Improving the predictive assessment of water biological quality using macrophytes: Empirical testing and method selection

Daniel Gebler

daniel.gebler@up.poznan.pl

Uniwersytet Przyrodniczy w Poznaniu https://orcid.org/0000-0002-6549-9418

Pedro Segurado

University of Lisbon School of Agriculture: Universidade de Lisboa Instituto Superior de Agronomia

Maria Teresa Ferreira

University of Lisbon Higher Institute of Agronomy: Universidade de Lisboa Instituto Superior de Agronomia

Francisca C. Aguiar

University of Lisbon Higher Institute of Agronomy: Universidade de Lisboa Instituto Superior de Agronomia

---

Research Article

Keywords: macrophytes, bioindication, river assessment, artificial neural networks, linear regression, boosted regression trees

Posted Date: April 2nd, 2024

DOI: https://doi.org/10.21203/rs.3.rs-3980058/v1

Abstract

Bioassessment in southern European rivers has been hampered by difficulties in reference data availability and the unknown effect of the interacting multiple stressors on plant communities. Predictive modelling may help to overcome this limitation. This study aims to develop and evaluate macrophyte-based predictive models of the biological status of rivers using various modelling techniques. We compared models based on multiple linear regression (MLR), boosted regression trees (BRT) and artificial neural networks (ANNs). Secondarily, we investigated the relationship between two macrophyte indices grounded in distinct conceptual premises (the Riparian Vegetation Index – RVI, and the Macrophyte Biological Index for Rivers – IBMR) and a set of environmental variables, including climatic conditions, geographical characteristics, land use, water chemistry and habitat quality of rivers. We assembled a dataset of 292 Mediterranean sampling locations on perennial rivers and streams (mainland Portugal) with macrophyte and environmental data. The quality of models for the IBMR was higher than for the RVI for all cases, which indicates a better ecological linkage of IBMR with the stressor and abiotic variables. The IBMR using ANN outperformed the BRT models, for which the r-Pearson correlation coefficients were 0.877 and 0.801, and the normalised root mean square errors were 10.0 and 11.3, respectively. Variable importance analysis revealed that longitude and geology, hydrological/climatic conditions, water body size, and land use had the highest impact on the IBMR model predictions. Despite the differences in the quality of the models, all showed similar importance to individual input variables, although in a different order. Despite some difficulties in model training for ANNs, our findings suggest that BRT and ANNs can be used to assess ecological quality, and for decision-making on the environmental management of rivers.

Introduction

Macrophytes are an important group of freshwater biota that commonly include vascular plants (ferns and angiosperms), bryophytes (mosses and liverworts) and macroscopic algae. These photosynthetic organisms are major primary producers and essential for several key processes, such as carbon and nutrient cycling and air-water-sediment exchanges (Haslam 1987). Macrophytes contribute to habitat creation, supporting other aquatic biota as refugia, nurseries and food sources and are providers of numerous ecosystem services (e.g., Gurnell et al. 2016; O'Hare et al. 2018).

Macrophyte species display consistent and stable responses to environmental change, especially concerning nutrient enrichment, sediment loading and hydrologic alterations (Hering et al. 2010; Aguiar et al. 2014 and references therein). In addition, a biogeographical homogeneity of aquatic plant species across the world and Europe is recognised, such as in the Central-Baltic region or the Mediterranean (Bonada and Resh 2013; Murphy et al. 2019). These characteristics contributed to their inclusion in freshwater bioassessment approaches worldwide and as biological quality elements for the implementation of the European Water Framework Directive (WFD; Birk et al. 2012; Feio et al. 2021).

Highly seasonal and temporary rivers frequently have few submerged and truly aquatic macrophytes (hydrophytes). For this reason, a line of research focusing on the utilisation of both aquatic and riparian plants for biomonitoring has emerged, yielding indices and models applicable to various Mediterranean countries (e.g., Aguiar et al. 2009; Papastergiadou et al. 2016). Nevertheless, the use of aquatic flora (mostly hydrophytes and emergent species) is mandatory under the WFD requirements, and official indices for Mediterranean rivers are grounded in bioindication (Haury et al. 2006). Bioassessment in Southern Europe using macrophytes has been hampered by difficulties in reference data availability and the unknown effect of the interacting stressors on plant communities (Aguiar et al. 2014; Dodkins et al. 2012). Moreover, multiple pressures impact European rivers due to organic pollution inflow via adjacent agricultural lands, hydrologic and geomorphologic alterations, damming and channelisation. The Mediterranean region is widely considered to be highly impaired by water scarcity, which becomes more challenging in the context of climate change with a projected decrease of 30% in annual mean precipitation (Bonada and Resh 2013; Feyen et al. 2020). Understanding how these different stressors interfere with aquatic organisms, including macrophytes, is essential for the development of effective methods to assess and monitor the ecological status of aquatic ecosystems (Hering et al. 2010; Polst et al. 2022). This involves researching innovative assessment systems and advanced analytical tools for enhanced data analysis (e.g., Hering et al. 2018; Rolim et al. 2023).

Regression models are recognized as being valuable tools in management, environmental decision-making, evaluation and ecosystem protection for environments exposed to multiple pressures (Lewis et al. 2021; Park and Lek 2016). They can be used to solve intricate relationships among ecosystem components and to predict their condition in response to changing environmental factors (Tiyasha et al. 2020). In recent decades, various predictive methods have rapidly developed, and they can be successfully used for environmental modelling (Elith et al. 2008; Poisot et al. 2016; Provata et al. 2008). These methods are derived from simple linear and nonlinear models. Moreover, more complex methods are used, such as neural networks, genetic algorithms, Bayesian networks, regression trees or random forests (Zhang et al. 2015). The different new methods of regression showed satisfactory quality concerning all major groups of aquatic biota, such as benthic macroinvertebrates, fishes, phytoplankton and macrophytes. Many authors have presented also comparisons between different regression methods. However, studies rarely refer to macrophyte indices, especially when compared to other biological indices used in freshwater ecology and bioassessment (Tiyasha et al. 2020 and references therein).

This study used various modelling techniques to develop and compare macrophyte-based predictive models for the bioassessment of Mediterranean rivers and streams. We also investigated the connections between two macrophyte-based indices widely used in Portugal, the Riparian Vegetation Index (RVI; Aguiar al. 2009), and the Macrophyte Biological Index for Rivers (IBMR; Haury et al. 2006) and environmental variables characterising Mediterranean perennial watercourses.

Material and methods

Site selection

The study concerns rivers and streams of mainland Portugal, on the western edge of the Iberian Peninsula, South Western Europe (Fig. 1). The country covers approximately 89 000 km2, with a large extension of the Atlantic Ocean coastline, and the influence of the Mediterranean Sea on the south region. Most of the study area has a temperate Mediterranean climate, with hot, dry summers and mild, wet winters. Inter-annual variability and multi-annual droughts are frequent. The coastal area is densely populated and impaired by forestry, agriculture and industry, and northern areas have a complex landscape of vineyards, orchards and small agricultural lands. South and inland areas have scattered settlements, extensive agricultural lands, Mediterranean scrubland and oak forests. Many rivers and streams in Southern regions have temporary streamflow regimes.

We used the macrophyte dataset (n = 402; 2004–2006; Portuguese Environmental Agency; https://www.apambiente.pt/dqa/macrofitos.html), developed for the WFD implementation in Portugal and the intercalibration of Mediterranean rivers and streams. For the modelling purposes, we defined a subset of sampling sites that agreed with the following criteria: i) rivers and streams having perennial streamflow; ii) abundance data on both aquatic and riparian plants; and iii) availability of data on water chemistry, habitat quality and hydrology. We attained a final dataset of 292 sites and a group of rivers of southern Portugal were removed from the database (Fig. 1). The main reason for site removal was the lack of aquatic macrophytes in streams with a temporary streamflow regime.

Sampling methods were based on the European standards EN14184:2003 and EN14996:2006 and followed a national protocol (Aguiar et al. 2014). Surveys were performed in late spring–early summer (May-June/July) in the in-stream part that is submerged most of the year, although it may be exposed temporarily under conditions of dry-water flow, usually in summer) or for more extended periods under certain natural (climatic, geological) conditions. Sampling involves wading into the water and following a zigzag pattern upstream along the reach length, usually 100 m long sections of the river channel. Exceptionally, surveys were made from one or both margins. The sites had 100 m of river length, with a minimum sampling area of 50 m^2. For the RVI, the surveys were performed both in-stream and on the riverbanks, including aquatic vegetation, riparian herbaceous and woody species (trees, shrubs and lianas). The superficial cover by each taxon was estimated as a percentage. The data included mostly bryophytes and vascular plants.

Environmental variables

The national wide-network abiotic database of rivers and streams allowed the selection of 24 environmental variables with standardised data representing the main characteristics of monitoring sites (Table 1). Eight variables characterised the geographical (latitude, longitude and altitude), climatic specificity of rivers (thermal conditions, annual precipitation and runoff) and catchment characteristics (distance to the source and catchment area). Land use of the catchment included the proportion of natural areas, artificial areas, extensive agriculture (pastures, non-irrigation crops), and intensive agriculture (irrigation crops, orchards, vineyards). Three indices characterise habitat quality, namely, the Riparian Forest Quality Index (Munné et al. 2003) and two hydromorphological indices (Habitat Quality Assessment score and Habitat Modification Score) calculated by the application of the River Habitat Survey method (Raven et al. 1998). Nine variables characterise the water quality, e.g., water temperature, oxygen concentration, pH, orthophosphates and different forms of nitrogen (Table 1).

Table 1
Descriptive statistics of the explanatory variables

Variable	Range	Mean ± SD
Geography and climate		
Latitude (km in EPSG 20790 system)	93.6-323.7	209.1 ± 52.0
Longitude (km in EPSG 20790 system)	24.6-572.1	391.5 ± 129.5
Altitude (m)	3.0-1414.4	249.8 ± 237.5
Thermal range (°C)	6.1–14.7	10.3 ± 1.3
Mean annual precipitation (mm)	489.0-2926.0	1156.6 ± 501.4
Mean annual runoff (mm)	75.0-2200.0	524.9 ± 389.4
Distance to source (km)	0.03–237.2	36.7 ± 43.1
Catchment area (km^2)	1.1-5401.6	483.4 ± 963.7
Land use		
Natural areas in the catchment (%)	2.0-100.0	61.4 ± 26.0
Artificial areas in the catchment (%)	0.0–40.0	3.4 ± 5.6
Extensive agriculture in the catchment (%)	0.0–76.0	6.1 ± 10.2
Intensive agriculture in the catchment (%)	0.0-100.0	30.2 ± 24.3
Habitat quality		
Habitat Quality Assessment (-)	14.0–64.0	41.8 ± 8.5
Habitat Modification Score (-)	0.0–67.0	11.4 ± 11.7
Riparian Forest Quality Index (-)	0.0-100.0	57.4 ± 25.5
Water quality		
Water temperature (°C)	5.7–29.2	16.0 ± 3.8
Dissolved oxygen concentration (mg O_2/L)	2.1–17.5	9.5 ± 2.1
pH (-)	5.0–9.0	7.0 ± 0.7
Conductivity (uS/cm)	9.2–1388.0	183.9 ± 243.1
Alkalinity (mg HCO_3^{2-}/L)	1.5–442.0	57.2 ± 77.2
Total suspended solids (mg/L)	0.0–88.0	9.0 ± 12.7
Nitrates (mg NO_3^-/L)	0.01—34.2	3.4 ± 4.7

Variable	Range	Mean ± SD
Ammonia (mg NH_4^+/L)	0.01-13.0	0.3 ± 1.0
Orthophosphates (mg PO_4^{3-}/L)	0.01-6.0	0.2 ± 0.5

Macrophyte-based indices

The IBMR was first described by Haury et al. (2006) as an index for assessing water trophy and organic pollution. The IBMR was accepted as an official national method for classifying the ecological status of highly seasonal rivers of all EU Member States of the Mediterranean Geographical Intercalibration Group, except Slovenia (Aguiar et al. 2014). Following the WFD intercalibration, the IBMR was tested by other Mediterranean countries, such as Turkey (Özbay et al. 2019) and Greece (Stefanidis et al. 2022). It can be calculated using the following formula:

$$IBMR = \frac{\sum_{i=1}^{N}(CS_i \times E_i \times K_i)}{\sum_{i=1}^{N}(E_i \times K_i)}$$

where i – bioindicator taxon, CS_i – indicator value for the i-th taxon (0–20) expressing the preferred trophy level, E_i – stenoecy coefficient (weighting factor) of the i-th taxon, expressing ecological tolerance (1–3), K_i – abundance of the i-th taxon (translated in 5 classes). We considered that a minimum number of four bioindicator taxa is needed for a reliable IBMR calculation in Mediterranean rivers (Aguiar et al. 2014).

The RVI is a multimetric macrophyte-based index that uses the responses of structural and functional parameters of aquatic and riparian vegetation to global disturbance (Aguiar et al. 2009). The RVI has been widely used for environmental impact assessment purposes and general hydromorphological diagnostic of rivers and streams in mainland Portugal. It includes compositional metrics (e.g. cover and number of alien and endemic species) and functional metrics associated with life cycle and reproduction (e.g. proportion of perennial species), and with the trophic status (e.g. proportion of nitrophilous species). RVI uses a table of conversion of metric values into dimensionless values through a scale of three scores: 1 - poor quality, 3 - fair quality, and 5 - good quality. Metric values were obtained using reference and non-reference site values for each parameter. The RVI for a site was obtained by the sum of the quality scores of all metrics, subtracted by the total number of metrics (for more details see Aguiar et al. 2009). Five ecological quality classes were assigned: high (best of five classes), good, moderate, poor and bad. The high/good boundary was established using the 25th percentile of the reference sites. The four remaining categories were derived by evenly dividing the range between the upper limit defined by the high/good boundary and the lower extremity of the gradient.

Modelling approaches

We tested three different types of models: multiple linear regression (MLR), boosted regression trees (BRT) and artificial neural networks (ANNs). For modelling purposes, the dataset was randomly divided into two or three datasets used in different phases of model creation. Seventy percent of the cases were used in the training of models. The remaining thirty percent of cases were then used to validate the MLR and the BRT. ANN modelling consists of three phases: training, additional testing, and validation. Therefore, the testing and validation datasets used 15% of cases each. Concerning the environmental variables dataset, we reduced their number and avoided collinearity between them using factor analysis (FA) (StatSoft Inc. 2017). The reduction of input variables in the model leads to fewer cases needed in the modelling process, simplifying the overall structure of models and eliminating redundant information without affecting the model's error (Dormann et al. 2013).

Multiple linear regression is a classical statistical approach used in an enormous number of historical and current studies. It is a method based on the assumption of a linear relationship between one dependent variable and a set of independent explanatory variables (Olive 2017). In our study, we fitted several candidate MLR models using all possible combinations of predictors. The selection of the final model was based on a multimodel inference procedure (Grueber et al. 2011) using the Akaike weight (Burnham and Anderson 2002) and the variable effect sizes, as measured by the absolute standardised regression coefficients. The coefficients of the final model were derived from averaging the estimated coefficients of models with the highest AIC weights and delta AIC < 2 (absolute difference between the AIC of each model and the best approximating model below two). MLR models were fitted using R version 4.2.0 (R Core Team 2022) and the multimodel inference was performed with the MuMIn package for R (Bartoń 2016).

Boosted regression trees are additive regression models that differ from traditional regression methods. BRT involves fitting a sequence of classification or regression trees, each seeking to explain the variation in data not explained by the previous tree, until a certain level of predictive performance is achieved. Each tree is fitted using a random sample of observations and each node within the tree is based on a random subset of variables. In this study, we fitted BRT models using the optimisation procedure implemented in the dismo package for R version 1.3-9 following the stepwise procedure recommended by Elith et al. (2008). To optimise the number of trees in each BRT model, we performed a stepwise process based on 10-fold cross-validations using mean deviance to measure predictive performance. The tree complexity, a parameter that controls the number of interactions among variables (i.e., the number of splits of individual trees), was set to two (pairwise interactions). A second parameter, the learning rate, which determines the contribution of each tree to the growing model, was set iteratively to ensure that at least 1000 trees were achieved after the stepwise process (Elith et al. 2008).

Artificial neural networks are a type of deep learning algorithm that consists of layers of interconnected nodes (artificial neurons) that can learn and extract features from data. ANNs use a process called backpropagation to update the weights and biases of the neurons in each layer, optimising the network to make better predictions. This study used the multilayer perceptron (MLP) type of ANN with a Broyden–Fletcher–Goldfarb–Shanno (BFGS) algorithm, available in STATISTICA 13 (StatSoft Inc. 2017). This type of network is trained with the supervised teacher technique called the delta rule. The ANN was previously recognised as a valuable method for modelling nonlinear relationships (Park and Lek 2016) and was used in similar studies (Gebler et al. 2018; Krtolica et al. 2021; Rocha et al. 2017). A three-layer neural network was used in this study. The first layer included factors obtained in the FA. The hidden layer consisted of neurons ranging from ($2n^{1/2} + m$) to ($2n + m$), where n and m are the number of input and output neurons, respectively (Fletcher and Goss 1993). The out neuron was always one (IBMR or RVI).

Due to the recommendations (e.g., Park and Lek 2016), the application of statistical models was preceded by output data range normalisation (min-max normalization) to a range of 0.1–0.9. The input variables were standardised by the autoscaling method (linear transformation carried out by scaling the values with mean = 0 and variance = 1) in factor analysis.

The quality of each model (i.e., MLR, BRT and ANN) was evaluated using three parameters, that are commonly used performance measures in similar studies (Hernandez-Suarez and Nejadhashem 2018): coefficient of determination (R^2, Eq. 1) representing the amount of explained variance, the r-Pearson correlation coefficient (r, Eq. 2) showing the fitness of data, and the normalised root mean square error (NRMSE, Eq. 3) based on values of biological indices and modelled values.

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - y_i')^2}{\sum_{i=1}^{n} (y_i - \bar{y}^{i})^2} \text{ (Eq. 1)}$$

$$r = \frac{\sum_{i=1}^{n} (y_i - \bar{y})(y_i' - \bar{y}')}{\sqrt{\sum_{i=1}^{n} (y_i - \bar{y})^2 \sum_{i=1}^{n} (y_i' - \bar{y}')^2}} \text{ (Eq. 2)}$$

$$NRMSE = \frac{\sqrt{\frac{\sum_{i=1}^{n} \left(y_i' - y_i\right)^2}{n}}}{y_{max} - y_{min}} \text{ (Eq. 3)}$$

where y_i – i-th value of the output variable, $\bar{y}$ – mean value of the output variable, y_i' – i-th value of the output variable derived from the model, $\bar{y}'$ – mean value of the output variable derived from the model, y_{max} – maximum value of the variable, y_{min} – minimum value of the variable and n – number of cases.

The important aspect of our work was identifying which environmental factors significantly influence the modelled values of the two macrophyte indicators. The relative importance of predictor variables in MLR was assessed based on the variable effect sizes. In the case of BRT models, the relative importance of each predictor was estimated by averaging the number of times each variable was selected for splitting a tree and the squared improvement resulting from these splits. For the ANNs, a sensitivity analysis was carried out. This analysis shows how much the model error would increase after removing a given variable. Higher values indicate greater importance of the variable in the model. To compare the variable importance derived by different models (MLR, BRT, ANN), in each case, the relative influence of each variable in each model was scaled to 100, with higher numbers indicating a higher contribution to the outputs (Elith et al. 2008).

Results

Factor analysis (FA) enabled us to reduce the original set of 24 environmental variables to eight latent factors (Table A1). The first factor combines longitude, alkalinity and conductivity information, indicating information on geographical gradient and geology. The second extracted factor refers to water quality regarding nutrient content. The third factor was an emanation of the water body's size described by the catchment's size and the associated distance from the source. The fourth factor represented hydrological and climatic conditions. The fifth factor indicated the quality of habitat, encompassing the naturalness and modifications of both the river channel and riparian zone. The next factor was related to different categories of catchment land use. The last two factors were related to the physicochemical quality of water (total suspended solids and pH) and oxygen conditions, respectively. The eight identified factors explained more than 70% of the variance in the environmental variables presented by the dataset. The acquired factors were subsequently used as explanatory (input) variables in all models for both IBMR and RVI indices.

Modelling performance

All methods used (MLR, BRT, ANN) for the IBMR prediction had an overall higher quality than the exact methods modelling the RVI. The analysis of model quality focused on evaluating the performance parameter values for the calibration-independent validation dataset. A comparison of the model quality of all models is presented in Table 2, and the plot of the modelled against observed values of the macrophyte indices is presented in Fig. 2. It's noteworthy that, regardless of the modelled index, the hierarchy of the efficacy of the different regression methods remained consistent and was ranked as follows: ANN > BRT > MLR. The ANN revealed the strongest correlation between the predicted and observed values, showing consistent similarity between them. This model also explained the highest percentage of the variance of the modelled variables and had the lowest modelling errors. This was demonstrated for the IBMR model by the value of the coefficient of determination for the final validation dataset (0.77), the r-Pearson correlation coefficient (0.88), and the normalised mean squared error value, which did not exceed 10.0%. The same type of model for the RVI was of correspondingly lower quality (R^2 = 0.67, r = 0.82, NRSME = 15.7%).

The prediction accuracy of BRT for both indices showed a diminished quality in the final validation procedure when compared to the ANN. Despite higher performance parameters during the training step, which marks the initial phase in model creation, BRT's overall validation performance fell short of ANN. Moreover, a similar dependence between BRT models for the two indices appeared, showing that the prediction of the IBMR was more accurate than that of the RVI. The values of the determination coefficient for the validation procedure were 0.64 and 0.60, and the fitness of the modelled and observed data determined by the r-Pearson correlation coefficient were 0.80 and 0.78, respectively. Moreover, the NRMSE values were 12.3% vs. 17.1%.

Both MLR models exhibited the lowest quality of prediction and demonstrated subpar performance parameters. The coefficients of determination for the IBMR and RVI models were below 0.5, signifying that these models accounted for less than half of the variance in the modelled variables. Furthermore, a less accurate fit of the modelled data (r = 0.74 and 0.68, respectively) and higher errors (NRMSE: 13.8 vs. 22.7%) were obtained indicating again a better model quality for the IBMR compared to the RVI.

Table 2
Quality statistics of the constructed ANNs

Index	Type of model	r			NRMSE (%)		
		Training	Testing	Validation	Training	Testing	Validation
IBMR	MLR	0.73	-	0.74	12.7	-	13.8
	BRT	0.94	-	0.80	6.5	-	12.3
	ANN	0.88	0.87	0.88	8.9	12.4	10.0
RVI	MLR	0.68	-	0.68	23.9	-	22.7
	BRT	0.88	-	0.78	12.8	-	17.1
	ANN	0.85	0.80	0.82	13.9	18.2	15.7

Importance analysis

The diverse modelling methods displayed similar importance to the individual input variables (Fig. 3) despite noticeable differences in the performance of individual models. In most cases, Factor 1, encompassing longitude and catchment geology, exerted the greatest influence on the modelled values of the IBMR index. This factor proved to be the most important in both the ANN and BRT models. Hydrological and climatic conditions (Factor 4), water body size (Factor 3) and land use (Factor 6) were also crucial input factors in models, albeit with different ranking positions. The hydrological and climatic factors were second in the ANN and third in BRT, and the water body size factor was third (value very close to Factor 4) in the ANN and fourth in BRT. Land use emerged as the fourth most influential in the networks but appeared second in BRT. In addition, two other factors, nutrients (Factor 2) and habitat quality (Factor 5), also affected the modelling values of the IBMR, but their importance was relatively lower. The final order of the first six variables in the ANNs was as follows: Factor 1 > Factor 4 > Factor 3 > Factor 6 > Factor 2 > Factor 5, while in the BRT, it was as follows: Factor 1 > Factor 6 > Factor 4 > Factor 3 > Factor 2 > Factor 5. The six input parameters selected for MLR overlapped with the most relevant parameters for the ANN and BRT, and their order closely resembled that of the previous methods: Factor 1 > Factor 3 > Factor 4 > Factor 6 > Factor 2 > Factor 5. The other parameters, namely oxygen conditions, TSS, and pH (Factors 8 and 7), were consistently ranked lower across all three methods, indicating their diminished importance in influencing the IBMR models.

In the case of the RVI, the various models displayed greater differences in the importance of specific factors compared to the IBMR (Fig. 3). Despite these differences, it is noteworthy that for the RVI, longitude and geology (Factor 1) were also essential parameters. Factors 3 and 6, encompassing water body size and land use, respectively, along with habitat quality (Factor 5) and nutrients (Factor 2), were also important. Similar to previous models for the IBMR, Factor 7 (TSS and pH) has little importance, and in contrast to the IBMR, Factor 4, i.e., climatic conditions, although still significant for MLR, is not as crucial for ANN and BRT models.

Discussion

The study showed that different modelling approaches can express and quantify the relationship between the macrophyte-based indices for water quality assessment and a range of environmental variables. This underscores the potential of these models as practical tools for predicting the biological quality of water bodies that are not subject to direct monitoring. The accuracy of such predictions relies on the careful selection of the most suitable model and its associated environmental variables. The quality of modelling by artificial neural networks (ANNs), boosted regression trees (BRT), and multiple linear regression (MLR) can vary according to the specific goals and the dataset (units, size, quality). However, machine-learning techniques often provide better solutions (Hernandez-Suarez and Nejadhashem 2018; Mata et al. 2021; Ren et al. 2020).

The results showed that there were noteworthy differences in the prediction performance of the obtained models. The higher performance quality of the ANN and BRT can be attributed to their capability to address the complexity of observed ecological mechanisms. This makes them well-suited for tackling these types of intricate problems (Hernandez-Suarez and Nejadhashem 2018; Mata et al. 2021; Ren et al. 2020). The higher quality neural networks and boosted regression trees may be because they can deal with the complexity of the observed ecological mechanisms, making them suitable for these kinds of problems (Lemm et al. 2021; Park and Lek 2016). Both methods use iterative algorithms and obtain results by weighing the predictors in the learning or boosting procedures and making them suitable for large data analysis. The methods are relatively easy and fast to train and can produce highly accurate predictions of nonlinear relationships and interactions between variables (El Bouchefry and de Souza 2020; Elith et al. 2008). Contrary to linear regression, they do not assume a linear relationship between the independent and dependent variables, the homogeneity of variances or the normality of data. For this reason, multilinear regression may not be appropriate for the considered problem because many studies indicate that the relationships observed in aquatic ecosystems are often highly complex and nonlinear (Boldina and Meninger 2016), and new analytical method provides better results (e.g., Satich et al. 2022; Schreiber et al. 2022).

Early work (e.g., Silver and Babtist 2000) pointed to the relatively small amount of ecological data available, which limited the application of ANNs. Similar inferences can be drawn for BRTs taking into account their nature (Elith et al. 2008). Although linear regression is the most common method, a notable increase in monitoring efforts has guaranteed the availability of data for the implementation of new analysis approaches. Data obtained within water monitoring programs cope with different sources of bias (Schreiber et al. 2022), which predisposes towards the use of an ANN and BRT over MLR in modelling these relationships (Elith et al. 2008; Park and Lek 2016). Artificial neural networks have been considered a suitable alternative for modelling biological indices in rivers (e.g., Krtolica et al. 2021) and lakes (e.g., Luo et al. 2019). The BRT also indicated satisfactory quality in similar aquatic ecosystem modelling issues (Elias et al. 2016; Lemm et al. 2021).

One of the analysed aspects in the study was the comparison of the modelling capabilities of two ecological assessment indices: IBMR and RVI. The obtained results unequivocally indicated higher modelling quality of the IBMR compared to the RVI, regardless of the applied model (ANN, BRT or MLR). This most likely arises from a more consistent relation between the aquatic vegetation and the environmental variables (input data), particularly those with the most significant influence on the modelled values (Gebler et al. 2018). Additionally, linkages of macrophyte indices of river assessment, including the IBMR, with these factors were presented in other works (Aguiar et al. 2014; Krtolica et al. 2021), which confirms our findings. In fact, IBMR bioindicator species are truly aquatic species or hydrophytes and emergent species (or helophytes). These species rely heavily on water (hydrophytes) or are adapted to both wet or waterlogged substrates (helophytes). In addition, these aquatic communities are far more homogeneous than the riparian communities, which have complex vertical and spatial zonation within the riparian zone and a diverse linkage to the environment. The RVI incorporates both aquatic and riparian vegetation, which has the advantage of expressing the overall condition of rivers and streams and likely expresses the influences of multiscale environmental processes. However, the connection to environmental variables can be obscured by the diverse vegetation units (individual plants, vegetation patches, plant communities, and riparian corridors) that are associated with specific spatial and temporal scales (González del Tánago et al. 2021). Some efforts have been made by Aguiar et al. (2011) to test the suitability of predictive modelling approaches for water quality assessment in Mediterranean rivers. They concluded that the performance of the diverse methods was difficult to compare as they express different types of disturbance acting at diverse spatial scales. Another limitation is the interannual variability affecting plant composition, which should also be incorporated.

Undoubtedly, there is a strong relationship between various environmental factors typical for aquatic ecosystems and riparian vegetation (Aguiar et al. 2011; González de Tánago et al. 2021; Rodrigues et al. 2019). However, the quality of our models revealed that this dependency is insufficient to be utilised in the models. Based on this, we can also assume that our input variables did not include factors strongly associated with the RVI. Despite the evident connection between riparian vegetation and the river environment itself, the influence of other factors and stressors typical for terrestrial environments (e.g., soil erosion, sedimentation), along with climatic factors, can significantly affect the riparian zones (Steiger and Gurnell 2003). Stella and Bendix (2019) highlighted the numerous pressures exerted on riparian vegetation. These authors mentioned natural and anthropogenic disturbances that are more typical for terrestrial ecosystems (land use, storms, temperature), including factors relevant in the Mediterranean or similar climate ecosystems (fires, droughts).

The importance analysis indicated that despite modelling two different indices using three different methods, comparable results were obtained regarding the most significant environmental factors influencing these indices. Significant influences were attributed to longitude and geology, which can be associated with the specific characteristics of the studied rivers. In Portugal, the coast is primarily sedimentary (calcareous), highly affected by urban settlements and organic pollution, and can affect aquatic vegetation (Aguiar et al. 2011). This factor was grouped with conductivity and alkalinity, factors often affecting river macrophyte development (Feio et al. 2012; Szoszkiewicz et al. 2020). Our research also suggested that hydrological conditions significantly influence macrophyte indices. The impact of water flow conditions in rivers (e.g., droughts and flash floods), especially in southern European rivers (Bonada and Resh 2013), was often highlighted as crucial for developing macrophytes and other aquatic organisms. This effect could be even more pronounced if temporary rivers were included in this study. These rivers have a large variability in hydrological conditions that greatly shape fluvial flora and fauna (Cid et al. 2017; Feio et al. 2012; Stefanidis et al. 2021, 2022). According to our results, other significant factors were water body size and land use, which can be linked to conditions within the catchment area, including pressures and their accumulation with the catchment area's size and their direct impact on river ecosystems (Aguiar et al. 2011). These factors also appeared to be important factors affecting inland waters in the multistressor study (Lemm et al. 2020).

There were some factors that ranked lower in the importance analysis, such as total suspended solids and pH. Typically, these factors play a prominent role in plant distribution and growth, which has been demonstrated elsewhere (Demars et al. 2012). Comparable outcomes were achieved in neural network models for temperate rivers in Central and Northern Europe. In these models, factors other than pH, such as nutrients or habitat quality, were notably more significant (Bucior et al. 2021; Gebler et al. 2018; Krtolica et al. 2021; Szoszkiewicz et al. 2020). Interestingly, in the various models implemented for the IBMR, nutrients and habitat conditions were also not among the most significant factors, ranking fifth and sixth in importance, although they were often considered crucial for macrophyte development and differentiation (Haury et al. 2006; Szoszkiewicz et al. 2020), including macrophytes of the Iberian and Mediterranean rivers (Aguiar et al. 2014; Papastergiadou et al. 2016; Stefanidis et al. 2022).

The undervaluation of water physicochemical characteristics may be attributed to the extensive dataset employed in model development, which encompasses large geographical variations across the country and shared ecological typologies that can bias the relevance of water quality variables. Notably, the dataset includes the longitudinal gradient (West—East), mirroring geological distinctions, ranging from sedimentary coastal rivers in a high-populated area to continental rivers characterised by geological formations such as granites, schists, and quartzites, each exhibiting diverse degrees of metamorphic alteration, modulated by varying altitude and climatic variables and less human pressure. Therefore, broad regional variables were more valued in the models and may obscure the local effects of nutrient loads from point sources (e.g., sewage) or nonpoint pollution sources (e.g., agriculture). From the management point of view, these findings indicate that there is a need to refine the river typology and the pressure data inputs to overcome biogeographical differences.

Conclusions

The presented study explored the potential of macrophyte-based indices to assess the water's biological quality and the modelling methods for decision-making on river environmental management. Our findings provide evidence that artificial neural networks and boosted regression trees are well-suited for modelling the intricate relationships between environmental variables and the plant biota of aquatic ecosystems. Almost all the developed models emphasised the crucial role of longitude and geology, indicating the importance of geographic factors (both anthropogenic and natural). Furthermore, the study revealed that in permanent rivers and streams of Southern Europe, hydrological conditions significantly impact macrophytes, surpassing the effect of nutrient levels or habitat quality. While the latter factors hold importance, they should be regarded as complementary factors in more comprehensive models.

Declarations

Authors contribution: D.G., F.C.A., P.S. and M.T.F. designed the research. F.C.A. provided and pre-processed data. D.G. and P.S. conducted the statistical analysis. All authors were involved in the interpretation of results and the manuscript preparation and adjustment.

Funding information: D.G. was granted by the Polish National Agency for Academic Exchange [grant number PPN/BEK/2018/1/00401]. This research was funded by the Forest Research Centre, a research unit funded by Fundação para a Ciência e a Tecnologia I.P. (FCT), Portugal [grant number UIDB/00239/2020]. F.C.A. is also funded via FCT with the reference DOI: 10.54499/DL57/2016/CP1382/CT0028.

Ethics approval consent to participate: Not applicable.

Consent for publication: Not applicable.

Conflict of interest: The authors declare that they have no conflict of interest.

Competing interests: The authors declare that they have no competing interests.

Data availability: Data will be made available on request.

References

1. Aguiar FC, Feio MJ, Ferreira MT (2011) Choosing the best method for stream bioassessment using macrophyte communities: indices and predictive models. Ecol Indic 11:379–388. https://doi.org/10.1016/j.ecolind.2010.06.006

2. Aguiar FC, Ferreira MT, Albuquerque A, Rodríguez-González P, Segurado P (2009) Structural and functional responses of riparian vegetation to human disturbance: performance and spatial scale-dependence. Fund Appl Limnol 175/3:249–263. https://doi.org/10.1127/1863-9135/2009/0175-0249
3. Aguiar FC, Segurado P, Urbanič G, Cambra J, Chauvin C, Ciadamidaro S, Dörflinger G, Ferreira J, Germ M, Manolaki P, Minciardi MR, Munné A, Papastergiadou E, Ferreira MT (2014) Comparability of river quality assessment using macrophytes: A multi-step procedure to overcome biogeographical differences. Sci Total Environ 476–477:757–767. https://doi.org/10.1016/j.scitotenv.2013.10.021
4. Birk S, Bonne W, Borja A, Brucet S, Courrat A, Poikane S, Solimini A, van de Bund W, Zampoukas N, Hering D (2012) Three hundred ways to assess Europe's surface waters: an almost complete overview of biological methods to implement the Water Framework Directive. Ecol Indic 18:31–41. https://doi.org/10.1016/j.ecolind.2011.10.009
5. Boldina I, Meninger PG (2016) Strengthening statistical usage in marine ecology: Linear regression. J Exp Mar Biol Ecol 474:81–91. https://doi.org/10.1016/j.jembe.2015.09.010
6. Bonada N, Resh VH (2013) Mediterranean-climate streams and rivers: Geographically separated but ecologically comparable freshwater systems. Hydrobiologia 719:1–29. https://doi.org/10.1007/s10750-013-1634-2
7. Bowden WB, Glime JM, Riis T (2017) Macrophytes and bryophytes. In: Hauer FR, Lamberti GA (eds) Methods in stream ecology, 3rd edn. Academic, San Diego, California., pp 243–271. https://doi.org/10.1016/B978-0-12-416558-8.00013
8. Bucior A, Rippey B, McElarney Y, Douglas R (2021) Evaluating macrophytes as indicators of anthropogenic pressures in rivers in Ireland. Hydrobiologia 848:1087–1099. https://doi.org/10.1007/s10750-021-04516-x
9. Burnham KP, Anderson DR (2002) Model Selection and Inference: A Practical Information-Theoretic Approach, 2nd edn. Springer-, New York. http://dx.doi.org/10.1007/b97636
10. Cid N, Bonada N, Carlson SM, Grantham TE, Gasith A, Resh VH (2017) High Variability Is a Defining Component of Mediterranean-Climate Rivers and Their Biota. Water 9:52. https://doi.org/10.3390/w9010052
11. Demars BOL, Potts JM, Tremolières M, Thiébaut G, Gougelin N, Nordmann V (2012) River macrophyte indices: not the Holy Grail! Freshw Biol 57:1745–1759. https://doi.org/10.1111/j.1365-2427.2012.02834.x
12. Dodkins I, Aguiar F, Rivaes R, Rodríguez-González P, Albuquerque A, Ferreira MT (2012) Measuring ecological change of aquatic macrophytes in Mediterranean Rivers. Limnologica 42:95–107. https://doi.org/10.1016/j.limno.2011.09.001
13. Dormann CF, Elith J, Bacher S, Buchmann C, Carl G, Carre G, Garcia Marquez JR, Gruber B, Lafoourcade B, Leitao PJ, Münkemüller T, Mcclean C, Osborne PE, Reineking B, Schröder B, Skidmore AK, Zurell D, Lautenbach S (2013) Collinearity: a review of methods to deal with it and a simulation

study evaluating their performance. Ecography 5:1–20. https://doi.org/10.1111/j.1600-0587.2012.07348.x

14. El Bouchefry K, de Souza RS (2020) Chap. 12 - Learning in Big Data: Introduction to Machine Learning. In: Skoda P, Adam F (Eds) Knowledge Discovery in Big Data from Astronomy and Earth Observation. 1st Edn. Elsevier, pp 225–49. https://doi.org/10.1016/B978-0-12-819154-5.00023-0
15. Elias CL, Calapez AR, Almeida SFP, Chessman B, Simões N, Feio MJ (2016) Predicting reference conditions for river bioassessment by incorporating boosted trees in the environmental filters method. Ecol Indic 69:239–251. https://doi.org/10.1016/j.ecolind.2016.04.027
16. Elith J, Leathwick JR, Hastie T (2008) Boosted regression trees – a new technique for modelling ecological data. J Anim Ecol 77:802–813. https://doi.org/10.1111/j.1365-2656.2008.01390.x
17. European Environment Agency (EEA) (2021) Water resources across Europe — confronting water stress: an updated assessment. Publications Office of the European Union, Luxembourg. https://doi:10.2800/320975
18. Feio MJ, Aguiar FC, Almeida SFP, Ferreira MT (2012) AQUAFLORA: A predictive model based on diatoms and macrophytes for streams water quality assessment. Ecol Indic 18:586–598. https://doi.org/10.1016/j.ecolind.2012.01.016
19. Feio MJ, Hughes RM, Callisto M, Nichols SJ, Odume ON, Quintella BR, Kuemmerlen M, Aguiar FC, Almeida SFP, Alonso-EguíaLis P, Arimoro FO, Dyer FJ, Harding JS, Jang S, Kaufmann PR, Lee S, Li J, Macedo DR, Mendes A, Mercado-Silva N, Monk W, Nakamura K, Ndiritu GG, Ogden R, Peat M, Reynoldson TB, Rios-Touma B, Segurado P, Yates AG (2021) The Biological Assessment and Rehabilitation of the World's Rivers: An Overview. Water 13:371. https://doi.org/10.3390/w13030371
20. Feyen L, Ciscar JC, Gosling S, Ibarreta D, Soria A (eds) (2020) Climate change impacts and adaptation in Europe: JRC Peseta IV final report. Joint Research Centre, Publications Office of the European Union, Luxembourg. https://op.europa.eu/en/publication-detail/-/publication/c707e646-99b7-11ea-aac4-01aa75ed71a1/language-en
21. Fletcher D, Goss E (1993) Forecasting with neural networks: an application using bankruptcy data. Inf Manag 24:159–167. https://doi.org/10.1016/0378-7206(93)90064-Z
22. Gebler D, Wiegleb G, Szoszkiewicz K (2018) Integrating river hydromorphology and water quality into ecological status modelling by artificial neural networks. Water Res 139:395–405. https://doi.org/10.1016/j.watres.2018.04.016
23. González del Tánago M, Martínez-Fernández V, Aguiar FC, Bertoldi W, Dufour S, García de Jalón D, Garófano-Gómez V, Mandzukovski D, Rodríguez-González PM (2021) Improving river hydromorphological assessment through better integration of riparian vegetation: Scientific evidence and guidelines. J Environ Manage 292:112730. https://doi.org/10.1016/j.jenvman.2021.112730
24. Grueber CE, Nakagawa S, Laws RJ, Jamieson IG (2011) Multimodel inference in ecology and evolution: challenges and solutions. J Evol Biol 24:699–711. https://doi.org/10.1111/j.1420-9101.2010.02210.x

25. Gurnell AM, Corenblit D, García de Jalón D, González del Tánago M, Grabowski RC, O'Hare MT, Szewczyk M (2016) A conceptual model of vegetation-hydrogeomorphology interactions within river corridors. River Res Appl 39:142–163. https://doi.org/10.1002/rra.2928
26. Haslam SM (1987) River Plants of Western Europe. Cambridge University Press, Cambridge
27. Haury J, Peltre M-C, Tremolieres M, Barbe J, Thiebaut G, Bernez I, Daniel H, Chatenet P, Haan-Archipof G, Muller S, Dutartre A, Laplace-Treyture C, Cazaubon A, Lambert-Servien E (2006) A new method to assess water trophy and organic pollution – the Macrophyte Biological Index for Rivers (IBMR): its application to different types of river and pollution. Hydrobiologia 570:153–158. https://doi.org/10.1007/978-1-4020-5390-0_22
28. Hering D, Borja A, Carstensen J, Carvalho L, Elliott M, Feld CK, Heiskanen A-S, Johnson RK, Moe J, Pont D, Solheim AL, van de Bund W (2010) The European Water Framework Directive at the age of 10: a critical review of the achievements with recommendations for the future. Sci Total Environ 408:4007–4019. https://doi.org/10.1016/j.scitotenv.2010.05.031
29. Hering D, Borja A, Jones IJ, Pont D, Boets P, Bouchez A, Bruce K, Drakare S, Hänfling B, Kahlert M, Leese F, Meissner K, Mergen P, Reyjol Y, Segurado P, Vogler A, Kelly M (2018) Implementation options for DNA-based identification into ecological status assessment under the European Water Framework Directive. Water Res 138:192–205. https://doi.org/10.1016/j.watres.2018.03.003
30. Hernandez-Suarez JS, Nejadhashemi AP (2018) A review of macroinvertebrate- and fish-based stream health modelling techniques. Ecohydrology 11:e2022. https://doi.org/10.1002/eco.2022
31. Krtolica I, Cvijanović D, Obradović Đ, Novković M, Milošević D, Savić D, Vojinović-Miloradov M, Radulović S (2021) Water quality and macrophytes in the Danube River: Artificial neural network modelling. Ecol Indic 121:107076. https://doi.org/10.1016/j.ecolind.2020.107076
32. Kubosova K, Brabec K, Jarkovsky J, Syrovatka V (2010) Selection of indicative taxa for river habitats: a case study on benthic macroinvertebrates using indicator species analysis and the random forest methods. Hydrobiologia 651:101–114. https://doi.org/10.1007/s10750-010-0280-1
33. Lemm JU, Venohr M, Globevnik L, Stefanidis K, Panagopoulos Y, van Gils J, Posthuma L, Kristensen P, Feld CK, Mahnkopf J, Hering D, Birk S (2021) Multiple stressors determine river ecological status at the European scale: towards an integrated understanding of river status deterioration. Glob Chang Biol 27:1962–1975. https://doi.org/10.1111/gcb.15504
34. Lewis KA, Rose KA, de Mutsert K, Sable S, Ainsworth C, Brady DC, Townsend H (2021) Using Multiple Ecological Models to Inform Environmental Decision-Making. Front Mar Sci 8:625790. https://doi.org/10.3389/fmars.2021.625790
35. Luo W, Zhu S, Wu S, Dai J (2019) Comparing artificial intelligence techniques for chlorophyll-a prediction in US lakes. Environ Sci Pollut Res 26:30524–30532. https://doi.org/10.1007/s11356-019-06360-y
36. Mata J, Salazar F, Barateiro J, Antunes A (2021) Validation of Machine Learning Models for Structural Dam Behaviour Interpretation and Prediction. Water 13:2717. https://doi.org/10.3390/w13192717

37. Munné A, Prat N, Solà C, Bonada N, Rieradevall M (2003) A simple field method for assessing the ecological quality of riparian habitat in rivers and streams: QBR index. Aquat Conserv: Mar Freshw Ecosyst 13:147–163. https://doi.org/10.1002/aqc.529
38. Murphy K, Efremov A, Davidson TA, Molina-Navarro E, Fidanza K, Betiol TCC, Chambers P, Grimaldo JT, Martins SM, Springuel I, Kennedy M, Mormul RP, Dibble E, Hofstra D, Lukács BA, Gebler D, Baastrup-Spohr L, Urrutia-Estrada J (2019) World distribution, diversity and endemism of aquatic macrophytes. Aquat Bot 158:103127. https://doi.org/10.1016/j.aquabot.2019.06.006
39. O'Hare MT, Aguiar FC, Asaeda T, Bakker ES, Chambers PA, Clayton JS, Elger A, Ferreira TM, Gross EM, Gunn IDW, Gurnell AM, Hellsten S, Hofstra DE, Li W, Mohr S, Puijalon S, Szoszkiewicz K, Willby NJ, Wood KA (2018) Plants in aquatic ecosystems: current trends and future directions. Hydrobiologia 812:1–11. https://doi.org/10.1007/s10750-017-3190-7
40. Olive DJ (2017) Multiple linear regression. In: Olive DJ (ed) Linear Regression. Springer, pp 17–83. https://doi.org/10.1007/978-3-319-55252-1_2
41. Özbay H, Yaprak AE, Turan N (2019) Assessing water quality in the Ceyhan River basin (Turkey) with the use of aquatic macrophytes. Chem Ecol 35:891–902. https://doi.org/10.1080/02757540.2019.1668928
42. Park Y-S, Lek S (2016) Artificial Neural Networks: Multilayer Perceptron for Ecological Modeling. Dev Environ Model 28:123–140. https://doi.org/10.1016/B978-0-444-63623-2.00007-4
43. Papastergiadou E, Stefanidis K, Dorflinger G, Giannouris E, Kostara K, Manolaki P (2016) Exploring biodiversity in riparian corridors of a Mediterranean island: Plant communities and environmental parameters in Cyprus rivers. Plant Biosyst 150:91–103. https://doi.org/10.1080/11263504.2014.941032
44. Poisot T, Stouffer DB, Kéfi S (2016) Describe, understand and predict: why do we need networks in ecology? Funct Ecol 30:1878–1882. https://doi.org/10.1111/1365-2435.12799
45. Polst BH, Hilt S, Stibor H, Hölker F, Allen J, Vijayaraj V, Kipferler N, Leflaive J, Gross EM, Schmitt-Jansen M (2022) Warming lowers critical thresholds for multiple stressor–induced shifts between aquatic primary producers. Sci Tot Environ 838:156511. https://doi.org/10.1016/j.scitotenv.2022.156511
46. Provata A, Sokolov IM, Spagnolo B (2008) Editorial: Ecological complex systems. Eur Phys J B 65:307–314. https://doi.org/10.1140/epjb/e2008-00380-9
47. Raven P, Holmes NTH, Dawson FH, Everard M (1998) Quality assessment using River Habitat Survey data. Aquat Conser: Mar Freshw Ecosyst 8:477–499. https://doi.org/10.1002/(SICI)1099-0755(199807/08)8:4<477::AID-AQC299>3.0.CO;2-K
48. Ren X, Mi Z, Georgopoulos P (2020) Comparison of Machine Learning and Land Use Regression for fine scale spatiotemporal estimation of ambient air pollution: Modeling ozone concentrations across the contiguous United States. Environ Int 142:105827. https://doi.org/10.1016/j.envint.2020.105827
49. Rocha JC, Peres CK, Buzzo JLL, de Souza V, Krause EA, Bispo PC, Frei F, Costa LSM, Branco CCZ (2017) Modeling the species richness and abundance of lotic macroalgae based on habitat

characteristics by artificial neural networks: a potentially useful tool for stream biomonitoring programs. J Appl Phycol 29:2145–2153. https://doi.org/10.1007/s10811-017-1107-5

50. Rodrigues C, Alves P, Bio A, Vieira C, Guimarães L, Pinheiro C, Vieira N (2019) Assessing the ecological status of small Mediterranean rivers using benthic macroinvertebrates and macrophytes as indicators. Environ Monit Assess 191:596. https://doi.org/10.1007/s10661-019-7766-8
51. Rolim SBA, Veettil BK, Vieiro AP, Kessler AB, Gonzatti C (2023) Remote sensing for mapping algal blooms in freshwater lakes: a review. Environ Sci Pollut Res 30:19602–19616. https://doi.org/10.1007/s11356-023-25230-2
52. Satish N, Anmala J, Varma MRR (2022) Prediction of stream water quality in Godavari River Basin, India using statistical and artificial neural network models. H2Open J 5:621–641. https://doi.org/10.2166/h2oj.2022.019
53. Schreiber SG, Schreiber S, Tanna RN, Roberts DR, Arciszewski TJ (2022) Statistical tools for water quality assessment and monitoring in river ecosystems – a scoping review and recommendations for data analysis. Water Qual Res J 57:40–57. https://doi.org/10.2166/wqrj.2022.028
54. StatSoft Inc (2017) STATISTICA (data analysis software system), version 10. www.statsoft.com
55. Stefanidis K, Dimitrellos G, Sarika M, Tsoukalas D, Papastergiadou E (2022) Ecological Quality Assessment of Greek Lowland Rivers with Aquatic Macrophytes in Compliance with the EU Water Framework Directive. Water 14:2771. https://doi.org/10.3390/w14182771
56. Stefanidis K, Oikonomou A, Papastergiadou E (2021) Responses of different facets of aquatic plant diversity along environmental gradients in Mediterranean streams: Results from rivers of Greece. J Environ Manage 296:113307. https://doi.org/10.1016/j.jenvman.2021.113307
57. Steiger J, Gurnell AM (2003) Spatial hydrogeomorphological influences on sediment and nutrient deposition in riparian zones: observations from the Garonne River, France. Geomorphology 49:1–23. https://doi.org/10.1016/S0169-555X(02)00144-7
58. Stella JC, Bendix J (2019) Chap. 5 - multiple stressors in riparian ecosystems. In: Sabater S, Elosegi A, Ludwig R (eds) Multiple Stressors in River Ecosystems. Elsevier, pp 81–110. https://doi.org/10.1016/B978-0-12-811713-2.00005-4
59. Szoszkiewicz K, Jusik S, Pietruczuk K, Gebler D (2020) The Macrophyte Index for Rivers (MIR) as an Advantageous Approach to Running Water Assessment in Local Geographical Conditions. Water 12:108. https://doi.org/10.3390/w12010108
60. Tiyasha, Tung TM, Yaseen ZM (2020) A survey on river water quality modelling using artificial intelligence models: 2000–2020. J Hydrol 585:124670. https://doi.org/10.1016/j.jhydrol.2020.124670
61. Zhang L, Liu J, Ho K (2015) Chap. 7 - Ecosystem risk assessment modelling method for emerging pollutants. In: Park Y-S, Lek S, Baehr C, Jørgensen SE (eds) Developments in Environmental Modelling, vol 27. Elsevier, pp 135–162. https://doi.org/10.1016/B978-0-444-63536-5.00006-5

Figures

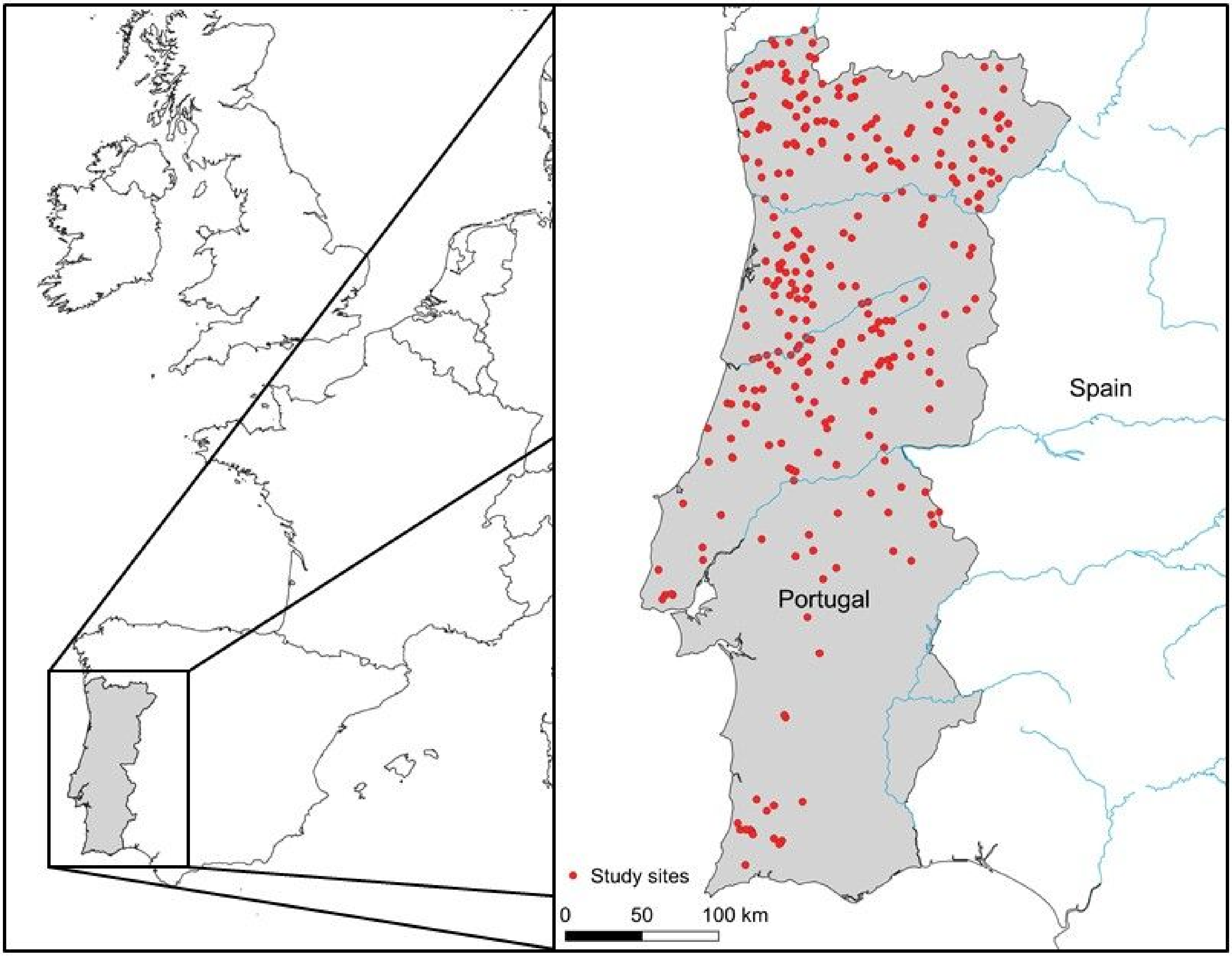

Figure 1

Map the distribution of sampling sites (n=292) in Portugal, SW Europe

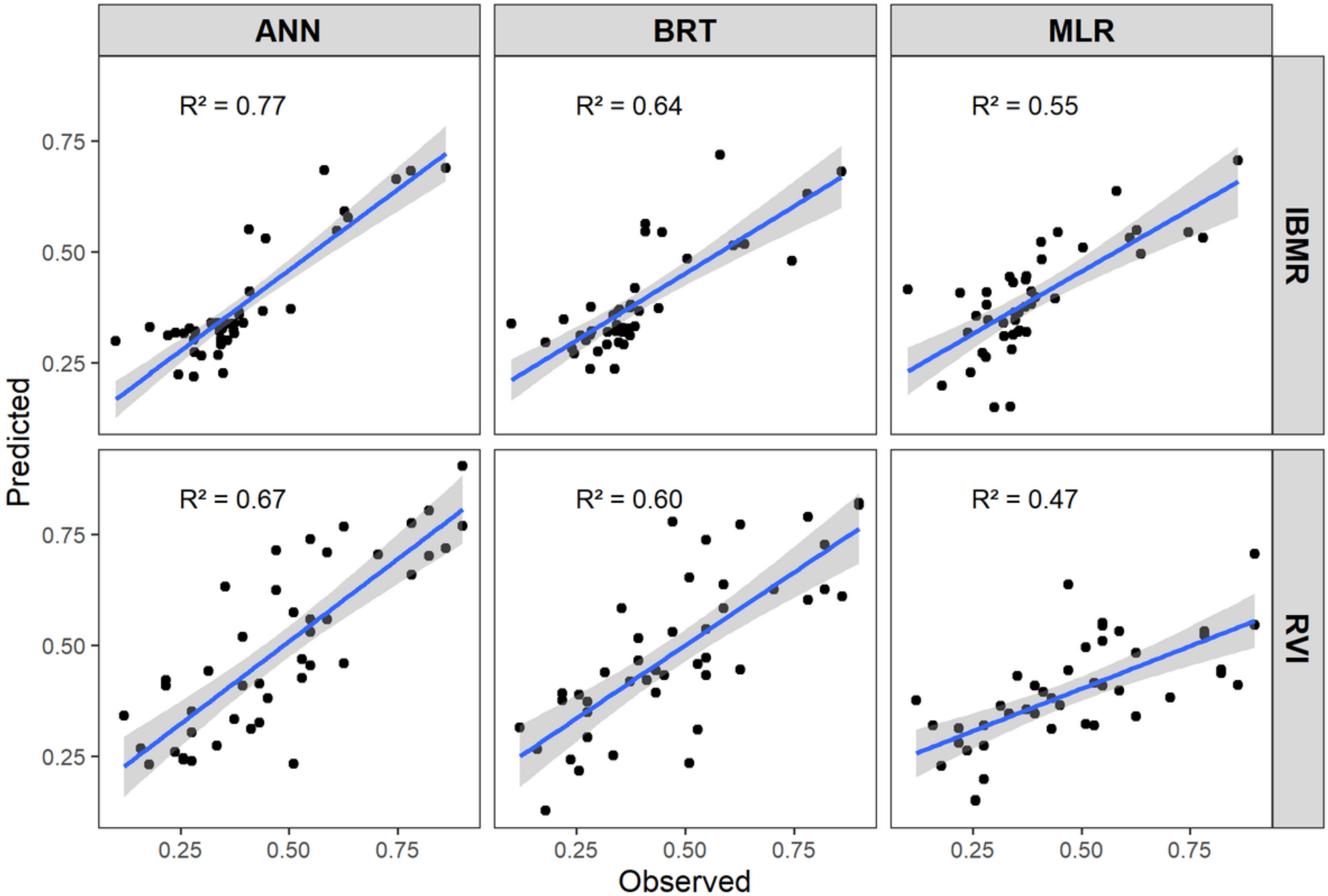

Figure 2

Values of IBMR and RVI (validation dataset) predicted and observed by different models

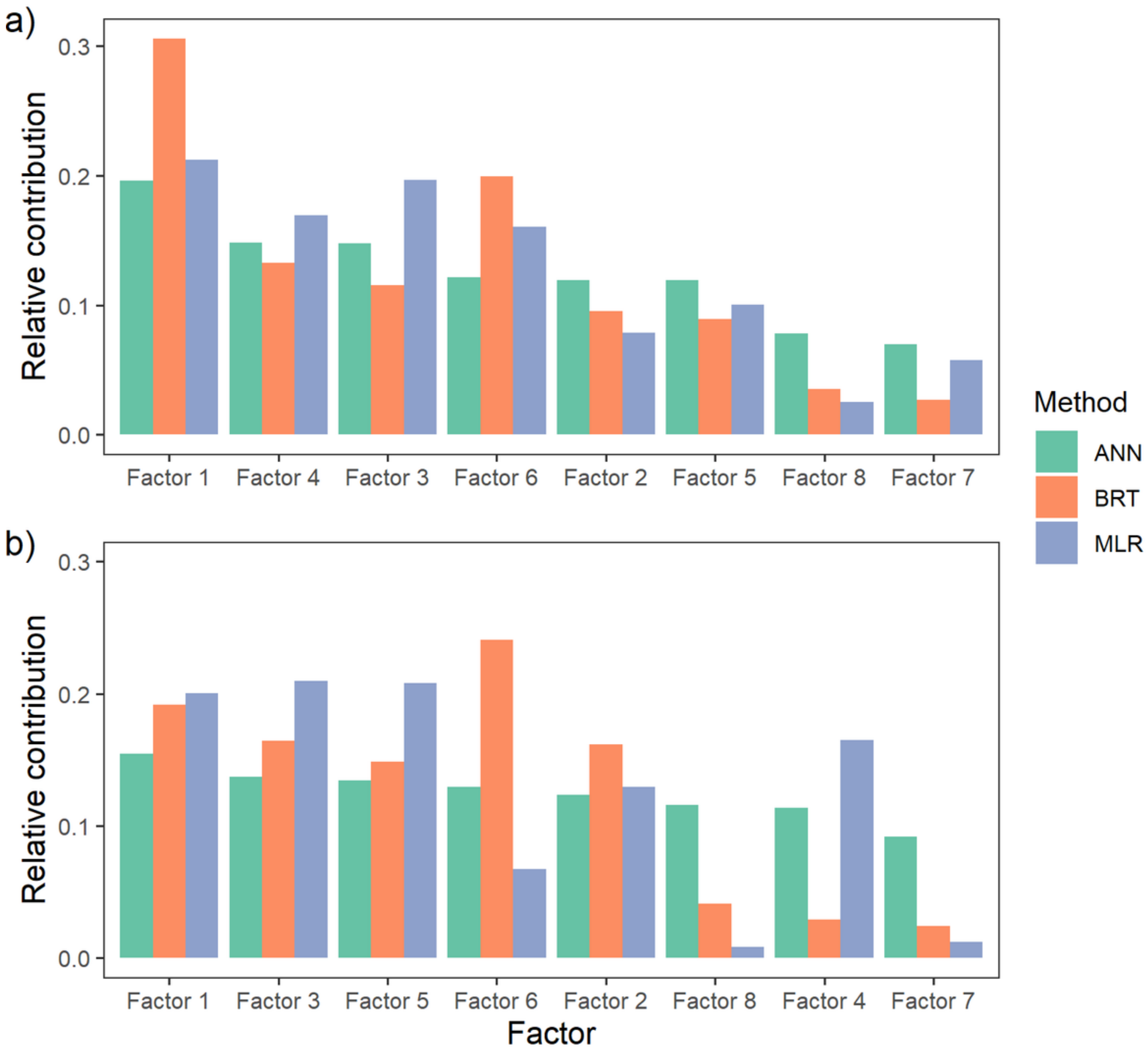

Figure 3

Relative contribution of input variables (PCA factors) in different IBMR (a) and RVI (b) models

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- AppendixGebleretalESPR.docx