We are IntechOpen, the world’s leading publisher of Open Access books
Built by scientists, for scientists

4,400
Open access books available

117,000
International authors and editors

130M
Downloads

154
Countries delivered to

TOP 1%
Our authors are among the most cited scientists

12.2%
Contributors from top 500 universities

WEB OF SCIENCE™
Selection of our books indexed in the Book Citation Index in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com
A Fuzzy Water Quality Index for Watershed Quality Analysis and Management

André Lermontov¹², Lidia Yokoyama¹, Mihail Lermontov³ and Maria Augusta Soares Machado⁴
¹Univerisidade Federal do Rio de Janeiro
²Grupo Águas do Brasil S/A
³Universidade Federal Fluminense
⁴IBMEC-RJ
Brazil

1. Introduction

Climate change and hydric stress are limiting the availability of clean water. Overexploitation of natural resources has led to environmental unbalance. Present decisions relative to the management of hydric resources will deeply affect the economy and our future environment. The use of indicators is a good alternative for the evaluation of environmental behavior as well as a management instrument, as long as the conceptual and structural parameters of the indicators are respected.

The use of fuzzy logic to study the influence and the consequences of environmental problems has increased significantly in recent years. According to Silvert (1997), most activities, either natural or anthropic, have multiple effects and any environmental index should offer a consistent meaning as well as a coherent quantitative and qualitative appraisal of all these effects.

Among the several reasons for applying fuzzy logic to complex situations, the most important is probably the need to combine different indicators. Maybe the most significant advantage of the use of fuzzy logic for the development of environmental indicators is that it combines different aspects with much more flexibility than other methods, such as, for example, binary indices of the kind “acceptable vs. unacceptable.”

Methods to integrate several variables related to water quality in a specific index are increasingly needed in national and international scenarios. Several authors have integrated water quality variables into indices, technically called Water Quality Indices (WQIs) (Bolton et al., 1978; Bhargava, 1983; House, 1989; Mitchell, 1996; Pesce and Wunderlin, 1999; Cude, 2001; Liou et al., 2004; Said et al., 2004; Silva and Jardim, 2006; Nasiri et al., 2007). Most are based in a concept developed by the U. S. National Sanitation Foundation (NSF, 2007).

There is an obvious need for more advanced techniques to assess the importance of water quality variables and to integrate the distinct parameters involved. In this context, new, alternative integration methods are being developed. Artificial Intelligence has thus become a tool for modeling water quality (Chau, 2006). Traditional methodologies cannot classify and quantify environmental effects of a subjective nature or even provide formalism for
dealing with missing data. Fuzzy Logic can combine these different approaches. In this context, new methodologies for the management of environmental variables are being developed (Silvert, 1997, 2000).

The main purpose of this research is to propose a new water quality index, called Fuzzy Water Quality Index (INQA – Índice Nebuloso de Qualidade da Água, originally in Portuguese), to be computed using Fuzzy Logic and Fuzzy Inference tools. A second goal is to compare statistically the INQA with other indices suggested in the literature using data from hydrographic surveys of four different watersheds, in São Paulo State, Brazil, from 2004 to 2006 (CETESB, 2004, 2005, 2006).

2. Background

2.1 Water quality indices

The purpose of an index is not to describe separately a pollutant’s concentration or the changes in a certain parameter. To synthesize a complex reality in a single number is the biggest challenge in the development of a water quality index (IQA – Índice de Qualidade de Água, originally in Portuguese), since it is directly affected by a large number of environmental variables. Therefore, a clear definition of the goals to be attained by the use of such an index is needed. The formulation of a IQA may be simplified if one considers only the variables which are deemed critical for a certain water body. Among their advantages, indices facilitate communication with lay people. They are considered more trustworthy than isolated variables. They also integrate several variables in a single number, combining different units of measurement.

In a groundbreaking work, Horton (1965) developed general water quality indices, selecting and weighting several parameters. This methodology was then improved by the U.S. National Sanitation Foundation (NSF, 2007). The conventional way to obtain a IQA is to compute the weighted average of some predefined parameters, normalized in a scale from 0 to 100 and multiplied by their respective weights.

Conesa (1995) modified the traditional method and created another index, called Subjective Water Quality Index (IQA_{sub}), that includes a subjective constant, k. This constant assumes values between 0.25 and 1.00 at intervals of 0.25, with 0.25 representing polluted water and 1.00 a not polluted one. The parameters used to calculate this index (eq. 1) must be previously normalized using curves given by Conesa (1995). The Objective Water Quality Index (IQA_{obj}) results from the elimination of the subjective constant k.

\[
IQA_{sub} = k \frac{\sum C_i \times P_i}{\sum P_i} \tag{1}
\]

where:
- k is the subjective constant (0.25, 0.50, 0.75, and 1.00);
- C_i, the value of the i<sup>th</sup> normalized parameter (Conesa, 1995);
- P_i, the relative weight of the i<sup>th</sup> parameter (Conesa, 1995).

The Brazilian IQA is an adaptation from the NSF index. Nine variables, being the most relevant for water quality evaluation, are computed as the weighted product (eq. 2) of the normalized values of these variables, n; Temperature (TEMP), pH, Dissolved Oxygen (DO), Biochemical Oxygen Demand (BOD_5), Thermotolerant Coliforms (TC), Dissolved Inorganic Nitrogen (DIN), Total Phosphorus (TP), Total Solids (TS) and Turbidity (T). Each parameter
is weighted by a value \( w_i \) between 0 and 1 and the sum of all weights is 1. The result is expressed by a number between 0 and 100, divided in 5 quality ranges: (100 - 79) - Excellent Quality; (79 - 51) - Good Quality; (51 - 36) - Fair Quality; (36 - 19) - Poor Quality; [19 - 0] - Bad Quality, normalization curves for each variable, as well as the respective weights, are available in the São Paulo’s State Water Quality Reports (CETESB, 2004, 2005 and 2006).

\[
IQ_{\text{CETESB}} = \sum_{i=1}^{n} w_i \cdot IQ_i
\]

Silva and Jardim (2006) used the concept of minimum operator to develop their index, called Water Quality Index for protection of aquatic life (IQAPAL). The IQAPAL (eq. 3) is based on only two parameters, Total Ammonia (TA) and Dissolved Oxygen (DO):

\[
IQ_{\text{PAL}} = \min (TA, DO) \quad (3)
\]

A fourth index, called IQA\(_{\text{min}}\), proposed by Pesce and Wunderlin (2000), is the arithmetic mean (eq. 4) of three environmental parameters, Dissolved Oxygen (DO), Turbidity (T) and Total Phosphorus (TP), normalized using Conesa's curves (Conesa, 1995).

\[
IQ_{\text{min}} = \frac{DO + T + TP}{3} \quad (4)
\]

Other indices are found in the literature and will not be considered in this study (Bordalo et al., 2001; SDD, 1976; Stambuk Giljanovic, 1999).

### 2.2 Fuzzy inference

One of the research fields involving Artificial Intelligence - AI is fuzzy logic, originally conceived as a way to represent intrinsically vague or linguistic knowledge. It is based on the mathematics of fuzzy sets (Zadeh, 1965). Fuzzy inference is the result of the combination of fuzzy logic with expert systems (Yager, 1994). The commonest models used to represent the process of classification of water bodies are called deterministic conceptual models. They are deterministic because they ignore the stochastic properties of the process and conceptual because they try to give a physical interpretation to the several subprocesses involved. These models often use a large number of parameters, making modeling a complex and time demanding task (Barreto, 2001).

Models based on fuzzy rules are seen as adequate tools to represent uncertainties and inaccuracies in knowledge and data. These models can represent qualitative aspects of knowledge and human inference processes without a precise quantitative analysis. They are, therefore, less accurate than conventional numerical models. However, the gains in simplicity, computational speed and flexibility that result from the use of these models may compensate an eventual loss in precision (Bárdossy, 1995).

There are at least six reasons why models based on fuzzy rules may be justified: first, they can be used to describe a large variety of nonlinear relations; second, they tend to be simple, since they are based on a set of local simple models; third, they can be interpreted verbally and this makes them analogous to AI models; fourth, they use information that other methods cannot include, such as individual knowledge and experience; fifth, the fuzzy approach has a big advantage over other indices, once they have the ability expand and combine quantitative and qualitative data that expresses the ecological status of a river,
allowing to avoid artificial precision and producing results that are more similar to the ecological complexity and real world problems in a more realistic panorama; and sixth, fuzzy logic can deal with and process missing data without compromising the final result. The way systems based on fuzzy rules have been successfully used to model dynamic systems in other fields of science and engineering suggests that this approach may become an effective and efficient way to build a meaningful IQA.

Fuzzy inference is the process that maps an input set into an output set using fuzzy logic. This mapping may be used for decision making or for pattern recognition. The fuzzy inference process involves four main steps: 1) fuzzy sets and membership functions; 2) fuzzy set operations; 3) fuzzy logic; and 4) inference rules. These concepts are discussed in depth in Bárdossy (1995), Yen e Langari (1999), Ross (2004), Cruz (2004) and Caldeira et al. (2007).

The concept of fuzzy sets for modeling water quality was considered by Dahiya (2007), Nasiri et al. (2007) Chau (2006), Ocampo-Duque et al. (2006), Icaga (2007), and Chang et al. (2001), Lermontov et al. (2009), Ramesh et al. (2010), Taner et al. (2011).

2.3 Development of the fuzzy water quality index (INQA)
The fuzzy sets were defined in terms of a membership function that maps a domain of interest to the interval [0,1]. Curves are used to map the membership function of each set. They show to which degree a specific value belongs to the corresponding set (eq. 5):

\[ \mu_A : X \rightarrow [0,1] \]  

(Fig. 1) are used in this study, for the same nine parameters used by CETESB to calculate its IQA, so that this methodology can be statistically compared and validated. The data shown in Tables 1 and 2 are used according to Figure 1 to create the fuzzy sets:

In a rule based fuzzy system, a linguistic description is attributed to each set. The sets are then named according to a perceived degree of quality, that ranges from very excellent to very bad (Tables 1 and 2). For the parameters temperature and pH, two sets for each linguistic variable are used. Temperature and pH sets have the same linguistic terms above and under the Very Excellent point while distancing from it. The sets under are marked with a ▼ symbol. The trapezoidal function is only used for the Very Excellent linguistic variable and the triangular for all others. This study uses the linguistic model of fuzzy inference, where the input data set, the water quality variables, called antecedents, are processed using linguistic if/then rules to yield an output data set, the so-called consequents.
Table 1. Fuzzy sets and linguistic terms for input parameters of Group 01, 02 and 03

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Symbol Unit</th>
<th>Gr01</th>
<th>Gr02</th>
<th>Gr03</th>
</tr>
</thead>
<tbody>
<tr>
<td>Temperature</td>
<td>Temp</td>
<td>a</td>
<td>b</td>
<td>c</td>
</tr>
<tr>
<td>pH</td>
<td>-6 - 45</td>
<td>a</td>
<td>b</td>
<td>c</td>
</tr>
<tr>
<td>Inorg. Nitrogen</td>
<td>mg/l</td>
<td>0 - 100</td>
<td>a</td>
<td>b</td>
</tr>
<tr>
<td>Phosphorus</td>
<td>mg/l</td>
<td>0 - 10</td>
<td>a</td>
<td>b</td>
</tr>
<tr>
<td>Total Solids</td>
<td>mg/l</td>
<td>0 - 150</td>
<td>a</td>
<td>b</td>
</tr>
<tr>
<td>Turbidity</td>
<td>mg/l</td>
<td>0 - 150</td>
<td>a</td>
<td>b</td>
</tr>
<tr>
<td>Coli Colonies/100ml</td>
<td></td>
<td>0 - 100</td>
<td>a</td>
<td>b</td>
</tr>
</tbody>
</table>

Table 2. Fuzzy sets and linguistic terms for input parameters of Group 04 and 05 and output parameters of all groups

www.intechopen.com
Figure 2 shows the flow graph of the process, where the individual quality variables are processed by inference systems, yielding several groups normalized between 0 and 100. The groups are then processed for a second time, using a new inference, and the end result is the Fuzzy Water Quality Index – INQA/FWQI.

In the traditional methods used to obtain a IQA, parameters are normalized with the help of tables or curves and weight factors (Conesa, 1995; Mitchell, 1996; Pesce and Wunderlin, 1999; CETESB, 2004, 2005 and 2006; NSF, 2007) and then calculated by conventional mathematical methods, while in this work, parameters are normalized and grouped through a fuzzy inference system.

In a fuzzy inference system a quantitative numerical value is fuzzyfied into a qualitative state and processed by an inference engine, through rules, sets and operators in a qualitative sphere, allowing the use of information that other methods cannot include, such as individual knowledge and experience (Balas et al., 2004), permitting qualitative environmental parameters and factors to be integrated and processed (Silvert, 2000) producing similar to the real world results.

A rule in the inference system is a mathematical formalism that translates expert judgment expressed in linguistic terms (as in NFS’s IQA formulation) and therefore is a subjective and qualitative weight factor in the inference engine. I.e.: Rule 1: if Thermotolerant Coliform is very high and pH is lower than average than index is very poor; Rule 2: if Thermotolerant Coliform is very high and pH is excellent than index is poor. One can notice that these rules have been designed as an expert system and a subjective and qualitative weight factor based on an
expert judgment has been introduced in the process scoop. In spite of the strong pH variation, the final score is not strongly affected.

The physical parameters pH and Temp are normalized and aggregated into the first group (Gr01). DO and BOD comprise Gr02. Thermotolerant coliforms (Coli) were independently normalized as Gr03. The nutrients DIN and TP make up Gr04; TS and Turb are grouped in Gr05. The water analyses results used in this research were taken from the CETESB reports for the years of 2004, 2005 and 2006 (CETESB, 2004, 2005 and 2006). Curves to help in the creation and normalization of the fuzzy sets were taken these reports for the parameters pH, BOD, Coli, DIN, TP, TS and Turb and from Conesa (1995) for Temp and DO.

The rules for normalization and aggregation followed the logic described below and the consequent always obeyed the prescription of the minimum operator:

\[
\begin{align*}
\text{If } FP & \text{ is VE and } SP \text{ is VE then GR output is VE} \\
\text{If } FP & \text{ is VE and } SP \text{ is E then GR output is E} \\
\text{If } FP & \text{ is E and } SP \text{ is VE then GR output is E} \\
\text{If } FP & \text{ is VE and } SP \text{ is VP then GR output is VP} \\
\text{If } FP & \text{ is VP and } SP \text{ is VE then GR output is VP}
\end{align*}
\]

where: FP - First Parameter  /  SP - Second Parameter  /  GR - Group

The INQA was developed from a fuzzy inference that had Groups 01 to 05 as input sets and a series or rules. The antecedent sets (Groups) and the consequent set (INQA) were created by trapezoid (Excellent and Poor sets) and triangular pertinence (all others) functions (Table 3, Figure 3); the INQA classes were the same as for the CETESB’s IQA quality standards (Table 3). For example, it was assumed that the boundary between Good and Excellent had a pertinence of 50% in the Excellent and Good fuzzy sets and so on, showing absence of a rigid boundary between classes.

Fig. 3. Output Membership Function

<table>
<thead>
<tr>
<th>Gr01, 02, 03, 04, 05 and INQA</th>
<th>IQA CETESB Classes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Excellent</td>
<td>65 90 100 100</td>
</tr>
<tr>
<td>Good</td>
<td>44 65 90 51</td>
</tr>
<tr>
<td>Fair</td>
<td>28 44 65 36</td>
</tr>
<tr>
<td>Bad</td>
<td>0 28 44 19</td>
</tr>
<tr>
<td>Poor</td>
<td>0 0 9 28 0</td>
</tr>
</tbody>
</table>

Table 3. Input and output fuzzy sets for inference IN06 and IQA\text{CETESB} classes
The fuzzy inference system used to compute the INQA has 3125 rules. Being impossible to write them all in this paper, some examples are given below:

Rule 01:
If Gr01 is Excellent and Gr02 is Excellent and Gr03 is Excellent and Gr04 is Excellent and Gr05 is Excellent then INQA is Excellent.

Rule 830:
If Gr01 is Excellent and Gr02 is Good and Gr03 is Bad and Gr04 is Excellent and Gr05 is Poor then INQA is Good.

Rule 1214:
If Gr01 is Good and Gr02 is Poor and Gr03 is Bad and Gr04 is Fair and Gr05 is Bad then INQA is Bad.

Rule 2445:
If Gr01 is Bad and Gr02 is Poor and Gr03 is Fair and Gr04 is Poor and Gr05 is Poor then INQA is Poor.

All the computations were processed using the “fuzzy logic toolbox” for MATLAB® (2006).

2.4 Study area

2.4.1 Ribeira do Iguape river – environmental conservation area

The watershed of Ribeira River and the Lagoone-Estuary Complex of Iguape, Cananéia and Paranaguá, called Ribeira Valley, comprises 32 counties and covers area of 28,306 km², with 9 cities and 12,238 km² in Paraná State and 23 cities and 16,068 km² in São Paulo State, Brasil. The economy of Ribeira Vally is based in livestock raising (200,421 hectares), fruticulture (49,942 hectares), silviculture (46,368 hectares), temporary cultures (15,965 hectares) and horticulture (2,773 hectares). Sand and turf extraction from low-lying areas are also significant. About 1% of the state population (396,684 people) live in this river basin, 68% of them in cities. About 56% of the effluents are collected and 49% are treated. It is estimated that approximately 8.8 tons of BOD₅ (remaining pollutant charge) are launched in rivers for disposal within this watershed (CETESB, 2006). The sampling points are given in Table 4 and an illustrative map for this area is shown in Figure 4.

Table 4. Sampling point locations in the Ribeira do Iguape river

2.4.2 Paranapanema river – farming area

Paranapanema River has a total extension of 929 km, with eight dams and barrages along its length. The area under study is about 29,114 km². Soil use is predominantly rural and thus the region is considered a farming area, occupied mainly by pastures (1,781,625 ha), followed by temporary cultures, such as sugar cane, soy and corn (764,476 ha) and silviculture (76,595 ha). Fruticulture occupies 40,917 ha and horticulture, 2,477 ha. The watershed comprises 63 counties, with a total population of 1,155,060, of which 88% is urban (CETESB, 2006). Approximately 95.5% of the effluents produced in this watershed are collected and about 79% of these are treated. It is estimated that approximately 20 tons of
BODs are dumped in reception bodies of this watershed for disposal (CETESB, 2006). The sampling points are given in Table 5 and an illustrative map for this area is shown in Figure 5.

Fig. 4. Map showing Ribeira do Iguape River in a conservation area.

Fig. 5. Map showing Paranapanema River in a farming area.
Table 5. Sampling point locations in Paranapanema River

2.4.3 Pardo river – industrializing area

Pardo River is born in a small spring in Minas Gerais state, crosses the northwest part of São Paulo state and, after running for 240 km with a watershed of 8,993 km², empties in the estuary of Mogi-Guaçu river. The main uses of the soil in this watershed are urban-industrial and farming, with predominance of sugar cane (329,924 ha), followed by pastures (261,999 ha), fruticulture (83,611 ha) and silviculture (46,640 ha). About 3% of the state population live in this UGRHI (1,056,658 people) with 97% of the population in urban areas, scattered over 23 cities. More than 99% of the effluents are collected and 51% are treated. It is estimated that approximately 31 tons of BOD₅ are dumped in reception bodies of this watershed for disposal (CETESB, 2006). The sampling points are given in Table 6 and an illustrative map for this area is shown in Figure 6.

Table 6. Sampling point locations in Pardo River

---

![Fig. 6. Map showing Pardo River in an industrializing area.](www.intechopen.com)
2.4.4 Paraíba do Sul river – industrial area

Paraíba do Sul River has an approximate length of 1,150 km (Jornal da ASEAC, 2001). Its watershed is located in the southwest region of Brazil and covers approximately 55,400 km², including the states of São Paulo (13,500 km²), Rio de Janeiro (21,000 km²) and Minas Gerais (20,900 km²). The watershed comprises 180 counties, with a total population of 5,588,237, 88.8% in urban areas. The river is used predominantly for irrigation (49.73 m³/s), without taking into account the transposition of the Paraíba do Sul (160 m³/s) and Pirai (20 m³/s) rivers to the metropolitan region of Rio de Janeiro. The urban supply amounts to about 16.5 m³/s, while the industrial sector uses 13.6 m³/s, surpassing only the cattle-raising sector, with less than 4 m³/s. The main uses of the soil are urban-industrial and rural, the second with pastures (545,156 ha), temporary cultures (57,709 ha), fruticulture (2,996 ha), horticulture (438) and silviculture (83,667 ha). About 5% of the state population (1,944,638) live in this watershed, with 91% in urban areas, scattered throughout 34 counties. Of the total effluents produced in this watershed, 89% are collected and 33% of these are treated. It is estimated that about 72 tons of BOD are dumped in this river for disposal (CETESB, 2006). The sampling points are given in Table 7 and an illustrative map for this area is shown in Figure 7.

Table 7. Sampling point locations in Paraíba do Sul River

<table>
<thead>
<tr>
<th>Sampling Point</th>
<th>CETESB</th>
<th>Latitude W</th>
<th>Longitude S</th>
<th>Sampling Site Description</th>
<th>County</th>
</tr>
</thead>
<tbody>
<tr>
<td>SP 01</td>
<td>PARB</td>
<td>23.83.30</td>
<td>45.55.52</td>
<td>at the water intake point in São Branca county</td>
<td>São Branca</td>
</tr>
<tr>
<td>SP 02</td>
<td>PARB</td>
<td>23.82.09</td>
<td>45.53.50</td>
<td>under a bridge on the highway SP-77, between Japeri and São Branca</td>
<td>São Branca</td>
</tr>
<tr>
<td>SP 03</td>
<td>PARB</td>
<td>23.81.46</td>
<td>45.56.08</td>
<td>at the Japeri water intake point</td>
<td>Japeri</td>
</tr>
<tr>
<td>SP 04</td>
<td>PARB</td>
<td>23.81.32</td>
<td>45.56.48</td>
<td>under a bridge that crosses the Araruama sub-basin, in São José dos Campos</td>
<td>São José dos Campos</td>
</tr>
<tr>
<td>SP 05</td>
<td>PARB</td>
<td>23.81.68</td>
<td>45.53.54</td>
<td>at the São José dos Campos water intake point</td>
<td>São José dos Campos</td>
</tr>
<tr>
<td>SP 06</td>
<td>PARB</td>
<td>23.81.48</td>
<td>45.52.42</td>
<td>under a bridge on the highway, between Cachoeira and Bragança</td>
<td>Cachoeira</td>
</tr>
<tr>
<td>SP 07</td>
<td>PARB</td>
<td>23.81.40</td>
<td>45.53.10</td>
<td>at the SABESP water intake point for Itatiba city</td>
<td>Itatiba</td>
</tr>
<tr>
<td>SP 08</td>
<td>PARB</td>
<td>23.81.02</td>
<td>45.56.58</td>
<td>under a bridge on the highway, between Pinheiros and Vargem Grande</td>
<td>Pinheiros</td>
</tr>
<tr>
<td>SP 09</td>
<td>PARB</td>
<td>23.81.60</td>
<td>45.56.04</td>
<td>at the water intake point for Anapuva</td>
<td>Anapuva</td>
</tr>
<tr>
<td>SP 10</td>
<td>PARB</td>
<td>23.82.32</td>
<td>45.54.52</td>
<td>under a bridge on the highway BR-499, between Lobos and Piratuba</td>
<td>Lobos</td>
</tr>
<tr>
<td>SP 11</td>
<td>PARB</td>
<td>23.81.10</td>
<td>45.48.26</td>
<td>under the bridge in Quiteria county</td>
<td>Quiteria</td>
</tr>
</tbody>
</table>

Fig. 7. Map showing Paraíba do Sul River in an industrial area.

www.intechopen.com
3. Index results and discussion

The IQA\textsubscript{CETESB} was taken from the Relatórios de Qualidade das Águas Interiores do Estado de São Paulo (CETESB, 2004, 2005, 2006). The IQA\textsubscript{sub} was calculated with a weight factor $k = 0.75$ for good quality water. The IQA\textsubscript{min} was calculated as described by Pesce and Wunderlin (2000) and the IQA\textsubscript{PAL} according to Silva e Jardim (2006), using the recommended technologies. The INQA was computed using the method previously outlined. In this work individual results will not be presented. The results will be graphically presented in the consolidated form of weighted averages. A statistical analysis of the results will then be performed. Factors or influences that lead to an increase or decrease of individual parameters will not be discussed, since this would take us too far afield. A discussion of the subject can be found in Lermontov (2009).

3.1 Ribeira do Iguape river indices – environmental conservation area

The annual averages of the indices for 2004, 2005 and 2006 are shown in Figure 8 for all sampling points. The IQA\textsubscript{CETESB}, IQA\textsubscript{sub} and INQA indices are strongly correlated. In most cases, the IQA\textsubscript{sub} index is the stricter and IQA\textsubscript{min} is the less strict, attributing a better quality to the same water sample.

![Fig. 8. Annual averages of the indices for the Ribeira do Iguape River.](image)

3.2 Paranapanema river indices – farming area

The results for the Parapanema River are shown in Figure 9. The IQA\textsubscript{min}, for 2004 is less strict than the other indices, while the IQA\textsubscript{min} is the stricter. The other the indices are very close for sampling points SP 03, 04 and 05, but diverge somewhat for sampling points SP 01 and 02.

In the case of 2005 data, the INQA stays close to the IQA\textsubscript{CETESB} for all sampling points but the two indices are weakly correlated, specially at sampling point SP 02. The IQA\textsubscript{sub} is again the stricter index and the IQA\textsubscript{min} the less strict. Data for 2006 confirm that the IQA\textsubscript{sub} is not the best indicator for the water quality of this river, since it diverges significantly from the other indices. The INQA is again very close to the IQA\textsubscript{CETESB}, although slightly less strict.

![image](image)
3.3 Pardo river indices – industrializing area
The results for the Pardo River are shown in Figure 10. For 2004, the IQA_{CETESB}, IQA_{sub} and INQA indices are very close. A $k = 0.75$ value for the IQA_{sub} index shows a less strict evaluation, while a $k = 1.00$ for the IQA_{obj} shows a stricter evaluation. The INQA is in general close to the IQA_{CETESB} albeit somewhat less strict for SP 04. The 2005 results show the INQA close to the IQA_{CETESB} for sampling points SP 01 and SP 02 but the indices diverge for SP 03 and SP 04. The IQA_{sub} is again the stricter index. The results for 2006 are similar.

3.4 Paraíba do Sul indices – industrial area
The results for the Paraíba do Sul River are shown in Figure 11. In the case, the IQA_{PAL} is the stricter index, while the IQA_{obj} and the IQA_{min} alternate as the less strict index, depending on the sampling point. The IQA_{CETESB}, IQA_{sub} and INQA are closely related.
4. Statistical results, discussion and conclusions

4.1 Statistical results

The purpose of statistical analysis of the results for each watershed was to validate the use of fuzzy methodology to develop a fuzzy water quality index (INQA). In this process, the results for 2004, 2005 and 2006 were not separately studied, but were grouped in a single data set for each index. The results are shown in Table 8.

Table 8. Statistical Data

The statistical data were computed using the StatSoft Statistica application and will be discussed in section 4.2. Figure 12 show the coefficient of variation of the indices. Table 9 shows the relative differences between the means of the indices and the official index (IQA\textsubscript{CETESB}) and the proposed new index (INQA), calculated using Equation 6:
% variation = (I_1 - I_2) / I_1 x 100 \hspace{1cm} (6)

Where:
I_1 – First index
I_2 – Second index

Fig. 12. Coefficients of variation of the indices.

Table 9. Relative differences between the means of the indices and IQA_{CETESB} and INQA.
The frequency histograms of the indices for the four watersheds are shown in Figure 13 and correspond to a visual representation of the frequency distribution tables. For analysis and interpretation of these graphs, see Lermontov (2009).

Fig. 13. Frequency histograms for the four watersheds.
Figures 14 and 15 show box & whiskers plots for all indices and watersheds. These plots are a convenient way to visualize the main trend and the data scatter and to show, in the same graph, the main results of a sampling.

![Box & Whiskers plots for Ribeira do Iguape, Paranapanema, Pardo, and Paraíba do Sul.](image)

**Ribeira do Iguape**

**Paranapanema**

**Pardo**

**Paraíba do Sul**

Fig. 14. Box & Whiskers plots of the mean, mean ± standard deviation and mean ± 1.96 times standard deviation for the four watersheds.
Ribeira do Iguape     Paranapanema

Pardo             Paraíba do Sul

Fig. 15. Box & Whiskers plots of the median, upper and lower quartile and maximum and minimum value for the four watersheds.

Table 10 shows the correlations between the fuzzy index (INQA) and the other indices. The best correlation, 0.8527 (a strong correlation), between the INQA and the IQA_{CETESB} for the Paranapanema River, is illustrated in Figure 16. The worst correlation, 0.3740, between the INQA and the IQA_{PAL} for the Ribeira do Iguape River, is illustrated in Figure 17.

<table>
<thead>
<tr>
<th></th>
<th>Ribeira do Iguape</th>
<th>Paranapanema</th>
<th>Pardo</th>
<th>Paraíba do Sul</th>
</tr>
</thead>
<tbody>
<tr>
<td>INQA x IQA_{CETESB}</td>
<td>0.79381</td>
<td>0.8527</td>
<td>0.8206</td>
<td>0.7943</td>
</tr>
<tr>
<td>INQA x IQA_{sub}</td>
<td>0.57937</td>
<td>0.7710</td>
<td>0.7107</td>
<td>0.8127</td>
</tr>
<tr>
<td>INQA x IQA_{obj}</td>
<td>0.57937</td>
<td>0.7710</td>
<td>0.7107</td>
<td>0.8742</td>
</tr>
<tr>
<td>INQA x IQA_{min}</td>
<td>0.59937</td>
<td>0.6444</td>
<td>0.6520</td>
<td>0.7483</td>
</tr>
<tr>
<td>INQA x IQA_{PAL}</td>
<td>0.37406</td>
<td>0.3924</td>
<td>0.4025</td>
<td>0.5191</td>
</tr>
</tbody>
</table>

Table 10. Correlations between the INQA and the other indices for the four watersheds.
Fig. 16. Best correlation – \( \text{INQA} \times \text{IQACETESB} \) – \( r = 0.8527 \) – Paranapanema River

Fig. 17. Worst correlation – \( \text{INQA} \times \text{IQAPal} \) – \( r = 0.3740 \) – Ribeira do Iguape River
4.2 Statistical discussion

The statistical data that were collected and presented in this work provide a rich field for discussion and analysis. However, our purpose here was only to validate the use of the fuzzy index (INQA). A simplified statistical analysis was implemented and fulfilled its purpose.

In the case of the Ribeira do Iguape River, we could compute all indices from the available data, except the IQA_{CETESB}, that was taken directly from reports.

In the case of the Paraíba do Sul River, since there was a minimum equal to zero, the geometric and harmonic means could not be computed.

For all watersheds and all indices, the geometric mean was lower than the arithmetic mean and the harmonic mean was lower than the arithmetic mean.

The geometric mean and the harmonic mean of the IQA_{PAL} could not be computed for the Paraíba do Sul River because, in the case, the minimum value was 0.

The coefficients of variation shown in the last column of Table 8 were plotted in Figure 12. In this kind of analysis, the statistical results are presented through a parameter that reflects the scattering of the data points. The worst coefficient of variation was that of the IQA_{PAL} and the best were those of the IQA_{sub} and the IQA_{obj}. When the results for the INQA and the IQA_{CETESB} are compared, one notices that the coefficient of variation of the INQA was smaller than that of the IQA_{CETESB} in three watersheds: Ribeira do Iguape, Paranapanema and Pardo. Only in the industrial area of the Paraíba do Sul River the coefficient of variation of the IQA_{CETESB} was smaller than that of the INQA. This is probably due to the fact that the Paraíba do Sul watershed is more polluted than the others, with low quality water.

The relative differences more relevant to our study, i.e. those between means of the other indices and the IQA_{CETESB} and the INQA means, were computed using Equation 6 and the results are shown in Table 9. In the case of the difference between the IQA_{CETESB} and the INQA, the main focus of our study, all the differences were smaller than 10%. The largest difference, 7.5%, was for the Paraíba do Sul watershed, an industrial area, and the smallest, 0.5%, was for the Paranapanema watershed, a farming area.

Examining the box and whiskers plots of Figures 14 and 15 along with the data from Table 9, one can draw the following conclusions:

- IQA_{obj} and IQA_{min} are the indices that diverge more sharply from the others, especially from IQA_{CETESB}, calculated using a well accepted method;
- INQA yielded satisfactory results when compared to a traditional method such as IQA_{CETESB};
- The results obtained using INQA and IQA_{CETESB} were closest for a farming region and were farthest for an industrial region.

The correlation data are shown in Table 10. The correlation coefficient r, or “Pearson’s r”, as it is also called, is used in this study to measure the degree of correlation between INQA and the other indices for each watershed. Values between 0.7 and 1.0 (positive or negative) indicate a strong correlation between two parameters. Examining the correlation data, one can draw the following conclusions:

- The worst correlation with INQA was that of IQA_{PAL} in all four watersheds. This is probably due to the fact that this indicator is based on only two parameters;
- The best correlation with INQA was that of IQA_{obj} in the industrial region (Paraíba do Sul watershed), but the correlation of IQA_{obj} with INQA was much weaker in the other regions;
- The best global correlation with INQA was that of IQA\textsubscript{CETESB}, a widely accepted index;
- The best individual correlation between INQA and IQA\textsubscript{CETESB} was in the farming region (Paranapanema watershed).

### 4.3 Statistical conclusions

The main conclusions of the statistical analysis are the following:
- There is a strong correlation between the proposed fuzzy index (INQA) and a widely accepted, traditional index (IQA\textsubscript{CETESB});
- The relative differences between the means of INQA and IQA\textsubscript{CETESB} were less than 8% for all four watersheds;
- The box and whiskers plots for the two indices are reasonably similar;
- The other statistical results for the two indices also were reasonably similar;
- The coefficients of variation of the INQA were smaller than those of the IQA\textsubscript{CETESB} for all four watersheds.

### 5. General conclusions

The use of several water quality indices and the development, application and evaluation of a new indexing method to assess river water quality using fuzzy inference is discussed. A new index, called Fuzzy Water Quality Index (INQA) is developed to correct perceived deficiencies in environmental monitoring, water quality classification and management of water resources in cases where the conventional, deterministic methods can be inaccurate or conceptually limited. This methodology differs from other fuzzy water quality indexing methodologies by incorporating the weight factor in qualitative sphere throughout the rules in the inference engine. This is only possible due to a high variety of rules inserted in the inference system. The practical applications of the new index is tested in a realistic case study carried out in Ribeira do Iguape River in São Paulo State, Brazil, showing that the proposed index is reliable and consistent with the traditional qualitative methods.

Most institutional players are not familiar with fuzzy logic concepts, therefore being unaware of the potential of this technique for the transfer of expert knowledge in a qualitative sphere into a formal system of environmental assessment. We think that this approach can and should be used as an alternate tool for the analysis of river water quality and for strategic planning and decision making in the context of integrated environmental management.

For this doctoral study, the same nine parameters used by CETESB State Organ to calculate its IQA were chosen for the methodology validation by statistical comparison. The authors also worked in the development of an index with additional parameters, such as heavy metals, organoleptic metals and toxic compounds, for a more realistic evaluation of the hydric bodies (Lermontov, 2009).

### 6. References


In recent years the topic of environmental management has become very common. In sustainable
development conditions, central and local governments much more often notice the need of acting in ways that
diminish negative impact on environment. Environmental management may take place on many different
levels - starting from global level, e.g. climate changes, through national and regional level (environmental
policy) and ending on micro level. This publication shows many examples of environmental management. The
diversity of presented aspects within environmental management and approaching the subject from the
perspective of various countries contributes greatly to the development of environmental management field of
research.

How to reference
In order to correctly reference this scholarly work, feel free to copy and paste the following:

André Lermontov, Lidia Yokoyama, Mihail Lermontov and Maria Augusta Soares Machado (2011). A Fuzzy
Water Quality Index for Watershed Quality Analysis and Management, Environmental Management in Practice,
http://www.intechopen.com/books/environmental-management-in-practice/a-fuzzy-water-quality-index-for-
watershed-quality-analysis-and-management