scientific reports

Check for updates

OPEN

Water quality assessment based on multivariate statistics and water quality index of a strategic river in the Brazilian Atlantic Forest

David de Andrade Costa[1,2]✉, José Paulo Soares de Azevedo[1]✉, Marco Aurélio dos Santos[1] & Rafaela dos Santos Facchetti Vinhaes Assumpção[3]

Fifty-four water samples were collected between July and December 2019 at nine monitoring stations and fifteen parameters were analysed to provide an updated diagnosis of the Piabanha River water quality. Further, forty years of monitoring were analysed, including government data and previous research projects. A georeferenced database was also built containing water management data. The Water Quality Index from the National Sanitation Foundation (WQI$_{NSF}$) was calculated using two datasets and showed an improvement in overall water quality, despite still presenting systematic violations to Brazilian standards. Principal components analysis (PCA) showed the most contributing parameters to water quality and enabled its association with the main pollution sources identified in the geodatabase. PCA showed that sewage discharge is still the main pollution source. The cluster analysis (CA) made possible to recommend the monitoring network optimization, thereby enabling the expansion of the monitoring to other rivers. Finally, the diagnosis provided by this research establishes the first step towards the Framing of water resources according to their intended uses, as established by the Brazilian National Water Resources Policy.

Aquatic systems have been significantly affected by human activities causing water quality deterioration, decreasing water availability and reducing the carrying capacity of aquatic life[1–4]. Water quality deterioration still persists in developed countries, while it is a major problem in developing countries in which a substantial amount of sewage is discharged directly into rivers[5–8]. Moreover, according to UNEP[9], water pollution has worsened since the 1990s in the majority of rivers in Latin America. The global concern with water availability and its quality has been growing, and it is estimated that the demand for water will increase between 20 and 30% by 2050[10,11]. In addition, spatial and temporal variations in the hydrological cycle and their uncertainties related to climate change may worsen this scenario[12–16].

Monitoring water quality in order to assess its spatial and temporal variations is essential for water management and pollution control[17]. On the other hand, monitoring programs generate large data sets that require interpretation techniques[18]. There are a number of methods for water quality assessment, including single-factor, multi-index, fuzzy mathematics, grey system evaluation, artificial neural network, multi-criteria analysis, geographical interpolation and multivariate statistical approach[18–25]. Among them, the most used are the Water Quality Indexes (WQI) that transform a complex set of data into a single value indicative of water quality[26,27] and reflect its suitability for different uses[28]. Multivariate statistics is another widely used approach[29,30], mainly with Principal Components Analysis (PCA) and Cluster Analysis (CA), helping to achieve a better understanding of the spatial and temporal dynamics of water quality.

A comparison of seven methods for assessing water quality indicated WQI as one of the best[20]. The assessment of Poyang Lake[28], China and the upper Selenga River[31], Mongolia showed that WQIs are suitable for the assessment of both interannual trends and seasonal variations[28]. Multivariate statistical techniques associated with WQI have been used for numerous water bodies world-wide, including the Nag River[30], India, the Paraíba

[1]Federal University of Rio de Janeiro (UFRJ), Alberto Luiz Coimbra Institute for Graduate Studies and Engineering Research (COPPE), Centro Tecnológico, Cidade Universitária, Rio de Janeiro, RJ, Brazil. [2]Federal Fluminense Institute, São João da Barra Advanced Campus, BR 356, KM 181, São João da Barra, RJ, Brazil. [3]Oswaldo Cruz Foundation (Fiocruz), National School of Public Health Sergio Arouca (ENSP), Rua Leopoldo Bulhões, 1.480, Manguinhos, Rio de Janeiro, RJ, Brazil. ✉email: davidcosta@id.uff.br; azevedo@poli.ufrj.br

do Sul River[32], Brazil, and the before mentioned Selenga River[31]. CA grouped the monitoring stations according to their similarities, while the PCA highlighted components that were related to its pollution sources[30–32].

In order to ensure water quantity and quality, the Brazilian National Water Resources Policy[33] has established a management tool called Framework, according to the main intended uses of water. It has also created participatory management committees, the so-called Basin Committees, which, together with its technical agency, are responsible for the Framework establishment. Unfortunately, even after two decades, Brazil has had very few successful experiences on the subject[34].

Brazil has a gigantic and complex hydrographic network present in many different ecosystems[34]. The Brazilian Atlantic Forest is one of the most biodiverse biomes on the planet[35,36], extending along the Brazilian coast and currently covering only 11.4% of its original territory[37] under constant threats[38–40]. The hydrographic basin of the Paraíba do Sul river is located in this environment, which is the integration axis of the most industrialized Brazilian states, São Paulo, Rio de Janeiro and Minas Gerais, and home to around 6.2 million people[41]. A water transfer system regularly supplies another 9 million people in the metropolitan region of Rio de Janeiro, through the Guandu system. Another water transfer system connects the Paraíba do Sul river to the Cantareira system, complementing with 5 m³/s the water supply to over 9 million people in the metropolitan region of São Paulo[41]. These systems went through an intense water scarcity between 2014 and 2016 with severe impacts on water quality and availability[32].

Our study is focused on the Piabanha River watershed, a strategic sub-basin of the Paraíba do Sul river, combining urban, industrial, rural characteristics, and large preserved fragments of Atlantic Forest[36,42]. The Piabanha Basin has been monitored for over 10 years with the Studies in Experimental and Representative Watersheds (EIBEX) project, a partnership between universities and government agencies[42–44]. The State Environmental Agency of Rio de Janeiro (INEA) has been monitoring the basin since 1980. Other studies in the region include the analysis of contamination by pesticides[45], energy generation[46] and dispersion of pollutants[47]. The Piabanha Basin received international attention in Nature's article on biodiversity[36]. But in addition to forest preservation, can the Piabanha River support biodiversity? How is its water quality today? In this way, the Piabanha Basin Committee defined the Framework as a priority in its management plan (2018–2020) and to accomplish this goal, established water monitoring as a strategic action[48].

Our study covers 40 years of monitoring, including government data, our research projects and, currently, a monitoring program that is being conducted with funding from the Piabanha Basin Committee. The main objectives were: (1) to carry out an updated diagnosis of water quality using multivariate techniques and WQI; (2) to examine the parameters that most influence water quality, and (3) to identify river stretches with similar water quality. Our study provides an extensive understanding of the Piabanha River and supports its Steering Committee in the application of public policies. This is a pilot project that can be a reference for other Framework programs for improving water quality in Brazil.

Results

Water uses. We have requested and received from INEA two water user databases of the Piabanha Basin. The first set corresponds to raw data from the National Water Resources Register (CNARH), with all the registrations until December 2017 and with 1549 registered interferences (water abstraction or effluent discharge). The second one is the registration validated by INEA until August 2018 by the Águas do Rio project comprising a total of 669 validated interferences. With these data, it was possible to build a georeferenced base. By so doing, it was possible to list the main effluent discharges by type for each monitoring station.

In the validated database, from the 669 interferences, 84% are water abstractions and 16% are effluent discharges. Water abstraction account for 425 m³ day⁻¹ with 75% from wells and 25% from rivers. On the other hand, effluent discharges are 89 m³ day⁻¹. The largest volume of effluents comes from the sanitation sector with 57% of the total, whereas industries account for 33%, aquaculture with 4% and mining for 3% of discharges.

When comparing the two databases, it is clear that the universe of registered users is much larger than the universe of validated users; in other words, those whose data were made up by the state environmental agency and, therefore, received a license. For example, the validated database has only six interferences related to agriculture, in contrast to 789 interferences awaiting validation. This is a serious obstacle for water resources management in the region, which threatens the sustainability of water resources.

Short time monitoring and water quality index. In order to assess and compare the water quality of the Piabanha River, we calculated the Water Quality Index from the National Sanitation Foundation (WQI$_{NSF}$) using two datasets, the first one from 2012 and the last one from 2019 (Tables 1 and Table 2). The 2012 results (Fig. 1A) oscillated between the bad and medium categories, generally with medium quality (50.5 ± 10.3). In 2019 (Fig. 1B), the results ranged between the medium and good categories, in general with medium quality (61.6 ± 10.8).

Data sets show significant seasonal behavior (p < 0.05) (Fig. 1C,D) between the end of the dry period (Jul, Aug, Sep) and the beginning of the rainy period (Oct, Nov, Dec) for the parameters DO, WT, pH, nitrate, phosphate and turbidity, while no significant seasonal difference (p > 0.05) was found for the parameters *E. coli*, BOD and TDS. The parameters that have most impacted the WQI$_{NSF}$ were coliforms and BOD. Ammonia and total phosphorus do not account to WQI$_{NSF}$, but their concentration has violated Brazilian legislation and their influence can be better understood by PCA.

Principal components and clusters analysis. The 2019 dataset (n = 48), comprising six monitoring campaigns at the eight monitoring stations along the Piabanha River with 15 parameters analysed, was grouped by the average value of each parameter at each station (n = 8). Pearson's correlation matrix is presented in Table 3,

Parameters	Abbreviation	Units	Quantification limits	Permissible limits—class 2	Method
Electrical conductivity	EC	µS/cm	0.01	–	Field measure
Water temperature	Temp	°C	0.1	–	Field measure
Turbidity	Turb	NTU	0.01	100	Field measure
Dissolved oxygen	DO	Mg L^{-1}	0.01	>5	Field measure
pH	pH	pH units	0.01	6–9	Field measure
Total dissolved solids	TDS	mg L^{-1}	10	500	SMWW 2540C
Suspended solids	SS	mg L^{-1}	1	–	SMWW 2540D
Alcalinity	Alcal	mg/L ($CaCO_3$)	3	–	SMWW 2320B
Biological oxygen demand	BOD	mg L^{-1}	2	5	SMWW 5210B
Chemical oxygen demand	COD	mg L^{-1}	10	–	SMWW 5220D
Escherichia coli	*E. coli*	CFU/100 mL	1	1000	SMWW 9223A/B
Phosphate	PO_4^{3-}	mg L^{-1}	0.02	–	SMWW 4500 P E
Total phosphorus	TP	mg L^{-1}	0.02	0.1	SMWW 4500 P E
Nitrate	NO_3^-	mg L^{-1}	1	10.0	SMWW 4500D
Ammonium	NH_3	mg L^{-1}	0.06	3.7 (pH ≤ 7.5)	SMWW 4500F
Total nitrogen (Kjeldahl)	TN	mg L^{-1}	2	2.18	SMWW 4500A

Table 1. Parameters, abbreviations, units, quantification limits, permissible limits to rivers class 2 and methods. Field measurements were performed using a multiparameter probe (YSI model 556) and a portable turbidimeter (HANNA model HI 98703-0). Laboratory analyses follows the Standard Methods for the Examination of Water and Wastewater (SMWW).

Dist (km)	Station	Season	DO	Water temp	*E. coli*	pH	BOD	NO_3^-	PO_4^{3-}	Turb	TDS	WQI$_{NSF}$	SS	Alcal	COD	TP	NH_3	TN
12	1	Dry	7.71	21.04	7467	7.26	22.67	1.00	0.58	4.72	173	52.03	26	67	59.33	0.83	8.00	8.97
		Wet	8.72	21.79	3673	6.73	3.33	1.00	0.43	5.85	158	62.77	10	57	16.33	0.74	5.92	6.73
15	2	Dry	7.83	20.42	67,867	7.38	10.67	1.16	0.83	8.26	167	55.77	14	77	31.00	1.21	9.75	12.00
		Wet	5.84	22.07	16,327	6.79	8.33	1.02	0.61	5.64	168	54.03	11	60	28.00	0.99	6.00	8.33
29	3	Dry	7.70	20.60	67,100	7.15	5.33	1.16	0.51	5.05	138	65.60	7	46	14.67	0.78	6.43	8.33
		Wet	6.47	21.87	7800	6.91	6.33	1.00	0.33	5.06	134	66.63	10	47	25.67	0.56	4.75	6.00
33	4	Dry	8.13	19.75	9500	7.40	13.00	1.05	0.41	10.47	118	52.63	26	45	36.33	0.63	4.53	6.53
		Wet	5.76	23.06	9007	7.01	5.67	1.00	0.19	13.31	96	60.17	15	38	25.33	0.41	2.67	4.23
52	5	Dry	8.31	19.31	3834	7.22	7.00	3.97	0.42	6.32	135	68.90	10	33	21.67	0.65	1.72	7.53
		Wet	6.69	24.29	10,367	7.02	7.00	2.22	0.18	34.93	86	57.37	41	20	30.67	0.44	1.01	5.27
58	6	Dry	8.01	19.90	4467	7.93	3.33	2.92	0.16	10.77	84	62.30	8	21	12.33	0.31	0.53	4.83
		Wet	6.84	24.89	5100	6.95	7.67	1.72	0.09	62.47	73	56.80	48	18	29.00	0.21	0.41	2.70
70	7	Dry	7.52	20.84	714	7.39	3.67	3.42	0.16	11.82	91	72.10	10	20	13.33	0.28	0.19	5.07
		Wet	6.75	24.59	4490	7.06	6.33	2.07	0.07	145.20	65	58.27	69	17	29.67	0.24	0.11	3.23
79	8	Dry	7.72	20.41	270	7.39	3.33	3.23	0.15	9.44	90	73.27	7	22	11.33	0.26	0.08	3.70
		Wet	6.81	24.59	627	7.07	5.67	2.23	0.09	45.60	64	66.93	32	18	22.67	0.25	0.11	4.00
–	9	Dry	9.04	17.75	1113	7.33	4.00	3.25	0.14	6.54	105	68.53	8	25	14.33	0.33	1.32	6.57
		Wet	5.91	22.37	10,967	7.15	15.00	2.01	0.05	34.53	72	51.63	51	22	48.33	0.30	0.67	4.23
Mean (n = 54)			7.32	21.64	12,816	7.17	7.69	1.97	0.30	23.67	112	61.43	22	36	26.11	0.52	3.01	6.01
Standard deviation			1.61	2.79	37,759	0.60	7.52	1.12	0.24	49.47	47	11.02	30	21	21.65	0.33	3.13	2.53
Maximum			11.72	26.12	200,000	8.75	45.00	4.44	0.93	330.00	248	88.10	147	91	114.00	1.65	10.00	13.00
Minimum			3.40	15.59	1	6.11	2.00	1.00	0.03	3.58	43	33.70	1	10	10.00	0.15	0.06	2.00

Table 2. Average seasonal results in 2019. Distances refers to measures from the source to the mouth of the Piabanha River. Station 9 is located on the Paquequer/Preto river. Abbreviations, units, methods, quantification limits and permissible limits to rivers class 2 are described in Table 1. The entire dataset can be found online as Supplementary Table S2.

most parameters showing a strong correlation (r > 0.5) with a confidence interval greater than 95% (α = 0.05). The KMO measures of sampling adequacy (n = 8) were near to 0.5 and the significance level of test of sphericity was less than 0.001, indicating that the data was fit for PCA and the correlation matrix is not an identity matrix and so the variables are significantly related. The Shapiro test confirmed the data normality (p > 0.01) for all parameters, except for *E. coli*.
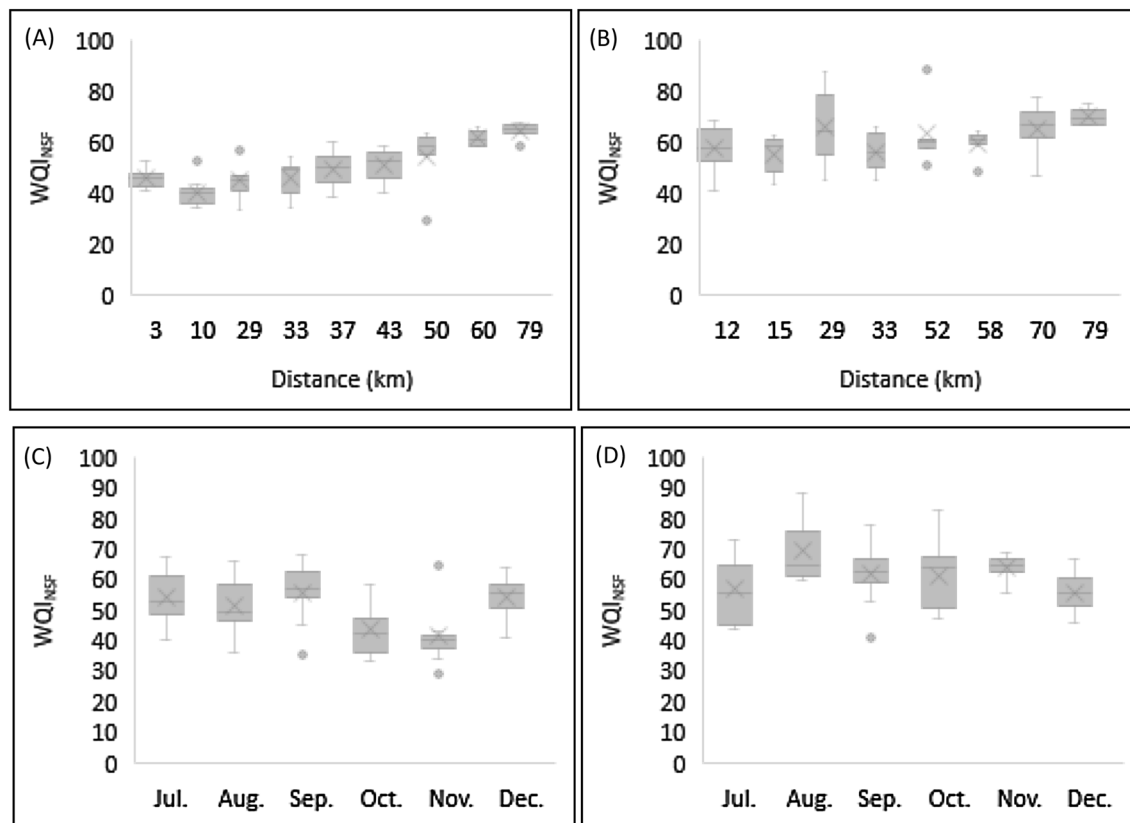
Figure 1. WQI$_{NSF}$ spatial variation over each station from July to December (**A**) 2012 and (**B**) 2019. WQI$_{NSF}$ seasonal variation over the entire length of the river (**C**) 2012 and (**D**) 2019. The entire dataset can be found online as Supplementary Table S1 and S2, respectively for 2012 and 2019.

r/p	DO	WT	*E. coli*	pH	BOD	NO$_3^-$	PO$_4^{3-}$	Turb	TDS	SS	Alcal	COD	TP	NH$_3$	TN
DO		0.93	0.22	0.62	0.26	0.89	0.95	0.78	0.64	0.83	0.87	0.27	0.94	0.90	0.98
WT	0.03		0.05	0.05	0.06	0.01	<0.01	<0.01	<0.01	<0.01	<0.01	0.10	<0.01	<0.01	0.01
E. coli	−0.49	−0.70		0.22	0.71	0.12	0.02	0.19	0.07	0.03	0.06	0.88	0.03	0.04	0.02
pH	−0.21	0.72	−0.49		0.14	0.21	0.04	0.15	0.01	0.11	0.04	0.18	0.04	0.03	0.03
BOD	0.45	−0.69	0.16	−0.57		0.06	0.05	0.13	0.02	0.31	0.01	<0.01	0.05	0.03	0.07
NO$_3^-$	0.06	0.82	−0.60	0.50	−0.68		0.04	0.07	0.03	0.04	<0.01	0.13	0.05	<0.01	0.09
PO$_4^{3-}$	−0.03	−0.86	0.77	−0.73	0.71	−0.72		0.05	<0.01	0.03	<0.01	0.09	0.00	<0.01	<0.01
Turb	−0.12	0.87	−0.52	0.55	−0.59	0.67	−0.70		0.04	<0.01	0.04	0.23	0.05	0.04	0.07
TDS	0.19	−0.87	0.67	−0.81	0.80	−0.75	0.97	−0.72		0.03	<0.01	0.04	<0.01	<0.01	<0.01
SS	0.09	0.85	−0.75	0.61	−0.41	0.74	−0.75	0.91	−0.74		0.03	0.52	0.04	0.02	0.05
Alcal	0.07	−0.87	0.68	−0.73	0.81	−0.85	0.96	−0.72	0.97	−0.76		0.04	<0.01	<0.01	<0.01
COD	0.44	−0.62	0.06	−0.53	0.98	−0.59	0.64	−0.48	0.72	−0.27	0.73		0.09	0.07	0.11
TP	−0.03	−0.86	0.77	−0.73	0.70	−0.70	1.00	−0.70	0.96	−0.73	0.95	0.64		<0.01	<0.01
NH$_3$	0.06	−0.89	0.73	−0.75	0.77	−0.87	0.96	−0.73	0.97	−0.80	0.99	0.67	0.94		<0.01
TN	−0.01	−0.83	0.77	−0.76	0.67	−0.64	0.99	−0.66	0.96	−0.71	0.93	0.60	0.99	0.92	

Table 3. Pearson correlation (r). Correlation values are given below the main diagonal of the matrix. The p-values (two-tailed probabilities that the variables are uncorrelated) are given above the main diagonal of the matrix.

	DO	WT	*E. coli*	pH	BOD	NO$_3^-$	PO$_4^{3-}$	Turb	TDS	SS	Alcal	COD	TP	NH$_3$	TN
PC 1	0.07	−0.94	0.72	−0.78	0.78	−0.83	0.97	−0.80	0.98	−0.82	0.98	0.70	0.96	0.99	0.94
PC 2	0.87	0.09	−0.66	−0.10	0.58	0.05	−0.09	0.04	0.10	0.33	0.04	0.64	−0.09	−0.02	−0.09

Table 4. PCA loadings, values greater than 0.50 or less than -0.50 are very significant.
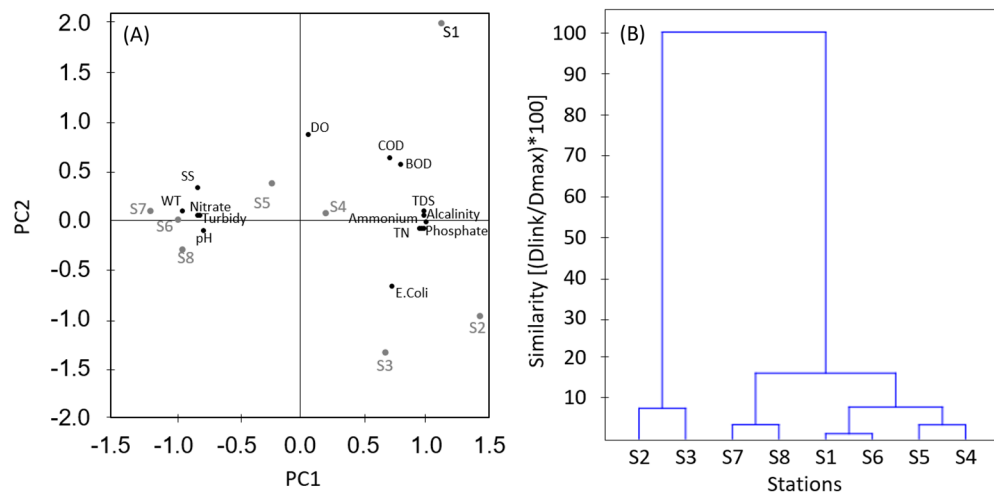
Figure 2. Multivariate techniques. (**A**) PCA plot with station scores and parameters loadings. (**B**) Hierarchical clustering by Ward linkage with Euclidean distance. The entire dataset can be found as Supplementary Table S2 online.

ACP was applied to identify groups of parameters that influence water quality. PC 1, PC 2 and PC3 account for 72% (eigenvalue 10.74), 14% (eigenvalue 13.94) and 5% (eigenvalue 0.8), respectively, of the data variance. Components with eigenvalues larger than the unit were selected. That is, the first two components together account for 86% of the total variance. The loadings that compose the first two components are presented in the Table 4 and the stations that most influence the results are represented in Fig. 2A.

PC1 was substantially correlated with practically all parameters. Stations number 1 to 4 loaded positively (loadings > 0.7) to PC1 with the parameters TDS, Alkalinity, Ammonia, Total Nitrogen, Phosphate, Total Phosphorus, DBO, COD, *E. coli*, while stations number 5 to 8 loaded negatively (loadings < − 0.7) with Nitrate, Turbidity, SS, pH and WT. PC2 was most influenced by stations in the urban area, notably station 1, and showed a positive correlation (loadings > 0.5) with OD, COD, BOD and less by SS (loading = 0.33), being more influenced by station 1 in the urban area. On the other hand, it was negatively correlated with *E. coli* (loading = − 0.66) with a large influence of station 3.

The sampling stations were grouped into three statistically significant clusters with 75% of similarity by agglomerative hierarchical clusterization based on the ward linkage by Euclidean distance (Fig. 2B): cluster 1 (Stations 2 and 3), cluster 2 (Stations 7 and 8) and cluster 3 (Stations 1, 4, 5 and 6).

Longtime monitoring assessment based on Mann–Kendall rank test and Fourier transform.
In a complementary way, in order to evaluate a possible trend on water quality and to detect the seasonal behavior of the basin, we used a time series with 40 years of monitoring. Since dissolved oxygen can be used as a surrogate variable for the general health of aquatic ecosystems[49–51], it was selected to perform the Mann–Kendall rank test of randomness for the station more upstream and further downstream of the Piabanha River, PB002 and PB011 respectively. The upstream station showed a statistically significant increasing trend (n = 166, S = 1507, Z = 2.10, p < 0.03), whereas the downstream station does not show a statistically significant trend (n = 198, S = 1179, Z = 1.27, p = 0.20). The entire dataset can be found as Supplementary Table S3 and S4.

To detect the seasonal behavior, we have applied a Fourier transform algorithm to the time series from 1980 to 2019 to the station PB011 (Fig. 3A, which does not display a tendency behavior and can be considered as representative of the entire basin because it is the most downstream station. The data were organized in quarterly averages for the DO parameter. The two most powerful signals correspond to the frequencies of 0.25 and 0.45, nearly (Fig. 3B) It corresponds to periods of 12 and 6 months, respectively. Taking into account this seasonality, we confirmed that our 2019 field campaigns are representative of seasonality comprising the final half of the dry season and the initial half of the rainy season.

Discussion
Water quality assessment.
The Piabanha River had a better water quality in 2019 than in 2012, according to WQI_NSF results (Fig. 1). The improvement was substantial over the first 40 km, rated as "bad" in most campaigns in 2012, while rated as medium in most campaigns in 2019 due to sewage collection and treatment system expansion. Since 2012, Petrópolis has built 50 km of sewage collection network and 7 new sewage treatment units[52]. These plants produce secondary level effluents through biological treatment, the plants flow capacity reaches about 800 L s⁻¹. These stations use different technologies such as: submerged aerated biofilters, anaerobic upflow reactor, moving bed biofilm reactor and upflow anaerobic sludge blanket reactor. Beside this, in some plants are used biosystems[53]. Water quality improved in stretches after 40 km due to self-purification processes and the contribution of clean tributaries. This is in line with findings from other rivers worldwide[31,54,55].
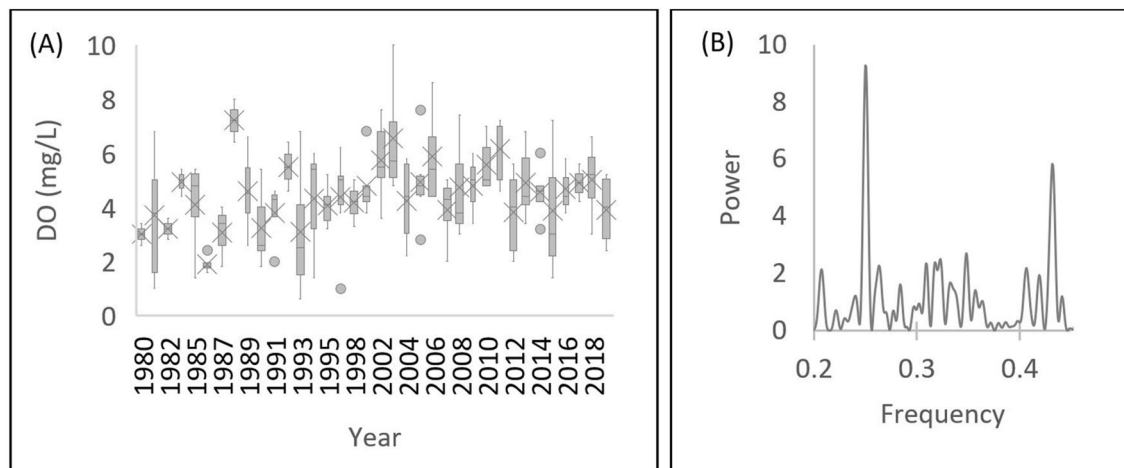
Figure 3. (**A**) Temporal distribution of dissolved oxygen from 1980 to 2019 at station PB002 (n = 160). (**B**) Periodogram. The entire dataset can be found in Supplementary Table S5.

Dry seasons, in general, presented better water quality indexes than rainy seasons. Other studies[28,56,57] have shown similar seasonal behavior, where water quality worsens in the rainy season due to sediments and pollutants input carried by the rain. In addition, most of the sewage network is the same network that collects rainwater. Thus, during rainy events, sewage is no longer treated and is discharged directly into rivers.

Although the WQI_{NSF} had a medium rating in 2019, BOD and Coliforms were substantially above the maximum allowed by Brazilian regulation. In addition, the index is limited to the parameters used in its calculation[58]. This is the case for the ammonium parameter, which presented concentrations up to three times higher than allowed in Brazilian regulation, reminding that only nitrate is used in the WQI_{NSF}. The same occurs with total phosphorus: only phosphate is considered, although it does not have a maximum value established by the Brazilian federal regulation. In what follows, we analyse these parameters in more detail.

Biochemical Oxygen Demand (BOD) is one of the most widely used criteria for water quality assessment. It provides information on the ready biodegradable fraction of the organic load in water[59]. High BOD concentrations reduce oxygen availability, mainly correlated to microbiological activity[60]. Its concentration ranged from 2.00 to 45 mg L^{-1} (average 7.69 ± 7.52) over the entire data, with its concentrations most of the time substantially above the maximum allowed by Brazilian regulation (5 mg L^{-1}). *Escherichia coli* is naturally present in the intestinal tracts of warm-blooded animals and it is widely used as an indicator of fecal contamination[61,62]. Villas-Boas[42] pointed to fecal coliforms as the most relevant water quality parameter in the urban area of Petrópolis, mainly related to pollution caused by untreated domestic sewage.

Phosphorus is an essential nutrient for all forms of life[63]. Its availability can be related to atmospheric deposition[64], anthropic uses of products such as detergents[65] and due to agricultural activities[66]. Orthophosphates are the most relevant in the aquatic environment as they are the main form of phosphate assimilated by aquatic vegetables[67]. Previous studies[42,68,69] in the Piabanha Basin found phosphate values in perfect agreement with ours. Alvim[68] points out that the main source of phosphorus for the Piabanha River is the sewage discharge and the higher concentrations are found during the rainy season.

Nitrate is a very common element in surface water since it is the end product of the aerobic decomposition of the organic nitrogenous compound[70,71]. Its sources are related to landscape composition, being influenced by both agricultural and urban uses[72]. Villas-Boas[42] found high concentration of nitrate and ammonium in the urban region of Piabanha River in agreement with this study. Alvim[68] reports that domestic sewage discharged into Piabanha River waters account for 43% of the nitrogen load, the atmospheric contribution for 31% and the farming activity for 15%.

The major contributors to water quality and stretches of river with similar water quality. The first two components together account for 86% of the total variance, indicating method high explanatory power of the method. It was far better than other similar studies around the world[29,30,71,73–75]. PC1 predominantly accounts for urban sewage pollution. This is clearly demonstrated by the fact that stations from 1 to 4, located in the urban area of Petrópolis, positively loaded PC1 with organic matter (BOD and COD), TDS and nutrients such as phosphorus and nitrogenous constituents, especially ammonia, indicating recent pollution. Even clearer is the fact that stations from 5 to 8 have negatively loaded with nitrate, showing the nitrogen compounds degradation in the downstream stretches of the urban area. On the other hand, the increase in nitrate concentrations in association with the increase in turbidity in stations outside the urban area may also be associated with land use, especially in agriculture.

PC2 is dominated by the dissolved oxygen parameter and other parameters that indicate the health of the river, as organic load and coliforms. It is explained by water pollution by organic matter and biological activity and reinforces the result of CP1. In the study region, sanitation is still a challenge to be faced by the government, especially in the first urban stretch, after 40 km from the source of the Piabanha River, this region has 26% of untreated sewage[53].

Cluster analysis was used to group sampling stations into similarity classes indicating the stretches of river with similar water quality. As pointed out by Singh[29], it implies that only one site in each cluster may serve as good in spatial assessment of the water quality as the whole cluster. So, the number of sampling sites can be reduced; hence, cost without losing any significance of the outcome. On the other hand, this interpretation should be done with caution since trends in different stretches can be very different, making future changes significant. Therefore, great care must be taken to reduce monitoring stations.

It is important to notice that the first cluster (S1, S6 and S4, S5) groups station 1 with station 6, the first one corresponding to the urban area of Petrópolis whose pollution stems from sewage and industrial effluents. Likewise, station 6 is located after the confluence of the Preto-Paquequer River, which crosses Teresópolis, the second largest city in the hydrographic basin, also with the presence of economic and industrial activities. Sand mining is the predominant activity near stations 4 and 5, which together receive the impact of five mining companies. Similarly, station 6, after the Preto River, receives the impact of seven sand mines. In fact, this group brings together economic activities whose impact on water quality is similar. Station 5 could be removed from the network monitoring in order to reduce costs.

The second cluster (S2 and S3) refers to the most urbanized section of the basin. When individually checking the quality parameters between these stations, one can conclude that they differ only by the diluting effect caused by the contribution of the Araras River, on the left bank, and of the Poço do Ferreira River, on the right bank, which receives its waters from the Bonfim River after its source in the Serra dos Órgãos National Park, an important federal conservation unit. Station 3 was introduced precisely to detect this diluting effect, but since the cluster analysis showed that it was not significant it is recommended to remove this station.

The third cluster (S7 and S8) has a very similar behavior: station 8 is just before the Piabanha River mouth and station 7 is located less than 10 km upstream of the mouth. In addition, on this stretch there are only three interferences registered as discharges. Thus, it is recommended to remove station 7, considering the importance of maintaining a station close to the river mouth.

Trend analysis and seasonal variation. Although it still presents systematic violations to Brazilian standards[76], the water quality, in general, has improved in the Piabanha River over the past 40 years (Fig. 3A,B). This statement is supported by the Mann–Kendall rank test of randomness, indicating a significant (p = 0.03) tendency to increase the values of the dissolved oxygen parameter at station PB002, located in the urban area of Petrópolis, which is highly impacted by effluent discharges, despite the fact that this region has municipal sewage treatment. PB011 presents high levels of DO, since the beginning of the time series exhibiting an almost monotonic behavior over time, thus it has no tendency. The high DO levels are due to both the river's reoxygenation process and the contribution of clean waters from its tributaries, such as the Fagundes River.

A strong annual and semi-annual seasonality was indicated by the power spectral density, which can be seen in the periodogram (Fig. 3B) resulting from the Fast Fourier Transform. The results are in accordance with the literature[77] indicating that more than 90% of the total variance of dissolved oxygen is accounted for by the annual periodicity and the next four higher harmonics (semi-annual; tri-annual, etc.). Seasonality follows the rainfall regime with a dry period from April to September, and a wet period from October to March, according to Araújo's[78] study carried out in the Piabanha River basin.

Water quality at point PB002 started to improve in 2000, when the first sewage treatment plant in the city of Petrópolis came into operation. Currently, 95% of the population has access to drinking water, and the coverage of treated urban sewage is 85%. The municipality has 26 sewage treatment units, responsible for the treatment of 56.2 million liters per day. In relation to the other municipalities in the basin, according to the National Sanitation Information System[79] (SNIS), the municipality of Três Rios treats 2.97% of its sewage, while the other municipalities, Teresópolis, Areal, São José do Vale do Rio Preto, Paty do Alferes and Paraíba do Sul did not report their data to SNIS, potentially indicating that they do not perform sewage treatment. In other words, about 50% of the population has no formal access to sewage treatment services.

Conclusion

The diagnosis provided by this research establishes the first step towards the Framing of water resources according to their intended uses, as established by the Brazilian National Water Resources Policy. In addition to the diagnosis which was carried out a georeferenced database was built. There are few cases of Framework in Brazil and none in the studied watershed. This makes this study relevant to Brazilian water resources management. The considerable number of users awaiting regularization from the State Environmental Institute is a limitation to implement the Framework and requires a joint effort of the watershed committee.

Answering our initial question, Piabanha River water quality is medium according to the WQI$_{NSF}$ and certainly is not able to support high levels of biodiversity. Some river stretches have quality compatible with class 4 according to the Brazilian regulation for the coliforms, BOD and TP parameters; hence, they cannot be used for irrigation, human or animal consumption, not even after treatment. On the other hand, the Framework must be carried out according to intended uses. Therefore, we recommend that the Piabanha Committee, in partnership with the State Public Ministry, lead actions to reduce the concentrations of these parameters, mainly in the sanitation sector.

It is recommended that the monitoring program be continued and expanded to stretches where conflicts between water uses occur, in order to implement the Framework to enforce the improvement of water quality. It is also important to point out that this study was financed with public resources from the Piabanha water resources fund and that the present analysis made possible to recommend the exclusion of three of the eight existing stations, thereby enabling the expansion of the monitoring to other tributaries of the Piabanha River under the influence of large population with practically no sanitation, notably the Rio Preto/Paquequer sub-basin.
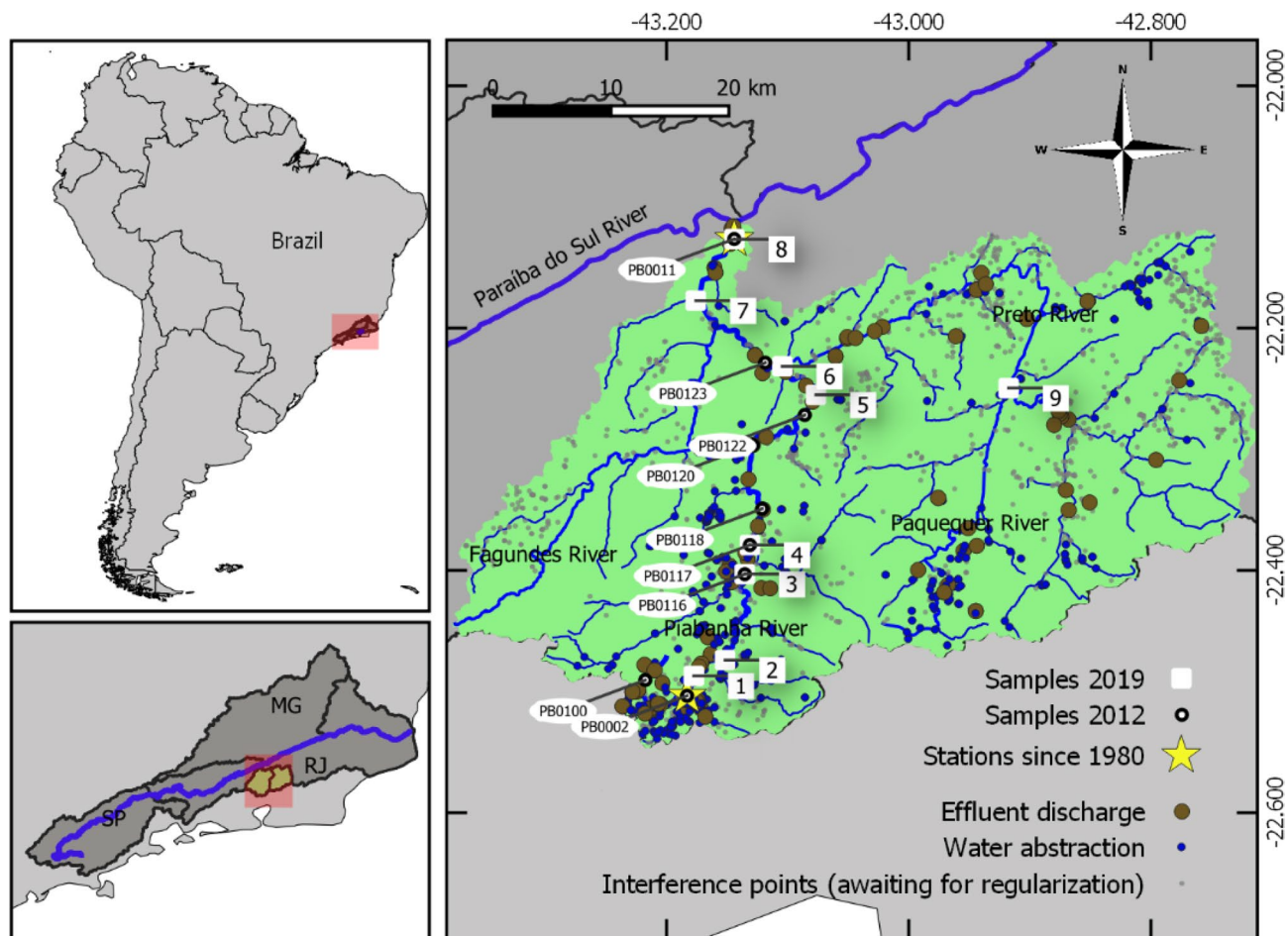
Figure 4. Study area, sample stations and interference points (water abstraction or effluxent discharge). This map was generated in the open source software QGIS version 3.14.15 (https://qgis.org/).

This work describes a methodological approach that can be useful for other researches in environmental science and management. We have applied an integrated approach using data from different sources combined with data analysis based on WQI, PCA, CA, frequency analysis and trend analysis, which were used in a complementary way to understand a research problem.

Materials and methods

Study area. The Piabanha Basin is located in southern Brazil, belonging to the mountainous region of the State of Rio de Janeiro with an area of 2050 km² (Fig. 4). The Piabanha River source is at 1150 m of altitude and runs down 80 km until it flows into the Paraíba do Sul River at an altitude of 260 m. The upper portion of the basin presents a humid tropical climate. With steep slopes, annual rainfall exceeds 2000 mm. The lower portion of the basin has a sub-humid climate and the average rainfall decreases to 1300 mm. The seasons are well defined throughout the basin and the rainfall regime has symmetry in its distribution between the periods from January to June and from July to December[78]. The territory is home to 535 thousand people in 2018[80]. The two largest cities in the region, Petrópolis and Teresópolis, are located in the headwaters of the basins and give rise to the Piabanha and Preto rivers, respectively. Additionally, because the sewage treatment is limited and the river flows are low, high constituent concentrations are observed (e.g., fecal coliform, nitrate, and BOD), especially in urban areas[42].

Datasets. Three sets of monitoring data have been used in this researchh (Fig. 4). The first and main one was the result of a monitoring program that is being conducted by the Piabanha watershed Committee, in which data from July to December 2019 have been analysed and are described in more details in the next item. The second were from 6 campaigns carried out in 2012 by HIDROECO project[44] also with financial resources from the Piabanha Committee which is used as a baseline for comparison purposes. The third was comprised of two stations of the basic monitoring network of the Rio de Janeiro Environmental Institute, with data from 1980 to the present, except for periods of data gaps.

A georeferenced database was also built containing water management data. Brazilian National Water Agency (ANA) has developed the National Water Resources Users Register (CNARH) for any bulk water user that

changes regime, quantity or quality of a water body. It is a federal platform, but it can be managed by each state. Registration is a prerequisite for the other stages of uses regularization.

Monitoring campaigns and analytical procedures.   Physical–chemical parameters were measured in situ using a multiparameter probe (YSI model 556) and a portable turbidimeter (HANNA model HI 98703-0), both previously calibrated and later verified. The samples were placed in specific containers for each analysis, for the necessary parameters the samples were preserved with H_2SO_4 and kept at a temperature below 4 °C. Laboratory analyses (Table 1) were performed according to Standard Methods for the Examination of Water and Wastewater (SMWW)[81]. The laboratory has an accreditation certificate issued by the State Environmental Agency (INEA CCL No. IN044710) and also complies to ISO/IEC 17025 (CRL 1035).

Water Quality Index.   A Water Quality Index (WQI) is an empirical expression which integrates significant physical, chemical and microbiological parameters of water quality into a single number[82]. It can be a powerful communication tool to simplify a complex set of parameters, whose individual interpretation can be difficult, into a single index representing the general water quality. A water quality index was initially proposed by Horton[26] and further developed by Brown[27,83] resulting in the National (USA) Sanitation Foundation Water Quality Index (WQI$_{NSF}$).

The original version of the WQI$_{NSF}$ established an additive expression[27]; on the other hand, field data analysis suggested that the additive WQI lacked sensitivity in adequately reflecting the effect of a single low value parameter on the overall water quality. As a result, a multiplicative form of WQI was proposed[82,83]:

$$WQI_{NSF} = \prod_{i=1}^{n} q_i^{w_i}$$

q_i is the quality class for the nth variable, a number between 0 and 100, obtained from the respective average quality variation curve[82], depending on the concentration of each nth variable. W_i is the relative weight for the nth variable, number between 0 and 1, assigned according to the importance of the variable for overall quality conformation. WQI$_{NSF}$ is the National Sanitation Foundation Water Quality Index, a number between 0 and 100, rated as "excellent" ($100 > WQI \geq 90$), "good," ($90 > WQI \geq 70$), "medium" ($70 > WQI \geq 50$), "bad" ($50 > WQI \geq 25$) or "very bad" ($25 > WQI \geq 0$).

The WQI$_{NSF}$ and its many adaptations have been widely used[84,85], however, its use is not uniform, replacing parameters without the necessary adaptation of the respective curve of the indicator. In Brazil, since 1975 the WQI$_{NSF}$ has been used by CETESB (Environmental Company of the State of São Paulo). In the following decades, other Brazilian states adopted, with minor adaptations, this index, which today is the most widely used in the country. In the present study, the weights (w_i) have been used according to the methodology established by INEA (Environmental Institute of the State of Rio de Janeiro): DO (0.17); Fecal coliforms (0.16); pH and BOD (0.11); Nitrates, Phosphate and Temperature (0.10); Turbidity (0.08) and TDS (0.07), rather than total solids.

The replacement of the total solids for dissolved solids parameter may cause an average variation of 0.2% in the final result of WQI$_{NSF}$, based on our estimates (n = 48, data 2019). In relation to microbiology, *E. coli* have been used instead of fecal coliforms, applying a correction factor[86] of 1.25 on the result of *E. coli*.

Principal component analysis and cluster analysis.   Principal component analysis (PCA), as defined by Hotelling[87], is a multivariate technique of covariance modeling that reduces the dimensionality of an originally correlated dataset, with the lowest possible information loss. A new set of variables containing new orthogonal, uncorrelated variables, is formed from a dataset of correlated variables, which are weighed linear combinations of the original variables[30].

PCA technique extracts the eigenvalues and eigenvectors from the covariance matrix of original variables. The PCs are obtained by multiplying the original correlated variables with the eigenvector, which is a list of coefficients, frequently called "loadings"[29,30,88,89]. A widely accepted and simple qualitative rule proposes that loadings greater than 0.30 or less than −0.30 are significant; loadings greater than 0.40 or less than −0.40 are more important, whereas loadings greater than 0.50 or less than −0.50 are very significant[90]. The suitability of data for PCA was evaluated by Kaiser–Meyer–Olkin[91,92] (KMO) measuring of sampling adequacy and Bartlett tests of sphericity[93]. The Shapiro test was evaluated to verify the data normality ($\alpha = 0.01$).

Cluster analysis reveals the latent behavior of a dataset to categorize the objects into groups or clusters on the basis of similarities[30,88,89]. Hierarchical agglomerative cluster analysis (CA) classifies objects by first putting each object in a separate cluster, and then joins the clusters together stepwise until a single cluster remains[29].

Timeseries analysis and trend detection.   Mann–Kendall trend test is a nonparametric test used to identify a trend in a series, first proposed by Mann[94] and further improved by Kendall[95] and Hirsch[96]. The null hypothesis (H_0) for these tests is that there is no trend in the series. The tests are based on the calculation of Kendall's tau measure of association between two samples, which is itself based on the ranks with the samples. The variables are ranked in pairs, and the difference of each variable to its antecessor is calculated. The total number of pairs that present negative differences is subtracted from the number of pairs with positive differences (S). A positive value of S indicates an upward trend, and a negative value of S a downward trend. For n > 10, a normal approximation is used to calculate Z statistic which is used to calculate p-value[96].

Fourier decomposition is a technique which allows the separation of frequency components from a data series with seasonal behavior from a complex water quality dataset[97]. Spectral analysis performed using a Fast Fourier Transform (FFT) algorithm is widely used in environmental studies, because it reveals the dominant

influences and their scales[50]. Power spectral density (PSD) obtained from FFT and represented by periodograms is a recommended procedure to detect seasonality[98,99].

Brazilian legal regulation. Brazilian fresh waters are divided into four classes, depending on the intended use[76]. The Special Class is intended mainly for the preservation of the natural balance of aquatic communities in fully protected conservation areas. Class 1 is designed for human consumption supply, after simplified treatment, for the protection of aquatic communities and for primary contact recreation. Class 2 requires conventional treatment for human consumption. Class 3 requires conventional or advanced treatment for human consumption and can be used to feed animals and irrigate some crops. Class 4 is intended only for navigation and landscape harmony. It is important to note that the Framework refers to the required water quality target according to water uses. The river basin committees are responsible for implementing the Framework, in accordance with the Brazilian National Water Resources Policy[33]. As long as the Framework is not established by the basin committee, fresh waters will be considered class 2 (Art. 42 CONAMA 357/2005)[76].

Data availability
All data generated or analysed during this study are included in this published article and its Supplementary Information files.

References
1. Martin, V. M. & Joel, A. T. *History of the Urban Environment* (University of Pittsburgh Press, Pittsburgh, 2012).
2. Wang, J., Liu, X. D. & Lu, J. Urban river pollution control and remediation. *Procedia Environ. Sci.* **13**, 1856–1862 (2012).
3. Zhang, X., Wu, Y. & Gu, B. Urban rivers as hotspots of regional nitrogen pollution. *Environ. Pollut.* **205**, 139–144 (2015).
4. Harding, L. W. *et al.* Long-term trends, current status, and transitions of water quality in Chesapeake Bay. *Sci. Rep.* **9**, 1–19 (2019).
5. John, V., Jain, P., Rahate, M. & Labhasetwar, P. Assessment of deterioration in water quality from source to household storage in semi-urban settings of developing countries. *Environ. Monit. Assess.* **186**, 725–734 (2014).
6. Mishra, B. K. *et al.* Assessment of Bagmati river pollution in Kathmandu Valley: scenario-based modeling and analysis for sustainable urban development. *Sustain. Water Qual. Ecol.* https://doi.org/10.1016/j.swaqe.2017.06.001 (2017).
7. Xu, Z. *et al.* Urban river pollution control in developing countries. *Nat. Sustain.* **2**, 158–160 (2019).
8. UN-Water. Sustainable Development Goal 6 Synthesis Report on Water and Sanitation 2018. *Un* (2018). https://doi.org/10.1126/science.278.5339.827.
9. UNEP. *A Snapshot of the World's Water Quality: Towards a global assessment.* (United Nations Environment Programme, 2016).
10. Wada, Y. *et al.* Modeling global water use for the 21st century: the Water Futures and Solutions (WFaS) initiative and its approaches. Geosci. Model Dev. https://doi.org/10.5194/gmd-9-175-2016 (2016).
11. WWAP. *The United Nations World Water Development Report 2019: Leaving No One Behind* (2019).
12. Fan, M. & Shibata, H. Simulation of watershed hydrology and stream water quality under land use and climate change scenarios in Teshio River watershed, northern Japan. *Ecol. Indic.* https://doi.org/10.1016/j.ecolind.2014.11.003 (2015).
13. Putro, B., Kjeldsen, T. R., Hutchins, M. G. & Miller, J. An empirical investigation of climate and land-use effects on water quantity and quality in two urbanising catchments in the southern United Kingdom. *Sci. Total Environ.* https://doi.org/10.1016/j.scitotenv.2015.12.132 (2016).
14. Li, B., Rodell, M., Sheffield, J., Wood, E. & Sutanudjaja, E. Long-term, non-anthropogenic groundwater storage changes simulated by three global-scale hydrological models. *Sci. Rep.* **9**, 10746 (2019).
15. Jaeger, W. K. *et al.* Scope and limitations of drought management within complex human–natural systems. *Nat. Sustain.* https://doi.org/10.1038/s41893-019-0326-y (2019).
16. Pastor, A. V. *et al.* The global nexus of food–trade–water sustaining environmental flows by 2050. *Nat. Sustain.* https://doi.org/10.1038/s41893-019-0287-1 (2019).
17. Melo, D. C. D. *et al.* The big picture of field hydrology studies in Brazil. *Hydrol. Sci. J.* **65**, 1262–1280 (2020).
18. Dixon, W. & Chiswell, B. Review of aquatic monitoring program design. *Water Res.* **30**, 1935–1948 (1996).
19. Wang, Y., Xiang, C., Zhao, P., Mao, G. & Du, H. A bibliometric analysis for the research on river water quality assessment and simulation during 2000–2014. *Scientometrics* **108**, 1333–1346 (2016).
20. Ji, X., Dahlgren, R. A. & Zhang, M. Comparison of seven water quality assessment methods for the characterization and management of highly impaired river systems. *Environ. Monit. Assess.* **188**, 15 (2016).
21. Deng, W. & Wang, G. A novel water quality data analysis framework based on time-series data mining. *J. Environ. Manag.* **196**, 365–375 (2017).
22. Singh, S. *et al.* Development of indices for surface and ground water quality assessment and characterization for Indian conditions. *Environ. Monit. Assess.* **191**, 182 (2019).
23. Mladenović-Ranisavljević, I. I., Takić, L. & Nikolić, D. Water quality assessment based on combined multi-criteria decision-making method with index method. *Water Resour. Manag.* **32**, 2261–2276 (2018).
24. Chen, S. K., Jang, C. S. & Chou, C. Y. Assessment of spatiotemporal variations in river water quality for sustainable environmental and recreational management in the highly urbanized Danshui River basin. *Environ. Monit. Assess.* **191**, 100 (2019).
25. Rakotondrabe, F. *et al.* Water quality assessment in the Bétaré-Oya gold mining area (East-Cameroon): multivariate statistical analysis approach. *Sci. Total Environ.* **610–611**, 831–844 (2018).
26. Horton, R. K. An Index Number System for Rating Water Quality. *J. Water Pollut. Control Fed.* (1965).
27. Brown, R. M., McClelland, N. I., Deininger, R. A. & Tozer, R. G. A water quality index—do we dare?. *Water Sew. Work* **117**, 339–343 (1970).
28. Wu, Z. *et al.* Water quality assessment based on the water quality index method in Lake Poyang: the largest freshwater lake in China. *Sci. Rep.* **7**, 1–10 (2017).
29. Singh, K. P., Malik, A., Mohan, D. & Sinha, S. Multivariate statistical techniques for the evaluation of spatial and temporal variations in water quality of Gomti River (India)—a case study. *Water Res.* **38**, 3980–3992 (2004).
30. Dutta, S., Dwivedi, A. & Suresh Kumar, M. Use of water quality index and multivariate statistical techniques for the assessment of spatial variations in water quality of a small river. *Environ. Monit. Assess.* **190**, 718 (2018).
31. Malsy, M., Flörke, M. & Borchardt, D. What drives the water quality changes in the Selenga Basin: climate change or socio-economic development?. *Reg. Environ. Change* **17**, 1977–1989 (2017).

32. Pacheco, F. S. *et al.* Water quality longitudinal profile of the Paraíba do Sul River, Brazil during an extreme drought event. *Limnol. Oceanogr.* **62**, S131–S146 (2017).
33. Brazilian National Congress. *Brazilian National Water Resources Policy. Federal Law n. 9433.* (1997).
34. ANA. *Brazilian Water Resources Report—2017* (National Water Agency (Brazil), 2018).
35. Myers, N., Mittermeier, R. A., Mittermeier, C. G., da Fonseca, G. A. B. & Kent, J. Biodiversity hotspots for conservation priorities. *Nature* **403**, 853–858 (2000).
36. Russo, G. Biodiversity: biodiversity's bright spot. *Nature* **462**, 266–269 (2009).
37. Ribeiro, M. C., Metzger, J. P., Martensen, A. C., Ponzoni, F. J. & Hirota, M. M. The Brazilian Atlantic Forest: How much is left, and how is the remaining forest distributed? Implications for conservation. *Biol. Conserv.* https://doi.org/10.1016/j.biocon.2009.02.021 (2009).
38. Tabarelli, M., Aguiar, A. V., Ribeiro, M. C., Metzger, J. P. & Peres, C. A. Prospects for biodiversity conservation in the Atlantic Forest: lessons from aging human-modified landscapes. *Biol. Conserv.* **143**, 2328–2340 (2010).
39. Bogoni, J. A., Pires, J. S. R., Graipel, M. E., Peroni, N. & Peres, C. A. Wish you were here: how defaunated is the Atlantic Forest biome of its medium- to large-bodied mammal fauna?. *PLoS ONE* **13**, e0204515 (2018).
40. Rezende, C. L. *et al.* From hotspot to hopespot: an opportunity for the Brazilian Atlantic Forest. *Perspect. Ecol. Conserv.* **16**, 208–214 (2018).
41. CEIVAP & PROFILL. *Plano de Bacia: Consolidação do diagnóstico.* http://54.94.199.16:8080/publicacoesArquivos/ceivap/arq_pubMidia_Processo_030-2018_AGVP_PS_PIRH-Atualizacao_TOMO_I_R03_2.pdf (2018).
42. Villas-Boas, M. D., Olivera, F. & de Azevedo, J. P. S. Assessment of the water quality monitoring network of the Piabanha River experimental watersheds in Rio de Janeiro, Brazil, using autoassociative neural networks. *Environ. Monit. Assess.* https://doi.org/10.1007/s10661-017-6134-9 (2017).
43. Morais, A., Villas-Boas, M., Bastos, A., Monteiro, A. & Araújo, L. Estudos para um diagnóstico quali-quantitativo em bacias experimentais – Estudo de Caso: Bacia do rio Piabanha. In *II Seminário de Recursos Hídricos da Bacia Hidrográfica do Paraíba do Sul: Recuperação de Áreas Degradadas, Serviços Ambientais e Sustentabilidade* 173–180 (2009). https://doi.org/10.4136/serhidro.23.
44. Azevedo, J. P. S. de. *Relatório final do projeto HIDROECO/Piabanha: Metodologia para Determinação de Vazões Ambientais na Região Serrana do RJ Integrando Aspectos Hidrometeorológicos, Ecológicos e Socioeconômico. Volume 1: Informações Quali-quanti-tativas* (2017).
45. de Mello, F. V. *et al.* Current state of contamination by persistent organic pollutants and trace elements on Piabanha River Basin—Rio de Janeiro, Brazil. *Orbital Electron. J. Chem.* **10**, 327–336 (2018).
46. Chiappori, D., Hora, M. & Azevedo, J. Interface between hydropower generation and other water uses in the Piabanha River Basin in Brazil. *Br. J. Appl. Sci. Technol.* https://doi.org/10.9734/bjast/2016/23935 (2016).
47. da Silva, P. V. R. M., Pecly, J. O. G. & de Azevedo, J. P. S. Uso de traçadores fluorescentes para determinar características de trans-porte e dispersão no Rio Piabanha (RJ) para a modelagem quali-quantitativa pelo HEC-RAS. *Eng. Sanit. Ambient.* https://doi.org/10.1590/s1413-41522017150187 (2017).
48. de Costa, D. A., dos Assumpção, R. S. F. V., de Azevedo, J. P. S. & dos Santos, M. A. On water resources management instruments—Framing—as a tool for river rehabilitation. *Saúde em Debate* **43**, 35–50 (2019).
49. Abdul-Aziz, O. I., Wilson, B. N. & Gulliver, J. S. An extended stochastic harmonic analysis algorithm: application for dissolved oxygen. *Water Resour. Res.* **43**, W08417 (2007).
50. Rajwa-Kuligiewicz, A., Bialik, R. J. & Rowiński, P. M. Dissolved oxygen and water temperature dynamics in lowland rivers over various timescales. *J. Hydrol. Hydromechanics* **63**, 353–363 (2015).
51. United States Environmental Protection Agency. *Quality criteria for water.* https://nepis.epa.gov/Exe/ZyPDF.cgi/00001MGA.PDF?Dockey=00001MGA.PDF (1986).
52. Imperador, Á. do. Our history. https://www.grupoaguasdobrasil.com.br/aguas-imperador/en/ (2020).
53. ANA. *Atlas Esgotos—Despoluição de bacias hidrográficas.* (Brazilian National Water Agency, 2017).
54. Karthe, D., Lin, P.-Y. & Westphal, K. Instream coliform gradients in the Holtemme, a small headwater stream in the Elbe River Basin, Northern Germany. *Front. Earth Sci.* **11**, 544–553 (2017).
55. von Sperling, M. & von Sperling, E. Challenges for bathing in rivers in terms of compliance with coliform standards. Case study in a large urbanized basin (das Velhas River, Brazil). *Water Sci. Technol.* **67**, 2534–2542 (2013).
56. Bae, H. Changes of river's water quality responded to rainfall events. *Environ. Ecol. Res.* **1**, 21–25 (2013).
57. Yu, S., Xu, Z., Wu, W. & Zuo, D. Effect of land use types on stream water quality under seasonal variation and topographic char-acteristics in the Wei River basin, China. *Ecol. Indic.* **60**, 202–212 (2016).
58. Lumb, A., Sharma, T. C. & Bibeault, J.-F. A review of genesis and evolution of Water Quality Index (WQI) and some future direc-tions. *Water Qual. Exposure Health* **3**, 11–24 (2011).
59. Jouanneau, S. *et al.* Methods for assessing biochemical oxygen demand (BOD): a review. *Water Res.* **49**, 62–82 (2014).
60. Vigiak, O. *et al.* Predicting biochemical oxygen demand in European freshwater bodies. *Sci. Total Environ.* **666**, 1089–1105 (2019).
61. Ishii, S. & Sadowsky, M. J. Escherichiacoli in the environment: implications for water quality and human health. *Microbes Environ.* **23**, 101–108 (2008).
62. Odonkor, S. T. & Ampofo, J. K. Escherichia coli as an indicator of bacteriological quality of water: an overview. *Microbiol. Res. (Pavia)* **4**, 2 (2013).
63. Schlesinger, W. H. & Bernhardt, E. S. *Biogeochemistry. Biogeochemistry: an analysis of global change* 3rd edn. (Elsevier, Amsterdam , 2013). https://doi.org/10.1016/C2010-0-66291-2.
64. Tipping, E. *et al.* Atmospheric deposition of phosphorus to land and freshwater. *Environ. Sci. Process. Impacts* https://doi.org/10.1039/c3em00641g (2014).
65. Withers, P. J. A. & Jarvie, H. P. Delivery and cycling of phosphorus in rivers: a review. *Sci. Total Environ.* https://doi.org/10.1016/j.scitotenv.2008.08.002 (2008).
66. Sharpley, A. Agricultural phosphorus, water quality, and poultry production: are they compatible?. *Poult. Sci.* https://doi.org/10.1093/ps/78.5.660 (1999).
67. House, W. A. & Denison, F. H. Exchange of inorganic phosphate between river waters and bed-sediments. *Environ. Sci. Technol.* https://doi.org/10.1021/es020039z (2002).
68. Alvim, B. R. Dinâmica do nitrogênio e fósforo em águas fluviais de uma bacia hidrográfica com diferentes usos do solo no Sudeste do Brasil (Universidade Federal Fluminense, 2016).
69. Molinari, B. S. Modelagem espacial da qualidade de água na bacia do rio Piabanha/RJ (Universidade Federal do Rio de Janeiro, 2015).
70. Jaji, M. O., Bamgbose, O., Odukoya, O. O. & Arowolo, T. A. Water quality assessment of Ogun river, South West Nigeria. *Environ. Monit. Assess.* **133**, 473–482 (2007).
71. Mitra, S. *et al.* Water quality assessment of the ecologically stressed Hooghly River Estuary, India: a multivariate approach. *Mar. Pollut. Bull.* **126**, 592–599 (2018).
72. Guo, H. Y., Wang, X. R. & Zhu, J. G. Quantification and index of non-point source pollution in Taihu Lake region with GIS. *Environ. Geochem. Health* https://doi.org/10.1023/B:EGAH.0000039577.67508.76 (2004).
73. Khuhawar, M. Y., Zaman Brohi, R. O., Jahangir, T. M. & Lanjwani, M. F. Water quality assessment of Ramser site, Indus Delta, Sindh, Pakistan. *Environ. Monit. Assess.* **190**, 492 (2018).

74. Alves, R. I. S. *et al.* Water quality assessment of the Pardo River Basin, Brazil: a multivariate approach using limnological parameters, metal concentrations and indicator bacteria. *Arch. Environ. Contam. Toxicol.* **75**, 199–212 (2018).
75. Liang, B. *et al.* Distribution, sources, and water quality assessment of dissolved heavy metals in the Jiulongjiang River water, Southeast China. *Int. J. Environ. Res. Public Health* **15**, 2752 (2018).
76. Brazil. *Brazilian National Environment Council* (CONAMA) *Resolution n. 357. Provides the classification of water bodies and environmental guidelines for their framework, as well as establishes the conditions and standards for effluents discharge* 1–27 (2005).
77. Thomann, R. V. Time-series analyses of water-quality data. *J. Sanit. Eng. Div.* **93**, 1–24 (1967).
78. Araújo, L. M. N. de. Identification of precipitation and soil moisture hydrological patterns at Piabanha river basin. Ph.D. thesis. (Federal University of Rio de Janiero, 2016).
79. Brazil. Brazilian National Sanitation Information System (SNIS). http://www.snis.gov.br/ (2020).
80. CEIVAP/PROFILL. *Integrated plan for water resources in the watershed of the Paraíba do Sul river.* http://sigaceivap.org.br:8080/publicacoesArquivos/ceivap/arq_pubMidia_AGVP_PS_PIRH_PP-06_REV03_FINAL.pdf (2020).
81. American Public Health Association. *Standard Method for Examination of Water and Wastewater.* (American Public Health Association, 2012).
82. McClelland, N. I. *Water Quality Index Application in the Kansas River Basin* (1974).
83. Brown, R.M., McClelland, N.I., Deininger, R.A., Landwehr, J. M. *Validating the WQI. National meeting of American Society of Civil Engineers on water resources engineering* (1973).
84. Noori, R., Berndtsson, R., Hosseinzadeh, M., Adamowski, J. F. & Abyaneh, M. R. A critical review on the application of the National Sanitation Foundation Water Quality Index. *Environ. Pollut.* **244**, 575–587 (2019).
85. Kachroud, M., Trolard, F., Kefi, M., Jebari, S. & Bourrié, G. Water quality indices: challenges and application limits in the literature. *Water (Switzerland)* **11**, 1–26 (2019).
86. CETESB, Companhia Ambiental do Estado de São Paulo. *Qualidade das Águas Interiores no Estado de São Paulo - Apêndice D—Índices de Qualidade das Águas.* https://cetesb.sp.gov.br/aguas-interiores/publicacoes-e-relatorios/ (2019).
87. Hotelling, H. Analysis of a complex of statistical variables into principal components. *J. Educ. Psychol.* **24**, 417–441 (1933).
88. Helena, B. *et al.* Temporal evolution of groundwater composition in an alluvial aquifer (Pisuerga River, Spain) by principal component analysis. *Water Res.* https://doi.org/10.1016/S0043-1354(99)00225-0 (2000).
89. Vega, M., Pardo, R., Barrado, E. & Debán, L. Assessment of seasonal and polluting effects on the quality of river water by exploratory data analysis. *Water Res.* **32**, 3581–3592 (1998).
90. Sergeant, C. J., Starkey, E. N., Bartz, K. K., Wilson, M. H. & Mueter, F. J. A practitioner's guide for exploring water quality patterns using principal components analysis and procrustes. *Environ. Monit. Assess.* **188**, 249 (2016).
91. Cerny, B. A. & Kaiser, H. F. A study of a measure of sampling adequacy for factor-analytic correlation matrices. *Multivariate Behav. Res.* https://doi.org/10.1207/s15327906mbr1201_3 (1977).
92. Kaiser, H. F. An index of factorial simplicity. *Psychometrika* https://doi.org/10.1007/BF02291575 (1974).
93. Arsham, H. & Lovric, M. Bartlett's Test. In *International encyclopedia of statistical science* (ed. Lovric, M.) 87–88 (Springer, Berlin, 2011). https://doi.org/10.1007/978-3-642-04898-2_132.
94. Mann, H. B. Nonparametric tests against trend. *Econometrica* **13**, 245 (1945).
95. Kendall, M. G. *Rank correlation methods* (Oxford University Press, Oxford, 1975).
96. Hirsch, R. M., Slack, J. R. & Smith, R. A. Techniques of trend analysis for monthly water quality data. *Water Resour. Res.* **18**, 107–121 (1982).
97. Whitfield, P. H. Identification and characterization of transient water quality events by Fourier analysis. *Environ. Int.* **21**, 571–575 (1995).
98. Harris, J., Loftis, J. C. & Montgomery, R. H. Statistical methods for characterizing ground-water quality. *Ground Water* **25**, 185–193 (1987).
99. Hipel, K. W. & McLeod, A. I. Time series modelling of water resources and environmental systems. *Time Ser. Model. Water Resour. Environ. Syst.* https://doi.org/10.1016/0022-1694(95)90010-1 (1994).

Author contributions
D.A.C. compiled the manuscript, performed the analysis, and generated the figures and tables in the main text. J.P.S.A. contributed to the discussions and carefully reviewed the manuscript. M.A.S., R.S.F.V.A. and J.P.S.A. made substantial contributions to the conception and design of the research. All the authors reviewed the manuscript.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-020-78563-0.

Correspondence and requests for materials should be addressed to D.A.C. or J.P.S.d.

Reprints and permissions information is available at www.nature.com/reprints.

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