Revista Brasileira de Ciências Ambientais
Brazilian Journal of Environmental Sciences

Modelling of the water quality index of the São Francisco River Integrating Project using Minimum Spanning Tree and SKATER algorithm

Modelagem do índice de qualidade de água do Projeto de Integração do Rio São Francisco usando *Minimum Spanning Tree* e algoritmo *SKATER*

Érika Alves Tavares Marques[1], *Anthony Epifanio Alves*[1], *Rogério Moreira Chagas*[2], *Arisvaldo Vieira Méllo Júnior*[2], *Maria do Carmo Sobral*[1]

ABSTRACT

This study modeled the water quality index in the East and North Axis of the São Francisco River Integration Project (PISF) using the statistical tools Minimum Spanning Tree and SKATER algorithm. This study is justified by presenting, for the first time, a historical series over 13 years, considering water quality data from the hydrographic basins included in the two axes of the transposition, between 2009 and 2022, showing the spatial-temporal evolution of water quality indices during this period. The spatial variation identified in each axis is useful for establishing sustainable management and planning programs for the water bodies under study. In both the North and East Axes, the WQI of the basins ranged from Excellent to Poor during the study period. It is noteworthy that the best WQI corresponded to the São Francisco River Basin, as it is the water-donating basin. It is noteworthy that no sampling points classified as Very Poor were recorded in either the North or East Axis during the study period. The findings highlighted that the use of WQI, associated with statistics and spatial modeling techniques was efficient to define that the waters of the studied sampling points were suitable for multiple uses, however, considering that they are located in the semi-arid region, it is necessary to intensify management actions in these basins, as climate change related issues could affect water quality in the future.

Keywords: transposition of the São Francisco River; environmental quality; spatial modeling; water quality index.

RESUMO

Neste estudo, foi feita a modelagem do índice de qualidade da água nos Eixos Leste e Norte do Projeto de Integração do Rio São Francisco (PISF) usando as ferramentas estatísticas *Minimum Spanning Tree* e algoritmo SKATER. A contribuição deste estudo justifica-se por apresentar, pela primeira vez, uma série histórica de 13 anos, considerando dados de monitoramento da qualidade da água das bacias hidrográficas incluídas nos dois eixos da transposição, entre 2009 e 2022, mostrando a evolução espaçotemporal dos índices de qualidade da água durante esse período. A variação espacial identificada em cada eixo é útil para estabelecer programas de gestão e planejamento sustentáveis para os corpos d'água em estudo. Tanto no Eixo Norte quanto no Leste, o índice de qualidade da água (IQA) das bacias hidrográficas variou de excelente a ruim durante o período do estudo. Vale destacar que o melhor IQA correspondeu à bacia do rio São Francisco, por ser esta a bacia doadora de água. É importante destacar que nenhum ponto de amostragem classificado como muito ruim foi registrado nos Eixos Norte ou Leste durante o período do estudo. Os resultados destacaram que o uso do IQA associado a técnicas estatísticas e de modelagem espacial foi eficiente para revelar que as águas dos pontos de amostragem estudados são adequadas para múltiplos usos. No entanto, considerando-se que estão localizados em uma região semiárida e em um contexto de mudanças climáticas, que podem comprometer a qualidade da água no futuro, é necessário intensificar as ações de gestão nessas bacias.

Palavras-chave: transposição do Rio São Francisco; qualidade ambiental; modelagem espacial; índice de qualidade de água.

[1]Universidade Federal de Pernambuco – Recife (PE), Brazil.

[2]Universidade de São Paulo – São Paulo (SP), Brazil.

Corresponding author: Érika Alves Tavares Marques – Avenida da Arquitetura, s/n. – Cidade Universitária – CEP: 50740-550 – Recife (PE), Brazil. E-mail: erikatmbio@gmail.com

Conflicts of interest: the authors declare there are no conflicts of interest.

Funding: National Council for Scientific and Technological Development (Universal Project 408231/2023-8) and Foundation for the Support of Science and Technology of Pernambuco.

Received on: 09/21/2025. Accepted on: 03/10/2026.

https://doi.org/10.5327/Z2176-94782799

Introduction

Water transposition systems between river basins are widely used strategies around the world to redistribute water resources from areas with greater availability to regions with water deficits (Bozorg-Haddad et al., 2024).

Inter-basin water transfer projects (IBWT) have significantly increased in number in recent decades due to the unremitting need to solve the problem of global water imbalance (Faúndez et al., 2022). IBWT involves the transport of water from one geographically distinct basin to another, balancing the distribution of water resources. Although the socio-economic benefits of implementing these projects are well recognized, little is known about the subsequent effects on the water quality of the receiving systems (Barbosa et al., 2021).

The Brazilian Northeast rivers are characterized by hydrological discontinuity, and therefore defined as non-perennial (Messager et al., 2021). This discontinuity can potentially affect the water quality both spatially and temporally in tropical semiarid watersheds in Brazil (Dantas et al., 2020).

Climate changes tend to intensify water scarcity in arid and semi-arid regions, making water management in these areas more challenging and directly influencing hydrological dynamics, causing impacts on ecosystems and society (Silveira et al., 2024). These changes have exacerbated droughts, for example, making them longer and more frequent, and with greater effects on water availability, especially in arid and semi-arid regions, although they are becoming more widespread across the country (Costa, 2025). In arid or semi-arid regions, the situation is even more alarming, since water resources are highly vulnerable and not readily accessible (Saadtpour, 2020).

In this sense, the Integration Project of the São Francisco River with the Hydrographic Basins of the Brazilian Northeast (PISF) aims to mitigate the serious problem of water scarcity, providing water security for the population residing in the receiving basins (Brasil, 2019). The project's water quality is of greatest importance as it impacts the health and water supply of approximately 12 million people across four Brazilian states.

The São Francisco River Transposition aims to capture water and direct it to two independent axes: North and East. With 270 km of canals, the North Axis will capture water from the São Francisco River, near the municipality of Cabrobó (PE), and will be used for the backlands of the states of Pernambuco, Rio Grande do Norte, Ceará, and Paraíba (Pires, 2019).

The São Francisco River crosses some of the driest parts of the Brazilian Semi-arid Region and brings life to ecosystems and to millions of people. Droughts are a recurrent problem in the basin. Since the second part of the twentieth century, several dams, irrigation projects, and water and sanitation systems have been built, and these developments have impacted river conditions, such as stream flow, sedimentation, silting, and water quality (Magalhães and Martins, 2021).

Monitoring water bodies over long periods and at multiple sampling stations produces a large and complex database covering various water quality parameters. This complexity makes it difficult to analyze and interpret the data, and to extract useful information for the proper management of water quality, resulting in the database often being underutilized (Trindade et al., 2017). Multivariate statistical methods are excellent exploration tools for interpreting this complex set of information and are frequently used in conjunction with trend analysis.

The behavior and impacts on the receiving basins are the subject of numerous studies, evaluations, and ongoing monitoring, both before and after the implementation of the PISF; however, these analyses have been conducted on a case-by-case basis. In this sense, there are no studies showing the temporal and spatial trends of water quality in the water bodies of the PISF. A robust database was created from nine parameters, 26 monitoring campaigns, in two axes, 13 river basins, and 86 sampling stations, totaling 523,224 observations. Later the data were analyzed, and multivariate statistical algorithms were used to generate comprehensive information of water quality tendencies of the basins.

This study is justified by presenting, for the first time, a historical series over 13 years, considering water quality monitoring data from the hydrographic basins included in the two axes of the project, between the years 2009 and 2022, showing the spatial-temporal evolution of water quality indices during this long-term period. To solve this problem, the Skater algorithm and MST statistical tools were used to reduce information and show the spatial distribution of the average WQI clusters for the monitoring points located on both North and East Axes of the PISF.

Access to water is one of the main challenges for global water security (Chloé et al., 2022). In Brazil, the water supply in semiarid regions depends largely on surface water accumulated in reservoirs, which are artificial ecosystems essential for the social and economic development of the region (Azevêdo et al., 2018). Water scarcity has frequently been pointed out as a key hindrance to regional development, especially during the recurrent droughts (Medeiros and Sivapalan, 2020).

Water security can be defined as the ability to offer water in sufficient quantity and quality to meet human needs, to carry out economic activities and to conserve aquatic ecosystems, at acceptable levels of risk, the implications of which are closely linked to food and nutritional security (Unesco, 2019). In Brazil, water security varies significantly between regions, with the Brazilian semiarid facing greater insecurity due to climatic factors, insufficient planning, and inadequate infrastructure (Cordão et al., 2020; Ramos Filho et al., 2023).

The study on water security in the area of influence of the PISF pointed to a risk of water scarcity for a total urban population of 4 million people, that the lack of water threatens 92% of the economic value of production in irrigated agriculture and 85% of industrial production. The adaptive capacity of the region is very limited by the fragile natural and artificial conditions of water reservation and management. Water quality also presents a critical environmental condition (Cerezini and Castro, 2023).

Usually, when water scarcity occurs, priority is allocated to human consumption and animal watering. This principle is also stated in the Brazilian "Water Law" (Federal Act 9,433/97). Therefore, any other type of use can be impacted in water scarcity cases (Thomaz et al., 2025).

The topic of water has received high visibility and attention on the global sustainability agenda. This is due to increasing pressure from factors such as economic development models, climate change, population growth and public health (Mehmood, 2019). Concurrently, climate change variability poses a substantial challenge to water security by intensifying the hydrological cycle (Tang et al., 2022).

Availability and easy access to safe and quality water is a fundamental human right and availability of clean water and sanitation for all has been listed as one of the goals to be achieved by the year 2030 for sustainable development by the United Nations General Assembly (UNGA) (United Nations, 2018). In this sense, water quality (WQ) assessment has become increasingly more important given the growing concerns over limited access to surface water resources (Das, 2025).

According to the National Water and Sanitation Agency (ANA, 2012), the water quality index (WQI) in the basins of the Northeast Atlantic Hydrographic Region were evaluated as excellent in 24% and as good in 71%, considering the sampled points monitored in the period from the years 2001 to 2010. Low rainfall levels in the period from 2012 to 2014 have meant that the main reservoirs in the basin have not received the expected volume of water for the São Francisco Hydrographic Region, which may be related to the severe drought that the Northeast region has been experiencing since 2012 (ANA, 2015). The sampling points in the Apodi rivers, near Mossoró, and in the Espinharas, a tributary of the Piranhas river were evaluated as regular, meanwhile the the sampling points in the Currais Novos river, in the state of Rio Grande do Norte were defined as poor. The other points monitored in the São Francisco River, in the stretch between the Sobradinho and Paulo Afonso reservoirs, presented WQI in the Good class. According to Caldas (2021), when comparing the results regarding the water quality index (WQI), before and after the arrival of the waters of the São Francisco River, the samples from the reservoirs in the receiving basin showed a "good" quality classification more frequently.

Water quality for natural uses is a component of the ecosystem dimension of the water security index - WSI (Cerezini and Castro, 2023). The Water Quality and Limnology Monitoring Program of the Basic Environmental Project 22 (PBA-22) was designed to meet the conditions presented in the PISF Preliminary License. The main objective was to develop and improve information supporting the formulation of environmental protection policies, as well as decision-making related to environmental management actions (MIDR, 2016). The WQI has been used to assess the environmental conditions of water bodies, their classification into use classes, and the management of river basins as a whole (Castro et al., 2023).

Recent international studies (2022–2026) in water quality assessment are increasingly adopting a "data-driven" paradigm, shifting from purely descriptive statistics to integrated, AI-enhanced, multivariate statistical approaches for interpretation of water quality data and spatial-explicit models to identify spatiotemporal patterns and pollution hotspots (Alssgeer et al., 2022; Begum et al., 2023; Lu et al., 2023; Wang et al., 2023; Ruckert et al., 2024; Shehab et al., 2025; Almegdadi et al., 2026).

Water quality monitoring can be done in a basin using statistical methods, such as regression analysis, clustering, forecasting and regionalization. Assunção et al. (2006) used the clustering algorithm, called the Minimum Spanning Tree (MST), which measures the similarity between points as a weighted average of proximity in geographic space and characteristics, as a method of spatial regionalization.

It should be noted that this study is aligned with the Sustainable Development Goals (SDGs), defined by the United Nations (2018) in the 2030 Agenda, to promote sustainable development. In this context, water management for the development of the region, through supply and access to basic sanitation, is part of SDG 6—Clean Water and Sanitation. In addition, the work is related to SDG 11—Sustainable Cities and Communities and SDG 14—Life Below Water.

The objective of this study is to analyze the temporal and spatial variability of the WQI in the river basins included in the PISF, in order to detect problematic locations of the indexes. This study will contribute to the understanding of the evolution and behavior of water quality indicators in the river basins included in the PISF, which may support management agencies in developing actions aimed at improving water quality and facilitate the necessary interventions to solve possible environmental problems.

Data and Methods

Study area

The PISF has two channels, the North and the East Axis. The North Axis, which is 260 km long, is collected from the São Francisco River, close to Assunção Island, in the municipality of Cabrobó, state of Pernambuco. The East Axis, which is 217 km long, is collected from the Itaparica reservoir, state of Pernambuco.

The North Axis canal was designed for a maximum capacity of 99 m^3/s and will operate with a continuous flow of 24.75 m^3/s, carrying water to the Salgado and Jaguaribe rivers, in Ceará; Apodi, in Rio Grande do Norte; and Piranhas, in Paraíba and Rio Grande do Norte. The East Axis canal was designed for a maximum capacity of 28 m^3/s and will operate with a continuous flow of 14 m^3/s, conveying water to the basins of the Pajeú and Moxotó rivers and to the Pernambuco backlands until reaching the Paraíba River, in the municipality of Monteiro, in the state of Paraíba (MIDR, 2025).

Monitoring water quality and limnology in the area of influence of PISF, is part of PBA-22, prepared based on the recommendations proposed by the Environmental Impact Study (EIS). Data on the nine water quality parameters were obtained in 26 collection campaigns, at 86 monitoring points, in the river basins included in the North and East Axes of the PISF, from the years 2009 to 2022. Figure 1 shows the spatial distribution of the points in the region.

WQI calculation

The WQI followed the National Sanitation Foundation (NSF) method, adapted by the Environmental Company of São Paulo (CETESB, 2024), based on nine (n) parameters: pH (p=0.12), biochemical oxygen demand ($BOD_{5,20}$, p=0.1), total nitrogen (TN, p=0.1), total phosphorus (on, p=0.1), temperature (T, p=0.1), turbidity (TUR, p=0.08), total dissolved solids (TDS, p=0.08), dissolved oxygen (DO, p=0.17) and thermotolerant coliforms (TC, p=0.15) (Equation 1).

The quality of the i-th parameter (qi), a number between 0 and 100, is obtained from the "average quality variation curves" or equations, as a function of its measured concentration. The weight corresponding to the i-th parameter (pi), a number between 0 and 1, is assigned according to its importance for the overall quality conformation, with the parameters with the greatest weight being DO and Coliforms. The WQI classification for raw water ranges on a scale from 0 to 100, according to the following class limits: Poor (WQI≤19), Bad (19<WQI≤36), Acceptable (36<WQI≤51), Good (51<WQI≤79) and Great (79<WQI≤100). The WQI results were organized by axis to better identify flow and receiving basin characteristics.

$$WQI = \prod_{i=1}^{n} {q_i}^{p_i} \quad (1)$$

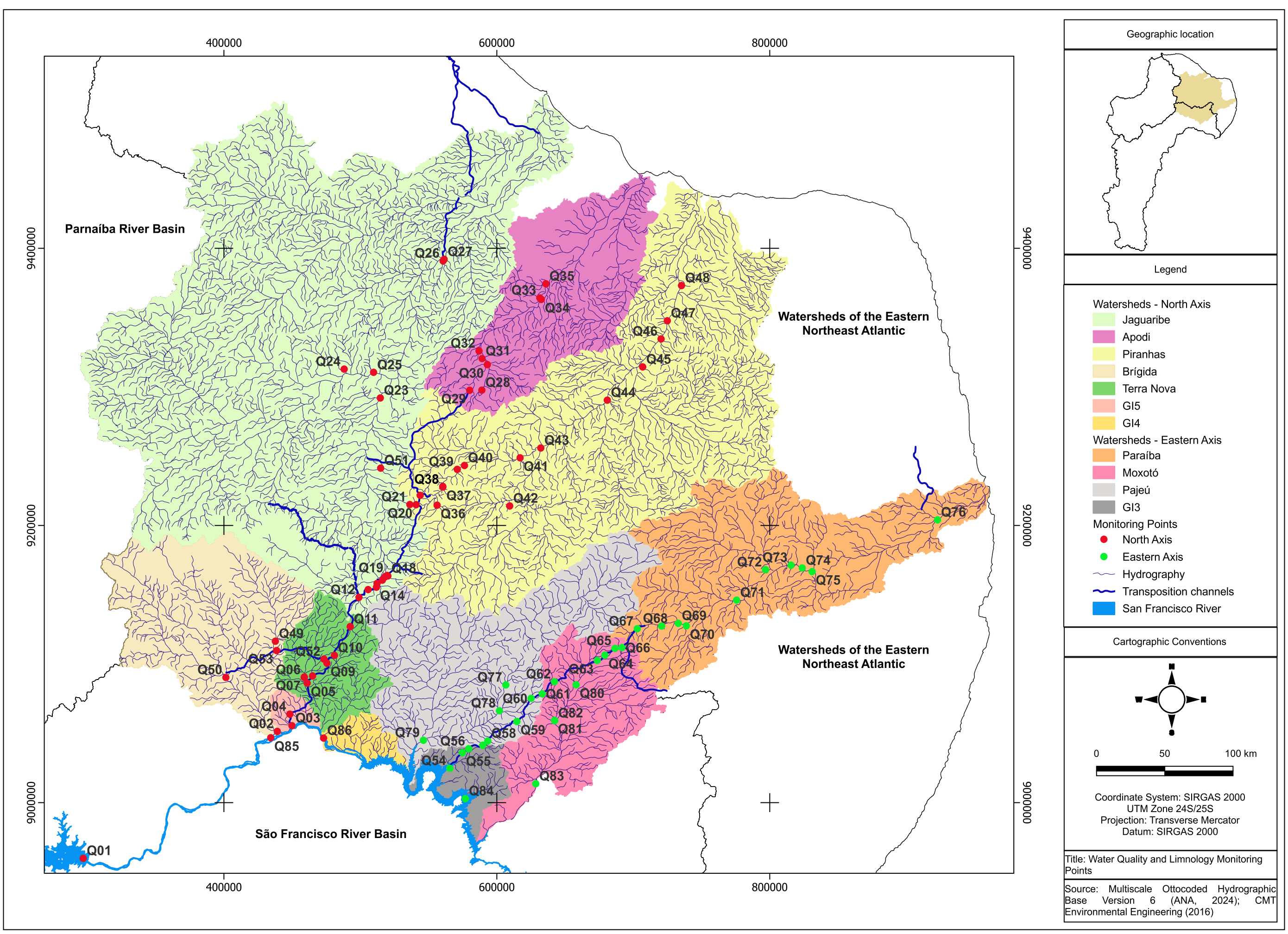


Figure 1 – Water quality monitoring points in the basins of the East and North Axes of PISF.

SKATER algorithm

The Spatial 'K'luster Analysis by Tree Edge Removal, is an algorithm based on the removal of unnecessary parts of the decision tree, which reflects the contiguity structure between observations (Assunção et al., 2006), with the purpose of reducing the complexity of the model, avoiding overlaps and improving efficiency.

Initially, it is necessary to generate the dissimilarity matrix that contains weights for contiguous observations W (Equation 2), with dimension n x n, expressing the neighborhood structure between observations in which the elements wij are different from zero when i and j are neighbors, and zero otherwise (Anselin, 2020). The self-neighborhood relation is deleted, so the diagonal elements of W are zero, wii=0. The W matrix expresses the existence of a neighborhood relation as a binary relation, with weights defined as 1 and 0. Each spatial unit is represented in the matrix by a row i, and the potential neighbors by the columns j, with j i. The existence of a neighborhood relationship between the spatial unit corresponding to row i and a corresponding column j. Thus, it can be said that the pairs are contiguous when wij=1.

$$W = \begin{bmatrix} w_{11} & w_{12} & \cdots & w_{1n} \\ w_{21} & w_{22} & \cdots & w_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ w_{n1} & w_{n2} & \cdots & w_{nn} \end{bmatrix} \tag{2}$$

The matrix is represented as a graph with observations as nodes and contiguity relations as edges. The proximity between contiguous neighboring points is given by the Euclidean distance dij between i and j, for vectors xi and xj, in the set of observations n, according to Equation 3.

$$d_{ij} = d(x_i, x_j) = \sum_{l=1}^{n} (x_{il} - x_{jl})^2 \tag{3}$$

The complete graph is reduced to a minimum spanning tree (MST) such that there is a path connecting all observations (nodes), but each node is only visited once. In other words, the n observations (nodes) are connected by n - 1 edges, so that the overall dissimilarity between nodes is minimized. After the MST is formed, SKATER performs a recursive partition of the MST to obtain a cluster. The result of clustering is the homogeneity of variables within the group. To form a partition, n - 1 edges are removed from the MST so that the overall dissimilarity between nodes is minimized.

This produces an initial sum of squared deviations between clusters, SDQ (Equation 4), where x_{ij} is the j-th attribute of the spatial object (node value), $\bar{x}_j$ is the average value of the jth attribute for all objects in the k-tree (overall node average), m is the number of attributes considered in the analysis, n_k is the number of spatial objects in the tree. The overall dissimilarity between nodes, in all partitions (Π), in T trees, should be minimized. Equation 5 represents the value associated with the partition quality, when $Q(\Pi)$ there is a value associated with the quality of a partition Π.

$$SQD = \sum_{j=1}^{m} \sum_{i=1}^{n_k} (x_{ij} - \bar{x}_j)^2 \tag{4}$$

$$Q(\Pi) = \sum_{i=0}^{k} SQD \tag{5}$$

In each iteration, an edge n is removed that will divide tree k into two trees from a set, dividing a connected region into two sub-regions. In the process, the edge that provides the greatest increase in the overall quality of the resulting clusters is selected, given by the objective function (Equation 6), where the arrangement S_l^T is produced by removing the edge L in the T tree, and the two trees T_a and T_b are produced by dividing T after cutting edge L. At each iteration $f_l(S_l^T)$, the number of clusters is increased until it reaches an optimal solution $f_l(S_*^{T_i})$ until a predetermined number of groupings is reached. Spatial modeling was performed using the GeoDa program (2025).

The number of clusters is a critical step in cluster analysis to ensure that the groupings are meaningful, stable, and not merely reflecting noise or overfitting. The rationale is to choose the point where the rate of decrease of inertia (WCSS) changes significantly, forming an "elbow" in the graph. This indicates that adding more clusters does not substantially improve the data modeling. The rationale for the number of clusters is that the number of groupings is related to the quantity of points and the range of variation of the points. As there were 31 points on the East Axis and 41 points on the North Axis, which presented maximum, median, and minimum IQA values, it was decided to include three more clusters initially, plus two more clusters that could represent extreme values exceeding the second and third quartiles, totaling five clusters. The five groups also facilitate the correlation of the results of the box-plot analysis presented, in an attempt to provide synergy and better understanding of the results.

Water resources planning and management require complete data sets of a number of hydrological variables. However, hydrologists are often faced with the issue of missing data in hydrological databases, remaining a major challenge in many data-scarce regions like the Brazilian semiarid. Hydrological studies require both spatially and temporally continuous datasets for model calibration and validation (Aissia et al., 2017).

Minimum Spanning Tree (MST) cluster analysis is a robust unsupervised learning technique, ideal for identifying groups with irregular shapes and varying densities, frequently used in Hydrology and Geosciences. However, the application of MST in hydrology scenarios, especially with dry reservoir data (lack of data/missing data), imposes significant limitations such as selection bias and disconnection, imprecision in distance measurement, inability to process faults and inconsistency in the edges. The implications for the results are:

- Underestimation of Variability: If dry reservoirs are imputed with a value of zero, the model may group them together (false "drought" cluster), masking real hydrological differences;
- Increased Uncertainty: "Breaking" the MST (edge trimming) in areas of missing data can lead to an incorrect number of clusters, ignoring the true characteristics of the region;
- Poor "Recognition" Results: If data density is very low (e.g., extreme drought), MST may not be able to identify the forms/groups that persist;
- Impact on Regionalization: Hydrological regionalization (definition of homogeneous basins) becomes less reliable if the database is discontinuous, generating errors in flow forecasting for dry areas.
- To mitigate these problems, technical articles suggest using data imputation methods before clustering, such as:
- Imputation Techniques (Filling): The use of methods such as nearest neighbors (KNN) or similarity-based models (such as correlation/distance) to fill in dry reservoir data based on neighboring regions;
- Partial Distance Evidential Clustering (PEC): A technique that models the imprecision generated by missing data, avoiding incorrect imputations;
- Use of R-MST (Representative-MST): Selection of representative points to build the tree, mitigating the effect of sparse data.

Spatial modeling was performed with the GeoDa program (2025) (Equation 6).

$$f_l(S_l^T) = SQD_T - (SQD_{Ta} + SQD_{Tb})$$

Results and Discussion

Data on water quality parameters obtained from several campaigns in the basins included in the PISF, between 2009 and 2022, were used to calculate the WQIs. The averages of the maximum, mean and minimum values are presented in the Supplementary Material for the East Axis and North Axis, respectively. The medians and interquartile ranges indicate a very similar pattern of mean annual WQI between the two axes. As shown in Figure 2, the interquartile range was greater on the North axis (31.25) than on the East axis (22.72), indicating greater dispersion of the data.

The medians on the North and East axes were 62.9 and 68, respectively, indicating a slightly better condition in the latter, but nevertheless, the WQI was still placed in the good class. On the North axis, the first quartile which represents the value that separates the 25% smallest data, corresponds to WQI, which falls into the regular class. Meanwhile, on the East axis, the WQI remains in the good class. On the North axis, the first quartile, the WQI falls into the regular class, while on the East axis, the WQI remains in the good class.

The monitoring of 33 points on the East Axis and 40 points on the North Axis of the PISF showed that there was no statistically significant difference in the average WQI between the East and North Axes of the PISF, observed from the years 2009 to 2022. The medians were 68 (East) and 62.9 (North), and the interquartile ranges were 22.7 (East) and 31.3 (North).

Spatial modeling performed on each axis identified five WQI clusters. In the East Axis, the highest average WQIs were located in the Itaparica reservoir, located in the group of small inland river basins (GI3), and also in the São Francisco River basin (82). The lowest WQI occurred in the Paraíba River basin (54.79). In the North Axis, the highest values (74.39) occurred in the São Francisco River, the donor basin, in the Brígida, GI4, GI5 basins, in the Serra do Livramento reservoir (TN) and in the São Gonçalo and Coremas Mãe d'Água reservoirs, in the Piranhas River basin. The lowest values occurred at the headwaters of the Apodi River (44.1), in the backwaters of the Boa Vista and the Engenheiros Ávidos Reservoirs, and in the Piranhas River Basin (46.4).

Spatially restricted cluster modeling, as proposed by Assunção et al. (2006) and Novianta *et al.* (2025), was performed separately for the East and North Axes. A summary of the results from the axis cluster modeling is presented in Table 1. Five clusters were determined with average, minimum and maximum WQIs values. On the East Axis, the total sum of squared deviations was 75, the sum of squares deviations (SSD) between the clusters was 48.782, and the relationship between them was 0.6504. On the North Axis, the total sum of squared deviations was 117, the SSD between the clusters was 85.257, and the relationship between them was 0.729. The SSD between clusters was smaller than the total SSD, indicating that the number of clusters was appropriate to represent the homogeneous regions of the WQI on both axes. The results were considered satisfactory for the optimization of the graph for the number of proposed clusters; according to Anselin (2020).

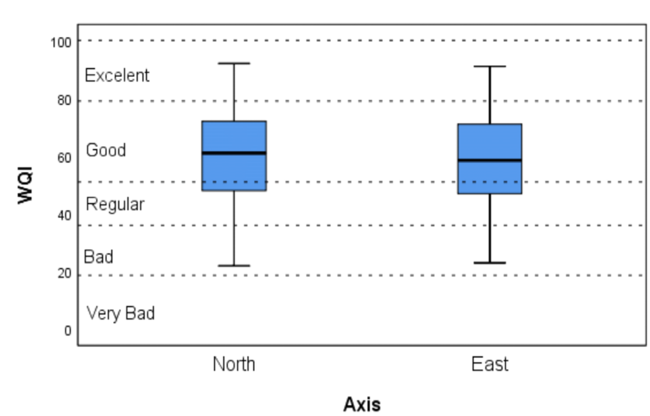


Figure 2 – Boxplot of the variation in the WQI in the East and North Axes, from 2009 to 2022.

The spatial distribution of the WQI cluster along the East Axis is shown in Figure 3. Of the 26 monitored points, seven were placed in group C1: Q69, Q71, Q72, Q73, Q74, Q75 and Q76, with lower average values of minimum (39.29), average (54.79) and maximum (71.57) WQI, all of which are located in the Paraíba River basin (Table 2 and Figure 3).

Twelve points were placed in group C2: Q63, Q64, Q65, Q68, Q70, Q77, Q78, Q79, Q80, Q81, Q82 and Q83, with mean values of minimum (41.42), medium (60.61) and maximum (79.83) WQI (Table 3 and Figure 3).

Cluster C3 presented five points: Q54, Q55, Q56, Q57 and Q59, all with average values of minimum (52.25), medium (63.08) and maximum (75.25) WQI (Table 4 and Figure 3).

Table 1 – Summary of WQI clustering analyses using the SKATER algorithm, for sampling points located on the East and North Axes of the PISF.

Groupment	East Axis WQI			North Axis WQI		
	max	avg	min	max	avg	min
C1	71.57	54.78	39.29	72.20	50.40	28.60
C2	79.20	60.66	36.00	79.43	55.31	32.29
C3	75.25	63.08	52.25	71.00	55.91	36.29
C4	82.67	66.03	52.33	79.75	64.17	42.00
C5	85.00	66.80	55.50	86.25	74.39	60.25
Total square sum	60			114		
Sum of squares between clusters	23.1893			68.8675		
Relationship between the total sum of squares	0.3865			0.6041		

Max: maximun; avg: average; min: minimum.

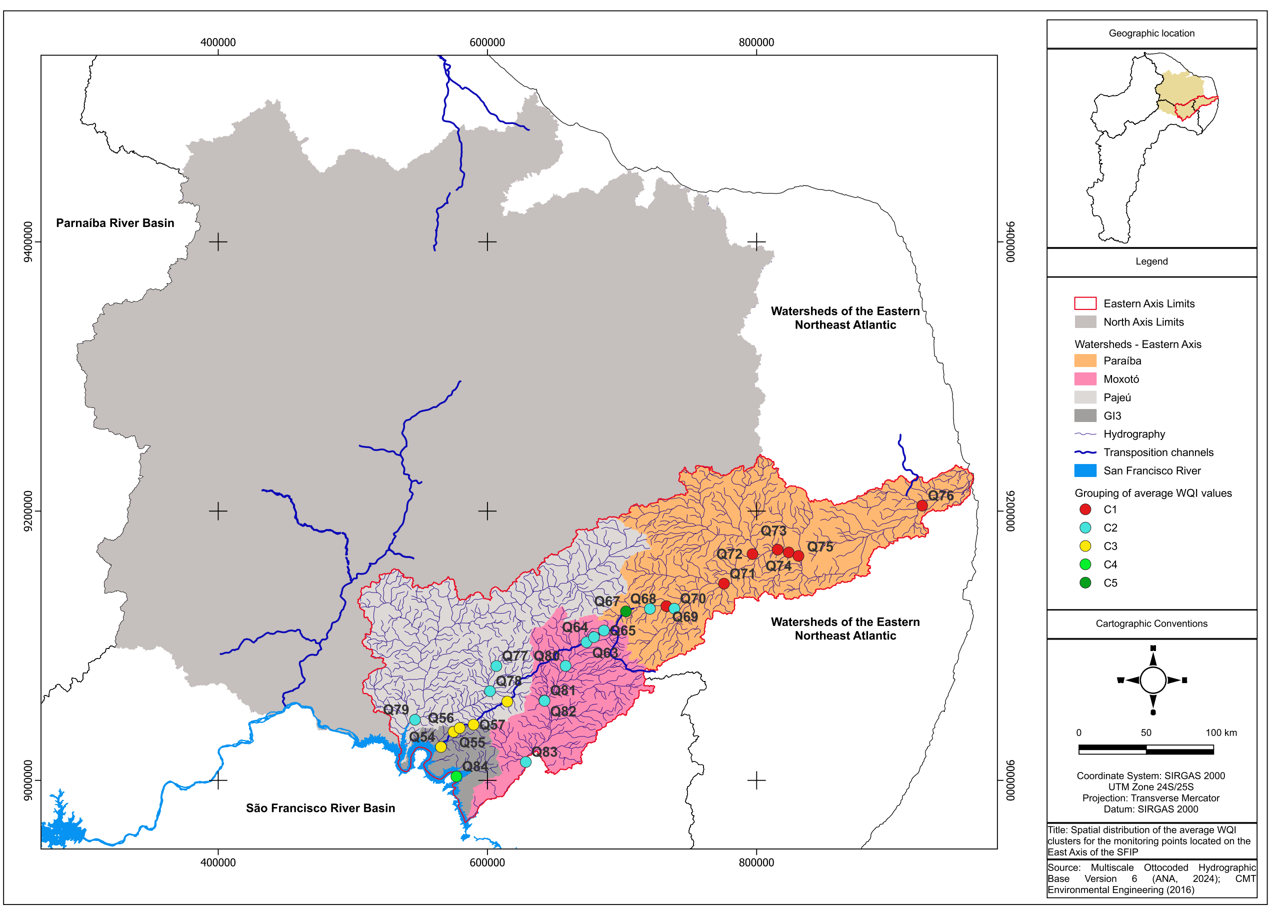


Figure 3 – Spatial distribution of the average WQI clusters for monitoring points located on the Eastern Axis of the PISF.

One point was grouped in group C4: Q84, with mean WQI values of minimum, medium, and maximum of 68, 81.8, and 94, respectively. In cluster C5, point Q67 was classified, with mean values of minimum, medium, and maximum WQI values of 82. It should be noted that this sampling point only had water in one campaign. (Table 5 and Figure 3).

Spatial distribution of the WQI cluster along the North Axis is shown in Figure 2. Out of the 40 monitored points, two presented lower average values than the minimum (29), average (44.1), and maximum (65.5) WQI and were grouped in group C1: Q31 and Q32 (Table 6 and Figure 4).

Table 2 – Spatial distribution of the average WQI of Group C1 at the monitoring points located on the East Axis of the PISF.

Grouping C1 (East Axis)					
Basin	SP	Location	WQI		
			max	avg	min
PB	Q69	Upstream of Camalaú Dam	71.57	54.79	39.29
	Q71	Do Meio River			
	Q72	Backwater of Epitácio Pessoa Dam			
	Q73	Epitácio Pessoa Dam			
	Q74	Downstream of Epitácio Pessoa Dam			
	Q75	Paraíba River			
	Q76	Downstream of Acauã Dam			

SP: Sampling point; PB: Paraíba; max: maximum; avg: average; min: minimum.

Table 3 – Spatial distribution of the average WQI of Group C2 at the monitoring points located on the East Axis of the PISF.

Grouping C2 (East Axis)					
Basin	SP	Location	WQI		
			max	avg	min
PJ	Q79	Pajeú River – after do Navio Stream	79.83	60.61	41.42
RN	Q77	Do Navio Stream			
	Q78	Barra do Juá Dam			
MX	Q63	Moxotó Reservoir			
	Q64	Barreiros Reservoir			
	Q65	Campos Reservoir			
	Q80	Affluent of Moxotó River			
	Q81	Poço da Cruz Dam- Axis			
	Q82	Moxotó River – downstream of Poço da Cruz Dam			
	Q83	Moxotó River			
PB	Q68	Acauã Dam			
	Q70	Camalaú Dam			

SP: Sampling point; PJ: Pajeú; RN: Riacho do Navio; MX: Moxotó; PB: Paraíba; max: maximum; avg: average; min: minimum.

Table 4 – Spatial distribution of the average WQI of Group C3 at the monitoring points located on the East Axis of the PISF.

Grouping C3 (East Axis)					
Basin	SP	Location	WQI		
			max	avg	min
PJ	Q59	Muquém Reservoir	76.0	74.94	74.40
GI3	Q54	Itaparica Reservoir			
	Q55	Areias Reservoir			
	Q56	Braúnas Reservoir			
	Q57	Mandantes Reservoir			

SP: Sampling point; PJ: Pajeú; GI3: Group of small inland river basins 3; max: maximum; avg: average; min: minimum.

Two points were placed in group C2, with mean WQI values of minimum, medium, and maximum values of 38, 46.4, and 57.5, respectively: Q21, Q36, Q05, Q06, Q07, Q49, Q50, Q13, Q14, and Q24 ~~and~~ (Table 7 and Figure 4).

Nineteen points were placed in group C3, with mean WQI values of minimum, medium, and maximum of 33.68, 57.22, 78.21, respectively: Q06, Q07, Q13, Q14, Q23, Q24, Q28, Q29, Q30, Q37, Q38, Q39, Q41, Q43, Q44, Q45, Q46, Q49, and Q50 (Table 8 and Figure 4).

Table 5 – Spatial distribution of the average WQI of Groups C4 and C5 at the monitoring points located on the East Axis of the PISF.

Grouping C4 (East Axis)					
Basin	SP	Location	WQI		
			max	avg	min
GI3	Q84	Itaparica Reservoir (Nova Petrolândia)	94.00	81.80	68.00
PB	Q67	Monteiros River – Upstream of Poções Dam	82.00	82.00	82.00

SP: Sampling point; GI3: Group of small inland river basins 3; PB: Paraíba; max: maximum; avg: average; min: minimum.

Table 6 – Spatial distribution of the average WQI of Group C1 at the monitoring points located on the North Axis of PISF.

Grouping C1 (North Axis)					
Basin	SP	Location	WQI		
			max	avg	min
AP	Q31	Pau dos Ferros Dam- axis	65.5	44.1	29.0
	Q32	Apodi River – Pau dos Ferros municipality/RN			

SP: Sampling point; AP: Apodi; RN: Rio Grande do Norte; max: maximum; avg: average; min: minimum.

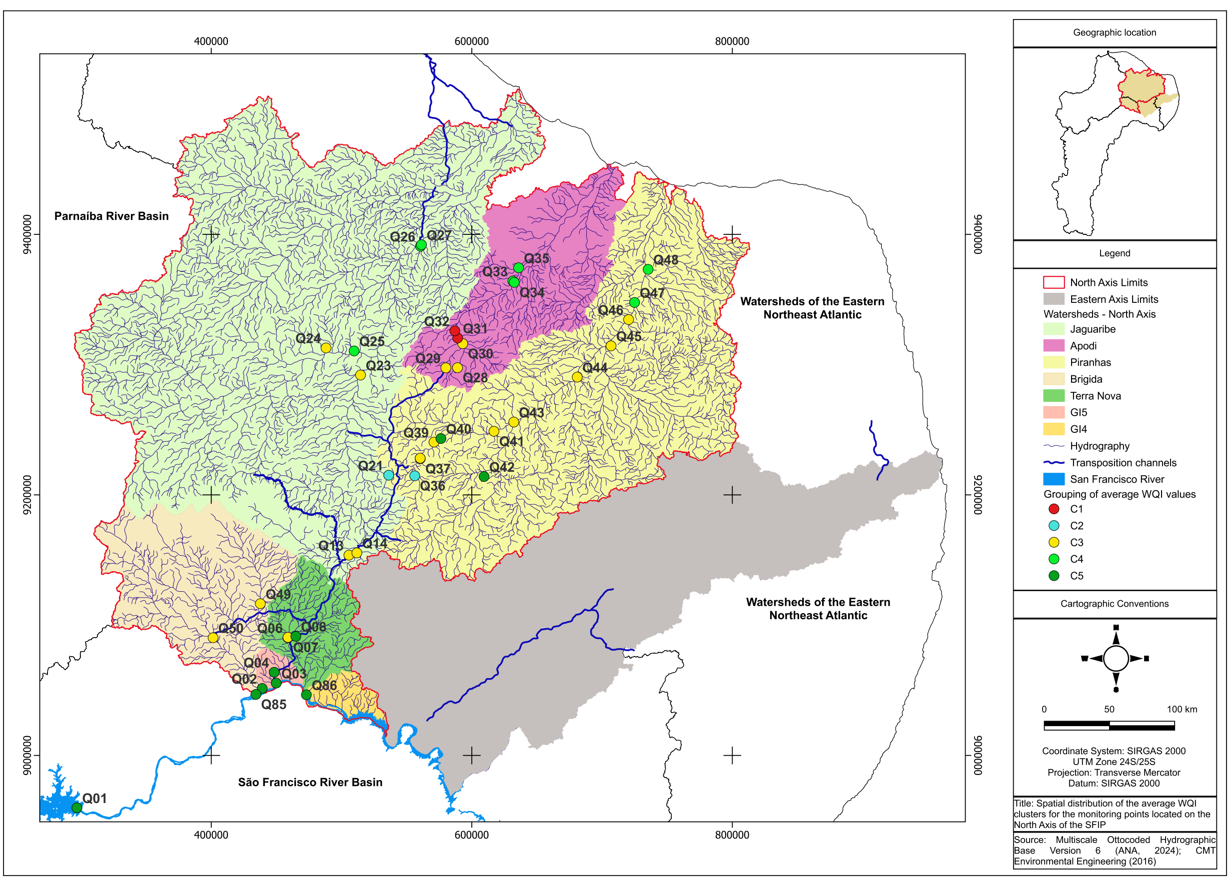


Figure 4 – Spatial distribution of the average WQI clusters for the monitoring points located on the North Axis of the PISF.

Cluster C4 comprises eight points with mean WQI minimum, medium, and maximum values of 44.13, 66.56 and 79,88, respectively: Q25, Q26, Q27, Q33, Q34, Q35, Q47, and Q48 (Table 9 and Figure 4).

Eight points were placed in group C5, with mean WQI of minimum, medium, and maximum values of 60.25, 74.31 and 86.25, respectively: Q01, Q02, Q03, Q04, Q08, Q40, Q42, Q85, and Q86 (Table 10 and Figure 4).

Table 11 presents the classification of the groups into WQI classes; according to CETESB (2024). In the East Axis channel, the lowest values were found for the minimum WQI of 39.29 and for the average WQI of 54.79 in seven points located at the Paraíba River Basin, corresponding to the good class for average WQI and regular for minimum WQI. In Riacho do Navio, the Pajeú River, the Moxotó River, the reservoirs installed in the area, and the Acauã and Camalaú dams, both in the Paraíba River Basin, the minimum WQI was classified as regular.

In the North Axis (Table 11), the two points located at the head of the Apodi River Basin, placed in group C1, were classified in the regular class, considering the average WQI, and in the poor class, considering the minimum WQI. For the other groups with an average WQI, all points fell into the good class. Considering the minimum WQI, groups C2 and C4 were classified in the regular class and were located in the basins of the Piranhas, Apodi, and Jaguaribe Rivers. In groups C1 and C3, the WQI was classified as poor for many points in all basins; however, two points located in the Apodi River Basin, in the Pau dos Ferros

Table 7 – Spatial distribution of the average WQI of Group C2 at the monitoring points located on the North Axis of PISF.

Grouping C2 (North Axis)					
Basin	SP	Location	WQI		
			max	avg	min
PR	Q21	Boa Vista Reservoir –upstream of the backwater of the dam	57.5	46.4	38.0
	Q36	Engenheiros Ávidos Dam - Backwater			

SP: Sampling point; PR: Piranhas; max: maximum; avg: average; min: minimum.

Table 8 – Spatial distribution of the average WQI of Group C3 at monitoring points located on the North Axis of the PISF.

Grouping C3 (North Axis)					
Basin	SP	Location	WQI		
			max	avg	min
TN	Q06	Terra Nova Reservoir	78.21	57.22	33.68
	Q07	Terra Nova River – downstream of Terra Nova reservoir			
BG	Q49	Chapéu Dam			
	Q50	Entremontes Dam			
JG	Q13	Jaguaribe River – Backwater of Atalho Dam			
	Q14	Atalho Dam			
	Q23	Salgado River - Icó Municipality, CE			
	Q24	Orós Dam			
AP	Q28	Angicos Dam - Backwater			
	Q29	Angicos Dam			
	Q30	Pau dos Ferros Dam - Backwater			
PR	Q37	Engenheiro Ávidos Dam			
	Q38	Engenheiro Ávidos Dam – downstream			
	Q39	Piranhas River - Backwater of São Gonçalo Dam			
	Q41	Piranhas River – São Domingos do Pombal Municipality, PB			
	Q43	Piancó River – Backwater of Piranhas River			
	Q44	Paraíba River – Border between PB/RN			
	Q45	Paraíba River – Oiticica II			
	Q46	Açu River - Backwater			

SP: Sampling point; TN: Newfoundland; BG: Brígida; JG: Jaguaribe; AP: Apodi; PR: Piranhas; max: maximum; avg: average; min: minimum.

Reservoir (Q31) and in the Apodi River (Q32), in the municipality of Pau dos Ferros, RN, presented lower values.

Spatial modeling highlighted a significant reduction in the WQI in the extreme clusters. In the East Axis, the reduction in the WQI between the C1 (lower Paraíba River) and C5 (Monteiros River, upstream of the Poções reservoir, in the same basin) groups was 49.7% for the average value and 109% for the minimum value. In the North Axis, the reduction in the WQI between C5 (São Francisco River, PISF catchment, and mouth of the Brígida River) and C1 (headwaters of the Apodi River) was 68.7% for the average value and 105% for the minimum value.

The equation used to calculate WQI uses exponents with the weights of the parameters ranging from 0 to 1. High concentrations of certain parameters resulted in lower values of the associated com-

Table 9 – Spatial distribution of the average WQI of Group C4 at monitoring points located on the North Axis of the PISF.

Grouping C4 (North Axis)					
Basin	SP	Location	WQI		
			max	avg	min
JG	Q25	Jaguaribe River - upstream of the confluence with Salgado River	79.88	66.56	44.13
	Q26	Castanhão Dam – center			
	Q27	Castanhão Dam – axis			
AP	Q33	Apodi River - Backwater of Santa Cruz Dam			
	Q34	Santa Cruz dam			
	Q35	Apodi River - Pedra de Abelhas (Apodi Swamp)			
	Q47	Armando Ribeiro Gonçalves Dam - center			
	Q48	Armando Ribeiro Gonçalves Dam – axis			

SP: Sampling point; JG: Jaguaribe; AP: Apodi; max: maximum; avg: average; min: minimum.

Table 10 – Spatial distribution of the average WQI of Group C5 at monitoring points located on the North Axis of PISF.

Grouping C5 (North Axis)					
Basin	SP	Location	WQI		
			max	avg	min
SF	Q01	Sobradinho Reservoir	86.89	74.39	59.44
	Q85	São Francisco River in Orocó/PE			
	Q03	São Francisco River - capture of the North Axis			
BG	Q02	Brígida River - Mouth of the Brígida River			
GI5	Q04	Tucutu Reservoir			
TN	Q08	Serra do Livramento Reservoir			
PR	Q40	São Gonçalo Dam – axis			
	Q42	Coremas – Mãe d’água Dam - axis			
GI4	Q86	São Francisco River in Ibó/PE			

SP: Sampling point; SF: São Francisco; BG: Brígida; GI5: Group of small inland river basins 5; TN: Terra Nova; PR: Piranhas; GI4: Group of small inland river basins; max: maximum; avg: average; min: minimum.

Table 11 – Classification of groups in WQI classes, at the East and North Axes.

Grouping	East Axis				North Axis			
	WQI avg	Class	WQI min	Class	WQI avg	Class	WQI min	Class
C1	54.79	Good	39.29	Regular	44.10	Regular	29.00	Very bad
C2	60.61	Good	41.42	Regular	46.40	Good	38.00	Regular
C3	74.94	Good	74.40	Good	57.22	Good	33.68	Very bad
C4	81.80	Excellent	68.00	Good	66.56	Good	44.13	Regular
C5	82.00	Excellent	82.00	Excellent	74.39	Good	59.44	Good

Min: minimum; avg: average.

ponents. For example, Q41, a point located on the North Axis, at the headwaters of the Piranhas River, showed average values of total phosphorus (2.91 mg/L), TDS (259.3 mg/L), and turbidity (57.3 UNT), which resulted in the calculated average WQI equal to 54.39 (Suppl. Mat.). Of these parameters, only total phosphorus (≤0.1 mg/L for lotic environments) exceeded the water quality standard limit for class 2 (Brasil, 2005).

Total phosphorus is an important indicator of the eutrophication process in aquatic bodies, and its concentration, together with that of nitrogen, is decisive for algal growth and cyanobacterial blooms (Costa et al., 2016). High total phosphorus concentrations are related to sewage discharges, drainage of agricultural and urban areas, and geological and soil formations (Castro et al., 2023). However, at the aforementioned point, the average values of BOD (28.8 mg/L O_2), Coliforms (1,790.8/100 mL H_2O), and total nitrogen (2.13 mg/L) exceeded the standard limit for class 2, but were not decisive for the reduction of the average WQI calculated for the point. High concentrations of these parameters corroborate the hypothesis of domestic sewage discharge (Von Sperling, 1996).

For several decades, numerous environmental stressors (both anthropogenic and natural factors) have caused significant deterioration of surface water quality. These include, but are not limited to, rapid proliferation of urbanization, industrialization, agricultural growth, and natural phenomena that cause the constant discharge of effluents, inorganic substances, and contaminants into surface water sources (Javed et al., 2017; Uddin et al., 2018; Cook et al., 2020). In this study, the field team reported the occurrence of areas intended for agricultural cultivation, pastures, silt strips, sand extraction, residences, sewage discharge and waste discharge, among others, on the banks of the reservoirs.

Siegmund-Schultze et al. (2018) stated that the increasing use of water and adjacent land constitutes a major potential threat to water quality. In scenarios of water scarcity, the loss of quality becomes even more critical, because, in addition to the low levels of availability, pollution by contaminants from human activities, such as irrigated agriculture and the release of untreated wastewater, can compromise domestic supply, which should be a priority.

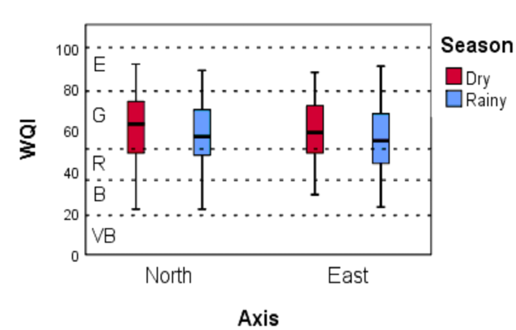
The implementation of artificial reservoirs has significant impacts on the physical, biotic, and socioeconomic environments and raises a series of issues related to planning and management actions. Changes in land use can decisively influence water quality and sediment input, compromising the useful life of reservoirs, often undermining the purpose for which they were built. Thus, the future of reservoirs depends on their initial characteristics (preserved areas, degraded soils, water quality, and biodiversity) and the way in which multiple uses of water and its surroundings occur. In general, artificial reservoirs attract economic development and promote a reorganization of local and regional systems, which, in turn, also brings greater risks of soil degradation and water quality (Tundisi et al., 2008).

The constituent parameters of the WQI may also vary according to the season. During the rainy season, surface runoff increases when the soil is waterproofed, carrying nutrients and organic matter to the water bodies. Figure 5 shows the seasonal variations in the WQI values during the dry and rainy seasons. On both axes, the quartiles were slightly higher in the dry season, indicating that the WQI worsened in the rainy season~~,~~ but without a change in class between the seasons. The WQI remained in the good class for the medians, and in the regular class for the first quartile~~,~~ in both seasons on both axes. During the dry season, many rivers ran dry, and some reservoirs had very low water volumes, preventing the quantification of water quality parameters. This may have distorted the representativeness of the WQI, as the lack of water reduces the dilution effect of chemical substances, increasing their concentration and, consequently, reducing the WQI.

The North Axis came into operation in February 2022, and the East Axis in April 2017 (MIDR, 2025). The monitoring data used in this study covered the period from the years 2009 to 2022. The flows provided by the PISF cover a short period of the historical series of water quality data. Therefore, it is not possible to determine the influence of the project flows on the WQI. However, there is a tendency for improvement, as spatial modeling has shown that the water in the São Francisco River is of much higher quality than that in the receiving basins and could contribute to improving the WQI in the future.

The WQI expresses water quality through a single value (CETESB, 2024) indicating the environmental conditions at the monitoring point in a classificatory manner. This may lead to incorrect interpretations of the parameters contributing to lower WQI and, consequently, the sources of pollution. To identify the interrelationships between the parameters, principal component analysis — PCA (Castro et al., 2023) should be applied; however, this was not the objective of this study.

Castro et al. (2023) pointed out several applications of WQI used for monitoring pollution sources and managing land use and occupation in river basins, aiming to improve water quality management.



VB: Very Bad; B: Bad; R: Regular; G: Good; E: Excellent.

Figure 5 – Seasonal variations in the WQI on the North and East Axes during the dry and rainy periods.

It is the main WQI used in Brazil to evaluate raw water for use in public supply after treatment (ANA, 2025). Spatial modeling assists management as it is a tool that allows regionalization of critical water quality regions, as presented by Novianta *et al.* (2025). One of the applications would be the identification of pollution sources and support for the implementation of environmental programs to mitigate adverse effects.

Spatial modeling performed on each axis identified five WQI clusters. In the East Axis, the highest average WQIs were located in the Itaparica reservoir and the lowest WQI occurred in the Paraíba River basin. In the North Axis, the highest values occurred in the São Francisco River, which is the donor basin. The lowest values occurred at the headwaters of the Apodi and Piranhas River Basins.

Based on the average WQI values, all monitored points on both axes presented a good class. Only the lowest group, located on the Apodi River, presented a regular class. Spatial variation of the average WQI between the extreme clusters was greater on the North Axis than on the East Axis.

In the dry season, the WQI presented a median and third quartile slightly higher than in the rainy season on both axes, indicating worsening water quality in the rainy season. Since the water from the São Francisco River is of superior quality, it is expected that this will improve the indicators in the receiving basins. However, this trend was not confirmed in this study, possibly due to the short period of operation of the PISF.

Water quality modeling allows for better decision making, more technically robust solutions among alternative possibilities for water quality management. Models are required to determine better alternatives for solving sustainable water quality problems in the long term. In addition, models are essential to provide a basis for economic analysis, and decision-makers can use the output to assess the environmental implications of a project and the cost–benefit ratio (Liu, 2018). There are several methods for water quality assessment, including single-factor, multi-index, fuzzy mathematics, gray system evaluation, artificial neural networks, multi-criteria analysis, geographical interpolation, and multivariate statistical approaches (Dixon and Chiswell, 1996; Rakotondrabe et al., 2018; Aires and Salgado, 2024; Souza et al., 2024; Singh et al., 2025).

Water quality models have been applied to assess water quality in various waterbodies. The temporal patterns, seasonal variations, and long-term trends in water quality parameters can be deciphered more accurately using Multivariate Analysis (MVA); and hence, an increasing number of statistical studies have used MVA for data analysis and interpretation (Malsy et al., 2017; Njugana et al., 2020; Patil et al., 2020; Fatima et al., 2022). These studies help diagnose and compare the quality of water sources and are important for the planning and management of waterbodies.

Nong et al. (2020) evaluated water quality in the South-to-North Water Diversion Project of China (SNWDPC) using the WQI method to evaluate the seasonal and spatial water quality changes during the monitoring period. A new WQImin model consisting of five crucial parameters, TP, *F. coli*, Hg, WT, and DO, was established using stepwise multiple linear regression analysis. The results demonstrated that the water quality status of the Middle Route of the SNWDPC has been steadily maintained at an "excellent" level during the monitoring period, with an overall average WQI value of 90.39 and twelve seasonal mean WQI values ranging from 87.67 to 91.82.

When compared to the traditional WQI, the study of Deng *et al.* (2022), the WQImin model, with the assistance of stepwise linear regression analysis, could exhibit more accurate explanation with the coefficient of determination (R2) and percentage error (PE) values being 0.895 and 5.515%, respectively. The proposed framework is of great importance to improve the spatiotemporal recognition of water quality patterns and further aid in developing efficient water management strategies at a reduced cost.

Xu et al. (2023) compared the differences in water quality in the Middle Route of China's South-to-North Water Diversion Project (MRP) between the initial stage (Nov. 2015 to Oct. 2017, low transfer volumes) and the current stage (Nov. 2017 to Oct. 2020, high transfer volumes), and the spatiotemporal water quality variations in the current stage were evaluated using multivariate statistical methods. For this purpose, approximately 12,528 observations, including the datasets of 12 water quality parameters collected from 29 monitoring sites, were used. The results showed that the water quality status improved significantly during the current stage.

Conclusions

The contribution of this study is justified by presenting, for the first time, a historical series over 13 years. It considered water quality monitoring data from the hydrographic basins included in the two axes of the transposition project, between 2009 and 2022, showing the spatiotemporal evolution of water quality indices during this long-term period. The spatial variation identified in each axis is useful for establishing sustainable management and planning programs for the water bodies under study.

The spatiotemporal analysis carried out in the São Francisco River Basin between 2009 and 2022 revealed a panorama of intense transformations in land use and occupation, which, along with the typical climatic effects of the semi-arid region, caused relevant environmental impacts on abiotic and biotic indicators. In both the North and East Axes, the WQI of the basins ranged from Excellent to Poor during the study period. Notably the best WQI corresponded to the São Francisco River Basin, given that it is defined as the water-donating basin. No sampling points were classified as Very Poor in either the North or East Axis during the study period.

It was concluded that the use of WQI, associated with statistical and spatial modeling techniques, is important for generating data about a given reality. In this sense, the WQI results indicated that the waters of the studied sampling points are suitable for multiple uses. However, considering that they are located in the semi-arid region and in the context of climate change, which may compromise water quality in the future, it is necessary to intensify management actions in these basins.

With the project's operation, modifications to the aquatic conditions of the studied rivers, weirs, and reservoirs, in relation to aquatic biota and physicochemical parameters, will be better evaluated through the mixing of water between the water bodies. Current data continue to demonstrate that the observed impact, mainly on physicochemical parameters, is a consequence of land use and occupation in the studied river basins. The actions implemented in the PISF to guarantee water quality will help improve this situation. The PISF is a key element for the semi-arid region, which suffers from water scarcity and therefore depends considerably on guaranteeing the multiple uses that are being defined.

Monitoring water quality is a complex issue that requires support tools to provide information on water-resource management. Ideally, there would be continuous monitoring of water quality however, this is not possible due to budgetary constraints. The use of statistical tools such as modeling proved to be effective, allowing the detection of critical sampling points, predicting trends, diagnosing river basins, and allowing for better decision-making. For future research, it is recommended that the continuity of monitoring in another long-time series be conducted to determine the behavior of the PISF WQI over time.

Authors' contributions

Marques, E. A. T.: conceptualization; data curation; formal analysis; statistical analysis; writing – original draft; writing – review & editing; **Alves**, A. E.: investigation; formal analysis; **Chagas,** R. M.: statistical analysis; methodology; validation; **Méllo Júnior**, A. V.: statistical analysis; methodology; validation; supervision; **Sobral**, M. C.: methodology; supervision.

References

Agência Nacional de Águas e Saneamento Básico (ANA), 2012. Panorama da qualidade das águas superficiais do Brasil. Brasília, ANA, 264 p.

Agência Nacional de Águas e Saneamento Básico (ANA), 2015. Encarte especial sobre a crise hídrica. Informe 2014. Conjuntura dos Recursos Hídricos no Brasil. Brasília, Ministério do Meio Ambiente, 30 p.

Agência Nacional de Águas e Saneamento Básico (ANA), 2025. Indicadores de qualidade - Índice de qualidade das águas (IQA). ANA (Accessed Apr. 25, 2025) at:. https://www.ana.gov.br/portalpnqa/indicadores-indice-aguas.aspx.

Aires, R.F.F.; Salgado, C.C.R., 2024. A TOPSIS-based multicriteria approach for reservoir assessment. Sociedade & Natureza, v. 36, (1), e70948. https://doi.org/10.14393/SN-v36-2024-70948

Aissia, M.A.B.; Chebana, F.; Ouarda, T.B.M., 2017. Multivariate missing data in hydrology – Review and applications. Advances in Water Resources, v. 110, (3), 299-309. https://doi.org/10.1016/j.advwatres.2017.10.002.

Almegdadi, O.; Marcelino, J.; Fakhreddine, S.; Manso, J.; Marques, N.C., 2026. Self-organizing maps for water quality assessment in reservoirs and lakes: A systematic literature review. Ecological Informatics, v. 93, 103542. https://doi.org/10.1016/j.ecoinf.2025.103542

Alssgeer, H.M.A.; Kamarudin, M.K.A.; Samah, M.A.A.; Toriman, M.E.; Gazim, M.B. et al., 2022. Spatial and temporal variations of river water quality using multivariate statistical techniques. Desalination and Water Treatment, v. 269, 106-122. https://doi.org/10.5004/dwt.2022.28677

Anselin, L., 2020. Contiguity-based spatial weights. Center for Spatial Data Science, University of Chicago (Accessed Apr. 20, 2025) at:. https://geodacenter.github.io/workbook/4a_contig_weights/lab4a.html#fn1.

Assunção, R.M.; Neves, M.C.; Câmara, G.; Freitas, C.C., 2006. Efficient regionalization techniques for socio-economic geographical units using minimum spanning trees. International Journal of Geographical Information Science, v. 20, (7), 797-811. https://doi.org/10.1080/13658810600665111.

Azevêdo, E.L.; Medeiros, C.R.; Gomes, W.I.A.; Azevêdo, D.J.S.; Alves, R.R.N. et al., 2018. The use of risk incidence and diversity indices to evaluate water quality of semi-arid reservoirs. Ecological Indicators, v. 90, 90-100. https://doi.org/10.1016/j.ecolind.2018.02.052.

Barbosa, J.E.L.; Severiano, J.; Cavalcante, H.; Lucena-Silva, D.; Mendes, C.F. et al., 2021. Impacts of inter-basin water transfer on the water quality of receiving reservoirs in a tropical semi-arid region. Hydrobiologia, v. 848, (1), 651-673. https://doi.org/10.1007/s10750-020-04471-z.

Begum, S.; Firdous, S.; Naeem, Z.; Chaudry, G.S.; Arshad, S. et al., 2023. Trop. Life Science Research, v. 34, (3), 129-149. https://doi.org/10.21315/tlsr2023.34.3.7.

Bozorg-Haddad, O.; Bahrami, M.; Gholami, A.; Chu, X.F.; Loáiciga, H.A., 2024. Investigation and classification of water resources management strategies: possible threats and solutions. Natural Hazards, v. 120, (11), 9867-9892. https://doi.org/10.1007/s11069-024-06589-y

Brasil, 2005. Resolução n. 357, de 18 de março de 2005. Diário Oficial da União.

Brasil, 2019. Projeto de Integração do Rio São Francisco. Ministério do Desenvolvimento Regional (Accessed Dec. 2019) at:. https://www.gov.br/mdr/pt-br/assuntos/seguranca-hidrica/projeto-sao-francisco.

Caldas, H.F.M., 2021. Análise da evolução espaço-temporal da qualidade da água no Eixo Leste do Projeto de Integração do Rio São Francisco. Dissertation, Programa de Pós-Graduação em Engenharia Civil, Universidade Federal de Pernambuco, Recife.

Castro, R.M.; Marcionilio, S.M.L.O.; Morais, W.A.; Alves, W.S.; Oliveira, L.D. et al., 2023. Spatiotemporal analysis of water quality indicators in a watershed in Brazilian Cerrado. Water Supply, v. 23, (3), 1161-1176. https://doi.org/10.2166/ws.2023.061.

Cerezini, M.T.; Castro, C.N., 2023. Projeto de integração do São Francisco e a segurança hídrica da região beneficiada. Texto para Discussão. Rio de Janeiro, Instituto de Pesquisa Econômica e Aplicada (IPEA).

Chloé, N.A.; Branco, G.; Bopp, C.; Carrasco, N., 2022. Modes of access to water for domestic use in rural Chile: a typological proposal. Water Policy, v. 24, (7), 1179-1194. https://doi.org/10.2166/wp.2022.026.

Companhia Ambiental do Estado de São Paulo (CETESB), 2024. Apêndice D: Metodologia de cálculo do índice de qualidade das águas. In: CETESB. Qualidade das águas interiores do estado de São Paulo. São Paulo, CETESB (Accessed Apr. 20, 2025) at:. https://cetesb.sp.gov.br/aguas-interiores/.

Cook, S.; Chan, H.L.; Abolfathi, S.; Bending, G.D.; Schäfer, H. et al., 2020. Longitudinal dispersion of microplastics in aquatic flows using fluorometric techniques. Water Research, v. 170, 115337. https://doi.org/10.1016/j.watres.2019.115337.

Cordão, M.J.S.; Rufino, I.A.A.; Alves, P.B.R.; Barros Filho, M.N.M., 2020. Water shortage risk mapping: a GIS-MCDA approach for a medium-sized city in the Brazilian semi-arid region. Urban Water Journal, v. 17, (7), 642-655. https://doi.org/10.1080/1573062X.2020.1804596.

Costa, M.R.A.; Attayde, J.L.; Becker, V., 2016. Effects of water level reduction on the dynamics of phytoplankton functional groups in tropical semi-arid shallow lakes. Hydrobiologia, v. 778, (1), 75-89. https://doi.org/10.1007/s10750-015-2593-6.

Costa, P.V.M., 2025. Secas: desigualdades e saúde no semiárido brasileiro. Centro de Integração de Dados e Conhecimento para Saúde (Cidacs) (Accessed Sep. 11, 2025) at:. https://cidacs.bahia.fiocruz.br/2025/07/secas-desigualdades-e-saude-no-semiarido-brasileiro/.

Dantas, J.C.; Silva, R.M.; Santos, C.A.G., 2020. Drought impacts, social organization, and public policies in northeastern Brazil: A case study of the upper Paraíba river basin. Environmental Monitoring and Assessment, v. 192, (5), 317. https://doi.org/10.1007/s10661-020-8219-0.

Das, A., 2025. Surface water quality assessment and its evaluation of potential pollution risks for drinking purposes employing water quality indices and various machine learning techniques. Desalination and Water Treatment, v. 324, 101548. https://doi.org/10.1016/j.dwt.2025.101548

Deng, L.; Chen, K.; Liu, Z.; Wu, B.; Chen, Z.; He, C., 2022. Spatiotemporal variation evaluation of water quality in middle and lower Han River, China. Science Report, v. 12, 14125. https://doi.org/10.1038/s41598-022-16808-w.

Dixon, W.; Chiswell, B., 1996. Review of aquatic monitoring program design. Water Research, v. 30, (9), 1935-1948. https://doi.org/10.1016/0043-1354(96)00087-5

Fatima, S.U., Khan, M.A.; Alamgir, A.; Mahmood, N.; Sulman, N., 2022. Multivariate and spatial methods-based water quality assessment of Chu Tran Valley, Gilgit Baltistan. Applied Water Science, v. 12, (6), 129. https://doi.org/10.1007/s13201-022-01661-z.

Faúndez, M.; Alcayaga, H.; Walters, J.; Pizarro, A.; Soto, M., 2022. Sustainability of water transfer projects: A systematic review. Science of the Total Environment, v. 860, (4), 160500. https://doi.org/10.1016/j.scitotenv.2022.160500.

Geoda Center, 2025. Geoda Release version 1.22.0.14. Campaign Inquiry & Impact, University of Chicago (Accessed Apr. 20, 2025) at:. https://geodacenter.github.io/download/.

Javed, S.; Ali, A.; Ullah, S., 2017. Spatial assessment of water quality parameters in Jhelum city (Pakistan). Environmental Monitoring and Assessment, v. 189, 119. https://doi.org/10.1007/s10661-017-5822-9.

Liu, L., 2018. Application of a hydrodynamic and water quality model for inland surface water systems. In: Malcangio, D. (Ed.), Applications in water systems management and modeling. InTech, pp. 87-109.

Lu, J.; Gu, J.; Han, J.; Xu, J.; Liu, Y. et al., 2023. Evaluation of spatiotemporal patterns and water quality conditions using multivariate statistical analysis in the Yangtze River, China. Water, v. 15, (18), 3242. https://doi.org/10.3390/w15183242.

Magalhães, A.R.; Martins, E.S.P.R., 2021. The São Francisco River Basin. In: Schmandt J.; Kibaroglu A.; eds. Sustainability of engineered rivers in arid lands: challenge and response. Cambridge University Press, Cambridge.

Malsy, M.; Flörke, M.; Borchardt, D., 2017. What drives the water quality changes in the Selenga Basin: climate change or socio-economic development? Regional Environmental Change, v. 17, (7), 1977-1989. https://doi.org/10.1007/s10113-016-1005-4.

Medeiros, P.; Sivapalan, M., 2020. From hard-path to soft-path solutions: slow–fast dynamics of human adaptation to droughts in a water scarce environment. Hydrological Sciences Journal, v. 65, (11), 1803-1814. https://doi.org/10.1080/02626667.2020.1770258.

Mehmood, H., 2019. Bibliometrics of water research: a global snapshot. UNU-INWEH Report Series, Issue 06. Hamilton, United Nations University Institute for Water, Environment and Health, 24 p.

Messager, M.L.; Lehner, B.; Cockburn, C.; Lamouroux, N.; Pella, H. et al., 2021. Global prevalence of non-perennial rivers and streams. Nature, v. 594, 391-397. https://doi.org/10.1038/s41586-021-03565-5.

Ministério da Integração e Desenvolvimento Regional (MIDR), 2016. Relatório da 20ª campanha de monitoramento da qualidade da água e limnologia do PISF: Relatório Técnico RT/PISF/CTD/006-16. Custódia, MIDR.

Ministério da Integração e Desenvolvimento Regional (MIDR), 2025. Projeto de Integração do Rio São Francisco. Sumário Executivo. Brasília, MIDR.

Njugana, S.M.; Onyango, J.A.; Githaiga, K.B.; Gituru, R.W.; Yan, X., 2020. Application of multivariate statistical analysis and water quality index in health risk assessment by domestic use of river water. Case study of Tana River in Kenya. Process Safety and Environmental Protection, v. 133, 149-158. https://doi.org/10.1016/j.psep.2019.11.006.

Nong, X.; Shao, D.; Zhong, H.; Liang, J., 2020. Evaluation of water quality in the South-to-North Water Diversion Project of China using the water quality index (WQI) method. Water Research, v. 178, 115781. https://doi.org/10.1016/j.watres.2020.115781.

Novianta, M.A.; Syafrudin, S.; Warsito, B.; Rachmawati, S., 2025. Mapping of river water quality through spatial K'luster analysis by tree edge removal. Ecological Engineering & Environmental Technology, v. 26, (1), 58-67. https://doi.org/10.12912/27197050/195296.

Patil, V.B.B.; Pinto, S.M.; Govindaraju, T.; Hebbalu, V.S.; Bhat, V. et al., 2020. Multivariate statistics and water quality index (WQI) approach for geochemical assessment of groundwater quality-a case study of Kanavi Halla Sub-Basin, Belagavi, India. Environmental Geochemistry and Health, v. 42, (9), 2667-2684. https://doi.org/10.1007/s10653-019-00500-6.

Pires, A.P.N., 2019. Estrutura e objetivos da transposição do rio São Francisco: versões de uma mesma história. Geousp – Espaço e Tempo, v. 23, (1), 182-197. https://doi.org/10.11606/issn.2179-0892.geousp.2019.122366

Rakotondrabe, F.; Ngoupayou, J.R.N.; Mfonka, Z.; Rasolomanana, E.H.; Abolo, A.J.N. et al., 2018. Water quality assessment in the Bétaré-Oya gold mining area (East-Cameroon): multivariate statistical analysis approach. Science of the Total Environment, v. 610-611, 831-844. https://doi.org/10.1016/j.scitotenv.2017.08.080

Ramos Filho, R.S.; Barros Filho, M.N.M.; Rufino, I.A.A.; Costa, F.F.; Cordão, M.J.S., 2023. Urban water insecurity drivers in the Brazilian semiarid region. Water Supply, v. 23, (2), 447-458.

Ruckert, G.V.; Figueredo, C.C.; Barros, J.E.F.; Silva, R.R., 2024. Water quality and land use in Ipanema Stream Watershed (Doce River Basin/Brazil): effects of urbanization. Revista Brasileira de Recursos Hídricos, v. 29, e25. https://doi.org/10.1590/2318-0331.292420230142

Saadtpour, M., 2020. An adaptive surrogate assisted CE-QUAL-W2 model embedded in hybrid NSGA-II AMOSA algorithm for reservoir water quality and quantity management. Water Resources Management, v. 34, (4), 1437-1451. https://doi.org/10.1007/s11269-020-02510-x.

Shehab, Z.N.; Algamal, Z.Y.; Farhhan, A.F., 2025. Water quality assessment and spatial variation analysis in Tigris River, Mosul, Iraq During 2023-2024. Egyptian Journal of Aquatic Biology & Fisheries, v. 29, (6), 3607-3626. https://doi.org/10.21608/ejabf.2025.440427.6955

Siegmund-Schultze, M.; Köppel, J.; Sobral, M.C., 2018. Unraveling the water and land nexus through inter and transdisciplinar research: sustainable land management in a semi-arid watershed in Brazil's Northeast. Regional Environmental Change, v. 18, (7), 1-13. https://doi.org/10.1007/s10113-018-1302-1

Silveira, N.T.; Soares, G.A.S.; Bandim, C.G.A.; Santana, S.H.C.; Galvíncio, J.D., 2024. Influência das mudanças climáticas na qualidade da água na bacia hidrográfica do Terra Nova, Pernambuco. Scientia Plena, v. 20, (9), 099901. https://doi.org/10.14808/sci.plena.2024.099901

Singh, P.; Dar, J.G.; Dhanorkar, M.; Warke, A., 2025. Multivariate statistical analysis of water quality of major rivers in pune: a case study. International Journal of Mathematical, Engineering and Management Sciences, v. 10, (4), 1013-1031. https://doi.org/10.33889/IJMEMS.2025.10.4.049.

Souza, S.D.G.; Souza, A.C.N.; Souza, M.L.M.; Nascimento, F.R., 2024. Uso, cobertura da terra e degradação ambiental na porção sedimentar da Bacia Hidrográfica do rio Apodi-Mossoró, RN, Brasil. Revista Equador, v. 13, (1), 169-187.

Tang, Q.; Liu, X.; Zhou, Y.; Wang, P.; Li, Z. et al., 2022. Climate change and water security in the northern slope of the Tianshan Mountains. Geography and Sustainability, v. 3, (3), 246-257. https://doi.org/10.1016/j.geosus.2022.08.004.

Thomaz, F.R.; Miguez, M.G.; Sá, J.G.S.R.; Alberto, G.W.M.; Fontes, J.P.M., 2025. Water scarcity risk index: a tool for strategic drought risk management. Water, v. 15, (2), 255. https://doi.org/10.3390/w15020255.

Trindade, A.L.C.; Almeida, K.C.B.; Barbosa, P.E.; Oliveira, S.M.A.C., 2017. Temporal and spatial trends of surface water quality of Velhas River sub-basin, Minas Gerais state. Engenhara Sanitária e Ambiental, v. 22, (1), 13-24. https://doi.org/10.1590/S1413-41522016131457.

Tundisi, J.G.; Matsumura-Tundisi, T.; Tundisi, J.E.M., 2008. Reservoirs and human well-being: new challenges for evaluating impacts and benefits in the Neotropics. Brazilian Journal of Biology, v. 68, (4 Suppl.), 1133-1135. https://doi.org/10.1590/S1519-69842008000500020.

Uddin, M.G.; Moniruzzaman, M.; Quader, M.A.; Hasan, M.A., 2018. Spatial variability in the distribution of trace metals in groundwater around the Rooppur nuclear power plant in Ishwardi, Bangladesh. Groundwater for Sustainable Development, v. 7, 220-231. https://doi.org/10.1016/j.gsd.2018.06.002.

United Nations, 2018. The 2030 Agenda and the Sustainable Development Goals: An opportunity for Latin America and the Caribbean (LC/G.2681-P/Rev.3). Santiago, United Nations.

United Nations Educational, Scientific and Cultural Organization (UNESCO), 2019. World Water Assessment Programme. Relatório Mundial das Nações Unidas sobre Desenvolvimento dos Recursos Hídricos 2019: Não deixar ninguém para trás, resumo executivo (Accessed Nov. 7, 2025) at:. www.unesco.org/ark:/48223/pf0000367303_por#.

Von Sperling, M., 1996. Introdução à qualidade das águas e ao tratamento de esgotos. 2. ed. Belo Horizonte, Editora UFMG.

Wang, Z.; Jia, D.; Song, S.; Sun, J., 2023. Assessments of surface water quality through the use of multivariate statistical techniques: A case study for the watershed of the Yuqiao Reservoir, China. Frontiers in Ecology and the Environment, v. 11, 1107591. https://doi.org/10.3389/fenvs.2023.1107591

Xu, H.; Lin, J.; Lei, X.; Zhang, D.; Peng, Q.; Wang, J.; Zhu, B., 2023. Assessment of the spatiotemporal water quality variations in the Middle Route of China's South-to-North Water Diversion Project by multivariate analysis. Environmental Science and Pollution Research, v. 30, (4), 1-17. https://doi.org/10.1007/s11356-022-25115-w.