Comparing the responses of two water quality indices using simulated and real data

Marco Antonio Ribeiro Pessoa¹, José Paulo Soares de Azevedo² e Patrícia Domingos³

¹, ²Civil Engineering Program, PEC/COPPE/UFRJ- Federal University of Rio de Janeiro, Rio de Janeiro, RJ, Brazil
pessoamarco54@gmail.com; zepaulo@coc.ufrj.br
³Department of Plant Biology, UERJ - Rio de Janeiro State University, Rio de Janeiro, RJ, Brazil
patvitesse@gmail.com

ABSTRACT

The dissemination of information on water quality for a non-specialized audience is essential to support programs and institutional policies aimed at the management of aquatic environments. In order to represent the extensive set of variables used to describe the water quality, this set can be synthesized into a single value or category information by using water quality indices. This research compares the Fuzzy Water Quality Index for Lotic Environments – IQAFAL, a new index of water quality based on fuzzy logic, with the traditional water quality index developed by the United States National Sanitation Foundation - WQI. This new index was developed for lotic environments through the collaboration of water quality experts of the Rio de Janeiro Environmental Agency - FEEMA. Both indices were applied to water quality data from the Paraíba do Sul River, obtained by INEA (the State Environmental Institute of Rio de Janeiro), from 2002 to 2009 in addition to a set of simulated water quality data to compare their responses in specific scenarios. The comparison between the results obtained using both IQAFAL and WQI showed that only IQAFAL was sensitive to the influence of a bad condition in one isolated variable, even if the other variables were in good condition. Thus, the IQAFAL methodology allowed the development of an index able to avoid the attenuation of the influence of a variable in critical condition by the influence of other variables in good conditions.

Keywords: Water Quality Index. Fuzzy Logic. Lotic System. Paraíba do Sul River.

Comparação das respostas de dois índices de qualidade de água usando dados simulados e reais

Recebido: 06/05/15 - Revisado: 19/06/15 - Aceito: 27/08/15

ABSTRACT

A divulgação de informações sobre qualidade das águas para um público não especialista é fundamental para subsidiar ações políticas e institucionais de gestão dos ambientes aquáticos. Para tanto, índices de qualidade de água têm sido propostos por serem capazes de sintetizar em um único valor ou categoria a informação, normalmente descrita a partir de um conjunto extenso de variáveis de qualidade de água. Este trabalho compara o Índice Fuzzy de Qualidade da Água para Ambientes Lóticos - IQAFAL, um novo índice de qualidade da água desenvolvido com base em lógica nebulosa, com o índice de qualidade da água tradicional desenvolvido pela United States National Sanitation Foundation - WQI. O novo índice, IQAFAL foi desenvolvido para ambientes lóticos com a colaboração dos especialistas em qualidade em água da Fundação Estadual de Engenheiria do Meio Ambiente do Rio de Janeiro. Ambos os índices foram aplicados a dados de qualidade da água do Rio Paraíba do Sul, obtidos pelo INEA (Instituto Estadual do Ambiente do Rio de Janeiro), 2002-2009, além de um conjunto de dados simulados de qualidade da água para comparar suas respostas em cenários específicos. A comparação entre os resultados obtidos utilizando tanto IQAFAL quanto o WQI mostrou que apenas IQAFAL foi sensível à influência de uma variável isolada em condição ruim, mesmo quando as outras variáveis se encontravam em boas condições. A metodologia utilizada na construção do IQAFAL permitiu o desenvolvimento de um índice capaz de evitar a atenuação da influência de uma variável em estado crítico pela influência de outras variáveis em boas condições.

INTRODUCTION

Water is a natural resource, essential for human life among other reasons because it is a substance whose daily consumption is vital for the survival of all individuals. Aside from being directly consumed, water is used in almost all activities developed by mankind such as irrigation, food production, industrial chemical processes, energy generation, sailing, landscaping, and more. Over the last decades, humanity has witnessed an increasing acceleration in the degradation of water resources available on the planet, a consequence of indiscriminate uses such as the discharge of liquid and solid residues, the destruction of flooded areas and riparian forests, and the growing reduction of the vegetation cover, which affects the availability of surface and underground waters (UNESCO, 2012).

Water resource management, therefore, has become an increasingly predominant issue and, among other purposes, it seeks to follow the conditions and availability of these resources, defines their uses, and suggests improvements.

However, the information obtained through the evaluation and interpretation of the data on water quality, which is necessary to follow the condition of bodies of water, is in general understandable almost exclusively by experts, making its use troublesome for non-specialized decision-makers who define the policies related to the use of water resources. Among the propositions that support the translation of the information produced by water quality experts into a language that is accessible to a non-technical audience is the development of water quality indices or indicators. These aim to integrate, in a single value, the original information obtained through several variables related to water quality.

Water Quality Indices

Horton (1965) was the first to develop water quality indices from a reflection on water quality variables.

The United States National Sanitation Foundation - NSF improved Horton (1965) methodology, developing the Water Quality Index – WQI by selecting water quality variables and their respective weights, based on the opinion of experts (PINHEIRO, 2004).

The selected variables were standardized in a scale of 0 to 100 and multiplied by their weighing factor as shown in equation 1.

\[
WQI = \frac{\sum C_i \times P_i}{\sum P_i}
\]

Where WQI is the weighted average of the pre-established parameters (Table 1) standardized on a scale from 0 to 100; C_i is the value of each parameter after standardizing; and P_i is the relative value of the weight for each parameter.

Fia et al. (2015) evaluate the spatial and temporal variation of water quality and trophic characteristics of the main watercourses in the watershed of Ribeirão Vermelho in Lavras – MG, Brazil using the Water Quality Index (IQA) and the Trophic State Index (IET) during the rainy season (summer) and the dry season (winter).

Almeida e Schwarzbold, (2003) evaluated water quality in Arroio da Cria, Montenegro, RS, Brazil between November 1997 and August 1998. The following variables were examined: dissolved oxygen saturation, biochemical oxygen demand, pH, total phosphorus, nitrate, fecal coliforms, turbidity, total suspended solids, electric conductivity, alkalinity, chloride, total chromium and temperature. The results were analyzed using the Water Quality Index from NSF.

Rosa, Oliveira e Saad (2014) evaluated the influence of urban expansion on the quality of the Cotia River water, in the municipality of Cotia, State of São Paulo, Brazil. The Water Quality Index and its parameters were used to look at statistical trends and fluctuations throughout the period of the study, using linear and polynomial regression and comparative correlation with urban expansion.

Most of the water quality indices were developed by water quality experts who applied statistical methodologies by selecting water quality variables and classifying their importance by attributing of weights.

However, traditional methodologies for calculating water quality indices have not proved to be efficient enough in representing a more subjective knowledge of the variables used to evaluate the quality of aquatic environments (LERMONTOV et al., 2009).

Fuzzy logic has been used as an alternative for modeling water quality indices, since it provides another approach in dealing with issues where the objectives are not well defined and the information is not precise (CHAU, 2006). Fuzzy logic is a methodology that has been used in the development of water quality indices, because it is able to capture with greater accuracy the knowledge of experts and the subjective perception they have acquired through their professional experience (ICAGA, 2007; LERMONTOV et al., 2009; OCAampo-DUQUE et al., 2006).

Moreover, this methodology is able to eliminate or minimize the eclipse effect, described as an attenuation of the influence of values of a given parameter in bad condition, in light of the balanced behavior of other parameters, a very common occurrence in traditional indices (SILVA; JARDIM, 2006).

Fuzzy Logic

Fuzzy Logic extended set theory and Boolean logic,
which treat the real world as having only two classes (true or false), by inserting the concept of partial truth, where an element belongs to a set with an associated degree of membership (ZADEH, 1965).

The concept of membership of an element within a set, in the classic set theory, is exclusive, so that for a given set A in a universe X, the elements of this universe simply belong or not to that set as shown in equation 2.

\[ f_\mu(x) = \begin{cases} 1 & \text{if and only if } x \in A \\ 0 & \text{if and only if } x \notin A \end{cases} \]  

(2)

As opposed to traditional set theory, the concept of membership in fuzzy logic allows an element to partially belong to more than one set at a time and not necessarily to a single set. In fuzzy logic, the function can assume any value in interval [0,1] so that a set A in a universe X is defined by the membership function \( \mu_A(x) : X \rightarrow [0,1] \) represented by a set of ordered pairs as shown in equation 3.

\[ A = \{ \mu_A(x)/x \} x \in A \]  

(3)

Where A determines the degree of membership of an element in set A.

Fuzzy logic introduces the concept of linguistic variables, whose values are linguistic terms such as “high”, “low”, “medium”, “very high”, etc., represented by fuzzy sets. When assuming linguistic terms as values, the linguistic variables enable a systematic characterization of complex or poorly defined phenomena. Thus, fuzzy logic allows us to process the subjective information of natural language, which is by nature vague or uncertain (ZADEH, 1973).

According to Zadeh (2009), “fuzzy logic adds an important capacity to bivalent logic, the capacity to precisely think about imperfect information”, and the author himself defines imperfect information as that which in one or more aspects is imprecise, uncertain, incomplete, vague, or partially true. One of the great advantages of this model lies in the possibility of capturing, with mathematical formalism, concepts derived from linguistic terms like, for instance, comfort, satisfaction, and others (OLIVEIRA, 1999).

When building a fuzzy system, the variables used are called linguistic variables. The linguistic terms represent fuzzy sets defined by membership functions. The membership functions are described by curves that define the outlines of the fuzzy sets. The semantic properties of the concept (linguistic term) are described by the outline of the respective fuzzy set; therefore, the closer the curve of the membership function is to the behavior of the phenomenon being studied, the better and more precise the performance of the fuzzy model in representing the real world.

Figure 1 graphically illustrates the concepts of linguistic variables, universe of discourse, linguistic terms, and membership function.

Figure 1 – Graphical representation of the concepts of linguistic variable, universe of discourse, linguistic terms, and membership function

The combination of these variables in its different conditions. This enables the creation of a model where the result from the combination of different variables has a non-linear behavior that is closer to the environmental phenomena.

This work compares a new index, the Fuzzy Water Quality Index for Lotic Environments - IQA_{FAL} (PESSOA, 2010) based on fuzzy logic and developed through the contribution of water quality experts from the “Instituto Estadual do Ambiente” - INEA (Rio de Janeiro State Environmental Agency) with the traditional Water Quality Index – WQI (PINHEIRO, 2004).

MATERIAL AND METHODS

The Fuzzy Water Quality Index for Lotic Environments - IQA_{FAL}.

The IQA_{FAL} uses seven water quality variables (Table 2) in its formulation.

Table 2 - Water quality variables used for IQA_{FAL}

<table>
<thead>
<tr>
<th>Type</th>
<th>Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>Biologic</td>
<td>Shannon-Weaver Diversity Index (SHANNON, 1948)</td>
</tr>
<tr>
<td></td>
<td>Density of Cyanobacteria</td>
</tr>
<tr>
<td>Nutrients</td>
<td>Total Phosphorus</td>
</tr>
<tr>
<td></td>
<td>Ammoniacal Nitrogen</td>
</tr>
<tr>
<td>Oxygen</td>
<td>Dissolved Oxygen</td>
</tr>
<tr>
<td></td>
<td>Biochemical Oxygen Demand</td>
</tr>
<tr>
<td>Bacteriological</td>
<td>Fecal Coliform (Thermo-tolerant)</td>
</tr>
</tbody>
</table>

Two biological variables describe the conditions of the aquatic ecosystem based on quantitative and qualitative fluctuations of the phytoplankton community. Two chemical variables...
variables represent the eutrophication potential, as well as the degree of contamination from domestic discharges. Two variables represent the oxygen dynamics (availability and consumption). One variable indicates the degree of water contamination from sewage, which indirectly represents the risk of contamination by waterborne diseases. The criteria to the selection of these parameters following the recommendations of a water quality experts group from INEA formed with aim of develop the IQAFAL index. Besides these parameters belong the water quality indicators from Brazilian legislation.

In the IQAFAL fuzzy system the linguistic terms with their fuzzy sets defined by membership function map a domain of interest to the interval [0, 1] (Figure 2).

The values of the input variables are fuzzified into a qualitative state and processed by an inference engine through a rules base named fuzzy associative memory (Table 3).

The IQAFAL uses fuzzy subsystems with two input variables, dividing the seven chosen water quality variables into groups of two, which generated fuzzy subsystems, called sub-indices, which in turn is used as input for the final index (Figure 3).

The IQAFAL, like the WQI, ranges from 0 to 100, where 0 represents the poorest quality and 100 the best quality. For the purpose of evaluating the results of these indices, this interval was divided into five categories, namely: VERY BAD, BAD, REGULAR, GOOD, and EXCELLENT, with the poorest results falling within the category VERY BAD, and the best results falling within the category EXCELLENT (Table 4).

**Figure 2. The linguistic terms with their fuzzy sets and respective membership function created for the IQAFAL input variable Thermotolerant Coliform**

**Figure 3. The IQAFAL - System Flow showing the input variables and the sub-index**

<table>
<thead>
<tr>
<th>Oxygen Subindex</th>
<th>Thermostolerant Coliform</th>
</tr>
</thead>
<tbody>
<tr>
<td>EXCELENT</td>
<td>Excellent</td>
</tr>
<tr>
<td>GOOD</td>
<td>Good</td>
</tr>
<tr>
<td>REGULAR</td>
<td>Good</td>
</tr>
<tr>
<td>BAD</td>
<td>Regular</td>
</tr>
<tr>
<td>VERY BAD</td>
<td>Regular</td>
</tr>
</tbody>
</table>

**Table 3. The rules base created for the IQAFAL Domestic Discharge Subindex with the input variables Thermotolerant Coliform and Oxygen Subindex**

**Table 4 - IQAFAL and WQI five categories**

<table>
<thead>
<tr>
<th>Category</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Excellent</td>
<td>90 &lt; Indicator ≤ 100</td>
</tr>
<tr>
<td>Good</td>
<td>70 &lt; Indicator ≤ 90</td>
</tr>
<tr>
<td>Regular</td>
<td>50 &lt; Indicator ≤ 70</td>
</tr>
<tr>
<td>Bad</td>
<td>25 &lt; Indicator ≤ 50</td>
</tr>
<tr>
<td>Very Bad</td>
<td>0 ≤ Indicator ≤ 25</td>
</tr>
</tbody>
</table>

Table 4 - IQAFAL and WQI five categories
RESULTS

Using simulated data

We compare the results of the behavior of IQA_FAL and WQI in different simulated water quality conditions of Fecal Coliform contamination represented by values varying from 0NPM/100mL to 4000NPM/100mL. The bacteriological variable Thermo-tolerant Coliform was chosen since it is an important indicator of contamination by warm-blooded animal feces. The presence of these organisms is a strong indication of the discharge of untreated sewage into the body of water. The importance of the presence of thermo-tolerant coliforms in the water is directly linked to the possibility of contamination by waterborne diseases. This is an important indicator of a public health risk and the level of sanitary conditions of the river basin.

The results of the behavior of IQA_FAL reveal that the fuzzy index is able to produce more sensitive responses than WQI to different levels of a single variable. Figures 4, 5 show the simulations of IQA_FAL and WQI results using Thermo-tolerant Coliform data ranging between 0 and 4000 (NPM/100mL).

In this simulation, the values used for the other index input variables are those observed in excellent and in regular water quality condition, respectively (Table 5). The values for excellent and regular condition are chosen based on the national legislation that regulates bodies of water (BRASIL, 2005) and

<table>
<thead>
<tr>
<th>Variables</th>
<th>Excellent water quality condition</th>
<th>Regular water quality condition</th>
</tr>
</thead>
<tbody>
<tr>
<td>DO (mg/L)</td>
<td>8,0</td>
<td>5,0</td>
</tr>
<tr>
<td>DBO (mg/L)</td>
<td>0,5</td>
<td>4,5</td>
</tr>
<tr>
<td>Total Phosphorus (mg/L)</td>
<td>0,001</td>
<td>0,1</td>
</tr>
<tr>
<td>Ammoniacal Nitrogen (mg/L)</td>
<td>0,01</td>
<td>0,5</td>
</tr>
<tr>
<td>Shannon-Weaver Diversity Index</td>
<td>4,5</td>
<td>2,5</td>
</tr>
<tr>
<td>Density of Cyanobacteria (cels/ml)</td>
<td>50</td>
<td>35000</td>
</tr>
<tr>
<td>Total Nitrogen (mg/L)</td>
<td>0,01</td>
<td>10,0</td>
</tr>
<tr>
<td>pH</td>
<td>7,1</td>
<td>6,0</td>
</tr>
<tr>
<td>Total Solids (mg/L)</td>
<td>80</td>
<td>400</td>
</tr>
<tr>
<td>Turbidity (NTU)</td>
<td>0,1</td>
<td>35</td>
</tr>
<tr>
<td>Water Temperature (°C)</td>
<td>22,0</td>
<td>22</td>
</tr>
<tr>
<td>Air Temperature (°C)</td>
<td>22,0</td>
<td>22</td>
</tr>
</tbody>
</table>
their potability (BRASIL, 2004). Besides that, the knowledge and expertise accumulated by INEA institutional team were considered.

We observe that the IQAFAL responds better than WQI to different levels of Thermo-tolerant Coliform when the other variables are in excellent conditions as well as in regular conditions. The simulation results show that the IQAFAL values decrease as the Thermo-tolerant Coliform levels increase, varying from 98 to 38 (Excellent to Bad) when the other index input variables are in excellent conditions, and from 59 to 10 (Regular to Very Bad) when they are in regular conditions. On the other hand, the results of WQI do not reflect variations in the Thermo-tolerant Coliform levels, staying within 90 and 70 (Good to Good) when the other index input variables are in excellent conditions, and 60 to 47 (Regular to Bad) when they are in regular conditions.

Using real data

According to Pessoa, 2010, the application of the IQAFAL to the data obtained by “Instituto Estadual do Ambiente” - INEA (Environmental Agency of the state of Rio de Janeiro) from the sampling stations along stretches of the Paraíba do Sul River and the Guandu River, Rio de Janeiro, Brazil in the period (2002-2009) showed that the index results were in line with the perception of water quality experts from this institution regarding the quality of these aquatic environments (FEEMA, 2002; INEA, 2013, 2014).

The Paraíba do Sul River flows across 3 States in Brazil. In the State of Rio de Janeiro, the Paraíba do Sul River flows through 37 cities with an extension reaching 500 km, making it the most important river in the state, as it supplies water to 85% of the population in the Rio de Janeiro Metropolitan Area. Problems resulting from the lack of domestic sewage treatment in most of the cities are an important factor in the degradation of the water quality in the rivers of the Paraíba do Sul Basin, as well as a health risk for the population. The data used in this study refers to the first stretch of the Paraíba do Sul river in the State of Rio de Janeiro, located between the Funil Reservoir and the city of Barra do Piraí (Figure 6).

The highest percentage frequency of the IQAFAL results for sampling stations located along the Paraíba do Sul and Guandu Rivers occurs in the category “Bad”. The sampling station located at the exit of the Funil Reservoir is an exception, and the highest percentage obtained from the IQAFAL results at that location falls within the category “Good” (Figure 7).

Figure 6 - Map localizing the sampling stations along the stretch of the Paraíba do Sul River where the water quality data were obtained
Figure 7 - Map with percentages from the IQAFAL results for each category at the sampling stations along a stretch of the Paraíba do Sul River in the period (2002-2009)

Figure 8 - Map with percentages from the WQI results for each category at the sampling stations along a stretch of the Paraíba do Sul River in the period (2002-2009)
In this study the IQA_FAL also proved to be more sensitive in the detection of changes in water quality, when compared to the traditional WQI. The IQA_FAL proved itself able to reflect the perception that some water quality parameters, depending on the range in which they are found, need to determine the index result in a preponderant way, regardless of the values of the other parameters. This is justified by the fact that for some parameters, certain values are determinant in the evaluation of water quality; in other words, it is not possible to allow the values of other parameters to attenuate the final result. The study’s results suggest that the methodology used in formulating the WQI, calculated through a weighted average of the normalized values for water quality parameters, was not able to avoid the eclipse effect. The eclipse effect is described as the attenuation of very poor values for a specific parameter, in the face of a balanced behavior by other parameters (SILVA; JARDIM, 2006). Although WQI incorporates the subjective knowledge of experts by attributing different weights to the various input parameters, this index has shown itself to be influenced by this effect.

When comparing the results of these two indices, we noticed that the WQI proved to be less restrictive than the IQA_FAL for the ensemble of data used. However, for the sampling station located at the exit of the Funil Reservoir, the highest percentage frequency from the WQI results follows the behavior of the other stations and occurs within the category “REGULAR”, while the highest percentage from the IQA_FAL results for this station fall within the category “GOOD” (Figure 8).

CONCLUSIONS

As described in (FEEMA, 2002), the water quality in this stretch of the river does not present critical conditions in terms of organic pollution, which is most likely due to the river’s own self-depuration capacity. The DO and BOD values, in general, fall within the standards of Class II (BRASIL, 2005) that establish 5 mg/l as the maximum acceptable concentrations of DBO and 5 mg/l as the minimum acceptable concentrations of DO. Nevertheless, the water quality is affected by the impact from the discharge of sanitary sewage, which is validated by the high concentrations of Thermo-tolerant Coliforms (FEEMA, 2002; INEA, 2013).

The IQA_FAL proved to be more sensitive in the detection of changes in water quality, when compared to the WQI for the data used in this study. Through the thematic maps, we observed that in almost all sampling stations the highest percentage frequency of the IQA_FAL results is in the classes “BAD” and “VERY BAD”, while the highest percentage frequency of the WQI results is in class “REGULAR”.

The WQI, whose calculation also includes the parameter Thermo-tolerant Coliform with the second biggest weight, shows a smaller sensitivity to the high values for this parameter in almost all samples. The study’s results suggest that the methodology used in formulating the WQI, calculated through a weighted average of the normalized values for water quality parameters, was not able to avoid the effect of attenuating very poor values for a specific parameter, in the face of a balanced behavior by other parameters, the eclipse effect (SILVA; JARDIM, 2006), even though it did incorporate the subjective knowledge of experts by attributing different weights to the various input parameters.

Throughout the development of the IQA_FAL, the importance of a methodology capable of incorporating the experts’ perceptions became clear. Their perceptions confirm that some water quality parameters, depending on the range in which they are found, need to determine the index’s result in a preponderant way, regardless of the values of the other parameters. This is justified by the fact that for some parameters, certain values are determinant in the evaluation of water quality; in other words, it is not acceptable to allow the values of other parameters to attenuate the final result. An example, in this report, was the parameter Thermo-tolerant Coliform. For the water quality experts involved in this study, the IQA_FAL result for samples where the Thermo-tolerant Coliform values were above 2500 NMP/100ml must necessarily be low, even if the results for other parameters used in the index were in ranges considered good.

The sampling station located at the exit of the Funil Reservoir is the only station where the highest percentage frequency of the IQA_FAL results differs from that of other stations, and occurs in the category “GOOD”. This station is located at the exit of the Funil Reservoir, allowing the rate of Thermo-tolerant Coliforms to drop. Since Thermo-tolerant Coliforms proved to be the parameter with the greatest influence over the IQA_FAL results, they placed in a better category for this station.

REFERENCES


Contribuição dos autores
Marco Antonio Ribeiro Pessoa: Idealização do tema, coordenação geral das atividades, desenvolvimento e utilização do IQA_TAL: Índice Fuzzy de Qualidade de Água, redação final do artigo.
José Paulo Soares de Azevedo: Idealização do tema, escolha dos valores utilizados na parametrização dos dados, redação do artigo.
Patrícia Domingos: Participação no desenvolvimento do IQA_TAL, coleta, revisão e análise dos dados, redação do artigo.