

Same species, same habitat preferences? The distribution of aquatic plants is not explained by the same predictors in lakes and streams

Morgane B. Gillard^{1*}, Jukka Aroviita², Janne Alahuhta^{2,3}

* Corresponding author

morgane.gillard35@gmail.com

1 USDA-ARS Invasive Species and Pollinator Health Research Unit, Department of Plant Sciences, University of California, Davis, CA 95616, USA

2 Finnish Environment Institute, Freshwater Centre, Oulu, Finland

3 Geography Research Unit, University of Oulu, PO Box 3000, FI-90014 University of Oulu, Finland

Keywords : boreal water bodies, freshwater ecosystems, macrophytes, rivers, species distribution models

Abstract

1. Studying the geographical distribution of species can reveal conditions and processes that may drive species presence and abundance. Organism distribution has frequently been explained by climate, but the relative role of local environmental predictors is not fully understood. Moreover, in the freshwater realm, intrinsic differences existing between different categories of water bodies can lead to significant differences in species-environment relationships. Here, we tested the relative importance of broad-scale climate and local environmental predictors in explaining plant species distributions in freshwater lakes and streams.
2. We built species distribution models to investigate which predictors best explain aquatic plant distribution in two categories of water bodies. We used species inventories and records of 3 climate and 8 local environmental predictors for 150 lakes and 150 streams in Finland.
3. We found that sets of predictors that explain the distribution of macrophyte species are unique depending on if species are in a lake or a stream. Overall, air temperature and ecosystem size were essential to predict aquatic plant species presence in both water body categories. Broad-scale climate predictors were always very important in explaining species distribution, while local environmental conditions such as water chemistry were of variable influence, depending on species and water body category.
4. These results are likely due to high spatial and temporal variability and range of water physico-chemical parameters, especially in streams. Nonetheless, despite a lower relative importance than climatic factors, local environmental predictors also strongly affected species distributions.
5. Our findings highlight that incorporating local environmental conditions to species distribution models in addition to climate predictors is necessary to improve predictions, particularly for distribution of stream flora. Considering the species-specific responses of aquatic plants to their environment, studying species individually with species distribution models represents a useful analysis.

1. Introduction

Understanding patterns and drivers of species distributions has intrigued ecologists and biogeographers for decades (Humboldt & Bonpland, 1805; Cox, Moore & Ladle, 2016). The study of geographical distributions allows researchers to estimate the ecological requirements or limitations of species, with the aim to disentangle the complexity of processes that determine their location (Brown, Stevens & Kaufman, 1996). In this way, study of factors associated with distributions contributes to better understanding of the ecological and evolutionary history of species, and can provide information crucial for conservation management. Organism distributions at large and regional scales have typically been explained by climatic factors, but less frequently by local environmental predictors like physico-chemical characteristics of water or soil (Dubuis *et al.*, 2013; Chappuis, Gacia & Ballesteros, 2014). Although data on local environmental factors over large geographic scales are often less accessible and/or available than climate data, these data are important for understanding the ecological niche of a species. For example, predictions of terrestrial plant species distributions can be improved by including local environmental factors such as soil properties (Dubuis *et al.*, 2013), and prediction of aquatic diatom distributions are improved by inclusion of water chemistry data in models (Potapova & Charles, 2002; Pajunen, Luoto & Soininen, 2016).

Climate has been found to have a stronger structuring effect on species distributions in terrestrial systems than it is in aquatic ecosystems (Santamaría, 2002). This difference is thought to result from a few fundamental differences between the environments. The most obvious and primary difference is the intrinsically different relationship of a terrestrial vs. an aquatic plant species to water. The survival of terrestrial species is dependent on their access to water and is largely influenced by climate. In contrast, aquatic macrophytes are obligate wetland species, but their relationship to water velocity, depth, or physico-chemical properties greatly influences their growth and distribution (Bornette & Puijalon, 2011; Anten & Sterck, 2012). Second, terrestrial species often experience high temperature fluctuations, whereas the aquatic environment narrows the temperature gradient for resident species (Nakano & Murakami, 2001; Santamaría, 2002). Nonetheless, the relative importance of climate in defining species habitat compared with the local environmental factors is somewhat poorly known in aquatic systems. Recently, King, Cheruvilil & Pollard (2019) showed that abiotic properties such as total phosphorus and total nitrogen values in US lakes, streams and wetlands were not dependent on water body categories; however, they demonstrated that percentage cover of aquatic vegetation was higher in lakes and wetlands, and lower in streams. Exploring factors that drive freshwater species distributions at various scales and in different ecosystems can improve our general understanding of species habitat preferences.

Although the influence of differently scaled environmental predictors (e.g., climate vs. local environment) in explaining species distributions has received limited attention in freshwater habitats compared to other ecosystems, even less is known about how species distributions differ among freshwater water body categories. Lakes and streams represent a major freshwater resource (Dudgeon *et al.*, 2006), and possess distinct physical properties resulting in substantial variation in their habitat characteristics. Streams present unidirectional running water, with frequent water level fluctuations due to intermittent discharges, which generate recurrent natural disturbances and large variations in water chemistry (Wetzel, 2001). On the other hand, lakes as lentic systems have much longer hydrologic residence times than streams, lower

amplitude of water level fluctuations, less temporal variation in water chemistry, and are thus overall more hydrologically stable than lotic systems.

Recent studies on differences in community-environment relationships between lakes and streams notably focused on planktonic species (Crump *et al.*, 2007; De Bie *et al.*, 2008) and macroinvertebrates (Johnson, Goedkoop & Sandin, 2004; Heino & Alahuhta, 2015). Nonetheless, the extent to which climatic factors, ecosystem size or water characteristics drive differences in species composition between lake and stream communities remains poorly understood. By studying diatoms in both lentic and lotic systems located in nearby study areas, Soininen & Weckström (2009) provided evidence that communities were structured by both local and large-scale environmental factors. Szoszkiewicz *et al.* (2014) demonstrated in Polish lakes that macrophyte communities were mainly structured by water quality parameters related to eutrophication (total organic carbon, total N, Secchi depth), whereas in streams they were primarily controlled by substrate type. However, to best of our knowledge, there are no existing studies investigating the influence of niche characteristics on same plant species growing in both lakes and streams.

The differences in the frequency and intensity of disturbances between and within lakes and streams can result in the presence of different life forms (Bornette & Puijalon, 2011). Overall, the different life forms of aquatic plants often exhibit variable responses to environmental factors (Netten *et al.*, 2011; Xia *et al.*, 2014), due to their specific morphological characteristics and their position in the water column. For example, free-floating species are strongly dependent on nutrient availability in the water column (Schneider *et al.*, 2018), while floating-leaved, submerged and helophyte species can obtain nutrients from sediments (Barko & Smart, 1986). However, the growth and survival of submerged species is dependent on under-water light availability, while the three other functional life forms have morphological primacy to obtain this resource. Therefore, recognition and evaluation of the different functional life form groups of macrophytes, and their responses to environmental resource variables that may limit their growth and distribution can reveal evidence on what shapes macrophyte communities.

In this study, we used species distribution modelling (SDM) to investigate the relative importance of local environmental and climate predictors in explaining freshwater plant distributions, using data from 150 lakes and 150 streams covering a latitudinal gradient of ca. 1000 km in Finland. A second objective of the work was to determine if plant species that occur in both lake and stream habitats respond differently to environmental factors in these aquatic habitats. Based on previous research (Kosten *et al.*, 2009; Alahuhta *et al.*, 2013; Rääpysjärvi, Hämäläinen & Aroviita, 2016), we hypothesised that i) regional climate will act predominantly on lakes and will be less important in streams due to higher variation in local environmental conditions (e.g., discharge) that dominate over regional climate, ii) the individual predictors that best predict aquatic plant distributions will be the same among lakes and streams, and iii) different life forms of aquatic plants will be sensitive to different local and climate predictors.

2. Methods

2.1. Study sites

The studied water bodies belong to the Finnish national monitoring network and were evenly distributed in latitude and among stream basins (Fig. 1). These 150 lakes and 150 streams are

variably impacted by human pressures, ranging from close to pristine headwaters to agricultural dominated freshwaters of lowlands. In addition, geological and soil properties within stream basins of the studied water bodies are different, with diverse degrees of organic peat, mineral and clay soils. The inclusion of different kinds of lentic and lotic systems enabled us to increase the range of local environmental conditions within the data set, leading to better comparability among sites across a relatively wide climate gradient.

2.2. Species data from aquatic plant surveys

Presence and absence data of aquatic plant species in the 300 studied water bodies were obtained from field observation data maintained by the Finnish Environment Institute. Lake plants were surveyed using a main belt transect method (Kanninen *et al.*, 2013), where each five-metre-wide transect extended, perpendicular to the shoreline, from the upper eulittoral to the outer depth limit of vegetation. Alternatively, if vegetation covered the entire lake surface, then the transect was drawn to the deepest point of the basin. The number of transects varied between 7 and 25 (mean = 14, SD = 3.6), depending on lake size and number required to secure proper coverage of species composition (Kanninen *et al.*, 2013). Lake plants were observed by wading or by boat, with the aid of rake and hydroscope. The plant surveys were done between June and September over the period 2006–2012. In addition to true aquatic plants (i.e., hydrophytes), we also recorded helophytes species, which are ecologically important in the boreal region (Toivonen & Huttunen, 1995).

Stream plants were surveyed from two different sections of 100m in each stream: a riffle and adjacent pool section (Rääpysjärvi *et al.*, 2016). Each section was divided into five 20 m long subsections, where the presence of stream plant species was assessed. Stream plants were primarily observed by wading but also by boat using rake and hydroscope in larger streams. The species recorded in the riffle and pool section were pooled together to represent total plant occurrence in the whole stream. The plants were surveyed once at each stream in July–August between 2009 and 2016. Only vascular plants (i.e., hydrophytes and helophytes) were used in this study. The different species recorded were classified by life form as per Toivonen & Huttunen (1995): free-floating, floating-leaved, submerged and helophyte species.

2.3. Explanatory predictors

Similar explanatory predictors, describing climate and local environment, were used to study the distribution of aquatic plant species in the lakes and streams. These predictors were either resource, direct or indirect variables, *sensu* Guisan & Zimmermann (2000), that influence the growth, development and distribution of aquatic plants (Fig. 2). Identical local environmental predictors measured in both lakes and streams were alkalinity (mmol.L⁻¹), colour (mg PT.L⁻¹), conductivity (mS.m⁻¹), pH, TN (µg.L⁻¹), TP (µg.L⁻¹) and turbidity (PTU). These water chemistry predictors are known to strongly influence the occurrence and growth of aquatic plants (Toivonen & Huttunen, 1995; Rääpysjärvi *et al.*, 2016), either because they are resource predictors, such as nutrients (e.g., TN and TP), or because they directly or indirectly have an impact on resource predictors (Fig. 2). In addition, ecosystem size (surface area (km²) in lakes and stream width (m) in streams) was used to represent habitat availability for plant species (Jones, Li & Maberly, 2003). In lakes, water chemistry predictors were mean values of multiple individual samples collected between 2006 and 2012. In streams, water chemistry data was based on mean values of multiple samples taken from 2006 through 2012 for majority of

streams (134 out of 150), overlapping with stream plant surveys (2009-2012). For mean water chemistry values of samples collected during ice-free period in all lakes and most streams ($n=134$), the number of individual samples varied among predictors and water bodies, ranging from three to 50 depending on local monitoring frequency. For 16 streams out of 150 surveyed in year 2016, the water chemistry values were single samples taken simultaneously with the plant surveys (see Rääpysjärvi *et al.*, 2016 and Toivanen *et al.*, 2019 for further details). Year-to-year variation in hydrology and water chemistry may have had some influence on mean water quality values when number of samples was low. Nevertheless, for these water bodies, no additional measurements on water chemistry were available during ice-free period, while the geographical balance between lakes and rivers within the different catchment areas was an important criterion in selecting the studied water bodies.

Climate predictors included growing degree days $> 5^{\circ}\text{C}$ (GDD5), temperature of the coldest month (January, $^{\circ}\text{C}$) and mean annual precipitation (mm). These climate data, obtained from the Finnish Meteorological Institute (Pirinen *et al.*, 2012), represent the most updated high-resolution climate data from Finland, with 1-km resolution from the period of 1981-2010. Growing degree days directly affects plant growth, whereas temperature of the coldest month was used as a proxy for negative effects of winter conditions on plants (e.g., ice erosion and freezing of sediments; Lind, Nilsson & Weber, 2014). Although atmospheric temperature was used, it is often known to closely correlate with water temperatures (O'Reilly *et al.*, 2015). Mean annual precipitation represented indirect effects of water level fluctuations and leaching of nutrients from land to water by runoff (Johnson *et al.*, 1997). In lakes, the climate predictors were averaged for lake surface area, whereas values of climate predictors were extracted for plant survey points in the study streams. The median and range values of the predictors are shown in Fig. S1.

2.4. Modelling method and statistical analyses

Among all the species inventoried (a total of 115 species for lakes, 65 species for streams), species were included in the analyses if they occurred in more than 5% and less than 95% of the 150 studied lakes or streams. Thus, 80 species were selected for lakes, and 41 for streams, with 38 species shared between the two water body categories (Table S1).

We performed the predictor selection and the distribution modelling described below using R version 3.4.3 (R Core Team, 2017), with six algorithms implemented within the biomod2 package (Thuiller *et al.*, 2016): two machine learning methods, (1) generalised boosted models (GBM) (Ridgeway, 1999) and (2) random forest (RF) (Breiman, 2001), and four regression methods, (3) generalised linear model (GLM) (McCullagh & Nelder, 1989), (4) generalised additive model (GAM) (Hastie & Tibshirani, 1990) (5) multivariate adaptive regression splines (MARS) (Friedman, 1991) and (6) flexible discriminant analysis (FDA) (Hastie, Tibshirani & Buja, 1994). RF is a classification method that grows multiple decision trees based on random subsets of the data and after a large number of trees is generated, they individually vote for a class, ultimately generating a prediction by the forest. GBM is based on a combination of boosting techniques and trees. For both machine learning algorithms, new trees take into account the error of previously built trees. GLM is a regression model for data with a non-normal distribution, fitted with maximum likelihood estimation. GAM is a multiple regression model, that uses non-parametric smoothing functions to model non-linear relationships between

the response and the predictors. MARS builds multiple linear regression models by partitioning the data and run a linear regression model on each different partition. Finally, FDA is a multigroup classification method based on a mixture of linear regression models and using optimal scoring to separate the groups. For all six modelling techniques, we used the default parameters as implemented in R. These different algorithms have different strengths and limitations; thus, when testing the same phenomenon, the use of multiple algorithms supports improved predictions. These algorithms require datasets with both presence and absence to perform, which was the case of our dataset as actual plant species presences and absences were recorded in the field. Therefore, we did not need to generate pseudo-absences.

2.4.1. Predictor selection

We applied a protocol to select best relevant predictors for each species among the 11 predictors (3 climate predictors and 8 local environmental predictors), separately for lakes and streams. Firstly, we identified intercorrelated predictors (Pearsons' $r < 0.70$; see Fig. S2 and S3) (Dormann *et al.*, 2013), and then within each group of intercorrelated predictors, we obtained the relative importance of each predictors with 3 runs of 10 random permutations, as implemented in biomod2. To allow comparison among the results from the six algorithms, the values of importance were then converted into a rank of importance, per run and per algorithm, rank 1 being attributed to the lowest value of importance. We computed the mean of rank values of a predictor for each run, and these values were then averaged to obtain a final rank value for each predictor. The predictor with the highest average rank was the most important of the intercorrelated predictors, and therefore was selected.

Finally, among the previously selected uncorrelated predictors, we kept those that were identified as important predictors (as computed by the variable importance procedure of biomod2; 4 runs) by at least three of the six modelling techniques. Important predictors did not always coincide among the six algorithms (Table S2 and S3). A detailed protocol of variable selection is documented in Leroy *et al.*, 2014. Using this established method, we selected a subset of predictors suited to each species in lakes or streams. As intercorrelated predictors were not the same for lakes and streams, we calculated the basic probability of each predictor to be incorporated into models after the first step of predictor selection (Table 1).

2.4.2. Modelling process

We calibrated the models with a random selection of 80% of the presence-absence data, and evaluated the model performance with the other 20%, with the six modelling techniques introduced above. Two metrics were used to evaluate model performance: the area under the receiver operating characteristic curve (ROC) and the true skills statistic (TSS) (Fielding & Bell, 1997; Allouche, Tsoar & Kadmon, 2006). The ROC analysis involves plotting sensitivity (i.e. the proportion of known presences predicted present, aka true presences) against 1-specificity (i.e., proportion of known absences predicted present, aka false presences) (Peterson, Papeş & Soberón, 2008). The area under the ROC curve is then compared against null expectations (ROC value of 0.5). ROC values range from 0 to 1 and higher the value, the better the algorithm perform at predicting species distribution. At 0, the algorithm never falsely identifies species absences, but it also fails at identifying all known presences; at 1, the algorithm correctly identifies all true presences, and never misclassifies a true absence as a

presence (i.e. no false presence). TSS is a threshold dependant measure of algorithm accuracy, that measures the difference between sensitivity and 1-specificity. It ranges from -1 to +1, where +1 indicates perfect agreement between predictions and observations and values of 0 or less indicate that the algorithm perform no better than random (Allouche *et al.*, 2006). Algorithms with TSS values below 0.4 are commonly considered as performing poorly.

The calibration and evaluation steps were replicated five times per modelling technique. Runs with ROC <0.5 were discarded in further analysis, as such models do not perform better than random predictions. For each of the five runs, the predicted probabilities of occurrence were transformed into binary prediction using the probability threshold that maximises the TSS score (Liu *et al.*, 2005; Allouche *et al.*, 2006). We then used committee averaging to obtain consensus results for each modelling technique (i.e., we attributed species presence in a cell when more than half of the runs predicted presence, otherwise we assigned species absence; Araújo & New, 2007).

3. Results

3.1. Evaluation of models

About 11.3% of the models from lake data and 8.3% of the models from stream data had ROC values below 0.5, indicating that such models perform no better than random, and were thus discarded. Calibrated models had mean ROC values above 0.7 and TSS values above 0.4 for both lakes and streams, but with a large dispersion of values depending on species and runs (Fig. S4). This pattern was similar among the six modelling techniques. Floating-leaved species had overall lower ROC values than the three other life forms, based on confidence intervals (Fig. S5).

3.2. Importance of predictors

For lakes, the groups of intercorrelated predictors were 1: GDD5 and January temperature, 2: TP and TN, and 3: alkalinity, conductivity and pH (Fig. 3, Fig. S2). For streams, intercorrelated predictors were 1: GDD5 and January temperature, and 2: TP and turbidity (Fig. 3, Fig. S3). The group with GDD5 and January temperature was the only one comparable between lakes and streams. For both the step of predictor selection among intercorrelated predictor groups (Fig. 3) and the step of final predictors selection (Fig. 4), the predictors that were considered most important by the models were sometimes life form dependent, or water body category dependent. Overall, alkalinity and pH were more frequently selected for plants in lakes than for those in streams, while turbidity was more frequently included for plants present in streams than for those in lakes, independently of life forms (Fig. 4),

3.2.1. Free-floating species

Within the first group of intercorrelated predictors, GDD5 and January temperature were comparably selected between lakes and streams, with January temperature being preferred in 2/3 of the cases (Fig. 3a,b). In lakes, TN and TP were similarly selected, which was also the case for the predictors of the third group, alkalinity, conductivity and pH. In streams, TP was selected over turbidity for two of the three free-floating species (Fig. 3b).

The final step of predictor selection showed that the three climate predictors as well as conductivity and turbidity were selected more times in lakes than in streams, while TN and

colour were evenly selected between the two water bodies category (Fig. 4a). TP was selected in 100% and 80% of the models in streams and lakes, respectively. Ecosystem size, represented by lake area and stream width, was noticeably more important for lakes, where it was selected by 89% of the models when available, than for streams where it was never selected (Fig. 4a).

3.2.2. Floating-leaved species

GDD5 and January temperature were evenly selected in lakes, and GDD5 was preferred over January temperature in streams (Fig. 3c,d). In lakes, TN was more frequently selected (78%) than TP. Within the third group of intercorrelated predictors in lakes, conductivity was noticeably more frequently selected. In streams, turbidity was more efficient than TP to predict the distribution of all three floating-leaved species (Fig. 3d).

The final step of predictor selection showed that GDD5, precipitation and TN were more frequently selected in lakes than in streams (Fig. 4a). TP was only available for lake species and had been selected by 50% of the models (Fig. 4b). January temperature, conductivity and colour were selected in the same proportions (100%, 67% and 67%, respectively) between lakes and streams. Ecosystem size was selected a bit more often in streams than in lakes (Fig. 4b).

3.2.3. Helophyte species

GDD5 was preferred over January temperature in lakes, and these two intercorrelated predictors were evenly selected in streams (Fig. 3e,f). In lakes, TN and TP were similarly selected. Within the third group of intercorrelated predictors in lakes, conductivity was preferred over alkalinity and pH. For the helophytes in streams, TP and turbidity were equivalently selected (Fig. 3f).

Whether helophytes were present in lakes or streams, the three climate predictors, colour and ecosystem size were selected in the same proportions among water body categories (Fig. 4c). TN, TP and conductivity were considered by the models to be more important to predict the distribution of helophyte species in lakes than in streams. Finally, turbidity was selected slightly more often for plants present in streams than for those in lakes (Fig. 4c),

3.2.4. Submerged species

The intercorrelated predictors GDD5 and January temperature were comparably selected between lakes and streams, with January temperature being preferred over GDD5 (Fig. 3g,h). In lakes, TN and TP were similarly selected, which was also the case for the predictors of the third group, alkalinity, conductivity and pH. For a large majority of submerged species in streams, TP was considered to be a better predictor than turbidity (Fig. 3h).

The final step of predictor selection showed that TN, GDD5, precipitation and turbidity were more frequently selected in streams than in lakes (Fig. 4d). On the contrary, colour and TP were selected more often in lakes than in streams. Conductivity and ecosystem size were selected similar proportions between lakes and streams (Fig. 4d).

3.3. Focus on species shared between lakes and streams

The patterns shown in Fig. 4 were mostly similar when restricted to the set of 38 species shared between lakes and streams (Fig. S6). On average, selected predictors were similar in both water body categories only for 39% of the species in common (Fig. 5). This percentage was highly variable among species, but none of the species displayed a 100% match of selected

predictors between lake and stream (Fig. 5). Interestingly, the model selections of the best predictor among correlated predictors did not always lead to the same results between lakes and streams. Thus, for example, the selected predictors were completely independent for *Sparganium erectum*, depending on whether the species was found in lakes or streams. Other striking cases are those of *Juncus filiformis*, *Potamogeton natans*, and *Utricularia intermedia*, for which none of the three climate predictors were selected to predict their presence in streams. In lakes, *Carex rostrata* was the only species that presented such a pattern. When comparing evaluation metrics between lakes and streams, predictor selection resulted in better models for species in streams than for lakes. For example, see results for *J. filiformis* and *U. intermedia* (Fig. S7). In addition to qualitative differences, there were also quantitative differences between lakes and streams in the number of predictors selected. For instance, for *Phragmites australis*, GDD5 and TP were the only two predictors selected, while six predictors had been selected for the species in streams (Fig. 5). Despite this difference, the prediction accuracy of models was similar between lakes and streams for *P. australis* (Fig. S7).

A few other patterns were noticeable when considering species life forms: for the three free-floating species shared between lakes and streams, both stream width and GDD5 were never considered important predictors (Fig. 5). Turbidity was selected for five out of the seven submerged species when they were present in lakes, but only for one species in streams.

To visually illustrate differences in predictive performances of species distribution between lakes and streams, we focused on the predictions of GAM for four helophyte species shared between both water body categories that differed in prediction accuracy. We chose to focus on GAM arbitrarily, any of the five other models would allow to show similar examples, although on different species. The model showed accurate predictions of the presence of *Glyceria fluitans* in both lakes and streams (Fig. 6a), whereas the occurrences of *Sparganium emersum* were not predicted accurately in either lakes or streams (Fig. 6b). *C. rostrata* and *Menyanthes trifoliata* are two other interesting examples, the presence of the former species was better predicted in streams than in lakes (Fig. 6c), while the opposite was observed for the latter species (Fig. 6d). The similarities and differences in prediction performances for these four species seem unrelated to differences in environmental predictors or number of predictors used (Fig. 5), or to the values of their evaluation metrics (Fig. S7). Moreover, the performances of the models were independent of species prevalence (Fig. S8). The model evaluation and predictions could be improved for some species, such as for *Elatine triandra*, *Hippuris vulgaris*, *Myriophyllum verticillatum*, *Nymphaea alba*, *P. australis*, *U. intermedia*. The predictions of these species' distribution could benefit from the use of additional predictors, that would widen the predicted fundamental niche (less false absence predicted), or on the contrary, that would narrow the predicted fundamental niche (less false presences predicted).

4. Discussion

In this study, we investigated the contribution of climate and local environmental predictors by modelling the distribution of boreal aquatic plant species using data set from 150 lakes and 150 streams. We found that air temperatures and ecosystem size were crucial predictors of species presence in both lakes and streams, whilst the importance of water chemistry was significant to a variable degree for lake and stream species. Moreover, by comparing the model outcomes for species shared between lakes and streams, we found that the set of best predictors

of a given species distribution can highly vary between the water body category and is also quite different depending on life forms.

4.1. Climate is an important predictor of macrophyte distributions in lakes and streams

We first hypothesised that regional climate would act predominantly on lakes, while local predictors will be more important in streams. This assumption was not supported by our models, as most local predictors did not dominantly structure aquatic plant distributions in streams, and climate predictors had similar importance in both water body categories. Instead, especially in streams, air temperature predictors were selected at least as often by models as water physico-chemical predictors. Mean annual precipitation was also frequently selected by models, both in lakes and streams, which is not surprising considering its large influence on water level fluctuation, changes in flow velocity and nutrient input from catchment area (Tuo *et al.*, 2016), and likely on plant propagule dispersal for those species which disperse via watercourses (Andersson & Nilsson, 2002; Merritt & Wohl, 2002). Thus, the three tested climate predictors indicate different broad-scale environmental conditions, ranging from direct temperature effects to indirect local hydrological and water chemical influences on species distributions, and distinguishing pure effects of any single predictor is most challenging. Similarly to our models outcome, both climate and local predictors have been found to explain community composition and species richness of aquatic plants in numerous previous studies (Heino, 2001; Kosten *et al.*, 2009; Alahuhta *et al.*, 2013; Grimaldo *et al.*, 2016). Furthermore, the global beta diversity of aquatic plants has been shown to be affected by temperature variation within the study region (Alahuhta *et al.*, 2017). However, these previous studies on aquatic plants focused on community composition, species richness, plant cover or beta diversity patterns. Here, we explicitly quantified that climate is an important characteristic in explaining distributions of individual taxa of aquatic plants.

We found that local environmental predictors were also useful and highly relevant to predict species distribution, similarly to the findings of Pont, Hugueny & Oberdorff (2005). Incorporating fine-scale predictors to models, such as physico-chemical predictors, can increase their fit for some species, but they do not necessarily improve, and can even decrease, the models' fit for other species (Gies *et al.*, 2015). This suggests that the relative importance of climate and local environmental predictors varies depending on species, as found in the present study. However, even the predictors that were selected the least frequently by the models were, independent of life forms, considered important for 40% of the species. Our results indicate that neither local environmental predictors nor climatic effects dominate in explaining aquatic plant distribution at regional extent, but that both contribute to species distributions.

Our results must be carefully interpreted and nuanced by the fact that some predictors differed in their probability to be selected by the models, due to differences in the number and nature of correlated predictors between the two water body categories. One limit of the present study could be that we used only three climate predictors, and eight local environmental predictors to predict species distributions. Due to this unbalanced number between the two kinds of predictors, we did not directly compare models built with either climate or local predictors. However, the chosen climate predictors are ecologically relevant to predict aquatic plant distributions (Alahuhta *et al.*, 2013). Indeed, the two temperature predictors had been chosen in accordance to the climatic conditions encountered by plants in high-latitudes, where

growth is mostly limited by low temperatures and where frost can damage plant propagules (Lind *et al.*, 2014). Nevertheless, the impacts of predictors such as extreme temperatures would be worth exploring (Feldmeier *et al.*, 2018), especially considering that climate change will increase the frequency of extreme events such as polar vortex, which could impact species survival and so their distribution. Some of the tested predictors are proxies for other environmental variables (light, nutrients, photosynthesis). Nonetheless, the addition of other predictors that have been shown to influence species distribution, such as soil variables (Dubuis *et al.*, 2013), geodiversity (Toivanen *et al.*, 2019), land cover (Luoto, Virkkala & Heikkinen, 2007; Gallardo & Aldridge, 2013), water temperature (Cianfrani, Satizábal & Randin, 2015), human footprint (Rodríguez-Merino *et al.*, 2018) or biotic interactions (Wisz *et al.*, 2013), could improve predictive performance at least for some species. Using SDMs to explore the statistical relationship between species occurrences and environmental predictors is useful but has limitations, as models are an estimation of the fundamental niche based on the realized niche, as observed distributions are constrained by biotic interactions and limiting resources (Guisan & Thuiller, 2005). Plus, the algorithms each have their specific limitations producing uncertainty and can be more or less suited to a given species. However, using multiple SDMs techniques allow to obtain consensus results, coping with some of the individual limits of the techniques. While in many studies pseudo-absences must be generated to perform SDMs (Guisan & Thuiller, 2005; Barbet-Massin *et al.*, 2012), our dataset included real absences, providing more strength to the predictions.

4.2. Different predictors shape aquatic plant distributions in lakes and streams

Our second hypothesis was that same predictors could explain aquatic plant distributions in lakes and streams. We found some support for this assumption, as climatic predictors were mostly equally important for both lake and stream plants. However, different local environmental predictors explained species distributions in these two habitats. Considering that the selected physico-chemical predictors varied strongly between lakes and streams even for same species, the finding may be related to the range of the values of the two predictor groups. Indeed, the values of local physico-chemical predictors can strongly differ among consecutive years and for geographically adjacent water bodies (Borghini *et al.*, 2008). The range in climate predictors also typically increases more with increasing scale than the gradient in local environmental variation (Willis & Whittaker, 2002). Moreover, GDD5 may also indirectly reflect catchment productivity, leading to increased influxes of carbon and inorganic nutrients to the surface water (Pajunen *et al.*, 2016). Thus, our results demonstrate that local environmental predictors with large ranges in values across water bodies are of high importance to predict aquatic plant distributions at regional scales, but their influence on species distributions is more species-specific compared to that of climate predictors.

Our study agreed with some recent findings, but also revealed novel insights on the importance of temperature on aquatic plant distributions. On average, GDD5 and temperature of the coldest month were both important predictors of the species distributions, similar to previous studies (Alahuhta, Heino & Luoto, 2011; Pajunen *et al.*, 2016). Nonetheless, for the set of 38 species found in both habitats, plants in lakes responded more often to GDD5, which is an indicator of heat accumulation, having a direct effect on metabolic processes and influences primary production (Brown *et al.*, 2004). Although we cannot distinguish direct

temperature effects of GDD5 from indirect catchment productivity influences of this predictor on aquatic plant distributions (Pajunen *et al.*, 2016), the majority of lake plants responding to the GDD5 were helophytes and floating-leaved species that at water surface are directly exposed to air temperatures. The temperature of the coldest month contributes to species distributions through, for example, ice erosion, freezing of bottom sediments, or by limiting light penetration and air-water gas exchanges due to thick ice and snow cover (Hellsten, 2001; Lacoul & Freedman, 2006). These harmful winter effects are often more severe in streams than in lakes, because greater moving of ice (e.g., through anchor ice formation) in streams remoulds sediments and thus removes plants (Lind *et al.*, 2014). This likely explains why stream plants responded strongly to the temperature of the coldest month in our study. Helophytes are especially vulnerable to harsh winter conditions (Hellsten, 2001), but species belonging to other growth forms also responded to the temperature of the coldest month in our investigation. Netten *et al.* (2011) showed that mild winters with less intense harmful effects give benefit to free-floating plants and to submerged plants with an evergreen overwintering strategy, over the submerged flora whose vascular system does not survive in winter. However, our findings do not clearly affirm or contradict these results.

Our third hypothesis was validated, as expected different life forms of aquatic plants were sensitive to different predictors. This was especially noticeable considering local environmental conditions, as their effects on the distributions of different life forms were easier to distinguish among the selected predictors. Availability and form of carbon is highly important for aquatic plants (Sand-Jensen, Binzer & Middelboe, 2007), and correlated predictors describing this phenomenon (i.e., conductivity, alkalinity and pH) were often important for lake plants but less so for stream plants. Similarly, TP had a higher influence on lake plant than on stream plant distributions, in accordance with findings by Szoszkiewicz *et al.* (2014). Thereby, our findings suggest that predictors indicating variation in water chemistry play a greater role in lakes with a slower rate of water replacement compared to streams. On the other hand, stream width and lake area were selected in most models, often equally between lakes and streams. Ecosystem size indeed explains the presence of aquatic plant species (Chambers & Kalf, 1984; O'Hare *et al.*, 2006; Alahuhta *et al.*, 2015), as it likely acts as a surrogate for habitat availability.

In comparison to lakes, streams subject aquatic plants to greater disturbances such as increased mechanical stress against stems, and resource allocation to rooting production for anchorage. Stream biota is also more prone to effects of multiple anthropogenic stressors (e.g. nutrient enrichment, sediments, alterations in hydrological morphological conditions (Hering *et al.*, 1986), while lakes are most strongly affected by nutrient enrichment and eutrophication (Szoszkiewicz *et al.*, 2014). These inherent differences between the water body categories can lead to differentiated impacts on the macrophyte communities, such as lower diversity from water discharge (Franklin, Dunbar & Whitehead, 2008) or increase in biomass production due to non-limiting nutrient concentrations (Hilton *et al.*, 2006). For these reasons, local predictors indicating physical environmental conditions in addition to water chemical properties are necessary when modelling species distributions of stream plants (Barendregt & Bio, 2003). Further studies could explore whether stream species present in riffles and pools respond to similar predictors, and if results for species in pools compare more closely to those that inhabit lakes.

To conclude, we found that different predictors shape the distribution of macrophyte species depending on if they are in a lake or a stream. In addition, our study indicates that broad-scale climate predictors are important environmental characteristics, often even overriding the effects of local conditions in explaining aquatic plant distributions at regional extent. However, our analysis also showed that despite lower relative importance, local environmental predictors strongly contribute to aquatic plant species distributions, complementarily to climate predictors. This emphasises the fact that both individual and joint effects of climate and local habitat variation are needed to accurately model aquatic plant distributions in freshwaters. Furthermore, different life forms were impacted differently by the tested predictors, and the way different species and life forms use the resources available at a given time and place could determine the local assemblage. We demonstrated that drivers of aquatic plant distributions vary strongly among species, meaning that species respond individualistically to environmental gradients. Thus, species-specific species distribution models offer a promising tool to explore aquatic plant distributions in relation to local environmental conditions and climate.

Acknowledgments

We thank Minna Kuoppala for maintaining the lake macrophyte database and Krister Karttunen, Juha Riihimäki, Jaana Rääpysjärvi, Jarno Turunen and the many trainees for the long days in the field in rain and shine. Parts of the data were collected in MaaMet-monitoring funded by the Finnish Ministry of Agriculture and Forestry. We warmly thank Brenda J. Grewell for the time she kindly dedicated to the English proofreading of this manuscript. M.B. Gillard was supported by a USDA-ARS funded post-doctoral researcher appointment at UC Davis while conducting this research. J. Aroviita was supported by the MARS (contract No. 603378) project funded under the 7th Framework Programme by the European Union, the Freshabit EU LIFE IP project (LIFE14/IPE/FI/023) and the Nordic Centre of Excellence “BIOWATER” (Nordforsk Project no. 82263). J. Alahuhta acknowledges support from the Academy of Finland.

Data availability

Data are available from the authors upon request.

References

- Alahuhta J., Heino J. & Luoto M. (2011). Climate change and the future distributions of aquatic macrophytes across boreal catchments. *Journal of Biogeography* **38**, 383–393. <https://doi.org/10.1111/j.1365-2699.2010.02412.x>
- Alahuhta J., Kanninen A., Hellsten S., Vuori K.-M., Kuoppala M. & Hämäläinen H. (2013). Environmental and spatial correlates of community composition, richness and status of boreal lake macrophytes. *Ecological Indicators* **32**, 172–181. <https://doi.org/10.1016/j.ecolind.2013.03.031>
- Alahuhta J., Kosten S., Akasaka M., Auderset D., Azzella M.M., Bolpagni R., *et al.* (2017). Global variation in the beta diversity of lake macrophytes is driven by environmental heterogeneity rather than latitude. *Journal of Biogeography* **44**, 1758–1769. <https://doi.org/10.1111/jbi.12978>
- Alahuhta J., Rääpysjärvi J., Hellsten S., Kuoppala M. & Aroviita J. (2015). Species sorting

- drives variation of boreal lake and river macrophyte communities. *Community Ecology* **16**, 76–85. <https://doi.org/https://doi.org/10.1556/168.2015.16.1.9>
- Allouche O., Tsoar A. & Kadmon R. (2006). Assessing the accuracy of species distribution models: Prevalence, kappa and the true skill statistic (TSS). *Journal of Applied Ecology* **43**, 1223–1232. <https://doi.org/10.1111/j.1365-2664.2006.01214.x>
- Andersson E. & Nilsson C. (2002). Temporal variation in the drift of plant litter and propagules in a small boreal river. *Freshwater Biology* **47**, 1674–1684. <https://doi.org/10.1046/j.1365-2427.2002.00925.x>
- Anten N.P.R. & Sterck F.J. (2012). Terrestrial vs aquatic plants: How general is the drag tolerance-avoidance trade-off? *New Phytologist* **193**, 6–8. <https://doi.org/10.1111/j.1469-8137.2011.03994.x>
- Araújo M.B. & New M. (2007). Ensemble forecasting of species distributions. *Trends in ecology & evolution* **22**, 42–7. <https://doi.org/10.1016/j.tree.2006.09.010>
- Barbet-Massin M., Jiguet F., Albert C.H. & Thuiller W. (2012). Selecting pseudo-absences for species distribution models: how, where and how many? *Methods in Ecology and Evolution* **3**, 327–338. <https://doi.org/10.1111/j.2041-210X.2011.00172.x>
- Barendregt A. & Bio A.M.F. (2003). Relevant variables to predict macrophyte communities in running waters. *Ecological Modelling* **160**, 205–217. [https://doi.org/10.1016/S0304-3800\(02\)00254-5](https://doi.org/10.1016/S0304-3800(02)00254-5)
- Barko J.W. & Smart R.M. (1986). Sediment-Related Mechanisms of Growth Limitation in Submersed Macrophytes. *Ecology* **67**, 1328–1340
- De Bie T., Declerck S., Martens K., De Meester L. & Brendonck L. (2008). A comparative analysis of cladoceran communities from different water body types: Patterns in community composition and diversity. *Hydrobiologia* **597**, 19–27. <https://doi.org/10.1007/s10750-007-9222-y>
- Borghini F., Colacevich A., Caruso T. & Bargagli R. (2008). Temporal variation in the water chemistry of northern Victoria Land lakes (Antarctica). *Aquatic Sciences* **70**, 134–141. <https://doi.org/10.1007/s00027-008-8026-0>
- Bornette G. & Puijalon S. (2011). Response of aquatic plants to abiotic factors: a review. *Aquatic Sciences* **73**, 1–14. <https://doi.org/10.1007/s00027-010-0162-7>
- Breiman L. (2001). Random forests. *Machine learning* **45**, 5–32. <https://doi.org/10.1023/A:1010933404324>
- Brown J.H., Gillooly J.F., Allen A.P., Savage V.M. & West G.B. (2004). Toward a metabolic theory of ecology. *Ecology* **85**, 1771–1789. [https://doi.org/Doi 10.1890/03-9000](https://doi.org/Doi%2010.1890/03-9000)
- Brown J.H., Stevens G.C. & Kaufman D.M. (1996). The geographic range: size, shape, and internal structure. *Annual Review of Ecology and Systematics* **27**, 597–623. <https://doi.org/10.1146/annurev.ecolsys.27.1.597>
- Chambers P.A. & Kalf J. (1984). Depth distribution and biomass of submerged aquatic macrophyte communities in relation to Secchi depth. *Canadian Journal of Fisheries and Aquatic Sciences* **42**, 701–709
- Chappuis E., Gacia E. & Ballesteros E. (2014). Environmental factors explaining the distribution and diversity of vascular aquatic macrophytes in a highly heterogeneous Mediterranean region. *Aquatic Botany* **113**, 72–82. <https://doi.org/10.1016/j.aquabot.2013.11.007>

- Cianfrani C., Satizábal H.F. & Randin C. (2015). A spatial modelling framework for assessing climate change impacts on freshwater ecosystems: Response of brown trout (*Salmo trutta* L.) biomass to warming water temperature. *Ecological Modelling* **313**, 1–12. <https://doi.org/10.1016/j.ecolmodel.2015.06.023>
- Cox C.B., Moore P.D. & Ladle R. (2016). *Biogeography: An Ecological and Evolutionary Approach, 9th Edition*. John Wiley & Sons.
- Crump R.C., Adams H.E., Hobbie J.E. & Kling G.W. (2007). Biogeography of bacterioplankton in lakes and streams of an arctic tundra catchment. *Ecology* **88**, 1365–1378
- Dormann C.F., Elith J., Bacher S., Buchmann C., Carl G., Carré G., *et al.* (2013). Collinearity: a review of methods to deal with it and a simulation study evaluating their performance. *Ecography* **36**, 27–46. <https://doi.org/10.1111/j.1600-0587.2012.07348.x>
- Dubuis A., Giovanettina S., Pellissier L., Pottier J., Vittoz P. & Guisan A. (2013). Improving the prediction of plant species distribution and community composition by adding edaphic to topo-climatic variables. *Journal of Vegetation Science* **24**, 593–606. <https://doi.org/10.1111/jvs.12002>
- Dudgeon D., Arthington A.H., Gessner M.O., Kawabata Z.-I., Knowler D.J., Lévêque C., *et al.* (2006). Freshwater biodiversity: importance, threats, status and conservation challenges. *Biological Reviews* **81**, 163–82. <https://doi.org/10.1017/S1464793105006950>
- Feldmeier S., Schefczyk L., Hochkirch A., Lötters S., Pfeifer M.A., Heinemann G., *et al.* (2018). Climate versus weather extremes: Temporal predictor resolution matters for future rather than current regional species distribution models. *Diversity and Distributions* **24**, 1047–1060. <https://doi.org/10.1111/ddi.12746>
- Fielding A.H. & Bell J.F. (1997). A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environmental Conservation* **24**, 38–49
- Franklin P., Dunbar M. & Whitehead P. (2008). Flow controls on lowland river macrophytes: A review. *Science of the Total Environment* **400**, 369–378. <https://doi.org/10.1016/j.scitotenv.2008.06.018>
- Friedman J.H. (1991). Multivariate adaptive regression splines. *The annals of statistics* **19**, 1–67
- Gallardo B. & Aldridge D.C. (2013). The ‘dirty dozen’: socio-economic factors amplify the invasion potential of 12 high-risk aquatic invasive species in Great Britain and Ireland. *Journal of Applied Ecology* **50**, 757–766. <https://doi.org/10.1111/1365-2664.12079>
- Gies M., Sondermann M., Hering D. & Feld C.K. (2015). Are species distribution models based on broad-scale environmental variables transferable across adjacent watersheds? A case study with eleven macroinvertebrate species. *Fundamental and Applied Limnology / Archiv für Hydrobiologie* **186**, 63–97. <https://doi.org/10.1127/fal/2014/0600>
- Grimaldo J.T., Bini L.M., Landeiro V.L., O’Hare M.T., Caffrey J., Spink A., *et al.* (2016). Spatial and environmental drivers of macrophyte diversity and community composition in temperate and tropical calcareous rivers. *Aquatic Botany* **132**, 49–61. <https://doi.org/10.1016/j.aquabot.2016.04.006>
- Guisan A. & Thuiller W. (2005). Predicting species distribution: offering more than simple habitat models. *Ecology Letters* **8**, 993–1009. <https://doi.org/10.1111/j.1461-0248.2005.00792.x>
- Guisan A. & Zimmermann N.E. (2000). Predictive habitat distribution models in ecology.

- Ecological Modelling* **135**, 147–186. [https://doi.org/10.1016/S0304-3800\(00\)00354-9](https://doi.org/10.1016/S0304-3800(00)00354-9)
- Hastie T. & Tibshirani R. (1990). *Generalized Additive Models*. Chapman and Hall, London.
- Hastie T., Tibshirani R. & Buja A. (1994). Flexible discriminant analysis by optimal scoring. *Journal of the American Statistical Association* **89**, 1255–1270. <https://doi.org/10.2307/2290989>
- Heino J. (2001). Regional gradient analysis of freshwater biota: Do similar biogeographic patterns exist among multiple taxonomic groups? *Journal of Biogeography* **28**, 69–76. <https://doi.org/10.1046/j.1365-2699.2001.00538.x>
- Heino J. & Alahuhta J. (2015). Elements of regional beetle faunas: Faunal variation and compositional breakpoints along climate, land cover and geographical gradients. *Journal of Animal Ecology* **84**, 427–441. <https://doi.org/10.1111/1365-2656.12287>
- Hellsten S. (2001). Effects of lake water level regulation on aquatic macrophyte stands in northern Finland and options to predict these impacts under varying conditions. *Acta Botanica Fennica*, (171) 1–47
- Hering D., Carvalho L., Argillier C., Beklioglu M., Borja A., Cardoso A.C., *et al.* (1986). Managing aquatic ecosystems and water resources under multiple stress - An introduction to the MARS project. *Science of the Total Environment* **67**, 1328–1340. <https://doi.org/10.1016/j.scitotenv.2014.06.106>
- Hilton J., O'Hare M., Bowes M.J. & Jones J.I. (2006). How green is my river? A new paradigm of eutrophication in rivers. *Science of the Total Environment* **365**, 66–83. <https://doi.org/10.1016/j.scitotenv.2006.02.055>
- Humboldt A. von & Bonpland A. (1805). *Essay on the geography of plants*. Chez Levrault, Schoell et compagnie, libraires, Paris.
- Johnson L., Richards C., Host G. & Arthur J. (1997). Landscape influences on water chemistry in Midwestern stream ecosystems. *Freshwater Biology* **37**, 193–208. <https://doi.org/10.1046/j.1365-2427.1997.d01-539.x>
- Johnson R.K., Goedkoop W. & Sandin L. (2004). Spatial scale and ecological relationships between the macroinvertebrate communities of stony habitats of streams and lakes. *Freshwater Biology* **49**, 1179–1194. <https://doi.org/10.1111/j.1365-2427.2004.01262.x>
- Jones J.I., Li W. & Maberly S.C. (2003). Area, altitude and aquatic plant diversity. *Ecography* **26**, 411–420. <https://doi.org/10.1034/j.1600-0587.2003.03554.x>
- Kanninen A., Vallinkoski V.M., Leka J., Marjomäki T.J., Hellsten S. & Hämäläinen H. (2013). A comparison of two methods for surveying aquatic macrophyte communities in boreal lakes: Implications for bioassessment. *Aquatic Botany* **104**, 88–100. <https://doi.org/10.1016/j.aquabot.2012.09.002>
- King K., Cheruvilil K.S. & Pollard A. (2019). Drivers and spatial structure of abiotic and biotic properties of lakes, wetlands, and streams at the national scale. *Ecological Applications* **29**, 1–13. <https://doi.org/10.1002/eap.1957>
- Kosten S., Kamarainen A., Jeppesen E., Van Nes E.H., Peeters E.T.H.M., Mazzeo N., *et al.* (2009). Climate-related differences in the dominance of submerged macrophytes in shallow lakes. *Global Change Biology* **15**, 2503–2517. <https://doi.org/10.1111/j.1365-2486.2009.01969.x>
- Lacoul P. & Freedman B. (2006). Environmental influences on aquatic plants in freshwater ecosystems. *Environmental Reviews* **14**, 89–136. <https://doi.org/10.1139/A06-001>

- Leroy B., Bellard C., Dubos N., Colliot A., Vasseur M., Courtial C., *et al.* (2014). Forecasted climate and land use changes, and protected areas: the contrasting case of spiders. *Diversity and Distributions* **20**, 686–697. <https://doi.org/10.1111/ddi.12191>
- Lind L., Nilsson C. & Weber C. (2014). Effects of ice and floods on vegetation in streams in cold regions: Implications for climate change. *Ecology and Evolution* **4**, 4173–4184. <https://doi.org/10.1002/ece3.1283>
- Liu C., Berry P.M., Dawson T.P. & Pearson R.G. (2005). Selecting thresholds of occurrence in the prediction of species distributions. *Ecography* **28**, 385–393. <https://doi.org/10.1111/j.0906-7590.2005.03957.x>
- Luoto M., Virkkala R. & Heikkinen R.K. (2007). The role of land cover in bioclimatic models depends on spatial resolution. *Global Ecology and Biogeography* **16**, 34–42. <https://doi.org/10.1111/j.1466-8238.2006.00262.x>
- McCullagh P. & Nelder J.A. (1989). *Generalized Linear Models, Second edition*. Chapman and Hall.
- Merritt D.M. & Wohl E.E. (2002). Processes governing hydrochory along rivers: Hydraulics, hydrology, and dispersal phenology. *Ecological Applications* **12**, 1071–1087. [https://doi.org/10.1890/1051-0761\(2002\)012\[1071:PGHARH\]2.0.CO;2](https://doi.org/10.1890/1051-0761(2002)012[1071:PGHARH]2.0.CO;2)
- Nakano S. & Murakami M. (2001). Reciprocal subsidies: Dynamic interdependence between terrestrial and aquatic food webs. *Proceedings of the National Academy of Sciences* **98**, 166–170. <https://doi.org/10.1073/pnas.98.1.166>
- Netten J.J.C., van Zuidam J., Kosten S. & Peeters E.T.H.M. (2011). Differential response to climatic variation of free-floating and submerged macrophytes in ditches. *Freshwater Biology* **56**, 1761–1768. <https://doi.org/10.1111/j.1365-2427.2011.02611.x>
- O'Hare M.T., Baattrup-Pedersen A., Nijboer R., Szoszkiewicz K. & Ferreira T. (2006). Macrophyte communities of European streams with altered physical habitat. *Hydrobiologia* **566**, 197–210. <https://doi.org/10.1007/s10750-006-0095-2>
- O'Reilly C.M., Rowley R.J., Schneider P., Lenters J.D., McIntyre P.B. & Kraemer B.M. (2015). Rapid and highly variable warming of lake surface waters around the globe. *Geophysical Research Letters* **42**, 10773–10781. <https://doi.org/10.1002/2015GL066235>
- Pajunen V., Luoto M. & Soininen J. (2016). Climate is an important driver for stream diatom distributions. *Global Ecology and Biogeography* **25**, 198–206. <https://doi.org/10.1111/geb.12399>
- Peterson A.T., Papeş M. & Soberón J. (2008). Rethinking receiver operating characteristic analysis applications in ecological niche modeling. *Ecological Modelling* **213**, 63–72. <https://doi.org/10.1016/j.ecolmodel.2007.11.008>
- Pirinen P., Simola H., Aalto J., Kaukoranta J.-P., Karlsson P. & Ruuhela R. (2012). *Tilastoja Suomen Ilmastosta 1981–2010 [Climatological statistics of Finland 1981–2010]*, Ilmatieteen laitos [Finnish Meteorological Institute].
- Pont D., Hugueny B. & Oberdorff T. (2005). Modelling habitat requirement of European fishes: do species have similar responses to local and regional environmental constraints? *Canadian Journal of Fisheries and Aquatic Sciences* **62**, 163–173. <https://doi.org/10.1139/f04-183>
- Potapova M.G. & Charles D.F. (2002). Benthic diatoms in USA rivers: Distributions along spatial and environmental gradients. *Journal of Biogeography* **29**, 167–187.

- <https://doi.org/10.1046/j.1365-2699.2002.00668.x>
- R Core Team (2017). *R: A language and environment for statistical computing*. version 3.4.3, R Foundation for Statistical Computing, Vienna, Austria.
- Rääpysjärvi J., Hämäläinen H. & Aroviita J. (2016). Macrophytes in boreal streams: Characterizing and predicting native occurrence and abundance to assess human impact. *Ecological Indicators* **64**, 309–318. <https://doi.org/10.1016/j.ecolind.2016.01.014>
- Ridgeway G. (1999). The state of boosting. *Computing Science and Statistics* **31**, 172–181. <https://doi.org/citeulike-article-id:7678637>
- Rodríguez-Merino A., García-Murillo P., Cirujano S. & Fernández-Zamudio R. (2018). Predicting the risk of aquatic plant invasions in Europe: How climatic factors and anthropogenic activity influence potential species distributions. *Journal for Nature Conservation* **45**, 58–71. <https://doi.org/10.1016/j.jnc.2018.08.007>
- Sand-Jensen K., Binzer T. & Middelboe A.L. (2007). Scaling of photosynthetic production of aquatic macrophytes - A review. *Oikos* **116**, 280–294. <https://doi.org/10.1111/j.2006.0030-1299.15093.x>
- Santamaría L. (2002). Why are most aquatic plants widely distributed? Dispersal, clonal growth and small-scale heterogeneity in a stressful environment. *Acta Oecologica* **23**, 137–154. [https://doi.org/10.1016/S1146-609X\(02\)01146-3](https://doi.org/10.1016/S1146-609X(02)01146-3)
- Schneider B., Cunha E.R., Marchese M. & Thomaz S.M. (2018). Associations between Macrophyte Life Forms and Environmental and Morphometric Factors in a Large Sub-tropical Floodplain. *Frontiers in Plant Science* **9**, 1–10. <https://doi.org/10.3389/fpls.2018.00195>
- Soininen J. & Weckström J. (2009). Diatom community structure along environmental and spatial gradients in lakes and streams. *Fundamental and Applied Limnology / Archiv für Hydrobiologie* **174**, 205–213. <https://doi.org/10.1127/1863-9135/2009/0174-0205>
- Szoszkiewicz K., Ciecierska H., Kolada A., Schneider S.C., Szwabińska M. & Ruszczyńska J. (2014). Parameters structuring macrophyte communities in rivers and lakes – results from a case study in North-Central Poland. *Knowledge and Management of Aquatic Ecosystems*, (415) 08. <https://doi.org/10.1051/kmae/2014034>
- Thuiller W., Georges D., Engler R. & Breiner F. (2016). *biomod2:Ensemble Platform for Species Distribution Modeling*. R package version 3.3-7.
- Toivanen M., Hjort J., Heino J., Tukiainen H., Aroviita J. & Alahuhta J. (2019). Is catchment geodiversity a useful surrogate of aquatic plant species richness? *Journal of Biogeography*. <https://doi.org/10.1111/jbi.13648>
- Toivonen H. & Huttunen P. (1995). Aquatic macrophytes and ecological gradients in 57 small lakes in southern Finland. *Aquatic Botany* **51**, 197–221. [https://doi.org/10.1016/0304-3770\(95\)00458-C](https://doi.org/10.1016/0304-3770(95)00458-C)
- Tuo Y., Duan Z., Disse M. & Chiogna G. (2016). Evaluation of precipitation input for SWAT modeling in Alpine catchment: A case study in the Adige river basin (Italy). *Science of the Total Environment* **573**, 66–82. <https://doi.org/10.1016/j.scitotenv.2016.08.034>
- Wetzel R.G. (2001). *Limnology: Lake and River Ecosystems (Third Edition)*. Academic Press, San Diego.
- Willis K.J. & Whittaker R.J. (2002). Species diversity - scale matters. *Science* **295**, 1245–1247. <https://doi.org/10.1126/science.1067335>

- Wisn M.S., Pottier J., Kissling W.D., Pellissier L., Lenoir J., Damgaard C.F., *et al.* (2013). The role of biotic interactions in shaping distributions and realised assemblages of species: Implications for species distribution modelling. *Biological Reviews* **88**, 15–30. <https://doi.org/10.1111/j.1469-185X.2012.00235.x>
- Xia C., Yu D., Wang Z. & Xie D. (2014). Stoichiometry patterns of leaf carbon, nitrogen and phosphorous in aquatic macrophytes in eastern China. *Ecological Engineering* **70**, 406–413. <https://doi.org/10.1016/j.ecoleng.2014.06.018>

Table 1 – Probability that each predictor had to be selected by a model, depending on if they were correlated with other predictor or not.

	Lakes	Streams
GDD5	0.5	0.5
January temperature	0.5	0.5
Precipitation	1	1
Conductivity	0.33	1
Alkalinity	0.33	1
pH	0.33	1
Total P	0.5	0.5
Total N	0.5	1
Turbidity	1	0.5
Colour	1	1
Area	1	NA
Width	NA	1

Figures

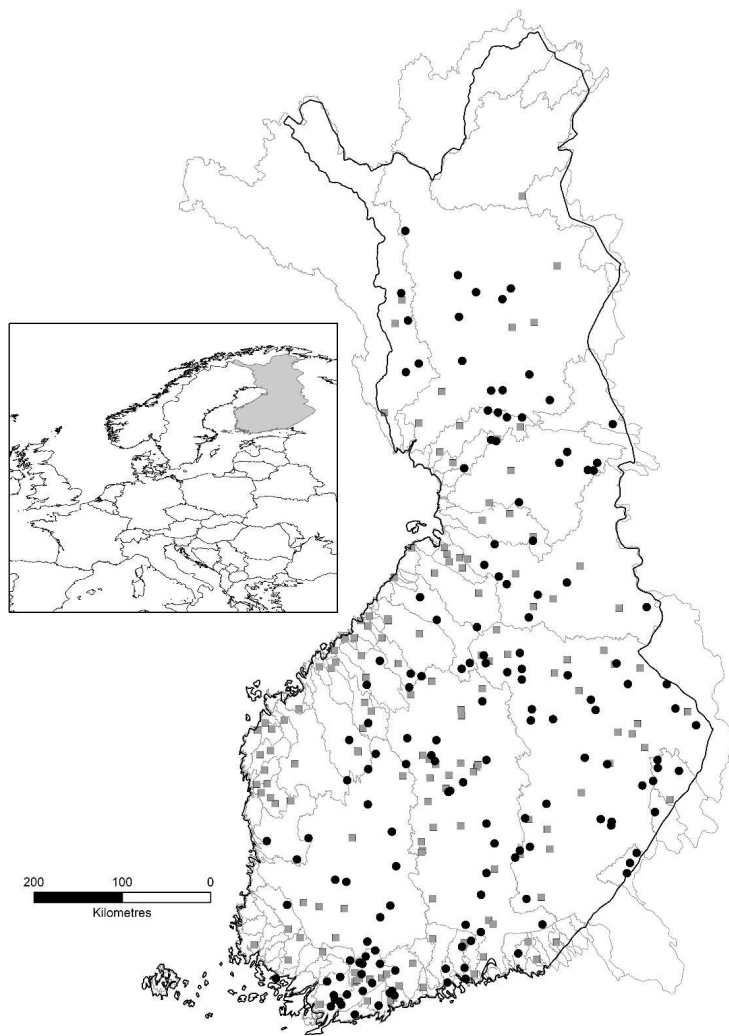


Fig. 1 Location of studied lakes and streams situated in different river basins across Finland. Gray squares represent the position of stream sites, and black circles are for lake sites

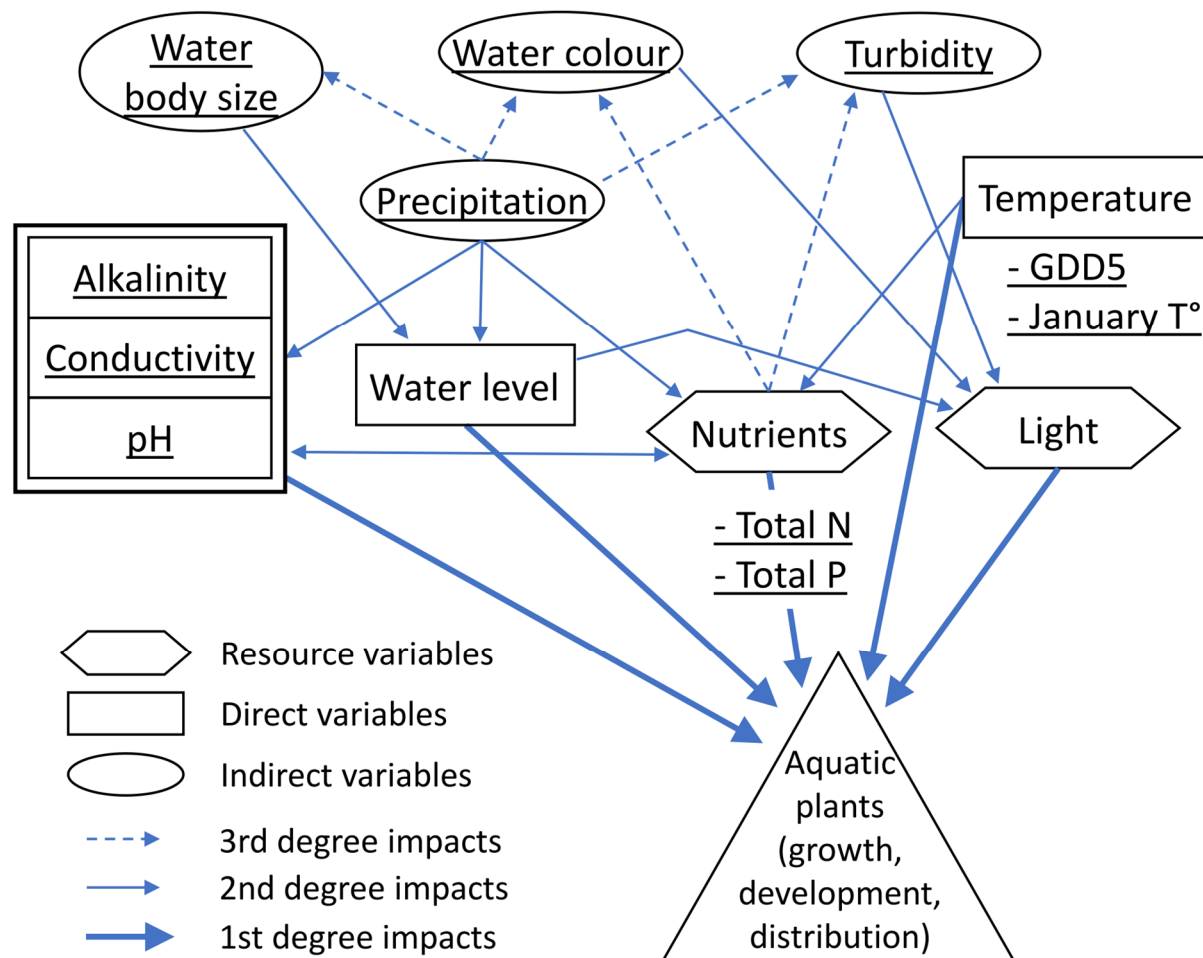


Fig. 2 Simplified conceptual framework of the relationships between resources, direct and indirect variables on aquatic plants. Underlined predictors are those that were used in the study. This figure was modified from Guisan & Zimmermann (2000) to be adapted to the case of aquatic plants and of our study

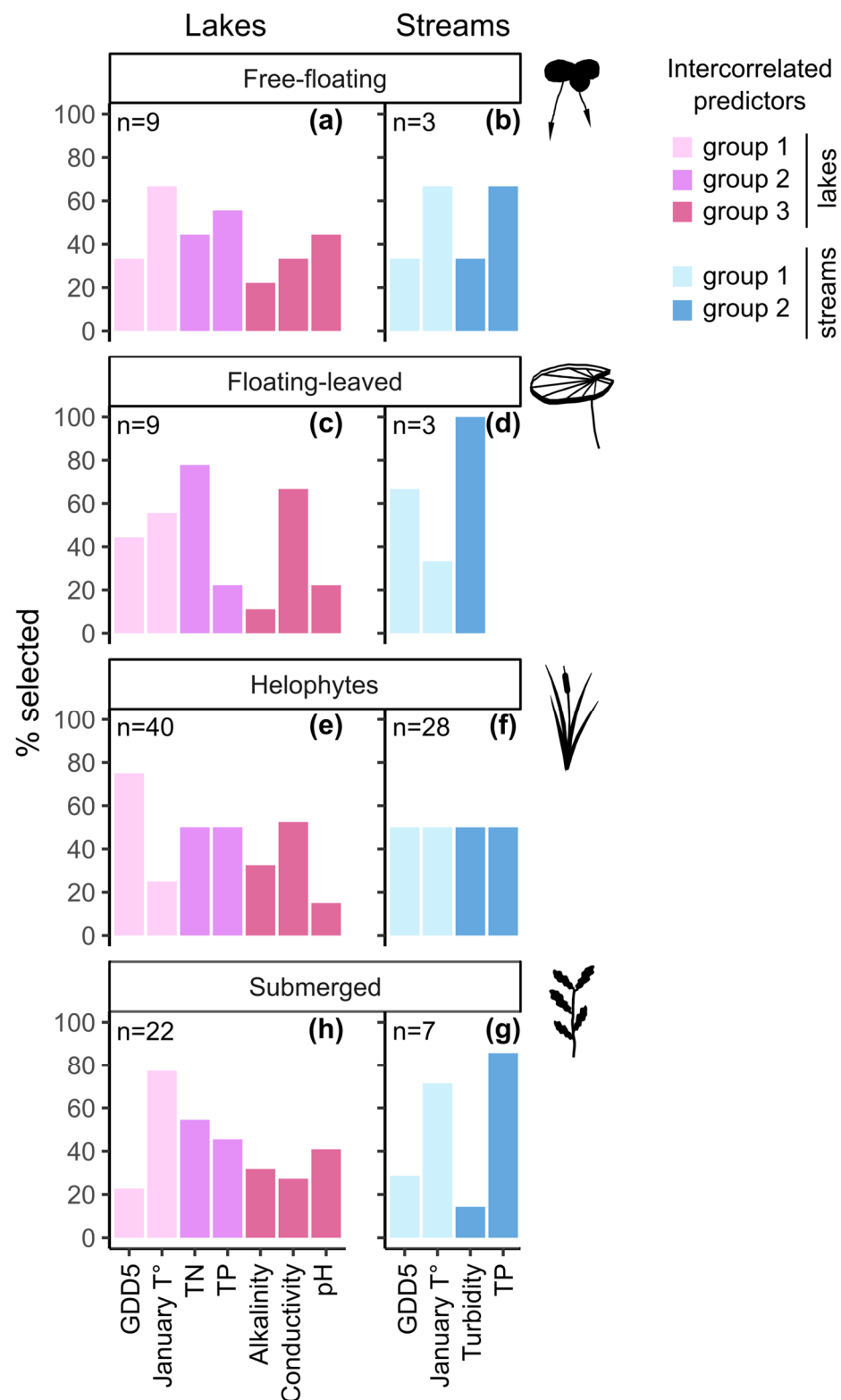


Fig. 3 Percentage of predictors selected within intercorrelated variables groups by the protocol of predictor selection, per life form, in lakes and streams. n=number of species in each life form group

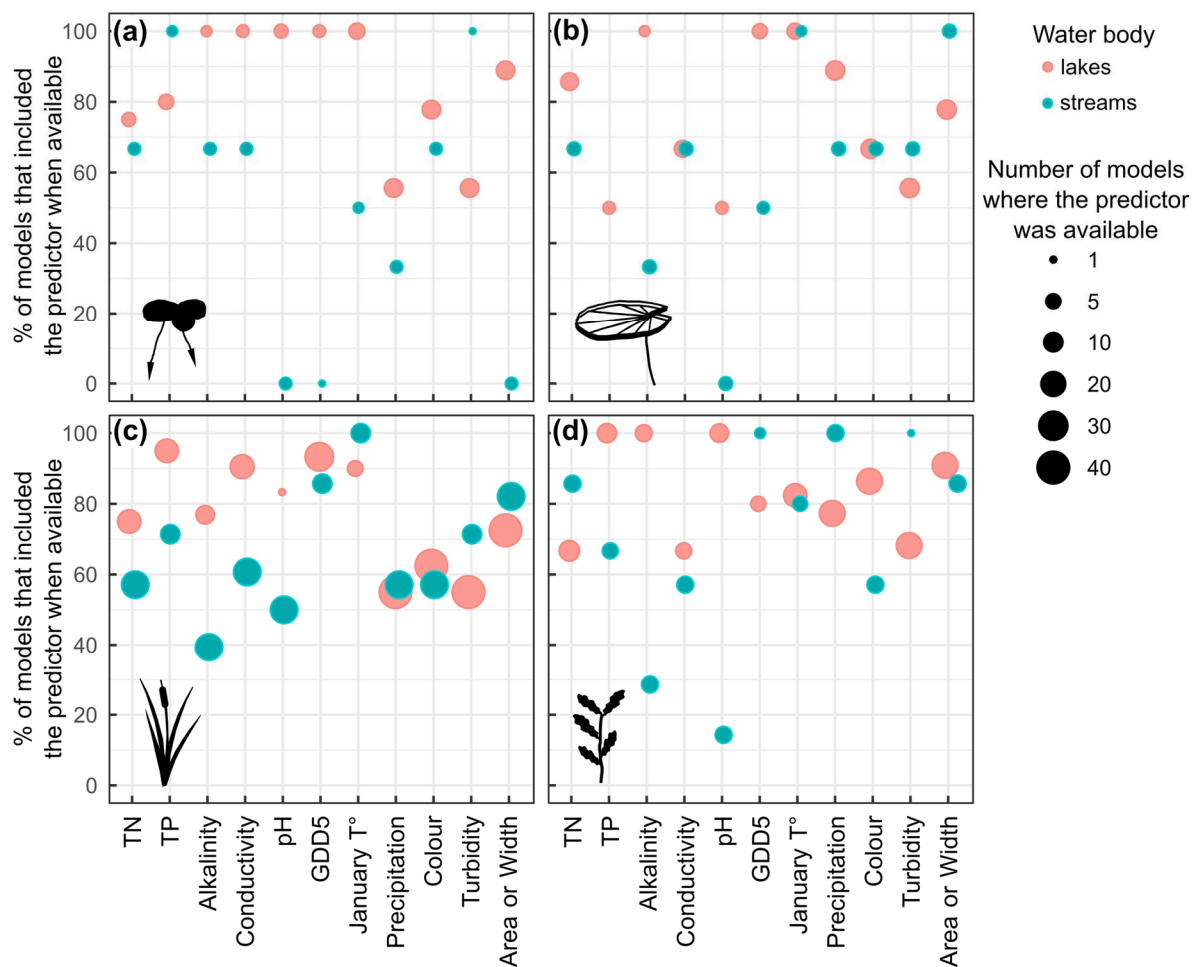


Fig. 4 Total percentage of models that selected a given predictor (when the predictor was available in predictors set), in relation to the number of models in which the predictor was available for selection, with results for all the species present in each water body category, presented by life form: free-floating (a), floating-leaved (b), helophytes (c), submerged (d) species. For each condition, the maximum number of models was that of the number of species present by life form type and by water body categories (see Fig. 3)

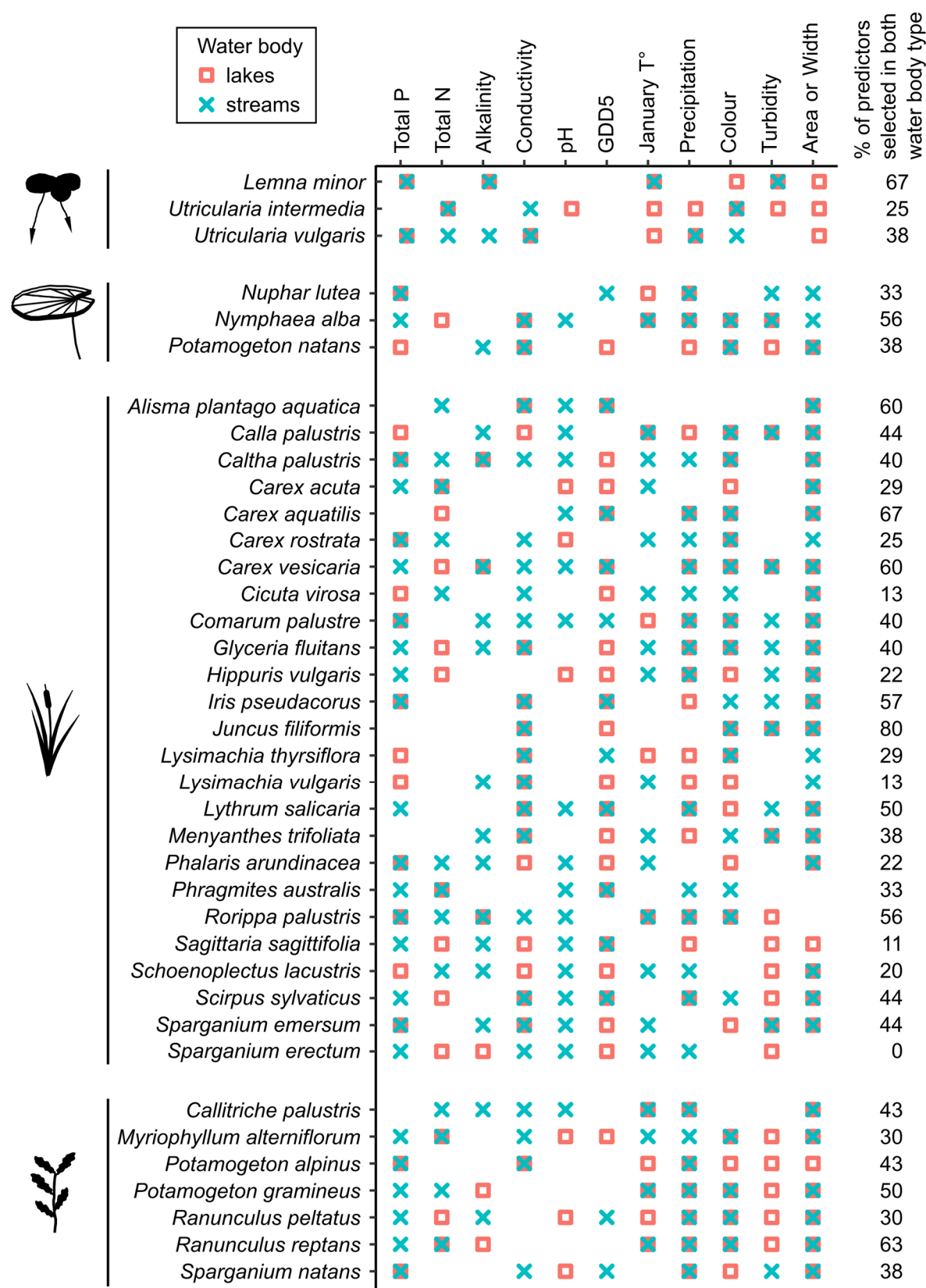


Fig. 5 Comparison of the predictor selected depending on the water body category for species shared between lakes and streams

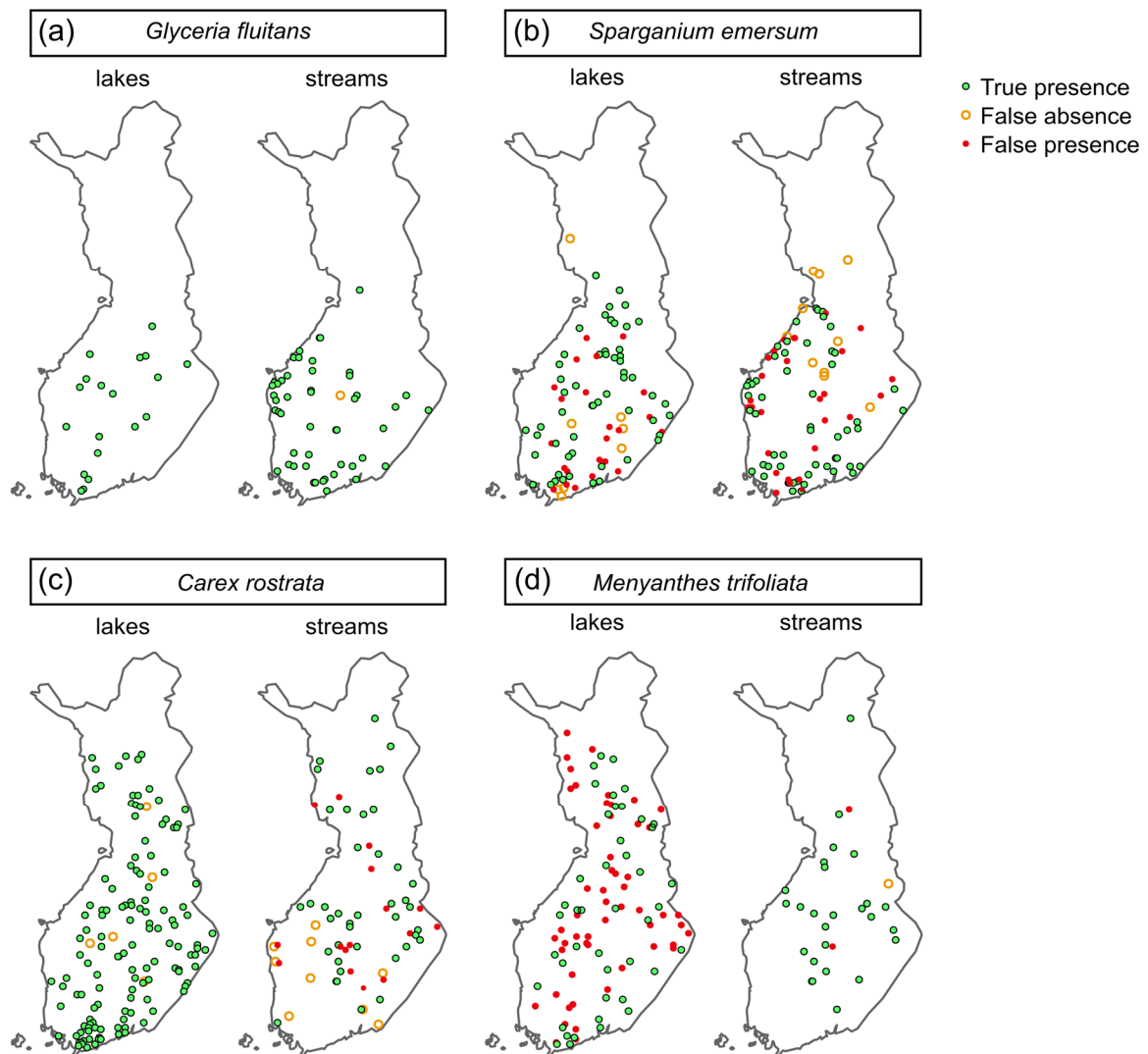


Fig. 6 Observed and predicted distribution of four species shared between lakes and streams. Predicted distribution were modelled with GAM and with each set of local environmental and climate predictors selected, specific to each species. (a) species well predicted for both habitats, (b) species not predicted well in neither habitat, (c) species well predicted in lakes but not in streams, and (d) species well predicted in streams, less in lakes