



Spatial optimization of the water quality monitoring network in São Paulo State (Brazil) to improve sampling efficiency and reduce bias in a developing sub-tropical region

Ricardo Gabriel Bandeira de Almeida¹ · Marta Condé Lamparelli² · Walter Kennedy Dodds³ · Davi Gasparini Fernandes Cunha¹

Received: 18 May 2021 / Accepted: 31 August 2021 / Published online: 18 September 2021
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2021

Abstract

Water quality monitoring networks (WQMN) are essential to provide good data for management decisions. Nevertheless, some WQMN may not appropriately reflect the conditions of the water bodies and their temporal/spatial dimensions, more particularly in developing countries. Also, some WQMN may use more resources to attain management goals than necessary and can be improved. Here we analyzed the São Paulo State (Brazil) WQMN design in order to evaluate and increase its spatial representativeness based on cluster analysis and stratified sampling strategy focused on clear monitoring goals. We selected water resources management units (UGRHs) representative of contrasting land uses in the state, with bimonthly data from 2004 to 2018 in 160 river/stream sites. Cluster analysis indicated monitoring site redundancy above 20% in most of the UGRHs. We identified heterogeneous spatial strata based on land use, hydrological, and geological features through a stratified sampling strategy. We identified that monitoring sites overrepresented more impacted areas. Thus, the network is biased against determination of baseline conditions and towards highly modified aquatic systems. Our proposed spatial strategy suggested the reduction of the number of sites up to 12% in the UGRHs with the highest population densities, while others would need expansions based on their environmental heterogeneity. The final densities ranged from 1.6 to 13.4 sites/1,000km². Our results illustrate a successful approach to be considered in the São Paulo WQMN strategy, as well as providing a methodology that can be broadly applied in other developing countries.

Keywords Developing countries · Monitoring efficiency · Rivers and streams · Spatial optimization · Water pollution · Water quality surveillance

Introduction

Water quality monitoring networks are essential to understanding and managing freshwater systems in the face of global anthropogenic impacts. While researchers and managers have monitored local water quality in some areas since the late 1800s (Worrall et al. 2015), the first formally designed water quality networks were established only in the late 1960s (Harmancioglu et al. 1998; Strobl and Robillard 2008). These networks should allow the detection of both spatial and temporal trends (Tavakol et al. 2017; Calazans et al. 2018; Peña-Guzmán et al. 2019), and possibly the determination of reference conditions to control and regulate human activities (Strobl and Robillard 2008; Cunha et al. 2011). Such networks can provide important information for establishing and evaluating policies to improve water quality, as well as linking costs and benefits of ecosystem services to specific management

Responsible Editor: Xianliang Yi

✉ Ricardo Gabriel Bandeira de Almeida
rgb Almeida@gmail.com

- ¹ Departamento de Hidráulica e Saneamento, Escola de Engenharia de São Carlos, Universidade de São Paulo, Avenida Trabalhador São-Carlense, 400. Centro, São Carlos, SP CEP 13566-590, Brazil
- ² Companhia Ambiental do Estado de São Paulo, Avenida Professor Frederico Hermann Júnior, 345. Alto de Pinheiros, São Paulo, SP CEP 05459-900, Brazil
- ³ Division of Biology, Kansas State University, 116 Ackert Hall, Manhattan, KS, USA

actions (Nel et al. 2009; Pynegar et al. 2018). Such networks have the general objectives of 1) accurately monitoring water quality of a specified region, 2) identifying problematic areas, and 3) accomplishing these first two goals as efficiently as possible.

Surface freshwaters (e.g., rivers and streams) provide resources for human activities, with specific water quality requirements associated with different uses (e.g., irrigation, navigation, energy generation, and water supply). The natural concentrations of solutes in the aquatic systems are influenced by their transport, cycling, and retention caused by physical, chemical, and biological processes as they move through watershed (e.g., atmospheric deposition, surface runoff, weathering of rocks, biological uptake, sorption, and desorption) (Meybeck and Helmer 1989; Stream Solute Workshop 1990). Intensification of anthropogenic activities can cause rapid land use shifts and create significant impacts on ecological processes (Meybeck 2003). Watersheds integrate large terrestrial areas, perturbations are transmitted downstream, and many rivers are dammed. Thus, water quality characterization is complex in drainage networks (Bostanmaneshrad et al. 2018; Rodrigues et al. 2018). Consequently, representative and well-structured monitoring programs are necessary to support water resources management.

Monitoring strategies in many countries were historically dependent on logistical aspects and based on subjective professional judgments for defining the temporal and spatial components of the water quality monitoring network (WQMN). This approach can neglect specific hydrologic aspects of the monitoring area and lack clear definitions of the monitoring goals (Harmancioglu et al. 1998; Strobl and Robillard 2008; Mei et al. 2011; Mavukkandy et al. 2014). These deficiencies, coupled with limited feedback and evaluation routines, frequently lead to inefficiency and production of monitoring data with poor cost-benefit relationships. Adaptive management approaches can address such inefficiencies and respond to changing conditions. Regular reassessment of a WQMN allows identification of adjustments required because of data mining problems (e.g., insufficient, missing, or unreliable data), environmental changes, availability of new monitoring technologies, and potentially changing monitoring goals (Harmancioglu et al. 1998; Strobl and Robillard 2008).

Researchers have previously assessed existing WQMNs (e.g., Mei et al. 2011; Do et al. 2013; Calazans et al. 2018; Peña-Guzmán et al. 2019), including the revision of monitoring sites locations, sampling frequency, and water quality parameters measured (Jiang et al. 2020). Numerous techniques have been applied for these purposes, including multivariate statistical analysis (Ouyang 2005; Kovács et al. 2015; Calazans et al. 2018; Peña-Guzmán et al. 2019), entropy analysis (Mahjouri and Kerachian 2011), genetic algorithms (Park et al. 2006), geospatial analysis (Strobl et al. 2006), water quality modeling (Chen et al. 2012), analytic hierarchical

process (Do et al. 2013), and fuzzy logic (Chang and Lin 2014). Although there are different methods available to optimize WQMN, the previous studies did not simultaneously integrate (or only partially integrated) some relevant aspects of the WQMN planning on the optimization methods, such as the monitoring goals and the environmental representativeness of the monitoring sites. Thus, providing a flexible approach to fill this gap was a motivation for the present study.

An efficient river WQMN is especially crucial in developing countries, where financial resources are usually scarce, but population growth and water quality degradation are rapid (Capps et al. 2016; Ma et al. 2020). However, technical guidelines for planning WQMN in such countries are still limited, potentially leading to high monitoring costs and insufficient data to guide water resources management programs (Harmancioglu et al. 1998; Mavukkandy et al. 2014; Camara et al. 2020). All these challenges are present in Brazil, where limited sanitation infrastructure (e.g., 27% of the population with no sewage collection and treatment, ANA 2017) represents the main cause of water pollution. According to the Brazilian Water Regulatory Agency (ANA, “Agência Nacional de Águas e Saneamento Básico”), 12% of all the 8,863 surface water samples collected in the country in 2015 had poor water quality (ANA 2017). Despite the water quality degradation, 8 out of the 27 Brazilian states had no active WQMNs until 2016 (ANA 2018). Sociological and biological issues can lead to conditions that vary from those in more developed areas such as North America and Europe. For example, riparian zones in small streams can dominate water quality in some watersheds (e.g., Dodds and Oakes 2008).

Other point and nonpoint pollution sources affect Brazilian rivers' water quality, in addition to sewage contributions, especially nutrient and pesticide inputs from agricultural areas (Maillard and Pinheiro Santos 2008; de Mello et al. 2018). However, urban development in tropical cities without centralized sewage treatment can lead to greater influences of riparian areas further down in the watershed (Tromboni and Dodds 2017). Additionally, sediments from deforested areas (Taniwaki et al. 2017), metal contamination from mining (Veado et al. 2006), industrial wastewater discharges from fertilizer, leather, metallurgy, and sugar/ethanol production (Schulz and Martins-Junior 2001; Gunkel et al. 2007; Martinelli et al. 2013; Silva et al. 2016) can jeopardize the water quality. These differing sources of impact on water quality demand different approaches to WQMNs.

São Paulo State (Southeastern Brazil) is the most populous in the country (IBGE 2019). The state has a large diversity of potential influences on water quality, some particular to tropical or subtropical areas, and others to rapidly developing countries. For example, it includes a megalopolis (the city of São Paulo and surrounding municipalities, with over 21 million inhabitants) and numerous other cities from small to over

a million people in each. The state has extensive agriculture, industry, forest harvest, reservoirs, and other land developments. Yet, there are still reference areas, and several protected areas exist. Typical of many developing countries, untreated sewage discharges are still the main cause of local surface water pollution (ANA 2005; CETESB 2019). São Paulo State is attempting to address these water quality challenges. In 2016, the São Paulo surface WQMN run by the State Environmental Company (CETESB, “Companhia Ambiental do Estado de São Paulo”) had almost 500 monitoring sites across the state, with a bimonthly or quarterly frequency, and quantification of up to 60 water quality parameters in each site (CETESB 2019). This represents a site density around 2.0 sites/1,000 km² (CETESB 2017), which is six times above the national average (ANA 2018), but still lower than in the European Union (7.0 sites/1,000 km²) (European Commission 2010). Traditionally in Brazil, the WQMN and the streamflow networks run separately and are not integrated. Such separation is also frequently observed in the São Paulo State, leading to unpaired water quality data and river/stream information on discharge, water velocity and level. Thus, in this paper we concentrate on the water quality network with the realization that future work should harmonize hydrologic and water quality monitoring station locations. The São Paulo WQMN has a strong concentration of sampling sites in the eastern portion of the state due higher anthropogenic pressures, more severe degradation of water resources, and also to logistical aspects (e.g., the proximity of analytical labs to process the samples) (Midaglia 2011). Thus, the development of a proposal to promote spatial optimization of the São Paulo’s WQMN is essential and could be a starting point for similar initiatives in other areas with common challenges.

Here we assess São Paulo’s WQMN with respect to its spatial representativeness with a more detailed focus on the regional scale and based on specific statistical procedures and criteria. We used both multivariate statistics and stratified sampling strategy technique and defined clear objective criteria to assess redundancies in the current network and potential under-represented areas not well covered by the network. We hypothesized that the evaluation of the studied WQMN would show spatial imbalances with biases toward specific areas. We expect our workflow and methodology could be useful in sub-tropical regions and could be particularly useful in other developing countries with financial constraints that aim to improve their WQMN.

Materials and methods

Study area

São Paulo is the most industrialized state in Brazil (IBGE 2020) (Fig. 1), and covers an area of approximately 248,200

km². The population of about 45 million inhabitants is strongly concentrated on the east portion of its territory (CETESB, 2019). According to the Köppen-Geiger classification (Kottek et al. 2006), the dominant climate is tropical wet with dry winters (Aw), with relatively high average annual air temperatures (15–25°C), and average annual rainfall between 1,250 and 2,250 mm (De Souza Rolim et al. 2007). The main land use is agricultural (40%), while approximately 3% of the state are covered by artificial areas (built-up) and 11% by forested areas, especially tropical ombrophilous forest (Kronka et al. 2005).

The São Paulo State is divided into 22 water resource management units (UGRHIs, “Unidades de Gerenciamento de Recursos Hídricos”) for administrative purposes, based on its watersheds and similarities of environmental features (geomorphology, geology, regional hydrology, and hydrogeology). The present study focused on seven UGRHIs (Fig. 1) characterized by different predominant land uses, as agriculture (UGRHIs 09, 14, and 15), artificial (i.e., modified and urban) areas (UGRHI 06), and grassland or forest vegetation (UGRHIs 01, 03, and 11). The selected UGRHIs represent 32% of the area, and more than 55% of the population of the state, ~ 25 million inhabitants (Table 1).

Database

Data on *Escherichia coli* (*E.coli*), pH, biochemical oxygen demand (BOD), total nitrogen (TN), total phosphorus (TP), turbidity, total solids (TS), temperature, and dissolved oxygen (DO) from 2004 to 2018 were compiled for 160 river or stream monitoring sites. These sites were monitored by the CETESB in the São Paulo state (InfoÁguas Online System) in the UGRHIs 01, 03, 06, 09, 11, 14 and 15 (Table 1), with a bimonthly sampling frequency. The laboratory analyses were performed by CETESB (with laboratory accreditation by the National Institute of Metrology, Standardization and Industrial Quality) following Standard Methods (APHA 1998; APHA 2005). The parameters were selected because they are used for the calculation of the Water Quality Index (WQI) in Brazil, which is an adaptation of the index developed by the National Sanitation Foundation of the United States (Finotti et al. 2015). Such index is broadly applied to classify surface water quality (Noori et al. 2019). It was developed to provide an overview of the surface water quality, to compare stream conditions on different spatial and temporal scales, and to measure the progress of water quality programs (Brown et al. 1970). The WQI values range from 0 to 100, with five water quality categories: very poor (WQI ≤ 19), poor (19 < WQI ≤ 36), regular (36 < WQI ≤ 51), good (51 < WQI ≤ 79) or excellent (WQI > 79). The WQI is calculated by multiplying weighted scores for these nine parameters (CETESB, 2019). The weights for each parameter were defined based on the feedback of a group of water quality experts, with values

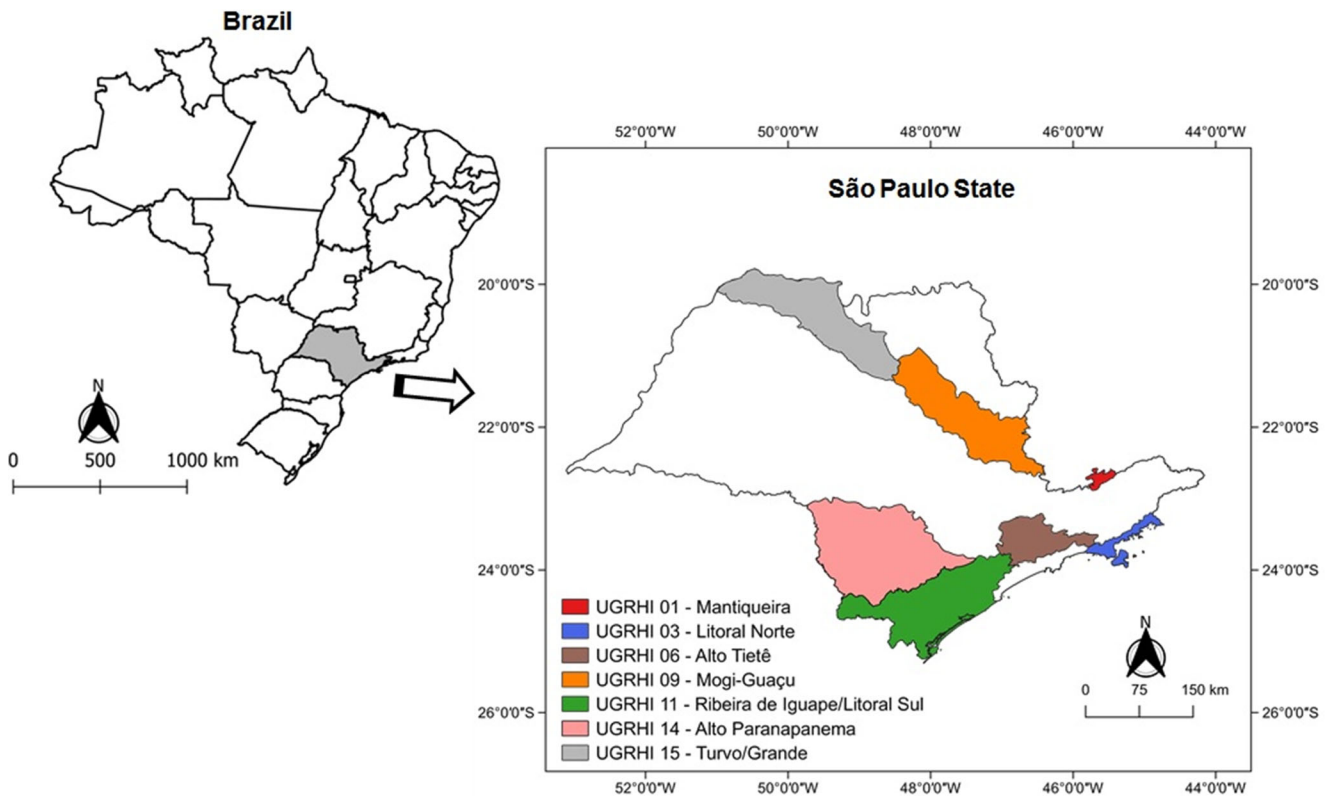


Fig. 1 Map of Brazil with the São Paulo State in gray (left) and map of the São Paulo State with the studied UGRHIs colored (right)

ranging from 0.08 (turbidity and total solids) to 0.15 (*E. coli*). The scores are obtained from variation curves which were also defined by a group of water quality experts. It is important to take into consideration that our study accounts only for the nine parameters selected, used for the WQI calculation. Other measurements included in CETESB’s monitoring program (e.g., metals, chlorophyll, pesticides, and toxicity), were not evaluated in this paper. However, the WQI variables used here are common indicators of water quality applied worldwide.

The raw database was initially submitted to a data treatment procedure for the selection of appropriate monitoring sites and parameters to perform further multivariate statistics. We detail the workflow of this procedure in Online Resource 1. All WQI parameters and 143 monitoring sites across 100 rivers and streams were considered suitable to the subsequent analyses (Table 2), and are hereafter referred to as “approved database”. The only exception was in UGRHI 11, where BOD and TN data were not consistent due to many censored

Table 1 Specific information about the studied UGRHIs and the São Paulo State in general, including total area (km²), population (10³ inhabitants), population density (inhabitants/km²) and predominant land uses in terms of area (Adapted from CETESB 2019)

UGRHI	Area (km ²)	Population (10 ³ inhab)	Population density (inhab/km ²)	Predominant land uses ^a
01—Mantiqueira	675	69	103	G (38%), MF (26%), F (19%), Others (17%)
03—Litoral Norte	1,948	331	170	F (86%), AA (6%), MF (3%), Others (5%)
06—Alto Tietê	5,868	21,386	3,645	AA (34%), F (29%), MF (28%), Others (9%)
09—Mogi-Guaçu	15,004	1,606	107	AG (68%), MF (15%), S (4%), Others (13%)
11—Ribeira de Iguape/ Litoral Sul	17,068	380	22	F (72%), MF (20%), S (3%), Others (5%)
14—Alto Paranapanema	22,689	774	34	AG (28%), MF (25%), S (19%), Others (28%)
15—Turvo/ Grande	15,925	1,355	85	AG (57%), MF (21%), MP (17%), Others (5%)
São Paulo State	248,222	45,513	183	AG (40%), MF (21%), MP (15%), Others (24%)

^a G grassland; MF mix of anthropogenic activities (e.g., agricultural, silviculture, managed pasture) and forested areas; F forest vegetation; AA artificial area (built-up); AG agricultural area; S silviculture (e.g., eucalyptus); MP managed pasture

Table 2 Summary of the data treatment, with the initial number of monitoring sites, initial site density, number of sites with suitable data for cluster analysis, and number of rivers and streams represented by sites with appropriate data for cluster analysis

UGRHI	Initial number of monitoring sites	Initial site density (sites/1,000km ²)	Number of sites with suitable data for cluster analysis	Represented rivers and streams
01	4	5.9	4	3
03	29	14.9	29	27
06	51	8.7	46	31
09	35	2.3	32	15
11	12	0.7	10	7
14	10	0.4	7	7
15	19	1.2	15	10
Total	160	2.0 ^a	143	100

^a Weighted average by the UGRHIs areas

data points. BOD was not used for the cluster analysis, and TN was replaced by the sum of total nitrogen Kjeldahl and nitrate in UGRHI 11.

Using primarily public sources (i.e., freely available), cartographic data were gathered to support the stratified sampling strategy. The cartographic variables (see details at Table S1 from Online Resource 2) were land use, average annual rainfall isohyets, soil types, hydrography, watershed coded by the Otto Pfafstetter method, protected areas, and digital elevation model (DEM). Contrasting time and spatial scales represent limitations from our cartographic database, and these constraints in data availability could be rectified in the future to further refine our analysis.

The multivariate statistics were performed in OriginPro 2016[®] and MATLAB 2015a, while geospatial analyses were performed in ArcGIS 10.3[®].

Redundancy identification

For each UGRHI, cluster analysis identified the degree of redundancy of the monitoring sites in relation to the studied WQI parameters, as recommended by Mavukkandy et al. (2014), CCME (2015), Calazans et al. (2018) and Peña-Guzmán et al. (2019). The agglomerative hierarchical clustering method was selected due to the advantages described in previous studies (Namratha and Prajwala 2012; Gomes et al. 2014). The water quality parameters were first transformed into logarithm base. After this step, the medians at each site for each parameter were obtained and standardized to Z-scale, which were the input data for the cluster analysis.

Linkage methods and similarity measures were chosen based on the cophenetic correlation coefficient (Sokal and Rohlf 1962), which has been employed to evaluate the

performance of different clustering techniques (da Silva and dos Dias 2013; Saraçlı et al. 2013). The cophenetic coefficients estimate the degree of distortion of the original distances between objects after the outputs of the cluster analysis, with a range from 0 to 1, and best results represented by higher coefficients (Saraçlı et al. 2013). Single, complete, average, median, Ward, and centroid linkage methods were tested associated with Euclidean and squared Euclidean distances.

The clustering criteria were based on the Silhouette Index, which can estimate the goodness-of-fit and validate the number of groups formed (Rousseeuw 1987). This index promotes comparisons between the dissimilarity of one object to its own group and the dissimilarity to other groups. A given object is well classified if the internal dissimilarity is low, but the external one is high. An average Silhouette Index greater than or equal to 0.71, as well as absence of negative values, were the criteria considered to delineate the most appropriate number of groups (Kaufman and Rousseeuw 2009).

The water quality monitoring sites allocated in the same group were considered to produce redundant information in relation to the WQI parameters. However, while the cluster analysis indicates the monitoring sites eligible for exclusion following a statistical approach, the method is not able to evaluate if these sites encompass different monitoring goals and therefore should not be excluded. Thus, we established the monitoring goals for each site to avoid network optimization based only on the statistical approach.

Definition of the monitoring goals

Four possible monitoring goals were considered: trend analysis, regulation, establishment of reference (pristine) conditions, and water body representativeness, with the respective criteria presented in Table 3. These goals are typically considered in surface water quality monitoring (Bartram and Ballance 1996; CCME 2015; Arle et al. 2016) and are bracketed by the general goals presented by CETESB (2019) for the São Paulo State WQMN.

Following the group formation in the cluster analysis and the assignment of the monitoring goals for each site, the following criteria were used to select monitoring sites that should not be excluded even with the statistical redundancy indicated by the cluster analysis:

- 1) In each group, keep the monitoring site which satisfies more goals;
- 2) Keep all monitoring sites associated to the regulation goal;
- 3) In each group, keep the monitoring site with the largest drainage area that meets the establishment of reference goal. For sites that do not meet the data requirements for the cluster analysis, keep all that satisfy the establishment of reference goal.

Table 3 Water quality monitoring goals and the respective classification criteria

Monitoring goal	Classification criteria
Trend analysis	Sites with at least a 10-year time series of available monitoring data.
Regulation	Sites downstream or upstream industries, cities, DWTP ^a or WWTP ^b .
Establishment of reference conditions	Sites with WQI greater than or equal to 70 (i.e., good or excellent conditions) in more than 80% of samples.
Water body representativeness	Sites with less than 10 years of monitoring, but located in locally relevant water bodies.

^a DWTP drinking water treatment plant; ^b WWTP wastewater treatment plant

- In each group, keep the monitoring sites with the largest drainage areas if another site located at the same river was not selected by previous criteria. The drainage areas data were obtained from the Hidroweb system from ANA. For sites with the absence of this information, the drainage area was calculated with the *Hydrology* tool from ArcGIS 10.3[®].

Monitoring network representativeness evaluation and spatial update proposal

We performed the representativeness evaluation with a stratified sampling strategy, which is based on the division of each UGRHI’s area into subgroups called strata. These spatial strata can be identified by the combination of relevant features for the study area, demonstrating high internal homogeneity (Gilbert 1987; Dobbie et al. 2008). The main advantages of this technique are the reduction of data redundancy and unsampled areas (Haining 2015); good performance to represent average and variance of data from heterogeneous areas (Cochran 1977; Wang et al. 2010); and unbiased sampling (Catherine et al. 2008; Dobbie et al. 2008).

The first step to create the strata was the management unit definition. Level 6 basins coded by the Otto Pfafstetter method were considered because it is the standard unit in the Brazilian Water Resources Policy (CNRH 2002) and recommended by the European Union (de Jager and Vogt 2010). For the strata definition, we performed cluster analysis, as suggested by Danz et al. (2005) and Catherine et al. (2008) for grouping basins based on relevant environmental features. The clustering methodology was the same described above, and the input data was the log-transformed percentages for each land use category (“anthropogenic factor”), average annual rainfall isohyets (“hydrological factor”), and soil type (“geological factor”) to reduce kurtosis and asymmetry. The use of average annual rainfall isohyets to define strata was possible after spatial interpolation by *Topo to Raster* tool from ArcGIS 10.3[®], followed by reclassification into three categories: below average rainfall, average rainfall, and above average rainfall. The rainfall values ranging between 90% and 110% of average annual rainfall were classified as “average rainfall”.

Each group of Pfafstetter level 6 basins formed in the cluster analysis (strata) represents different environmental features for which monitoring sites can be allocated. However, as advised by Danz et al. (2005) and Dobbie et al. (2008), a large number of groups formed by small units could make the water quality monitoring program impracticable and produce a high level of redundancy. Thus, we selected as representative for the trend analysis goal (see Table 3) the strata with a minimum area of 10% in comparison to the largest strata.

Whenever the existing monitoring sites were insufficient to represent all strata for trend analysis, priority areas for network expansion were investigated. The proposal consisted of indicating rivers reaches with Strahler order (Strahler 1952) greater than or equal to three and obviously located in the unrepresented strata. We considered rivers with order greater than or equal to three because they integrate larger areas of land, and the buffering mechanisms of downstream portions of a river network reduce the influences in water quality by unusual local conditions (Wohl 2017). However, it was not the objective of our study to define the micro-location of sites because this is mostly constrained by field inspection, taking into account access and sampling safety, mixing conditions, point sources of pollution, presence of discharge monitoring sites etc (Sanders 1988).

The present study also identified representative strata for sites aiming at the establishment of reference conditions (Table 3). We indicated areas with low human disturbance, representative of the trend analysis strata, and with small drainage areas, as recommended by Helmer (1994) and WMO (2013). We generated such strata by overlapping the layer of trend analysis strata with more than 50% of the area covered by forest vegetation or grassland with the layer of protected areas at São Paulo State. The protected areas are instituted by law to protect the natural resources within its own limits (Brasil 2000). In São Paulo State, they are frequently designed to promote the maintenance or improvement of rivers and streams water quality with a focus on water supply (Mello-Théry 2011; Dib et al. 2020), playing an important role as buffer zones to reduce the impacts of anthropogenic activities on water quality (e.g., nutrients and organic pollutant loads, sediments inputs) (Kuhlmann et al. 2014; Cunha et al. 2016).

Similarly to the protocol for the trend analysis goal, when the existing monitoring sites were insufficient to represent all

strata for the establishment of reference conditions, a proposal for priority areas for network expansion was presented. This consisted of rivers reaches with Strahler order (Strahler 1952) less than or equal to two located in the unrepresented strata. We prioritized lower Strahler order reaches in this case because the increase of the drainage areas reduces the probability of identifying reaches with low human disturbance (Dodds and Oakes 2004), making it more complex to detect the effects of land use changes on surface water quality (Thomas et al. 2004). We also evaluated if the suggested river reaches for network expansion to trend analysis goal intersected the unrepresented strata for the establishment of reference. In positive case, these reaches were considered sufficient to meet both goals.

The final WQMN spatial update proposal was composed by the monitoring sites selected on the steps of redundancy identification, goals definition, and representativeness evaluation, with the addition of suitable rivers reaches for new

monitoring sites that meet the trend analysis and/or the establishment of reference goals. Our study did not aim at proposing new areas for the regulation goal because it represents specific demands from the environmental agency (e.g., regulation of point source pollution and applied studies). New sites for the goal of water body representativeness were not proposed either because this goal only differs from the trend analysis goal by the duration of the data series. Finally, as a preliminary assessment of our proposal reliability, we ran the Mann-Whitney nonparametric test with a significance level of 0.05 for each WQI parameter considering the approved database in each UGRHI. This hypothesis test is frequently employed to statistically indicate if two groups belong to the same population or not. Thus, we compared the data series from the previously existing monitoring sites with the selected monitoring sites to look for statistical difference or similarity on the data structure before and after the spatial optimization. The workflow is summarized in Fig. 2.

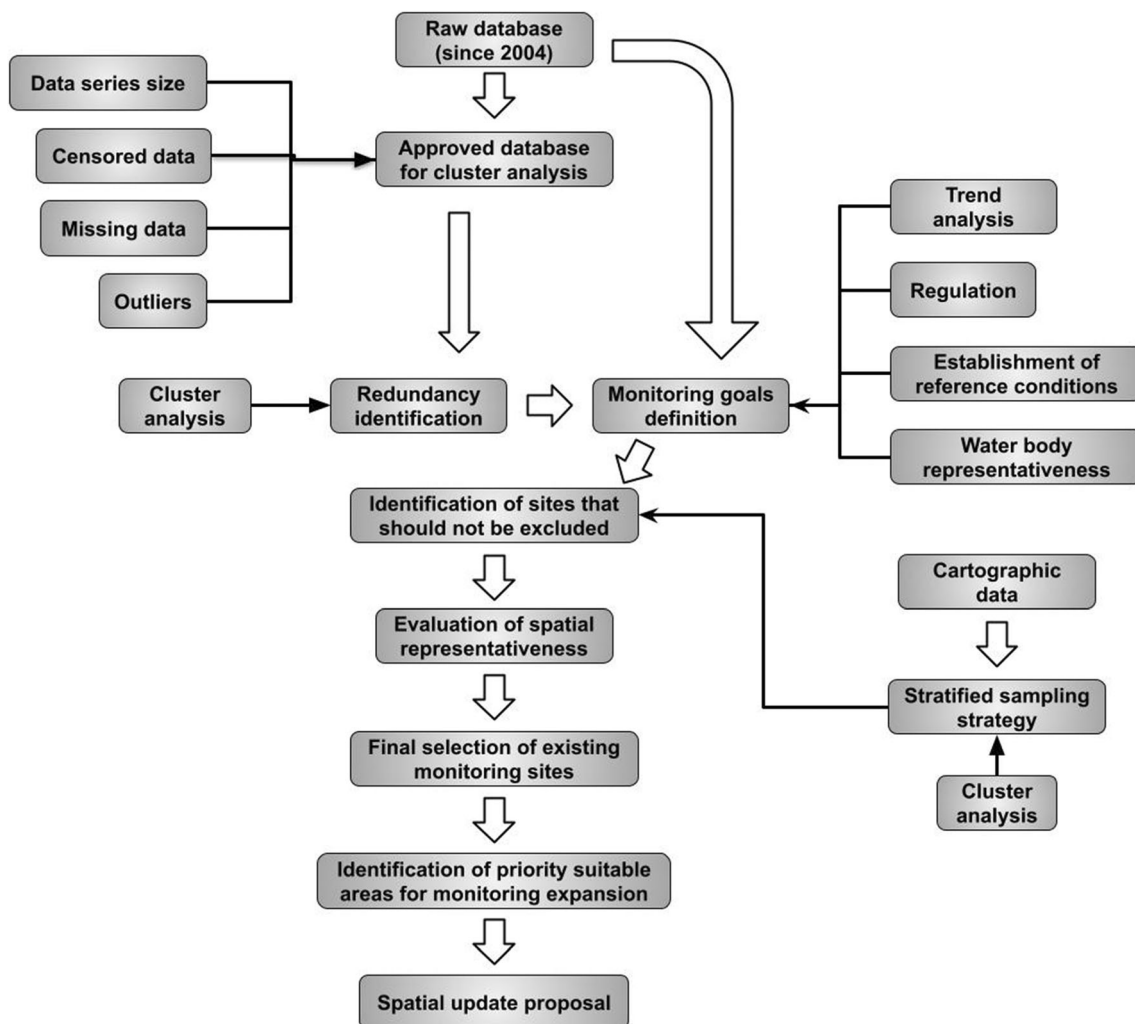


Fig. 2 Workflow summary for the spatial update proposal for each water resources management unit (UGRHI)

Results

Approved database

Following the initial data treatment procedure, more than 76,000 data points for the nine water quality parameters were approved, resulting in an average of ~1,200 data points per parameter in each UGRHI (Table 4). The descriptive statistics (median and percentiles) suggested a considerable variation in the water quality across the selected UGRHIs. In addition, data availability was heterogeneous due to different monitoring site densities and to the contrasting starting dates of operation.

Redundancy identification

For all the UGRHIs, the cophenetic correlation coefficient indicated the Euclidean distance associated with average, single, and complete linkage methods provided the lowest distortions of the original distances (higher coefficients). The average linkage had the best performance so we considered it the most appropriate method in five of the seven UGRHIs (Table 5). The cluster analysis suggested sampling sites produced redundant information in six of the seven UGRHIs, in which percentages of redundant sites varied from 13 to 59% (Table 5). The numbers of clusters formed in each UGRHI varied from 4 to 40, reaching the maximum in the UGRHI with the highest population density and the worst water quality (UGRHI 06). No redundancy was observed for UGRHI 01, with an average Silhouette Index of 1.00, indicating that all monitoring sites were classified in different groups.

Definition of monitoring goals

The most common monitoring goal we identified was the trend analysis, with 108 monitoring sites (Table 6) concentrated (more than 70%) in the three UGRHIs with the highest population density. On the other hand, the less represented goal was the establishment of reference, with only eight monitoring sites meeting this goal in four UGRHIs. Half of these sites were located in the UGRHI with the largest forest cover (UGRHI 03). The regulation goal was relevant in the network, with more than 21% of monitoring sites meeting this goal. The UGRHI with the highest number of monitoring sites for regulation (UGRHI 09) also presented the highest redundancy (Tables 5 and 6).

Considering the monitoring goals, the potential reduction in the number of sites in the UGRHIs with redundancy ranged from 12 to 46% of the existing monitoring sites. In general, the increase in site redundancy was followed by an increase in the potential of site reduction, regardless the definition of the monitoring goals for each site. This analysis indicates that of the initial 160 sites, 37 (23%) could be excluded, considering

the redundancy in relation to the WQI parameters and also the established goals.

Monitoring network representativeness evaluation and spatial update proposal

The stratified sampling strategy showed that the number of strata for trend analysis were different across the UGRHIs, with a minimum of three and a maximum of 64 strata (Table 7). In general, the UGRHIs with larger area had a greater number of strata, except for UGRHI 15 that had the highest number of strata but was not the largest in terms of area. The representativeness evaluation for trend analysis strata indicated that strata were more represented by existing monitoring sites in smaller UGRHIs (up to 100%), while strata representativeness below 20% was observed in larger UGRHIs.

More than 89% of the strata for the establishment of reference goal were concentrated in the two UGRHIs with more than 57% of the area covered by protected areas (UGRHI 03 and 11). In addition, less than 35% of the strata for the establishment of reference were represented by the existing monitoring sites. For the establishment of the reference goal, the identified strata showed that four out of the seven UGRHIs presented eligible river reaches to have reference monitoring sites (Table 7).

The final proposal for network expansion showed different patterns regarding the continuity of the existing monitoring sites and the number of suitable strata for network expansion in each UGRHI (Table 8). The continuity of the existing monitoring sites ranged from 63% to 100% of initial sites, with the lowest percentage in the UGRHI with the highest monitoring site redundancy (UGRHI 09). The number of suitable strata for network expansion ranged from 0 to 53 for trend analysis, 0 to 3 for the establishment of reference, and 0 to 15 for both goals. The UGRHI that had all strata represented was also the only one that had no redundant sites (UGRHI 01).

The final proposal suggested the reduction of the monitoring site densities up to 12% in the four UGRHIs with the highest initial monitoring site densities and the highest population densities. On the other hand, expansions in the number of monitoring sites from +125% to +390% were suggested in the three UGRHIs with the lowest population densities and the lowest initial monitoring site densities. The final monitoring site densities varied from 1.6 to 13.4 sites/1,000km² (Table 8). From the initial 160 existing monitoring sites, 126 (79%) were maintained and 127 new potential sites were identified (assuming only one monitoring site for each unrepresented strata). Interestingly, the results for the Mann-Whitney nonparametric tests ($\alpha=0.05$) showed that the proposed exclusion of monitoring sites did not significantly change the data series for most of the studied UGRHIs and parameters. This indicated that the data series before and after optimization remained

Table 4 Total number of data points, missing data, censored data, median, 10% percentile, and 90% percentile for data on *Escherichia coli* (*E. coli*), pH, Biochemical Oxygen Demand (BOD), total nitrogen (TN), total phosphorus (TP), temperature (T), turbidity (Turb), total solids (TS), and dissolved oxygen (DO) for all water resources management units (UGRHIs)

Parameter	UGRHI	<i>E. coli</i> CFU/ 100mL	pH	BOD mg/L	TN mg/L	TP mg/L	T °C	Turb NTU	TS mg/L	DO mg/L
Total number of data points	1	98	100	99	96	99	100	100	100	100
	3	2,061	2,064	2,057	1,908	2,062	2,064	2,064	2,063	2,064
	6	2,368	2,473	2,466	2,370	2,437	2,473	2,414	2,436	2,472
	9	2,204	2,240	2,218	2,179	2,237	2,241	2,221	2,217	2,235
	11	574	575	570	557	573	576	576	569	573
	14	463	469	458	464	464	472	472	471	472
	15	699	698	695	693	698	699	699	699	698
Missing data (%)	1	2	0	1	4	1	0	0	0	0
	3	1	1	2	8	1	1	1	1	1
	6	5	0	1	4	2	0	3	2	0
	9	2	0	1	1	0	0	1	1	0
	11	0	0	1	3	1	0	0	1	1
	14	2	1	0	2	2	0	0	0	0
	15	0	0	1	1	0	0	0	0	0
Censored data (%)	1	0	0	43	56	1	0	0	47	0
	3	0	0	55	69	32	0	0	25	1
	6	0	0	15	70	1	0	0	16	9
	9	0	0	47	33	8	0	0	11	0
	11	0	0	89 ^a	91 ^a	16	0	1	9	0
	14	0	0	62	59	7	0	0	2	0
	15	1	0	55	33	11	0	0	2	0
Median	1	9,350	6.9	3	1.8	0.09	18	14	51	6.5
	3	760	6.7	3	0.8	0.03	22	5	114	7.3
	6	410,000	7.1	20	10.8	0.76	22	29	250	1.7
	9	1,580	6.8	2	3.6	0.09	23	21	106	5.8
	11	720	7.1	5	1.3	0.04	22	18	134	7.6
	14	1,160	6.9	2	1.1	0.07	21	27	106	6.7
	15	680	7.0	2	1.6	0.10	23	26	140	5.3
10% Percent- ile	1	2,170	6.5	3	0.9	0.04	14	7	50	4.7
	3	71	6.3	2	0.4	0.01	19	2	50	3.3
	6	1,670	6.5	3	1.2	0.06	18	8	100	0.2
	9	120	6.1	2	2.5	0.02	18	9	58	2.7
	11	84	6.2	2	0.8	0.01	18	4	58	5.6
	14	145	6.6	2	0.6	0.02	17	12	62	4.5
	15	70	6.0	2	0.7	0.02	18	11	84	1.9
90% Percent- ile	1	54,900	7.3	6	3.5	0.19	20	65	176	8.2
	3	6,300	7.2	5	2.8	0.16	25	19	3,836	8.6
	6	4,300,000	7.6	94	28.9	2.81	26	90	482	6.4
	9	51,000	7.3	8	12.4	0.41	26	80	196	7.9
	11	5,570	7.9	5	1.6	0.48	26	64	522	9.3
	14	39,800	7.4	8	4.0	0.26	25	95	188	8.2
	15	60,000	7.4	8	5.9	0.59	27	75	242	7.5

^a Data did not meet the minimum requirements for the cluster analysis. Specifically for this UGRHI, BOD was not used and TN was replaced by the sum of total nitrogen Kjeldahl and nitrate concentrations

Table 5 Summary of the cluster analysis results, with most appropriate linkage method, cophenetic correlation coefficient, average Silhouette Index, number of groups formed, and number of sites with redundancy for each water resources management unit (UGRHI)

UGRHI	Linkage method/ cophenetic correlation coefficient	Average Silhouette Index	Number of groups formed	Sites with redundancy (% of approved)
01	Average/0.79	1.00	4	0 (0%)
03	Average/0.91	0.78	22	7 (24%)
06	Complete/0.81	0.87	40	6 (13%)
09	Single/0.90	0.72	13	19 (59%)
11	Average /0.84	0.77	8	2 (20%)
14	Average 0.89	0.75	4	3 (43%)
15	Average /0.86	0.75	10	5 (33%)

similar even with the exclusion of monitoring sites. Out of the 62 comparisons for the WQI parameters, only 10 (16%) presented a significant statistical difference considering the data series before and after the spatial optimization. The UGRHIs 06 and 09 presented all the observed differences, with seven and three in total, respectively. For the UGRHI 06, only pH and temperature did not present statistical differences, while the UGRHI 09 showed differences for BOD, temperature, and DO.

The final spatial update proposed for the UGRHIs 01, 03, 06, 09, 11, 14, and 15 (see an illustrative map in Fig. 3) is fully available in Online Resource 3 (Figures S1 to S7). For the UGRHIs with suitable areas for network expansion, new monitoring sites in the identified rivers reaches have the potential to promote better spatial homogeneity and representativeness.

Discussion

The monitoring site redundancy with respect to the selected WQI parameters was relatively high for most of the studied

UGRHIs, frequently above 20% (Table 5). Redundancy is also common in the WQMN of other developing countries (e.g., India, Colombia, and Iran), where more than 40% of water quality monitoring sites with high similarity were reported (Mavukkandy et al. 2014; Tavakol et al. 2017; Peña-Guzmán et al. 2019). This situation coupled with under representation of some watersheds can lead to WQMN with non-optimal cost-benefit relationships (Peña-Guzmán et al. 2019; Camara et al. 2020) that may fail to identify spatial and temporal changes in the water quality (Mahjouri and Kerachian 2011; Chen et al. 2012). Our results here indicate a perspective for spatial reassessment of the São Paulo State WQMN to reduce operational costs with redundant monitoring sites, to allow the expansion to areas with lack of data, and to promote a better coverage of the proposed monitoring goals.

The number of sites associated to the regulation goal (Table 6) contributed to the redundancy of the network. Such sites were mainly located to regulate point sources of pollution and not designed to provide spatial representativeness of water quality in the watersheds. However, we highlight that the parameters not analyzed in this study, such as heavy metals and toxicity, may throw another light in the importance of point source regulation sites. According to the land use cartographic base, the artificial areas represent less than 4% of the UGRHIs’ area. However, more than 65% of the existing monitoring sites are located up to 2 km far from the artificial areas’ polygons. Such spatial concentration may lead to a low representation of the identified strata for some of the selected UGRHIs (Table 7), especially for those more heterogeneous in relation to the environmental features. We highlight the case of the UGRHI 09 that presented the highest number of sites meeting the regulation goal, and 12 monitoring sites in the same river (Mogi-Guaçu River). This UGRHI had the highest statistical redundancy (59%) associated with the highest potential of exclusion of existing monitoring sites (37%).

The lack of monitoring sites for the establishment of reference conditions is another common deficiency in WQMN

Table 6 Results for monitoring goals definition, with the number of sites that met each goal (trend analysis, regulation, establishment of reference conditions, and water body representativeness) and the number of sites that could be excluded based on redundancy to WQI parameters and monitoring goals. A monitoring site could meet more than one goal at the same time

UGRHI	Trend analysis	Regulation	Establishment of reference conditions	Water body representativeness	Sites that could be excluded (% of total)
01	1	2	0	3	0 (0%)
03	28	5	4	1	6 (21%)
06	31	10	0	20	6 (12%)
09	23	11	1	12	16 (46%)
11	8	2	0	4	2 (17%)
14	6	1	1	4	2 (20%)
15	11	3	2	8	5 (26%)
total	108	34	8	52	37 (23%)

Table 7 Results for stratified sampling strategy for each water resources management unit (UGRHI), with the number of identified strata for trend analysis goal and the number of identified strata for the establishment of reference goal. Representativeness evaluation is also presented for each UGRHI, with strata for trend analysis goal already represented by the existing monitoring sites and strata for the establishment of reference goal already represented by the existing monitoring sites

UGRHI	Stratified sampling strategy		Strata representativeness evaluation for the existing monitoring sites	
	Number of identified STA ^a	Number of identified SER ^b	STA already represented (% of total)	SER already represented (% of total)
01	3	1	3 (100%)	1 (100%)
03	4	4	4 (100%)	1 (25%)
06	15	2	11 (73%)	1 (50%)
09	28	0	16 (57%)	* ^c
11	24	22	8 (33%)	7 (32%)
14	46	0	7 (15%)	* ^c
15	64	0	11 (17%)	* ^c
total	184	29	60 (33%)	10(34%)

^a STA strata for trend analysis goal; ^b SER strata for establishment of reference goal; ^c *: absence of SER

worldwide. Such issue is frequently attributed to the rapid development of human activities and the consequent changes in river water quality (Cunha et al. 2011; Davies-Colley et al. 2011; Huo et al. 2013; Almeida et al. 2014; Kaboré et al. 2018). Researchers have developed several methodologies to estimate the reference conditions to alleviate the problem of defining monitoring sites in rivers and streams with low human disturbance in developed regions. These methods include the trisection method, statistical relationships, and models (e.g., Smith et al. 2003). However, most authors suggest that data collected at pristine sites provide the best indicator of baseline conditions since contrasting estimation methods could present differences up to 50% in the reference concentrations (Dodds and Oakes 2004; Huo et al. 2013; Hsieh et al. 2016). Our results reinforced the difficulty of finding suitable river reaches for establishing reference conditions. For

example, we identified no strata for the establishment of reference in the predominantly agricultural UGRHIs (main land use of São Paulo State). These are characterized by the presence of large regions without protected areas, and intense deforestation as in other Brazilian agricultural areas in general (Sparovek et al. 2010; Calaboni et al. 2018). These areas associated with intensive agriculture have extensive use of pesticides and fertilizers, as well as land disturbance. These uses lead to problems such as high levels of nitrate, suspended solids, and turbidity that are found in São Paulo State agricultural catchments (Mori et al. 2015; Simedo et al. 2018).

The non-representativeness with respect to the reference sites could relate to the absence of this goal in the initial design of the São Paulo WQMN and to the concentration of sites in areas classified as artificial in the land use layer. Such monitoring strategy is possibly due to the need for water quality

Table 8 Results for the spatial update proposal for each water resources management unit (UGRHI), with existing monitoring maintained on the final proposal, number of strata for trend analysis goal suitable for network expansion, number of strata for the establishment of reference goal suitable for network expansion, number of strata suitable for network expansion with rivers reaches that meet both trend analysis and establishment of reference goals, and final proposal for monitoring site density

UGRHI	Spatial update proposal				
	Existing monitoring sites maintained (% of total)	Number of STA ^a suitable for network expansion	Number of SER ^b suitable for network expansion	Number of STR ^c suitable for network expansion	Final proposal for monitoring site density ^d (site/1,000km ²)/(% change from initial sites)
01	4 (100%)	0	0	0	5.9/ (0%)
03	23 (79%)	0	3	0	13.4/ (-10%)
06	41 (80%)	3	0	1	7.7/ (-12%)
09	22 (63%)	12	0	0	2.3/ (-3%)
11	11 (92%)	1	0	15	1.6 (+125%)
14	10 (100%)	39	0	0	2.2/ (+390%)
15	15 (79%)	53	0	0	4.3/ (+258%)
total	126 (79%)	108	3	16	3.2 (+58%)

^a STA strata for trend analysis goal; ^b SER strata for establishment of reference goal; ^c STR: strata with rivers reaches that simultaneously meet the goals of trend analysis and establishment of reference; ^d Assuming only one monitoring site for each unrepresented strata

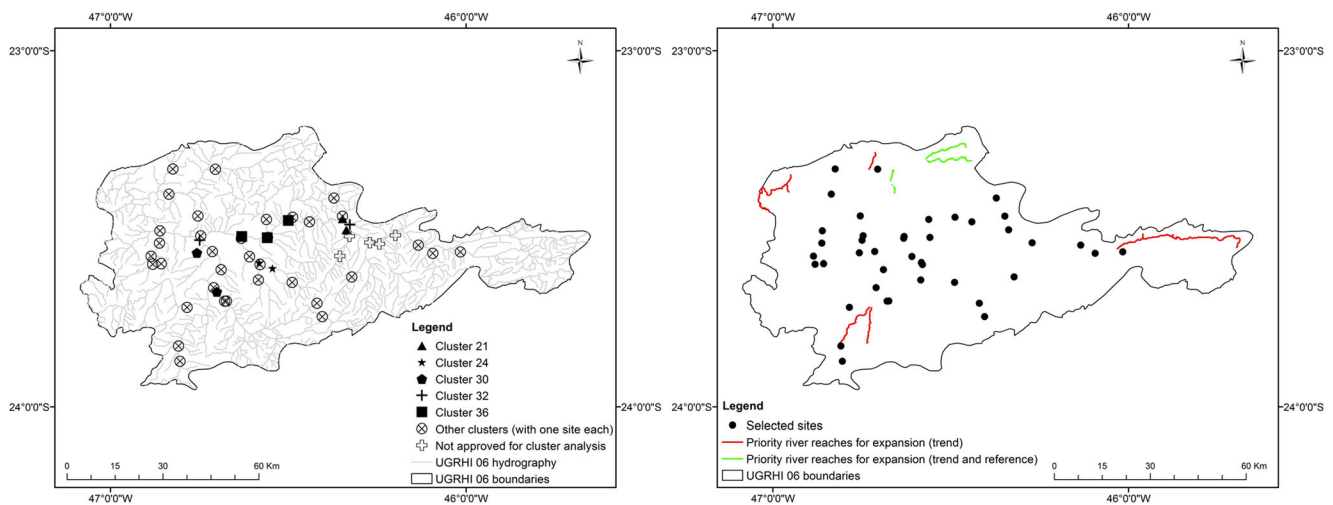


Fig. 3 Initial distribution of existing monitoring sites in UGRHI 06 and the respective groups formed after the cluster analysis (left). The final proposed network spatial structure is shown (right) with the selected sites

from the existing monitoring network and also the priority river reaches for network expansion (if applicable)

data in areas with higher water demand for human uses, associated with more conflicts on the use of water resources, coupled with financial and operational constraints. According to our results from the steps of goal definition and representativeness evaluation, only 12% of the existing monitoring sites met the reference goal. However, the WFD (2003) strongly recommends the design of networks aiming at establishing reference conditions in the different types of water bodies (classified according to geomorphology, drainage area, physicochemical parameters) as a tool to support the definition of the ecological status, and to assess the anthropogenic impacts in water bodies (Pardo et al. 2012; Voulvoulis et al. 2017) as these factors may alter baseline water quality as well as responses to anthropogenic pressures. This is a crucial concern for the São Paulo WQMN where the percentage of monitoring sites with poor and very poor water quality is above 16% since 2013 (CETESB 2019). Thus, our approach of selecting small streams with low disturbance as representative of the establishment of reference goal could be an important initiative towards the establishment of baseline conditions in the study area.

In the São Paulo State WQMN, all monitoring sites include at least the WQI parameters, always sampled in the same frequency. The same procedure is adopted at the National WQMN in Brazil. Nevertheless, the São Paulo State WQMN has a different approach for water supply sites and for evaluating aquatic life protection, including parameters such as cyanobacteria cells, chlorophyll and ecotoxicity assays. A partially similar approach is followed by other WQMN worldwide, which have flexible monitoring methods depending on the monitoring goals of each site, usually providing better cost-benefit (Strobl and Robillard 2008).

In Europe, the WFD (2003) suggested differentiating the monitoring methods according to the monitoring goals into

“surveillance”, “operational”, and “investigative”, each with different frequencies and parameters. The USA follows a similar approach, where the National Water Quality Assessment Program divides efforts into customized projects that aim at monitoring the national status of water quality, establishing reference conditions, and assessing changes in water quality by natural or anthropogenic features (Coles et al. 2019). In 2013 ANA started to build a National WQMN, run by the states, leading to a revision of the São Paulo sampling sites in the last five years. The novelty was the incorporation of specific monitoring goals for network design (“impact”, “strategic”, and “reference”) (ANA 2013) based on a larger scale methodology (basin levels 3 and 4 from the Otto Pfafstetter coding system). A more detailed design and different approaches based on environmental heterogeneity, and goals, such as proposed in this paper, could result in a more regionalized São Paulo State WQMN. Moreover, once a balanced and representative WQMN is established, spot sampling can fill gaps and deal with finer scale determinations, or control actions if necessary. Thus, our results regarding the proposed monitoring goals could be a starting point towards a water adaptive management in São Paulo and Brazil in general.

Our final proposed network structure presented an average site density weighted by the UGRHIs’ areas of 3.2 sites/1,000 km², which is compatible with the minimum site density of 1.0 site/1,000 km² recommended for the National WQMN (ANA 2013). This is comparable to other proposals (Table 9) that followed contrasting methodologies and considered different objectives for the WQMN optimization in the United States (Ouyang 2005), China (Ning and Bin 2004), Brazil (Calazans et al. 2018), and Malaysia (Camara et al. 2020), where average site densities varied from 2.6 to 4.6 sites/1,000km². However, other studies (Table 9) indicated

Table 9 Overview of some studies on water quality monitoring network optimization. For each reference, study area (river/country), watershed characteristics, optimization methods, drainage area, initial number of monitoring sites, proposed number of monitoring sites, initial site density, and proposed site density are shown

River/Country	Watershed characteristics	Methods	Drainage area (km ²)	Initial number of monitoring sites	Proposed number of monitoring sites	Initial site density (site/1,000km ²)	Proposed site density (site/1,000km ²)	References
Tisza/Hungary	Several large cities along the river. Anthropogenic factors strongly influence river water quality.	CA ^a , DA ^b	47,000	14	11	0.3	0.2	Tanos et al. 2015
Tunjuelo/Colombia	Drainage area is predominant urban. The river water quality is compromised by industrial and domestic wastewater.	PCA ^c , CA ^a , and nonparametric tests	457	10	7	21.9	15.3	Peña-Guzmán et al. 2019
Liaohu/China	The river is an important source for water supply. Industrial and domestic wastewaters affect the river water quality.	Modified approaching degree model	69,200	8	5	0.1	0.1	Wang et al. 2020
Jajrood/Iran	The river is an important source for water supply. Domestic wastewater affects river water quality.	Entropy	710	(Saedi et al. 2011)	33	30	46.5	42.2
Mahjouri and Kerachian 2011								
St Johns/United State	The drainage area is covered by forested, urban and industrial areas. Nutrients, hydrocarbons, pesticides and heavy metals strongly influence river water quality.	PCA ^c , FA ^d	7,192	22	19	3.0	2.6	Ouyang 2005
Ping/China	The drainage area is covered by agricultural and urban areas. Recently have been experiencing industrial expansion.	Fuzzy analysis	3,256	7	15	2.2	4.6	Ning and Bin 2004
Paraobebe/Brazil	The river has been employed to wastewater dilution and water supply. Anthropogenic activities as agriculture and industry compromise river water quality.	CA ^a , PCA ^c , FA ^d , and nonparametric tests	12,054	33	32	2.7	2.6	Calazans et al. 2018
Selangor/Malaysia	The river is an important source of water supply. The drainage area is covered by agricultural and forested areas. Recently have been experiencing industrial expansion.	Geo-statistical modelling, fuzzy analysis, AHP ^e , Kendall W test, Cellular Automata and Markov chain model.	2,100	21	6	10.0	2.9	Camara et al. 2020
Karoon/Iran	The river is an important source of water supply.	PCA ^c	67,000	17	13	0.2	0.2	Noori et al. 2010

Table 9 (continued)

River/Country	Watershed characteristics	Methods	Drainage area (km ²)	Initial number of monitoring sites	Proposed number of monitoring sites	Initial site density (site/1,000km ²)	Proposed site density (site/1,000km ²)	References
Several rivers/Brazil	Industrial, urban, and agricultural wastewaters strongly influence river water quality. Drainage areas covered by artificial, agricultural, grassland, and forested areas. Untreated sewage discharges strongly influence rivers water quality.	CA ^a , definition of monitoring goals, stratified sampling strategy	79,177	160	253	2.0	3.2	Our study

^a CA cluster analysis; ^b DA discriminant analysis; ^c PCA principal component analysis; ^d FA factor analysis. ^e AHP analytical hierarchical process

contrasting densities as in watersheds of Hungary (Tanos et al. 2015), Colombia (Peña-Guzmán et al. 2019), Iran (Mahjouri and Kerachian 2011), and China (Wang et al. 2020), with average site densities from 0.1 to 42.2 sites/1,000 km². In general, for most studies compiled in Table 9, the higher the initial monitoring site density, the higher the final proposed monitoring site density. We did not observe this pattern in most of the studied UGRHIs, as greater initial density did not necessarily lead to greater proposed final density of sites. Such differences are probably related to the evaluation of spatial representativeness based on environmental features we performed in contrast to most of the studies in Table 9. Our analyses also showed that site density can vary according to the spatial heterogeneity and monitoring goals, resulting in densities from 1.6 to 13.4 sites/1,000 km² across different UGRHIs.

The data series of the existing monitoring sites versus the optimized sites were consistent for most of the UGRHIs, with no significant statistical differences for the WQI parameters. This indicated that the data structure was not significantly changed after the spatial update proposal, suggesting that the optimized network could provide a similar degree of information even with a reduction of 21% of the total number of monitoring sites. Only UGRHIs 06 and 09 presented statistical differences for some parameters, but still the optimization does not necessarily imply the provision of statistically similar data series as the exclusion of redundant sites can significantly change the means of the WQI parameters and thus the structure of the dataset in general. We argue our methodology could be incorporated in other countries to promote a better link among network design, monitoring goals, and representativeness of land use, hydrological features, and geomorphological characteristics in the watersheds. However, if these data are used to simply cut back the total number of stations, no improvement

would be seen. The resources would need to be re-allocated to create a better coverage of stations. This re-allocation should also consider overlapping water quality and hydrologic monitoring stations in areas where they are not congruent.

In the present study, there was a greater need for network expansions in the three UGRHIs with the lowest initial site densities and the lowest population densities. The UGRHI 15, despite not being the largest one, presented the highest demand for monitoring expansion, but also potential exclusion of 21% of the existing monitoring sites. This suggests the current spatial distribution of the monitoring sites is specially unbalanced in this UGRHI, which is highly heterogeneous to environmental features. Therefore, our results showed that the area was not the single driver of the demand for new monitoring sites. The number of identified strata and the spatial representativeness of existing monitoring sites also influenced projected needs for new sites. Chen et al. (2012) also reported a lack of spatial representativeness and suggested a need for relocation of more than 20% of existing monitoring sites in the Heilongjiang River in Northeast China. On the other hand, it is clear that the resulting expansion of 53 new sampling sites only in one UGRHI may be unrealistic due to the financial constraints and the increment in operational costs, but prioritizing river reaches for the expansion can be a starting point toward a more representative network.

The congruence of water quality and hydrologic monitoring sites is an additional consideration: in São Paulo state they are not always co-located. Integrating water quality and quantity is crucial for effectively evaluate and manage this resource. While baseflow conditions are extremely important for some aspects of water quality at each site (e.g., average conditions experienced by biota, average drinking water quality), considering high flow events can also be important. If the

concern is mass of pollutants transported downstream, then flow weighting is necessary and co-location of hydrologic and water quality sites is particularly important.

The three UGRHIs with the highest population density had the highest potential reduction of the number of monitoring sites. High population density has a direct relationship with surface water pollution and can create unusual local conditions influencing water quality (Chen et al. 2016; Liyanage and Yamada 2017; Diamantini et al. 2018). Therefore, many authors recommend intensifying monitoring in such areas with poor surface water quality (Liyanage et al. 2016; Calazans et al. 2018). However, for the UGRHIs of interest, the strong presence of human activities generated a high number of monitoring sites meeting the same monitoring goals, leading to redundancy in relation to the WQI parameters. Brazilian regulation establishes five classes of water quality (Classes Special, 1, 2, 3, and 4), according to its expected uses, with specific water quality guidelines allowing increasing levels of pollution. In urban areas of the São Paulo State, poor quality river reaches (“Class 4”) are common and influence the quality of the same river downstream with higher water quality requirements, demanding a higher number of sites in the same river. Therefore, such legal point of view may partially explain the greater density of monitoring sites in urban areas. Although this was not one of the goals considered, our results confirm the initial hypothesis that the São Paulo State WQMN presents areas that may be either over or sub represented.

A crucial assumption of using homogeneous areas (e.g., strata, ecoregions) as a tool for the evaluation of spatial representativeness is that there is a spatial correlation between the water quality parameters and the environmental features used as input for the definition of the homogenous areas. Thus, the choice of inappropriate scales or layers that have no correlation with water quality can lead to misclassifications of water bodies (Bailey 2004; Cheruvilil et al. 2008) as well as a misrepresentation of the overall water quality. It is well known that hydrological, geological, and anthropogenic characteristics are important drivers of water quality composition (Khatri and Tyagi 2015; Igwe et al. 2017). However, there is a great variety of layers that represent such characteristics, and it would be impracticable to select all of them. In our study, we considered the land use (anthropogenic), the rainfall (hydrological), and the soil types (geological) as input features, but additional parameters could control water quality. Our choice was based on data availability for the studied area, and on previous investigations that showed high correlations between such features and surface water quality (Shehane et al. 2005; Taka et al. 2016; Smedo et al. 2018). Potentially relevant additional layers proposed in other studies include watershed morphometry (Catherine et al. 2008), atmospheric deposition, population density, sources of pollution (Danz et al. 2005), climate, and geology (Omernik et al. 2000).

Conclusions

Our study indicated that monitoring sites of the São Paulo WQMN are concentrated in areas with higher human disturbance levels. This result, using the proposed methodology, generated redundancies above 20% of the existing monitoring sites in most of the studied watersheds, thus raising the possibility for their exclusion, as well as pinpointed the lack of sites suitable for establishment of reference conditions. Especially in watersheds with more heterogeneous environmental features, such as UGRHIs 11, 14, and 15, a low spatial representativeness was established, with less than 35% of the identified strata represented. Moreover, we reported the possibility of exclusion of the existing monitoring sites in some of the UGRHIs, with a potential reduction of up to 37%. The UGRHIs with the highest population densities presented potential reduction of the number of monitoring sites of up to 12%, while the UGRHIs with the lowest population densities presented a potential of expansion up to 390%, resulting in final monitoring site densities from 1.6 to 13.4 sites/1,000km². We argue that the WQMN optimization does not necessarily imply reduction in the number of monitoring sites. In practical terms, the need for expansion of the network is even a more critical aspect in comparison to the elimination of redundant monitoring. Redundant monitoring sites still provide data to support water quality management in the represented area. Conversely, areas with a lack of monitoring sites can provide limited data to water quality programs. Obviously, the feasibility of such expansions needs further evaluation from the environmental agency because it brings a financial impact, and additional criteria to prioritize the river reaches for expansion could be necessary, such as population density and/or strata’s area. We argue that the WQMN optimization is especially necessary for developing countries like Brazil, where i) there are financial constraints for the investments in design, operation and maintenance of the WQMN; ii) surface water quality is strongly affected by point and nonpoint sources of pollution; iii) fast population growth and limited sanitation infrastructure exacerbate the presence of conflicts for water uses (e.g., irrigation, energy generation, public water supply, industry).

Although there is no scientific consensus about the best method for the WQMN spatial optimization, the cluster analysis followed by monitoring goals definition and stratified sampling strategy presented here provided consistent results with other studies in literature. The main advantage of our approach is the consideration of relevant environmental features affecting water quality (anthropogenic, hydrological, and geological aspects) in a more detailed scale to guide proposed reduction or expansion in the number of monitoring sites towards increase representativeness, while taking into account the monitoring goals. Our approach could benefit from the integration between water quality data, streamflow, and other physical data (e.g., discharge, water velocity and river/stream morphometry). Because such integration is not common in the São Paulo State and in Brazil in general, data was not available, and these aspects were not considered in the redundancy analysis, but

we are aware that this is another important criteria to be included in WQMN optimization. We expect our flexible methodology could serve as a basis for the WQMN optimization in other developing countries, with the possibility of adapting the parameters for cluster analysis, the monitoring goals, and the input features for the stratified sampling strategy procedure. Future research initiatives are still needed to indicate the efficiency and advantages of the proposed strategy. Further comparisons with widely used optimizing WQNs methodologies could represent a starting point in this direction.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s11356-021-16344-6>.

Acknowledgements DGFC thanks the National Council for Scientific and Technological Development (CNPq, Conselho Nacional de Desenvolvimento Científico e Tecnológico) in Brazil for the research productivity grant (Process #310844/2020-7). RGBA thanks the Geological Survey of Brazil (CPRM, Companhia de Pesquisa de Recursos Minerais) for the in-kind support for this research.

Author contribution RGBA and DGFC contributed to the study conception and design. Material preparation, data collection and analysis were performed by RGBA and DGFC. Interpretation of the results and discussions were performed by RGBA, MCL, WKD and DGFC. The first draft of the manuscript was written by RGBA and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding DGFC received a research productivity grant from the National Council for Scientific and Technological Development (CNPq, Conselho Nacional de Desenvolvimento Científico e Tecnológico) in Brazil (Process #310844/2020-7).

Data availability The datasets used and analyzed during the current study are available from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate Not applicable.

Consent for publication Not applicable.

Competing interests The authors declare no competing interests.

References

- Agência Nacional de Águas (ANA) e Saneamento Básico (2017) Conjuntura dos Recursos Hídricos no Brasil 2017: Relatório Pleno. Brasília. Available at: http://www.snirh.gov.br/portal/snirh/centrais-de-conteudos/conjuntura-dos-recursos-hidricos/conj2017_rel-1.pdf. Accessed 10 June 2020 (in Portuguese)
- Agência Nacional de Águas e Saneamento Básico (ANA) (2005) Panorama da Qualidade das Águas Superficiais no Brasil Caderno de Recursos Hídricos. Brasília. Available at: http://portalpnqa.ana.gov.br/Publicacao/PANORAMA_DA_QUALIDADE_DAS_AGUAS.pdf. Accessed 05 May 2020 (in Portuguese)
- Agência Nacional de Águas e Saneamento Básico (ANA) (2013) Resolução N° 903, de 22 de julho de 2013. Cria a Rede Nacional de Monitoramento de Qualidade das Águas Superficiais-RNQA e estabelece suas diretrizes. Available at: <http://arquivos.ana.gov.br/resolucoes/2013/903-2013.pdf>. Accessed 20 October 2019 (in Portuguese)
- Agência Nacional de Águas e Saneamento Básico (ANA) (2018) Conjuntura dos Recursos hídricos no Brasil 2018: informe anual. Brasília. Available at: <http://arquivos.ana.gov.br/portal/publicacao/Conjuntura2018.pdf>. Accessed 10 June 2020 (in Portuguese)
- Almeida SFP, Elias C, Ferreira J, Tornés E, Puccinelli C, Delmas F, Dörflinger G, Urbanić G, Marcheggiani S, Rosebery J, Mancini L, Sabater S (2014) Water quality assessment of rivers using diatom metrics across Mediterranean Europe: A methods intercalibration exercise. *Sci Total Environ* 476–477:768–776. <https://doi.org/10.1016/j.scitotenv.2013.11.144>
- APHA (1998) Standard Methods for the Examination of Water and Wastewater, 20th edn. American Public Health Association, American Water Works Association and Water Environmental Federation, Washington DC
- APHA (2005) Standard Methods for the Examination of Water and Wastewater, 21st edn. American Public Health Association/American Water Works Association/Water Environment Federation, Washington DC
- Arle J, Mohaupt V, Kirst I (2016) Monitoring of Surface Waters in Germany under the Water Framework Directive—A Review of Approaches, Methods and Results. *Water* 8:217. <https://doi.org/10.3390/w8060217>
- Bailey RG (2004) Identifying ecoregion boundaries. *Environ Manage* 34: 14–26. <https://doi.org/10.1007/s00267-003-0163-6>
- Bartram J, Ballance R (eds) (1996) Water quality monitoring: a practical guide to the design and implementation of freshwater quality studies and monitoring programmes. CRC Press, Boca Raton
- Bostanmaneshrad F, Partani S, Noori R, Nachtnebel HP, Berndtsson R, Adamowski JF (2018) Relationship between water quality and macro-scale parameters (land use, erosion, geology, and population density) in the Siminehrood River Basin. *Sci Total Environ* 639: 1588–1600. <https://doi.org/10.1016/j.scitotenv.2018.05.244>
- Brasil (2000) Lei n 9.985, de 18 de julho de 2000. Regulamenta o art. 225, § 1o, incisos I, II, III e VII da Constituição Federal, institui o Sistema Nacional de Unidades de Conservação da Natureza e dá outras providências. Diário Oficial da República Federativa do Brasil, Brasília (in Portuguese)
- Brown RM, McClelland NI, Deininger RA, Tozer RG (1970) A water quality index—do we dare. *Water Sew Work* 117:339–343
- Calaboni A, Tambosi LR, Igari AT, Farinaci JS, Metzger JP, Uriarte M (2018) The forest transition in São Paulo, Brazil: Historical patterns and potential drivers. *Ecol Soc* 23. <https://doi.org/10.5751/ES-10270-230407>
- Calazans GM, Pinto CC, da Costa EP, Perini AF, Oliveira SC (2018) The use of multivariate statistical methods for optimization of the surface water quality network monitoring in the Paraopeba river basin, Brazil. *Environ Monit Assess* 190:491. <https://doi.org/10.1007/s10661-018-6873-2>
- Camara M, Jamil NR, Abdullah AF Bin, et al (2020) Economic and efficiency based optimisation of water quality monitoring network for land use impact assessment. *Sci Total Environ* 737:. <https://doi.org/10.1016/j.scitotenv.2020.139800>
- Canadian Council of Ministers of the Environment (CCME) (2015) Guidance Manual for Optimizing Water Quality Monitoring Program Design Executive Summary. In C. C. O. M. O. T Environment, ed., p. 88
- Capps KA, Bentsen CN, Ramirez A (2016) Poverty, urbanization, and environmental degradation: Urban streams in the developing world. *Freshw Sci* 35:429–435. <https://doi.org/10.1086/684945>
- Catherine A, Troussellier M, Bernard C (2008) Design and application of a stratified sampling strategy to study the regional distribution of

- cyanobacteria (Ile-de-France, France). *Water Res* 42:4989–5001. <https://doi.org/10.1016/j.watres.2008.09.028>
- Chang CL, Lin YT (2014) A water quality monitoring network design using fuzzy theory and multiple criteria analysis. *Environ Monit Assess* 186:6459–6469. <https://doi.org/10.1007/s10661-014-3867-6>
- Chen Q, Wu W, Blanckaert K, Ma J, Huang G (2012) Optimization of water quality monitoring network in a large river by combining measurements, a numerical model and matter-element analyses. *J Environ Manage* 110:116–124. <https://doi.org/10.1016/j.jenvman.2012.05.024>
- Chen Q, Mei K, Dahlgren RA, Wang T, Gong J, Zhang M (2016) Impacts of land use and population density on seasonal surface water quality using a modified geographically weighted regression. *Sci Total Environ* 572:450–466. <https://doi.org/10.1016/j.scitotenv.2016.08.052>
- Cheruvetil KS, Soranno PA, Bremigan MT, Wagner T, Martin SL (2008) Grouping lakes for water quality assessment and monitoring: The roles of regionalization and spatial scale. *Environ Manage* 41:425–440. <https://doi.org/10.1007/s00267-007-9045-7>
- Cochran WC (1977) Sampling techniques, third edn. John Wiley and Sons, New York
- Coles JF, Riva-Murray K, Van Metre PC, et al (2019) Design and methods of the US Geological Survey Northeast Stream Quality Assessment (NESQA), 2016. US Geological Survey
- Companhia Ambiental do Estado de São Paulo (CETESB) (2017) Qualidade das águas interiores no estado de São Paulo 2016. São Paulo: CETESB. Available at: https://cetesb.sp.gov.br/aguas-interiores/wp-content/uploads/sites/12/2013/11/Cetesb_QualidadeAguasInteriores_2016_corre%C3%A7%C3%A3o02-11.pdf. Accessed: 15 August 2020 (in Portuguese)
- Companhia Ambiental do Estado de São Paulo (CETESB) (2019) Qualidade das águas interiores no estado de São Paulo 2018. São Paulo: CETESB. Available at: <https://cetesb.sp.gov.br/aguas-interiores/wp-content/uploads/sites/12/2019/10/Relat%C3%B3rio-de-Qualidade-das-%C3%81guas-Interiores-no-Estado-de-SP-2018.pdf>. Accessed 23 August 2020 (in Portuguese)
- Conselho Nacional de Recursos Hídricos (CNRH) (2002). Resolução nº 30 de 11 de dezembro de 2002. Available at: <https://cnrh.mdr.gov.br/divisao-hidrografica-nacional/73-resolucao-n-30-de-11-de-dezembro-de-2002/file>. Accessed 01 August 2020 (in Portuguese)
- Cunha DGF, Dodds WK, Do CCM (2011) Defining nutrient and biochemical oxygen demand baselines for tropical rivers and streams in São Paulo State (Brazil): A comparison between reference and impacted sites. *Environ Manage* 48:945–956. <https://doi.org/10.1007/s00267-011-9739-8>
- Cunha DGF, Sabogal-Paz LP, Dodds WK (2016) Land use influence on raw surface water quality and treatment costs for drinking supply in São Paulo State (Brazil). *Ecol Eng* 94:516–524. <https://doi.org/10.1016/j.ecoleng.2016.06.063>
- da Silva AR, dos Dias SCT (2013) A cophenetic correlation coefficient for tocher's method. *Pesqui Agropecu Bras* 48:589–596. <https://doi.org/10.1590/S0100-204X2013000600003>
- Danz NP, Regal RR, Niemi GJ, Brady VJ, Hollenhorst T, Johnson LB, Host GE, Hanowski JM, Johnston CA, Brown T, Kingston J, Kelly JR (2005) Environmentally stratified sampling design for the development of Great Lakes environmental indicators. *Environ Monit Assess* 102:41–65. <https://doi.org/10.1007/s10661-005-1594-8>
- Davies-Colley RJ, Smith DG, Ward RC, Bryers GG, McBride GB, Quinn JM, Scarsbrook MR (2011) Twenty years of New Zealand's national rivers water quality network: Benefits of careful design and consistent operation. *J Am Water Resour Assoc* 47:750–771. <https://doi.org/10.1111/j.1752-1688.2011.00554.x>
- de Jager AL, Vogt JV (2010) Développement et implémentation d'un système de codification d'entités hydrologiques structurées en Europe. *Hydrol Sci J* 55:661–675. <https://doi.org/10.1080/02626667.2010.490786>
- de Mello K, Valente RA, Randhir TO, Vettorazzi CA (2018) Impacts of tropical forest cover on water quality in agricultural watersheds in southeastern Brazil. *Ecol Indic* 93:1293–1301. <https://doi.org/10.1016/j.ecolind.2018.06.030>
- De Souza Rolim G, De Camargo MBP, Grosseliliana D, De Moraes JFL (2007) Climatic classification of köppen and thomthwaite systems and their applicability in the determination of agroclimatic zoning for the state of São Paulo, Brazil. *Bragantia* 66:711–720. <https://doi.org/10.1590/s0006-87052007000400022> (in Portuguese)
- Diamantini E, Lutz SR, Mallucci S, Majone B, Merz R, Bellin A (2018) Driver detection of water quality trends in three large European river basins. *Sci Total Environ* 612:49–62. <https://doi.org/10.1016/j.scitotenv.2017.08.172>
- Dib V, Nalon MA, Amazonas NT, Vidal CY, Ortiz-Rodríguez IA, Daněk J, Oliveira MF, Alberti P, Silva RA, Precinoto RS, Gomes TF (2020) Drivers of change in biodiversity and ecosystem services in the cantareira system protected area: a prospective analysis of the implementation of public policies. *Biota Neotrop* 20:1–12. <https://doi.org/10.1590/1676-0611-BN-2019-0915>
- Do HT, Lo SL, Phan Thi LA (2013) Calculating of river water quality sampling frequency by the analytic hierarchy process (AHP). *Environ Monit Assess* 185:909–916. <https://doi.org/10.1007/s10661-012-2600-6>
- Dobbie MJ, Henderson BL, Stevens DL (2008) Sparse sampling: Spatial design for monitoring stream networks. *Stat Surv* 2:113–153. <https://doi.org/10.1214/07-SS032>
- Dodds WK, Oakes RM (2004) A technique for establishing reference nutrient concentrations across watersheds affected by humans. *Limnol Oceanogr Methods* 2:333–341. <https://doi.org/10.4319/lom.2004.2.333>
- Dodds WK, Oakes RM (2008) Headwater influences on downstream water quality. *Environ Manage* 41:367–377. <https://doi.org/10.1007/s00267-007-9033-y>
- European Commission (2010) On Implementation of Council Directive 91/676/EEC Concerning the Protection of Waters Against Pollution Caused by Nitrates From Agricultural Sources Based on Member State Reports for the Period 2004–2007. In Commission Staff Working Document SEC (2010) 118, 41 pp
- Finotti AR, Finkler R, Susin N, Schneider VE (2015) Use of water quality index as a tool for urban water resources management. *Int J Sustain Dev Plan* 10:781–794. <https://doi.org/10.2495/SDP-V10-N6-781-794>
- Gilbert RO (1987) Statistical methods for environmental pollution monitoring. John Wiley & Sons, New York
- Gomes AI, Pires JCM, Figueiredo SA, Boaventura RAR (2014) Optimization of River Water Quality Surveys by Multivariate Analysis of Physicochemical, Bacteriological and Ecotoxicological Data. *Water Resour Manag* 28:1345–1361. <https://doi.org/10.1007/s11269-014-0547-9>
- Gunkel G, Kosmol J, Sobral M, Rohn H, Montenegro S, Aureliano J (2007) Sugar cane industry as a source of water pollution - Case study on the situation in Ipojuca river, Pernambuco, Brazil. *Water Air Soil Pollut* 180:261–269. <https://doi.org/10.1007/s11270-006-9268-x>
- Haining R (2015) Spatial Sampling. *International Encyclopedia of the Social & Behavioral Sciences* 23:185–190. <https://doi.org/10.1016/B978-0-08-097086-8.72065-4>
- Harmancioglu NB, Alpaslan MN., Singh VP (1998) Needs for Environmental Data Management. In: Harmancioglu NB., Singh VP, Alpaslan MN (eds) *Environmental Data Management. Water Science and Technology Library*, vol 27. Springer, Dordrecht
- Helmer R (1994) Water quality monitoring: national and international approaches. *Hydrol Chem Biol Process Transform Transp Contam Aquat Environ Proc Symp Rostov-on-Don* 1993:3–17

- Hsieh PY, Shiu HY, Te Chiueh P (2016) Reconstructing nutrient criteria for source water areas using reference conditions. *Sustain Environ Res* 26:243–248. <https://doi.org/10.1016/j.serj.2016.05.002>
- Huo S, Xi B, Su J, Zan F, Chen Q, Ji D, Ma C (2013) Determining reference conditions for TN, TP, SD and Chl-a in eastern plain ecoregion lakes, China. *J Environ Sci (China)* 25:1001–1006. [https://doi.org/10.1016/S1001-0742\(12\)60135-1](https://doi.org/10.1016/S1001-0742(12)60135-1)
- Igwe PU, Chukwudi CC, Ifenatuorah FC et al (2017) A Review of Environmental Effects of Surface Water Pollution. *Int J Adv Eng Res Sci* 4:128–137. <https://doi.org/10.22161/ijaers.4.12.21>
- Instituto Brasileiro de Geografia e Estatística (IBGE) (2019) População estimada: IBGE, Diretoria de Pesquisas, Coordenação de População e Indicadores Sociais, Estimativas da população residente com data de referência 1° de julho de 2019. In IBGE. Available at: <https://www.ibge.gov.br/estatisticas/sociais/populacao/9103-estimativas-de-populacao.html?=&t=resultados>. Accessed: 02 July 2020 (in Portuguese)
- Instituto Brasileiro de Geografia e Estatística (IBGE) (2020). Cadastro Central de Empresas. Available at: <https://cidades.ibge.gov.br/brasil/sp/pesquisa/19/29765?indicador=59927&tipo=ranking>. Accessed: 20 September 2020 (in Portuguese)
- Jiang J, Tang S, Han D, Fu G, Solomatine D, Zheng Y (2020) A comprehensive review on the design and optimization of surface water quality monitoring networks. *Environ Model Softw* 132:104792. <https://doi.org/10.1016/j.envsoft.2020.104792>
- Kaboré I, Moog O, Ouéda A, Sendzimir J, Ouédraogo R, Guenda W, Melcher AH (2018) Developing reference criteria for the ecological status of West African rivers. *Environ Monit Assess* 190:2. <https://doi.org/10.1007/s10661-017-6360-1>
- Kaufman L, Rousseeuw PJ (2009) Finding groups in data: an introduction to cluster analysis, vol 344. John Wiley & Sons, New York
- Khatri N, Tyagi S (2015) Influences of natural and anthropogenic factors on surface and groundwater quality in rural and urban areas. *Front Life Sci* 8:23–39. <https://doi.org/10.1080/21553769.2014.933716>
- Kottek M, Grieser J, Beck C, Rudolf B, Rubel F (2006) World map of the Köppen-Geiger climate classification updated. *Meteorol Zeitschrift* 15:259–263. <https://doi.org/10.1127/0941-2948/2006/0130>
- Kovács J, Kovács S, Hatvani IG, Magyar N, Tanos P, Korponai J, Blaschke AP (2015) Spatial optimization of monitoring networks on the examples of a river, a Lake-Wetland system and a Sub-Surface water system. *Water Resour Manag* 29:5275–5294. <https://doi.org/10.1007/s11269-015-1117-5>
- Kronka FJN, Nalon MA, Matsukuma CK et al (2005) Monitoramento da vegetação natural e do reflorestamento no Estado de São Paulo. XII Simpósio Bras Sensoriamento Remoto:1569–1576. <https://doi.org/10.1017/CBO9781107415324.004>
- Kuhlmann ML, Imbimbo HRV, Ogura LL, Villani JP, Starzynski R, Robim MJ (2014) Effects of human activities on rivers located in protected areas of the Atlantic Forest. *Acta Limnol Bras* 26:60–72. <https://doi.org/10.1590/s2179-975x2014000100008>
- Liyanage CP, Yamada K (2017) Impact of population growth on the water quality of natural water bodies. *Sustain* 9. <https://doi.org/10.3390/su9081405>
- Liyanage CP, Marasinghe A, Yamada K (2016) Comparison of optimized selection methods of sampling sites network for water quality monitoring in a river. *Int J Affect Eng* 15:195–204. <https://doi.org/10.5057/ijae.ijae-d-15-00043>
- Ma T, Sun S, Fu G, Hall JW, Ni Y, He L, Yi J, Zhao N, du Y, Pei T, Cheng W, Song C, Fang C, Zhou C (2020) Pollution exacerbates China's water scarcity and its regional inequality. *Nat Commun* 11:1–9. <https://doi.org/10.1038/s41467-020-14532-5>
- Mahjouri N, Kerachian R (2011) Revising river water quality monitoring networks using discrete entropy theory: The Jajrood River experience. *Environ Monit Assess* 175:291–302. <https://doi.org/10.1007/s10661-010-1512-6>
- Maillard P, Pinheiro Santos NA (2008) A spatial-statistical approach for modeling the effect of non-point source pollution on different water quality parameters in the Velhas river watershed - Brazil. *J Environ Manage* 86:158–170. <https://doi.org/10.1016/j.jenvman.2006.12.009>
- Martinelli LA, Filoso S, De ACB et al (2013) Water Use in Sugar and Ethanol Industry in the State of São Paulo (Southeast Brazil). *J Sustain Bioenergy Syst* 03:135–142. <https://doi.org/10.4236/jsbs.2013.2019>
- Mavukkandy MO, Karmakar S, Harikumar PS (2014) Assessment and rationalization of water quality monitoring network: A multivariate statistical approach to the Kabbini River (India). *Environ Sci Pollut Res* 21:10045–10066. <https://doi.org/10.1007/s11356-014-3000-y>
- Mei K, Zhu Y, Liao L, Dahlgren R, Shang X, Zhang M (2011) Optimizing water quality monitoring networks using continuous longitudinal monitoring data: A case study of Wen-Rui Tang River, Wenzhou, China. *J Environ Monit* 13:2755–2762. <https://doi.org/10.1039/c1em10352k>
- Mello-Théry NA (2011) Conservation of natural areas in São Paulo. *Estud Avancados* 25:175–188. <https://doi.org/10.1590/S0103-40142011000100012>
- Meybeck M (2003) Global analysis of river systems: From Earth system controls to Anthropocene syndromes. *Philos Trans R Soc B Biol Sci* 358:1935–1955. <https://doi.org/10.1098/rstb.2003.1379>
- Meybeck M, Helmer R (1989) The Quality of Rivers: From Pristine Stage To Global Pollution. *Palaeogeogr Palaeoclimatol Palaeoecol (Global Planet Chang Sect Elsevier Sci Publ BV* 75:283–309
- Midaglia CLV (2011) Proposta de Implantação do índice de abrangência espacial de monitoramento - IAEM por meio da evolução da rede de qualidade das águas superficiais do Estado de São Paulo. Doctoral Theses, University of São Paulo (in Portuguese)
- Mori GB, De Paula FR, De Ferraz SFB et al (2015) Influence of landscape properties on stream water quality in agricultural catchments in Southeastern Brazil. *Ann Limnol* 51:11–21. <https://doi.org/10.1051/limn/2014029>
- Namratha M, Prajwala TR (2012) A comprehensive overview of clustering algorithms in pattern recognition. *IOR J Comput Eng* 4:23–30
- Nel JL, Roux DJ, Abell R, Ashton PJ, Cowling RM, Higgins JV, Thieme M, Viers JH (2009) Progress and challenges in freshwater conservation planning. *Aquat Conserv Mar Freshw Ecosyst* 19:474–485
- Ning SK, Bin CN (2004) Optimal expansion of water quality monitoring network by fuzzy optimization approach. *Environ Monit Assess* 91:145–170. <https://doi.org/10.1023/B:EMAS.0000009233.98215.1f>
- Noori R, Sabahi MS, Karbassi AR, Baghvand A, Taati Zadeh H (2010) Multivariate statistical analysis of surface water quality based on correlations and variations in the data set. *Desalination* 260:129–136. <https://doi.org/10.1016/j.desal.2010.04.053>
- Noori R, Berndtsson R, Hosseinzadeh M, Adamowski JF, Abyaneh MR (2019) A critical review on the application of the National Sanitation Foundation Water Quality Index. *Environ Pollut* 244:575–587. <https://doi.org/10.1016/j.envpol.2018.10.076>
- Omernik JM, Chapman SS, Lillie RA, Dumke RT (2000) Ecoregions of Wisconsin. *Trans Wisconsin Acad Sci Arts Lett* 88:77–103
- Ouyang Y (2005) Evaluation of river water quality monitoring stations by principal component analysis. *Water Res* 39:2621–2635. <https://doi.org/10.1016/j.watres.2005.04.024>
- Pardo I, Gómez-Rodríguez C, Wasson JG, Owen R, van de Bund W, Kelly M, Bennett C, Birk S, Buffagni A, Erba S, Mengin N, Murray-Bligh J, Ofenböck G (2012) The European reference condition concept: A scientific and technical approach to identify minimally-impacted river ecosystems. *Sci Total Environ* 420:33–42. <https://doi.org/10.1016/j.scitotenv.2012.01.026>
- Park SY, Choi JH, Wang S, Park SS (2006) Design of a water quality monitoring network in a large river system using the genetic algorithm. *Ecol Model* 199:289–297. <https://doi.org/10.1016/j.ecolmodel.2006.06.002>

- Peña-Guzmán CA, Soto L, Diaz A (2019) A proposal for redesigning the water quality network of the Tunjuelo River in Bogotá. Colombia through a spatio-temporal analysis. *Resources* 8. <https://doi.org/10.3390/resources8020064>
- Pynegar EL, Jones JPG, Gibbons JM, Asquith NM (2018) The effectiveness of Payments for Ecosystem Services at delivering improvements in water quality: Lessons for experiments at the landscape scale. *PeerJ* 2018:1–29. <https://doi.org/10.7717/peerj.5753>
- Rodrigues V, Estrany J, Ranzini M, de Cicco V, Martín-Benito JMT, Hedo J, Lucas-Borja ME (2018) Effects of land use and seasonality on stream water quality in a small tropical catchment: The headwater of Córrego Água Limpa, São Paulo (Brazil). *Sci Total Environ* 622–623:1553–1561. <https://doi.org/10.1016/j.scitotenv.2017.10.028>
- Rousseeuw PJ (1987) Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *J Comput Appl Math* 20:53–65. [https://doi.org/10.1016/0377-0427\(87\)90125-7](https://doi.org/10.1016/0377-0427(87)90125-7)
- Saeedi M, Hosseinzadeh M, Rajabzadeh M (2011) Competitive heavy metals adsorption on natural bed sediments of Jajrood River, Iran. *Environ Earth Sci* 62:519–527. <https://doi.org/10.1007/s12665-010-0544-0>
- Sanders TG (1988) Chapter 13 Water Quality Monitoring Networks. In: Stephenson DBT-D in WS (ed) *Water and Wastewater System Analysis*. Elsevier, pp 204–216
- Saraçlı S, Doğan N, Doğan I (2013) Comparison of hierarchical cluster analysis methods by cophenetic correlation. *J Inequalities Appl* 2013:1–8. <https://doi.org/10.1186/1029-242X-2013-203>
- Schulz UH, Martins-Junior H (2001) *Astyanax fasciatus* as bioindicator of water pollution of Rio dos Sinos, RS, Brazil. *Brazilian J Biol* 61: 615–622. <https://doi.org/10.1590/s1519-69842001000400010>
- Shehane SD, Harwood VJ, Whitlock JE, Rose JB (2005) The influence of rainfall on the incidence of microbial faecal indicators and the dominant sources of faecal pollution in a Florida river. *J Appl Microbiol* 98:1127–1136. <https://doi.org/10.1111/j.1365-2672.2005.02554.x>
- Silva SVS, Dias AHC, Dutra ES et al (2016) The impact of water pollution on fish species in southeast region of Goiás, Brazil. *J Toxicol Environ Heal - Part A Curr Issues* 79:8–16. <https://doi.org/10.1080/15287394.2015.1099484>
- Simedo MBL, Martins ALM, Pissarra TCT, Lopes MC, Costa RCA, Valle-Junior RF, Campanelli LC, Rojas NET, Finoto EL (2018) Effect of watershed land use on water quality: A case study in córrego da olaria basin, são paulo state, Brazil. *Brazilian J Biol* 78: 625–635. <https://doi.org/10.1590/1519-6984.168423>
- Smith RA, Alexander RB, Schwarz GE (2003) Natural background concentrations of nutrients in streams and rivers of the conterminous United States. *Environ Sci Technol* 37:3039–3047. <https://doi.org/10.1021/es020663b>
- Sokal RR, Rohlf FJ (1962) The Comparison of dendrograms by objective methods. *Taxon* 11:33–40. <https://doi.org/10.2307/1217208>
- Sparovek G, Berndes G, Klug ILF, Barretto AGOP (2010) Brazilian agriculture and environmental legislation: Status and future challenges. *Environ Sci Technol* 44:6046–6053. <https://doi.org/10.1021/es1007824>
- Strahler AN (1952) Dynamic Basis of Geomorphology. *GSA Bull* 63: 923–938. [https://doi.org/10.1130/0016-7606\(1952\)63\[923:DBOG\]2.0.CO;2](https://doi.org/10.1130/0016-7606(1952)63[923:DBOG]2.0.CO;2)
- Stream Solute Workshop (1990) Concepts and methods for assessing solute dynamics in stream ecosystems. *J North Am Benthol Soc* 9: 95–119. <https://doi.org/10.2307/1467445>
- Strobl RO, Robillard PD (2008) Network design for water quality monitoring of surface freshwaters: A review. *J Environ Manage* 87:639–648. <https://doi.org/10.1016/j.jenvman.2007.03.001>
- Strobl RO, Robillard PD, Shannon RD, Day RL, McDonnell AJ (2006) A water quality monitoring network design methodology for the selection of critical sampling points: Part I. *Environ Monit Assess* 112:137–158. <https://doi.org/10.1007/s10661-006-0774-5>
- Taka M, Aalto J, Virkanen J, Luoto M (2016) The direct and indirect effects of watershed land use and soil type on stream water metal concentrations. *Water Resour Res* 52:7711–7725. <https://doi.org/10.1002/2016WR019226>
- Taniwaki RH, Cassiano CC, Filoso S, SFB F, Camargo PB, Martinelli LA (2017) Impacts of converting low-intensity pastureland to high-intensity bioenergy cropland on the water quality of tropical streams in Brazil. *Sci Total Environ* 584–585:339–347. <https://doi.org/10.1016/j.scitotenv.2016.12.150>
- Tanos P, Kovács J, Kovács S, Anda A, Hatvani IG (2015) Optimization of the monitoring network on the River Tisza (Central Europe, Hungary) using combined cluster and discriminant analysis, taking seasonality into account. *Environ Monit Assess* 187. <https://doi.org/10.1007/s10661-015-4777-y>
- Tavakol M, Arjmandi R, Shayeghi M, Monavari SM, Karbassi A (2017) Developing an environmental water quality monitoring program for Haraz River in Northern Iran. *Environ Monit Assess* 189:410. <https://doi.org/10.1007/s10661-017-6125-x>
- Thomas SM, Neill C, Deegan LA, Krusche AV, Ballester VM, Victoria RL (2004) Influences of land use and stream size on particulate and dissolved materials in a small Amazonian stream network. *Biogeochemistry* 68:135–151. <https://doi.org/10.1023/B:BIOG.0000025734.66083.b7>
- Tromboni F, Dodds WK (2017) Relationships Between Land Use and Stream Nutrient Concentrations in a Highly Urbanized Tropical Region of Brazil: Thresholds and Riparian Zones. *Environ Manage* 60:30–40. <https://doi.org/10.1007/s00267-017-0858-8>
- Veado MARV, Arantes IA, Oliveira AH, Almeida MRMG, Miguel RA, Severo MI, Cabaleiro HL (2006) Metal pollution in the environment of minas gerais state - Brazil. *Environ Monit Assess* 117:157–172. <https://doi.org/10.1007/s10661-006-8716-9>
- Voulvoulis N, Arpon KD, Giakoumis T (2017) The EU Water Framework Directive: From great expectations to problems with implementation. *Sci Total Environ* 575:358–366. <https://doi.org/10.1016/j.scitotenv.2016.09.228>
- Wang J, Haining R, Cao Z (2010) Sample surveying to estimate the mean of a heterogeneous surface: Reducing the error variance through zoning. *Int J Geogr Inf Sci* 24:523–543. <https://doi.org/10.1080/13658810902873512>
- Wang H, Jiao Z, Wang L, Wang Y, Luo Q, Wu H, Wang X, Sun L (2020) The study on optimal design of river monitoring network using modified approaching degree model: a case study of the Liaohe River, Northeast China. *Environ Sci Pollut Res*. 27:41515–41523. <https://doi.org/10.1007/s11356-020-10178-4>
- Water Framework Directive (WFD) 2003. Water Framework Directive, Common Implementation Strategy Working Group 2.7. Monitoring: guidance on monitoring for the water framework directive (2003), Final version 23 January 2003
- Wohl E (2017) The significance of small streams. *Front Earth Sci* 11: 447–456. <https://doi.org/10.1007/s11707-017-0647-y>
- World Meteorological Organization (WMO) (2013) Planning of water quality monitoring systems, Technical report series No. 3, WMO-No.1113, Geneva, Switzerland
- Worrall F, Howden NJK, Burt TP (2015) Time series analysis of the world's longest fluvial nitrate record: evidence for changing states of catchment saturation. *Hydrol Process* 29:434–444

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