



Hydrogeochemical Evaluation of Water Quality Suitable for Human Consumption and Comparative Interpretation for Water Quality Index Studies

D. Sierra-Porta¹ 

Received: 15 July 2019 / Accepted: 10 March 2020 / Published online: 21 May 2020
© Springer Nature Switzerland AG 2020

Abstract

To assess the general conditions of water quality for human consumption, specific parameters are estimated to determine its viability in terms of multiple variables, which together measure water quality through the calculation of a single water quality index. In general, many variables related to mineral and bacteriological contents of water are used to calculate these indices. This study aims to assess the feasibility and importance of a quality index for drinking water commonly used in Latin America compared to other international approaches. For this purpose, the evaluation of this index is done using two different methodologies in the same study area. Fourteen water quality parameters have been selected to calculate the index and determine the consumption conditions of 87 samples collected throughout the territory. Although it is true that there are many ways to evaluate the quality of water for human consumption, it is also necessary to assess the suitability of each methodology used for uniformity of the criteria. The results confirm that a locally used index overestimates the number of water samples evaluated as of good quality for consumption while underestimating the number of samples evaluated as not suitable for consumption. These results are beneficial for water quality management and could be used in the evaluation of the methodologies used to calculate the water quality index. In addition, it is suggested that the weights of the environmental parameters should be reconsidered in water quality assessments when the two methods considered are used.

Keywords Water quality index · Water resources · Drinking water · Colombian northwest | Index assessment and comparison

✉ D. Sierra-Porta
d.sierrap@uniandes.edu.co

¹ Departamento de Física, Universidad de los Andes, Cra. 1 No. 18A-10, Edificio IP, Bogotá, Colombia

1 Introduction

The increase in the demand of water entails serious pollution problems, both chemical and biological, whose correction requires costly control systems. In this regard, it is estimated that diarrheal diseases of the total daily global disease burden and cause 1.5 million deaths each year (https://www.who.int/water_sanitation_health/diseases-risks/en/). According to estimates, 58% of this disease burden - that is 842,000 deaths per year - is due to the absence of safe water and poor sanitation and hygiene, and includes 361,000 deaths of children under five, most of them in low-income countries (<https://www.who.int/news-room/fact-sheets/detail/drinking-water>). According to this information, it is important to establish mechanisms and actions aiming at conserving and protecting the quality of continental and underground waters from the effects of pollution sources and climate change (Whitehead et al. 2009; Vörosmary et al. 2000; Murdoch et al. 2000; Delpla et al. 2009; Wilby et al. 2006) to establish a balance of the aquatic ecosystem.

The necessary process of monitoring the quality of water in natural bodies is carried out in the countries through the efforts of national environmental authorities, which allows the observation, evaluation and generation of reports so that decision-makers take the necessary actions. Thus, public institutions have been monitoring the water quality of natural resources for various purposes through several criteria and methodologies established in water quality monitoring protocols.

Actually, there are several monitoring protocols whose content is similar in the technical aspects, which establish the appropriate procedures with different criteria for taking water samples, either from industrial discharges (mining, cement industries, brewing, paper, etc.), domestic discharges or natural water bodies (continental and marine) (Kanu and Achi 2011; Tüfekci et al. 2007; Bengraime and Marhaba 2003). They also present the criteria for the selection of parameters, monitoring points, frequency of monitoring, analysis methodology, flow measurement and procedures for quality assurance.

The water quality index (WQI) is a tool that allows to classify the water quality of a surface or underground body in a given time. In general, the WQI incorporates data from multiple physical, chemical and biological parameters into a mathematical equation, through which the state of a body of water is evaluated (Cude 2001; Pesce and Wunderlin 2000; Poonam et al. 2013; Fernández et al. 2004; Ocampo-Duque et al. 2013). Through the WQI, a general analysis of water quality can be carried out at different levels, and the vulnerability of the body can be determined in the face of potential threats. This tool emerges as an alternative for the evaluation of water bodies, allowing the processes of formulating public policies and monitoring impacts to be more effective.

Historically, different organizations from different countries involved in the control of aquatic resources, and especially in the control and monitoring of water for drinking purposes, have used different methodologies for the calculation of indices for the quantification of water quality (Horton 1965; Brown et al. 1970; S'tambuk-Giljanovi'c N 1999; Tyagi et al. 2013). Many modifications have been considered to the WQI concept through various scientists and experts. However, a general water quality index (WQI) is based on the most common factors, which are described in the following five steps:

1. Establish and select a number of factors or parameters that have a strong correlation with water quality, i.e., physical parameters, chemical parameters, composition and biological factors that imply contamination.

2. Apply transformations to standardize each of the previous factors on a common scale of measurement.
3. Decide on the relative weight of each of the factors in the incidence of the quality of the resource.
4. Establish an algebraic function for the calculation of a single factor that determines the quality of the resource.
5. Establish a qualitative scale based on the values of the calculated index.

Despite this, the water quality index can often be very subjectively defined. This is because sometimes it can be generalized and assigned greater or lesser weights to different variables depending on particular choices; or it could be not indicative of the quality of the total dynamics of the system but be suitable for the measurement of some kind of water use; or the index could be limited in space and time which could give wrong readings at specific points or times. If two water quality indices that use the same risk criteria for each of the measured variables, even if they use different logical-mathematical expressions, offer different scenarios, then it is clear that some of these indices may not reflect the dynamics of the analyzed ecosystem. This is a big problem for communication of the results of water quality estimates. Regardless of the methods used and knowing that there is no universal index of water quality, many methodologies could converge in common scenarios. One purpose of this article is to fairly assess how representative are those two types of water quality indices to evaluate their use for human consumption. Finally, this article reviews the relevant methodologies for calculating water quality indices as an important input for the actions that could be taken by the agencies and institutions responsible for decision-making while discussing the suitability of water for human consumption, based on compliance with drinking water standards.

2 Study Area

This study is carried out throughout the department of Santander, located in the northwestern part of Colombia. Santander is one of the thirty-two departments of the country. Its capital is Bucaramanga. It is located in the Andean region and borders the Cesar and Norte de Santander departments on the north, Boyacá on the southeast, Antioquia on the west, and Bolivar on the northwest. The population was 2,060,000 inhabitants in 2015, and it is the sixth department by population size. The area, which comprises approximately 30,537 km², is delimited within the latitudes 5° 41' 0.8988" N to 8° 8' 38.292" N and longitudes 74° 32' 24.0756" W to 72° 28' 10.83" W (Fig. 1).

The climate of Santander is affected by the diversity of altitudes, its land is distributed in warm, temperate and moorland bioclimatic thermal floors. In the lower Magdalena valley, average temperatures are in the order of 29 °C and there is abundant rainfall, registering up to 3800 mm per year; in the flank of the mountain range, the temperature decreases and the annual average precipitation is between 1500 and 2000 mm; with the exception of the south, especially in the Chicamocha Canyon where precipitation is less than 500 mm, and temperature is high, reaching values up to 32 °C; the moorland area records temperatures below 7 °C and low precipitation.

The hydrographic network of Santander (Guzmán 2016; Fernández-Méndez et al. 2013) is composed of numerous rivers, streams and minor streams, including the Magdalena, Cararé, Lebrija, Opón, Sogamoso rivers (formed by the confluence of the Chicamocha and the Suarez

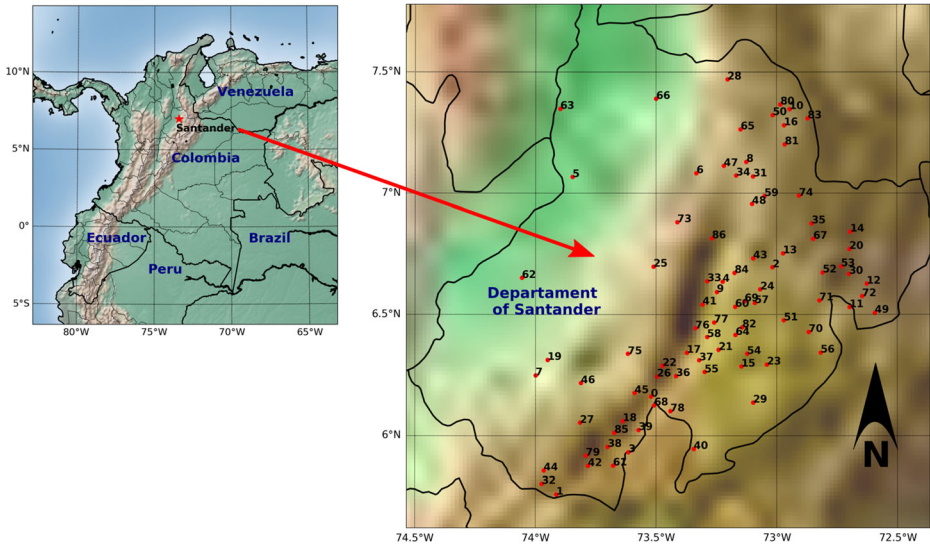


Fig. 1 **a** Location of the Department of Santander in Colombia. **b** Distribution of the 87 municipalities of the Santander Department in which samples have been collected for the study

River), Cachira, Chucurí, Ermitaño, Fonce, Guaca, Guayabito, Horta, La Colorada, Nevado, Onzaga, Paturia, San Juan and Servita. There are also several marshes located in nearby areas of the Magdalena River; among them, the most notable are Colorada, Doncella, El Llanito, Opon, Paredes, Rabon, Redonda, San Silvestre and Yariguies.

3 Water Quality Index Calculation

The National Institute of Health (INS: Instituto Nacional de Salud), in compliance with Decree 1575 of 2007 and its regulatory resolutions, which establish the system for the protection and control of water for human consumption, developed the application “Information System of Surveillance of Water Quality for Human Consumption-SIVICAP”, which allows all departmental health authorities to report the data of the monitoring of water quality, based on their inspection activities, surveillance and control in the country.

The SIVICAP WEB system allows the online reporting of water quality information and includes the calculation of the IRCA (for its acronym in Spanish: Índice de Riesgo de Calidad del Agua or Water Quality Index), IRABA (for its acronym in Spanish: Índice de Riesgo Municipal por Abastecimiento de Agua or Municipality Risk Index for Water Supplying), BPS (for its acronym in Spanish: Buenas Prácticas Sanitarias or Good Sanitary Practices) and RISK MAP indicators. Also, it allows sharing more efficiently the information generated and updated with different direct or indirect users of the sector.

Technically the IRCA index is defined as the degree of risk of disease occurrence related to non-compliance with the physical, chemical and microbiological characteristics of water for human consumption. This indicator is the result of assigning the risk score contemplated in Table 2 to the characteristics contemplated there for non-compliance with the acceptable values established. When the resulting score is between 0% and 5%, the water distributed is suitable for human consumption and is rated at the No-Risk level. When the IRCA is between

5.1% and 14%, is rated with a low-risk level and it is no longer suitable for human consumption; between 14.1% and 35% it qualifies with Medium level of irrigation and is not suitable for human consumption; between 35.1% and 80% the risk level is High and between 80.1% and 100% the water distributed is sanitary nonviable. When the monthly IRCA indicates that water is not suitable for human consumption, sanitary authority orders a series of actions for its improvement, being the most drastic when water is sanitarly unfeasible.

IRABA, on the other hand, is the weighting of two factors: (1) Treatment and continuity of the service of the aqueduct systems; and (2) Water distribution in the area of the corresponding municipality, which may indirectly affect the quality of water for human consumption and therefore, human health. This index aims to associate the risk to human health caused by the supply systems and establish the respective levels of risk, both at the level of the Provider (IRABApp) and at the level of municipality (IRABAm), where it takes into account the sum of the risk indices of all the suppliers for every municipality.

According to SIVICAP and INS (Nava et al. 2011) and National Decree (MADS 2007), the calculation of the risk index of water quality for human consumption - IRCA, is based on the following formulas:

$$IRCA(\%) = \frac{\sum_j A_j}{\sum_j N_j} \times 100\% \quad (1)$$

where A_j is the risk scores assigned to unacceptable characteristics, N_j is the risk scores assigned to all the characteristics analyzed, and the monthly average

$$IRCA(\%)_{prom} = \frac{\sum_j B_j}{N} \quad (2)$$

where B_j is the IRCAs obtained in each sample carried out in the month, and N is the total number of samples taken in the month.

In this sense, an acceptable characteristic (or acceptable measurement variable) is considered to be the one in which measurement is below the maximum acceptable value according to INS criteria. Of course, an unacceptable characteristic is that for which the value is above the maximum acceptable. Such criteria are shown in Table 1.

For the calculation of the IRCA, the risk score is assigned to each physical, chemical and microbiological characteristic, for non-compliance with the established acceptable values (see Table 2).

The value of the IRCA is zero (0) when it complies with the acceptable values for each of the physical, chemical and microbiological characteristics contemplated in the resolution and one hundred (100) for the highest risk when none of them is met.

According to the IRCA result per sample and the monthly IRCA, the following classification of the risk level of the water supplied for human consumption by the provider is defined, and the action that the competent health authority must carry out are indicated (see Table 3).

On the other hand, the IDEAM (for its acronym in Spanish: Instituto de Hidrología, Meteorología y Estudios Ambientales, or Institute of Hydrology, Meteorology and Environmental Studies) adopted the UWQI methodology (Universal Water Quality Index), which was developed and applied in order to obtain a simplified index to establish the quality of water used for human consumption. For its calculation, there is an additive-type equation or weighted sum was developed (Lozada et al. 2009), whose structure is the one presented in Eq. (3):

Table 1 Maximum acceptable values for each of the physical, chemical and biological characteristics used to calculate the IRCA (MADS 2007)

Parameter	Maximum permissible value	Unit
Apparent color	15	Cobalt Platinum Units (UPC)
Turbidity	2	Nephelometric Turbidity Units (UNT)
pH	6.5–9.0	mg/L
Residual chlorine	0.3–2.0	mg/L
Alkalinity	200	mg/L (CaCO ₃)
Calcium	60	mg/L (Ca)
Magnesium	36	mg/L (Mg)
Total hardness	300	mg/L (CaCO ₃)
Sulfates	250	mg/L (SO ₄ ²⁻)
Iron	0.3	mg/L (Fe ⁴)
Chlorides	250	mg/L (Cl ⁻)
Nitrites	0.1	mg/L (NO ₂ ⁻)
Total coliforms	0	UFC/100 cm ³
<i>Escherichia coli</i>	0	UFC/100 cm ³

$$UWQI = \sum_{j=1}^n W_j I_j \quad (3)$$

where W_j is the weight or percentage assigned to the j -th parameter and I_j is the subscript of the j -th parameter. The IDEAM (2015) adopted six basic variables for the determination of the IRCA in the bodies of water: one state (dissolved oxygen) and five pressure (chemical oxygen demand, COD; electrical conductivity, EC; total suspended solids, TSS; pH; and total nitrogen to total phosphorus ratio).

The values of the IRCA of the IDEAM comprise a scale from zero to one, in five categories: very bad, between 0.00 and 0.25 (represented with a red color); bad, between 0.26 and 0.50 (orange color); regular, between 0.51 and 0.70 (yellow color); acceptable, between 0.71 and 0.90 (green); and good, between 0.91 and 1.00 (blue color).

In this study, the WQI, which is calculated using the weighted arithmetic index method (Brown et al. 1970) is used to determine the effect of waste dumping on the immediate ground and surface water- bodies to the dumpsite, as it is deemed the most appropriate, based on the prevailing conditions.

Table 2 Risk score for the calculation of the IRCA (Nava et al. 2011; MADS 2007). Sum of assigned scores is equal to 100

Characteristic	Risk score	Characteristic	Risk score
Apparent color	6	Total hardness	1
Turbidity	15	Sulfates	1
pH	1.5	Total Iron	1.5
Free Residual Chlorine	15	Chlorides	1
Total Alkalinity	1	Nitrates	1
Calcium	1	Nitrites	3
Phosphates	1	Aluminum (Al ³⁺)	3
Manganese	1	Fluorides	1
Molybdenum	1	COT	3
Magnesium	1	Total coliforms	15
Zinc	1	<i>Escherichia Coli</i>	25

Table 3 Classification of the level of health risk according to the IRCA per sample and the monthly IRCA and actions that must be carried out (Nava et al. 2011; MADS 2007)

Classification IRCA (%)	Risk level	IRCA month (Requested actions)
80.1–100	Unviable sanitarily	Water not suitable for human consumption, direct management according to its competence of the provider, mayors, governors and entities of the national order.
35.1–80	High	Water not suitable for human consumption, direct management according to its competence of the person providing the service and of the respective mayors and governors.
14.4–35	Medium	Water not suitable for human consumption, direct management of provider.
5.1–14	Low	Water not suitable for human consumption, susceptible to improvement.
0–5	Without risk	Water suitable for human consumption. Continue the surveillance.

The WQI is given as:

$$WQI = \frac{\sum_j q_j w_j}{\sum_j w_j} \tag{4}$$

where q_j = quality rating (sub index) of j -th water quality parameter, w_j = unit weight of j -th water quality parameter (sum of all w_j is equal to one). Also, q_j , which relates the value of the parameter in polluted water to the standard permissible value is obtained as follows:

$$q_j = \frac{v_j - v_{lim}}{s_j - v_{lim}} \times 100\% \tag{5}$$

where v_j = estimated value of the j -th parameter, v_{lim} = ideal value of the j -th parameter, s_j = standard permissible value of the j -th parameter.

In most cases, $v_{lim} = 0$ except for pH and DO. In fact, for pH, $v_{lim} = 7$, and for DO, $v_{lim} = 14.6$ mg/L. The unit weight (w_j), which is inversely proportional to the values of the recommended standards is obtained as:

$$w_j = \frac{k}{s_j} \tag{6}$$

where $k = \left(\sum_{j=1}^n 1/s_j\right)^{-1}$.

Table 4 Rating of Water Quality for various WQIs (Brown et al. 1972; Chatterjee and Raziuddin 2002)

WQI value	Rating of Water Quality
0–25	Excellent
26–50	Good
51–75	Poor
16–100	Very poor
Above 100	Unsuitable for Drinking

The rating of the water quality using the above method is shown in Table 4 (Brown et al. 1972; Chatterjee and Raziuddin 2002).

4 Data, Methods and Parameter Selection

4.1 Data Available

The data used comes from the Open Data website of the Ministry of Information and Communication Technologies (www.datos.gov.co) supplied by the National Institute of Health. The data corresponding to the year 2015 has been selected in the study. The original data consists of 1019 records with 47 fields or columns of values for 40 characteristics or measured parameters. From this dataset, there were selected 14 parameters, namely: appearance, color, turbidity, pH, residual chlorine, alkalinity, calcium, magnesium, total hardness, sulfates, iron, chlorides, nitrites, total coliforms and Escherichia Coli.

In the filtering process, 87 records have been kept corresponding to each municipality of the department. In addition to this, a subset of features has been selected, due to a large amount of missing data for some parameters that have not been measured or have not been measured correctly. The selected parameters for water quality index in this study are those used for the development and sensitivity analysis of the global drinking water index (Drinking 2007) with greater weight and importance. These parameters were carefully selected to accurately reflect the main problems of acceptability and health related to water quality. In addition, factors including the level of detection and the overall ability of researchers and stakeholders to accurately measure parameters in most parts of the world were considered.

4.2 Water Quality Index Calculation

For the set of original data, the index for quality of water for drinking use was calculated from Eqs. (1) and (2) by the National Ministry of Health. However, in order to make comparisons for the calculation of WQI in its international standard form, Eqs. (4), (5) and (6) have been used. At the beginning was important to carry out filtering and cleaning of the data, because there was a lot of data lost concerning some of the variables considered. In this way, the variables with the greatest number of complete results were selected and the data set was homogenized. For these data, 14 variables of physical, chemical and biological parameters were selected for the study: apparent color, turbidity, pH, residual chlorine, alkalinity, calcium, magnesium, total hardness, sulfates, iron, chlorides, nitrites, total Coliforms and Escherichia Coli.

Likewise, the maximum values allowed for each parameter measured in the water samples were taken into account. These values are defined in the Colombian standard (MADS 2007) and are described in the Table 1.

4.3 Analysis of Water Samples

To evaluate the quality of surface water, eighty-seven different sampling stations were selected based on the position and geolocation of each of the municipalities of Santander with areas of probable impact and specific effluent discharge points. For each of the municipalities, they took several samples at different geographical points. The number of samples that were taken

into account with good measurement results can be seen in Fig. 2. The standard analytic procedures recommended by the Colombian standard in the manual of basic analytic methods for the analysis of waters for human consumption were used in the present study.

The obtained data is part of a citizen science strategy, in which the corresponding authorities provide and supply institutions and organize society with the material, technical elements and also measuring instruments for the collection and analysis of water samples. The data analyzed in this study correspond to those collected in the first quarter of 2015. Then, each institution and groups of citizen scientists take samples and measure the characteristics of the samples collected following a manual of procedures designed for this purpose. Samples and the results of the measurements are sent through a digital platform to central processing and are also sent to the central laboratories of the health institution, and a preliminary analysis of the conditions and quality of collected data is carried out. Of all the samples collected, a preliminary analysis is made to determine which of them meet the requirements of the procedure manual. Data that do not meet the quality criteria are discarded and those that meet the standards defined in the procedures manual are maintained. At this point, data that does not meet a minimum of parameters measured for the determination of the water quality index is also discarded. Health authorities geographically determine sample collection points, so that all participants must collect samples at specifically defined points. Those samples that do not meet this requirement are also discarded. Despite having different geographic points for data collection, the health authority has determined all these collection sites based on the geographical distance between points which does not exceed the amount of 1 km and on the same tributary, whereby it can be assumed that geographical dispersion is not an error factor in the calculation of water quality indices.

At this point a homogeneity and regularity of the data collected are ensured. Therefore, information is obtained that represents the sample collection area regardless of the positions in which they have been measured. Thus, before calculating the water quality index, averages of each characteristic measured are taken for all samples, and based on these averages the water quality index is calculated according to the methodology of Eqs. (1) and (2).

4.4 Multivariate Analysis Methods

Multivariate Factor Analysis is a class of multivariate statistical method whose main purpose is to expose the underlying structure in a data matrix. Analyzing the structure of the interrelations between a large number of variables does not require any distinction between dependent and

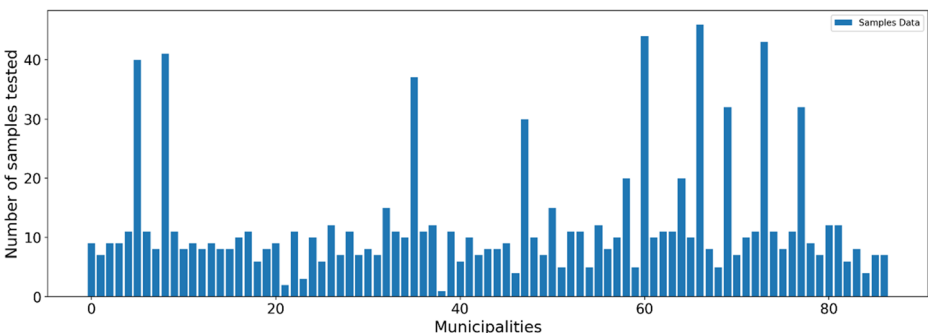


Fig. 2 Number of samples tested at each location in each municipalities

independent variables. Using this information, one can calculate a set of latent dimensions, known as factors, that seek to explain these interrelationships. It is, therefore, a data reduction technique, the information contained in the data matrix can be expressed without much distortion in a smaller number of dimensions represented by said factors.

In order to evaluate the significant differences between the sites for all the water quality variables, the data were analyzed through the analysis of variance. The multivariate analysis of the water quality data sets was done through hierarchical group analysis (HCA) and principal components analysis (PCA) (Jolliffe 2011). The objective of clustering is to divide the objects into homogeneous groups so that the similarities within the group are large compared to the similarities between groups. The Principal Components, on the other hand, are extracted to represent the patterns that encode the highest variance in the data set and not to maximize the separation between groups of samples directly. The statistical package used in this case was R version 3.4.4 (2018-03-15) (R Core Team 2018; Bunn 2008, 2010). The software was used for both the HCA and the PCA.

The elaboration of risk maps in the studied area as well as the reduction of the data was possible thanks to the use of a script written completely in Python 3.6, using standard libraries of Pandas 0.24.0 (McKinney et al. 2010), Matplotlib 3.0.2 (Barrett et al. 2005), and Scipy 1.0.2 (Oliphant 2007). Several steps were taken to construct a water table contour map; in this case for this work, we used Basemap 1.0.8 extension (Whitaker 2017) taking into consideration the geographical location of all samples.

5 Results and Discussion

As mentioned earlier, quantification through a WQI is a valuable tool to examine and establish criteria to evaluate the general state of water quality in a single factor that is useful for the selection of the appropriate treatment technique to solve problems and inform to authorities and decision-makers about the actions that must be taken to solve such problems in terms of drinking water consumption. In this study, the WQI has been calculated for the 87 municipalities of the Department of Santander for an equivalent number of samples between 3 and 45 in each station.

The graphs of Fig. 3 indicate the variation of the WQI in the sampling stations using both the conventional and official methodology for the Colombian State, and the international standard methodology for the calculation of the WQI. As expected, most sites are within acceptable limits with any of the calculated indices.

In Figs. 3 and 4 one can observe that during this period of the study, most of the stations reported water either with low risk or risk free ($\approx 57\%$), or of excellent and good water quality ($\approx 73\%$) according to the methodology used for the quality index. However, it is evident that according to the index measured by the Colombian standard methodology, none or very few samples have been reported of high risk or sanitary infeasibility, while with the standard international calculation method, a little more than 25% of the stations report poor or nonviable sanitary water quality, with about 16% of these unsuitable for human consumption. In the method used to calculate water quality in this study, the most affected stations measuring water quality from poor to unfit were stations 1–4, 11, 14, 28–30, 34, 37, 58, 63–64, 68, 73, 75, 79, 81, 84 and 85. An inevitable consequence of this result is that the methodology used by local agencies in charge of monitoring water quality in Colombia may be underestimating the high risk levels for water quality.

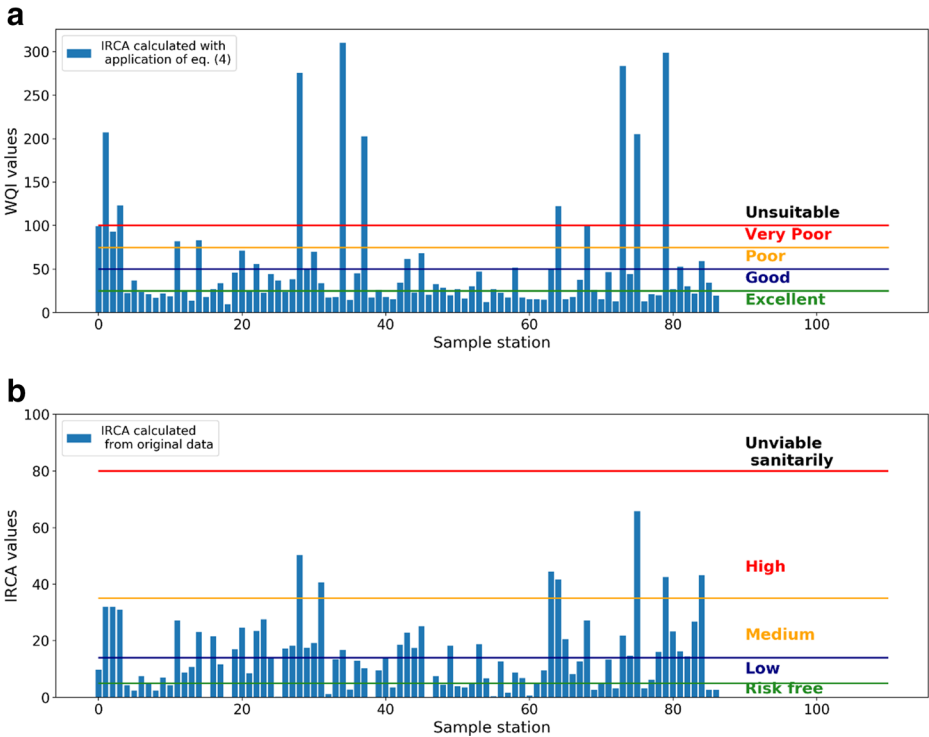


Fig. 3 Water quality index variation in the two methodologies applied

Table 5 presents the average values of the physical and chemical parameters measured at the different sites. The analysis of this table shows that there is a highly significant difference ($p \leq 0.001$) among the sites when considering all the physical and chemical parameters.

In addition, by examining Table 5 and comparing it with Table 1 for the maximum permissible values, we notice that some of the characteristics have third quartiles above these maximum permissible values. In effect, the measured biological parameters report that at least 25% of the samples have values above the maximum allowable and alarmingly the total coliform content exceeds the barrier of 400 UFC/100 mL. In addition, the content of *E. coli* also exceeds the maximum allowable in 50% of the samples.

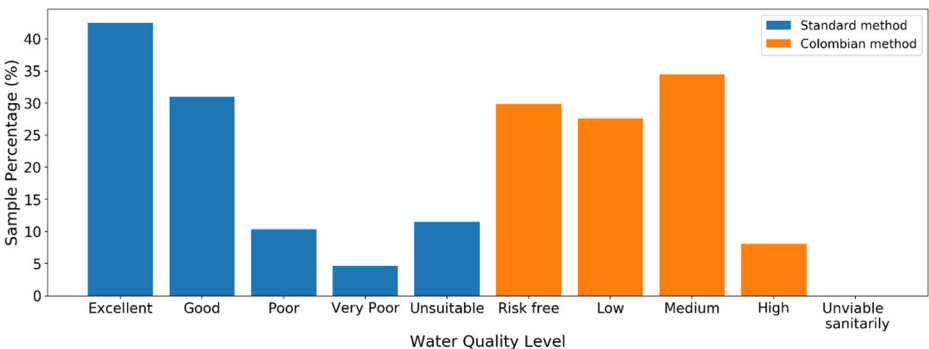


Fig. 4 Comparison of criteria and levels of Water quality index variation in the two methodologies applied

Table 5 Results of analysis of variance showing the effect of stations on physico-chemical parameters

Parameter	Mean	StD	Min	25%	75%	Max	t-statistic	p value two-tailed
App Color	2.8021	5.6824	0.0000	0.5025	3.0000	49.0000	4.5995	1.4463E-05
Turbidity	1.7141	3.5346	0.0046	0.3521	1.6677	28.9580	4.5234	1.9360E-05
pH	7.6364	3.7404	5.8050	7.1191	7.4400	41.9750	19.0426	1.3523E-32
Res Chloride	1.1263	1.2254	0.0000	0.7253	1.2774	11.0000	8.5727	3.6456E-13
Alkalinity	46.1460	34.1445	3.6363	16.3625	67.8848	136.375	12.6058	3.0271E-21
Calcium	15.7052	10.2095	0.0000	7.0142	23.9982	40.9250	14.3481	1.5398E-24
Magnesium	11.7982	8.1094	0.9350	5.2829	16.3498	31.1514	13.5702	4.3524E-23
Hardness	66.6996	45.7495	4.3000	29.0942	94.8154	231.2437	13.5986	3.8469E-23
Sulfates	27.5906	15.6011	0.1400	17.2241	37.5190	69.8700	16.4953	2.2535E-28
Iron	0.1039	0.0902	0.0100	0.0541	0.1178	0.6663	10.7519	1.3818E-17
Chloride	7.2903	2.9891	3.3275	5.5896	8.2826	23.1110	22.7486	3.8033E-38
Nitrites	0.0148	0.0061	0.0000	0.0111	0.0177	0.0412	22.6598	5.0779E-38
Coliforms	386.6222	790.6936	0.0000	1.0539	407.4318	3838.9512	4.5607	1.6786E-05
E Coli	58.8289	229.5989	0.0000	0.0000	8.7575	1443.3333	2.3899	1.9037E-02
IRCA CAL	53.2801	65.8380	9.5876	18.3022	51.2033	310.0423	7.5482	4.2963E-11

The results of the principal component analysis reveal that the first two independent variables that result from the decomposition study add up to 46% of the variability in the influence of physicochemical parameters at all stations.

The analysis of the correlation between the physicochemical parameters and their contributions to these variables (Fig. 5) shows that the biological parameters total coliforms and *E. coli* and the physical variables apparent color and turbidity, and the chemical variable nitrites, are positively correlated with dimensions 1 and 2. Likewise, the content of sulphates, magnesium and calcium, as well as alkalinity and total hardness are strongly correlated positively with dimension 1 and negatively with dimension 2.

The analysis of the projection of the different groups of samples (Fig. 6) on the dimensions of independent variables shows that there is a very well identified group which corresponds to the sampling sites 2, 35, 69, 76 and 80. This group identified as cluster 3, is presented with very high values of content of biological parameters total coliforms and *E. coli*, physical variables apparent color and turbidity, and chemical variables such as nitrites, while with low values of the remaining variables. Similarly, a second group identified as cluster 2, is presented with high values of sulfate, magnesium and calcium content, as well as alkalinity and hardness, and low values of the other parameters.

In the current data set, there are 6 factors (Tables 6 and 7) that can be designated as intermediate compounds. The first intermediate factor includes alkalinity (with a weight of 0.20), calcium (weight of 0.17), magnesium (weight of 0.21) and hardness (weight of 0.21). The second intermediate compound includes apparent color (weight 0.28), turbidity (weight of 0.24) and nitrites (weight 0.13). The third factor includes the biological variables coliforms (with a weight of 0.19) and *E. coli* (with a weight of 0.26). The last three intermediate compounds include pH (weight 0.33), residual chloride (weight 0.44), iron (weight 0.19), sulfates (weight 0.31), and chlorides (weight 0.44). In addition, six intermediate compounds are added with weights assigned in proportion to the variance explained. After the determination of the weights, the final aggregation can be performed using a hybrid aggregation of weighted arithmetic and weighted harmonic mean (Panda et al. 2016; Esty et al. 2005; Mohd Ali et al. 2013; Tripathi and Singal 2019). In this case, the weighted harmonic average of the individual factor loads and the weighted arithmetic average are calculated to obtain the final

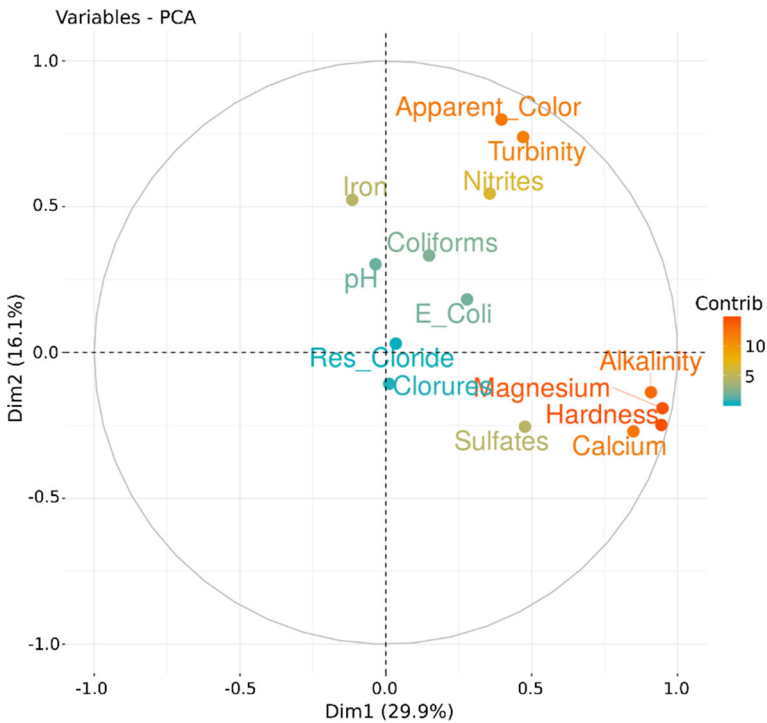


Fig. 5 Projection of physical and chemical parameters on the factorial planes determined for principal component analysis

results. This methodology could give clues about a modification of the core weights for the risk score for the calculation of the IRCA (Table 2).

The water table contour maps for the samples and the calculated water quality index were constructed using a cubic interpolation method with the help of GIS software (Surfer) and Basemap tool for Python. In this case it can be seen that the equally spaced contours define areas with a higher water quality index, reflecting very poor quality of this element. Moreover, according to Fig. 7, it can be noted that the areas with the worst water quality are located towards the center of the Department of Santander, which is dominated by topographic conditions of high altitudes and quite irregular territories but which coincide with municipalities of high oil activity and mining, as are the municipalities of Albania, San Vicente De Chucurí, Sucre, El Playon, Gambita and Guadalupe. So, the high rates evidenced in these municipalities could be linked to the commercial activity of each one respectively. Besides this, these municipalities are generally small, not exceeding 12 thousand inhabitants, and also far from the capital Bucaramanga.

6 Conclusions

This study has shown that WQI is a powerful and simple tool, which can be used to accurately determine the impact of different pollution activities on ground and surface waters. We have made water quality analysis for the northwest area in the Andean region of Colombia. For this

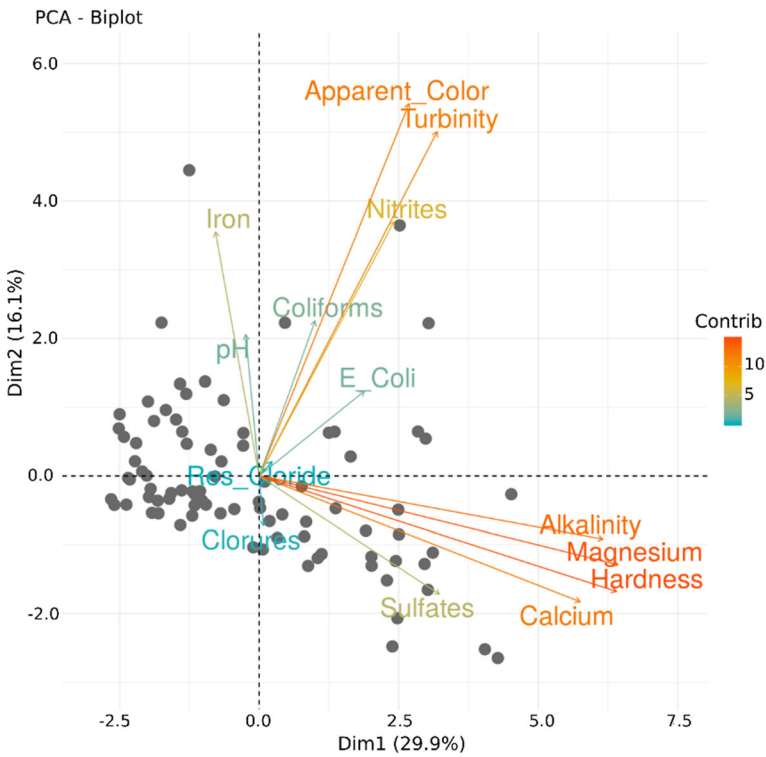


Fig. 6 Projection of stations and groups on factorial determined for principal component analysis

analysis, a comparison of two different methods for the calculation of WQI was made demonstrating that the method used locally in Colombia generally leaves very good water quality in most of the area, compared to the estimate using an international version of WQI.

Table 6 Squared Factor loadings of water quality data based on 6 PCs

	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5	Dim.6
Apparent Color	0.16	0.64	0.00	0.10	0.00	0.02
Turbidity	0.22	0.55	0.00	0.10	0.00	0.04
pH	0.00	0.09	0.24	0.51	0.02	0.00
Res Chloride	0.00	0.00	0.26	0.01	0.45	0.09
Alkalinity	0.83	0.02	0.03	0.00	0.01	0.02
Calcium	0.72	0.07	0.01	0.00	0.01	0.00
Magnesium	0.90	0.04	0.01	0.00	0.00	0.00
Hardness	0.89	0.06	0.01	0.00	0.00	0.00
Sulfates	0.23	0.06	0.01	0.08	0.01	0.26
Iron	0.01	0.27	0.24	0.29	0.03	0.01
Chloride	0.00	0.01	0.09	0.14	0.45	0.19
Nitrites	0.13	0.30	0.03	0.06	0.00	0.01
Coliforms	0.02	0.11	0.32	0.13	0.04	0.15
E Coli	0.08	0.03	0.44	0.13	0.01	0.05
Sum of sq. load	4.19	2.26	1.69	1.54	1.02	0.85
Weights of factors	29.89	16.14	12.10	11.01	7.30	6.07
Cumulative weights of factors	29.89	46.03	58.13	69.14	76.44	82.51

Table 7 Squared Factor loadings (scaled to unity sum)

	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5	Dim.6
Apparent Color	0.04	0.28	0.00	0.06	0.00	0.03
Turbidity	0.05	0.24	0.00	0.06	0.00	0.05
pH	0.00	0.04	0.14	0.33	0.02	0.00
Res Chloride	0.00	0.00	0.15	0.01	0.44	0.10
Alkalinity	0.20	0.01	0.02	0.00	0.01	0.02
Calcium	0.17	0.03	0.01	0.00	0.01	0.00
Magnesium	0.21	0.02	0.00	0.00	0.00	0.01
Hardness	0.21	0.03	0.01	0.00	0.00	0.00
Sulfates	0.05	0.03	0.01	0.05	0.01	0.31
Iron	0.00	0.12	0.14	0.19	0.03	0.01
Chloride	0.00	0.01	0.05	0.09	0.44	0.23
Nitrites	0.03	0.13	0.02	0.04	0.00	0.01
Coliforms	0.01	0.05	0.19	0.08	0.04	0.17
E Coli	0.02	0.01	0.26	0.08	0.01	0.06

According to the latter, severely deteriorated conditions were detected during the year in the stations of the rivers and sources of urban water supply, implying that this quality has been overestimated. It has been shown that surface water in the sampling stations is not suitable for drinking and for other daily uses of the farmers who live there. It was found that most surface

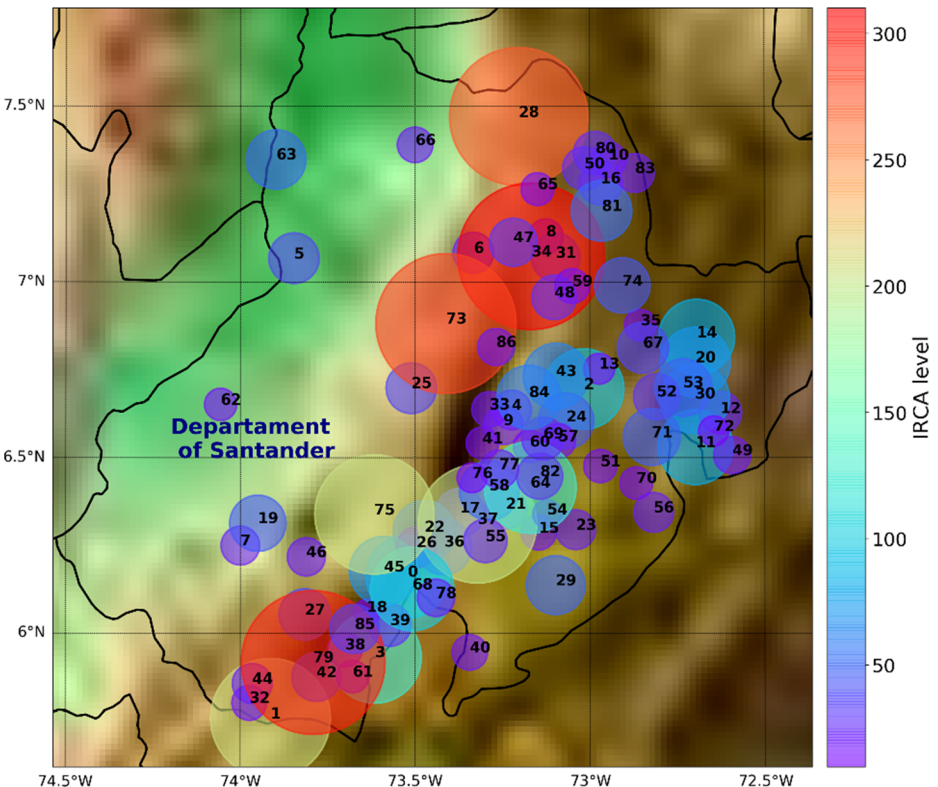


Fig. 7 Heatmap for the WQI evaluation on sample points

water sampling sites have a moderately satisfactory or poor water quality, with several exceptions.

This study showed the efficiency of multivariate statistical techniques and the water quality index to analyze and interpret a set of data for an effective evaluation of surface waters. From the above, it is necessary to implement standards and more thorough and rigorous evaluation for the preservation of the health of people who depend on the water resource as well as the preservation of the mineral necessary for the sustainability of the population.

Individual evaluations for the analysis of the water quality of the stations has been possible thanks to the examination of surface water quality parameters found in three groups. This clustering allows efficient management of surface water quality. The importance of the parameters in the studies of contamination and evaluation of water quality from a PCA analysis shows that the parameters that are most important are in factor 1 and the least important are in the second factor.

This article uses advanced statistical techniques of factor analysis in the estimation of weights for the development of a new water quality index. Statistical techniques make the final index more objective by nature, and therefore, reduce any bias in its application in different locations. In addition, factor analysis (FA) uses these groups to provide weight to the individual parameters, as well as to the “loads” of factors that represent the extension of the principal components analysis (PCA) to FA. In this study, FA was applied to 9 parameters preselected by PCA. This study it can be concluded that alkalinity, calcium, magnesium and hardness contribute to Factor 1, apparent color, turbidity and nitrites to Factor 2, coliforms and *E. coli* to Factor 3, pH, residual chloride, iron, sulfates and chlorides to other factors. In addition to the conclusions of this study, the results will be used in the future for the development of a new Water Quality Index.

Acknowledgements The main author would like to thank the National Institute of Health of Colombia for having openly provided the data (through the open data portal www.datos.gov.co) that have been used for this study. The data, as well as the Python code developed for the treatment, organization, analysis and visualization of the data is available for use by anyone from the author’s Github account: <https://github.com/sierraporta/Water-Quality>. The main author would also like to thank A. G. Muñoz S. and J. Díaz-Lobatón for a critical review of this manuscript and for their helpful comments.

References

- Barrett P, Hunter J, Miller JT, Hsu JC, Greenfield P (2005) Matplotlib—a portable python plotting package. *Astronomical Data Analysis Software and Systems XIV* 347:91
- Bengraïne K, Marhaba TF (2003) Using principal component analysis to monitor spatial and temporal changes in water quality. *J Hazard Mater* 100(1–3):179–195. [https://doi.org/10.1016/S0304-3894\(03\)00104-3](https://doi.org/10.1016/S0304-3894(03)00104-3)
- Brown RM, McClelland NI, Deininger RA, Tozer RG (1970) A water quality index- do we dare. *Water Sewage Works* 117(10):339–343. <https://doi.org/10.1007/s13201-015-0318-7>
- Brown R. M, McClelland N. I, Deininger R. A, O’Connor M. F (1972) A water quality index—crashing the psychological barrier. In: *Indicators of Environmental Quality*, Springer, pp 173–182. https://doi.org/10.1007/978-1-4684-2856-8_15, A Water Quality Index — Crashing the Psychological Barrier
- Bunn AG (2008) A dendrochronology program library in r (dplR). *Dendrochronologia* 26(2):115–124. <https://doi.org/10.1016/j.dendro.2008.01.002>
- Bunn AG (2010) Statistical and visual crossdating in r using the dplR library. *Dendrochronologia* 28(4):251–258. <https://doi.org/10.1016/j.dendro.2009.12.001>
- Chatterjee C, Raziuddin M (2002) Determination of water quality index (WQI) of a degraded river in Asansol industrial area (West Bengal). *Nat Environ Pollut Technol* 1(2):181–189

- Cude CG (2001) Oregon water quality index a tool for evaluating water quality management effectiveness. *J Am Water Resour Assoc* 37(1):125–137. <https://doi.org/10.1111/j.1752-1688.2001.tb05480.x>
- Delpla I, Jung AV, Baures E, Clement M, Thomas O (2009) Impacts of climate change on surface water quality in relation to drinking water production. *Environ Int* 35(8):1225–1233. <https://doi.org/10.1016/j.envint.2009.07.001>
- Drinking G (2007) Water quality index development and sensitivity analysis report. United Nations Environment Program Global Environment Monitoring System/Water Program. <http://www.gemswater.org>. Accessed 26 June 2018
- Esty D. C, Levy, M. A, Srebotnjak T, de Sherbinin, A. (2005) 2005 environmental sustainability index: benchmarking National Environmental Stewardship. New Haven, CT: Yale Center for Environmental Law & Policy
- Fernández N, Ramírez A, Solano F (2004) Physico-chemical water quality indices—a comparative review. *Bistua Revista de la Facultad de Ciencias Básicas* 2(1)
- Fernández-Méndez F, Bernate-Peña JF, Melo O (2013) Diversidad arbórea y prioridades de conservación de los bosques secos tropicales del sur del Departamento del Tolima en el Valle del río Magdalena. *Actualidades Biológicas* 35(99):161–83.
- Guzmán G (2016) The Chicamocha River Canyon. In: Hermelin M (ed) *Landscapes and Landforms of Colombia*. World Geomorphological Landscapes. Springer, Cham. https://doi.org/10.1007/978-3-319-11800-0_6
- Horton RK (1965) An index number system for rating water quality. *Journal of Water Pollution Control Federation* 37(3):300–306. <https://doi.org/10.12691/ajwr-1-3-3>
- Jolliffe I (2011) Principal component analysis. *International Encyclopedia of Statistical Science*, Springer, In, pp 1094–1096. https://doi.org/10.1007/978-3-642-04898-2_455
- Kanu I, Achi O (2011) Industrial effluents and their impact on water quality of receiving rivers in Nigeria. *Journal of Applied Technology in Environmental Sanitation* 1(1):75–86
- Lozada P. T, Vélez C. H. C, Patino P (2009) Indices de calidad de agua en fuentes superficiales utilizadas en la producción de agua para consumo humano. una revisión crítica. *Revista de Ingenierías: Universidad de Medellín* 8(15):3
- MADS (2007) Ministerio de Ambiente y Desarrollo Sostenible, Resolución Número 2115 (22
- McKinney W, et al. (2010) Data structures for statistical computing in python. In: proceedings of the 9th Python in science
- Mohd Ali Z, Ibrahim NA, Mengersen K, Shitan M, Juahir H (2013) The Langat river water quality index based on principal component analysis. *AIP Conference Proceedings*, AIP, vol 1522:1322–1336. <https://doi.org/10.1063/1.4801283>
- Murdoch PS, Baron JS, Miller TL (2000) Potential effects of climate change on surface-water quality in North America. *J Am Water Resour Assoc* 36(2):347–366. <https://doi.org/10.1111/j.1752-1688.2000.tb04273.x>
- Nava G, Riaño D, Martínez L, Murillo C, Peralta A (2011) Manual de métodos fisicoquímicos básicos para el análisis de aguas para consumo humano. URL <http://www.ins.gov.co/sivicap/Documentacin%20SIVICAP/2011%20Manual%20análisis%20fisico%20químico%20aguas.pdf>
- Ocampo-Duque W, Osorio C, Piamba C, Schuhmacher M, Domingo JL (2013) Water quality analysis in rivers with non-parametric probability distributions and fuzzy inference systems: application to the Cauca river, Colombia. *Environ Int* 52:17–28. <https://doi.org/10.1016/j.envint.2012.11.007>
- Oliphant TE (2007) Scipy: open source scientific tools for python. *Computing in Science and Engineering* 9:10–20. <https://doi.org/10.1109/MCSE.2007.58>
- Panda S, Chakraborty M, Misra S (2016) Assessment of social sustainable development in urban India by a composite index. *Int J Sustain Built Environ* 5(2):435–450. <https://doi.org/10.1016/j.ijbs.2016.08.001>
- Pesce SF, Wunderlin DA (2000) Use of water quality indices to verify the impact of Cordoba city (Argentina) on Suquia River. *Water Res* 34(11):2915–2926. [https://doi.org/10.1016/S0043-1354\(00\)00036-1](https://doi.org/10.1016/S0043-1354(00)00036-1)
- Poonam T, Tanushree B, Sukalyan C (2013) Water quality indices—important tools for water quality assessment: a review. *International Journal of Advances in Chemistry* 1(1):15–28. <https://doi.org/10.5121/ijac.2015.1102>
- R Core Team (2018) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, URL <https://www.R-project.org/>
- S'tambuk-Giljanovi'c N (1999) Water quality evaluation by index in Dalmatia. *Water Res* 33(16):3423–3440. [https://doi.org/10.1016/S0043-1354\(99\)00063-9](https://doi.org/10.1016/S0043-1354(99)00063-9)
- Tripathi M, Singal SK (2019) Allocation of weights using factor analysis for development of a novel water quality index. *Ecotoxicol Environ Saf* 183:109510. <https://doi.org/10.1016/j.ecoenv.2019.109510>
- Tüfekci N, Sivri N, Toroz I (2007) Pollutants of textile industry wastewater and assessment of its discharge limits by water quality standards. *Turk J Fish Aquat Sci* 7(2):97–103

- Tyagi S, Sharma B, Singh P, Dobhal R (2013) Water quality assessment in terms of water quality index. *American Journal of Water Resources* 1(3):34–38. <https://doi.org/10.12691/ajwr-1-3-3>
- Vörosarmaty CJ, Green P, Salisbury J, Lammers RB (2000) Global water resources: vulnerability from climate change and population growth. *Science* 289(5477):284–288. <https://doi.org/10.1126/science.289.5477.284>
- Whitaker J (2017) Basemap matplotlib toolkit 1.0. 8 documentation, online. <http://matplotlib.org/basemap>
- Whitehead P, Wilby R, Battarbee R, Kernan M, Wade AJ (2009) A review of the potential impacts of climate change on surface water quality. *Hydrol Sci J* 54(1):101–123. <https://doi.org/10.1623/hysj.54.1.101>
- Wilby R, Whitehead P, Wade A, Butterfield D, Davis R, Watts G (2006) Integrated modelling of climate change impacts on water resources and quality in a lowland catchment: river Kennet, UK. *J Hydrol* 330(1–2):204–220. <https://doi.org/10.1016/j.jhydrol.2006.04.033>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.