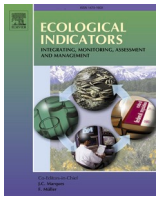
Ecological Indicators 126 (2021) 107706

Contents lists available at ScienceDirect

Ecological Indicators

journal homepage: www.elsevier.com/locate/ecolind



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

Zinsou Cosme Koudenoukpo [a,b,1], Olaniran Hamed Odountan [b,c,*,1], Prudenciène Ablawa Agboho [d], Tatenda Dalu [e,f], Bert Van Bocxlaer [g], Luc Janssens de Bistoven [h], Antoine Chikou [a], Thierry Backeljau [h,i]

[a] Laboratory of Hydrobiology and Aquaculture, Faculty of Agronomic Sciences, University of Abomey–Calavi, 01 BP 526 Cotonou, Benin
[b] Cercle d'Action pour la Protection de l'Environnement et de la Biodiversité (CAPE BIO–ONG), Cotonou, Abomey–Calavi, Benin
[c] Laboratory of Ecology and Aquatic Ecosystem Management, Department of Zoology, Faculty of Science and Technics, University of Abomey–Calavi, 01 BP 526 Cotonou, Benin
[d] Centre de Recherche Entomologique de Cotonou (CREC), 06 BP 2604 Cotonou, Benin
[e] School of Biology and Environmental Sciences, University of Mpumalanga, Nelspruit 1201, South Africa
[f] South African Institute for Aquatic Biodiversity, Grahamstown 6140, South Africa
[g] CNRS, Univ. Lille, UMR 8198 – Evo–Eco–Paleo, F–59000 Lille, France
[h] Royal Belgian Institute of Natural Sciences, Vautierstraat 29, B–1000 Brussels, Belgium
[i] Evolutionary Ecology Group, University of Antwerp, Universiteitsplein 1, B–2610 Antwerp, Belgium

ARTICLE INFO

Keywords:
Artificial neural network
Ecology
Freshwater biodiversity
Modelling
Mollusc community
Tropical river systems

ABSTRACT

The poor understanding of changes in mollusc ecology along rivers, especially in West Africa, hampers the implementation of management measures. We used a self–organizing map, indicator species analysis, linear discriminant analysis and a random forest model to distinguish mollusc assemblages, to determine the ecological preferences of individual mollusc species and to associate major physicochemical variables with mollusc assemblages and occurrences in the Sô River Basin, Benin. We identified four mollusc assemblages along an upstream–downstream gradient. Dissolved oxygen (DO), biochemical oxygen demand (BOD), salinity, calcium (Ca), total nitrogen (TN), copper (Cu), lead (Pb), nickel (Ni), cadmium (Cd) and mercury (Hg) were the major physicochemical variables responsible for structuring these mollusc assemblages. However, the physicochemical factors responsible for shaping the distribution of individual species varied per species. Upstream sites (assemblage I) showed high DO and low BOD and mineral compounds (i.e., TN, salinity, and Ca), which are primarily responsible for structuring the occurrences of bivalves (*Afropisidium pirothi*, *Etheria elliptica*, *Sphaerium hartmanni*) and the gastropod *Lanistes varicus*. Sites along the middle reach (assemblage II) were characterised by a high degree of organic pollution but low heavy metal pollution; we detected no specific mollusc indicator species. Downstream sites (assemblage III) displayed high mineral and heavy metal concentrations and a fauna without specific indicator species. Finally, downstream sites associated with brackish water (assemblage IV) displayed important levels of organic and heavy metal pollution. These sites are dominated by diverse gastropods (i.e., *Bulinus* spp., *Gabbiella africana*, *Indoplanorbis exustus*, *Pachymelania fusca*, *Radix natalensis*, *Stenophysa marmorata* and *Tympanotonos fuscatus*). Our results highlight that mollusc communities in the Sô River Basin are structured by key physicochemical variables related to the river–estuary continuum. Habitats that are progressively more downstream are confronted with increasing anthropogenic stress. Conservation and management plans should focus on downstream habitats.

* Corresponding author at: Laboratory of Ecology and Aquatic Ecosystem Management, Department of Zoology, Faculty of Science and Technics, University of Abomey–Calavi, 01 BP 526 Cotonou, Benin.
E-mail address: Odountan.hamed@gmail.com (O.H. Odountan).
[1] Contributed equally as the first authors.

https://doi.org/10.1016/j.ecolind.2021.107706
Received 14 October 2020; Received in revised form 6 March 2021; Accepted 7 April 2021
Available online 18 April 2021

1. Introduction

Mollusca are the second largest animal phylum after Arthropoda, and comprise about 75,000 extant described species of which about 6000–7000 are valid freshwater species (MolluscaBase, 2019). Freshwater molluscs are a main food source for higher trophic levels and exhibit complex and multiple ecological interactions with their environment (Seddon et al., 2011). They are also very suitable for biological monitoring (Roméo et al., 2005) as they are sensitive to water and sediment chemistry changes (Itsukushima et al., 2018). Molluscs have the capacity to capture substantial amounts of suspended matter, to control primary production and nutrient recycling, and to accumulate several contaminants, all of which underscores their role in ecosystems and their use as biomonitoring proxies (Usero et al., 2005).

Given the present and future challenges to manage river ecosystems sustainably (e.g., ecological restoration, relicensing of hydropower facilities, invasive species, climatic variability, urbanization and other land-use changes), there is a need for economic and reliable assessment strategies to understand multifactorial anthropogenic stress in freshwater ecosystems. Moreover, there is growing demand to develop robust biomonitoring tools based on organismal groups with high perceived societal value (Milošević et al., 2017). Molluscs are such a group given their aforementioned ecosystem functions and the fact that several species have direct societal and economic relevance as edible species and intermediate hosts of parasites that cause human and veterinary diseases (Seddon et al., 2011). Thus, there is a need to better understand how the physical and chemical changes caused by a variety of anthropogenic stressors affect freshwater ecosystems. Incomplete knowledge of these factors hampers the sustainable management of freshwater ecosystems and biodiversity globally, especially in developing countries.

Whereas the relationship between fauna and environment has been studied in various Nearctic and Palearctic freshwaters, our understanding of how changes in various water physical and chemical variables affect freshwater invertebrate species, notably molluscs, remains very fragmentary for the Afrotropics. Two important limiting factors are the lack of accurate taxonomic identifications (Odountan et al., 2019b) and the challenges to develop large spatial and temporal datasets (biological and physicochemical), which both result in a lack of systematic monitoring. As a result efforts often focus on the genus and/or family level rather than on the species level, so that differences and changes in faunal assemblages between stressed and pristine conditions are generally less evident (Jones, 2008; Bevilacqua et al., 2009). To enhance ecological knowledge of West African Mollusca at the species level, especially for Gastropoda (for the taxonomic aspects see Koudenoukpo et al., 2020), we focus here on developing innovative approaches to illustrate the relationships between physicochemical variables and mollusc communities of the Sô River Basin, Benin. This approach provides powerful tools for data reduction and the meaningful interpretation of ecological assessments.

Self–organizing maps (SOM) or Kohonen networks (Kohonen, 1982), involve a non–linear projection mapping based on an unsupervised pattern recognition method (Voyslavov et al., 2012). SOM is particularly useful and appropriate to investigate environment–species relationships, especially for grouping species (Tchakonté et al., 2014). To understand how inferred community groups are structured by environmental factors, it is important to predict relationships between environmental variables and the composition of biotic communities (Poff, 1997). Machine learning (ML) are predictive models that are increasingly used in the field of ecology and conservation (Humphries et al., 2018). Random forests (RF), derived from classification and regression trees (CART), are an ensemble modelling method that is probably the most successful ML algorithm used in ecology (Humphries et al., 2018) because it can handle both regression and classification through unsupervised learning (Cutler et al., 2012). Considering sensitivity, specificity and overall classification accuracy, linear discriminant analysis (LDA) is, next to RF, a good model approach for prediction (Maroco et al., 2011). In fact, LDA is sometime even preferred to RF for classification (Adandedjan et al., 2013) owing to the ease of the approach and its intuitive interpretation. Hence, using SOM for grouping, LDA for classification and RF for regression is an excellent strategy for the present study to explore the relationships between physicochemical variables and mollusc communities throughout a West African river.

Our study focuses on mollusc communities of the Sô River Basin (SRB) in Benin, which shows substantial mollusc diversity (Koudenoukpo et al., 2020), including some schistosome–transmitting species. Therefore, the SRB is an excellent study system to investigate environment–mollusc species relationships with general relevance to human health and ecosystem management. In this study, we (i) used repeated sampling at various locations throughout the SRB and a SOM to group mollusc communities along the river–estuary continuum into mollusc assemblages, (ii) used biodiversity indices and indicator species analysis (IndVal) to extract the major malacological features of these assemblages, and (iii) performed LDA and RF modelling to examine how physicochemical variables explain the distribution of species and how they shape mollusc assemblages. Overall, we investigated whether the mollusc communities change along an upstream–downstream gradient related to the various physicochemical properties of habitats along the Sô River Basin.

2. Material and methods

2.1. Study area

The Sô River (6°24′ to 6°32′ N, 2°27′ to 2°30′ E) originates from Lake Hlan and flows south for 85 km via a river-estuary continuum into Lake Nokoué near the towns of Ganvié and Vêkky, Benin (Fig. 1). The river has a mean depth, maximum depth and mean width of 3.7 m, 9 m and 68 m, respectively, with a catchment area of approximately 3000 km^2 (Hazoume, 2018). The Sô River is part of Ramsar site no. 1018 and is hydrologically connected with the Ouémé River, at least during extreme flooding in wet seasons via backwaters and canals, whereas downstream salt water intrusions from the Atlantic Ocean through Lake Nokoué may affect sites up to 5–10 km upstream of the river mouth, mostly during the dry seasons (Koudenoukpo, 2018). Previously, the Sô River system was a tributary of the Ouémé River, the longest river of Benin, from which it has since been largely separated. The main type of substrate in the SRB varies from sand upstream to mud downstream. The Sô River is an important source of water for domestic purposes, irrigation, and industrial use to local communities and furthermore serves as flood control for the Ouémé River (Hazoume, 2018; Koudenoukpo, 2018). As such, the Sô River suffers from the typical anthropogenic contamination that is impacting many West African river systems, such as petroleum leaking during transport (downstream), agricultural run-off, and domestic and industrial discharge (Calamari and Naeve, 1994; Koudenoukpo, 2018). The northern part of the SRB suffers less from these threats, is more pristine and may thus serve as reference area.

2.2. Data collection

Mollusc sampling and physicochemical surveys were conducted monthly from September 2017 to August 2018. Twelve sampling sites were chosen along the river-estuary continuum based on the geographic position of main tributaries, land use and/or human activities, as well as the hydrology of the river (Fig. 1). In total, we collected 144 samples for this study.

2.2.1. Physicochemical variables

Twenty–two physicochemical variables were measured according to quality standards and technical requirements for monitoring surface waters (Rodier et al., 2009; Rice et al., 2012), including meticulous procedures for collecting, preserving and processing samples, careful

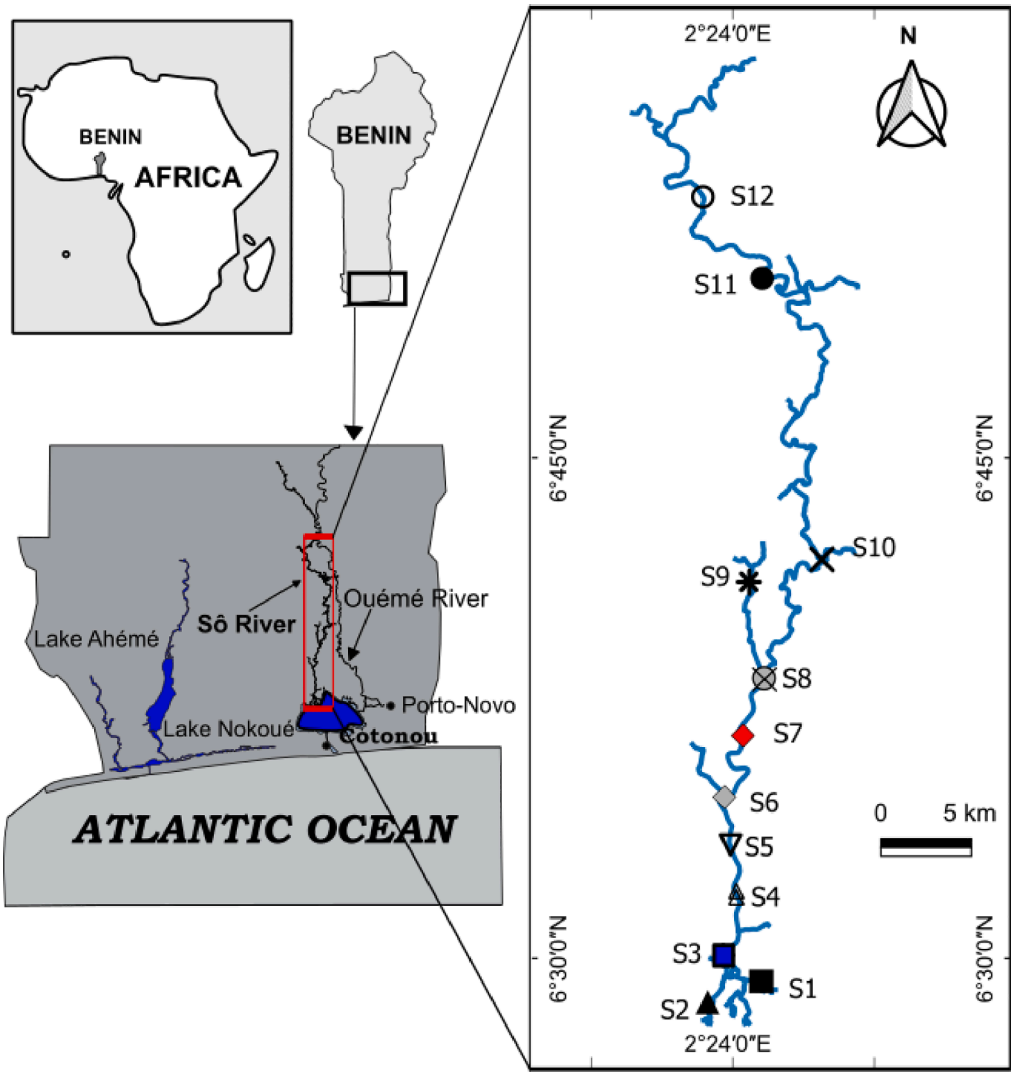


Fig. 1. Map of the Sô River Basin, Benin, indicating the locations of sites monitored in this study. Specific symbols were used for sites to present certain results.

standardization, procedural blank measurements and duplicate sampling. Duplicate 1 L water samples were collected at each site during each month in prelabelled plastic bottles, which had previously been treated with hydrochloric acid (ISO 5667-6, 2014). One duplicate was filtered with 0.45 μm mixed cellulose ester millipore filter membranes and fixed for measurement of nutrients, while the second duplicate was filtered with 0.45 μm cellulose acetate filter membranes and acidified for the determination of heavy metals (Li et al., 2018). Collected and processed water samples were kept on ice during transport to the laboratory and preserved in the refrigerator prior to analyses. Eight nutrients (Table 1) were measured with spectrophotometry (Model HACH DR 3900) as described in Rodier et al. (2009) and Rice et al. (2012). An inductively coupled plasma mass spectrometer (ICP–MS; Drawell MS–2000) was used for the detection of heavy metals (Table 1) except for mercury (Hg) concentrations which were measured with an atomic fluorescence spectrometer (AFS; Drawell AF–630A). Water depth and Secchi disk depth were measured with a graduated gauge and Secchi disk, respectively. Water temperature, salinity, pH, electric conductivity, total dissolved solids, and dissolved oxygen were measured *in situ* using a portable multi–probe water analyser (SX736 pH/mV/Conductivity/DO, Yanhe Instruments, China). The biochemical oxygen demand was measured using Winkler's method (Rodier et al., 2009).

2.2.2. *Mollusc sampling*

Three to five replicate mollusc samples were collected at each site using a long–handled kick net (250 μm mesh, 0.0625 m^2 opening) for 15 min, covering all suitable environments including littoral, benthic zones, and aquatic macrophytes (Koudenoukpo et al., 2020). Generally, the samples were collected by towing the net through vegetation and over the sediment up to 10 cm depth into sediment over an average distance of 5–10 m. All samples of the same site collected in each month were pooled and preserved in 4% formalin solution in 1.5 L plastic bottles. Molluscs were identified under a stereomicroscope (SFX–31 20×/40×, OPTIKA Microscopes, Italy) using morphological and anatomical characters following the identification keys by Nicklès (1950), Durand and Lévêque (1981), Brown and Kristensen (1993), Brown (1994) and Bony et al. (2008). The specimens were further compared with reference collections at the Royal Belgian Institute of Natural Sciences (RBINS, Brussels) and the Royal Museum for Central Africa (RMCA, Tervuren) to ensure accurate identification.

2.3. *Data analysis*

Our approach for data analysis is summarised in the workflow shown in Fig. 2. All analyses were performed in R version 3.6.0 (R Core Team, 2019).

Table 1
Summary statistics of physicochemical variables measured during our study in the Sô River. Statistics were based on values recorded per sample (n = 144). CV = Coefficient of Variation.

Variable	Acronym	Units	Range	Median	CV (%)
General variables					
Water depth	WD	[m]	1.25–10.25	3.355	49.98
Secchi disk depth	SD	[m]	0.22–1.97	0.75	50.19
Biochemical Oxygen Demand	BOD	[mg L^{-1}]	0.001–1.9	0.61	88.76
Temperature	Temp	°C	22.4–32.0	28.3	5.67
Dissolved Oxygen	DO	[mg L^{-1}]	0.34–13.5	4.15	64.94
pH	pH	–	5.84–8.22	7.22	7.97
Salinity	–	[PSU]	0.01–4.71	0.165	153.73
Electric Conductivity	EC	[µS Cm^{-1}]	49.5–12432	278	179.08
total dissolved solids	TDS	[mg L^{-1}]	22.2–9720	113	289.42
Inorganic ions		[mg L^{-1}]			
Nitrite	NO_2^-		0.001–1.849	0.118	126.78
Nitrate	NO_3^-		0.002–2.878	0.2335	124.54
Ammonium	NH_4^+		0.005–6.041	0.483	144.65
Total Nitrogen	TN		0.001–5.849	1.544	78.94
Orthophosphates	PO_4^{3-}		0.001–6.361	0.2015	247.56
Total Phosphorus	TP		0.004–5.670	0.4545	130.06
Magnesium	Mg		1.319–87.36	11.6575	91.09
Calcium	Ca		2.039–80.15	16.031	82.38
Heavy metals		[µg L^{-1}]			
Copper	Cu		0.12–1910	0.12	167.97
Cadmium	Cd		0.08–1700	0.08	157.95
Lead	Pb		0.08–1980	0.08	169.57
Nickel	Ni		0.65–980	0.65	166.45
Mercury	Hg		0.19–1760	0.19	162.54

2.3.1. Mollusc species-environment relationship analysis

Prior to in depth analysis, we examined whether the measured physicochemical variables affect mollusc community composition (Fig. 2, step 2) with multiple regressions distance matrices (MRM) on species and environmental. MRM is a flexible non-parametric and non-linear multiple regression method in which the response matrix (here, mollusc species occurrences) is regressed on an environmental matrix (physicochemical variables) and subjected to permutation to evaluate whether the observed correlation differs from the expectation under randomisation (Manly, 2006). The MRM analysis was performed with the function *multi.mantel* in the "Phytools" package (Revell, 2012). The strength of the relationship between physicochemical variables and occurrence data of species was evaluated with the R-squared statistic and *p*-value (corrected for the false discovery rate; Benjamini and Hochberg, 1995). If significant correlations between physicochemical variables and mollusc occurrence data were observed, the data can be subjected to in-depth analyses.

2.3.2. Self–organising maps (SOM)

As correlation between physicochemical variables and occurrence data were significant, we subsequently explored the association between mollusc species through the SOM. The SOM is a powerful data-mining tool used for clustering high-dimensional data (here mollusc abundances). Detailed information on the SOM approach (Kohonen, 1990, 2013) and how to construct a robust SOM can be found in the literature (e.g. Adandedjan et al., 2013; Liao et al., 2019). Here, the batch algorithm of the "Kohonen" 3.0 package (Wehrens and Kruisselbrink, 2018) was used to train a set of 2160 elements (15 mollusc taxa for each of the 144 samples) (Fig. 2, step 3). Mollusc abundance data were subjected to range normalisation to standardize weights of abundant and rare species (Park et al., 2018). Based on the relevant heuristic rule of Vesanto et al. (2000), a grid of 6 × 10 cells was considered appropriate for our data. We selected the number of clusters from a diversity of scenarios with the Davies–Bouldin index (DB), by identifying the minimal index for the sum of squares within clusters. Hierarchical cluster analysis, using the *cutree* and *hclust* functions of "dendextend" package (Galili, 2015), was performed to define clusters (based on SOM output), that were delineated by different colours.

2.3.3. Biological features of SOM clusters

To quantify differences in mollusc communities among the clusters obtained with the SOM, i.e., the mollusc assemblages, four biodiversity indices were used: species richness, taxon abundance, the Reciprocal

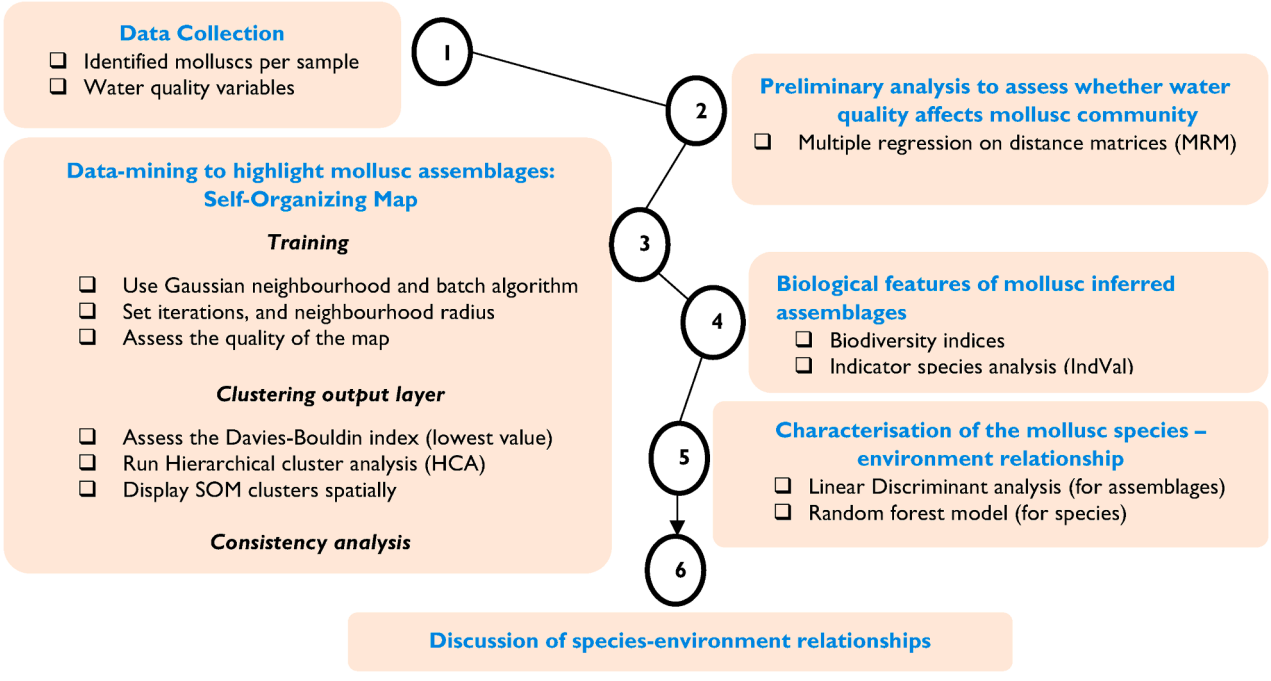


Fig. 2. Workflow followed in this study. Steps 1–5 relate to the analyses that were undertaken and step 6 to the discussion of the results.

Berger-Parker index and the Shannon-Wiener index (Fig. 2, step 4). Differences in the biodiversity indices among clusters were tested with a Kruskal–Wallis test from the "dplyr" package (Wickham et al., 2020), followed by pairwise Dunn's tests, using the *dunnTest* function in the "FSA" package, with Benjamini-Hochberg correction (Derek and Wheeler, 2020). To identify mollusc species that are characteristic for each mollusc assemblage we used IndVal (Dufrêne and Legendre, 1997). This method interrogates the affinity of taxa (here mollusc species) to groups of samples (here the mollusc assemblages) by analysing occurrence data and relative abundances of individual taxa. The algorithm provides per species and per assemblage an IndVal statistic, which ranges between 0 (not an indicator) and 1 (good indicator), with an associated probability. The indicator analysis was performed with the "indicspecies" package version 1.7.9 (De Caceres and Jansen, 2016) using 1000 permutations and $\alpha = 0.05$ significance level.

2.3.4. Mollusc species and water quality variables relationships

To determine which of the 22 physicochemical variables explain the differences in mollusc composition among mollusc assemblages, we performed LDA (Legendre and Legendre, 2012) as implemented in the "ade4" package version 1.7–15 (Dray et al., 2020). Prior to LDA, we tested for multivariate homogeneity of within–group variances (Borcard et al., 2018) and the heavy metals required square root transformation. Using standardized factorial coefficients, we identified the ten physicochemical variables that explain most of the differences among mollusc assemblages. Cross-validation was used to evaluate the predictive performance of the LDA model. The overall quality of the LDA model and the accuracy of the classification prediction for the samples of each cluster were evaluated using Cohen's Kappa Statistic with the "caret" package (Kuhn et al., 2020). Significant differences in the ten retained variables with respect to SOM clusters were tested using a Kruskal–Wallis test followed by pairwise Dunn's tests with Benjamini – Hochberg correction.

Afterwards, an RF model was used to predict how mollusc species abundances varied with physicochemical variables, and to evaluate the contribution of each environmental variable to the distribution of mollusc species. RF is a non-parametric method commonly used for the prediction/assessment of the relationship between some potential predictor variables (here physicochemical variables) and a response variable (here the abundance of mollusc species) (Breiman, 2001). We used the "randomForest" package version 4.6–14 (Liaw and Wiener, 2002) and maintained default parameter values except for the size of variables which were 3 and 7 for classification and regression, respectively (i.e., around the default values). To measure the importance of physicochemical variables, we used mean decrease in accuracy (% IncMSE) for regression. The parameter values were rescaled between 0 and 1 (Bae and Park, 2020). Partial dependence plots were constructed to visualize the effect of physicochemical variables to the fitted regression function from the RF.

3. Results

3.1. Physicochemical variables

DO and BOD were generally low across sites and over the 12 sampling months. Together with temperature, pH, WD, SD, these variables showed less variation among sites and months compared to salinity, EC, TDS, inorganic ions (i.e., NO_2^-, NO_3^-, NH_4^+, PO_4^{3-}, and TP) and heavy metals (i.e., Cu, Cd, Pb, Ni, and Hg) (Table 1). Temperature and pH showed less variation compared to other physicochemical factors.

3.2. Multiple regressions distance matrices analysis

The MRM analysis indicated that the correlation between mollusc species abundance and individual physicochemical variables was generally weak ($0.00 < R^2 < 0.25$) (Table 2) but significant for 14 of the 15 species, and the strongest correlations were usually with Cd, BOD, DO and EC. Therefore, the data can be subjected to in-depth analyses.

3.3. SOM application and sampling distribution

The SOM allowed us to recognize four mollusc assemblages based on the minimum value of the DB index (Fig. 3A). Fig. 3B shows how the 144 samples are distributed among the 60 nodes (=grid cells), with the boundaries between assemblages in bold. Assemblages I, II, III and IV contained 30 samples in 10 cells, 66 samples in 29 cells, 27 samples in 12 cells and 21 samples in 9 cells, respectively. The assignment of samples and sampling stations to mollusc assemblages follows a geographic pattern. Assemblage I contained only upstream sites (S10–S12), assemblage II grouped mainly sites situated in the middle reaches (sites S5–S9), and downstream sites (S1–S4) were distributed among assemblages III and IV. Whereas all samples of S4 were attributed to assemblage III and most of S2 to assemblage IV, the assignment of S2 and S3 to assemblages III or IV varied seasonally, as expected for sites that are seasonally influenced by salt water intrusions along a river – estuary continuum.

The component planes in Fig. 4 indicate how individual mollusc species are distributed over the grid cells and assemblages. The majority of species was confined to a specific assemblage, but *Pachymelania fusca*, *Pachymelania byronensis* and *Melanoides tuberculata* are generalists that occurred across all assemblages.

3.4. Mollusc species assemblages and indicator taxa

The four mollusc assemblages reconstructed with the SOM analysis show significant differences in mollusc richness (Kruskal–Wallis $H = 106.21$, $df = 3$, $p < 0.001$), abundance (Kruskal–Wallis $H = 89.25$, $df = 3$, $p < 0.001$), reciprocal Berger–Parker index (Kruskal–Wallis $H = 58.01$, $df = 3$, $p < 0.001$), and Shannon–Wiener diversity (Kruskal–Wallis $H = 94.86$, $df = 3$, $p < 0.01$) (Fig. 5). Generally, higher values for biodiversity indices were found downstream (assemblages III and IV), whereas values were intermediate upstream (assemblage I) and lowest along the middle reaches (assemblage II). Results of pairwise comparisons among all assemblages per biodiversity indices are indicated in Fig. 5.

The IndVal analysis suggested that 13 out of the 15 recovered species have an indicator value, but these indicator species only characterise assemblages I and IV (Table 3). Four species are indicators for

Table 2
Correlations between the Euclidean–based distance of physicochemical variables and the Bray–Curtis based distance abundance of mollusc species in Sô River, Benin as examined with MRN. Physicochemical variables with minimal and maximal values are presented in parenthesis. Significant correlations are indicated in bold.

Species	Minimum	Median	Maximum
Bivalvia			
Sphaerium hartmanni (Jickeli, 1874)	0.00 (SD)	0.05	**0.21** (BOD)
sAfropisidium pirothi (Jickeli, 1881)	0.00 (SD)	0.05	**0.21** (BOD)
Etheria elliptica (Lamarck, 1807)	0.00 (SD)	0.06	**0.22** (BOD)
Gastropoda			
Lanistes varicus varicus (O. F. Müller, 1774)	0 (BOD)	0.00	**0.03** (EC)
Gabbiella Africana (Frauenfeld, 1862)	0.00 (T)	0.01	**0.16** (Cd)
Melanoides tuberculata (O. F. Müller, 1774)	0.00 (BOD)	0.00	**0.01** (TDS)
Pachymelania fusca (Gmelin, 1791)	0.00 (TDS)	0.00	0.01 (WD)
Pachymelania byronensis (W. Wood, 1828)	0.00 (Hg)	0.00	**0.07** (EC)
Tympanotonos fuscatus (Linnaeus, 1758)	0.00 (T)	0.00	**0.24** (Cd)
Radix natalensis (Krauss, 1848)	0.00 (T)	0.01	**0.19** (DO)
Indoplanorbis exustus (Deshayes, 1834)	0.00 (NH_4^+)	0.01	**0.17** (Cd)
Bulinus truncates (Audouin, 1827)	0.00 (NH_4^+)	0.01	**0.16** (Cd)
Bulinus globosus (Morelet, 1866)	0.00 (NH_4^+)	0.01	**0.16** (Cd)
Bulinus forskalii (Ehrenberg, 1831)	0.00 (Ca^{2+})	0.01	**0.18** (Cd)
Stenophysa marmorata (Guilding, 1828)	0.000 (T)	0.01	**0.20** (DO)

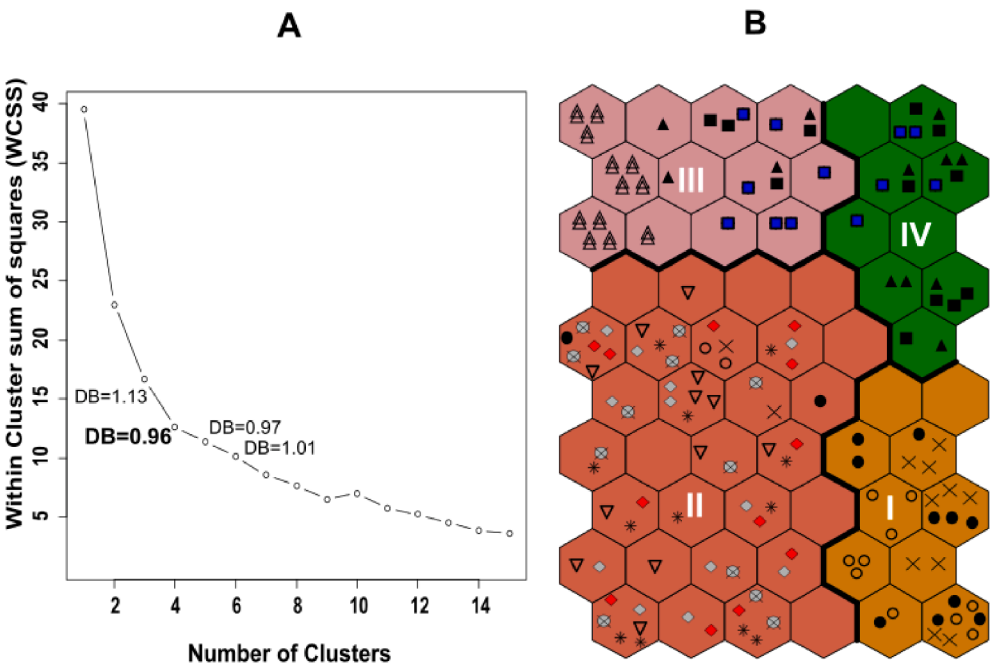


Fig. 3. Grouping of mollusc samples (n = 144) and sampling stations (n = 12) into the mollusc assemblages that are recognized on the basis of SOM along the river-estuary continuum of the Sô River, Benin: (A) Plot of within-cluster sum of squares over the total number of clusters considered, with minimal values of the Davies–Boulding (DB) index indicating preferable solutions (i.e., four clusters in this case); (B) distribution of sampling sites along the 60 grid cells (neurons) of our SOM, with in bold the delineation of the four mollusc assemblages (I, II, III IV). Symbols of sampling sites follow the representation in Fig. 1: S1: black-filled squares; S2: black-filled triangles; S3: blue-filled squares; S4: double triangles; S5: empty inverted triangles; S6: grey diamonds; S7: red diamonds; S8: grey-filled and crossed circles; S9: stars; S10: crosses; S11: solid circles; S12: empty circles. Empty neurons show that none of our samples have the pattern corresponding to that neuron. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

assemblage I, i.e., three bivalves and a gastropod, whereas assemblage IV has nine gastropod indicators (Table 3).

3.5. LDA and mollusc abundance–physicochemical variables relationships

How water quality variables differ among mollusc assemblages inferred with SOM analysis was examined with LDA (Fig. 6). The accuracy of the LDA was high (0.91) with a Kappa coefficient of 0.87 and a predictive reliability for assemblages I, II, III and IV of 1, 0.86, 0.93 and 0.90, respectively. The factorial axes 1, 2 and 3 explained 50.83%, 26.51%, and 22.66%, respectively, of the total variance of mollusc community data. Assemblages I and II were positively associated with axis 1, but separated along axis 2, whereas assemblages III and IV were negatively associated with axis 1, but separated along axis 2 (Fig. 6A). Many physicochemical variables differ significantly among the four molluscs assemblages. Mainly assemblage I, but also assemblage II, were correlated with high values of DO and SD, whereas assemblages III and IV were positively correlated with the other 20 water quality variables (Fig. 6B).

Detailed information of the ten water quality variables that contribute most to the discrimination between mollusc assemblages is provided in Table 4. Assemblage I is characterised by high DO, low BOD and mineral compounds (i.e., TN, salinity, and Ca^{2+}) and low concentrations of heavy metals (Cu, Pb, Cd, Hg and Ni). Assemblage II also groups samples with high DO and low concentrations of heavy metals, but higher BOD and concentrations of TN, salinity and Ca^{2+} compared to Assemblage I. Assemblage III and IV group samples with low DO and generally high mineral and heavy metal concentrations, and both differ mainly in salinity (high for assemblage IV, intermediate for III).

The RF model predictions of mollusc species abundances in relation to physicochemical variables are presented in Figs. 7 and 8. Overall, the RF models have high predictive power for the distribution of mollusc species (R^2: 0.56–0.93) except for *Lanistes varicus, Gabbiella africana, Melanoides tuberculata, Pachymelania fusca* and *Pachymelania byronensis* (R^2: 0.05–0.35) (Fig. 7). Overall, heavy metals, BOD and salinity were the main factors influencing the distribution of mollusc species, but the relative contribution of these variables to predictions differed among species. For example, BOD was by far the most predictive factor (rescaled IncMSE = 1) as to the occurrence of the bivalve species (i.e., *Sphaerium hartmanni*, *Afropisidium pirothi* and *Etheria elliptica*). Salinity (rescaled IncMSE = 1), and heavy metals (especially Cd and Hg) were most predictive for *Bulinus* spp. and *Indoplanorbis exustus* abundances. Heavy metals (rescaled IncMSE: 0.60–1) were by far the most important factor in structuring the occurrences of *Tympanotonos fuscatus, Radix natalensis* and the invasive *Stenophysa marmorata*. The pH (rescaled IncMSE = 1), mineral compounds EC (rescaled IncMSE = 1) and NO_2^- (rescaled IncMSE = 1) were influential in determining the occurrences of *Lanistes varicus*, *Melanoides tuberculata* and *Pachymelania byronensis*, respectively.

The partial dependence plot of two major physicochemical variables that influence all assemblages (i.e., BOD and salinity) as identified with LDA and RF are shown in Fig. 8A and B, respectively. Generally, bivalves showed an abundance increase with decreasing BOD and salinity, whereas *Radix natalensis* and *Stenophysa marmorata* showed an opposite trend. *Bulinus* spp., *Tympanotonos fuscatus*, and *Indoplanorbis exustus* abundances were higher at lower BOD ($\leq$1 mgL^{-1}) and higher salinity values ($\geq$2 PSU). The high abundance of *Melanoides tuberculata* and *Lanistes varicus* was related to low salinity values ($\leq$0.5 PSU), whereas the influence of BOD on their abundance remains unclear.

4. Discussion

Using a SOM, we revealed four distinct mollusc assemblages along the river–estuary continuum of the Sô River in Benin, which are characterised by different biodiversity indices and indicators species. LDA furthermore demonstrated that these assemblages differ markedly in physicochemical variables along an upstream–downstream gradient, while RF showed that the principal physicochemical variables influencing the distribution and abundance of each mollusc species vary among species. As such, these results illustrate the multifactorial nature of water quality effects on freshwater mollusc communities and individual mollusc species.

This study indicates that non–supervised machine learning SOM in conjunction with LDA and RF models are effective tools to characterise mollusc assemblages in relation with various physicochemical factors along rivers. This approach can most likely be applied to characterise a wider variety of freshwater systems. More generally, this study extends the use of these techniques to analyse multivariate biological data across various taxa and habitats (Lek-Ang et al., 2007; Adandedjan et al., 2013; Tsai et al., 2017).

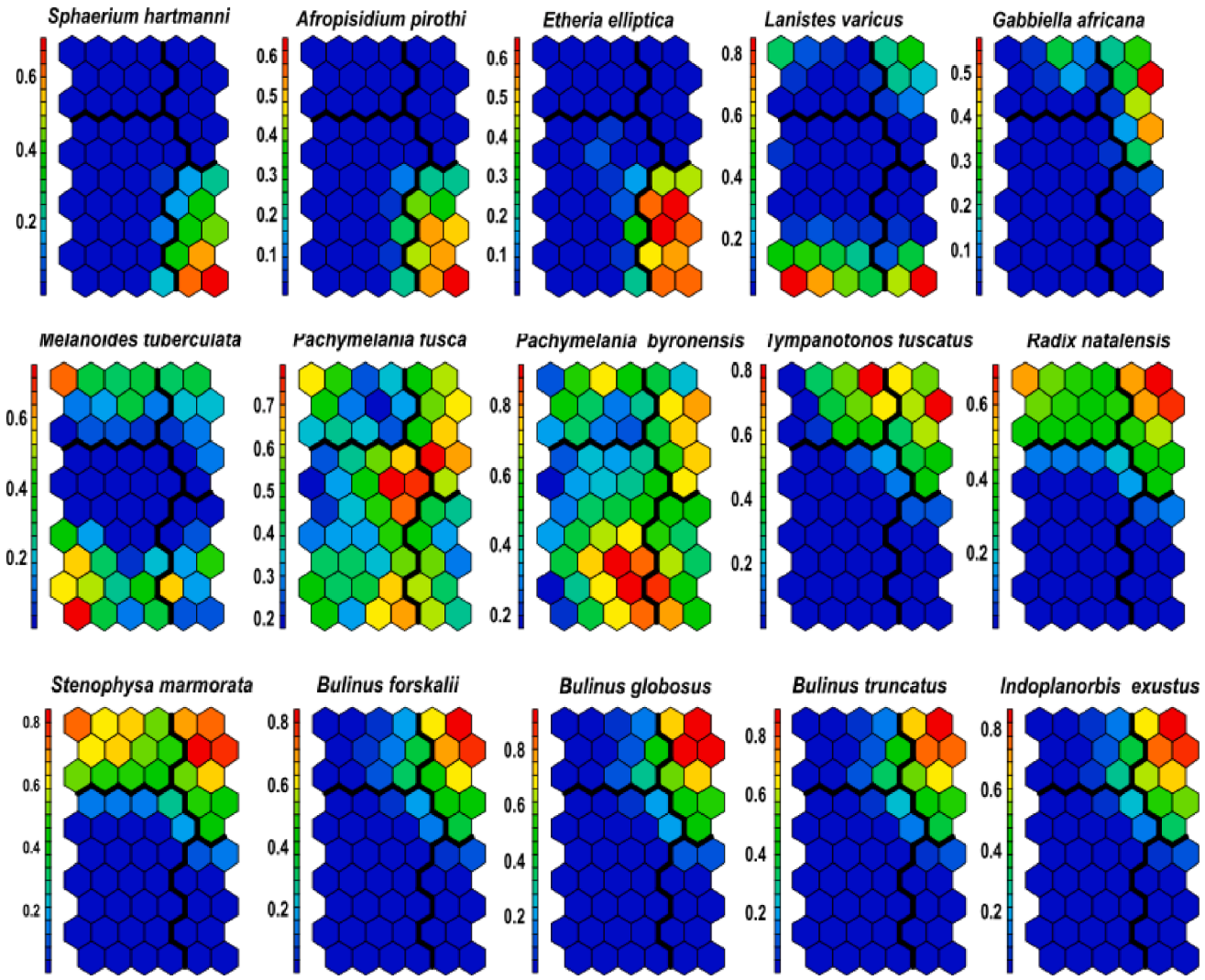


Fig. 4. Component planes for the mollusc species analysed with our self–organising map (SOM) for the Sô River, Benin. Neurons are color-coded according to the relative abundance of the species in that node. Warm colours (i.e., red) indicate high abundance, while cold colours (i.e., blue) represent low values. Slightly different scales were used depending on the species to better highlight variation in the abundance of each species per neuron. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

4.1. Molluscs patterns along the river–estuary continuum of the Sô River

Our data covered both spatial and temporal variation, but the SOM results indicated that spatial characteristics mainly influenced how mollusc assemblages are structured in the Sô River, particularly with respect to the upstream–downstream gradient. However, the assignment of certain sampling sites to assemblages indicates also seasonal variation (assemblages III and IV). Indeed, for most of the year a strong environmental gradient exists between S4 and downstream sites (S1–S3). This gradient is altered during periods of flooding (freshwater) or strong saltwater intrusion, so that their characteristics temporarily become similar. This finding is similar to observations reported in streams of Cameroon (Tchakonté et al., 2014) and Ivory Coast (Camara et al., 2012), where gastropod patterns are both affected by spatial heterogeneity of sampling sites and seasonal variation. Moreover, it was reported that spatial variation contributes to mollusc assemblages than seasonal variation (Tchakonté et al., 2014) because of the molluscs specific life history traits and ecology (e.g., limited mobility; but see Section 4.2 below where we provide information on seasonal variation).

Freshwater mollusc communities changed along the upstream–downstream gradient with respect to species composition, abundance and trophic structure, similar to previous findings of Tchakonté et al. (2014). The dominance of *Etheria elliptica*, *Afropisidium pirothi*, *Sphaerium hartmanni* and *Lanistes varicus* in assemblage I suggests that this assemblage consists mainly of collector–filterer organisms. Low mollusc richness, abundance and diversity were observed in assemblage II (middle reach). As shown by IndVal, this assemblage and assemblage III did not contain any indicator species and could be considered as transitional. This may suggest that there have been few changes in the species composition of mollusc communities along the areas covered by these assemblages. Assemblage IV (downstream) exhibited the highest biodiversity, and is occupied mainly by gastropods (polyphagia and scrapers), indicating important trophic changes in mollusc communities along the river-estuary continuum. Whereas these trophic changes may relate to changes in habitat availability along this continuum, we suspect that physicochemical variables (water quality) are important in shaping these communities. Therefore, we discuss their influence on the mollusc community composition of the four assemblages here below.

4.2. Relationship between mollusc species and physicochemical variables

Based on the LDA and RF model, the ten physicochemical variables that are most influential in structuring the upstream–downstream assemblages and mollusc species occurrence were DO, BOD, salinity, Ca^{2+},

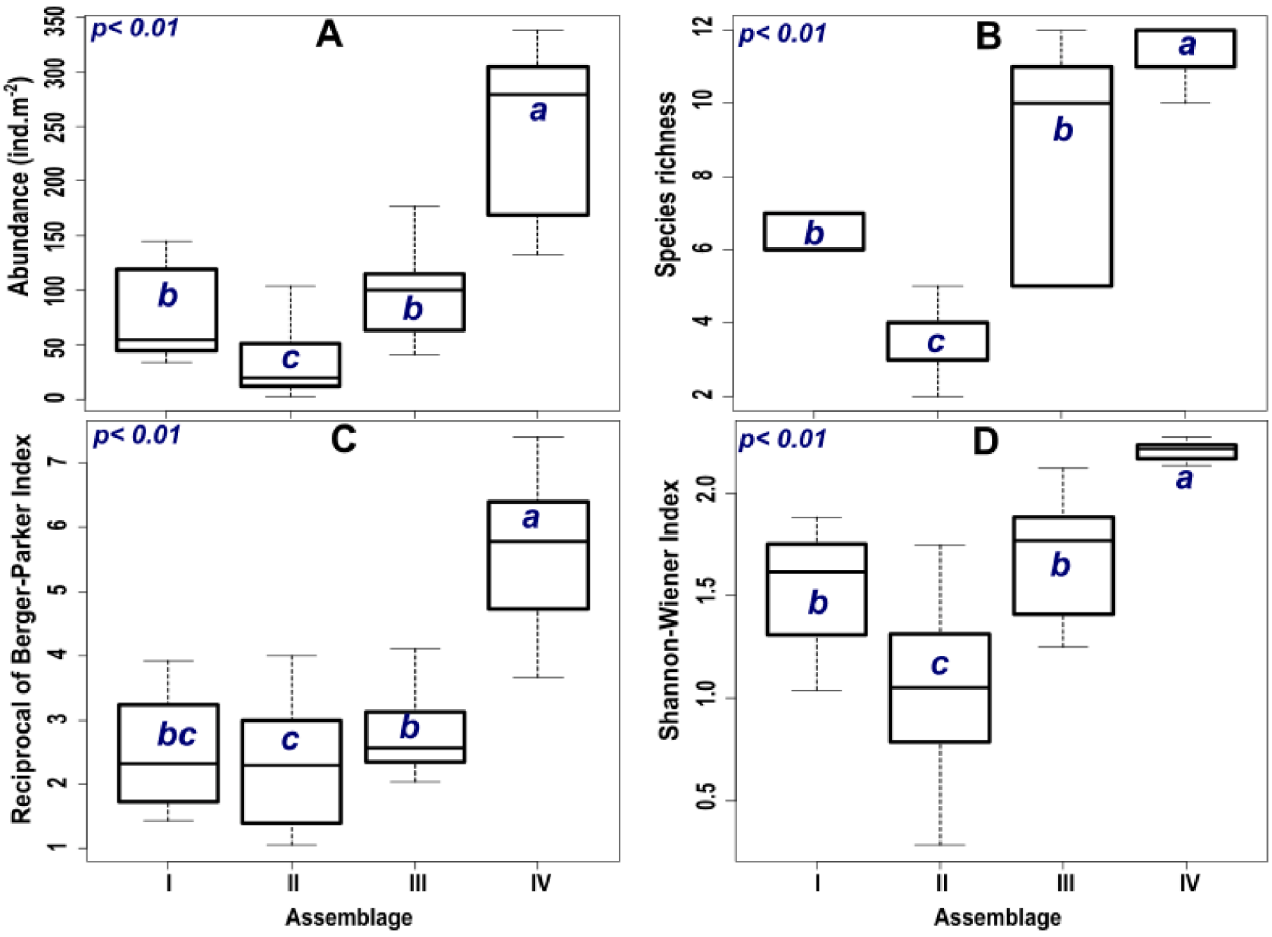


Fig. 5. Box plots of (A) species richness, (B) abundance, (C) Reciprocal of Berger-Parker and (D) Shannon diversity of mollusc species assemblages as reconstructed with SOM of the Sô River, Benin. The top, middle, and bottom of each box represent the 75th, 50th and 25th percentiles, respectively. The whiskers indicate the 90th and 10th percentiles, respectively. Letters within boxes summarize the results of pairwise statistical tests: assemblages with different letters differ significantly ($p < 0.01$) for the considered biodiversity index.

Table 3
Indicator species of each mollusc assemblage of the Sô River, Benin based on their indicator value and significant *p*–values.

SOM–Assemblages	IndVal	*p*		IndVal	*p*
Assemblage I					
Etheria elliptica	0.98	<0.001	*Afropisidium pirothi*	0.99	<0.001
Sphaerium hartmanni	0.99	<0.001	*Lanistes varicus*	0.62	0.032
Assemblage II					
None					
Assemblage III					
None					
Assemblage IV					
Indoplanorbis exustus	0.95	<0.001	*Tympanotonos fuscatus*	0.79	<0.001
Bulinus globosus	0.94	<0.001	*Radix natalensis*	0.75	<0.001
Bulinus truncatus	0.94	<0.001	*Stenophysa marmorata*	0.73	<0.001
Bulinus forskalii	0.94	<0.001	*Pachymelania fusca*	0.59	<0.001
Gabbiella africana	0.87	<0.001			

TN, Cu, Pb, Ni, Cd and Hg. However, the RF model showed that the main physicochemical factors influencing species occurrences varied among species. For example, *Bulinus* spp. and *Stenophysa marmorata*, which characterise assemblage IV, differ markedly in their sensitivity to BOD. This finding was linked to the fact that BOD was not the predominant predicting factor for assemblage IV. Instead, salinity as a major predictor for assemblage IV showed a similar effect on these two species. The combination of LDA and RF highlights that mollusc species of the same assemblage may differ markedly in their sensitivity to different physicochemical factors and as such how mollusc communities may be shaped by the multivariate interaction of various environmental variables. Phrased differently, environmental conditions act jointly as filters through which groups of species at the regional scale must pass to be potentially present at the locale scale (Bae et al., 2011). The cases in point are *Etheria elliptica* and *Lanistes varicus* from assemblage I, which differ markedly in their sensitivity to BOD, although BOD is one of the factors that characterizes assemblage I (Table 4, Fig. 8). Hence, their joint assignment to assemblage I can only be explained by considering interactions among several physicochemical variables. Thirteen out of the 15 species are indicators for two (I and IV) of the four assemblages. In contrast, assemblages II and III have a transitional nature, no indicator species and are not discussed further.

In assemblage I the presence and abundance of *Etheria elliptica*, *Afropisidium pirothi*, *Sphaerium hartmanni* and *Lanistes varicus* is correlated with high DO, and low BOD, TN, salinity, and Ca^{2+} concentrations, as well as low levels of organic and heavy metal pollution. Mainly the three bivalve species occur typically in oxygenated fairly fast running freshwater streams, with stony and turbulent habitats (Yonge, 1962; Seddon et al., 2018) as usually found in headwaters (Dobson and Frid, 2009), similar to those of the Sô River. The Sô River headwaters flow through the Djigbé forest which provides suitable habitat for macrophytes such as *Nimphea lotus* and *Lemna paucicostata*, which increase

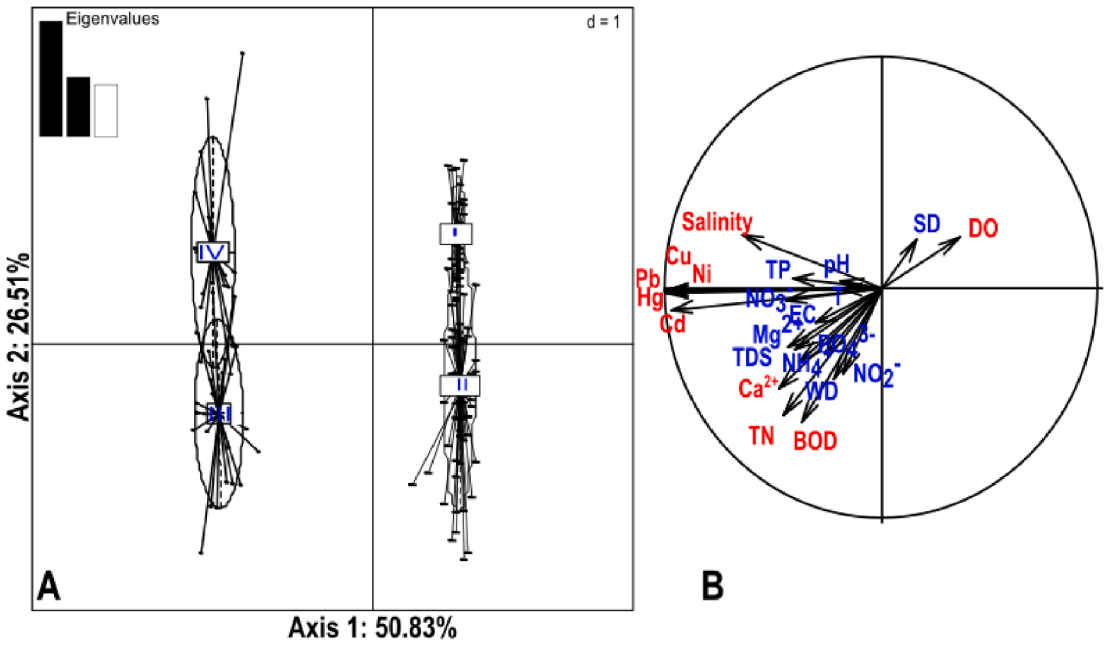


Fig. 6. The LDA discriminating the four molluscs assemblages (SOM clusters: I, II, III, IV) in the Sô River, Benin along axes 1 and 2 (A), and correlations of the water quality variables with the corresponding axes (B). The ten most influential physicochemical variables are indicated in red. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 4
The range (+median) for the ten physicochemical variables that contribute most to the discrimination among mollusc assemblages. Legend and units of variables are indicated in Table 2.

Variable	I	II	III	IV
DO	6.41–13.50 (9.43)^{a}	0.34–13.5 (4.76)^{b}	0.45–6.91 (2.08)^{c}	1.23–4.40 (2.44)^{c}
BOD	0.00–0.06 (0.02)^{b}	0.01–1.90 (0.69)^{a}	0.15–1.83 (0.83)^{a}	0.25–1.80 (0.75)^{a}
TN	0.00–1.75 (0.15)^{c}	0.02–5.85 (1.80)^{b}	0.22–5.15 (2.90)^{a}	1.00–4.96 (2.10)^{ab}
Ca	2.04–12.75 (6.03)^{c}	4.00–80.16 (16.03)^{b}	8.02–80.16 (24.05)^{a}	10.02–79.20 (28.06)^{a}
Salinity	0.01–1.35 (0.06)^{d}	0.01–1.54 (0.12)^{c}	0.07–3.68 (0.82)^{b}	0.18–4.71 (2.29)^{a}
Pb	0.00–0.00 (0.00)^{b}	0.00–0.00 (0.00)^{b}	280–1950 (670)^{a}	230–1980 (790)^{a}
Cu	0.00–0.00 (0.00)^{b}	0.00–0.00 (0.00)^{b}	190–1910 (690)^{a}	130–1670 (670)^{a}
Hg	0.00–0.00 (0.00)^{b}	0.00–6.00 (0.00)^{b}	360–1760 (760)^{a}	120–1490 (710)^{a}
Cd	0.00–0.00 (0.00)^{b}	0.00–1.00 (0.00)^{b}	180–1700 (810)^{a}	360–1290 (790)^{a}
Ni	0.00–0.00 (0.00)^{b}	0.00–78 (0.00)^{b}	60–980 (550)^{a}	110–780 (310)^{a}

For each variable superscripts, i.e., ^{a,b,c,d} summarize the significance of pairwise comparisons among assemblages. Assemblages with the same letter for a given variable do not differ significantly for that variable at $p > 0.05$.

habitat complexity and provide more food and refuge from predator for juveniles (Zaabar et al., 2018). Additionally, the high canopy linked to the high density of riparian vegetation (Konan et al., 2006; Hazoume, 2018), buffer fluctuations in water temperature and DO, making the habitat suitable for the bivalve molluscs. These bivalves are also highly sensitive to pollution (Adandedjan et al., 2013), which is limited in headwaters but increases markedly further downstream. Besides bivalves, *Lanistes varicus* was mostly found in sites belonging to assemblage I, resulting in low gastropod biodiversity, beyond low abundance, in assemblage I.

In summary, the bivalve species we recovered can be used as indicators for unpolluted river environments. Although *Lanistes varicus* occurred together with these bivalves, the bivalve species abundances are more structured by salinity, EC and pH rather than DO or BOD. Assemblage IV (downstream), for which we found nine indicator species, displayed high salinity, BOD and heavy metals, and low DO gradients, indicating highly eutrophic conditions. These nine species are eurytopic and typically inhabit less oxygenated fresh to brackish aquatic habitats (Ibikounlé et al., 2009; Koudenoukpo et al., 2020) typical of the downstream reach of rivers, which is also the case in Sô River system. Indeed, the sites of assemblage IV are first characterised by important seasonal variation in salinity depending on the level of freshwater discharge and salt water intrusion from the brackish Lake Nokoué (Odountan et al., 2019a). Additionally, these sites are characterised by a variety of anthropogenic disturbances, such as contamination caused by gasoline spills, illegal sand extraction, garbage littering, household discharges and industrial discharge. The fact that nine out of 15 mollusc species were associated with this assemblage implies that on a general scale salinity is the major physicochemical variable affecting the mollusc community along the downstream reach of the river system. The mollusc species whose distribution is most affected by salinity are *Bulinus* spp., *Tympanotonos fuscatus, Indoplanorbis exustus, Stenophysa marmorata* and *Radix natalensis* (Figs. 7 and 8).

Somewhat surprisingly, our study suggests the highest α–diversity in assemblage IV, which is also most affected by pollution. However, this high diversity results largely from seasonal variation in salinity levels, allowing marine, brackish and fresh water species to occupy the same area in different moments of the year rather than simultaneously (Bony, 2007; Odountan et al., 2019a). The presence of heterobranch gastropods in this assemblage is related to their capacity to obtain atmospheric oxygen via their mantle cavity, making them more tolerant to organic pollution (Tchakonté et al., 2014) or even high heavy metal pollution (Bae and Park, 2020). Increasing nutrient concentrations may favour primary productivity, thus increasing the amount of food resources (Bae and Park, 2020), but also temporal oxygen stress. Thereby it may promote gastropods resistant to oxygen stress. However, with increasing anthropogenic stress in the downstream reaches of Sô River system, pollution–sensitive taxa disappear, whereas pollution–tolerant and/or saprophytic gastropod species persist (Tachet et al., 2010). These results

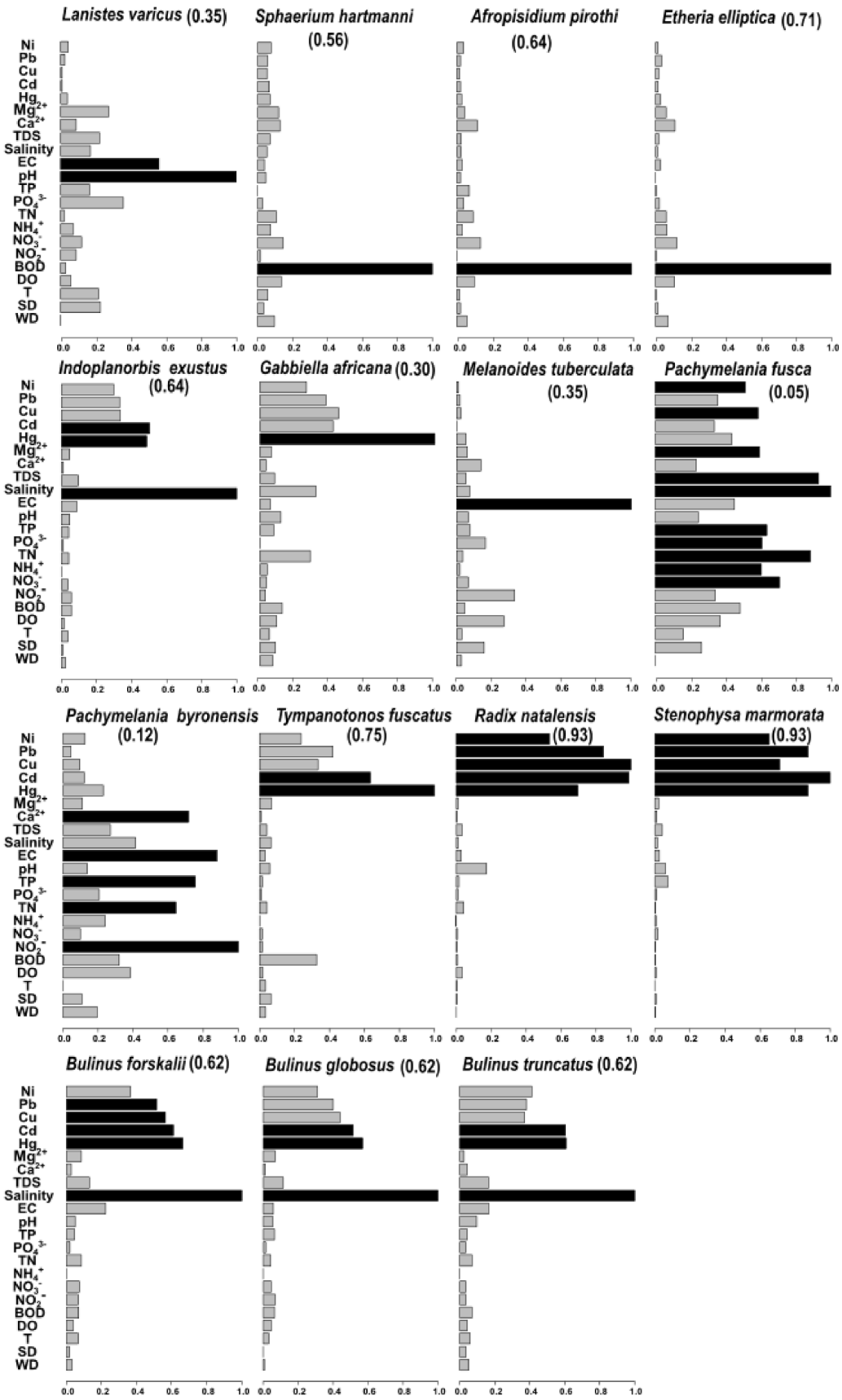


Fig. 7. Relative importance of physicochemical variables for predicting the distribution of mollusc species in the random forest model. Values in brackets represent the R-squared of the model for each species. Black bars indicate relative importance values ≥0.50, whereas grey bars < 0.50.

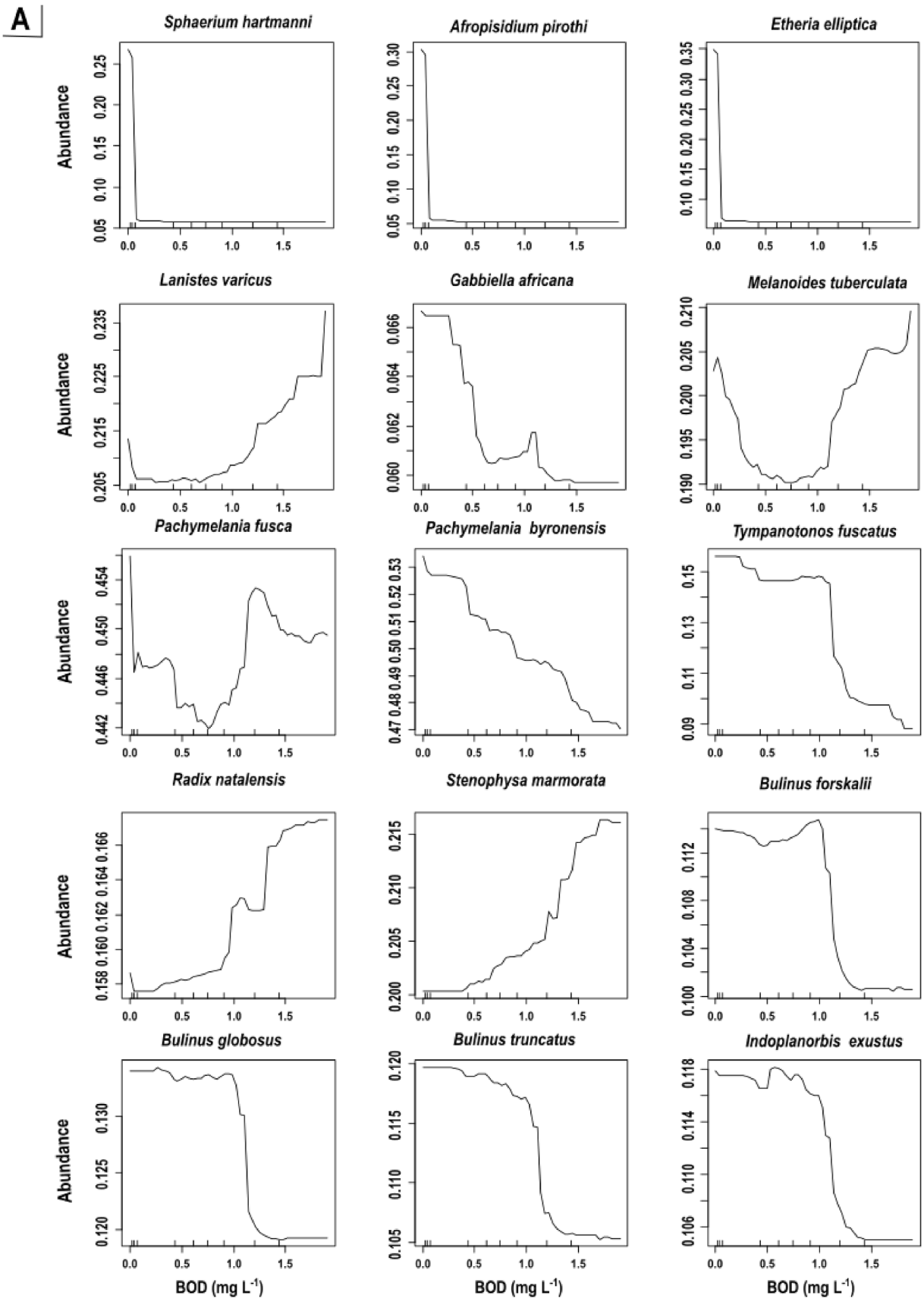


Fig. 8. Partial dependence plot of the mollusc species based on Random Forest prediction of BOD (A) and salinity (B). Units in brackets.

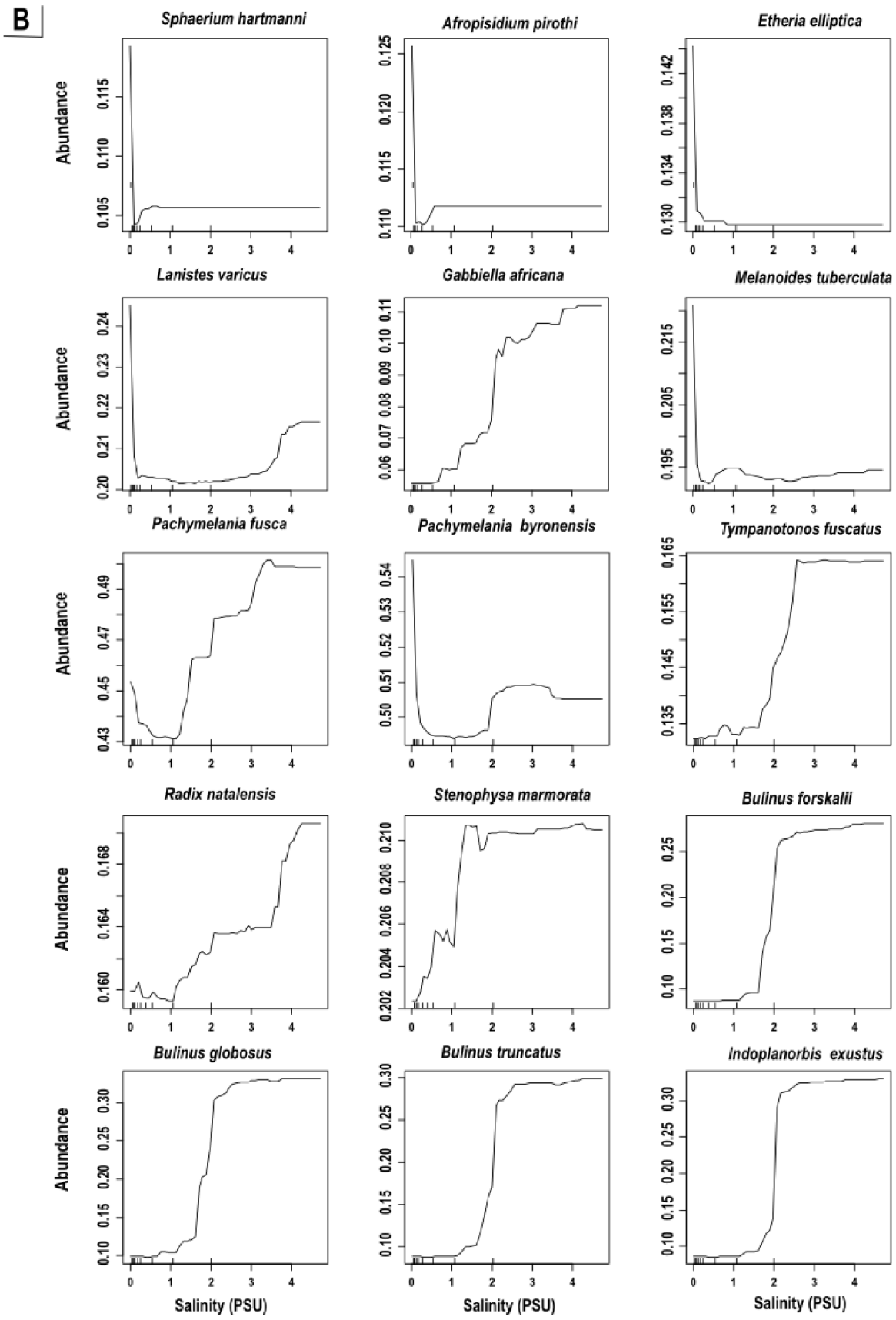


Fig. 8. (*continued*).

suggest that bivalves preferred relatively pristine ecological conditions, which occur mainly along the upstream reaches of the SRB, whereas stress-resistant gastropods dominated downstream despite the variety of anthropogenic impacts. Note that the increased eutrophication downstream has augmented the abundance of stress-resistant *Bulinus* spp., and *Radix natalensis* (Ibikounlé et al., 2014), which are intermediate hosts for trematodes that cause diseases in humans and livestock. As such, eutrophication in assemblage IV is susceptible of increasing the abundance of schistosome-transmitting snails, which may in turn increase the prevalence of schistosomiasis, as has been observed in local communities elsewhere in Africa (Van Bocxlaer et al., 2014).

5. Conclusion

The combination of SOM, LDA and RF modelling showed complex interrelations between physicochemical variables and mollusc species abundance and diversity. This combination of different analytical methods was useful in simplifying the complex ecological information, and notably to identify key variables for water resources management and ecosystem restoration in Afrotropical river systems. The distribution and composition of freshwater molluscs (gastropods and bivalves) in the Sô River reflected the physicochemical features, with good ecological conditions being observed in upstream sites, which were dominated by bivalves. Downstream environments showed a progressive increase of organic and heavy metal pollution. Moreover, high gastropod biodiversity was observed downstream because the seasonal variation in salinity allowed the same site near-estuary sites to be inhabited by a variety of taxa throughout the year. Nevertheless, all of these taxa are stress-resistant, and some transmit parasitic diseases to human and livestock indicating that further ecological deterioration may have various consequences to human health. Based on these findings, we recommend that conservation and management plans focus primarily on improving ecosystem health in downstream habitats along the Sô River. Further ecological degradation may have a strong impact on ecosystem functioning, the sustainability of the associated freshwater resources and human and veterinary health in riparian communities.

CRediT authorship contribution statement

Zinsou Cosme Koudenoukpo: Conceptualization, Methodology, Investigation, Funding acquisition, Resources, Data curation, Writing - review & editing. **Olaniran Hamed Odountan:** Conceptualization, Methodology, Data curation, Formal analysis, Software, Validation, Visualization, Writing - original draft, Writing - review & editing. **Prudencième Ablawa Agboho:** Conceptualization, Methodology, Investigation, Writing - review & editing. **Tatenda Dalu:** Methodology, Validation, Writing - review & editing. **Bert Van Bocxlaer:** Methodology, Data curation, Validation, Writing - review & editing. **Luc Janssens de Bistoven:** Methodology, Writing - review & editing. **Antoine Chikou:** Conceptualization, Methodology, Funding acquisition, Supervision, Writing - review & editing. **Thierry Backeljau:** Conceptualization, Methodology, Resources, Supervision, Writing - review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This study was financially supported by the ministry of higher education and scientific research of Benin. ZCK and OHO thank the Belgian National Focal Point to the Global Taxonomy Initiative of the CEBioS programme, at the Royal Belgian Institute of Natural Sciences, and financed by the Belgian cooperation for development (DGD). BVB thanks the French Agence Nationale de la Recherche (ANR-17-CE02-0015).

References

Adandedjan, D., Ahouansou Montcho, S., Chikou, A., Laleye, P., Gourene, G., 2013. Caractérisation des peuplements de macroinvertébrés benthiques à l'aide de la carte auto-organisatrice (SOM). C. R. Biol. 336 (5-6), 244–248. https://doi.org/10.1016/j.crvi.2013.04.009.

Bae, M.-J., Kwon, Y., Hwang, S.-J., Chon, T.-S., Yang, H.-J., Kwak, I.-S., Park, J.-H., Ham, S.-A., Park, Y.-S., 2011. Relationships between three major stream assemblages and their environmental factors in multiple spatial scales. Ann. Limnol. 47, S91–S105. https://doi.org/10.1051/limn/2011022.

Bae, M.-J., Park, Y.-S., 2020. Key determinants of freshwater gastropod diversity and distribution: the implications for conservation and management. Water 12, 1908. https://doi.org/10.3390/w12071908.

Benjamini, Y., Hochberg, Y., 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. R. Stat. Soc. 57 (1), 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x.

Bevilacqua, S., Fraschetti, S., Musco, L., Terlizzi, A., 2009. Taxonomic sufficiency in the detection of natural and human-induced changes in marine assemblages: a comparison of habitats and taxonomic groups. Mar. Pollut. Bull. 58 (12), 1850–1859. https://doi.org/10.1016/j.marpolbul.2009.07.018.

Bony, K.Y., 2007. Biodiversité et écologie des mollusques gastéropodes d'eau douce en milieu continental ivoirien (bassins de la Mé, de l'Agnéby et du Banco). Traits d'histoire de vie d'une espèce invasive Indoplanorbis exustus (Deshayes, 1834). PhD Thesis, Ecole Pratique des Hautes Etudes, Perpignan France.

Bony, Y.K., Kouassi, N.C., Diomandé, D., Gourene, G., Verdoit-Jarraya, M., Pointier, J.P., 2008. Ecological conditions for spread of the invasive snail Physa marmorata (Pulmonata : Physidae) in the Ivory Coast. African Zool. 43, 53–60. https://doi.org/10.1080/15627020.2008.11407406.

Borcard, D., Gillet, F., Legendre, P., 2018. Numerical Ecology with R, 2nd ed. Springer International Publishing AG. https://doi.org/10.1007/978-3-319-71404-2.

Breiman, L., 2001. Random forests. Mach. Learn. 45, 5–32.

Brown, D.S., 1994. Freshwater Snails of Africa and their Medical Importance, Revised 2n. ed. Taylor & Francis Ltd, London, UK.

Brown, D.S., Kristensen, T.K., 1993. A Field Guide to African Freshwater Snails. 1. West African Species. Danish Bilharziasis Laboratory, Charlottenlund, Denmark.

Calamari, D., Naeve, H., 1994. Revue de la pollution dans l'environnement aquatique africain, Document Technique du CPCA. No. 25. FAO, Rome, FAO.

Camara, I.A., Bony, Y.K., Diomandé, D., Edia, O.E., Konan, F.K., Kouassi, C.N., Gourène, G., Pointier, J.P., 2012. Freshwater snail distribution related to environmental factors in Banco National Park, an urban reserve in the Ivory Coast (West Africa). African Zool. 47 (1), 160–168. https://doi.org/10.3377/004.047.0106.

Cutler, A., Cutler, D.R., Stevens, J.R., 2012. Random Forests, in: Zhang, C., Ma, Y. (Eds.), Ensemble Machine Learning: Methods and Applications. pp. 157–175. https://doi.org/10.1007/978-1-4419-9326-7.

De Caceres, M., Jansen, F., 2016. Package 'indicspecies': Relationship Between Species and Groups of Sites. R Packag. version 1.7.6.

Derek, A., Wheeler, P., 2020. Package 'FSA': Simple Fisheries Stock Assessment Methods. R Packag. version 0.8.30.

Dobson, M., Frid, C., 2009. Ecology of Aquatic Systems, second ed. Oxford University Press.

Dray, S., Dufour, A.-B., Thioulouse, J., Jombart, T., Pavoine, S., Lobry, J.R., Ollier, S., Borcard, D., Legendre, P., Bougeard, S., Siberchicot, A., Chessel, D., 2020. Package 'ade4':Analysis of Ecological Data: Exploratory and Euclidean Methods in Environmental Sciences. R Packag. version 1.7-15.

Dufrêne, M., Legendre, P., 1997. Species assemblages and indicator species: the need for a flexible asymmetrical approach. Ecol. Monogr. 67 (3), 345–366.

Durand, J.R., Lévèque, C., 1981. Flore et faune aquatiques de l'Afrique sahelo-soudanienne., Tome 2. ed. ORSTOM, Paris, France.

Galili, T., 2015. dendextend: an R package for visualizing, adjusting and comparing trees of hierarchical clustering. Bioinformatics 31 (22), 3718–3720. https://doi.org/10.1093/bioinformatics/btv428.

Hazoume, R.U.S., 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.

Humphries, G.R.W., Magness, D.R., Huettmann, F., 2018. Machine Learning for Ecology and Sustainable Natural Resource Management. Springer Nature Switzerland AG, Cham, Switzerland, 10.1007/978-3-319-96978-7.

Ibikounlé, M., Gbédjissi, L.G., Ogouyèmi-Hounto, A., Batcho, W., Kindé-Gazard, D., Massougbodji, A., 2014. Schistosomose et géohelminthoses dans le nord-est du Bénin: cas des écoliers des communes de Nikki et de PèrèrèSchistosomiasis and soil-transmitted helminthiasis among schoolchildren of Nikki and Pèrèrè, two northeastern towns of Benin. Bull. la Soc. Pathol. Exot. 107 (3), 171–176. https://doi.org/10.1007/s13149-014-0344-y.

Ibikounlé, M., Mouahid, G., Sakiti, N.G., Massougbodji, A., Moné, H., 2009. Freshwater snail diversity in Benin (West Africa) with a focus on human schistosomiasis. Acta Trop. 111 (1), 29–34. https://doi.org/10.1016/j.actatropica.2009.02.001.

ISO 5667-6, 2014. Water quality — Sampling — Part 6: Guidance on sampling of rivers and streams.

Itsukushima, R., Yoshikawa, H., Morita, K., 2018. A dataset of molluscan fauna sampled in river estuaries of medium and small size river in Kyushu island, Japan. Biodivers. Data J. 6, e26101. https://doi.org/10.3897/BDJ.6.e26101.

Jones, F.C., 2008. Taxonomic sufficiency: the influence of taxonomic resolution on freshwater bioassessments using benthic macroinvertebrates. Environ. Rev. 16 (NA), 45–69. https://doi.org/10.1139/A07-010.
Kohonen, T., 2013. Essentials of the self-organizing map. Neural Networks 37, 52–65. https://doi.org/10.1016/j.neunet.2012.09.018.
Kohonen, T., 1990. The self-organizing map. Proc. IEEE 78, 1464–1480. https://doi.org/10.1109/5.58325.
Kohonen, T., 1982. Self-organized formation of topologically correct feature maps. Biol. Cybern. 43 (1), 59–69. https://doi.org/10.1007/BF00337288.
Konan, F.K., Leprieur, F., Ouattara, A., Brosse, S., Grenouillet, G., Gourène, G., Winterton, P., Lek, S., 2006. Spatio-temporal patterns of fish assemblages in coastal West African rivers: a self-organizing map approach. Aquat. Living Resour. 19 (4), 361–370.
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, Z.C., Odountan, O.H., Van Bocxlaer, B., Sablon, R., Chikou, A., Backeljau, T., 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.
Kuhn, M., Wing, J., Weston, S., Williams, A., Keefer, C., Engelhardt, A., Cooper, T., Mayer, Z., Kenkel, B., R Core Team, Benesty, M., Lescarbeau, R., Ziem, A., Scrucca, L., Tang, Y., Candan, C., Hunt, T., 2020. Package "caret": Classification and Regression Training. Version 6.0-86.
Legendre, P., Legendre, L., 2012. Numerical ecology, 3rd ed, Developments in Environmental Modeling 24. Elsevier Science Publishers B.V., Amsterdam, Netherlands.
Lek-Ang, S., Park, Y.-S., Ait-Mouloud, S., Deharveng, L., 2007. Collembolan communities in a peat bog versus surrounding forest analyzed by using self-organizing map. Ecol. Modell. 203 (1-2), 9–17. https://doi.org/10.1016/j.ecolmodel.2006.01.007.
Li, T., Sun, G., Yang, C., Liang, K., Ma, S., Huang, L., 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 (10), 1610–1624. https://doi.org/10.1007/s11442-019-1644-8.
Liaw, A., Wiener, M., 2002. Classification and Regression by randomForest. R News 2, 18–22.
Manly, B.F.J., 2006. Randomization, Bootstrap and Monte Carlo Methods in Biology. Chapman and Hall/CRC, USA.
Maroco, J., Silva, D., Rodrigues, A., Guerreiro, M., Santana, I., de Mendonça, A., 2011. Data mining methods in the prediction of Dementia: a real-data comparison of the accuracy, sensitivity and specificity of linear discriminant analysis, logistic regression, neural networks, support vector machines, classification trees and random forests. BMC Res. Notes 4, 299.
Milošević, D., Piperac, M.S., Petrović, A., Čerba, D., Mančev, D., Paunović, M., Simić, V., 2017. Community concordance in lotic ecosystems: how to establish unbiased congruence between macroinvertebrate and fish communities. Ecol. Indic. 83, 474–481. https://doi.org/10.1016/j.ecolind.2017.08.024.
MolluscaBase, 2019. MolluscaBase [WWW Document]. URL http://www.molluscabase.org (accessed 10.1.19).
Nicklès, M., 1950. Mollusques Testacés Marins de la Côte Occidentale d'Afrique, secons ed. Manuels Ouest-Africains, Paris, France.
Odountan, O.H., de Bisthoven, L.J., Koudenoukpo, C.Z., Abou, Y., 2019a. Spatio-temporal variation of environmental variables and aquatic macroinvertebrate assemblages in Lake Nokoué, a RAMSAR site of Benin. African J. Aquat. Sci. 44 (3), 219–231. https://doi.org/10.2989/16085914.2019.1629272.
Odountan, O.H., Janssens de Bisthoven, L., Abou, Y., Eggermont, H., 2019b. Biomonitoring of lakes using macroinvertebrates: recommended indices and metrics for use in West Africa and developing countries. Hydrobiologia 826 (1), 1–23. https://doi.org/10.1007/s10750-018-3745-2.
Park, Y.-S., Chon, T.-S., Bae, M.-J., Kim, D.-H., Lek, S., 2018. Multivariate data analysis by means of Self-Organizing Maps, in: Recknagel, F., W K Michener (Eds.), Ecological Informatics. Springer International Publishing, Cham, Switzerland, pp. 251–272. https://doi.org/10.1007/978-3-319-59928-1.
Poff, N.L.R., 1997. Landscape filters and species traits: towards mechanistic understanding and prediction in stream ecology. J. North Am. Benthol. Soc. 16 (2), 391–409. https://doi.org/10.2307/1468026.
R Core Team, 2019. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.
Revell, L.J., 2012. phytools: an R package for phylogenetic comparative biology (and other things). Methods Ecol. Evol. 3, 217–223. https://doi.org/10.1111/j.2041-210X.2011.00169.x.
Rice, E.W., Baird, R.B., Eaton, A.D., Clesceri, L.S., 2012. Standard Methods for the Examination of Water and Wastewater, 22nd ed. American Public Health Association, American Water Works Association, Water Environment Federation, Washington, D.C.
Rodier, J., Legube, B., Merlet, N., Brunet, R., Mialocq, J.-C., Leroy, P., Houssin, M., Lavison, G., Bechemin, C., Vincent, M., Rebouillon, P., Moulin, L., Chomodé, P., Dujardin, P., Gosselin, S., Seux, R., Al Mardini, F., 2009. L'analyse de l'eau. 9ème edition entièrement mise à jour. Dunod, Paris, France, France.
Roméo, M., Frasila, C., Gnassia-Barelli, M., Damiens, G., Micu, D., Mustata, G., 2005. Biomonitoring of trace metals in the Black Sea (Romania) using mussels Mytilus galloprovincialis. Water Res. 39 (4), 596–604.
Seddon, M., Appleton, C., Van Damme, D., Graf, D., 2011. Chapter 4. Freshwater molluscs of Africa: diversity, distribution, and conservation, in: Darwall, W., K, S., Allen, D., Holland, R., Harrison, I., Brooks, E. (Eds.), The Diversity of Life in African Freshwaters: Under Water, Under Threat. An Analysis of the Status and Distribution of Freshwater Species Throughout Mainland Africa. IUCN, Gland, Cambridge, pp. 94–125.
Seddon, M.B., Lange, C., Van Damme, D., 2018. Pisidium pirothi. The IUCN Red List of Threatened Species 2018: e.T165798A120110923. https://dx.doi.org/10.2305/IUCN.UK.2018-2.RLTS.T165798A120110923.en.
Tachet, H., Richoux, P., Bournaud, M., Usseglio-Polatera, P., 2010. Invertébrés d'eau douce: systématique, biologie, écologie. CNRS éditions, Paris, France.
Tchakonté, S., Ajeagah, G.A., Diomandé, D., Camara, A.I., Ngassam, P., 2014. Diversity, dynamic and ecology of freshwater snails related to environmental factors in urban and suburban streams in Douala-Cameroon (Central Africa). Aquat. Ecol. 48 (4), 379–395.
Tsai, W.-P., Huang, S.-P., Cheng, S.-T., Shao, K.-T., Chang, F.-J., 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.
Usero, J., Morillo, J., Gracia, I., 2005. Heavy metal concentrations in molluscs from the Atlantic coast of southern Spain. Chemosphere 59 (8), 1175–1181. https://doi.org/10.1016/j.chemosphere.2004.11.089.
Van Bocxlaer, B., Albrecht, C., Stauffer, J.R., 2014. Growing population and ecosystem change increase human schistosomiasis around Lake Malawi. Trends Parasitol. 30 (5), 217–220. https://doi.org/10.1016/j.pt.2014.02.006.
Vesanto, J., Himberg, J., Alhoniemi, E., Parhankangas, J., 2000. SOM Toolbox for Matlab 5. Technical Report A57. Helsinki, Finland.
Voyslavov, T., Tsakovski, S., Simeonov, V., 2012. Surface water quality assessment using self-organizing maps and Hasse diagram technique. Chemom. Intell. Lab. Syst. 118, 280–286. https://doi.org/10.1016/j.chemolab.2012.05.011.
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.
Wickham, H., François, R., Henry, L., Müller, K., 2020. Package 'dplyr': A Grammar of Data Manipulation. R Packag. Version 0.8.5.
Yonge, C.M., 1962. On *Etheria elliptica* LAM. and the course of evolution, including assumption of Monomyarianism, in the family Etheriidae (Bivalvia: Unionacea). Philos. Trans. R. Soc. Lond. B. Biol. Sci. 244, 423–458.
Zaabar, W., Zakhama-Sraieb, R., Charfi-Cheikhrouha, F., Achouri, M.S., 2018. Composition of a molluscan assemblage associated with macrophytes in Menzel Jemil (Bizerte lagoon, SW Mediterranean Sea). Afr. J. Ecol. 56 (3), 537–547. https://doi.org/10.1111/aje.12490.