Environmental Science and Pollution Research
https://doi.org/10.1007/s11356-022-22784-5

RESEARCH ARTICLE

Understanding the patterns and processes underlying water quality and pollution risk in West–Africa River using self-organizing maps and multivariate analyses

Zinsou Cosme Koudenoukpo[1,2] · Olaniran Hamed Odountan[2,3,4] · Chuanbo Guo[5] · Regis Céréghino[6] · Antoine Chikou[1] · Young-Seuk Park[7]

Received: 17 March 2022 / Accepted: 24 August 2022

Abstract

Rivers are dynamic systems in complex interactions with their surrounding environments. Reliable and fast interpretation of water quality is therefore needed for sustainable river management. Unfortunately, water quality and environmental status interactions have not yet been documented sufficiently in West–Africa. This study explored the spatial–latitudinal and seasonal features of water quality along the Sô River Basin (SRB, West Africa) using self-organizing map (SOM) and principal component analysis. Twenty-two water quality variables were measured in the surface layer at 12 different sampling sites during a twenty-four-month period from July 2016 to June 2018. The results revealed three water quality groups, following an upstream-downstream pollution gradient: (1) upstream and middle reach sites with high dissolved oxygen and Secchi disk depth values, which are more suitable for the aquatic biota; (2) downstream sites with high concentrations of ammonium, biochemical oxygen demand, and heavy metals especially in flood period, reflecting both high organic and heavy metal pollution; and (3) brackish downstream sites characterized by less heavy metal and organic pollutions. No significant variation was observed between seasons. However, the SRB relatively suffered from higher risks of heavy metal contamination and organic pollution in wet seasons. Although hydroclimatic processes affect the water quality, anthropogenic inputs of point and non-point sources were identified and discussed as a more prominent factor contributing to variation in the water quality condition. These results offer insights into the water quality dynamics in river–estuary system as well as potential pollution sources, crucial for defining sanitation, and management measures.

Keywords Eutrophication · Heavy metals · Artificial neural network · Multivariate analyses · West–Africa River systems

---

Zinsou Cosme Koudenoukpo and Olaniran Hamed Odountan are equal contributors.

Responsible Editor: Xianliang Yi

✉ Olaniran Hamed Odountan
Odountan.hamed@gmail.com

[1] Laboratoire d'Hydrobiologie et d'Aquaculture, Faculté des Sciences Agronomiques, Université d'Abomey-Calavi, 01 BP 526, Cotonou, Abomey-Calavi, Bénin

[2] Cercle d'Action pour la Protection de l'Environnement et de la Biodiversité (CAPE BIO-ONG), 10 PO Box 336, Cotonou, Abomey-Calavi, Benin

[3] Laboratory of Ecology and Aquatic Ecosystem Management, Department of Zoology, Faculty of Sciences and Technics, University of Abomey-Calavi, Abomey-Calavi, Republic of Benin

[4] Laboratory of Research on Wetlands, Department of Zoology, Faculty of Science and Technics, University of Abomey-Calavi, Abomey-Calavi, Benin

[5] State Key Laboratory of Freshwater Ecology and Biotechnology, Institute of Hydrobiology, Chinese Academy of Sciences, Wuhan 430072, Hubei, China

[6] Laboratoire Ecologie Fonctionnelle et Environnement, CNRS, Université de Toulouse, 118 route de Narbonne, F-31062 Toulouse Cedex 9, France

[7] Department of Biology, Kyung Hee University, Seoul 02447, Korea

Introduction

Promoting good quality of ground to surface waters is crucial to preserve the biodiversity of freshwater ecosystems, as well as the goods and services they provide. This is especially true for rivers that are of vital importance to human health and activities (Aboelnour et al. 2020). Variation of water quality within a river system relates to natural processes such as nutrient cycling or self-purification, human activities in the associated watershed, and landscape features (Ren et al. 2003). Industrial-urban sewage discharge and agricultural runoff are key sources of metal- and non-metal solid and liquid waste pollution (Vadde et al. 2018; Aniyikaiye et al. 2019). Without treatment, materials in dissolved and particulate phases originating from these activities enrich the aquatic environment with pollutants and undesirable or potentially dangerous elements for public health (Mamun and An 2018; Odountan et al. 2019).

In West–Africa, water quality and sanitation receive little attention, fostering direct and indirect waterborne diseases such as cholera, diarrhoea, Buruli ulcer, bilharzias, and schistosomiasis that are recurrent in the region (Calamari and Naeve 1994; UNEP 2010; Ibikounlé et al. 2014). As result, maintaining a healthy relationship between human, nature, and water, is of paramount interest than ever for the region. To overcome with the aquatic ecosystem degradation, West–Africa governments are devoting a substantial share of their national budgets. For example, the Benin government made considerable efforts to ensure sustainable management of country's major watershed. This is reflected in the roughly US$62 million budget of the OmiDelta programme, whose core activities include water sanitation and Integrated Water Resources Management (IWRM) with emphasis on the development of a delta plan promoting a safer environment for 600,000 people (OmiDelta–INE 2018). Noticeably, the cost to promote IWRM and sanitation requires a huge sum of money especially when the ecosystem is heavily polluted. Prime example is Singapore, where US$159.8 million have been invested for the restoration of heavily polluted rivers (Joshi et al. 2012). Therefore, it is crucial for ecologists in West–Africa, a region with limited resources, to provide decision-makers with a timely and reliable interpretation of river systems quality for adequate actions to keep these systems healthy. In the absence of such timely tool to understand water quality patterns allowing to control the evolution of water quality, large budget will be required for good quality recovery.

The assessment of the relationship, often complex among sources, water quality, and environmental status in an ecosystem requires the use of multivariate statistical methods (Wang et al. 2017; Dai et al. 2018; Ustaoğlu and Tepe 2019). The application of different multivariate statistical techniques, such as factor analysis (FA), principal component analysis (PCA), cluster analysis (CA), and discriminant analysis (DA), notably helped to identify possible sources that influence water systems (Zhao et al. 2012; Olawoyin et al. 2013; Berrada et al. 2016; Li et al. 2018; Gu et al. 2019). A much more efficient approach for classification, modelling, and interpretation of data obtained from monitoring studies of surface water is the one combining intelligent data analysis such as self-organizing map (SOM) and traditional multivariate analyses (Zhao et al. 2012; Olawoyin et al. 2013; Berrada et al. 2016; Li et al. 2018; Gu et al. 2019). Tsai et al. (2017) pointed out that the SOM is a cost-effective alternative to more traditional methods for identifying spatial characteristics of water quality parameters and to reflect the interaction with biological resources. One of the reasons is that SOMs allow simultaneous clustering of multidimensional dataset of variables and objects (samples). This enables wide applications in assessing relationships between samples and environmental issues as well as patterning and predicting water quality in aquatic ecological studies from surface waters (Chea et al. 2016; Pacheco et al. 2017; Wang and Zhang 2018; Gu et al. 2019) to groundwaters (Haselbeck et al. 2019; Zhu et al. 2020). Another reason is that SOMs preserve the topological structure of the input data. Moreover, Alvarez-Guerra et al. (2008) indicated that the powerful visualization tools of the SOM approach provided more information and is an easier way to classify sediment quality, which facilitates the task of establishing an order of priority between the distinguished groups of sites depending on their need for further investigations or remediation actions in subsequent management steps. Olkowska et al. (2014) argued that in the area of interest where numerous input parameters must be considered, SOM could allow clear distinction of groups of pollution as well as the seasonal variation of pollution. Although the growing environmental awareness has motivated the necessity to understand relationships among sources, water quality, and environmental status in order to achieve sustainability in rivers' management, few studies have explored these interactions and evolution with respect to spatial and temporal factors in West–Africa.

The Sô River Basin (SRB) of Benin is internationally known as both part of Ramsar site no. 1018 and Ouémé Man and Biosphere (MAB-UNESCO) reserve. Constant water quality degradation may threaten the health of thousands of people living around the ecosystem as well as the long-term sustainability of the fishery. The SRB shares several species with adjacent regional ecosystems (Koudenoukpo et al. 2020) and suffers from the typical anthropogenic contamination observed in many West–Africa river systems, such as petroleum leaking being a key corridor for the smuggling, water hyacinth proliferation, agricultural run-off, and domestic and industrial discharge (Mama 2010); all of which make the river an ideal place to draw up valid lessons on the water quality and pollution risk in the West–Africa context. Moreover, it is still upstream imbedded in a narrow strip of forest, which could serve as reference habitat. More importantly, SOM in conjunction with other analytical methods showed their power by successfully allowing to simplify the complex ecological relationship between key environmental variables and mollusc community for the river sustainable management and restoration (Koudenoukpo et al. 2021). Although the water degradation of the river is indisputable, a comprehensive understanding of the patterns and processes underlying its quality and pollution risk were needed for sustainable management.

This study contributes with novelties about water quality and pollution risk assessment in the West–Africa context using self-organizing maps and multivariate analyses that until now, to our knowledge, has not yet been published in the literature for this region. For this purpose, the specific objectives of this study were, (1) to bring out spatio-temporal variation patterns in the water quality, (2) to evaluate water quality, and identify the relative importance of the variables for clustering, and (3) to assess the pollution risk and sources by interpreting the water quality variable values and interrelations among them. We hypothesised that the water quality follows and upstream downstream gradient with respect to seasons. Documenting pollution of surface water in relation to abiotic variables is of paramount interest in developing countries, where expertise, techniques and facilities are limited for such study. Our approach could most likely be applied to characterise a wider variety of freshwater systems.

Material and methods

Study area

The Sô River, located between 6°24′ to 6°32′ N and 2°27′ to 2°30′ E, originates from the Lake Hlan and then flows over 85 km into the Lake Nokoué at northwest area. Considered as the second longest river in southern Benin after the Ouémé River, the Sô River occupies a drainage area of more than 3,000 km^2. It is one of the former tributaries of the Ouémé River. However, connection is observed between the two rivers during flooding via some marigots or canals. It is for local communities an important water resource for irrigation, domestic and industry uses as well as for the Ouémé River flood control purposes. The study area benefits from a sub-equatorial climate marked by two wet seasons and two dry seasons. The long dry season (LDS) expanded from mid-November to mid-March, whereas the long-wet season (LWS) occurred from mid-March to mid-July. The short dry season (SDS) is observed from mid-July to mid-September, and a short-wet season (SWS) reported from mid-September to mid-November (Gnohossou 2006). Locally, the annual mean temperature varies between 27 and 29 °C whereas the annual mean precipitation of 900 to 1 100 mm is reported (Odountan et al. 2019). SRB water resources are extremely important for eight riparian municipalities of South-East Benin namely Abomey-Calavi, Adjohoun, Aguégués, Cotonou, Dangbo, Sô-Ava, Toffo, and Zè. SRB serves for a variety of water-related activities such as livestock farming, fisheries, agriculture, navigation and transport of goods, trade, tourism and recreation, which support the livelihood of more than 300,000 people living at the vicinity of the basin.

Prior to the present study, we undertook a three-month (April to June 2016) exploratory phase to choose latitudinally distributed sampling sites along the river, and to assess unsuspected factor which could affect water quality and biodiversity. We ensured that sampling sites distribution was based on proximity to facilities, representation of the main tributaries, human activities and hydrological regime of the river. As such, 12 sampling sites were chosen (Fig. 1).

Sampling and water analysis

From July 2016 to June 2018, 12 sites were monthly sampled, leading to a total of 288 samples. Eight sampling events occurred during the LWS and LDS seasons against four sampling events in the SDS and SWS seasons. Water quality measurements were made according to environmental quality standards and technical specifications requirements for monitoring surface water (Rice et al. 2012). Operational process of some items can refer to the literature (Rodier et al. 2009; Hong et al. 2020; Xu et al. 2022). Briefly, the quality assurance of analytical data has been ensured by strict collecting, preserving, and processing procedures, careful standardization, procedural blank measurements, and duplicate samples. The selected water quality indicators measured in situ were water depth, Secchi disk depth, temperature, salinity, pH, electric conductivity, total dissolved solids, and dissolved oxygen, using YANHE multi-probe (Table 1). The sensors of the multi-probe were calibrated prior to each sampling campaign. In addition to in situ measurements, duplicate 1 L water samples were collected at each site in the surface layer from a depth of ~20 cm (ISO 5667-5 2006; ISO 5667-6 2014) in labelled plastic bottles that were previously treated with hydrochloric acid and carefully rinsed with the respective site water sample. One duplicate was filtered with 0.45 µm mixed cellulose ester millipore filter membranes and fixed for measurement of nutrients. The second duplicate was filtered with 0.45 µm cellulose acetate filter membranes and acidified to pH = 2 by addition of nitric acid for the determination of heavy metals (Li et al. 2018). The samples were transported in ice chest

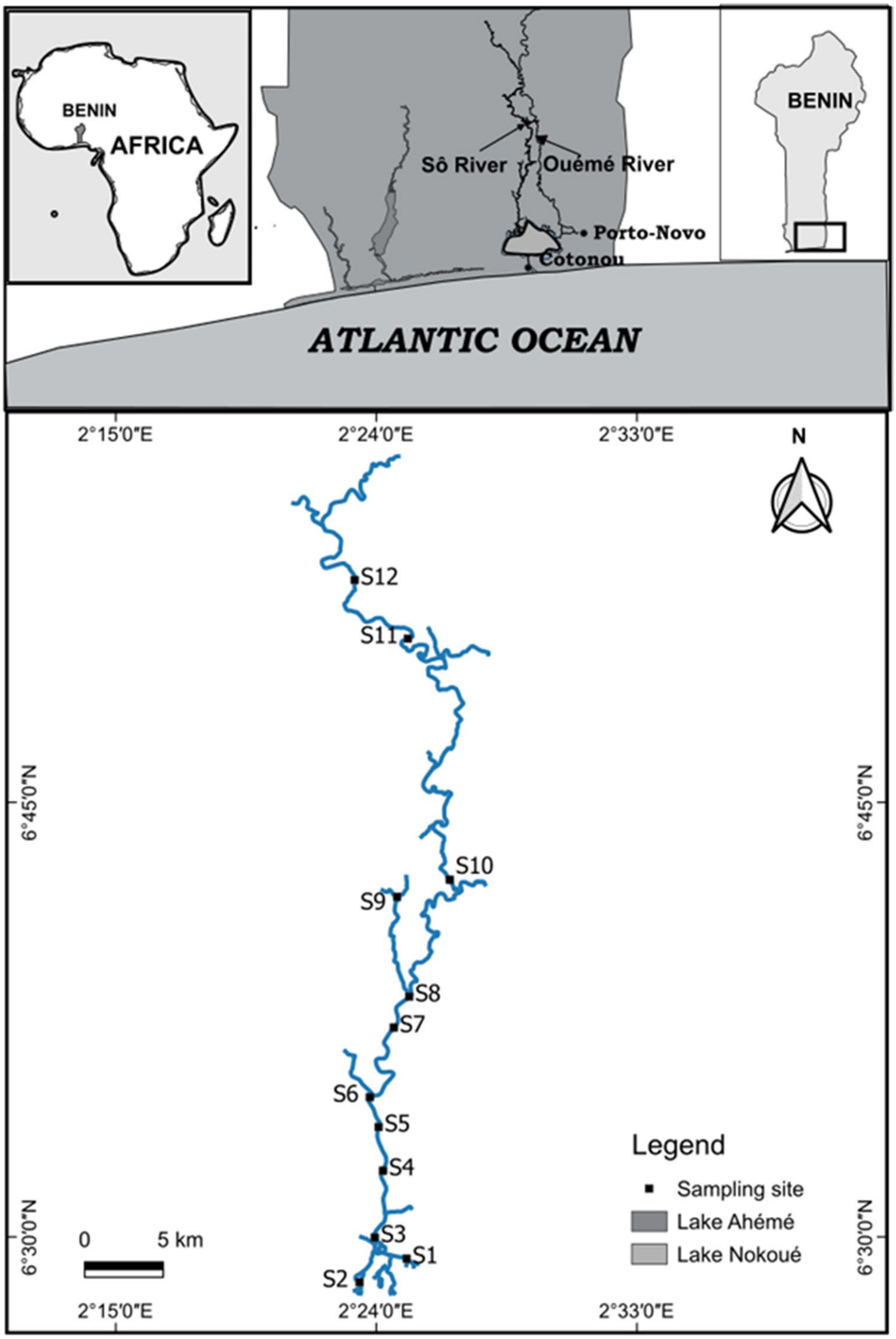


Fig. 1 Map of the Sô River Bassin with the locations of monitoring sites

to the laboratory and preserved in the refrigerator prior to analyses. The nutrients nitrite (NO_2^-), nitrate (NO_3^-), ammonium (NH_4^+), total nitrogen (TN), orthophosphates (PO_4^{3-}), total phosphorus (TP), magnesium (Mg^{2+}), and calcium (Ca^{2+}) were measured using the spectrophotometric method (Model HACH DR 3900). The reading was conducted with appropriate reagent at a specified wavelength (Rodier et al. 2009; Rice et al. 2012). An inductively coupled plasma mass spectrometer (ICP-MS; Drawell MS-2000) was used for the detection of heavy metals such as copper (Cu), cadmium (Cd), lead (Pb), and nickel (Ni), whereas an atomic fluorescence spectrometer (AFS; Drawell AF-630A) was used to measure the concentration of mercury (Hg). The detailed units, acronyms, analytical methods, and the limit of detection (if applicable) of the 22 water quality variables can be found in the Table 1.

Table 1 Overview of methods for determining core environmental variable used in the study.

Variable	Acronym	Units	Analytical methods	Limit of detection
In situ variables				
Water depth	WD	[m]	Water depth gauge	
Secchi disk depth	SD	[m]	Secchi disk	
Biochemical oxygen demand	BOD	[mg L^{-1}]	Respirometric BOD OxiTop method	
Temperature	T	°C	YANHE multi-probe water analyser (model SX736 pH/mV/ conductivity/DO)	
Dissolved oxygen	DO	[mg L^{-1}]		
pH				
Salinity		[PSU]		
Electric conductivity	EC	[µS Cm^{-1}]		
Total dissolved solids	TDS	[mg L^{-1}]		
Inorganic ions		[mg L^{-1}]		
Nitrite	NO_2^-		Diazotization, TNT 839 method	0.05
Nitrate	NO_3^-		Cadmium reduction method	0.3
Ammonium	NH_4^+		Nessler method	0.025
Total nitrogen	TN		Persulfate digestion method and Nessler method	0.01
Orthophosphates	PO_4^{3-}		Molybdate, sulfuric and ascorbic acids method	0.05
Total phosphorus	TP		Persulfate digestion method coupled to molybdate method	0.5
Magnesium	Mg^{2+}		Calmagite colorimetric method	0.012
Calcium	Ca^{2+}		Calmagite colorimetric method	0.02
Heavy metals		[µg L^{-1}]		
Copper	Cu		ICP-MS	0.02
Cadmium	Cd		ICP-MS	0.05
Lead	Pb		ICP-MS	0.005
Nickel	Ni		ICP-MS	0.005
Mercury	Hg		AFS	0.001

Data analysis

SOM, also known as the Kohonen map, is a nonlinear projection mapping based on unsupervised competition learning method which is used to describe and display multidimensional datasets onto topological structure preserving the topology of input data (Kohonen 1990, 2013). SOM can be used to partition the data with easy two-dimensional visualization of expression patterns while reducing computational requirements; all of which explains its preference compared to other nonlinear methods.

In this study, the SOM procedure was trained using the batch algorithm of the Kohonen 3.0 package (Wehrens and Kruisselbrink 2018) in R environment (R Core Team 2018). The details of the followed procedures for exploring the patterns by means of SOM and the other statistical tools are summarised in the flowchart presented in Fig. 2. Steps 1–6 will be addressed in the next sections whereas step 7 constitutes the results section. Most of the analyses were carried out in R environment (R Core Team 2018) with specific packages.

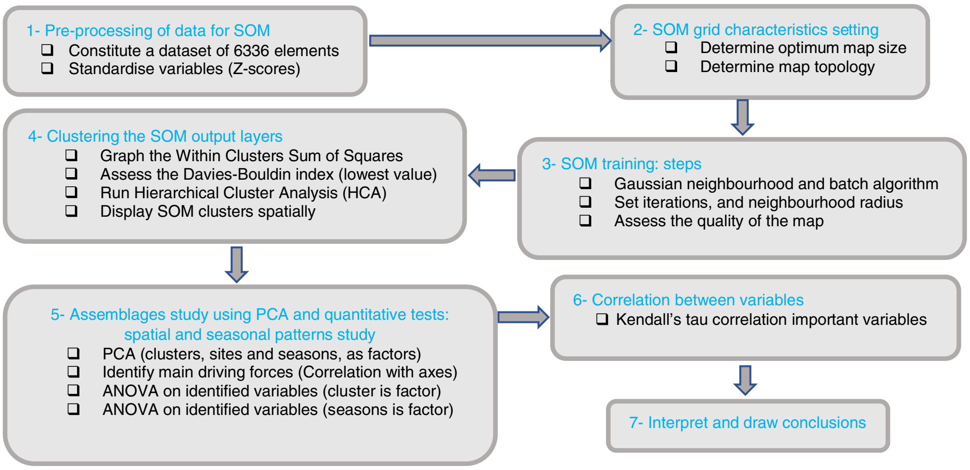
Data pre-processing for SOM

The first step of data analysis consisted in the preliminary selection of environmental variables and data preparation (Fig. 2; step 1). The SOM was applied on the SRB data set of 6,336 elements (22 environmental variables for each of the 288 samples). Standardization transformation (linear transformation that scales values with a mean = 0 and variance = 1, thus making the data dimensionless for each variable), very useful for studying environmental variables in ecological studies (Park et al. 2018), were applied to these variables in order to both normalise their frequency distributions and to ensure that variables showing different measurement units take values over roughly the same limits (Subida et al. 2013).

SOM grids characteristics

Map size selection is a key issue in SOM implementations (Fig. 2; step 2). Too big map size results in too detailed patterns with confusion sometimes due to the small difference between groups while too small map size generates too general patterns with limited details (Céréghino and Park 2009).

Fig. 2 Flowchart of the data analysis.



Following the heuristic rule suggested by Vesanto et al. (2000), the optimum size of the map is close to $5\sqrt{n}$ where n is the number of training samples. With 288 observations, this results in a SOM grid with approximately 85 neurons for this study. To more accurately select the map size, the networks were trained adjusting map dimensions sizes ranging from 80 to 90 and the quality of the trained maps were evaluated with the quantization error (QE) and topographic error (TE) (Vesanto et al. 2000). Given that the optimum SOM grid must minimize the QE and TE (Céréghino and Park 2009; Kohonen 2013), and that more edge effects is observed with square shaped SOM grids (Kohonen 1990),

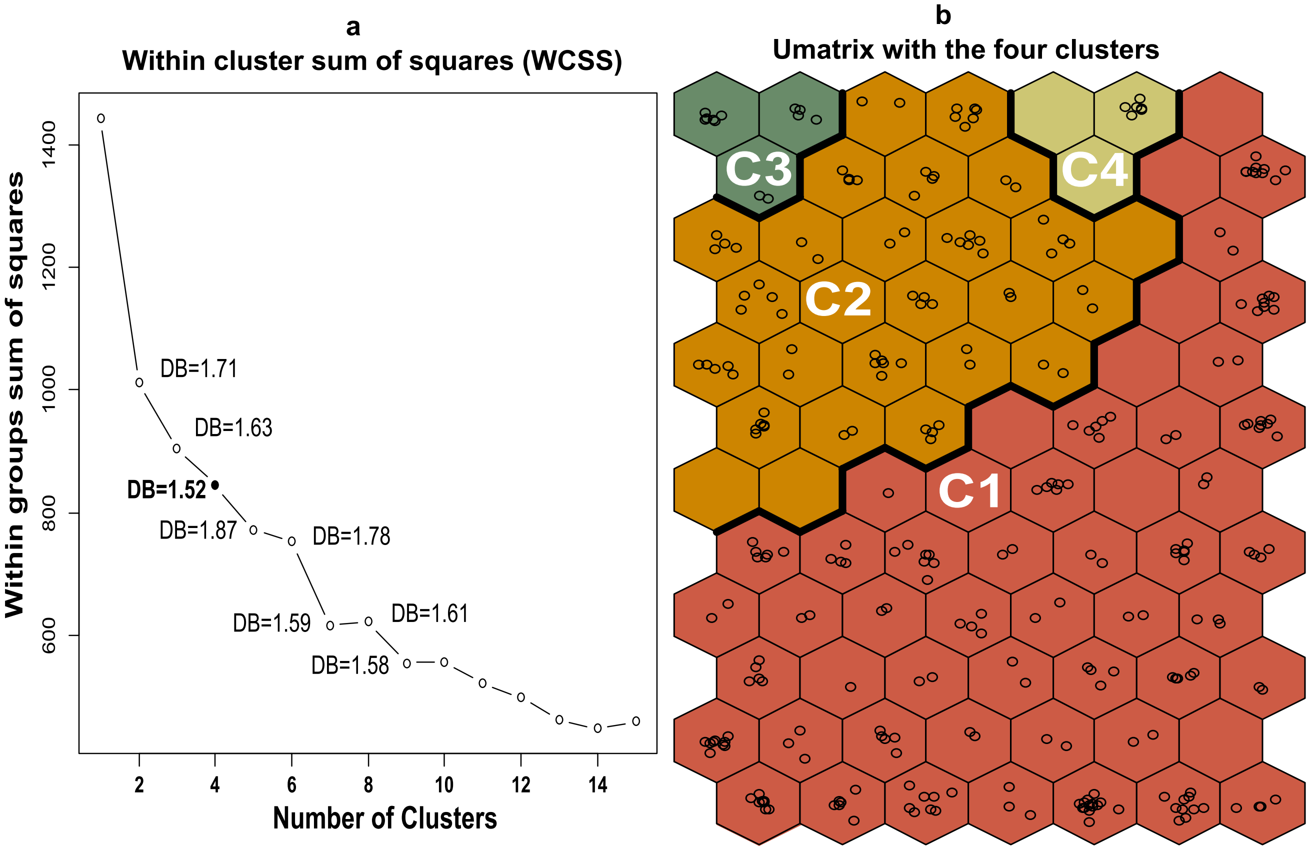


Fig. 3 Grouping of water samples (n = 288) covering all sampling sites (n = 12) into clusters based on SOM analysis: **a** within-cluster sum of squares for hierarchical clustering with average linkage and Davies-Bouldin index value (DB) highlighting the optimal clusters which is four in this study; **b** the map showing the clustering of samples according to their environmental variables features. C1–C4 represent different clusters. The small empty circles in the hexagonal units represent different samples.

Table 2 The spatial and seasonal distribution of the samples inside each cluster. *LDS*, long dry season; *LWS*, long wet season; *SDS*, short dry season; *SWS*, short wet season.

Variable	C1	C2	C3	C4
Site factor				
S1	0	18	6	0
S2	0	22	2	0
S3	0	22	2	0
S4	0	16	2	6
S5	24	0	0	0
S6	24	0	0	0
S7	24	0	0	0
S8	24	0	0	0
S9	24	0	0	0
S10	24	0	0	0
S11	24	0	0	0
S12	24	0	0	0
Season factor				
LDS	64	28	1	3
LWS	64	24	6	2
SDS	32	15	0	1
SWS	32	11	5	0
Total	192	78	12	6

our output layers were asymmetrical grids to reduce edge effects. Each neuron from the output layer has a 2-dimensional topologic structure of a regular hexagonal grid.

SOM training

A Gaussian neighbourhood for training the SOM was used to encounter best result (Wendel and Buttenfield 2010). Iterations (500), and neighbourhood radius (2/3 of all unit-to-unit distances) were adjusted to have a plateau in the training process. The quality of the nodes was assessed using the distance to the codebook vector within a node. The goal being less distance within node (Fig. 2; step 3).

Clustering SOM units using hierarchical cluster analysis (HCA)

Although the SOM itself is already a clustering method, data and map analyses are more accurate and easier when similar units of the SOM are grouped together as cluster (Fig. 2; step 4). Therefore, clustering within the output layer, using mathematical tool, has become a widely-used technique to deep relationships among input layers and to show the aggregation pattern of samples (Pacheco et al. 2017; Liao et al. 2019). The number of clusters was chosen based on the graph of the relationship between the number of cluster (2-15) and *Within Clusters Sum of Squares* (WCSS) metric. The so-called elbow-point (where the change in WCSS begins to level off) was observed around 5 clusters (Fig. 3). In addition, the Davies-Bouldin index was performed on 3, 4, 5, 6, 7, 8, and

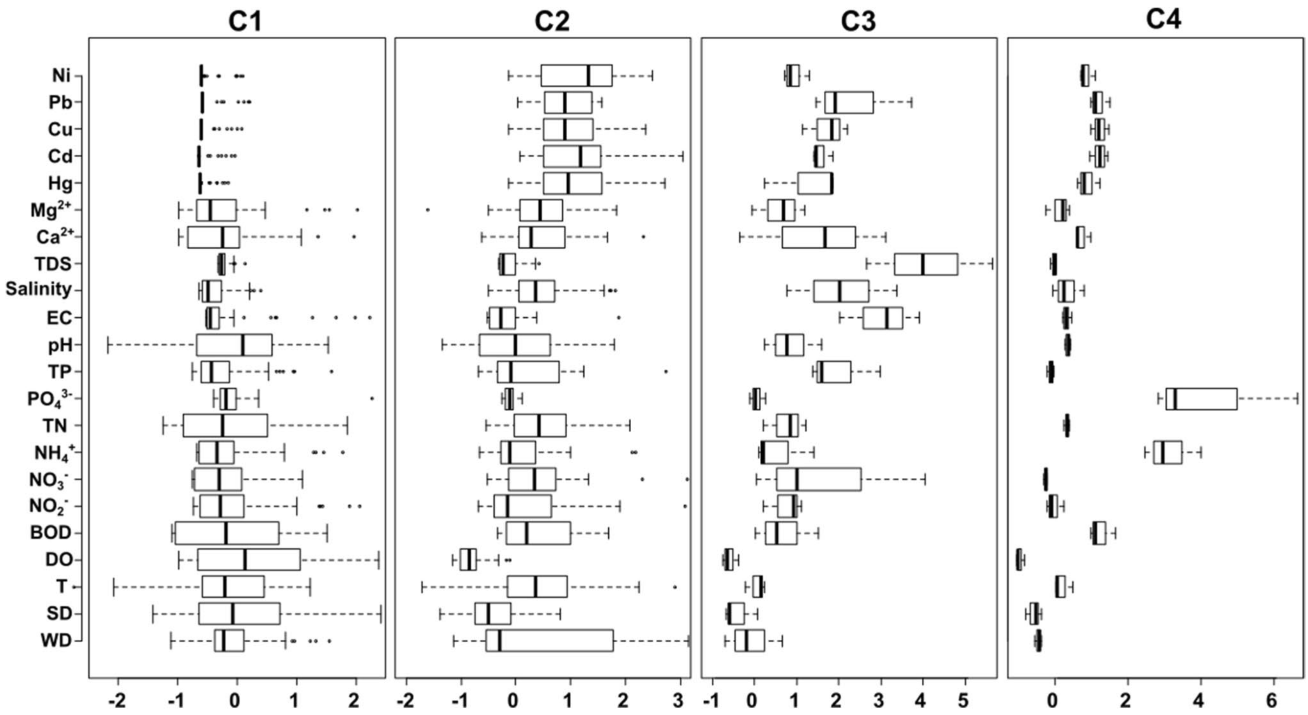


Fig. 4 Values of the weight vectors in SOM showing the variations in water quality among the four SOM clusters. The boxplots represent the 25th, 50th, and 75th percentiles, respectively, of Euclidean distance between SOM neurons whereas the whiskers (vertical dash of each boxplot) represent the 10^{th} and 90^{th} percentiles, respectively. Large boxes indicate large variations between neighbourhood neurons of each cluster, and the values for the variables (positive or negative) in each plot denote their importance in contributing to all neurons associated to the clusters.

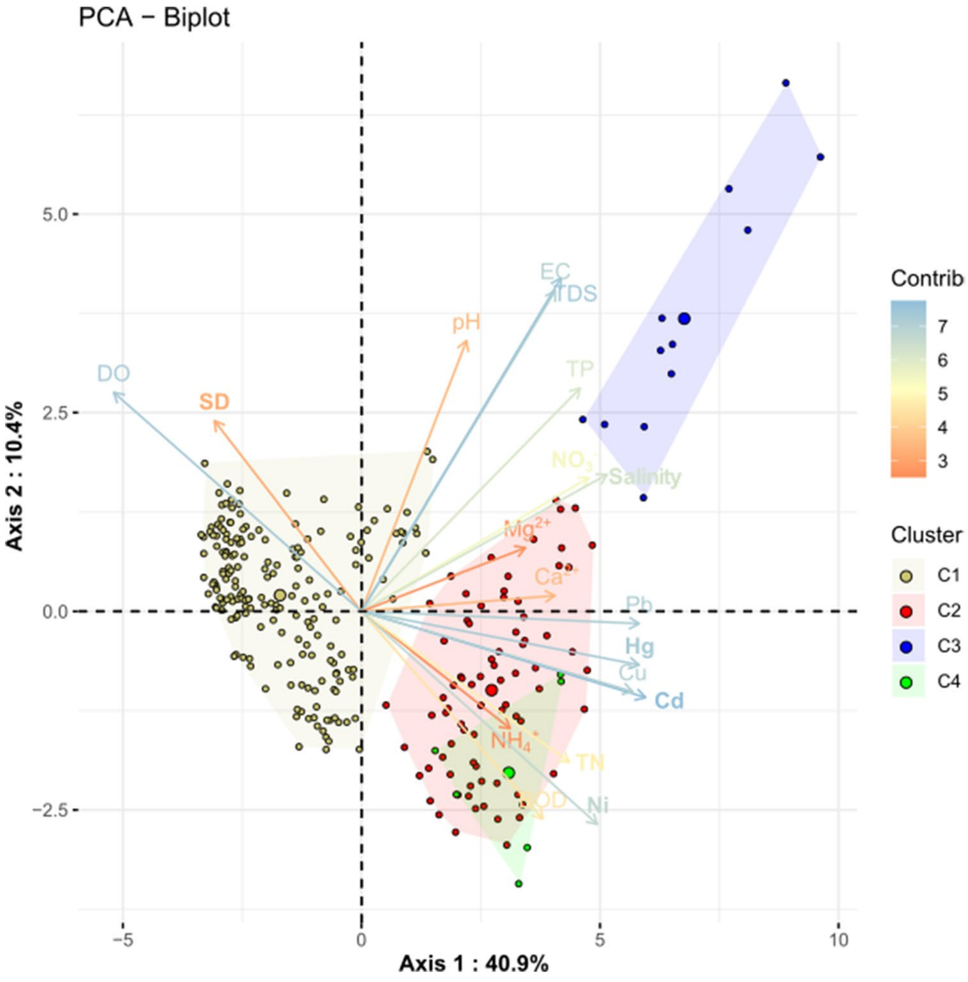


Fig. 5 The first two axes of the PCA of the samples differentiated by clusters identified with SOM.

9 looking for the lowest value indicating the best result (Gu et al. 2019). Afterwards, a hierarchical cluster analysis (HCA) was conducted with the number of clusters identified with the Davies-Bouldin index according to the Ward linkage method using Euclidean distance to define the cluster boundaries in the SOM units (Kohonen 2001). Different colours were used to each cluster to show their boundaries.

Deeper study of water quality assemblages using PCA and quantitative tests

Although SOM clusters can be used to study spatial and seasonal patterns of environmental data when the dataset submitted for analysis is built by considering these two factors, additional specific analyses (Fig. 2; step 5) are required to well examine contribution of each factor to the observed assemblages (Guo et al. 2020). The PCA was performed on all environmental variables to summarize their assemblages in the SRB and identify the main driving forces related to the variability observed in the clusters identified with the SOM. Correlations between the first two axes of PCA principle components (PCs) and the environmental variables were considered significant when the distances to the centre of the plane were $d > \sqrt{(2/n)}$, where = number of variables (Legendre and Legendre 2012). Therefore, variables such as WD, T, NO$_2^-$, and PO$_4^{3-}$ with $d \leq \sqrt{(2/22)} = 0.315$, were removed from the dataset and 18 variables were finally submitted to PCA. In addition, the PCA results were also visualized using sites and seasons as criteria of grouping to characterize each sites and seasons. The PCA was run using the package "FactoMineR" (Husson et al. 2020) while the package "factoextra" (Kassambara and Mundt 2019) allowed to extract and visualize the results in the R environment (R Core Team 2018). Standardized to a mean of zero and a standard deviation of each environmental variable were used in PCA (Li et al. 2018). In addition to PCA, the nonparametric Kruskal–Wallis and Dunn's tests were carried out considering two factors (clusters identified with SOM, and seasons) to further highlight the water quality features. The Kruskal–Wallis was performed with the *kruskal.test* function in the native "stats" package available in "dplyr" package (Wickham et al. 2020) allowing only to confirm or deny an existence of a difference with respect to a given variable among clusters and seasons. It was followed by Dunn's test, when adequate confirmation was obtained, using the *dunnTest* function in the “FSA” package (Derek and Wheeler 2020).

Pollution risk and sources

To assess the pollution risk of the SRB for local communities and aquatic biota, water quality variables were compared to the drinking-water standards at national scale (Decree N° 2001-094 of 20-02-2001) and international scale (WHO 2011) and to the ecological standards of surface ecosystems (Rodier et al. 2009), respectively.

Assessment of correlations among all variables extracted with the PCA will provide insights of the sources and migration of certain elements in the environment (e.g., Guo et al. 2020). Variables showing similar patterns and high correlations may share common sources, or analogous behaviours (Li et al. 2018) during transformation and migration under certain environmental conditions. This is especially true for variables having similar hydro-chemical characteristics in the water body such as nutrients (Pacheco et al. 2017; Li et al. 2018) and heavy metals (Zhang et al. 2016). In this study, we used the Kendall's Tau-b correlation to further explore correlations between heavy metals variables and with the other variables in the surface water layer. The function *cor.test* of the package “PerformanceAnalytics” (Peterson et al. 2020) was used for the correlation.

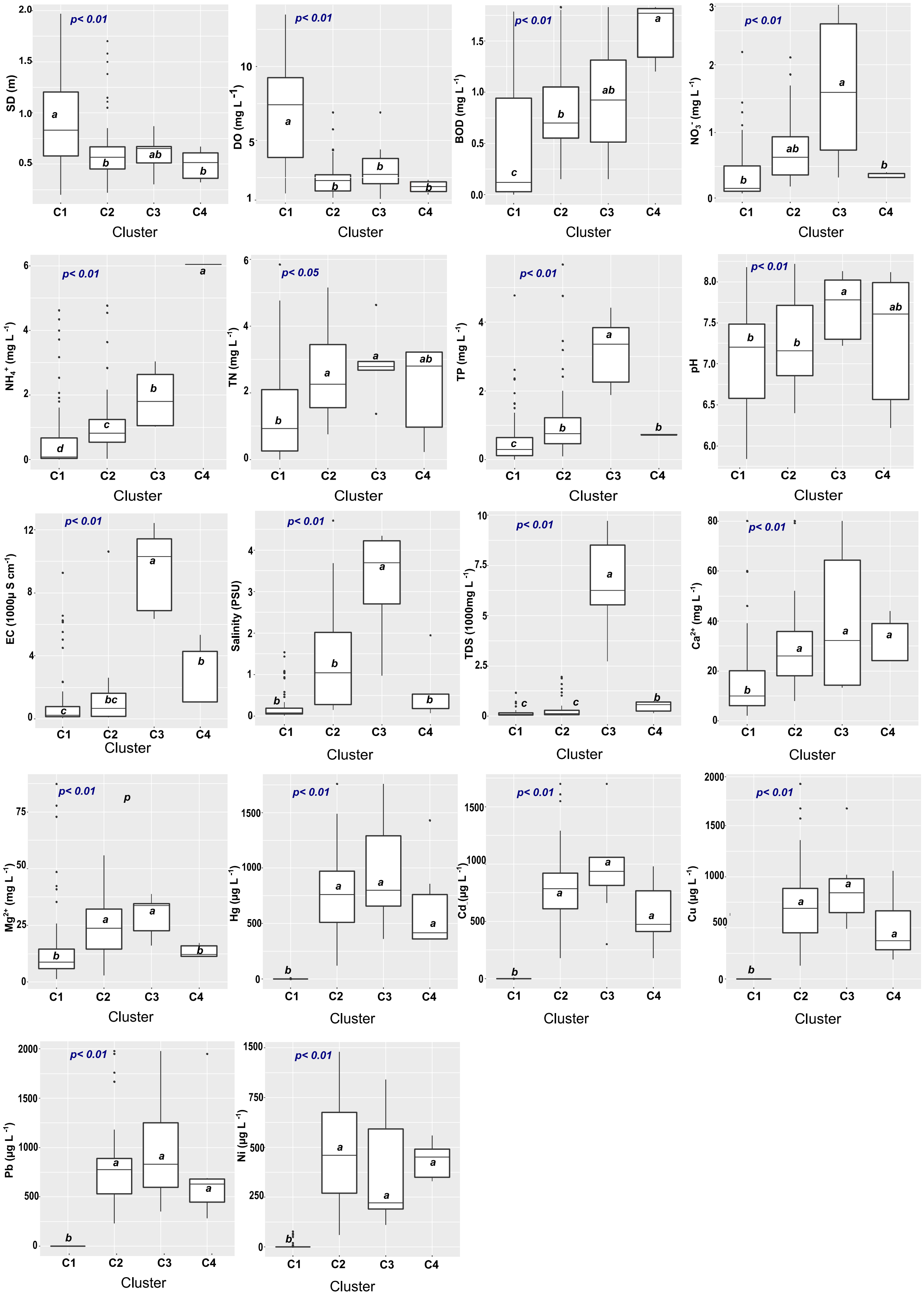
Results

Spatial patterns of water quality in the SRB

The SOM training at different map sizes revealed that an 84-units map organized in an array with 12 rows and 7 columns is the best compromise between the lowest QE (0.49) and TE (0.01) values, the number of neurons close to the 85 neurons fitted well Vesanto’s rule and the lowest number of empty output neurons. This grid size resulted in a distribution of 1 to 9 observations per neuron. Based on the minimum value of the Davies-Bouldin index (1.52), the optimal number of clusters is four (Fig. 3a). Therefore, to assess the spatial and seasonal variation of water quality, the 288 samples patterned in different neurons of the SOM were clustered into four groups: C1, C2, C3, and C4 with similar water quality (Fig. 3b) using the HCA. The cluster boundaries showed that two major clusters (C1 and C2) containing most of the samples (Fig. 3b).

The distribution of the samples within clusters mainly showed a geographical pattern (Fig. 3b, Table 2). Cluster C1, in the lower area of the SOM map, included 67% of the samples, mainly those located in the northern and middle area of the river. All samples of the four seasons from S5, S6, S7, S8, S9, S10, S11, and S12 are included while the whole samples of S1–S4 are missed. Cluster C2, included 27% of samples representing certain samples of southern part of the river (S1–S4). The Clusters C3 and C4 aggregated remaining samples of the southern area of the river (S1–S4), representing 4% and 2% of the whole samples, respectively. However, only S4 samples were represented in C4 whereas mainly S1 during LDS and SDS predominated in C3. Based on the contribution of each parameter (weight vector) to the identified clusters (Fig. 4), cluster C1 contributed relatively less by heavy metals, but was characterized by the highest level of DO, and SD among the four clusters (with weight vectors larger than −0.2). Clusters C2, C3, and C4 were characterized by high levels (i.e., with weight vectors generally larger than 0.5) of nutrients and heavy metals. However, levels of TDS, salinity, EC, and TP were the highest within C3 samples. Cluster C4 was in the transitional zone which was characterized by moderate metal levels (with weight vectors between 1 and 1.4) and the highest level of BOD, NH_4^+, and PO_4^{3-} (weight vector of variables larger than 1.6). A colour gradient of component planes (input environmental variables) showed that all variables excepted WD, T, NO_2^-, TN, and pH had a clear trend (Fig. S1; supplement file). SD and DO showed similar gradients. Their highest values were found in C1 and the lowest were observed mainly in C2 and C4. BOD, PO_4^{3-}, and NH_4^+ showed opposite trends. Heavy metals, Hg, Cd, Cu, Pb, and Ni, showed pronounced minima at the lower area (C1) and generally increased mainly towards the upper side (C4, C3, and C2). Ca^{2+}, and Mg^{2+} also paralleled this trend.

Further analysis of the clusters features was based on the 18 variables significantly correlated to the first two axes’ variables and contributing to the PCA structure. The total variance of the water quality variables explained by the first two axes was 51.28% (Fig. 5) with eigenvalues of $\delta 1_{=}$ 7.36, and $\delta 2$= 1.87. The spatial ordering of the C-planes and clusters validated the similarity between them. Generally, significantly distinct spatial patterns of water quality were found for all clusters except C2 and C4 which were overlapped. The highest variability of water quality indicated by the largest convex hull was observed in the C2 despite its intermediate number of samples. The first axis grouped samples from cluster C1 of the SOM on the left side, which represented mainly samples with a positive correlation with SD and DO and negative correlation with heavy metals (Hg, Ni, Pb, Cu, and Cd) and organic pollution indicators such as BOD, pH, NH_4^+, and TN. The clusters C2 and C4 features were opposite to the previous cluster and represented samples which have the highest positive correlation with heavy metals (Hg, Ni, Pb, Cu, and Cd) and high correlation with organic pollution indicators (BOD, pH, NH_4^+, and TN). However, the C4 is more positively correlated to the previous minerals’ variables than C2. Meanwhile, some samples of cluster C2 were strongly correlated to Ca^{2+} and Mg^{2+}. C3 was placed


SD (m)
2.0
1.5
1.0
0.5
p< 0.01
a
b
ab
b
C1
C2
C3
C4
Cluster
DO (mg L -1)
10
5
1
p< 0.01
a
b
b
b
C1
C2
C3
C4
Cluster
BOD (mg L -1)
1.5
1.0
0.5
0.0
p< 0.01
c
b
ab
a
C1
C2
C3
C4
Cluster
NO3- (mg L -1)
3
2
1
0
p< 0.01
b
ab
a
b
C1
C2
C3
C4
Cluster
NH4+ (mg L -1)
6
4
2
0
p< 0.01
d
c
b
a
C1
C2
C3
C4
Cluster
TN (mg L -1)
6
4
2
0
p< 0.05
b
a
a
ab
C1
C2
C3
C4
Cluster
TP (mg L -1)
4
2
0
p< 0.01
c
b
a
b
C1
C2
C3
C4
Cluster
pH
8.0
7.5
7.0
6.5
6.0
p< 0.01
b
b
a
ab
C1
C2
C3
C4
Cluster
EC (1000µ S cm-1)
12
8
4
0
p< 0.01
c
bc
a
b
C1
C2
C3
C4
Cluster
Salinity (PSU)
4
3
2
1
0
p< 0.01
b
b
a
b
C1
C2
C3
C4
Cluster
TDS (1000mg L -1)
10
7.5
5
2.5
0
p< 0.01
c
c
a
b
C1
C2
C3
C4
Cluster
Ca2+ (mg L -1)
80
60
40
20
0
p< 0.01
b
a
a
a
C1
C2
C3
C4
Cluster
Mg2+ (mg L -1)
75
50
25
0
p< 0.01
p
b
a
a
b
C1
C2
C3
C4
Cluster
Hg (µg L -1)
1500
1000
500
0
p< 0.01
b
a
a
a
C1
C2
C3
C4
Cluster
Cd (µg L -1)
1500
1000
500
0
p< 0.01
b
a
a
a
C1
C2
C3
C4
Cluster
Cu (µg L -1)
2000
1500
1000
500
0
p< 0.01
b
a
a
a
C1
C2
C3
C4
Cluster
Pb (µg L -1)
2000
1500
1000
500
0
p< 0.01
b
a
a
a
C1
C2
C3
C4
Cluster
Ni (µg L -1)
1500
750
500
250
0
p< 0.01
b
a
a
a
C1
C2
C3
C4
Cluster

◂**Fig. 6** Box plots comparing water quality variables in the four clusters derived from the SOM. The top-, mid-, and bottom-line of each box-plot represent the 75th, 50th, and 25th percentiles, respectively. The top and bottom-point of the vertical bar of each box-plot represent the 90th and 10th percentiles, respectively. Letters above or below median (50th) indicate statistical significance of differences between sites (post hoc pairwise comparisons): only sites with different letters differed significantly ($p < 0.05$) from each other.

by the two axes at the right side, representing most saline samples with intermediate heavy metals polluted. The main water quality variables positively correlated to axis 1 was Cd (0.83), Hg (0.81), Pb (0.81), Cu (0.79), Salinity (0.72), Ni (0.69), NO_3^- (0.67), TP (0.64), TN (0.61), whereas DO (−0.73), and SD (−0.43) were negatively correlated to the axis 1. The variables EC (0.56), TDS (0.59), pH (0.48), TP (0.39), DO (0.38), and Ni (−0.37) represented the main water quality variables correlated to axis 2.

Quantitative analyses revealed that there were significant differences in all water quality variables among the four clusters, which confirmed the effectiveness of the classification (Kruskal-Wallis test, $df = 3$, $p < 0.005$). Each cluster had its unique features (Fig. 6) representing specific water quality. The results of the pair-wise comparison of all clusters are included in Fig. 6.

Cluster CI was characterized by the lowest values of NH_4^+, TP, Salinity, TDS, and heavy metals (Hg, Cd, Cu, Pb, Ni) along with significantly high levels of SD and DO. The clusters C2, and C4 shared similar water quality features; these clusters had high levels of heavy metals (Cd, Cu, Pb, and Ni) and along with significantly low levels of SD and DO and intermediate values of salinity. Similar water compositions were seen in C2 and C3; these clusters had intermediate values for many water quality parameters such as BOD, and NH_4^+ along with the highest levels of Mg^{2+}, Ca^{2+}, and Hg. The median values Cd, Cu, Pb, and Ni were approximately equal in C2 and C3, although their concentration ranges were not significantly different to C4. C4 has the highest values of minerals variables such as BOD and NH_4^+. The cluster C3 was also distinguishable from the other clusters with its highest values of NO_3^-, TP, pH, EC, salinity, and TDS.

Latitudinal variation of water quality in the SRB

The PCA was used to further analyse the latitudinal variation of water quality, using sites as criteria of clustering (Fig. 7). All samples of the same site were restricted to specific area of the Axes 1 and 2. Overall, significantly different latitudinal patterns of water quality were found. Sites were distributed from the northern sites to the southern sites along the axis 1 from the left to right; left side was positively correlated to DO and SD while the opposite side was positively correlated to macronutrient and heavy metals. However, two clear clusters of sites could be observed. The southern sites (S1, S2, S3, and S4) were grouped and located at the right side of axis 1. They showed the highest variability in water quality especially for S1, S2, and S3. The middle and northern sites were grouped at the left side of axis 1 with two sub-groups and the site S8 overlapping between the two sub-groups. The far-left sub-group comprised S9, S10, S11, and S12 while the second sub-group contained S5, S6, and S7.

Seasonal variations of water quality in the SRB

Based on SOM output, generally, different trends were observed with the represented seasons within each cluster. The long seasons (LWS and LDS) were represented in all clusters while SDS and SWS samples were not observed in C3 and C4, respectively.

Based on the PCA results (Fig. 8) and using the seasons as criteria of grouping, it was observed that the highest variabilities of water quality indicated by the largest convex hull were observed in the wet seasons (SWS and LWS). Generally, most of samples of the four seasons are grouped at the left side of the axis 1 positively correlated to DO and SD. No significant characteristics of water quality were observed for the four seasons. However, most of the samples located at the upper right side of the PCA and positively correlated to the two axes are collected during wet seasons.

There were not significant differences in all water quality variables between seasons (Kruskal-Wallis, $df= 3$, $p > 0.05$), except for SD, TP, pH, and salinity (Kruskal-Wallis, $df =3$, $p < 0.05$). Specifically, water quality in the SDS was characterized by the highest levels of SD whereas lowest levels were observed for TP and pH (Fig. 9). The highest levels of TP and pH were observed in the wet seasons (SWS and LWS).

Pollution risk and correlation analysis of water quality variables in the SRB

None of the variables used as an indicator met the ecological standards for all samples, except for BOD, NO_2^- , and NO_3^- (Table 3). Generally, the samples of southern area (S1–S4; C3–C4) did not meet any of the standards for drinking-water quality (both Benin and WHO) and the ecological surface water quality. The concentrations of the heavy metals detected in the SRB did not meet the guidelines for drinking-water quality (both Benin and WHO) and the ecological surface water quality standards at the southern sites (S1–S4) and during the four seasons (considering means of sites values per season). However, the pH standards were not met only on sites S6 and S12 during dry seasons.

Kendall correlations among the 18 variables retained with the PCA showed 145 (~95 %) significant pair correlations (Fig. 10). One hundred and thirty-eight correlations were very significant ($p < 0.001$). Overall, relationships were

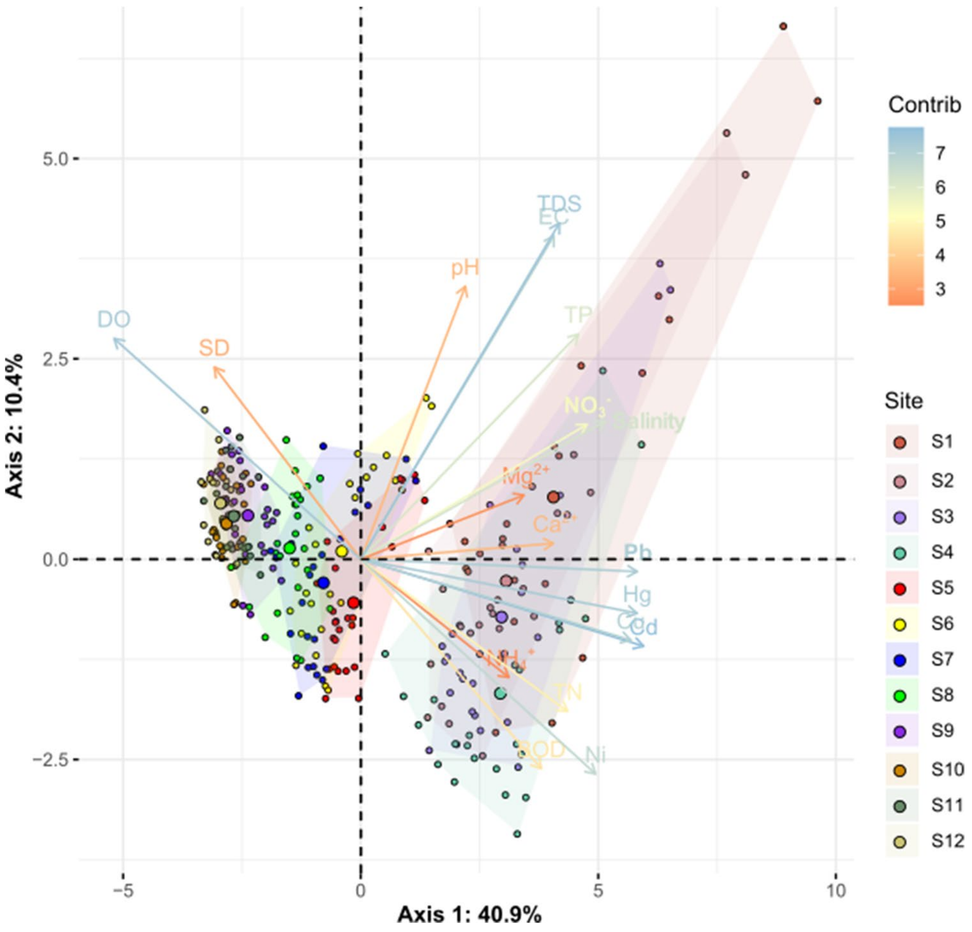


Fig. 7 PCA score plot of water quality samples differentiated by sites distributed along latitudinal of the SRB.

strong among: (i) organic pollution indicators, (ii) each pair of heavy metals and (iii) heavy metals and mineral variables. Heavy metals (Hg, Pb, Cu, Cd, and Ni) showed a very strong positive correlation ($0.89 \leq r \leq 0.93$). A moderate to fairly strong ($0.30 \leq r \leq 0.70$) moderate positive relationship was observed between heavy metals and nutrient enrichment variables (EC, TDS, salinity, Ca^{2+}, Mg^{2+}, BOD, NO_3^-, NH_4^+, TN, and TP). Each pair of the variables BOD, NO_3^-, NH_4^+, TN, TP, Salinity, Ca^{2+}, and Mg^{2+} showed moderate positive correlation against a moderate negative correlation observed between DO and SD was mainly observed.

Discussion

The SOM analysis alone yielded four clear clusters of samples irrespective of sites and seasons. However, spatial and latitudinal assemblages were evidenced. Moreover, water quality variables that contributed most to the discrimination between the four clusters were highlighted. As to the PCA, it revealed that 18 variables out 22 were most important. The PCA also depicted specific features of each cluster and the interrelation between C2 and C4, which might be considered unique cluster. More importantly, PCA showed that northern and middle reach sites grouped in C1 could be split in two clusters although the difference was very thin. Relying on these findings, three main clusters I (=C1), II (=C2 and C4) and III (=C3) will be considered in this discussion. SOM combined with multivariate analyses (HCA and PCA), and quantitative tests (Kruskal-Wallis and Kendall correlation) revealed a spatial-latitudinal pattern of surface water quality and pollution risk in a West–Africa River basin, the SRB. Correlation analysis allowed to identify potential sources and hotspots of organic and heavy metal pollutions which are important for sustainable management of aquatic ecosystems. Our study shows the benefit of SOM in conjunction with traditional statistical tools to analyse nonlinear ecological data, that could not be effectively evidenced by a single method or only with traditional methods. As such, our methodology is effective for assessing spatial and seasonal water quality dynamics, and can be used to understand the dynamic of aquatic processes based on relationships among numerous environmental variables. This is very important in the West African context, where insufficient expertise and a lack of

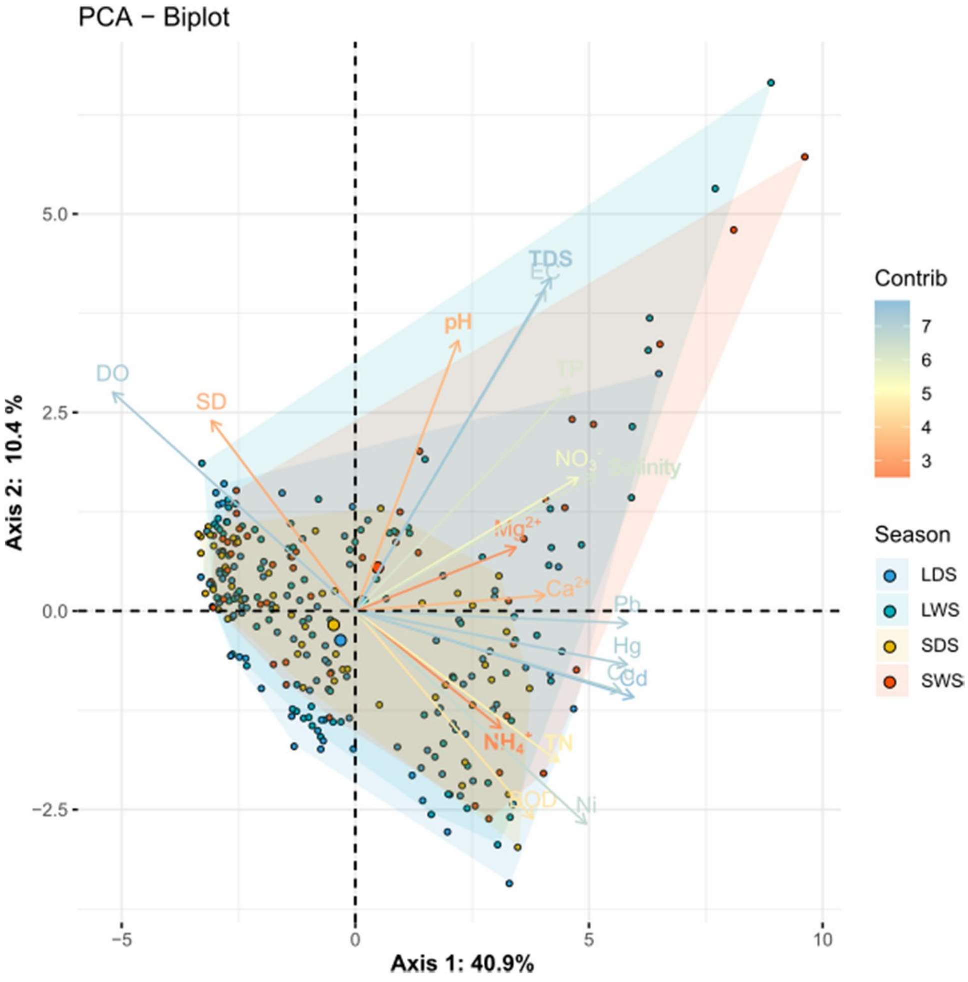
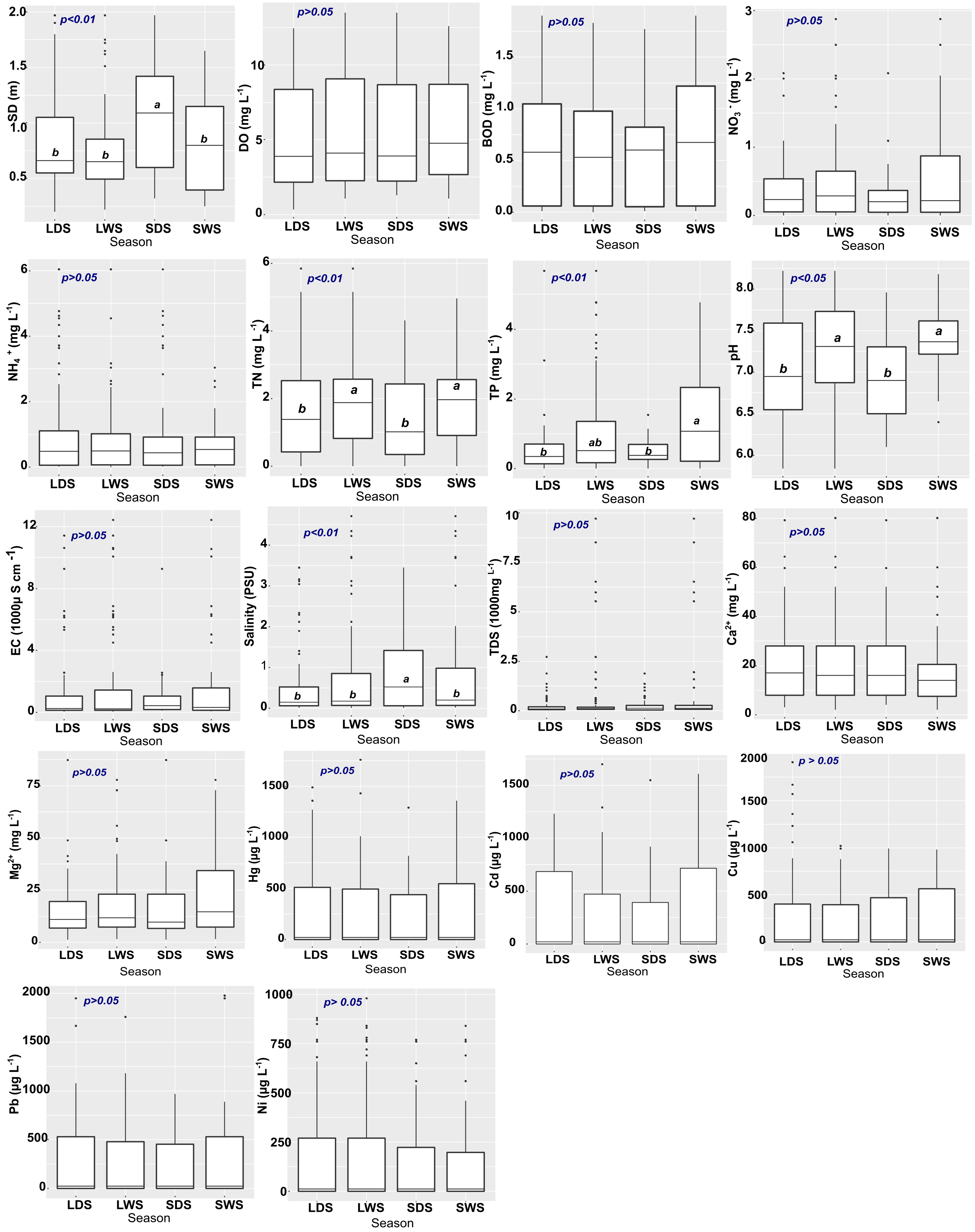


Fig. 8 Biplot based on the seasonal variations in water quality in the SRB.

data on water resources degradation by factors like pollution hampers the sustainable management of aquatic ecosystems (IUCN-PACO 2018).

Pattern of water quality from the headwater to estuary area

The SOM, HCA, and PCA eased the identification of pollution hotspots and clustering of samples following a latitudinal (upstream–downstream) gradient. The variability of water characteristics was defined by similar and complementary environmental features in clusters observed with SOM (I–III) and the PCA biplot based on sites (S1–S12). The highest values of DO and SD and the lowest values of almost all other variables in the middle and northern sites (I;S5–S12) suggested a relatively less impaired riverscape with eutrophication and heavy metals in the upstream reach compared to the southern sites' status present in groups II and III. Meanwhile, samples of the southern area grouped in group III was moderately polluted with eutrophication transitioning between the samples of group I (middle and northern sites) and II (majority of southern sites). These findings highlighted the latitudinal features of the river. Middle upper area sites (S9 and S10) are located in a less densely populated area, and constitutes relatively isolated areas that do not continually receive inflows and runoff from canal- and tributaries-associated pollution. As a result, the nearby downstream sites such as S8, S7, S6, and S5 showed less impaired characteristics. More importantly, S11 and S12 located in the northern part are close to the Djigbé forest which provides suitable habitat for many forest species (see Kakpo et al. 2019). The area has low navigation activities which could negatively affect the water quality. These factors could have contributed to maintain the water quality as backed in previous studies around the world (Pacheco et al. 2017; Guo et al. 2020). The relative pristine state of sites S5 – S12 is typical of headwaters (Frid and Dobson 2013). Many West–Africa rivers are probably in relative pristine state as river water quality is mainly determined by its catchment and human activities (Dobson and Frid 2009) and regional countries share similar challenges in terms of pollution type (i.e., plastic waste, organic pollution, untreated sewage, and deforestation) (IUCN-PACO 2018; McClure 2021), although intensity and amplitude could be


SD (m)
p<0.01
b b a b
DO (mg L⁻¹)
p>0.05
BOD (mg L⁻¹)
p>0.05
NO₃⁻ (mg L⁻¹)
p>0.05
NH₄⁺ (mg L⁻¹)
p>0.05
TN (mg L⁻¹)
p<0.01
b a b a
TP (mg L⁻¹)
p<0.01
b ab b a
pH
p<0.05
b a b a
EC (1000µ S cm⁻¹)
p>0.05
Salinity (PSU)
p<0.01
b b a b
TDS (1000mg L⁻¹)
p>0.05
Ca²⁺ (mg L⁻¹)
p>0.05
Mg²⁺ (mg L⁻¹)
p>0.05
Hg (µg L⁻¹)
p>0.05
Cd (µg L⁻¹)
p>0.05
Cu (µg L⁻¹)
p > 0.05
Pb (µg L⁻¹)
p>0.05
Ni (µg L⁻¹)
p> 0.05
LDS
LWS
SDS
SWS
Season

◂**Fig. 9** Box plot of the water quality variables in the four seasons. The top-, mid-, and bottom-line of each box-plot represent the 75th, 50th, and 25th percentiles, respectively. The top and bottom-point of the vertical bar of each box-plot represent the 90th and 10th percentiles, respectively. Letters above or below median (50th) vertical bars indicate statistical significance of differences between sites (post hoc pairwise comparisons): only sites with different letters differ significantly ($p < 0.05$).

different. The high presence of fishes such as *Brycinus nurse* (Rüppell, 1832), *Brycinus longipinnis* (Günther, 1864), and *Distichodus rostratus* Günther, 1864 in head water of Boubo River (Ivory Coast) and Ikpoba River (Nigeria) as observed in SRB (Koné et al. 2003; Kouamélan et al. 2003; Hazoume 2018) could be considered as evidence of both the relatively low pollution of headwaters and similarities between West–African rivers. Sustainable aquaculture initiatives could be promoted in head water, which are areas in relative pristine state and predisposing to high fishery potential yield.

On the contrary, in the most polluted southern sites (II; S1–S4), the very high levels of nitrogen and phosphorus compounds (above 0.5 mg L^{-1}) observed is related to environmental impacts of nutrient enrichment, which is sufficient to trigger excessive vegetation growth (WFD 2005). Aquatic ecosystem with high level of phosphorus and nitrates benefits the rapid multiplication of macrophytes as observed in the SRB and algal bloom which induce depletion of oxygen concentration in the water layers (~30 times less than in the same volume of air), especially when biological process demand exceeds the oxygen amount produced by photosynthesis and atmospheric oxygen dilution (Koudenoukpo 2018). This southern area is relatively densely populated and characterised many coastal West–African countries. The lifestyle of people living in the area and their activities were suspected to impact both the organic and heavy metal pollutions observed. Discharge wastewaters from municipal including primary and secondary school, domestic and churches, and agricultural sources could be the cause of the contaminated water leading to a low water quality. In addition, years of unchecked pollution from industry and inefficient and degraded sewage infrastructure resulting in raw sewage causing the water system to be polluted (McClure 2021). To overcome with this scourge pollution, investment is needed in urban infrastructure that limits water pollution, such as adequate sewerage and wastewater treatment facilities, household and industrial waste disposal and functioning storm drains.

In the group III (mainly S1), the pollution level in organic matter and heavy metals were observed along with high level of mineral compounds (i.e., salinity, TDS, and EC). These variables usually correlated (Rodier et al. 2009) highlighted a salt water intrusion via the Lake Nokoué. The frequent water exchanges of the SRB with the Nokoué Lake is an important factor determining the water quality of both ecosystems (Hazoume 2018; Koudenoukpo 2018) as observed in Ogun River connected with the Lagos lagoon (Fajemila et al. 2020). In such cases, it is necessary to maintain not only the integrity of the rivers but also that of the lakes and lagoons with which they interact, especially as the latter are densely populated areas with a high concentration of human activities.

Although the spatial variation mainly contributed to water quality, the assignment of certain sampling sites to groups indicated also seasonal variation (II and III) and discussed here below.

Seasonal pattern of water quality

We found non-significant seasonal variations of water quality in the SRB. Although our data covered both spatial and temporal variation, the SOM mainly highlighted the spatial patterns, confirming the PCA result with seasonal factor. Specific conditions and activities occurring at each location mostly contributed to water quality. These findings for seasonal variations were somewhat inconsistent with many other studies, which identified seasons as crucial factor affecting water quality condition (Jaji et al. 2007; Rodrigues et al. 2018; Cruz et al. 2019). For example, in Ogun River, South West Nigeria (Jaji et al. 2007), it was observed that the river was more polluted during dry season than wet season. Inconsistence with our results may be related to difference in hydrology and features of basins examined in the latter study and the one of SRB. Although Ogun River like SRB was in a West–Africa, difference in sampling period and climate changes could have contributed to the observed results. Climate change has strong impact on the seasonal distributions of the streamflow and nutrient load (Tu 2009) and become more important during last decade. The climatic changes have led, among others, to changes in seasonal onset, duration and patterns, increase in rainfall variability, with more unexpected dry period and heavy rainfall periods (Baudoin et al. 2014). Rains become more important during a period previously considered dry and less during wet seasons (Totin 2010). These unexpected weather patterns make it difficult to allocate the samples properly between seasons without introducing bias because expected temperature, rainfall and flood levels are no longer valid. As result, temporal variation, even if present, can be difficult to highlight through a seasonal sampling.

A year sampling on the same ecosystem showed more clearly a seasonal variation in water quality (Koudenoukpo et al. 2021). The fact that with two years, the seasonal variation is less clear, meaning that seasonal variation is not regular across years. So, a higher order temporal scale (year-to-year) variation should be envisaged in future studies. However, higher among-sites variability in water quality parameters as well as high levels of TP and pH were observed in wet seasons (SWS and LWS), indicating a high risk of organic pollution and eutrophication (Djihouessi and Aina 2018). Permanent high risk of eutrophication

Table 3 Criteria of water quality indicators used for water quality assessment with threshold values.

Variable	Range	Median	Beninese standards[*]	WHO standards[#]	Ecological standards[Ø]
BOD (mg L^{-1})	0.001–1.9	0.61	≤ 25	–	< 8
DO (mg L^{-1})	0.34–13.5	4.15	–	–	> 5
pH	5.84–8.22	7.22	6.5 – 8.5	6.5 – 8.5	6 – 9
EC (µS Cm^{-1})	49.5–12432	278	–	–	< 1500
TDS (mg L^{-1})	22.2–9720	113	–	< 1000	–
Nitrite (mg L^{-1})	0.001–1.849	0.118	< 3.2	< 03	< 03
Nitrate (mg L^{-1})	0.002–2.878	0.2335	< 45	< 50	< 50
Ammonium (mg L^{-1})	0.005–6.041	0.483	–	–	< 1
Total Nitrogen (mg L^{-1})	0.001–5.849	1.544	< 10	–	<3
Total Phosphorus (mg L^{-1})	0.004–5.676	0.4545	< 01	–	–
Magnesium (mg L^{-1})	1.319–87.36	11.65	< 50	–	< 36
Calcium (mg L^{-1})	2.039–80.15	16.031	< 100	–	< 60
Copper (µg L^{-1})	0.12–1910	0.12	< 2000	< 2000	< 1000
Cadmium (µg L^{-1})	0.08–1700	0.18	< 5	< 3	< 200
Lead (µg L^{-1})	0.08–1980	0.08	< 50	< 10	< 100
Nickel (µg L^{-1})	0.65–980	0.65	< 20	< 70	< 200
Mercury (µg L^{-1})	0.02–1760	0.19	< 1	< 1	< 2

[*] The quality standards for drinking-water in Benin (Decree No. 2001-094 of 20-02-2001)

[#] Guidelines for drinking-water quality (WHO 2011)

[Ø] Guidelines for aquatic ecosystems water quality to support biodiversity (Rodier et al. 2009)

may be attributed to the non-point pollution sources which accumulated and discharged from the canals and Ouémé River (Mama 2010). The specificity of the southern Benin's aquatic ecosystems such as SRB is that the SWS corresponds to the single wet season of the Northern Benin, considered as the cotton basin where large quantities of chemical fertilisers were used for the cotton cropping (Odountan et al. 2019). Large volumes of water loaded in nutrient (e.g., phosphorus and nitrogen) from the northern area enter the SRB and contribute to the river dilution. As a result, the water hyacinth and others macrophytes are abundant especially in SWS. The increased salinity during dry seasons, especially at southern sites (group III) were related to their connection with Lake Nokoué. The Lake Nokoué acted as reservoir of large proportion of pollutants, including the nutrients contributing to eutrophication during wet seasons for the SRB river–estuary system (Djihouessi and Aina 2018). During dry seasons with salt intrusion, the portion of pollutants not flushed into the Atlantic Ocean were returned to the river. This included rainfall and the dissolved matter carried from the littoral and agricultural areas into the ecosystem. Therefore, less difference in water quality was observed across pre-defined seasons and between sites.

Water quality, human activities, and pollution risk

Assessment of environmental variables such as WD, T, SD, DO, BOD, nitrogen compounds (NO_2^-, NO_3^-, NH_4^+, and TN), phosphorus compounds (PO_4^{3-} and TP), minerals variables (pH, TDS, EC, TDS, Ca^{2+}, and Mg^{2+}), and heavy metals is important to the characterization of surface waters because each variable reflects an aspect of the water body, has a different origin, with a specific distribution with respect to other factors (Aguirre et al. 2019). Pollution of surface water in relation to these variables is of great environmental concern in developing countries which have limited expertise, techniques, and facilities to address these threats. In this study, water quality variables used as indicators did not meet neither the drinking-water standards nor the ecological standards. The head water is just less polluted compared to southern sites and is not really in pristine state. These findings were consistent with those backed in two other aquatic ecosystems communicating with the SRB, the Ouémé River –the longest and largest river of Benin (Hounsou et al. 2011) and the Lake Nokoué (Dovonou et al. 2011). Pollution of surface water with traces elements and excess nutrients leading to eutrophication (coupled to toxic metals issues) is of paramount concern in Benin. The direct human-consumption of the water of these ecosystems without any treatment is not recommendable and prejudicial (Hounsou et al. 2011). As to the potential sources of eutrophication and heavy metal pollution, the very strong Kendall correlation between heavy metals (Hg, Cd, Cu, Pb, and Ni) pointed out the close relation between the five traces elements. Moreover, the high correlation of heavy metals with minerals contents allows to presume non-point

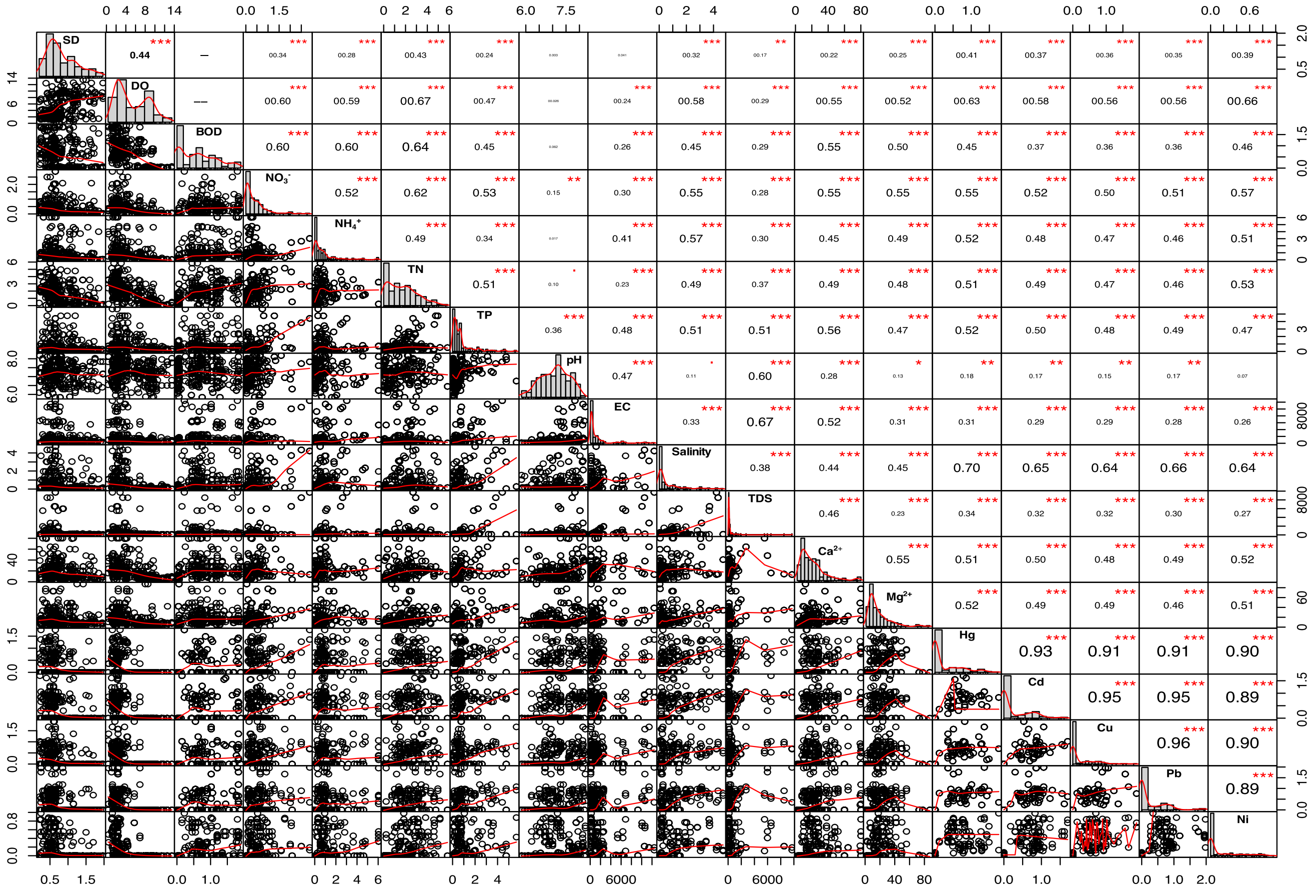


Fig. 10 Correlation matrix of water quality variables of SRB. (The distribution of each variable is shown on the diagonal. Bivariate scatter plots with fitted lines are displayed in the lower triangular portion of the matrix; the correlation values are shown in the upper triangular portion of the matrix. Statistical significance levels are denoted as "***", "**", "*", ".", and " " corresponding to p-values in the intervals of (0, 0.001, 0.01, 0.05, 0.1, 1))

sources including agricultural activities and domestic sewage discharges that are usually pointed out to contribute to nutrient enrichment (Najar and Khan 2012; Djihouessi and Aina 2018; Odountan et al. 2019; Guo et al. 2020) and heavy metal pollution (Calamari and Naeve 1994). The highest values of heavy metals around the locations of housing in the river villages (south of SRB) seem be related to the illegal traffic of gasoline from Nigeria with inappropriate boats and cans full of gasoline discharging in the river result in leaks into the water (Kaki et al. 2011; Djihouessi and Aina 2018). This hypothesis is also supported by the highest correlation between salinity and heavy metals. Indeed, sites with the highest salinity rate are located downstream and directly connected to the Lake Nokoué, receiving during dry season Atlantic Ocean water. As to nutrients, our study supports that, inputs from Ouémé River, domestic sewage discharge and Lake Nokoué releasing during dry seasons are greatest sources of nutrients (Mama 2010; Djihouessi and Aina 2018). The strong positive correlations observed between nutrient enrichment variables such as BOD, NH_4^+, NO_3^-, and TP, with salinity (higher downstream especially at southern sites connected to Lake Nokoué) confirmed this hypothesis. Nutrients from agricultural activities originate from non-point pollution sources are known to be the major sources of water quality impairment (Park et al. 2014).

Overall, heavy metal pollution can be linked to (i) the "illegal" solid waste dumps occurrence from the communities along the river's watershed; (ii) the destruction of the riparian vegetation promoting proliferation of landfills on the river's shores; (iii) the salt intrusion of Lake Nokoué water contaminated by various domestic and toxic wastes from the Dantokpa Market (situated south of the lake), as well as inputs of effluents and solid waste from Cotonou and Abomey-Calavi urban activities; and (iv) inappropriate transport of gasoline. As to the nutrients they are determined primarily by external inputs and modified by hydroclimatic processes. Sustainability in agriculture and measures of sanitation and water treatment are required to prevent further degradation of the ecosystem.

Conclusion

The spatial-latitudinal and seasonal patterns as well as the water pollution processes in SRB were assessed using SOM and multivariate analyses. The combination of these tools highlighted complex interrelations between physicochemical variables within SRB. This finding is useful in simplifying the complex ecological information, and notably to identify key variables to overcome with eutrophication and heavy metal pollution in West–Africa River systems. Latitudinal patterns revealed significant difference in water quality condition between upstream and downstream. Specifically, head water area showed less impaired features and can be suitable for sustainable aquaculture whereas middle reach and downstream area showed a progressive increase of organic and heavy metal pollution. As to the seasonal pattern, no significant difference in water quality was observed across seasons and does not mean no temporal variation in water quality. Due to climate changes which have led, among others, to changes in seasonal onset, duration and patterns, seasonal sampling might not be able to highlight temporal variations. Although, the results indicate that SRB water quality is strongly dependent on anthropogenic activities, the hydroclimatic processes could have also contributed to the patterns observed. The level of eutrophication and heavy metal pollution were high, and the ecosystem did not meet neither drinking-water standards nor ecological standards. Based on these findings, an efficient water management plan for the watershed is needed to reduce the impact of human activities on the water quality especially in the downstream area, and consequentially, to ensure fewer negative effects on human and aquatic biota.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s11356-022-22784-5.

Acknowledgements ZCK and OHO acknowledge CAPE BIO NGO through the Biodiversity and Sustainable Productivity of Aquatic Ecosystem of Ouémé Delta project. Special thanks are due to Dèdéou Apocalypse Tchokponhoue for the valuable comments on the first draft of the manuscript.

Author contribution ZCK: conceptualization, methodology, investigation, funding acquisition, resources, data curation, writing—review and editing; OHO: conceptualization, methodology, data curation, formal analysis, software, validation, visualization, writing—original draft, writing—review and editing; CG: methodology, validation, writing—review and editing; RC: methodology, data curation, validation, visualization, writing—review and editing; AC: conceptualization, methodology, funding acquisition, resources, writing—review and editing; Y-SP: conceptualization, methodology, writing—review and editing. All authors read and approved the manuscript.

Funding This study was supported by the grant provided by the Benin Higher Ministry of Education and Scientific Research to ZCK.

Data availability The datasets used and/or analysed in the current study are available on reasonable request from the corresponding author.

Declarations

Ethics approval and consent to participate Not applicable.

Consent for publication Not applicable.

Competing interests The authors declare no competing interests.

References

Aboelnour M, Gitau MW, Engel BA (2020) A comparison of streamflow and baseflow responses to land-use change and the variation in climate parameters using SWAT. Water (Basel) 12:191. https://doi.org/10.3390/w12010191

Aguirre BP, Masachessi G, Ferreyra LJ et al (2019) Searching variables to assess recreational water quality: the presence of infectious human enterovirus and its correlation with the main variables of water pollution by multivariate statistical approach in Córdoba, Argentina. Environ Sci Pollut Res 26:6586–6601. https://doi.org/10.1007/s11356-019-04124-2

Alvarez-Guerra M, González-Piñuela C, Andrés A et al (2008) Assessment of Self-Organizing Map artificial neural networks for the classification of sediment quality. Environ Int 34:782–790. https://doi.org/10.1016/j.envint.2008.01.006

Aniyikaiye TE, Oluseyi T, Odiyo JO, Edokpayi JN (2019) Physico-chemical analysis of wastewater discharge from selected paint industries in Lagos, Nigeria. Int J Environ Res Public Health 16:1235. https://doi.org/10.3390/ijerph16071235

Baudoin MA, Sanchez AC, Fandohan B (2014) Small scale farmers' vulnerability to climatic changes in southern Benin: the importance of farmers' perceptions of existing institutions. Mitig Adapt Strateg Glob Chang 19:1195–1207. https://doi.org/10.1007/s11027-013-9468-9

Berrada M, El Hmaidi A, Monyr N et al (2016) Self-organizing map for the detection of seasonal variations in Sidi Chahed Dam sediments (Northern Morocco). Hydrol Sci J 61:628–635. https://doi.org/10.1080/02626667.2014.964717

Calamari D, Naeve H (1994) Revue de la pollution dans l'environnement aquatique africain. FAO, Rome

Céréghino R, Park Y-S (2009) Review of the Self-Organizing Map (SOM) approach in water resources: commentary. Environ Model Softw 24:945–947. https://doi.org/10.1016/j.envsoft.2009.01.008

Chea R, Grenouillet G, Lek S (2016) Evidence of water quality degradation in lower Mekong basin revealed by Self-Organizing Map. PLoS One 11:e0145527. https://doi.org/10.1371/journal.pone.0145527

Cruz MAS, Gonçalves AD, de Aragão R, de Amorim JR et al (2019) Spatial and seasonal variability of the water quality characteristics of a river in Northeast Brazil. Environ Earth Sci 78:68. https://doi.org/10.1007/s12665-019-8087-5

Dai L, Wang L, Li L et al (2018) Multivariate geostatistical analysis and source identification of heavy metals in the sediment of Poyang Lake in China. Sci Total Environ 621:1433–1444. https://doi.org/10.1016/j.scitotenv.2017.10.085

Derek A, Wheeler P (2020) Package 'FSA': simple fisheries stock assessment methods. R package version 0.8.30. https://www.rdocumentation.org/packages/FSA/versions/0.8.30

Djihouessi MB, Aina MP (2018) A review of hydrodynamics and water quality of Lake Nokoué: current state of knowledge and prospects for further research. Reg Stud Mar Sci 18:57–67. https://doi.org/10.1016/j.rsma.2018.01.002

Dobson M, Frid C (2009) Ecology of Aquatic Systems, 2nd edn. Oxford University Press, Oxford

Dovonou F, Aina M, Boukari M, Alassane A (2011) Pollution physico-chimique et bactériologique d'un écosystème aquatique et ses risques écotoxicologiques: cas du lac Nokoue au Sud Benin. Int J Biol Chem Sci 5:1590–1602. https://doi.org/10.4314/ijbcs.v5i4.23

Fajemila OT, Sariaslan N, Langer MR (2020) Spatial distribution of benthic foraminifera in the Lagos Lagoon (Nigeria): tracing the impact of environmental perturbations. PLoS One 15:e0243481. https://doi.org/10.1371/journal.pone.0243481

Frid C, Dobson M (2013) Ecology of aquatic management, 2nd edn. Oxford University Press, Oxford

Gnohossou PM (2006) La faune benthique d'une lagune ouest africaine (le lac Nokoue au Benin), diversité, abondance, variations temporelles et spatiales, place dans la chaine trophique. PhD thesis. Institut National Polytechnique de Toulouse, France

Gu Q, Hu H, Ma L et al (2019) Characterizing the spatial variations of the relationship between land use and surface water quality using self-organizing map approach. Ecol Indic 102:633–643. https://doi.org/10.1016/j.ecolind.2019.03.017

Guo C, Chen Y, Xia W et al (2020) Eutrophication and heavy metal pollution patterns in the water suppling lakes of China's south-to-north water diversion project. Sci Total Environ 711:134543. https://doi.org/10.1016/j.scitotenv.2019.134543

Haselbeck V, Kordilla J, Krause F, Sauter M (2019) Self-organizing maps for the identification of groundwater salinity sources based on hydrochemical data. J Hydrol 576:610–619. https://doi.org/10.1016/j.jhydrol.2019.06.053

Hazoume RUS (2018) Diversité, organisation trophique et exploitation des poissons de la rivière Sô au Benin (Afrique de l'Ouest). PhD Thesis. Université d'Abomey-Calavi, Bénin

Hong H, Zhang Z, Guo A et al (2020) Radial basis function artificial neural network (RBF ANN) as well as the hybrid method of RBF ANN and grey relational analysis able to well predict trihalomethanes levels in tap water. J Hydrol 591:125574. https://doi.org/10.1016/j.jhydrol.2020.125574

Hounsou MB, Ahamide B, Agbossou EK, Gaiser T (2011) Evaluation of water quality in the Ouémé River (Bénin). Environmentalist 31:407–415. https://doi.org/10.1007/s10669-011-9356-3

Husson F, Josse J, Le S, Mazet J (2020) Package "FactoMineR": multivariate exploratory data analysis and data mining. R package version 2.3. https://cran.microsoft.com/snapshot/2020-04-13/web/packages/FactoMineR/index.html

Ibikounlé M, Gbédjissi LG, Ogouyèmi-Hounto A et al (2014) Schistosomose et géohelminthoses dans le nord-est du Bénin: cas des écoliers des communes de Nikki et de Pèrèrè. Bull de la Société de Pathologie Exotique 107:171–176. https://doi.org/10.1007/s13149-014-0344-y

ISO 5667-5 (2006) Water quality — Sampling — Part 5: Guidance on sampling of drinking water from treatment works and piped distribution systems. https://www.iso.org/standard/36694.html

ISO 5667-6 (2014) Water quality — Sampling — Part 6: Guidance on sampling of rivers and streams. https://www.iso.org/standard/55451.html

IUCN-PACO (2018) Partnership for Environmental Governance in West Africa. West Africa Water Resources Policy (WAWRP), Burkina Faso

Jaji MO, Bamgbose O, Odukoya OO (2007) Water quality assessment of Ogun river, South West Nigeria. Environ Monit Assess 133:473–482. https://doi.org/10.1007/s10661-006-9602-1

Joshi YK, Tortajada C, Biswas AK (2012) Cleaning of the Singapore River and Kallang Basin in Singapore: human and environmental dimensions. Ambio 41:777–781. https://doi.org/10.1007/s13280-012-0279-0

Kaki C, Guedenon P, Kelome N et al (2011) Evaluation of heavy metals pollution of Nokoue Lake. Afr J Environ Sci Technol 5:255–261

Kakpo SB, Aoudji AKN, Gnanguènon-Guéssè D et al (2019) Spatial distribution and impacts of climate change on Milicia excelsa in Benin, West Africa. J For Res. https://doi.org/10.1007/s11676-019-01069-7

Kassambara A, Mundt F (2019) Package "factoextra": extract and visualize the Results of multivariate data analyses. R package version 1.0.6. https://cran.r-project.org/src/contrib/Archive/factoextra/factoextra_1.0.6.tar.gz

Kohonen T (1990) The Self-Organizing map. Proc IEEE 78:1464–1480. https://doi.org/10.1109/5.58325

Kohonen V (2001) Towards experiential foreign language education. In: Kohonen V, Jaatinen R, Kaikkonen P, Lehtovaara J (eds) Experiential learning in foreign language education. Routledge: Taylor and Francis Group, Milton Park, Oxfordshire, pp 8–60

Kohonen T (2013) Essentials of the self-organizing map. Neural Netw 37:52–65. https://doi.org/10.1016/j.neunet.2012.09.018

Koné T, Teugels GG, N'Douba V et al (2003) Fish assemblages in relation to environmental gradients along a small west African coastal basin, the San Pedro River, Ivory Coast. Afr J Aquat Sci 28:163–168. https://doi.org/10.2989/16085910309503780

Kouamélan EP, Teugels GG, N'Douba VN et al (2003) Fish diversity and its relationships with environmental variables in a West African basin. Hydrobiologia 505:139–146. https://doi.org/10.1023/b:hydr.0000007302.74296.84

Koudenoukpo C (2018) Evaluation de la qualité écologique de la rivière So au Sud Benin: diversité et distribution des assemblages de zooplancton et des macroinvertébrés aquatiques. PhD Thesis. Université d'Abomey-Calavi, Bénin

Koudenoukpo ZC, Odountan OH, Van Bocxlaer B et al (2020) Checklist of the fresh and brackish water snails (Mollusca, Gastropoda) of Bénin and adjacent West African ecoregions. Zookeys 942:21–64. https://doi.org/10.3897/zookeys.942.52722

Koudenoukpo ZC, Odountan OH, Agboho PA et al (2021) Using self–organizing maps and machine learning models to assess mollusc community structure in relation to physicochemical variables in a West Africa river–estuary system. Ecol Indic 126:107706. https://doi.org/10.1016/j.ecolind.2021.107706

Legendre P, Legendre L (2012) Numerical ecology, 3rd edn. Elsevier Science Publishers B.V, Amsterdam

Li T, Sun G, Yang C et al (2018) Using self-organizing map for coastal water quality classification: Towards a better understanding of patterns and processes. Sci Total Environ 628–629:1446–1459. https://doi.org/10.1016/j.scitotenv.2018.02.163

Liao X, Tao H, Gong X, Li Y (2019) Exploring the database of a soil environmental survey using a geo-self-organizing map: a pilot study. J Geogr Sci 29:1610–1624. https://doi.org/10.1007/s11442-019-1644-8

Mama D (2010) Méthodologie et résultats du diagnostic de l'eutrophisation du Lac Nokoue (Bénin). PhD Thesis. Université de Limoges, France

Mamun M, An K-G (2018) Ecological health assessments of 72 streams and rivers in relation to water chemistry and land-use patterns in South Korea. Turk J Fish Aquat Sci 18:871–880. https://doi.org/10.4194/1303-2712-v18_7_05

McClure M (2021) What are the causes and effects of water pollution in Africa?

Najar IA, Khan AB (2012) Assessment of water quality and identification of pollution sources of three lakes in Kashmir, India, using multivariate analysis. Environ Earth Sci 66:2367–2378. https://doi.org/10.1007/s12665-011-1458-1

Odountan OH, Janssens de Bisthoven L, Koudenoukpo CZ, Abou Y (2019) Spatio-temporal variation of environmental variables and aquatic macroinvertebrate assemblages in Lake Nokoué, a RAMSAR site of Benin. Afr J Aquat Sci 44:219–231. https://doi.org/10.2989/16085914.2019.1629272

Olawoyin R, Nieto A, Grayson RL et al (2013) Application of artificial neural network (ANN)–self-organizing map (SOM) for the categorization of water, soil and sediment quality in petrochemical regions. Expert Syst Appl 40:3634–3648. https://doi.org/10.1016/j.eswa.2012.12.069

Olkowska E, Kudłak B, Tsakovski S et al (2014) Assessment of the water quality of Kłodnica River catchment using self-organizing maps. Sci Total Environ 476–477:477–484. https://doi.org/10.1016/j.scitotenv.2014.01.044

OmiDelta–INE (2018) Planification Delta: Préparation et organisation. Abomey-Calavi, Benin

Pacheco FS, Miranda M, Pezzi LP et al (2017) Water quality longitudinal profile of the Paraíba do Sul River, Brazil during an extreme drought event. Limnol Oceanogr 62:S131–S146. https://doi.org/10.1002/lno.10586

Park Y-S, Kwon Y-S, Hwang S-J, Park S (2014) Characterizing effects of landscape and morphometric factors on water quality of reservoirs using a self-organizing map. Environ Model Softw 55:214–221. https://doi.org/10.1016/j.envsoft.2014.01.031

Park Y-S, Chon T-S, Bae M-J et al (2018) Multivariate data analysis by means of Self-Organizing Maps. In: Recknagel F, Michener WK (eds) Ecological Informatics. Springer International Publishing, Cham, pp 251–272

Peterson BG, Carl P, Boudt K et al (2020) Package "Performance Analytics": econometric tools for performance and risk analysis. R package version 2.0.4. https://cran.r-project.org/web/packages/PerformanceAnalytics/index.html

R Core Team (2018) R: a language and environment for statistical computing. Version 3.5.2. https://cran.r-project.org/bin/windows/base/old/3.5.2/

Ren W, Zhong Y, Meligrana J et al (2003) Urbanization, land use, and water quality in Shanghai 1947–1996. Environ Int 29:649–659. https://doi.org/10.1016/S0160-4120(03)00051-5

Rice EW, Baird RB, Eaton AD, Clesceri LS (2012) Standard methods for the examination of water and wastewater. American Public Health Association, 22nd edn. American Water Works Association, Water Environment Federation, Washington

Rodier J, Legube B, Merlet N et al (2009) L'analyse de l'eau. 9ème édition entièrement mise à jour. Dunod, Paris

Rodrigues V, Estrany J, Ranzini M et al (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

Subida MD, Berihuete A, Drake P, Blasco J (2013) Multivariate methods and artificial neural networks in the assessment of the response of infaunal assemblages to sediment metal contamination and organic enrichment. Sci Total Environ 450–451:289–300. https://doi.org/10.1016/j.scitotenv.2013.02.009

Totin VSH (2010) Sensibilité des eaux souterraines du bassin sédimentaire côtier du Bénin à l'évolution du climat et aux modes d'exploitation : stratégies de gestion durable. PhD Thesis. Université d'Abomey-Calavi, Bénin

Tsai W-P, Huang S-P, Cheng S-T et al (2017) A data-mining framework for exploring the multi-relation between fish species and water quality through self-organizing map. Sci Total Environ 579:474–483. https://doi.org/10.1016/j.scitotenv.2016.11.071

Tu J (2009) Combined impact of climate and land use changes on streamflow and water quality in eastern Massachusetts, USA. J Hydrol 379:268–283. https://doi.org/10.1016/j.jhydrol.2009.10.009

UNEP (2010) Africa Water Atlas. UNEP, Nairobi, Kenya

Ustaoğlu F, Tepe Y (2019) Water quality and sediment contamination assessment of Pazarsuyu Stream, Turkey using multivariate statistical methods and pollution indicators. Int Soil Water Conserv Res 7:47–56. https://doi.org/10.1016/j.iswcr.2018.09.001

Vadde K, Wang J, Cao L et al (2018) Assessment of water quality and identification of pollution risk locations in Tiaoxi River (Taihu Watershed), China. Water (Basel) 10:183. https://doi.org/10.3390/w10020183

Vesanto J, Himberg J, Alhoniemi E, Parhankangas J (2000) SOM Toolbox for Matlab 5. Technical Report A57, Helsinki

Wang X, Zhang F (2018) Effects of land use/cover on surface water pollution based on remote sensing and 3D-EEM fluorescence data in the Jinghe Oasis. Sci Rep 8:1–13. https://doi.org/10.1038/s41598-018-31265-0

Wang J, Liu G, Liu H, Lam PKS (2017) Multivariate statistical evaluation of dissolved trace elements and a water quality assessment in the middle reaches of Huaihe River, Anhui, China. Sci Total Environ 583:421–431. https://doi.org/10.1016/j.scitotenv.2017.01.088

Wehrens R, Kruisselbrink J (2018) Flexible Self-Organizing Maps in Kohonen 3.0. J Stat Softw 87:1–18. https://doi.org/10.18637/jss.v087.i07

Wendel J, Buttenfield BP (2010) Formalizing guidelines for building meaningful Self- Organizing Maps. In: Sixth International Conference on Geographic Information Science. Zurich, Swithzerland, p 6

WFD (2005) Overall approach to the classification of ecological status and ecological potential. Bruxelles, Belgique

WHO (2011) Guidelines for drinking-water quality, 4th edn. Geneva, Switzerland

Wickham H, François R, Henry L, Müller K (2020) Package 'dplyr': A Grammar of Data Manipulation. R package Version 0.8.5. https://cran.r-project.org/src/contrib/Archive/dplyr/dplyr_0.8.5.tar.gz

Xu Z, Shen J, Qu Y et al (2022) Using simple and easy water quality parameters to predict trihalomethane occurrence in tap water. Chemosphere 286:131586. https://doi.org/10.1016/j.chemosphere.2021.131586

Zhang Z, Juying L, Mamat Z, Qingfu Y (2016) Sources identification and pollution evaluation of heavy metals in the surface sediments of Bortala River, Northwest China. Ecotoxicol Environ Saf 126:94–101. https://doi.org/10.1016/j.ecoenv.2015.12.025

Zhao Y, Xia XH, Yang ZF, Wang F (2012) Assessment of water quality in Baiyangdian Lake using multivariate statistical techniques. Procedia Environ Sci 13:1213–1226. https://doi.org/10.1016/j.proenv.2012.01.115

Zhu G, Wu X, Ge J et al (2020) Influence of mining activities on groundwater hydrochemistry and heavy metal migration using a self-organizing map (SOM). J Clean Prod 257:120664. https://doi.org/10.1016/j.jclepro.2020.120664