Reach hydromorphology: a crucial environmental variable for the occurrence of riverine macrophytes

Willem Kaijser · Daniel Hering · Armin W. Lorenz

Received: 2 February 2022 / Revised: 9 August 2022 / Accepted: 9 August 2022 / Published online: 3 September 2022

© The Author(s) 2022

Abstract Management measures targeting riverine macrophyte species often address dissolved nitrogen and phosphorus compounds. With recent reductions of nutrient pollution, effects of other variables have become more apparent. This study aims to shed light on the hierarchy of variables determining riverine macrophyte distribution, covering both surface water nutrients and a range of other variables, in a Central European landscape. We used a dataset of 1,371 monitoring samples to predict the presence and absence of 30 common macrophyte species with multiple random forest models. Besides surface water chemical variables (P-PO₄, TN and pH), five reach hydromorphological variables (substrate, flow velocity, shading, width and depth) and a variable related to river size (distance from the river source) were used as predictors. The most important variables for the prediction of the species’ presence and absence were substrate, distance from the river source and flow velocity. Additionally, P-PO₄ and shade were also important, whilst TN, pH, width and depth played only a minor role. River hydromorphology was more important than surface water chemistry for the occurrences of individual riverine macrophyte species, in particular under less extreme eutrophic conditions. Species occurrences were related to multiple interacting variables, amongst which hydromorphology often prevailed. Presence/absence of species and composition of assemblages cannot be attributed to surface water nutrients alone, even when considering river typologies. Therefore, the results support the conjecture that a broader approach considering a range of stressors is necessary to design restoration measures for management and enhancement of macrophyte diversity.

Keywords Random Forest · Hydromorphology · Nutrients · Macrophytes · Rivers

Introduction

Macrophytes are a major component and structural element of surface water bodies (Gurnell et al., 2016). From a management perspective, macrophyte assemblages are often used to predict the trophic state, in particular dissolved nitrogen and phosphorus...
compounds (e.g. Bucior et al., 2021; Poikane et al., 2021; Weekes et al., 2021). In lakes, increased surface water nutrients shift macrophyte assemblages from meadow forming species (isoetids or Characeae) to canopy forming species (Myriophyllum spicatum L. or Ceratophyllum demersum L.) and eventually to the dominance of phytoplankton (van Zuidam & Peeters, 2013; Verhofstad et al., 2017; Poikane et al., 2018). Rivers also suffer from excessive nutrient inputs that modify macrophyte assemblages through competition with algae and a multitude of indirect pathways (Hilton et al., 2006; Novak & Chambers, 2014; Kaijser et al., 2021). This is also reflected by the development of a multitude of trophic indices (e.g. Holmes et al., 1999; Schneider and Melzer, 2003; Szoszkiewicz et al., 2019).

Given the large number of existing indices that use macrophytes to predict trophic state (Demars et al., 2012), macrophytes appear to many practitioners (particularly in Europe) to reflect mainly nutrient concentrations, although there is rising evidence of marginal correlations (Bucior et al., 2021; Weekes et al., 2021). However, a range of other environmental variables (e.g. flow velocity, depth and habitat diversity), besides surface water nutrients shape the habitat, potentially determine the presence or absence of macrophyte species, and are relevant for management. According to Riis & Biggs (2003), hydrological disturbance, such as frequency of floods, can reduce diversity and macrophyte cover of free-floating and fragile species (e.g. of Lemna minor L. or Elodea canadensis Michx.). Flow velocity also affects substrate stability, where vascular macrophytes prefer fine substrates that accumulate under low flow conditions, whilst mosses prefer coarse positionally stable substrate, particularly in rivers with higher flow velocities (Biggs, 1996; Downes et al., 2003; Riis & Biggs, 2003). Substrate heterogeneity within a reach reflects the diversity of hydraulic conditions and affects macrophyte diversity (Baattrup-Pedersen & Riis, 1999; Lorenz et al., 2011; Janauer et al., 2018).

Species are often adapted to specific substrates and current velocities and unable to tolerate strongly deviating conditions (Riis et al., 2001; Riis & Biggs, 2003; O’Brien et al., 2017).

Current velocities also affect bank morphology and vice versa. Closer to the banks and at point bars, flow velocity and erosion are often low, supporting the occurrence of lentic macrophyte species (Kemp et al., 1999). Banks with shallow slopes create littoral zones of reduced current, and many shorter macrophytes prefer to grow at these locations (Riis et al., 2001; Lorenz et al., 2011; O’Brien et al., 2017). Woody riparian vegetation affects individual macrophytes species by shading and reduces overall macrophyte biomass (Köhler et al., 2010). Whilst mosses tolerate shaded conditions, vascular plants preferably grow in the absence of shade. Light limitation can also suppress the effects of elevated nutrient concentrations (Canfield Jr. & Hoyer, 1988; Köhler et al., 2010).

Whilst these general effects of individual hydro-morphological variables are well understood (Biggs, 1996; Riis & Biggs, 2003) and described for between-species variation via multivariate statistics (Tremp, 2007; Janauer et al., 2018), an exploratory quantification of the relevance for individual species is missing. In particular, it is not yet possible to quantify the relevance of hydromorphological variables against surface water nutrients, which are still widely considered by management as the prime variable affecting macrophyte distribution and thereby the ecological status of a waterbodies (Poikane et al., 2021). At the same time, degradation of hydromorphology is amongst the prime stressors affecting aquatic assemblages (Reid et al., 2019; Birk et al., 2020; Lemm et al., 2021) but rarely considered as a priority stressor for macrophytes.

Here we use a dataset of macrophyte samples taken in mountainous and lowland rivers in North Rhine-Westphalia (Germany), which encompass hydromorphological and surface water chemical variables at each sample location. This dataset provides the opportunity to predict species presence–absence and explore the relative importance of the different variables. For each individual species, we addressed the questions: (1) Which environmental variables best predict their occurrence? (2) How do the individual variables, in particular hydromorphological and nutrient components, relate to each other in determining the individual species’ distribution? We achieved a quantification of relative variable importance. This is a prerequisite for the targeted planning of restoration measures, as it quantifies how strongly individual species respond to hydromorphological (or physical and chemical) modifications.
Methods

Dataset

We used a dataset of macrophytes, water chemistry and hydromorphology obtained from monitoring programmes of streams and rivers in North Rhine-Westphalia (Germany) for the requirements of the Water Framework Directive (WFD). This dataset was gathered by the State Agency for Nature, Environment and Consumer Protection of North Rhine-Westphalia (LANUV) between 2010 and 2019. The dataset covers sites ranging from mountainous streams to lowland rivers. North Rhine-Westphalia includes both densely populated urban and large rural areas. Whilst most of the lowland regions are intensively used for agriculture, the mountainous regions in the South and East are predominantly covered by deciduous and non-native coniferous forests.

Macrophytes were sampled between May and October according to the standards described by Schaumburg et al. (2004). Along a representative stream reach of 100 m, all macrophyte species were recorded on species level. Abundance was measured with the ordinal Kohler scale ranging from 1 (lowest) to 5 (highest). However, we transferred the abundance data to presence/absence data for our analysis, as pre-analyses raised questions about the reproducibility of the abundance estimations. For almost all species, the abundance reflected by the Kohler scale was not mirrored by increasing P-PO₄ concentrations (Fig. S1). One would expect an increase in the abundance (higher Kohler scale) of, e.g., *Stuckenia pectinata* (L.) Börner or *L. minor* along the P-PO₄ gradient and a decrease in abundance for *Fontinalis antipyretica* Hedw. (Fig. S2A), and similar for other variables. Hence, it is easier to discriminate where a species would be more likely present/absent along a gradient because they are clearly positioned differently (Fig. S2B).

A reach hydromorphology protocol was available for all sampling sites, including the percentage of eight substrate types, shading, depth, width and flow velocity categories (Table 1). We converted the fractions of different substrate types to a single number reflecting substrate coarseness within the sample reach. For this, we assigned numbers to each of the substrate types and multiplied this number with the fraction of the respective substrate type recorded at the site. The results of the individual multiplications were added up and divided by the number of classes (see the example in Table S1). The resulting substrate coarseness can range from 0 and 1, with 0 indicating the dominance of fine organic material, whereas 1...
indicates the dominance of large rocks, boulders and bedrock. The distance of the river source to each sample site was included as continuous variable in kilometres (km) (e.g. a sample site can be located close to the river source, such as 5 km, or far away from the source, such as 50 km). Both shade and velocity comprised five categories, width comprised eight and depth three categories. The shading categories were as follows: no shade = in full sun during daytime (no riparian vegetation or other structures), slightly shaded = only shaded during sunrise and sunset, half shaded = shaded in the morning and afternoon, but in full sun during the day, largely shaded = more than half the day in the shade, fully shaded = completely shaded the entire day. For flow velocity, width and depth, the maxima observed within a 100 m reach were recorded.

Water samples were taken four times per year and analysed for phosphate (P-PO₄), total nitrogen (TN) and pH. In some cases, water and macrophytes samples were taken at different dates. To account for these inconsistencies, we calculated the median of the chemical variables recorded one month before until one month after a macrophyte sample. This median can consist of one to three observations.

From the macrophyte samples, only the most recent samples per site were selected and duplicates were excluded. Species occurring less than 50 times in the dataset were removed, under the arbitrary assumption that 50 observations are needed to highlight patterns. This procedure resulted in a dataset with 30 species, 1,371 unique sample locations and 4,453 individual species observations.

Random forest models

We applied Random Forest (RF) models with the randomForest package for R (Liaw & Wiener, 2002) to examine, which variables have the highest relative importance for the prediction of presence and absence of each species. The relative importance was expressed as decrease in GINI importance of each variable, in relation to the sum of decrease in GINI for all variables. It is important not to confuse a high/low relative importance with causality or effect. The relative importance is a way to describe the relative contribution of the loss function for each variable in the model.

We controlled the models based on the accuracy calculated with the caret package (Kuhn, 2020) and Cohen’s κ, calculated with rel package (LoMartire, 2020) from the Out-Of-Bag (OOB) predictions. The accuracy is the fraction of the presences or absences of a species that are correctly classified. Additionally to the accuracy, we used Cohen’s κ, which takes the hypothetical probability of chance agreement into account, i.e. how far the model deviates from random guessing. The ranges of Cohen’s κ can be interpret as follows: < 0 = poor agreement, 0–0.19 = slight agreement, 0.2–0.39 = fair agreement, 0.4–0.59 = moderate agreement, 0.6–0.79 = substantial agreement and > 0.8 = almost perfect agreement (Landis & Koch, 1977).

The number of random bootstrapped predictor variables at each node (mtry) of the model was set to default (square root of the number of variables) and the number of generated trees created was set to 1,000. The lowest amount of observation of either presences or absences was used to under sample the majority class used in the models, so they occur in balanced proportions in each generated tree. The nodesize was set to 20% of the minority class variable, indicating how many observations can end up in the terminal node, thus regulating the depth of each tree. If predictor variables were missing, the median of all other predictor values was imputed. The optimal combination of predictor variables with the highest value for Cohen’s κ was sought by recursive elimination.

Nonmetric multi-dimensional scaling (nMDS)

To visualize the results of the RF models, we calculated a dissimilarity matrix between species based on the relative importance of the variables created by the RF models. These dissimilarities were displayed in an nMDS biplot. RF algorithms are non-linear and do not return a direction coefficient. However, as signs (plus, minus) were required for the calculation of the dissimilarity matrix, we used the fact that predicted occurrences of the species along a gradient were positively or negatively monotonous. We created partial dependency plots (PDPs) for every species and each individual variable with the pdp package for R (Greenwell, 2017). The PDPs show the voting fraction of the models along a single gradient. These PDPs are created by keeping all predictor
variables constant, whilst changing only the selected variable. The voting fraction is the average number of trees predicting a species to be present. Based on these PDPs, we calculated if this was a positive or negative monotonous correlation with Spearman’s ρ. A positive response was defined if Spearman’s ρ was > 0 (+1), and a negative response if Spearman’s ρ was < 0 (−1). If a response was negative for a particular variable, the relative importance of this variable was set negative (e.g. −23), if positive, the relative importance was set as positive (e.g. 23). If a specific variable was not present in the table, the relative importance was set to zero (0). Based on these values, we created a dissimilarity matrix with Euclidean distances. The corresponding vectors were introduced as the positive or negative relative importance. The nMDS biplot with corresponding vectors shows the dissimilarity between species based on positive and negative relative importance from the RF models. The vectors’ direction represent if the prediction of a species is positive or negative in relation to the corresponding variable. The nMDS and vector calculation was performed with the vegan package for R (Oksanen et al., 2017).

Manipulation of variable importance

As the GINI importance can be biased (Strobl et al., 2007) the RF models were applied under multiple settings to estimate the bias’ magnitude. The RF models were applied (1) on the total dataset with all variables, (2) on the total dataset with only hydromorphological variables and (3) on the total dataset with only the chemical variables (P-PO₄, pH and TN). This process was performed with and without recursive elimination. For each step, the average OOB accuracy and Cohen’s κ with minimum and maximum were recorded. Additionally, the dataset was split into mountainous (siliceous and calcareous streams) and lowland streams and the same process was repeated. The stream type separation had little impact in the dataset, thus, we decided to combine the types. Additionally, this preserved a sufficient number of observations per species (>50). The relative importance of the total dataset with recursive elimination is displayed in the main body and the OOB accuracy and Cohen’s κ of other setting are given in the supplementary material. If the GINI importance would introduce extreme bias towards specific variables, the models using the sub datasets of the hydromorphological or chemical variables should reflect the contribution of these variables on OOB accuracy and Cohen’s κ.

Moreover, we added a “random variable” between 0 and 1 to the total dataset. This was also performed by incorporating a constant variable only consisting of the integer 1. The consideration of the random and constant variable also indicates how the GINI feature importance ranks the overall variables. If the random or constant variable have extremely high variable importance, this would be cause for concern. All calculations were performed in R (R Core Team, 2021) and plotted with ggplot2 (Wickham, 2009).

Results

In Fig. 1, the results of each individual species model are given. Each row contains the relative importance of each species expressed as a value between 0 and 100%. The last two rows contain two ways of calculating the average, based on which the predictor variables were ranked from highest to lowest relative importance. Substrate coarseness was the predictor with on average the highest relative importance (28%) followed by distance from the river source (15%), flow velocity (14%) and shade (12%). On average, reach hydromorphological variables had a higher relative importance in the models than surface water chemistry variables represented by P-PO₄ (12%), pH (10%) and TN (9%). River size-related variables (width and depth) were only important for a few species. However, the relative importance of the predictors is species specific, i.e. for 22 species substrate coarseness was the most important variable but for, e.g. Nuphar lutea (Smith.) flow velocity was the crucial variable (22%). The average OOB model accuracy was 74% (min = 64% and max = 87%) and Cohen’s κ 0.25 (min = 0.08 and max = 0.52). For the results of full models without recursive elimination see Fig. S3.

The nMDS biplot, which is based on the relative importance of the variables created by the RF models, shows a clear separation of the species along the first and second axes (Fig. 2). The mosses [Amblystegium fluviatile (Hedw.) Schimp., Chiloscyphus polyanthos (L.) Corda, F. antipyretica and Platyhypnidium riparioides (Hedw.) Dixon] aggregate on the far left of the x-axis, which mainly represents increasing
substrate coarseness and high flow velocity. Most vascular macrophytes are spread along the second axis. The second axis mainly expresses the prediction of species according to increasing distance from the river source, width and depth. Especially *C. demersum*, *Elodea nuttallii* (Planch.) H. St. John, *M. spicatum*, *N. lutea*, *Sagittaria sagittifolia* L., *Spirigora polypyrhiza* (L.) Schleid. and *S. pectinata* were predicted further downstream or/and in wider and deeper rivers.

Contrastingly, *Berula erecta* (Huds.) Coville, *Callitriche platycarpa* Kutz., *Glyceria fluitans* (L.) R.Br. and *Nasturtium officinale* W.T. Aiton were predicted closer to the river source.

Shading affected relative importance of mosses positively, whereas fast growing macrophytes were negatively affected (e.g. *E. nuttallii*, *N. lutea*, *S. sagittifolia*, *Sparganium emersum* Rehmann). For P-PO$_4$ we observed the opposite pattern, and predictions for most species were positive, with exception of the mosses and *Ranunculus peltatus* Schrank. The chemical variable pH only contributed to a minor degree to the predictions and the mosses were predicted more often under higher pH, whereas vascular plants, such as *Callitriche obtusangula* Le Gall, and *Lemna gibba* L. were predicted under lower pH. *Berula erecta*, *C. platycarpa*, *G. fluitans* and *N. officinale* L. were predicted under lower P-PO$_4$. For example, Fig. 3
Fig. 2 nMDS biplot based the relative importance of the variables created by the RF models for the individual macrophyte species. The vectors represent the environmental variables and their lengths the relative importance. Abbreviations consist of the first letter of the genus and first four letters of the species name (Table S2)

displays the contrasting predictions for three species (C. polyanthos, C. platycarpa and Potamogeton crispus L.) in relation to substrate coarseness, distance from the river source and flow velocity (PDPs for all other species can be found in Figs. S4–S6).

We repeated the modelling approach only on hydromorphological and on chemical variables separately, with and without recursive elimination. This was also applied on only lowland and mountainous streams to see how the average OOB accuracy and Cohen’s $\kappa$ changes (Table S3). Removing the chemical variables resulted only in a minor decline of the average OOB accuracy and Cohen’s $\kappa$, whilst removing hydromorphological variables caused a larger drop in the OOB accuracy and Cohen’s $\kappa$. However, both hydromorphological and chemical variables were required to obtain the maximum OOB accuracy and Cohen’s $\kappa$. Additionally, we repeated the same procedure for the total dataset with recursive elimination, when incorporating a random variable between 0 and 1 and with 1 as additional variable (Figs. S7, S8). However, in the first approach there was an inflation of the relative importance of the random variable for some species.

Discussion

The models show that reach scale hydromorphological variables were most determinant for the prediction of riverine macrophyte species. Surface water chemical variables had only subordinate importance. However, predictions were species specific.

The mosses were predicted to occur more often under conditions of high substrate coarseness and high flow velocity. Compared to fine substrate, coarser substrates are more stable and less frequently dislocated under high flow velocities. Mosses like F. antipyretica attach themselves with rhizoids on coarse substrate, thus tolerating high flow velocities (Biggs, 1996; Downes et al., 2003; Lang & Murphy, 2012). On the contrary, species such as L. minor, S. polyrhiza and C. demersum are free-floating species, swept away by even moderate flow velocity and experience higher hydraulic stress (Butcher, 1933; Schutten & Davy, 2000). Consequently, these species were predicted to occur frequently when flow velocity and substrate coarseness were low, where sedimentation potentially occurs (Butcher, 1933; Madsen et al., 2001). The association of free-floating species to substrate coarseness is indirect, reflecting lower flow
velocities and the consequential changes in sediment type. *Nuphar lutea* has a well-developed root system and can grow on a multitude of substrates. Yet, the leaves of this species experience a large amount of drag and can be easily damaged (Puijalon et al., 2011). Consequently, this species was also predicted to occur under lower flow velocities and finer substrate. For other species (*M. spicatum*) flow velocity was not selected. *Myriophyllum spicatum* roots deep in the substrate, has streamlined leaves and its shoots are relatively flexible; in combination, these traits well explain the predicted occurrence patterns (Schutten & Davy, 2000; Willby et al., 2000; Bal et al., 2011). Unexpectedly, *S. emersum* L. was predicted to occur more often under lower flow velocities, even though it displays lower hydraulic resistance due to its streamlined growth form and strong fixation in the substrate by an the extensive root system (O’Briain et al., 2017). These results may be caused in the macro-scale sampling procedure that complicates the detection of interferences, since flow velocity and substrate are heterogenic allies distributed within a reach, whilst a micro-scale sampling procedure might have resulted in better model performance (Demars & Harper, 1998; Kemp et al., 1999; Tremp, 2007). Hence, there is an intricate connection between

Fig. 3 Partial dependency plots for the prediction of *C. polyanthos*, *C. platycarpa* and *P. crispus* in relation to substrate coarseness (A), distance from source (B) and flow velocity (C)
species occurrence and the combined relation of flow velocity and substrate characteristics (Butcher, 1933; Madsen et al., 2001; Janauer et al., 2018).

Distance from the river source was another important predictor. Although here regarded as a hydro-morphological variable, it is better represented as a spatial variable. Several species had their main distribution in downstream sections (e.g. C. demersum, E. nuttallii, M. spicatum and N. lutea), whereas others were predicted closer to the river source (e.g. B. erecta, C. platycarpa, G. fluitans and N. officinale). Downstream sections reflect, in general, deeper cross-sections, constant flow velocity and less variable hydrological regimes, whilst upstream sections have a relatively shallow but more variable depth. Riis et al. (2001) discussed that the abundance of C. platycarpa declined with stream size, in line with our results. Callitriche species are fast colonizers often quickly re-occurring after disturbance. Also corresponding to our results, Riis et al. (2001) found that the abundance of S. emersum and large Potamogeton species (i.e. P. crispus and S. pectinata) increased with stream size.

Additionally, colonization success from upstream sections could increase with distance from the river source, if shallow, slow flowing parts or obstructions are present to retain fragments or seeds (Riis, 2008). This can also reflect the results of fragmentation (Bolgagni et al., 2016), whereby seedbanks of whole river sections are altered. Modifications of the hydrological regime could result in altered distribution patterns, considering dispersal is likely to take place (Cellot et al., 1998). The specific quantitative contribution of connectivity to macrophyte occurrences still remains uncertain.

The combination of distance from the river source, width and depth can represent a combination of natural and anthropogenic morphological characteristics (Riis et al., 2001; Janauer et al., 2018). With distance from the river source banks may get steeper due to channelization, losing the habitats that can be potentially occupied by shorter and emergent species (Lorenz et al., 2011; Dorotovičová, 2013). Closer to the river source, smaller amphibious species, such as B. erecta, Veronica beccabunga L. or N. officinale benefit from shallow and low bank slope (Lorenz et al., 2011; O’Briain et al., 2017; Janauer et al., 2018). Thus, widespread anthropogenic alteration of riverbanks could have driven these species to solely remain in upstream habitats.

Shading affected the prediction of several species, such as Elodea spp., S. emersum and S. pectinata. Canfield Jr. & Hoyer (1988) showed that biomass of macrophytes was strongly affected by openings in the canopy of riparian vegetation. Also, Köhler et al. (2010) observed that the biomass of productive species (C. demersum, M. spicatum, N. lutea, S. emersum, S. sagittifolia and S. pectinata) was negatively affected by shading of riparian vegetation. This suggests that the increase in shading can lead to the disappearance of highly productive and competitive species such as Elodea spp., S. emersum and S. pectinata. The chemical surface water variables P-PO₄, pH and TN had only low influence on the prediction of species occurrences (see also, Demars & Harper, 1998; Weekes et al., 2021). In the total dataset, P-PO₄ was on average more important than the other two variables. In general, the vascular plants were predicted to occur more frequently when P-PO₄ concentrations increased and mosses to occur less frequently. Higher concentrations of P-PO₄ may be correlated to depth and lower flow velocities and, in lakes, can lead to dominance of larger species (Verhofstad et al., 2017), but can also promote epiphytic algae growth preventing light reaching leaves of macrophytes (Köhler et al., 2010; Novak & Chambers, 2014). However, the range of P-PO₄ could be considered small (95% = 0.4 mg l⁻¹) although the values are relatively high according to some standards (Novak & Chambers, 2014; Poikane et al., 2021). If the range of this variable is small, this might result in a reduction of the relative importance. Yet, it is difficult to elude whether this is the case and exclude other variables from the equation in order to assess the direct independent impact of only surface water P-PO₄. Otherwise, P-PO₄ might also reflect nutrient concentration in the sediment. Carignan (1985) showed the strong nutrient uptake via the roots of M. spicatum and Barko et al. (1991) addressed this in a review of sediment interactions with macrophytes (compare also Butcher, 1933). Yet, information on sediment nutrients was absent from this dataset.

pH was also a predictor of species occurrence, but to a lesser degree. pH can reflect dissolved inorganic carbon (DIC) concentrations. At a pH of 8.4 all carbonic acid (H₂CO₃) has been transferred to bicarbonate (HCO₃⁻), whilst CO₂ concentration is low. This restricts the carbon source for species, which are only able to utilize CO₂ (Maberly et al., 2015; Svitok et al.,
Nitrogen compounds have been shown to correlate negatively with the number of macrophyte species (James et al., 2005) and relate to between-species variation (Leyssen et al., 2014). In our analysis, species such as *B. erecta* and *G. fluitans* were predicted to occur under higher TN concentrations. Smaller lowland streams are often located in agricultural areas, where the input of nitrogen compounds is high. Denitrification has likely not reduced concentrations over large spatial scales, DIC related variables (e.g. CO$_2$, HCO$_3^-$, GH, KH; all absent in this dataset) can lead to clearly observable difference (Iversen et al., 2019; Kajser et al., 2021) and correlation (Demars et al., 2012). The input of nutrients might positively relate to the depletion of CO$_2$ and an increase of HCO$_3^-$ by increasing the biomass of primary producers and altering biogeochemical pathways; in such a case, the relation of macrophyte assemblages to surface water nutrients is obvious, but might not be causal (Demars et al., 2012; Kajser et al., 2021).

Nitrogen compounds have been shown to correlate negatively with the number of macrophyte species (James et al., 2005) and relate to between-species variation (Leyssen et al., 2014). In our analysis, species such as *B. erecta* and *G. fluitans* were predicted to occur under higher TN concentrations. Smaller lowland streams are often located in agricultural areas, where the input of nitrogen compounds is high. Denitrification has likely not reduced concentration of NO$_3^-$, due to relative lower denitrification rates upstream as compared to downstream and due to high input of TN. Hence, NO$_3^-$ can decrease when anaerobic sediment leads to denitrification (García-Ruiz et al., 1998). As such, it is unclear if occurrence of species is causally linked to these variables. Under laboratory conditions excluding the effects of the other variables, however, *L. gibba* starts dominating over *C. demersum* under increasing nitrogen (TN) concentrations (Szabó et al., 2021). Yet, if TN is high, flow velocity within a river reach can still have a remarkable impact on macrophyte abundance and occurrence. In lakes, symptoms of eutrophication are therefore more likely observable than in rivers (Birk et al., 2020).

The underlying data source is subject to some discrepancies, such as (1) missing values for some variables that have been replaced (imputed) with the median, which can influence the OOB predictions (van Echelpoel & Goethals, 2018); (2) the use of GINI feature importance (Strobl et al., 2007); (3) recursively eliminating variables might bias the relative importance; (4) limited information on the spatial aspect of the sample locations within a the different river systems and (5) confounding variables, which need to be taken in consideration. The GINI feature importance favours continuous and cardinal variables with many categories over variables with few categories, inflating their relative importance; this effect can be observed in Figs. S7 and S8. Especially the relative importance of the random variable ranging between 0 and 1 (Fig. S7) related to the more common species (e.g. *N. officinale* and *Iris pseudacorus*) is inflated. Thus, when relations of predictors to the target variable become obscure, the relative importance of random variables tends to become inflated. However, the effects were only moderately affecting the overall results. In addition, predictor variables can be correlated affecting their relative importance and influencing interference from the models, which seems a general concern in spatial analyses of macrophytes (Demars & Harper, 1998; Demars et al., 2012). Despite these limitations, the position of the first five variables seems relatively stable regarding the different approaches (see Fig. 1; Figs. S7, S8). Moreover, regardless how the dataset was divided and whether recursive elimination was applied, hydromorphological variables always had a higher or similar OOB accuracy and Cohen’s $\kappa$ than the chemical variables. However, if hydromorphological and chemical variables were included, both contributed to the prediction as noted by Gebler et al. (2018). We used presence/absence data, omitting differences in quantitative community composition. However, with the Kohler scale abundance of the dataset we were unable to detect and model any patterns regarding abundance (Figs. S1, S2).

Our results support the conjecture that hydromorphological variables should always be regarded in addition to chemical variables when evaluating macrophyte occurrences, as hydromorphological variability contributes substantially. Even within stream types, the range of specific variables can still strongly relate to species occurrence. In North Rhine-Westphalia, the vast majority of streams and rivers has a degraded hydromorphology (LANUV, 2013), so hydromorphological variables should be much more strongly considered in restoration planning addressing aquatic macrophytes. The former focus on surface water nutrients in explaining macrophyte distribution seems not to be justified any longer, as nutrient gradients have declined in large parts of Central Europe in recent decades: extreme nutrient pollution has vanished, whilst nutrient-poor sites are still
a rare exception. Under these conditions, the contributions of other variables are getting more relevant. Hence, we should search for better ways to quantify spatial diversity within a river reach, as possible predictors of macrophyte community composition. Our results also outline the limitations of river assessment systems with macrophytes: most indices currently applied in Europe focus on surface water nutrient pollution (e.g. Schneider and Melzer, 2003; Haury et al., 2006; Poikane et al., 2021), which obviously impact macrophyte assemblages only to a minor degree. Restoration measures aiming at enhancing riverine macrophyte diversity need to include various hydro-morphological variables, such as increasing substrate and flow diversity, whilst further reductions of surface water nutrient concentrations may have limited effects.

Acknowledgements We are grateful to the public authorities of North Rhine-Westphalia (LANUV NRW) that provided the data for this study.

Author contributions WK: Writing—Original draft preparation, methodology design and software; AL: Data curation, supervision, writing—reviewing, methodology and editing; DH: Supervision, writing—reviewing and editing.

Funding Open Access funding enabled and organized by Projekt DEAL. The authors have not disclosed any funding.

Data availability The data is available from https://github.com/snwikaij/Data.

Declarations

Conflict of interest We declare no conflict of interest.

References

Baattrup-Pedersen, A. & T. Riis, 1999. Macrophyte diversity and composition in relation to substratum characteristics in regulated and unregulated Danish streams. Freshwater Biology 42: 375–385.

Bal, K. D., T. J. Bouma, K. Buis, E. Struyf, S. Jonas, H. Backx & P. Meire, 2011. Trade-off between drag reduction and light interception of macrophytes: comparing five aquatic plants with contrasting morphology: macrophytes and drag coefficients. Functional Ecology 25: 1197–1205.

Barko, J. W., D. Gunnison & S. R. Carpenter, 1991. Sediment interactions with submerged macrophyte growth and community dynamics. Aquatic Botany 41: 41–65.

Biggs, B. J. F., 1996. Hydraulic habitat of plants in streams. Regulated Rivers: Research and Management 12: 131–144.

Birk, S., D. Chapman, L. Carvalho, B. M. Spears, H. E. Andersen, C. Argillier, S. Auer, A. Baattrup-Pedersen, L. Banin, M. Beklioglu, E. Bondar-Kunze, A. Borja, P. Branco, T. Bucak, A. D. Buïsje, A. C. Cardoso, R.-M. Couture, F. Cremona, D. de Zwart, C. K. Feld, M. T. Ferreira, H. Feuchtmayr, M. O. Gessner, A. Gieswein, L. Globevnik, D. Graeber, W. Graf, C. Gutiérrez-Cánovas, J. Hanganu, U. Işık, M. Järvinen, E. Jeppesen, N. Kotamiki, M. Kuiper, J. U. Lemm, S. Lu, A. L. Solheim, U. Mischke, S. J. Moe, P. Nõges, T. Nõges, S. J. Ormerod, Y. Panagopoulos, G. Phillips, L. Posthuma, S. Pouso, C. Prudhomme, K. Rankinen, J. J. Rasmussen, J. Richardson, A. Sagouis, J. M. Santos, R. B. Schäfer, R. Schinegger, S. Schmutz, S. C. Schneider, L. Schütting, P. Segurado, K. Stefanidis, B. Sures, S. J. Thackeray, J. Turunen, M. C. Uyarra, M. Venohr, P. C. von der Ohe, N. Willby & D. Hering, 2020. Impacts of multiple stressors on freshwater biota across spatial scales and ecosystems. Nature Ecology and Evolution 4: 1060–1068.

Bolpagni, R., E. Racchetti & A. Laini, 2016. Fragmentation and groundwater supply as major drivers of algal and plant diversity and relative cover dynamics along a highly modified lowland river. Science of the Total Environment 568: 875–884.

Bucior, A., B. Rippey, Y. McElarney & R. Douglas, 2021. Evaluating macrophytes as indicators of anthropogenic pressures in rivers in Ireland. Hydrobiology 848: 1087–1099.

Butcher, R. W., 1933. Studies on the ecology of rivers: I. On the distribution of macrophytic vegetation in the rivers of Britain. The Journal of Ecology 21: 58.

Canfield, D. E., Jr. & M. V. Hoyer, 1988. Influence of nutrient enrichment and light availability on the abundance of aquatic macrophytes in Florida streams. Canadian Journal of Fisheries and Aquatic Sciences 45: 1467–1472.

Carignan, R., 1985. Nutrient dynamics in a littoral sediment colonized by the submerged macrophyte Myriophyllum spicatum. Canadian Journal of Fisheries and Aquatic Sciences 42: 1303–1311.

Cellot, B., F. Mouillot & C. P. Henry, 1998. Flood drift and propagule bank of aquatic macrophytes in a riverine wetland. Journal of Vegetation Science 9: 631–640.
Demars, B. O. L. & D. M. Harper, 1998. The aquatic macrophytes of an English lowland river system: assessing response to nutrient enrichment. Hydrobiologia 384: 75–88.

Demars, B. O. L., J. M. Potts, M. Trêmollières, G. Thièbaut, N. Gougelin & V. Nordmann, 2012. River macrophyte indices: not the Holy Grail! Freshwater Biology 57: 1745–1759.

Dorotovičová, Ľ., 2013. Man-made canals as a hotspot of aquatic macrophyte biodiversity in Slovakia. Limnologica 43: 277–287.

Downes, B. J., T. J. Entwisle & P. Reich, 2003. Effects of flow regulation on disturbance frequencies and in-channel bryophytes and macroalgae in some upland streams. River Research and Applications 19: 27–42.

García-Ruiz, R., S. N. Pattinson & B. A. Whitten, 1998. Denitrification in river sediments: relationship between process rate and properties of water and sediment. Freshwater Biology 39: 467–476.

Gebler, D., G. Wiegleb & K. Szoszkiewicz, 2018. Integrating river hydromorphology and water quality into ecological status modelling by artificial neural networks. Water Research 139: 395–405.

Greenwell, M., 2017. pdp: an R package for constructing partial dependence plots. The R Journal 9: 421–436.

Gurnell, A. M., D. Corenblit, D. García de Jalón, M. González del Tánago, R. C. Grabowski, M. T. O’Hare & M. Szewczyk, 2016. A conceptual model of vegetation–hydrogeomorphology interactions within river corridors. River Research and Applications 32: 142–163.

Haury, J., P. Chatenet, G. Haan-Archipof, S. Muller, A. Dutartre, C. Laplace-Treytoure, A. Cazaubon & E. Lambert-Servien, 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.

Hilton, J., M. O’Hare, M. J. Bowes & J. J. Jones, 2006. How green is my river? A new paradigm of eutrophication in rivers. Science of the Total Environment 365: 66–83.

Holmes, N. T. H., J. R. Newman, S. Chad, K. J. Rouen, L. Saint & F. H. Dawson, 1999. Mean Trophic Rank: A User’s Manual. Environment Agency.

Iversen, L. L., A. Winkel, L. Baastrup-Spohr, A. B. Hinke, M. Alahulta, A. Baattrup-Pedersen, S. Birk, P. Brodersen, P. A. Chambers, F. Ecke, T. Feldmann, D. Gebler, J. Heino, T. S. Jespersen, S. J. Moe, T. Riis, L. Sass, O. Vestergaard, S. C. Maberly, K. Sand-Jensen & O. Pedersen, 2019. Catchment properties and the photosynthetic trait composition of freshwater plant communities. Science 366: 878–881.

James, C., J. Fisher, V. Russell, S. Collings & B. Moss, 2005. Nitrate availability and hydrophtye species richness in shallow lakes. Freshwater Biology 50: 1049–1063.

Janauer, G. A., A. Gabersčik, J. Květ, M. Germ & N. Exler, 2018. Macrophytes of the River Danube Basin, Academia, Praha.

Kajiser, W., A. W. Lorenz, S. Birk & D. Hering, 2021. The interplay of nutrients, dissolved inorganic carbon and algae in determining macrophyte occurrences in rivers.

Science of the Total Environment. https://doi.org/10.1016/j.scitotenv.2021.146728.

Kemp, J. L., D. A. T. Harper & C. A. Giuseppe, 1999. Use of ‘functional habitats’ to link ecology with morphology and hydrology in river rehabilitation. Aquatic Conservation: Marine and Freshwater Ecosystems 9: 159–178.

Köhler, J., J. Hachol & S. Hilt, 2010. Regulation of submerged macrophyte biomass in a temperate lowland river: interactions between shading by bank vegetation, epiphyton and water turbidity. Aquatic Botany 92: 129–136.

Kuhn, M., 2020. CARET: Classification and Regression Training. R Package Version 6.0-85.

Landis, J. R. & G. G. Koch, 1977. The measurement of observer agreement for categorical data. Biometrics 33: 159.

Lang, P. & K. J. Murphy, 2012. Environmental drivers, life strategies and bioindicator capacity of bryophyte communities in high-latitude headwater streams. Hydrobiologia 679: 1–17.

LANUV, 2013. Gewässerstrukturen Nordrhein-Westfalen. Landesamt für Natur, Umwelt und Verbraucherschutz Nordrhein-Westfalen.

Lemm, J. U., M. Venohr, L. Globevnik, K. Stefanidis, Y. Panagopoulos, J. Gils, L. Posthuma, P. Kristensen, C. K. Feld, J. Mahnkopf, D. Hering & S. Birk, 2021. Multiple stressors determine river ecological status at the European scale: towards an integrated understanding of river status deterioration. Global Change Biology 27: 1962–1975.

Leyssen, A., L. Denys, A. Schneider & A. M. Mouton, 2014. Distribution and environmental requirements of stream habitat with Ranunculus fluitans and Callitricho-Batrachion vegetation in lower Belgium (Flanders). Aquatic Conservation: Marine and Freshwater Ecosystems 24: 601–622.

Liaw, A. & M. Wiener, 2002. Classification and regression by randomForest. R News 2: 18–22.

LoMartire, R., 2020. rel: Reliability Coefficients. R Package Version 1.4.2.

Lorenz, A. W., T. Korte, A. Sundermann, K. Januschke & P. Haase, 2011. Macrophytes respond to reach-scale river restorations. Journal of Applied Ecology. https://doi.org/10.1111/j.1365-2664.2011.02082.x.

Maberly, S. C., S. A. Berthelot, A. W. Stott & B. Gontero, 2015. Adaptation by macrophytes to inorganic carbon down a river with naturally variable concentrations of CO2. Journal of Plant Physiology 172: 120–127.

Madsen, J. D., P. A. Chambers, W. F. James, E. W. Koch & D. F. Westlake, 2001. The interaction between water movement, sediment dynamics and submerged macrophytes. Hydrobiologia 444: 71–84.

Novak, P. A. & J. M. Chambers, 2014. Investigation of nutrient thresholds to guide restoration and management of two impounded rivers in south-western Australia. Ecological Engineering 68: 116–123.

O’ bribery, R., S. Shepard & B. Coghlan, 2017. Pioneer macrophyte species engineer fine-scale physical heterogeneity in a shallow lowland river. Ecological Engineering 102: 451–458.

Oksanen, J., F. B. Guillaume, M. Friendly, R. Kindt, P. Legendre, D. McGlinn, P. R. Minchin, R. B. O’Hara, G. L.
Simpson, P. Solymos, M. H. H. Stevens, E. Szöcs & H. Wagner, 2017. Vegan: Community Ecology Package. R Package Version 2.4-5.

Poikane, S., R. Portièlje, L. Denys, D. Elferts, M. Kelly, A. Kolada, H. Mäemets, G. Phillips, M. Sondergaard, N. Willby & M. S. van den Berg, 2018. Macrophyte assessment in European lakes: diverse approaches but convergent views of ‘good’ ecological status. Ecological Indicators 94: 185–197.

Poikane, S., G. Várbiró, M. G. Kelly, S. Birk & G. Phillips, 2021. Estimating river nutrient concentrations consistent with good ecological condition: more stringent nutrient thresholds needed. Ecological Indicators 121: 107017.

Puijalon, S., T. J. Bouma, C. J. Douady, J. van Groenendael, N. P. R. Anten, E. Martel & G. Bornette, 2011. Plant resistance to mechanical stress: evidence of an avoidance-tolerance trade-off. New Phytologist 191: 1141–1149.

R Core Team, 2021. R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna:

Reid, A. J., A. K. Carlson, I. F. Creed, E. J. Eliason, P. A. Gell, P. T. J. Johnson, K. A. Kidd, T. J. MacCormack, J. D. Olden, S. J. Ormerod, J. P. Smol, W. W. Taylor, K. Tockner, J. C. Vermaire, D. Dudgeon & S. J. Cooke, 2019. Emerging threats and persistent conservation challenges for freshwater biodiversity. Biological Reviews 94: 849–873.

Riis, T., 2008. Dispersal and colonisation of plants in lowland streams: success rates and bottlenecks. Hydrobiologia 596: 341–351.

Riis, T. & B. J. F. Biggs, 2003. Hydrologic and hydraulic control of macrophyte establishment and performance in streams. Limnology and Oceanography 48: 1488–1497.

Riis, T., K. Sand-Jensen & S. E. Larsen, 2001. Plant distribution and abundance in relation to physical conditions and location within Danish stream systems. Hydrobiologia 448: 217–228.

Schaumburg, J., C. Schranz, G. Hofmann, D. Stelzer, S. Schneider & U. Schmedtje, 2004. Macrophytes and phytothons as indicators of ecological status in German lakes – a contribution to the implementation of the Water Framework Directive. Limnologica 34: 302–314.

Schneider, S. & A. Melzer, 2003. The trophic index of macrophytes (TIM) – a new tool for indicating the trophic state of running waters. International Review of Hydrobiology 88: 49–67.

Schuten, J. & A. J. Davy, 2000. Predicting the hydraulic forces on submerged macrophytes from current velocity, biomass and morphology. Oecologia 123: 445–452.

Strobl, C., A.-L. Boulesteix, A. Zeileis & T. Hothorn, 2007. Bias in random forest variable importance measures: illustrations, sources and a solution. BMC Bioinformatics 8: 25.

Svitok, M., R. Hrivnák, J. Kochjarová, H. Otahefová & P. Pařove-Balog, 2016. Environmental thresholds and predictors of macrophyte species richness in aquatic habitats in central Europe. Folia Geobotanica 51: 227–238.

Szabó, S., G. Koleszár, M. Braun, Z. Nagy, T. T. Viczi & E. T. H. M. Peeters, 2021. Submerged rootless macrophytes sustain a stable state against free-floating plants. Ecosystmes. https://doi.org/10.1007/s10021-021-00637-5.

Szoszkiewicz, K., S. Jusik, K. Pietruckzuk & D. Gebl, 2019. The macrophyte index for rivers (MIR) as an advantageous approach to running water assessment in local geographical conditions. Water 12: 108.

Tremp, H., 2007. Spatial and environmental effects on hydrophytic macrophyte occurrence in the Upper Rhine floodplain (Germany). Hydrobiologia 586: 167–177.

van Echelpoel, W. & P. L. M. Goethals, 2018. Variable importance for sustaining macrophyte presence via random forests: data imputation and model settings. Scientific Reports 8: 14557.

van Zuidam, J. P. & E. T. Peeters, 2013. Occurrence of macrophyte monocultures in drainage ditches relates to phosphorus in both sediment and water. SpringerPlus 2: 564.

Vorhofstad, M. J. J. M., M. M. Alirangues Núñez, E. P. Reichman, E. van Donk, L. P. M. Lamers & E. S. Bakker, 2017. Mass development of monospecific submerged macrophyte vegetation after the restoration of shallow lakes: roles of light, sediment nutrient levels, and propagule density. Aquatic Botany 141: 29–38.

Weekes, L., Ú. Fitzpatrick & M. Kelly-Quinn, 2021. Assessment of the efficiency of river macrophytes to detect water-column nutrient levels and other environmental conditions in Irish rivers. Hydrobiologia 848: 2797–2814.

Wickham, H., 2009. ggplot2: Elegant Graphics for Data Analysis, Springer, New York:

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