety of models: from simpler one-dimensional advection dispersion equation (1D ADE) models to complex eutrophication models. Gaining an understanding of uncertainties within catchment water quality models across different spatial and temporal scales for the evaluation and regulation of water compliance is still required. Furthermore, the use of incorrect model structures for water quality evaluation and regulation leads to considerable sources of uncertainty when applying duration over threshold regulation within the first 100s of meters and sub hourly time scale.

Erfan Babaee Tirkolaee, Ali Asghar Rahmani Hosseinabadi, Mehdi Soltani, Arun Kumar Sangaiah, ORCID andJin Wang (2018) Greenhouse gases (GHG) are the main reason for the global warming during the past decades. On the other hand, establishing a well-structured transportation system will yield to create least cost-pollution. This paper addresses a novel model for the multi-trip Green Capacitated Arc Routing Problem (G-CARP) with the aim of minimizing total cost including the cost of generation and emission of greenhouse gases, the cost of vehicle usage and routing cost. The cost of generation and emission of greenhouse gases is based on the calculation of the amount of carbon dioxide emitted from vehicles, which depends on such factors as the vehicle speed, weather conditions, load on the vehicle and traveled distance.

3. WATER ENVIRONMENTAL SYSTEM

Water Environmental System In order to assess the suitability of river water for various intended beneficial uses, it is most desired that an accurate and rational assessment of river quality should be made afl:er the confluence point of wastewater effluent outfalls into the river. Therefore, a suitable model was selected for DO and BOD prediction along the natural river system and it was further extended by taking the sedimentation of bio-flocculated particulate matters into account for the prediction of BOD. In this present research program, an attempt has been made to develop an algorithm and computer simulation for the prediction of water quality in the multi-wastewater outfall river system by taking mass balance, heat balance, gas transfer, sedimentation, solar radiation, velocity variation, variation in river dimensions, mixing distance, intake quantity of water for industrial use, effluent discharge and pH effects.

4. SYMMETRIC CROSS-SECTIONAL RIVER SYSTEM

Let us consider a symmetric cross sectional river system whose maximum surface velocity is 0.3m/s and its width and maximum depth are 50m, 3m respectively. Since the cross section of the river is symmetric, the depth is uniformly distributed against the width of the river system. Therefore, the domain of the width axis may be considered in the interval (-25, 25) and the maximum depth is at the origin. Now we have = -25, by = 25, D = 3 and V = 0.3, The velocity distribution at different layer over the cross section of the river has been shown in **Figure 1.**

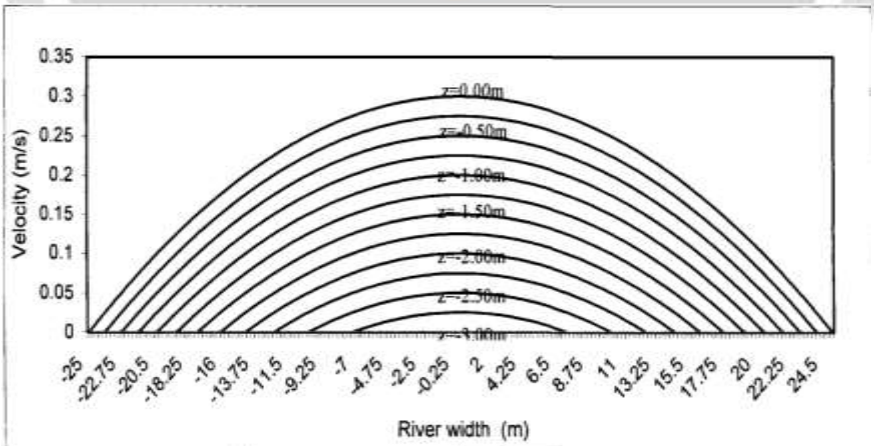

Figure 1: Plot of stream velocity at different points over the symmetric cross section of the river system

5. DEVELOPMENT OF RQMINFOSYS

Appropriate models for river flow estimation and river quality prediction were selected for developing River Quality Management Information System (RQM Infosys). The river system was divided into finite number of stretches and various river dimensions such as average depth, width and critical mixing distance, maximum stream velocity were monitored at each outfall point.

Spearman's Rank Correlation coefficient is a non-parametric measure of statisticaldependence between two variables. It assesses how well the relationship between two variables can be described using monotonic function. For the second test for trend, the entire data series for each parameter, paired with cumulative season for temporal variation, were analyzed using non-parametric Spearman's criterion to detect the existence of trends in water quality parameters. The existence of a trend in temporal variation of data was checked for significance at a level of 5% from statistical tables of Student's t-distribution (Antonopoulouset al., 1998).

The Spearman's Rank Correlation coefficient (R_{sp}) can be described by as :

$$R_{sp} = 1 - \frac{6\sum_{i=1}^{n}(D_i D_i)}{n(n^2-1)}, \quad (1)$$

where n = number of values in each set of water quality data, D= the difference, and I = chronological order number. The difference between rankings can be computed as:

$$D_i = K_{xi} - K_{yi}, \quad (2)$$

where K_{xi}= Rank of measured variable in chronological order, K_{yi}= series of measurements transformed to its rank equivalents by assigning the chronological order number of the measured variable in the original series; x to the corresponding order number in the ranked series, y.

Table 1: Permissible Limits for Drinking Water Quality (IS 10500-1991, CPCB)

Water Quality Parameter	Permissible Ranges
pH	7.0-8.5
DO	4.0-6.0
BOD	2.0-3.0
Conductivity	0-1000
Nitrate-N	45-100
TC	50-5000
FC	50-5000
COD	18-30
Ammonia-N	0-5.0
TA as CaCO3	200-600
TH as CaCO3	200-600

where v_i= value of the ith water quality parameter at a given sampling station and s_i= standard permissible value of ith water quality parameter. This equation ensures that q_i= 0 when a pollutant (the ith water quality parameter) is absent in the water while q_i= 100 if the value of this parameter is just equal to its permissible value for drinking water. Thus, the larger the value of *qi*, the more polluted is the river water with the *i*th pollutant. However, water quality ratings for pH and DO require special handling and care. The permissible range of pH for the drinking water is 7.0 to 8.5.

6. THE VISUAL APPROACH FOR DEFINING CORRELATION

A visual aspect of correlation can be obtained by representing each one of a pair of variables as an axis in Cartesian coordinate system as shown in Figure 2. The visual aspects of understanding may supply crucial insight about structures in the data being analysed and prevented the potential biases that may arise by directly interpreting numerical results yieldedby running computational procedures.

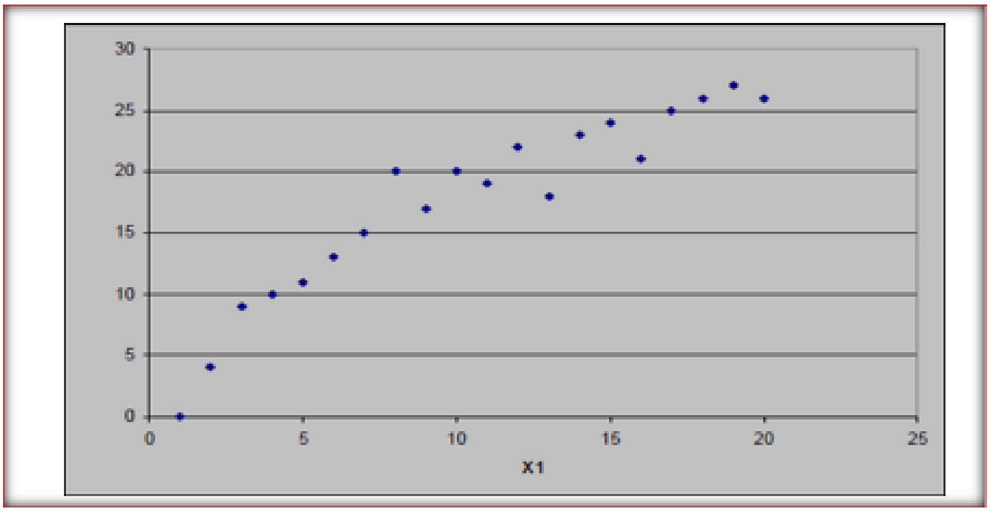

Figure 2: Scatter plot of variables

7. MULTIVARIATE ANALYSIS OF VARIANCE (MANOVA) WITH DISCRIMINANTANALYSIS

The summary of descriptive statistics of the results of the analysis is presented in Table 2, indicating the maximum, minimum, mean values of the parameters and the standard deviation. In summer season, TH as $CaCO_3$ has the high value of 132 mg/L, where as BOD has the highest value of 8.90 mg/L. Likewise in monsoon, TH as $CaCO_3$ has the high value of111 mg/L and Ammonia-N has the maximum value of 4.48 mg/L. Similarly, in winter season again TH as $CaCO_3$ has the high value of 120 mg/L and BOD has a maximum value of 7 mg/L. The standard deviation around the means is substantially high and random. This may be the results of temporal as well as spatial changes and also the different anthropogenic activities surrounding the study area.

In order to explore the spatial variation among different gauging stations and seasonal changes, MANOVA is used to group these on the basis of spatial and temporal similarities as shown in Table 2. Analysis between and within the water quality parameters shows the small significant differences at a significant level of α=0.05. From these results it can be concluded that the spatial sampling interval of 400 km is too a long distance to focus properly in Brahmani River Basin for the gauging stations situated.

The river is impacted by the same or similar influences over the spatial sampling interval which explain the lack of large significant variation in the properties in the study area in three consecutive seasons. This will be useful in monitoring protocol, but there should not be muchchange in the properties will be not be expected in the river. Any observed significant change could suggest a point source polluting it.

Table 2: Model for Multivariate tests for all sample seasons on River Brahmani

Model	Value (S)	F (S)	p-value (Sig.) (S)	Value(M)	F(M)	p-value (Sig.) (M)	Value (W)	F (W)	p-value (Sig.) (W)
Pillai's Trace	1.33	8.49	0	1.27	5.85	0	1.11	8.35	0
Wilks' Lambda	0.15	10.25	0	0.13	8.34	0	0.22	10.11	0
Hotelling's Trace	2.96	12.36	0	4.07	12.34	0	2.32	12.32	0
Roy's Largest Root	2.06	35.27	0	3.44	43.15	0	1.74	37.59	0

NOTE: S: Summer, M: Monsoon, W: Winter

8. CONCLUSION

Water pollution is one of the worldwide challenges facing both developed and developing countries. The cities mainly have the highest pollution rates because of inadequate waste management systems and urban runoff pollution. Water pollution problems are usually due to economic growth and have an impact on both the environment and human health. The causes of water contamination include soil erosion, deforestation, habitat destruction, improper waste management, overgrazing, lack of awareness in management, shortage of decision support tools, and an inadequate organized database system. In the current situation, the ecosystems require a sustainable management solution for improved socioeconomic development. For a better understanding of how the environment and water systems are polluted and to make fruitful decisions and directions, concrete knowledge of water quality modeling is needed. This review described an overview of water quality modeling emphasizing modeling application, commonly used water quality models, and model selection, application, and limitations for different waterbodies. Water quality modeling is a significant tool that helps to water managers and policymakers applying for unified water and environmental management. Scholars, policymakers, and designers through rules and regulations significantly require the practice and application of water quality modeling. Models have been applied to simulate various water quality characteristics and evaluate water quality changes as a result of wastewater discharge to the ecosystem. Various water quality models have been developed and applied in some countries to study the quality of water in various waterbodies with 1D, 2D, and 3D simulations. Different types of water quality models including some extensions of model software have been developed for predicting in different topography, waterbodies, and pollutants at different space and time scales. Water quality model studies are very important for providing solutions and directions towards sustainable planning and management of waterbodies. According to the performance evaluation and error analysis of the models, there is least error in case of ANFIS when compared with that of ANN and MCS. Therefore, it can be said that ANFIS predicted WQI with reasonable accuracy. From the results of ANFIS based analysis, it can be concluded that if the present conditions prevail the future years also WQI values will have the same trend as those from 2003 to 2012.

9. REFERENCES

1. Ping Liu, Jin Wang, Arun Kumar Sangaiah, Yang Xie and Xinchun Yin, “Analysis and Prediction of Water Quality Using LSTM Deep Neural Networks in IoT Environment**”,** Sustainability 2019, 11, 2058
2. Hmoud Al-Adhaileh, M.; Waselallah Alsaade, F. Modelling and Prediction of Water Quality by Using Artificial Intelligence. Sustainability 2021, 13, 4259. https://doi.org/ 10.3390/su13084259

3. Sabrina Sorlini, Daniela Palazzini, Joseph M. Sieliechi, Martin B. Ngassoum, “Assessment of Physical-Chemical Drinking Water Quality in the Logone Valley (Chad-Cameroon)”, Sustainability 2013, 5, 3060-3076; doi:10.3390/su5073060
4. Fawaz Al-Badaii, Mohammad Shuhaimi-Othman, and Muhd Barzani Gasim, “Water Quality Assessment of the Semenyih River, Selangor, Malaysia”, Hindawi Publishing Corporation Journal of Chemistry Volume 2013, Article ID 871056, 10 pages
5. Camacho Suarez, Vivian V., “Evaluating River Water Quality Modelling Uncertainties at Multiple Time and Space Scales”, Department of Civil and Structural Engineering, March 2020
6. Mishra, D.R.; D’Sa, E.J.; Mishra, S. Preface: Remote sensing of water resources. Remote Sens. 2018, 10, 115.
7. Gleason, C.J.; Wada, Y.; Wang, J. A hybrid of optical remote sensing and hydrological modelling improves water balance estimation. J. Adv. Model. Earth Syst. 2017, 10, 2–17.
8. Ren, Y.; Liu, Y.; Ji, S.; Sangaiah, A.K.; Wang, J. Incentive mechanism of data storage based on blockchain for wireless sensor networks. Mob. Inf. Syst. 2018, 2018, 6874158.
9. Gao, F.; Feng, M.Q.; Teng, S.F. On the way for forecasting the water quality by BP neural network based on the PSO. J. Saf. Environ. 2015, 15, 338–341.
10. You, G.; Xu, B.; Su, H.; Zhang, S.; Pan, J.; Hou, X.; Li, J.; Ding, R. Evaluation of Aquaculture Water Quality Based on Improved Fuzzy Comprehensive Evaluation Method. Water 2021, 13, 1019. https:// doi.org/10.3390/w13081019
11. Liu, S.; Tai, H.; Ding, Q.; Li, D.; Xu, L.; Wei, Y. A hybrid approach of support vector regression with genetic algorithm optimization for aquaculture water quality prediction. Math. Comput. Model. 2013, 58, 458–465.
12. Gazzaz, N.M.; Yusoff, M.K.; Aris, A.Z.; Juahir, H.; Ramli, M.F. Artificial neural network modeling of the water quality index for Kinta River (Malaysia) using water quality variables as predictors. Mar. Pollut. Bull. 2012, 64, 2409–2420.
13. Liu, D.J.; Zou, Z.H. Application of weighted combination model on forecasting water quality. Acta Scientiae Circumstantiae 2012, 32, 3128–3132.
14. Clark, R.; Hakim, S.; Ostfeld, A. Handbook of Water and Wastewater Systems Protection (Protecting Critical Infrastructure); Springer: New, York, NY, USA, 2011.
15. Maier, H.R.; Jain, A.; Dandy, G.C.; Sudheer, K.P. Methods used for the development of neural networks for the prediction of water resource variables in river systems: Current status and future directions. Environ. Model. Softw. 2010, 25, 891–909.