Development of a water quality index model -a comparative analysis of various weighting methods

Conference Paper · December 2021

4 authors:

Md Galal Uddin
University of Galway
62 PUBLICATIONS 1,429 CITATIONS

Agnieszka Indiana Olbert
University of Galway
74 PUBLICATIONS 1,803 CITATIONS

Stephen Nash
University of Galway
112 PUBLICATIONS 2,224 CITATIONS

Azizur Rahman
Charles Sturt University
259 PUBLICATIONS 3,247 CITATIONS

MedGU Mediterranean Geosciences Union
Annual Meeting | Istanbul, 25–28 November 2021
Springer | Publishing Partner

Development of a water quality index model - a comparative analysis of various weighting methods

Md. Galal Uddin[1,2,3,*], Agnieszka I. Olbert[1,2,3], Stephen Nash[1,2,3], Azizur Rahman[4,5]

[1] Department of Civil Engineering, College of Science and Engineering, National University of Ireland Galway, Ireland

[2] Ryan Institute for Environmental, Marine and Energy Research, National University of Ireland Galway, Ireland

[3] MaREI Research Centre for Energy, Climate and Marine, National University of Ireland Galway, Ireland

[4] School of Computing and Mathematics, Charles Sturt University, Wagga Wagga, Australia

[5] Institute for Land, Water and Society, Charles Sturt University, Wagga Wagga, Australia

*u.mdgalal1@nuigalway.ie

Abstract. The Water Quality Index (WQI) model is a popular tool for the assessment of water quality (WQ) and its status classification. Over the last 50 years, many different WQI models have been developed; while the four-component structure remains common across the models, they vary in approaches in terms of parameter selection and their associated weightings, and the methods used for sub-indexing and aggregation. In this research, the parameter weight factors are investigated. The parameters are assigned weightings depending on their significance to the assessment. Recent research has revealed that that the WQI models generate a significant amount of uncertainty as a result of this process. Most of the WQI models widely used the expert opinion or analytical hierarchy process (AHP) to estimate the parameters' weight values. Many researchers have argued that these methods do not accurately represent the importance of parameters.

The aim of this study is to compare various statistical weighting approaches and propose the most accurate method for calculating weight values in terms of uncertainty. To estimate the weight value of parameters, four statistical attribute weighting methods were used: (i) rank sum (RS), (ii) rank reciprocal (RR), (iii) rank order centroid (ROC), and (iv) equivalent (EQ). In this study, the Guide to the Expression of Measurement Uncertainty (GUM) approach was used to quantify uncertainty of the weighting process by implementing Monte Carlo simulation (MCS) technique. The finding of this study show that the ROC method produces the highest uncertainty, while the RS and EQ methods generate relatively lower uncertainties. In terms of uncertainty, the results suggest that the rank sum method is an effective method to estimate the weight value in developing a WQI model.

Keywords: Water Quality Index Model; Monte Carlo simulation; Weight Values; Weighting Methods; Uncertainty; Random Forest, Rating Methods

1 Introduction

Water quality index (WQI) model is a popular tool due to its simple architecture and easy to application. Commonly, an ideal WQI model comprises including four components: (i) parameter selection, (ii) sub-index process, (iii) parameter weighting and (iv) aggregation function. This model is becoming increasingly popular and utilized to assess and classify water quality. Nevertheless, recent research has revealed that WQI has uncertainty problems in different stages [1-2]. The weighting technique is one of them. The present WQI used a variety of methods (such as expert opinions, analytical hierarchy process, literatures) for generating parameters weight values. These methods have received much criticism due to its reliability in terms of uncertainty.

The present study have been carried out using amalgamation approaches with machine-learning (ML) and multi-attributes (MA) statistical techniques. This approach was implemented by following two sequential steps. The first step of this process was to obtain the water quality rank for assigning degree of significant, then parameters weight values were estimated using four common rank-based weighting methods.

However, the purpose of this study is to explore an effective weighting method in order to reduce model uncertainty in developing a WQI model.

2 Materials and Methods

2.1 Data Description

For this study, data on water quality obtained from https://www.catchments.ie/data/. The Irish Environmental Protection Agency (EPA) has been collecting water quality data from 32 different monitoring sites in the Cork harbour in 2020 that was used in the current research as a case study. This study used the average concentration of water quality data from eleven parameters: ammonia (AMN), biological oxygen demand (BOD), dissolved oxygen (DO), chlorophyll a (CHL), molybdate reactive phosphorus (MRP), water pH, total organic carbon (TOC), total organic nitrogen (TON), water transparency (TRAN), dissolved inorganic nitrogen (DIN), and water temperature (TEMP)

2.2 Parameters ranking procedures

In this research, we conducted a study using ten widely used machine-learning (ML) algorithms for ranking water quality parameters based on their real-time relative importance. For ranking, this study utilized surface water quality guidelines that adopted by the Environmental Protection Agency (EPA) of Ireland. The random forest (RF) algorithm was found to be the most efficient for assigning parameter rankings [3]. In this research, the RF is used in order to provide the rank of water quality parameters. The RF prediction function defined as follows:

$$f(x) = \sum_{m=1}^{M} c_m \prod (x, R_m) \qquad (1)$$

where *M* denotes the number of regions in the features space, R_m is the region that corresponds to m, c_m is a constant that relates to m.

$$\prod(x, R_m) = \begin{cases} 1, & if\ x, R_m \\ 0, & otherwise \end{cases} \qquad (2)$$

The Gini scores are obtained by voting scores of each decision tree. Finally, the important features are chosen from the new subset based on the Gini score of the features [3].

2.3 Estimating parameters weight values

The present study was used the four direct rating subjective methods to estimate weight value of water quality parameters. These methods may avoid the problems highlighted by recent studies. These functions are defined as follow:

(i) Rank Sum (RS) method: $w_i = \frac{n+1-i}{\sum_{j=1}^{n} j} = \frac{2(n+1-i)}{n(n+1)}, i = 1,2,3 \ldots \ldots., n$ (3)

(ii) Rank Reciprocal (RR) method: $w_i = \frac{1/i}{\sum_{j=1}^{n} 1/j}, i = 1,2,3 \ldots \ldots \ldots, n$ (4)

(iii) Rank Order Centroid (ROC) method: $w_i = \frac{1}{n}.\sum_{j=1}^{n} \frac{1}{j}, i = 1, 2, 3 \ldots \ldots \ldots, n$ (5)

(iv) Equal Weight (EQ) method: $w_i = \frac{1}{n}, i = 1,2,3 \ldots \ldots \ldots, n$ (6)

where, *n* is the total number of attributes; *i* is the i^{th} attribute rank; *j* is the summation of rank and *w* is the weight value.

The present study was used Monte Carlo simulation (MCS) approaches by following the Guide to the Expression of Measurement Uncertainty (GUM) method to calculate uncertainties in different weighting methods. In this study, the MCS run 103 trials for MC simulation.

3 Results

3.1 Ranks of water quality parameters

Figure 1 presents the rank of different water quality parameters were obtained from the RF method. The results showed that the TOC, TON, TRAN and DIN are most important parameters in terms of water quality standards. AMN and temperature were found to have no significance in the analysis of water quality for this study.

3.2 Comparative analysis of weighting methods

In this study, parameter weight values were estimated using four weighting methods. In terms of weight variability of parameter, there was a significant difference between the four weighting methods. Table 1 shows the weight values of different water quality variables for different techniques respectively.

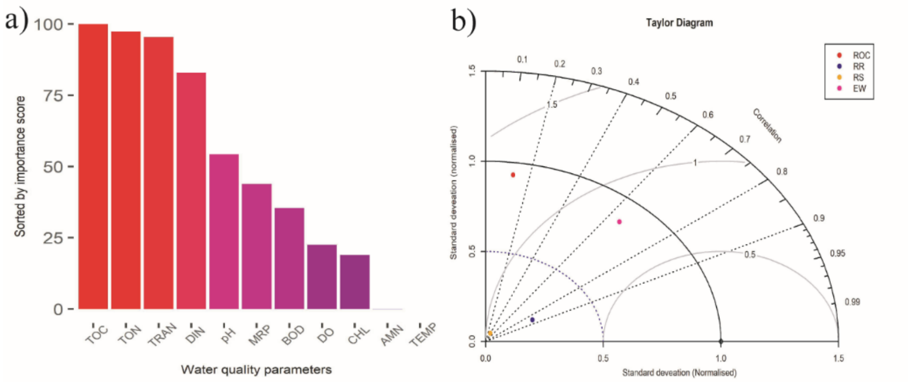

Fig. 1. (a) Importance of water quality parameters; and (b) Cross-validation of the four weighting methods using Taylor diagram.

Table 1. Estimating weight values of water quality parameters.

WQ parameters	Rank	ROC	RR	RS	EW
TOC	1	0.4083	0.4082	0.2857	0.1667
TON	2	0.2417	0.2041	0.2381	0.1667
TRAN	3	0.1583	0.1361	0.1905	0.1667
DIN	4	0.1028	0.1020	0.1429	0.1667
pH	5	0.0611	0.0816	0.0952	0.1667
MRP	6	0.0278	0.0680	0.0476	0.1667

3.3 Uncertainty analysis

The current study investigated into a significant difference in weight variation of water quality parameters between weighting methods. Table 2 shows the summary statistics of probability distributions of ROC,RR, and RS weighting methods were triangular and EW method was normal distribution.

Table 2. Different-weighting methods and related PDFs

Weighting methods	PDFs	PDF parameters
ROC	Triangular	a = 0.0278; b = 0.4083; c = 0.16667
RR	Triangular	a = 0.0680; b = 0.4082; c = 0.16667
RS	Triangular	a = 0.0476; b = 0.2857; c = 0.16667
EW	Normal	Mean = 0.1667; SD = 0.00

As seen in Figure 2, it shows the summary of MCS results. Table 3 provides comparative statistical scenarios between various weighting procedures. According to MCS results, the ROC method produced the highest weighting uncertainty, while the RS method produced the lowest. Table 4 presents the uncertainty results obtained from the MCS analysis for different weighting methods.

Table 3. Descriptive statistics of different weighting techniques.

Ranking methods	Descriptive statistics				
	Minimum	Maximum	Mean	SE	Std.
ROC	0.0278	0.4083	0.1667	0.0573	0.1404
RR	0.0680	0.4082	0.1667	0.0522	0.1279
RS	0.0476	0.2857	0.1667	0.0364	0.0891
EQ	0.1667	0.1667	0.1667	0.0000	0.000

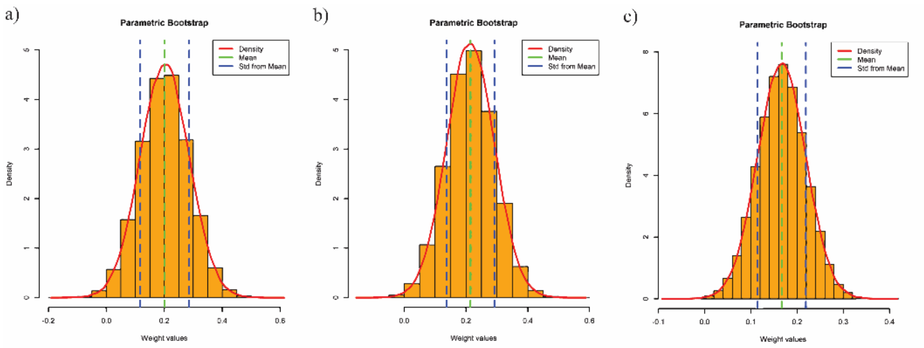

Fig. 2. Density plots for input PDFs of water quality parameters obtained from by MCS method: (a) ROC; (b) RR; (c) RS

Table 4. Uncertainty with 95% coverage probability of different weighting methods.

Ranking methods	Uncertainty parameters			
	SU[1]	EDF[2]	CF[3]	EU[4]
ROC	0.0320	5	2.57	0.0823
RR	0.0292	5	2.57	0.0751
RS	0.0198	5	2.57	0.0509
EQ	0.0	5	2.57	0.0

3.4 Cross-validation of uncertainty results

The effectiveness of the methods was compared and visualized using a Tylor diagram. The correlation coefficient (R), root mean square difference (RMSD), and standard deviation (SD) between any prediction and observation weight values are shown in this diagram. The RS method, as shown in Figure 1b, falls into lines with the lowest RMSD and SD between predicted and observed weighted values. The findings suggest that the RS method could help improve the weighting system's accuracy.

4 Discussion

As mentioned in the recent research, the weighting method contributes to the uncertainty in the WQI model. Six water quality parameters out of eleven were found to be significant in the current study in order to assess water quality in Cork Harbour, Ireland. The finding confirms that different weighting methods resulted in significant differences in parameter weight values. Figure 1b shows the summary statistics for different weighting methods. In contrast to four weighting methods, different methods show significant variation, which is consistent with previous literature. From the figure 1b, it can be seen that the RS method produced the lowest uncertainty relatively the highest weight variation was produced by ROC methods. However, recent research has revealed that the rank order centroid (ROC) weights lead to the highest performance in identifying the best alternative under the ranked attribute weights due to its sharpness and non- linear function. The findings of this study suggest that the rank sum method is an effective technique for estimating weight value in developing a WQI model.

5 Conclusions

The present study examined the performance of four statistical weighting methods in order to estimate weight values of water quality parameters. In terms of contribution to uncertainty in WQI model, the results indicate that the rank sum method is robust method to estimate the weight value than other methods. This new study could aid in reducing uncertainty and improving the WQI model's performance. The present study has confirmed the finding of recent studies, which revealed that the weighting method contributes to the uncertainty in WQI model. Therefore, according to analysis the weighting values had a slight significant impact on producing uncertainty in WQI model. Statistically, it could be held statistically accountable to estimate weight values in the WQI model due to this study was calculated less than two percent uncertainty in weighting process.

References

1. Sutadian, A.D., Muttil, N., Yilmaz, A.G., Perera, B.J.C., 2017. Using the Analytic Hierarchy Process to identify parameter weights for developing a water quality index. Ecol. Indic. 75, 220–233. https://doi.org/10.1016/j.ecolind.2016.12.043
2. Uddin, M.G., Nash, S., Olbert, A.I., 2021. A review of water quality index models and their use for assessing surface water quality. Ecol. Indic. 122, 107218. https://doi.org/10.1016/j.ecolind.2020.107218
3. Jović, A., Brkić, K., Bogunović, N., 2015. A review of feature selection methods with applications. 2015 38th Int. Conv. Inf. Commun. Technol. Electron. Microelectron. MIPRO 2015 - Proc. 1200–1205. https://doi.org/10.1109/MIPRO.2015.7160458

Acknowledgement

This research was funded by the Hardiman Research Scholarship of the National University of Ireland, Galway, which funded the first author as part of his PhD program. The authors would like to thank the Environmental Protection Agency of Ireland for providing water quality data.