DEVELOPMENT OF WATER QUALITY INDEX FOR MANAGEMENT OF RIVER

Ph.D THESIS

by

MANSI TRIPATHI

DEPARTMENT OF HYDRO AND RENEWABLE ENERGY
INDIAN INSTITUTE OF TECHNOLOGY ROORKEE
ROORKEE – 247667 (INDIA)

AUGUST, 2019

DEVELOPMENT OF WATER QUALITY INDEX FOR MANAGEMENT OF RIVER

A THESIS

Submitted in partial fulfilment of the requirements for the award of the degree

of

DOCTOR OF PHILOSOPHY

in

HYDRO AND RENEWABLE ENERGY

by

MANSI TRIPATHI

DEPARTMENT OF HYDRO AND RENEWABLE ENERGY
INDIAN INSTITUTE OF TECHNOLOGY ROORKEE
ROORKEE – 247667 (INDIA)

AUGUST, 2019

ABSTRACT

Due to its limited availability, the quality of consumable surface water becomes very important. Water quality includes anything and everything that it might have picked up during its journey; in colloidal, dissolved or even in suspended state. The voluminous data involved in the water quality assessment and management can often be quite overwhelming for a lay-man. In an attempt to transfer the information in more precise and comprehendible way, a new approach has been used to integrate the data pool in the form of simple numbers termed as water quality indexes (WQIs or indices). This concept was originally developed in Germany by Horton (1965) and since this pioneer attempt, several other initiatives have been made to develop a suitable index for water quality in different regions and also for global usage.

There are some limitations associated with WQIs in terms of providing the complete picture of sample in all forms of pollutants including micro pollutants and persistent pollutants but it benefits by far exceed its limitations. The benefits of WQIs are – (i) It provides a generic overall picture of water quality (ii) Gives its spatio- temporal variation to judge the effectiveness of a water quality program, (iii) It can be used for easy representation of the water for policy makers and stakeholders, (iv) It can be used as a decision making tool by government authorities (v) WQI helps in drawing comparison between water quality at different locations and thus reporting in an easily comprehendible manner.

Building on the base of earlier literature and its comprehensive review, it has been found that most of the previous researches have been directed towards specific regions, hence specific assessment, monitoring and planning is required. Most of the commonly used WQIs are found not to be in sync with water quality standards and do not provide flexibility in terms of number of parameters. It can also be seen that the earlier WQIs are either theoretical or statistical in nature and a delicate balance of both approaches is the need of the hour. Most of the commonly used indices are high on 'eclipsing' i.e. a degraded parameter's effect is often compensated by the positive effect of another parameter. Besides this, currently used indices classify a historical water quality status, where prediction of future parameters and issuing warning for future has not been observed.

A WQI forms basis of a systematic study and helps to identify the effects of various Government interventions. Keeping the above requirements in view, this study makes an attempt for development of a WQI for river Ganga flowing through Uttar Pradesh state in India with the followings objectives:

1. To procure historical data from monitoring agencies and carry out water sampling for trend analysis in the selected study area.
2. To select/reduce parameters using principal component analysis (PCA) and to provide weightage to these parameters using factor analysis (FA).
3. To generate sub- index rating curves, hybrid aggregation and sensitivity analysis of the developed Ganga WQI.
4. To conduct comparative analysis of developed GWQI along with cost-benefit analysis.
5. To develop regression and ANN model from secondary data and to validate it using primary data in order to find best predictive modelling approach.

In order to achieve the above objectives, the stretch of river Ganga from Garhmukteshwar to Ghazipur (967 km.) having 15 monitoring stations has been selected as the study area. The methodology adopted for this study is as follows-

1. Selection of study area and sampling locations
2. Procurement of water quality data from Central Pollution Control Board (CPCB) and Central Water Commission (CWC).
3. Collection of a set of post and pre monsoon samples from the sampling locations and their analysis
4. Selection of WQI parameters by using PCA and estimating weights of the shortlisted parameters by applying FA to the shortlisted parameters
5. Development of rating curves using national and international water quality standards
6. Aggregation of weighted sub-index values using a combination of harmonic and arithmetic means to obtain the final GWQI
7. Sensitivity Analysis of the developed GWQI
8. Analysis of historical trends and cost- benefit analysis of the GWQI
9. Predictive modelling using regression analysis and ANN.

For parameter reduction, PCA reiterates the data variability dispersed in several dimensions into a decreased number of uncorrelated dimensions called principal components (PCs) which can be expressed in the form of linear equations. Kaiser–Meyer–Olkin (KMO) and the Bartlett's tests of Sphericity have been performed on the available dataset to check its suitability before carrying out PCA. Both of these tests have been carried out by using SPSS software and nine parameters were selected for further analysis.

For providing weights to the parameters after PCA, the parameters contributing maximum (> 0.35; positive or negative) to the first five principal components were shortlisted. These PCs account for 90.36 % of the total variance, have individual eigen values >1 and their individual contribution to the overall variance is > 10%. These PCs were subjected to FA. The next step is rotation (commonly Varimax rotation) and post rotation, a matrix of square of factor loadings is prepared. Squaring gives the weights of individual factors while weights of individual parameters are estimated once the loadings are scaled to unity, where weights are equal to the maximum contribution of a parameter to an individual factor.

For sub-index development, each parameter has been assigned a value ranging from 1 to 100. The rating curves were developed on the basis of BIS (2012), WHO (2017) and DBU (standards developed by CPCB) and plotted using MATLAB software. The final aggregation has been done using a hybrid aggregation of weighted arithmetic and weighted harmonic mean.

After developing the final Ganga Water Quality Index (GWQI), its sensitivity analysis was carried out by removing individual parameters and all the reduced and increased GWQI values were found to be significantly correlated with the developed GWQI values ($R^2 > 0.95$). It was found that the developed GWQI is not biased towards a single parameter and the procedure can be used as the results are not affected by a single parameter.

For obtaining primary data for the study area, river water sampling was done at the 15 monitoring sites. The mean of the values obtained from the analysis of post-monsoon and pre- monsoon samples (2016) has been plotted along with the mean values obtained from the secondary data. The plots show the water quality trend in terms of individual parameters as well as GWQI values. The variability in the GWQI values is represented as Box and Whisker plots. Based on the analysis, the trend is found to be more or less similar over the years. In order to make a comparative assessment, CCMEWQI and NSFWQI were also estimated using primary data and it is found that GWQI and CCMEWQI are more sensitive than NSFWQI as their trend shows greater variation at different sites.

Further, the information and analysis in the form of WQIs is translated into socio-economic benefits and subsequently monetized considering the annual average income per house hold along the considered stretch of the river. It has been further analysed how the monetary benefit varies under different aggregation conditions. BOD appeared to be the most critical parameter as a 10 point reduction in BOD sub index value leads to lowest monetary benefits.

A Correlation analysis established the relation between various water quality parameters and parameters found to be dependent on it. Out of these relations, the three strongest relations were found i.e. DO – BOD (R = -0.8738), DO- GWQI (R= 0.9063) and BOD - GWQI (R= -0.8255). Considering these relations, a linear regression model and an ANN model were developed for prediction of GWQI from DO and BOD individually. For developing ANN model, the MATLAB software is used for training, validating and testing by using the data set as the secondary data collected for years 1996-2015 by giving one input and one output. It gives the best training function as TRAINR and adaption learning function is LEARNGD. The transfer function used is TRANSIG and number of hidden neurons are 10. The data samples were divided into training sets (60% of the total -1996 to 2015), validation sets (20% of the total- 1996 to 2015) and testing sets (20% of the total - 1996 to 2015). The results show that the proposed ANN model has better and greater potential to simulate and predict GWQI from DO and BOD individually with acceptable accuracy.

From this study, it is concluded that the procedure of development of GWQI can be customised as per the specific needs and historical data availability in a region in order to develop a regional WQI to address regional problems. It can also be used to compare and express water quality universally depending on availability of historical data as the number and nature of parameters can be customised for every case which is its major advantage over the most widely used and accepted WQIs worldwide. The conclusions drawn from this study are as follows-

1. In the first step for development of GWQI, 9 parameters - DO, BOD, TDS, pH, sulphate, conductivity, TC, magnesium and chlorides were identified using PCA. These parameters are the ones contributing > 0.35 to the variance of the retained 5PCs.
2. Weights have also been allocated to these 9 parameters individual factor loadings obtained from retained 5 PCs contributing more than 90% of the total variance.
3. Rating curves have been developed after considering national and international standards.
4. Final equation for estimating GWQI has been developed using a combination of arithmetic and harmonic means and the sensitivity analysis of this equation reveals that it is not biased towards a single parameter. The final equation obtained for estimation of GWQI is as follows-

$$\text{GWQI} = \left[\frac{0.28}{\frac{0.28}{I_{DO}}+\frac{0.36}{I_{BOD}}+\frac{0.24}{I_{TDS}}}\right] + \left[\frac{0.16}{\frac{0.25}{I_{pH}}+\frac{0.63}{I_{Sulphate}}}\right] + \left[\frac{0.14}{\frac{0.47}{I_{Cond}}+\frac{0.29}{I_{TC}}}\right] + \left[\frac{0.10}{\frac{0.6}{I_{Mg}}}\right] + \left[\frac{0.12}{\frac{0.83}{I_{Cl}}}\right]$$

5. The results of analysis of the monitoring stations' water quality data for 20 years by applying developed GWQI showed that most of the samples fall under the "average" category indicating not a very good health of the river. Besides this, no significant improvement in the river can be observed over these periods regardless several efforts by the Government probably because of the exponentially growing population pressure in this area.
6. From trend analysis of various parameters and various WQIs namely NSFWQI, CCMEWQI and the newly developed GWQI; Kanpur d/s (Jajmau) and Varanasi d/s (Malviya Bridge) have been identified as pollution hotspots.
7. The cost-benefit-analysis of the GWQI identifies BOD as the most sensitive parameter and its reduction or increment drastically affects the monetary benefits associated with GWQI.
8. The correlation analysis of the 9 parameters and GWQI shows that the most correlated factors are-
    - DO – BOD: Strong negative correlation (R = -0.8738)
    - DO- GWQI : Strong positive correlation (R= 0.9063)

- BOD - GWQI : Strong positive correlation (R= -0.8255)

9. As GWQI shows maximum correlation with DO and BOD, therefore these parameters are used to develop linear regression and ANN models in order to predict GWQI values corresponding to a DO or BOD value. The effectiveness of both the models have been compared and ANN proves to be a better model due to lesser error in predicted GWQI values, The MSE corresponding to DO- GWQI and BOD- GWQI linear regression models and DO- GWQI and BOD- GWQI ANN models were estimated to be 17.1250 e^{-0}, 35.7256 e^{-0}, 7.6830 e^{-0} and 24.1811e^{-0} respectively.
10. The lowest MSE value is observed for DO-GWQI ANN model and thus it is the most accurate out of these models, while ANN proves to be better predictive model than linear regression model in any case.
11. It can be concluded that with increased cost effectiveness and reduced subjectivity, GWQI can be used effectively to evaluate status of water quality of river Ganga, India and at the same time the methodology can also be effectively utilized for any other basin/river/ water body after relevant customization. Besides this, the developed index is also flexible in terms of number of parameters.
12. The developed GWQI can also be used in case of extreme financial limitations too by using the DO- ANN model to predict associated GWQI value.

From this study, it can be seen that the developed GWQI has several advantages over its contemporaries and can be effectively used to represent the water quality of river Ganga in the study area. The developed index strikes a delicate balance between subjective and objective techniques. Thus the work carried out in this thesis can form the foundation of future research by forming the basis of customized indices for other parts of the world and other types of indices. There is great scope for GWQI in Indian context as it can form the basis of identification of pollution hotspots and thus basis of evaluation of Government interventions.

Keywords – Ganga; water quality index; principal component analysis; factor analysis; rating curves; arithmetic mean; harmonic mean; trend; cost- benefit analysis; linear regression; artificial neural network.