Engenharia Agrícola

ISSN: 1809-4430 (on-line)

www.engenhariaagricola.org.br

Scientific Paper
Doi: http://dx.doi.org/10.1590/1809-4430-Eng.Agric.v42n3e20220006/2022

DEVELOPMENT OF A WATER QUALITY INDEX WITH A REDUCED NUMBER OF PARAMETERS

Amanda R. M. de Oliveira[1], Alisson C. Borges[1*], Greicelene J. da Silva[1], Moysés Nascimento[1]

[1*]Corresponding author. Federal University of Viçosa/Viçosa - MG, Brazil.
E-mail: borges@ufv.br | ORCID ID: https://orcid.org/0000-0002-9729-6439

KEYWORDS

Cluster analysis, analytic cost, monitoring in real time.

ABSTRACT

The development of water quality indices (WQIs), duly modified and using a small number of parameters, may be of great economic use. The purpose of this work was to develop a WQI with a reduced number of parameters (WQI_{red}), in relation to what is used in the Brazilian state of Minas Gerais (WQI_{IGAM}), in the basin of the Doce River. Four different scenarios were analysed, with the absence of the following parameters for the calculation of the WQI_{red}: BOD, *E. coli* and phosphate; BOD and *E. coli*; BOD, *E. coli* and total solids (TS); BOD, *E. coli*, phosphate and TS. The redistribution of the weight attributed to the eliminated parameters was done according to the following methods: (i) in weighted form, through optimisation of the correlation between the WQI_{red} and the WQI_{IGAM}; (ii) based on the cluster analysis. The best scenario found was that of the WQI_{red} considering only DO, pH, temperature variation, nitrates and turbidity, with the weights being redistributed based on cluster analysis. This scenario maintained satisfactory performance, with a minimum difference being observed between WQI_{IGAM} and WQI_{red}, in terms of management purpose. The WQI_{red} as proposed makes it possible to establish a monitoring control system operating in real time in the Doce river basin, as all the parameters considered can be obtained with the use of multiparameter probes.

INTRODUCTION

Over the last few years, water quality has been a major cause for concern throughout the world. Among the consequences of the increase in the water footprint as necessary for human activities, we have the adverse effects on the availability of drinking water for different uses (Fraga et al., 2020; Shah et al., 2019). According to the Brazilian National Waters Agency (ANA), over 100 thousand kilometres of sections of Brazilian rivers have their quality at risk due to the organic loads that have been poured into the water bodies (ANA 2020). This means that water quality has become a top priority as a way to ensure health and public safety (Dao et al. 2020).

Regular monitoring of water quality is very important for efficient management of water resources. On carrying out the monitoring, the environmental authorities usually implement methods that supply a generalised appraisal of surface water and group them into categories according to their state of quality (Yotova et al. 2021). This assessment is usually implemented based on water quality indices (WQIs), this being a tool often used to combine and convert the values of many environmental variables into one single number, which is able to reflect the quality of the water in many different classes (Zotou et al. 2020). In this way, water quality can be easily compared between different regions, instead of the confrontation between the numerical values of different environmental parameters (Nayak et al. 2020). Indeed, the application of the index allows discretion with regard to water quality, both by the interested public, as also by competent authorities that draw up public policies (Avigliano & Schenone, 2016).

The NSF index is known internationally, as well as being the most publicised. This index was developed by Brown et al. (1970) using the technique known as Delphi – an opinion research study to extract information from specialists in the area. In Brazil, this index is currently

[1] Federal University of Viçosa/Viçosa - MG, Brazil.

Area Editor: Fernando António Leal Pacheco
Received in: 1-14-2022
Accepted in: 6-10-2022

used by authorities responsible for the management of water resources in the state of Minas Gerais (Minas Gerais Water Management Institute – IGAM) with some changes (IGAM 2020). However, the high number of parameters to be obtained in the process to establish the water quality index generates financial costs, as well as environmental impacts arising from the analytic subproducts generated by the process (Oliveira et al. 2018a). The use of minimum indices simplifies the assessment of water quality for monitoring purposes (Oliveira et al., 2018a; Wu et al., 2018). Based on the use of fewer parameters, the use of minimum indices is even more beneficial, especially in emerging countries, as this reduces the analytic cost of measurement of analytic parameters (Wu et al., 2018; Pak et al., 2021). Results of the minimum WQI as calculated based on the NSF WQI have shown themselves to be highly linearly correlated, which shows that the minimum index approach is powerful when it comes to the quick establishment of the WQI (Wu et al. 2018). Well aware of the importance of the use of minimum indices, especially for emerging countries, in this work the main aim was to develop a water quality index with a reduced number of parameters (WQI_{red}) which is practical, easy to obtain, and equivalent to the one currently practised in the state of Minas Gerais (WQI_{IGAM}).

Considering that the analysis of certain parameters may take a long time and incur significant labour costs, as well as costs with purchase of chemical reagents and laboratory infrastructure, we propose an WQI_{red} that, apart from being easier to obtain – as parameters are obtained directly using multiparameter probes –, also makes it possible to set up a monitoring control system in real time, thus allowing an increased frequency of water quality monitoring in the basin of the Doce River.

MATERIAL AND METHODS

Area of study

The basin of the Doce River lies in Southeastern Brazil, between the states of Minas Gerais and Espírito Santo. This river basin is part of the Southeast Atlantic river basin, and lies between latitudes 17°45' and 21°15' South, and longitudes 39°30' and 43°45' West. The Doce River starts in the state of Minas Gerais, and is then taken to the coastline of the state of Espírito Santo, flowing into the Atlantic Ocean (Figure 1) (ANA 2013). It has a total drainage area of some 86,715 km². Within the state of Minas Gerais, the Doce River basin is divided into six Units for Planning and Management of Water Resources (UPGRHs) (see Figure 1): River Basin Committees of the Piranga River (DO1), Piracicaba River (DO2), Santo Antônio River (DO3), Suaçuí River (DO4), Caratinga River (DO5), and the Manhuaçu River (DO6) (ANA 2013).

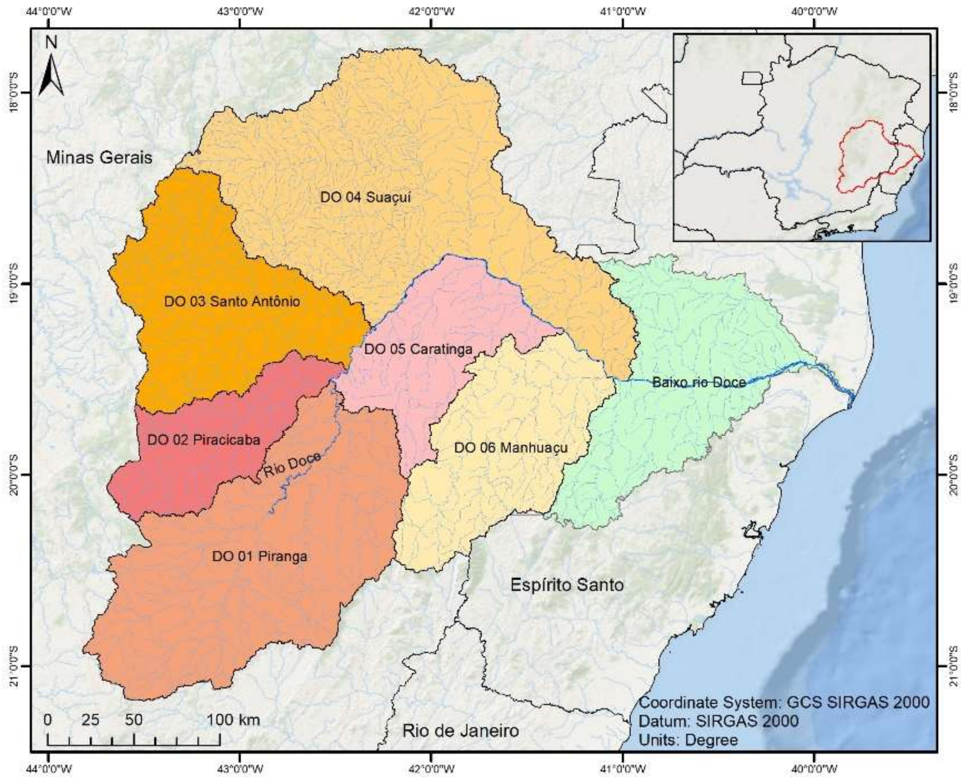

FIGURE 1. Map of the location of the water basin of the Doce River and its subdivision into UPGRHs.

The basin of the Doce River houses a population of some 3.3 million people, distributed in 229 different municipalities, of which 203 are in the state of Minas Gerais and 26 in the neighbouring state of Espírito Santo. This is a river basin with a diverse range of economic activities, such as livestock raising, mining, industrial, service provision, and electrical energy generation. The basins of the Piranga and Piracicaba rivers (DO1 and DO2) have the highest industrial GDP and account for some 48% of the total population of the total basin of the Doce River. In DO2, we have 50% of the total basin of the Doce River, and here we highlight industrial activities and extraction of iron ore (ANA 2013). For the reasons mentioned before, the basin of the Doce River, especially the units on the Piranga and Piracicaba rivers, were chosen as an area of study in the present paper.

Database

For the development of the WQI_{red} that shall be equivalent to WQI_{IGAM}, made use of the database on quality of surface water, as made available at the InfoHidro portal. This information is supplied by the Minas Gerais Water Management Institute (IGAM), which has monitored the quality of surface and underground waters of the state of Minas Gerais since 1997 – known as the 'Minas Gerais Waters Project'. At present, IGAM has sixty-four (64) sampling stations in its qualitative monitoring network in the basin of the Doce River, of which fifteen are at the Piranga River UPGRH and thirteen are at the UPGRH of the Piracicaba River (IGAM 2020). Taken together, these two UPGRHs account for 44% of the stations in the river basin. For assessment of water quality, 56 parameters are taken into consideration, including the nine used for the calculation of the WQI_{IGAM}.

In general, environmental data bring censured and lost values (Ngouna et al. 2020). To avoid any possible problems with statistical analysis, we have proceeded with the preparation of the database. In the pre-treatment of the censured values, the methodologies used were those used in works such as McLaughlin & Flinders (2016) and Oliveira et al. (2018a). According to this methodology, the values found below the minimum limit for detection are replaced by half the minimum limit for detection. In contrast, the values above the maximum as measured by the institution responsible are maintained (Oliveira et al., 2018a; Oliveira et al., 2018b). In the case of lost data, any sample that did not present one of the nine parameters, or did not show any parameters at all, were excluded. Well, in this academic paper we sought to simulate the absence of certain parameters in a controlled manner. In this study, we worked on the data of DO1 and DO2, considering the period between 2009 and 2015.

Preparation of WQI_{red}

Parameters to be adopted in WQI_{red}

After the prior treatment of the data came the preparation of the WQI_{red} equivalent to the WQI_{IGAM}. Initially, the WQI_{IGAM} was calculated for each point as sampled on the treated database. For this reason, we used [eq. (1)] (IGAM 2020).

$$WQI_{IGAM} = \prod_{i=1}^{n} q_i^{w_i} \quad (1)$$

In this equation, WQI_{IGAM} – Water Quality Index, ranging from 0 to 100; q_i – quality of the parameter i, obtained through the average specific quality curve; w_i – weight assigned to the parameter, due to its importance in quality, between 0 and 1; n – number of parameters. IGAM adopts the weights for dissolved oxygen (DO), thermotolerant coliforms / *E. coli*, pH, biochemical oxygen demand (BOD), temperature variation (ΔT), phosphates, nitrates, turbidity, and total solids (TS) equivalent to: 0.17; 0.15; 0.12; 0.10; 0.10; 0.10; 0.10; 0.08 and 0.08, respectively. The parameter of temperature variation is constant, and at a value of 92 (IGAM 2020). The classification of water quality was made based on that adopted by IGAM, meaning that the levels of excellent, good, fair/average, poor and very poor are assigned to point score brackets as follows: $90 < WQI \leq 100$, $70 < WQI \leq 90$, $50 < WQI \leq 70$, $25 < WQI \leq 50$ e $WQI \leq 25$, respectively.

The calculation of WQI_{red} follows the same line of thought as for WQI_{IGAM}, but some of the parameters have been removed from the calculations. We have analysed four different scenarios for the calculation of WQI_{red}, namely: i) absence of the parameters BOD, *E. coli* and phosphates; ii) absence of the parameters BOD and *E. coli*; iii) absence of parameters BOD, *E. coli* and TS; iv) absence of the parameters BOD, *E. coli*, phosphates and TS. In each of these scenarios, there was also the assessment of the presence and absence of the temperature variation parameter, because the quality of such a parameter is always constant.

On selecting the parameters to be part of WQI_{red}, the key factor for the decision was the practicality of obtaining the data. The parameters DO, pH, nitrates, phosphates and turbidity can also be measured with the use of multiparameter water quality probes. Temperature is also easily obtained and measured; however, the quality thereof is always constant. For this reason, there was analysis of inclusion and non-inclusion of this parameter in the calculation of WQI_{red}, in each scenario as here considered.

In the case of TS, it can be easily obtained indirectly through other parameters that are measured with the use of probes. This means that, for the calculation of the WQI_{red}, the TS were obtained through the sum of total solids in suspension (TSS) – estimated by the models as obtained in Oliveira et al (2018b) – and total dissolved solids (TDS) – estimated by the relationships established between TDS and electrical conductivity (EC) as presented by Walton (1989). This author has proposed conversion factors between the measurement of EC and TDS (*k*) which range from 0.50 to 0.75, according to the value of C as measured. The *k* factor of 0.50 is used for low values of electrical conductivity, while it is raised to $k = 0.75$ with the increase in this water quality parameter (Walton 1989).

Distribution of the weights of missing parameters

On leaving out parameters from the calculation of WQI, the weights of the missing parameters need to be redistributed to that the sum of the weights (w_i) is equal to 1, as shown by Wu et al. (2018). In this academic paper, the weights have been reallocated in three different ways: i) weights redistributed in weighted form among the other parameters; ii) weights redistributed between the other parameters through optimisation of the correlation between the WQI_{red} and the WQI_{IGAM}; iii) weights redistributed based on cluster analysis of the parameters used in the calculation of the WQI_{IGAM}. In cases where there is absence of the temperature variation parameter, in all cases the decision made was to carry out the distribution of the weight of this parameter in a weighted fashion, for the other parameters. The redistribution of the weights through optimisation of the correlation between the WQI_{red} and the WQI_{IGAM} was carried out through using the iterative method. There was a search of distribution of the weights for the other parameters, seeking to obtain the highest possible value for the correlation between the WQI_{red} and the WQI_{IGAM}, while respecting the condition that the sum of these weights should equal 1.

One significant development in recent years has been the introduction of multivariate statistical analysis, specifically cluster analysis, to identify the parameters that are important in the assessment of water quality (Fraga et al., 2020; Rahman et al., 2020; Pak et al., 2021). For the execution of cluster analysis, there was the standardisation of the data with a view to equivalence of the different parameters for water quality, as recommended by Kassambara (2015). In the standardisation process, for each parameter of water quality and each element within the sample, the values observed were subtracted from the mean and then divided by the standard deviation. Later, the Euclidean distance was used to measure the dissimilarity between the parameters of water quality. Ward's algorithm was used to group the parameters.

Classification bands for WQI_{red}

The classification of water quality based on the WQI_{red} was carried out using the value bands as set for the WQI. Two alternative classifications were also considered: i) BAND I – using the correlation between the WQI_{red} and the WQI_{IGAM} – this means that models were obtained, correlating the WQI_{IGAM} with each one of the best scenarios of the WQI_{red}, after which the values of the WQI for the bands in the models were applied, so as to obtain new bands; ii) BAND II – there was a verification of points of the WQI_{IGAM} found in each WQI class, and later these percentages were applied to the points of the WQI_{red} in a hierarchical way, meaning that the values that limited one class and another started to be defined as a new band. The values of the reduced indices were appraised using Pearson's 'r' correlation coefficient. There was also a verification of the equivalence of the reduced indices with the WQI_{IGAM} considering the percentage of points with compatibility and incompatibility of classes.

RESULTS AND DISCUSSION

Quality of the waters of the Doce river basin

A preliminary analysis of water quality in the basin of the Doce River was carried out, considering DO1 and DO2. Figure 2 shows the percentage of points within each class of WQI_{IGAM}, considering the complete data series and considering each year monitored. We see that, between the years of 2009 and 2015, the water quality was reasonable – 63% of the points showed a 'Fair' classification, while 29% were "Good" and 8% were "Poor". The analysis for each year shows that there is no homogeneous quality profile for the river basin. The quality of the water was hampered, in most cases, due to the parameters of turbidity and *E. coli*, due to the launching of treated effluents or sanitary sewage poured *in natura* in waterbodies. This is because the quality of these parameters, obtained through the specific medium quality curve, on average, was respectively 20 and 55. IGAM (2020) says that WQI is, above all, sensitive to contamination by sewage, as *E. coli* is an important indicator of faecal contamination (Grieco et al. 2017), while turbidity is also linked to bacterial contamination (Kataržytė et al. 2018).

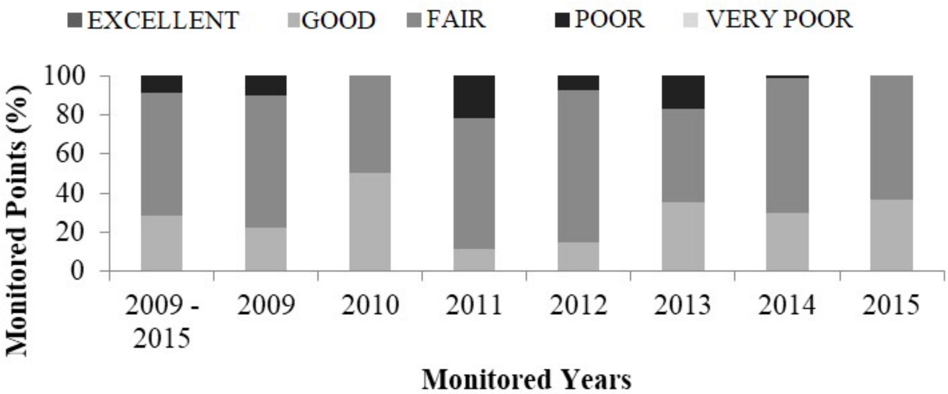

FIGURE 2. Water quality in the Doce River basin – WQI_{IGAM}.

Water quality index with a reduced number of parameters – WQI_{red}

Redistribution of weights in a weighted fashion

The scenarios were simulated by suppressing the parameters as previously listed, and redistributing the weights in a weighted fashion among the other parameters, a procedure also followed by IGAM (2020). Table 1 shows the Pearson correlation coefficient for the connection between the WQI_{red} and the WQI_{IGAM}, as well as the percentage of points that show compatibility and incompatibility of classes with the WQI_{IGAM}.

TABLE 1. Pearson's Correlation Coefficients for the WQI_{red} and the WQI_{IGAM} and percentage of points with compatibility and incompatibility of classes – WQI_{red} calculated considering the redistribution of the weights, in a weighted fashion for the other parameters.

Missing Parameters	r	Classes that were maintained (%)	Over-estimated (%)	Under-estimated (%)
BOD, *E. coli* and phosphates	0.73*	22	78	0
BOD, *E. coli*, phosphates and ΔT	0.73*	27	73	0
BOD and *E. coli*	0.75*	21	79	0
BOD, *E. coli* and ΔT	0.75*	23	77	0
BOD, *E. coli* and total solids	0.75*	20	80	0
BOD, *E. coli*, total solids and ΔT	0.75*	22	78	0
BOD, *E. coli*, phosphates and total solids	0.73*	21	79	0
BOD, *E. coli*, phosphates, total solids and ΔT	0.73*	24	76	0

*: Significant correlation at a probability level of 0.001 by the t-test.

For Momber et al. (2017), $0 < | r | < 0.19$ indicates a very weak correlation, while $0.20 < | r | < 0.39$ suggests a weak correlation; $0.40 < | r | < 0.59$, a moderate correlation; $0.60 < | r | < 0.79$, a strong correlation; and $0.80 < | r | < 1.0$, a very strong correlation. Therefore, we see that each of the eight scenarios involving the WQI_{red} and the WQI_{IGAM} are strongly correlated, this correlation being both positive and statistically significant. Considering only the strong correlation, the application of WQI_{red}, calculated based on the redistribution of weights in a weighted manner, just as an indicator of the profile regarding the quality of the waters, would be of great importance for the management of water resources. Indeed, should there be the observation of any non-conformity in the water quality, then analyses with greater precision could be carried out so that, then, management plans could be drawn up. However, we see that, even though the WQI_{red} has a strong correlation with the WQI_{IGAM}, the percentage of points that stayed in the class was low. The WQI_{red} overestimated the quality of the water at most of the points. Therefore, for the finalisation of the management, the WQI_{red} should be improved.

The methodology of calculation of the IGAM considers the premise that, in the absence of parameters *E. coli* or DO, one proceeds with the calculation of the WQI for a certain sampling point. In relation to the concentration of *E. coli*, this is probably due to the fact that this is a parameter of high concentration in the region of the study (Serrano et al. 2020). The non-inclusion of *E. coli* in the reduced index could have jeopardised the good equivalence between the index as proposed and the index as currently practised, as the quality of the water in the Doce River basin was in many cases harmed, mostly due to the presence of *E. coli*, for having, on average, the lowest quantity (q = 20). However, the option made was not to include this parameter, as its analysis takes a lot of time and incurs significant costs with labour, acquisition of reagents, and laboratory infrastructure.

Srivastava & Kumar (2013) proceeded to calculate the WQI using the additive form of the original index, as proposed by Brown et al. (1970). On eliminating the BOD parameter and redistributing the weight among the other parameters in a weighted manner, these authors could observe a small difference in the value of the WQI; however, they were able to confirm the conformity of classes. It is also worth mentioning that Srivastava & Kumar (2013) showed results of suppression of only one parameter in one specific point. In this academic paper, we simulated the suppression of between three and five parameters, which could have jeopardised the result, but was necessary in order to reach the target as proposed for the present work.

Redistribution of weights through optimisation of correlation

Seeking to improve the equivalence between WQI_{red} and WQI_{IGAM}, we proceeded with the analysis of the same scenarios for suppression of parameters, as previously listed. In this case, the weights were redistributed among the other parameters through the optimisation of the correlation between the WQI_{red} and the WQI_{IGAM}. The new results showed that the reduced indices also have strong positive correlations with the WQI_{IGAM} – a small increase in the Pearson correlation coefficient, being between 0.74 and 0.76, with all correlations significant. It was also confirmed that there was a significant reduction of percentage of points, in compliance of classes with the WQI_{IGAM} – the percentage of points where the classes were maintained ranged from 12 % to 13 %.

Redistribution of weights by cluster analysis

According to the cluster, the weights of the missing parameters were distributed to those parameters 'closer by'. In Figure 3, we can see the dendrogram of quality parameters, for the water used in the calculation of the WQI. The temperature variation parameter is not present, as it is not measured, with the adoption of a constant value.

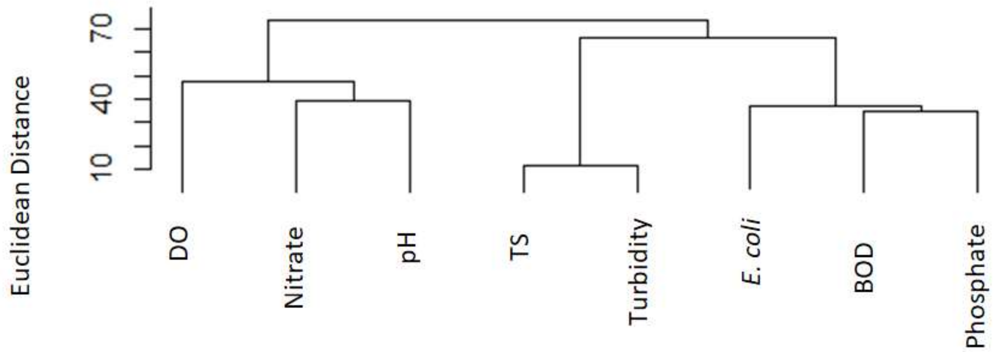

FIGURE 3. Dendrogram of the parameters for water quality.

The results shown in Figure 3 allow the grouping of the parameters in two distinct groups, as follows: i) DO, nitrates, and pH; ii) TS, turbidity, *E. coli*, BOD, and phosphates. In the second group, with turbidity being the only parameter not to be suppressed in any of the four simulated scenarios, it was decided to assign all the weights of the missing parameters to this parameter. In Table 2, we can see the correlation matrix for the weighted value of the quality (q_i^{wi}) of the parameters for water quality. Based on the results, we see that the distribution of the weights of the missing parameters for turbidity was the best option, as the value of q_i^{wi} of the parameters eliminated in the calculation of the reduced indices showed a significantly better correlation with the q_i^{wi} of turbidity. In Table 3, we present the Pearson correlation coefficient for the WQI_{red} calculated considering the redistribution of weights based on cluster analysis and the WQI_{IGAM}, as well as the percentage of points that presented compatibility and incompatibility of classes with the WQI_{IGAM}.

TABLE 2. Pearson Correlation between the values of weighted qualities (q_i^{wi}) of the parameters of water quality in the basin of the Doce River.

Quality Parameters	DO	*E. coli*	pH	BOD	Nitrates	Phosphates	ΔT	Turbidity
E. coli	0.10*							
pH	0.10*	0.03						
BOD	0.08*	0.10*	-0,.01					
Nitrates	-0.02	-0.01	-0.26*	0.08				
Phosphates	-0.04	0.25*	-0.07	**0.17***	0.19*			
ΔT	0.00	0.00	0.00	0.00	0.00	0.00		
Turbidity	-0.09*	**0.34***	0.11*	0.10*	-0.15*	**0.52***	0.00	
TS	-0.03	0.28*	0.05	0.15*	-0.03	0.48*	0.00	**0.65***

Note: In **bold**, we present the best correlations for the missing parameters.
*: Correlation is significant at a probability level of 0.05 according to the t-test.

TABLE 3. Pearson correlation coefficients for WQI_{red} and WQI_{IGAM} and percentage of points with compatibility and incompatibility of classes – the WQI_{red} calculated considering the redistribution of the weights, based on cluster analysis.

Missing Parameters	r	Classes that were maintained (%)	Over-estimated (%)	Under-estimated (%)
BOD, *E. coli* and phosphates	0.71*	54	37	9
BOD, *E. coli*, phosphates and ΔT	0.71*	50	37	13
BOD and *E. coli*	0.73*	52	43	5
BOD, *E. coli* and ΔT	0.72*	55	39	6
BOD, *E. coli* and total solids	0.72*	55	38	7
BOD, *E. coli*, total solids and ΔT	0.72*	48	39	13
BOD, *E. coli*, phosphates and total solids	0.70*	50	36	14
BOD, *E. coli*, phosphates, total solids and ΔT	0.70*	48	34	18

*: Correlation is significant at the probability level of 0.001 by the t-test.

Based on analysis of Table 3, we see that every one of the eight scenarios of WQI_{red} and WQI_{IGAM} show strong correlation, this correlation being positive and statistically significant. It is provisionally said that the best scenario corresponds to the calculation of the WQI in the absence of the parameters BOD, *E. coli* and TS, as this scenario provides the greatest percentage of points in compliance for classes and has one of the highest Pearson correlation coefficients. This means a better equivalence with the WQI_{IGAM}. However, one obstacle for the application of this type of calculation, which includes parameters DO, pH, temperature variation, phosphates, nitrates, and turbidity, would be the increase in the price of monitoring due to the measurement of phosphate content, which is carried out using costly probes.

This means that, analysing the data shown in Table 3, we see that one good alternative is the calculation of the WQI in the absence of the parameters: BOD, *E. coli*, phosphates and TS, as the difference, both in terms of correlation coefficient as also in terms of percentage of points in compliance with classes, is very small compared with the method mentioned previously. This factor is an advantage for emerging countries that have limited financial resources. We also mention that, in the consideration of the parameter of temperature variation, even though this is a parameter that shows little variation, the same has little influence on the practicality of a reduced index. This was therefore considered to be able to get better correlations between the WQI_{red} and the WQI_{IGAM} and better percentages of class conformities (Table 3).

Classification bands for the WQI_{red}

In Figure 4a, we see the classification of water quality through the use of the WQI_{red} based on the bands adopted by the IGAM. Through an analysis of the classification we can see that, despite the two scenarios as previously described having a better equivalence with the WQI_{IGAM}, they still show a significant overestimate regarding the water quality. We see that, in general, the water quality of the basin of the Doce River has been overestimated. For the two scenarios of the WQI_{red}, we have seen an increase in points with a 'Good' classification from 29% (WQI_{IGAM}) to more than 50% (WQI_{red}) and a reduction in those receiving a classification of 'Fair' from 63% to less than 35%. In addition, over 5% of the points showed a classification of 'Excellent'.

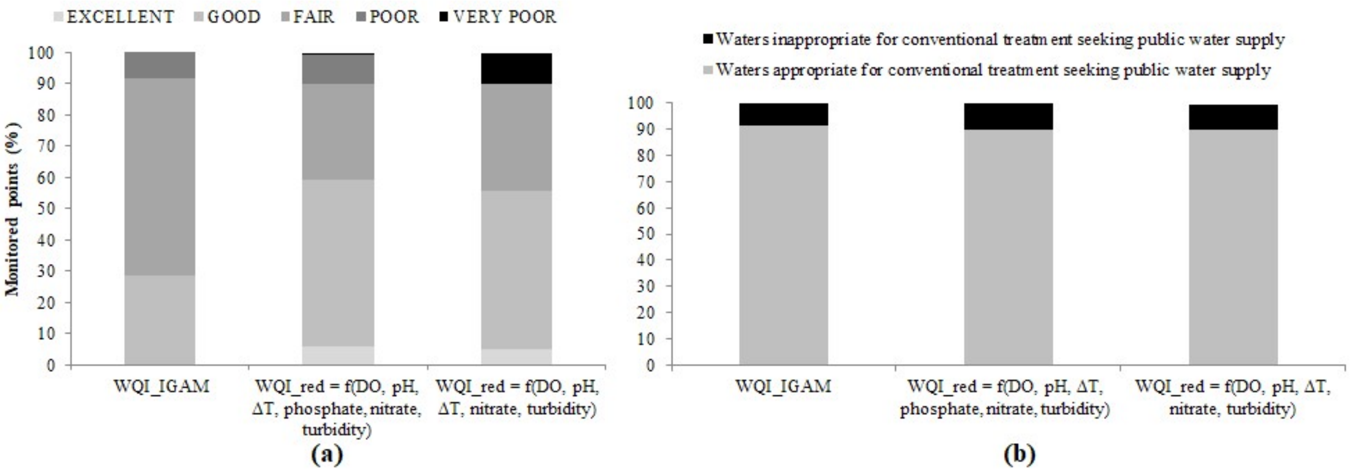

FIGURE 4. Water quality in the basin of the Doce River between 2009 and 2015, shown in three classes, namely WQI_{IGAM} and the two best scenarios for WQI_{red} (a) and the meaning of water quality for management of water resources in the basin of the Doce river, between 2009 and 2015, based on calculation by the WQI_{IGAM} and by the two best scenarios for WQI_{red} (b).

This overestimate of water quality was also reported by Naveedullah et al. (2016). The authors have also proposed a minimum index made up of five parameters: DO, EC, turbidity, temperature, and pH, and have checked that the index also showed an overestimation of water quality, backing up the results obtained in this academic work. However, Naveedullah et al. (2016) found a better equivalence between the reference index and the minimum index, that can be ascribed to the difference in the reference WQI. In this work, we used the WQI as currently practiced by IGAM, which had its origin in the WQI as proposed by Brown et al. (1970). The authors as previously mentioned used the Bascarán index as their reference. The way in which each one is calculated, and the parameters selected, may have favoured the obtaining of better equivalence. In addition, in this work, different from the work discussed, there was the analysis of a long data period (7 years), and a hydrographic basin with highly diversified activities.

According to IGAM (2020), the classes of the WQI_{IGAM} according to the bands of WQI values, have a meaning for the management of water resources, with the classes 'Excellent', 'Good' and 'Fair' meaning that the waters are appropriate for conventional treatment, seeking use for public water supply. The classes known as 'Poor' and 'Very Poor' suggest water not suitable for conventional treatment, seeking public water supply, meaning that more advanced types of treatment are then necessary (IGAM, 2020). In Figure 4b, we see the classification of water quality by the WQI_{red}, following this prerogative.

Based on the analysis of Figure 4b, with regard to the use of the WQI for management purposes, we see a minimum difference between WQI_{IGAM} and the two scenarios of the WQI_{red}, with the quality of the water having been underestimated in less than 2% of the sample points. Therefore, we can say that the use of WQI_{red} gives feasible results, for analysis of the water quality profile.

However, to obtain more reliable results using the WQI_{red}, their values should be well correlated with the index currently practised. This means that it shall be necessary to maintain regular monitoring of the routine with the use of the quarterly WQI_{IGAM}, as the IGAM already does, merely to guarantee the validity of the results obtained through prior monitoring of the WQI_{red}.

With a view to improvement of the equivalence between the reduced index as here developed and the WQI_{IGAM}, an attempt was made to classify water quality through use of the new bands. Table 4 shows the new bands for the two best scenarios of WQI_{red}. In Table 5, we see the percentage of points that showed conformity and non-conformity of classes with the WQI_{IGAM}, for each of the two best scenarios, in each of the two bands. Analysing Table 5, we observed an increase in the percentage of points showing conformity with the class (10% to 13%), with the new proposed band using the percentage of points within each class (Band II) showing the best results. However, we advise that there be an adjustment of the new bands occasionally, as the correlation between the WQI_{IGAM} and the WQI_{red} may change.

TABLE 4. New classification bands for water quality with the WQI_{red} = f(DO, pH, ΔT, phosphate, nitrate, turbidity) and WQI_{red} = f(DO, pH, ΔT, nitrate, turbidity).

WQI_{red} = f(DO, pH, ΔT, phosphates, nitrates, turbidity)			WQI_{red} = f(DO, pH, ΔT, nitrates, turbidity).	
Level	**BAND I**	**BAND II**	**BAND I**	**BAND II**
Excellent	95* < WQI ≤ 100*	95 < WQI ≤ 100	97* < WQI ≤ 100*	97 < WQI ≤ 100
Good	77 < WQI ≤ 95*	83 < WQI ≤ 95	75 < WQI ≤ 97*	82 < WQI ≤ 97
Fair	50 < WQI ≤ 77	27 < WQI ≤ 83	46 < WQI ≤ 75	25 < WQI ≤ 82
Poor	17 < WQI ≤ 50	20 < WQI ≤ 27	10 < WQI ≤ 46	15 < WQI ≤ 25
Very Poor	WQI ≤ 17	WQI ≤ 20	WQI ≤ 10	WQI ≤ 15

*Values established manually so that 100 is not exceeded.

TABLE 5. Conformity and non-conformity based on WQI_{red} = f(DO, pH, ΔT, phosphates, nitrates, turbidity) for Bands I and II; and based on WQI_{red} = f(DO, pH, ΔT, nitrates, turbidity) for Bands I and II.

	WQI_{red} = f(DO, pH, ΔT, phosphates, nitrates, turbidity)		WQI_{red} = f(DO, pH, ΔT, nitrates, turbidity)	
	BAND I	**BAND II**	**BAND I**	**BAND II**
Class that was maintained	65	67	65	68
Overestimated	27	18	26	15
Underestimated	8	15	9	17

CONCLUSIONS

The redistribution of weights based on cluster analysis provided a better equivalence between the WQI_{red} and the WQI_{IGAM}, being the best option for the preparation of the WQI_{red}. The best scenario is the calculation of the WQI_{red} considering only the parameters DO, pH, temperature variation, phosphates, nitrates, and turbidity. As an even more economic option, the calculation of the WQI_{red} can be made only considering the parameters DO, pH, temperature variation, nitrates, and turbidity. The use of the WQI_{red} produced good results for the analysis of the water quality profile of the Doce River, even though there was some overestimation when compared to the value obtained by the WQI_{IGAM}. However, to get more reliable results using the WQI_{red}, its values should be well correlated with the index as currently practised, with the need to maintain a regular monitoring of routine. In addition, the WQI_{red} enables the establishment of a monitoring control system in real time, as all the parameters considered can be obtained directly, using multiparameter probes, according to the two best scenarios.

ACKNOWLEDGMENTS

This work was funded partially by CNPq and Coordination for the Improvement of Higher Education Personnel (CAPES Finance Code 001).

REFERENCES

ANA - Agência Nacional de Águas e Saneamento Básico (2020) Atlas Esgotos revela mais de 110 mil km de rios com comprometimento da qualidade da água por carga orgânica. Available: https://www.ana.gov.br/noticias/atlas-esgotos-revela-mais-de-110-mil-km-de-rios-com-comprometimento-da-qualidade-da-agua-por-carga-organica#:~:text=De acordo com o Atlas, público devido à poluição hídrica. Accessed Dez 4, 2021.

ANA - Agência Nacional de Águas e Saneamento Básico (2013) Plano integrado de recursos hídricos da bacia hídrico do Rio Doce: relatório executivo. Available: http://www.cbhdoce.org.br/wp-content/uploads/2014/10/PIRH_Doce_Volume_I.pdf. Accessed Dez 4, 2021.

Avigliano E, Schenone N (2016) Water quality in Atlantic rainforest mountain rivers (South America): quality indices assessment, nutrients distribution, and consumption effect. Environmental Science and Pollution Research 23:15063–15075. DOI: https://doi.org/10.1007/s11356-016-6646-9

Brown RM, McClelland NI, Deininger RA, Tozer RG (1970) A water quality index - Do we dare? Water & Sewage Works 117(10):339-343.

Dao V, Urban W, Hazra SB (2020) Introducing the modification of the Canadian Water Quality Index. Groundwater for Sustainable Development 11:100457.

Fraga M de S, Reis GB, Silva DD da, Guedes HAS, Elesbon AAA (2020) Use of multivariate statistical methods to analyse the monitoring of surface water quality in the Doce River basin, Minas Gerais, Brazil. Environmental Science and Pollution Research 27:35303–35318.

Grieco AA, Fregonesi BM, Tonani KA de A, Silva TV, Celere BS, Trevilato TMB, Segura-Muñoz SI, Alves RI da S (2017) A tecnologia de remoção de fósforo: Gerenciamento do elemento em resíduos industriais. Revista Ambiente e Água 12:1–17.

IGAM (2020) Índice de Qualidade Das Águas. Available: http://portalinfohidro.igam.mg.gov.br/sem-categoria/319-indice-de-qualidade-das-aguas-WQI#:~:text= Accessed Dez 4, 2021.

Kassambara A (2015) Multivariate Analysis 1: Practical Guide To Cluster Analysis in R. 1ed. Ed. STHDA, 187p.

Kataržytė M, Mėžinė J, Vaičiūtė D, Liaugaudaitė S, Mukauskaitė K, Umgiesser G, Schernewski G (2018) Faecal contamination in shallow temperate estuarine lagoon: Source of the pollution and environmental factors. Marine Pollution Bulletin 133:762–772.

McLaughlin DB, Flinders CA (2016) Quantifying variability in four US streams using a long-term data set: Patterns in Water Quality Endpoints. Environmental Management 57:368–388.

Momber AW, Buchbach S, Plagemann P, Marquardt T (2017) Edge coverage of organic coatings and corrosion protection over edges under simulated ballast water tank conditions. Progress in Organic Coatings 108:90–92.

Nayak JG, Patil LG, Patki VK (2020) Development of water quality index for Godavari River (India) based on fuzzy inference system. Groundwater for Sustainable Development 10:100350.

Naveedullah N, Hashmi MZ, Yu C, Shen C, Muhammad N, Shen H, Chen Y (2016) Water quality characterization of the Siling reservoir (Zhejiang, China) using water quality index. Clean – Soil, Air, Water 44(5):553–562

Ngouna RH, Ratolojanahary R, Medjaher K, Dauriac F, Sebilo M, Junca-Bourié J (2020) A data-driven method for detecting and diagnosing causes of water quality contamination in a dataset with a high rate of missing values. Engineering Applications of Artificial Intelligence 95:103822.

Oliveira ARM, Borges AC, Matos AT, Nascimento M (2018a) Viability of the use of minimum water quality indices: A comparison of methods. Engenharia Agrícola 38:616–623.

Oliveira ARM, Borges AC, Matos AT, Nascimento M (2018b) Estimation on the concentration of suspended solids from turbidity in the water of two sub-basins in the Doce River basin. Engenharia Agrícola 38:751–759

Pak HY, Chuah CJ, Tan ML, Yong EL, Snyder SA (2021) A framework for assessing the adequacy of Water Quality Index – Quantifying parameter sensitivity and uncertainties in missing values distribution. Science of the Total Environment 751:141982.

Rahman MM, Howladar MF, Hossain MA, Shahidul Huqe Muzemder ATM, Al Numanbakth MA (2020) Impact assessment of anthropogenic activities on water environment of Tillai River and its surroundings, Barapukuria Thermal Power Plant, Dinajpur, Bangladesh. Groundwater for Sustainable Development 10:100310.

Serrano L de O, Borges AC, Pruski FF, de Melo MC (2020) A new approach to use load duration curves to evaluate water quality: A study in the Doce River Basin, Brazil. Water (Switzerland) 12:1–21.

Shah B, Kansara B, Shankar J, Soni M, Bhimjiyani P, Bhanushali T, Shah M, Sircar A (2019) Reckoning of water quality for irrigation and drinking purposes in the konkan geothermal provinces, Maharashtra, India. Groundwater for Sustainable Development 9:100247.

Srivastava G, Kumar P (2013) Water Quality Index With Missing Parameters. International Journal of Research in Engineering and Technology 02(04):609–614.

Walton NRG (1989) Electrical conductivity and total dissolved solids - What is their precise relationship? Desalination 72:275–292.

Wu Z, Wang X, Chen Y, Cai Y, Deng J (2018) Assessing river water quality using water quality index in Lake Taihu Basin, China. Science of the Total Environment 612:914-922.

Yotova G, Varbanov M, Tcherkezova E, Tsakovski S (2021) Water quality assessment of a river catchment by the composite water quality index and self-organising maps. Ecological Indicators 120(1):106872.

Zotou I, Tsihrintzis VA, Gikas GD (2020) Water quality evaluation of a lacustrine water body in the Mediterranean based on different water quality index (WQI) methodologies. Journal of Environmental Science and Health, Part A 55(5):537–548.