



OPEN

Distribution and co-occurrence patterns of charophytes and angiosperms in the northern Baltic Sea

Kristjan Herkül , Kaire Torn , Tiia Möller-Raid  & Georg Martin

The distribution data of 11 soft substrate charophyte and angiosperm species were analyzed. Our study aimed to elucidate the co-occurrence patterns among these sympatric macrophyte species and quantify their distribution areas. The central hypothesis of this study proposed that the observed co-occurrence patterns among the studied species deviate from what would be expected by random chance. Macrophyte occurrence data was derived from an extensive field sampling database. Environmental variables available as georeferenced raster layers including topographical, hydrodynamic, geological, physical, chemical, and biological variables were used as predictor variables in the random forest models to predict the spatial distribution of the species. Permutation tests revealed statistically significant deviations from random co-occurrence patterns. The analysis demonstrated that species tended to co-occur more frequently within their taxonomic groups (i.e., within charophytes and within angiosperms) than between these groups. The most extensive distribution overlap was observed between *Chara aspera* Willd. and *Chara canescens* Loisel., while *Zostera marina* L. exhibited the least overlap with the other species. The mean number of co-occurring species was the highest in *Chara baltica* (Hartman) Bruzelius while *Z. marina* had the largest share of single-species occurrences. Based on the distribution models, *Stuckenia pectinata* (L.) Börner had the largest distribution area.

Marine ecosystems are hard to access and thus there are many knowledge gaps, especially regarding biodiversity^{1,2}. For example, in many cases we still lack spatially continuous data on the distribution area of marine benthic species though information on the benthic community structure, composition, and environmental variables has been collected over decades. In the marine environment the species distribution models (SDM) have gained attention and popularity since the 1990s with significant increase through time^{3,4}; on the global scale the most studied groups have been fish, molluscs, and marine mammals⁵. A seamless map of distribution gives a significantly more relevant insight into species' spatial patterns than a simple plotting of sampling sites on a map. Particularly, when considering that field sampling sites are usually spatially unequally distributed over extensive areas. Distribution patterns and areal estimations of benthic habitat formers are also a viable part of successful nature conservation and marine spatial planning and serve the aims of many related conventions and directives, e.g., Convention on biological diversity⁶, The Convention for the Protection of the Marine Environment in the Northeast Atlantic⁷, The Convention on the Protection of the Marine Environment of the Baltic Sea Area⁸, Council Directive 92/43/EEC of 21 May 1992 on the conservation of natural habitats and of wild fauna and flora, Directive 2008/56/EC of the European Parliament and of the Council of 17 June 2008 establishing a framework for community action in the field of marine environmental policy (Marine Strategy Framework Directive), Recommendation of the European Parliament and of the Council of 30 May 2002 concerning the implementation of Integrated Coastal Zone Management in Europe.

Marine environment is under a severe human pressure and coastal benthic communities are not an exception^{9–11}. The shallow coastal waters are amongst the most productive ecosystems in the world and these habitats are also most affected by changes in oceanographic and atmospheric conditions and by various human impacts^{12–16}. For the shallow soft-bottom marine habitats the presence of submerged aquatic vegetation is crucial as it serves various structural and functional roles e.g., benthic primary production, enhancing biodiversity, habitat provisioning, providing coastal protection, and stabilizing sediments^{17,18}. Charophytes and angiosperms are the most characteristic species on the soft bottoms in the brackish Baltic Sea¹⁹; in total 23 species of charophytes

Estonian Marine Institute, University of Tartu, Mäealuse 14, 12618 Tallinn, Estonia. ✉ email: kristjan.herkul@ut.ee

and 91 species of angiosperms have been reported from this region²⁰. Both monospecies and multispecies communities of charophytes and angiosperms are common^{21–23}. Most of these species are of freshwater origin, somewhat adapted to brackish environments. Only *Zostera marina* L. represents the true marine species that is widespread also in lower salinity conditions^{24,25}. Charophytes and eelgrass *Z. marina* are considered to be most important habitat formers in the soft bottom shallow coastal areas of the Baltic Sea^{21,26}.

Decrease of submerged aquatic vegetation has been detected in both marine and freshwater ecosystems worldwide^{27–30}. Charophytes are sensitive to pollutants and eutrophication and on the European scale have undergone a severe decline³¹. Similarly, in the submerged angiosperms group, seagrasses are affected by a wide variety of human activities and the accelerating loss of seagrass ecosystems across the globe is threatening coastal ecosystems^{32,33}. Despite their significant ecological importance and declining distribution, research on the distribution and ecological preferences of submerged aquatic vegetation remains notably limited in comparison to studies on terrestrial vegetation^{34–38}. More effort is needed to evaluate the status and trends of the benthic habitats and species in order to fully understand the mechanisms behind the changes³³. However, the information on distribution patterns of soft bottom plant species is scarce and mostly in grey literature.

Set of abiotic environmental condition combined with biotic interactions, suitable for species to maintain the stable conditions in habitat, define the niche of the species^{39,40}. Niche width determines habitat specialization which may indicate the species vulnerability to global changes, as worldwide decline in specialist species has been noticed⁴¹. Knowledge about species specialization enables a better understanding of the mechanisms and consequences of environmental change, especially in brackish and estuarine environments where species live close to their physiological limits⁴². It is known that spatially and temporally variable salinity, underwater light environment, and availability of suitable substrate are the main drivers of submerged vegetation in estuaries and brackish water ecosystems^{19,43,44}. The distinct habitat preferences of charophytes and angiosperms along environmental gradients have been previously described by Herkül et al.⁴⁵. Although the abundance of charophytes and angiosperms increase proportionally with increase of soft substrate¹⁹, the distributions of individual species along the salinity gradient are dependent on species specific salinity tolerance^{43,45}. Previously, analysis of the niche breadth of the submerged soft-bottom macrophyte species showed a substantial variation of the niche breadths among the species. Furthermore, the angiosperms showed higher species-specific variability of niche breadths compared to those of the charophytes⁴⁵. Most of the work so far has been done on species level e.g.,^{46,47}, but studies comparing multidimensional niche overlaps of soft-bottom submerged vegetation are lacking. Further investigations involving the geographical aspects, for instance the species' distribution areas and co-occurrence patterns, are needed to gain a more complete understanding of the patterns and the mechanisms of the regional coexistence of aquatic macrophytes. To date, only a limited number of studies have explored the co-occurrence of aquatic macrophytes while examining the hypothesis of non-random patterns. However, these studies focus on freshwater systems and make inferences at the community level as a whole, e.g.,^{48–50}, rather than delving into species-specific pairwise patterns. There are only a few studies that investigate macrophytes to assess whether pairwise co-occurrence deviates from random patterns^{51,52} but these are also from freshwater environments. This study aims to address the research gap concerning co-occurrence patterns of soft substrate macrophytes within a brackish water environment. We highlight the importance of conducting pair-wise co-occurrence analysis as it serves as a fundamental step towards deducing the potential inter-specific mechanisms influencing the observed co-occurrence patterns.

This study focuses on spatial distribution patterns of characteristic plant species of soft-bottom habitat in the northern Baltic Sea region. Our central hypothesis suggested that the observed co-occurrence patterns among the studied species deviate from those expected by random chance. More specifically, our objectives were: (1) to provide an overview of the geographic distribution of angiosperms and charophytes in the Estonian sea area, northern Baltic Sea; (2) to elucidate the co-occurrence patterns of the studied charophytes and angiosperms, both within and between these groups. To address these objectives, we employed an extensive database of field inventories and species distribution modeling approach⁵³. The use of species distribution modeling was crucial for transitioning from site-specific data to spatially continuous datasets. This transition, in turn, facilitated the creation of species distribution maps and the quantification of species distribution areas.

Methods

Study area

The Baltic Sea is a tideless brackish water body. The phytobenthos data for this study originated from the Estonian marine area, northern Baltic Sea (Fig. 1). The study area included the full extent of the Estonian marine area from the coastline to the outer border of the exclusive economic zone (36,800 km²). The study area includes three large sub-basins of the Baltic Sea: The Baltic Proper, the Gulf of Finland, and the Gulf of Riga. The western study area around islands Saaremaa and Hiiumaa and the Gulf of Riga is characterized by extensive shallow coastal zone with numerous islands, islets, bays, and peninsulas. Contrastingly, the shallow coastal zone is limited due to steep coastal slope in the Gulf of Finland. Strong gradients of water depth, salinity and wave exposure exist in all sub-basins. The westernmost region of the study area is exposed to the open Baltic Proper with a wave fetch of hundreds of kilometers while the inner reaches of bays of the mainland are very sheltered, both by the mainland and by the islands. Likewise, the salinity gradient generally follows an east–west direction: salinity reaches seven in the Baltic Proper, while it falls to almost zero in the inner parts of the bays with riverine inflow. Hard limestone bedrock and granite boulders prevail in the most exposed areas. Mixed sediments of sand, gravel, and pebbles are typical in the mid-range of the wave exposure gradient. Fine sand and mud dominate in the most sheltered bays.

The abiotic environmental gradients drive the structure of macrophyte communities in the study area. Species of marine origin are common in the areas of the highest salinity, while freshwater species dominate in coastal areas with strong freshwater inflow. The brown alga *Fucus vesiculosus* L. and the red alga *Furcellaria lumbicalis*

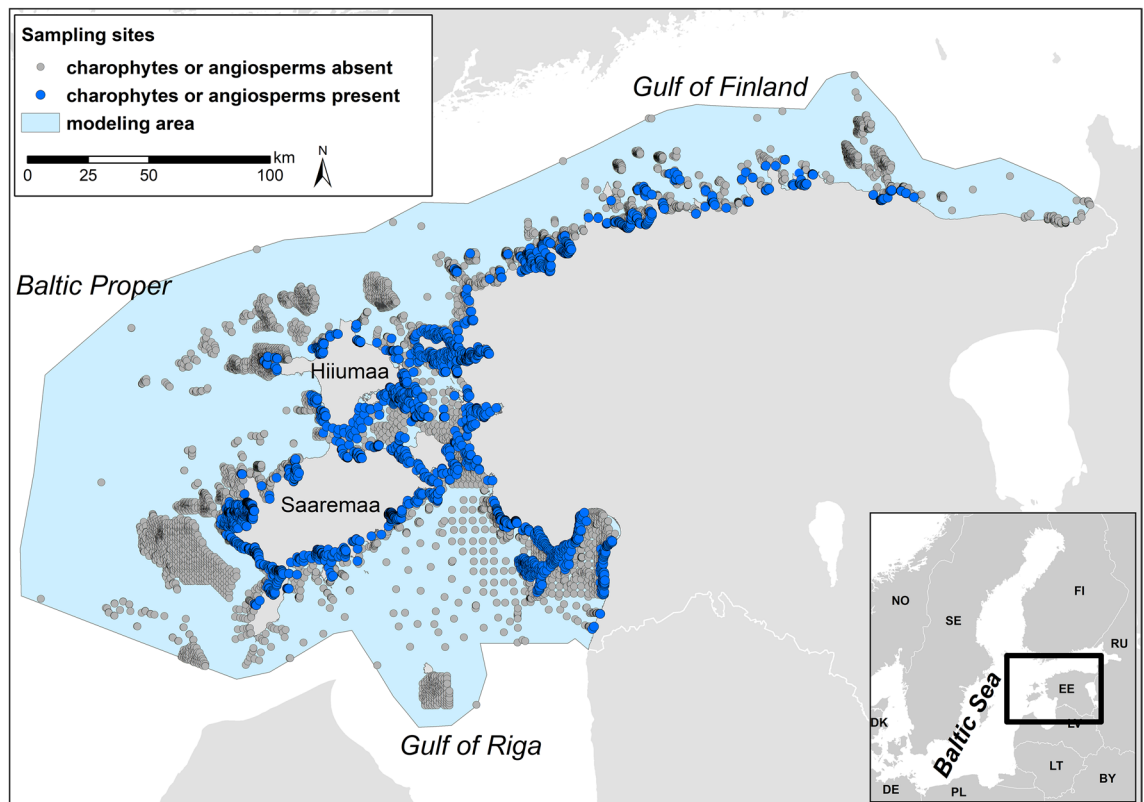


Figure 1. Study area and macrophyte sampling sites. The modeling area covered the full extent of the Estonian marine waters from the coastline to the outer border of the exclusive economic zone. *ESRI ArcMap* version 10.8.1 (<https://esri.com>) was used to generate this map.

(Hudson) J.V. Lamouroux are the most important perennial habitat-forming algal species on hard substrate. Annual and perennial filamentous green, brown, and red algae like *Ulva intestinalis* L., *Cladophora glomerata* (Linnaeus) Kützing, *Pylaiella littoralis* (Linnaeus) Kjellman, *Ceramium tenuicorne* (Kützing) Waern, *Vertebrata fucoides* (Hudson) Kuntze are common on hard substrate. Several species of angiosperm plants are common on soft substrate, e.g., *Zostera marina* L., *Stuckenia pectinata* (L.) Börner, *Potamogeton perfoliatus* L., *Zannichellia palustris* L., *Myriophyllum spicatum* L. and *Ruppia maritima* L. The charophytes (*Chara* spp. L., *Tolypella nidifica* (O.F.Müll.) Leonh.) are common on sandy and muddy sediments in shallow waters of sheltered bays. As a result of eutrophication extensive growth of annual filamentous algae and formation of drift algal mats are common phenomena in the study area⁵⁴.

Biological sampling

Data for the study originated from a macrobenthos database of the Estonian Marine Institute, University of Tartu. The database stores an extensive amount of macrobenthos data starting from 1950s and including data from more than 55,000 benthic samples. The database incorporates georeferenced benthos data originating from various projects like basic scientific research, national marine monitoring, benthic habitat mapping studies, surveys related to environmental impact assessments etc. Data from the database was filtered to include only years 2005–2020 because samples from that period represent methodologically coherent dataset with good spatial coverage. Both biomass and percentage coverage data were used in this study. The quantitative biomass samples were collected using Ekman and Van Veen type bottom grab samplers or by scuba divers, who collected all the flora and fauna inside a 0.04 m² metal frame. The samples were stored deep frozen (−18 °C) until analysis. In a laboratory, the samples were sorted and all macrobenthic organisms were identified under a microscope. Dry weights of all taxa were measured after two weeks drying at 60 °C. Biomasses were calculated per square meter. Data from replicate biomass samples (n = 3) of each sampling station were averaged. Coverage estimates were done either by scuba divers who assessed the percentage cover of all plant species or by using the so-called drop-camera, which was let down above the bottom of the sea by a cable. Videos were later analyzed on computer screen by estimating the average percentage coverage of substrate types and coverage of benthic macrophyte species seen in a video recording. Both video estimates and in situ estimates by scuba divers covered an area of about 2 m radius. Data from visual estimates were used only if visual information was sufficient for species level determination or the species determinations were confirmed by biomass samples. For the purposes of this study, the biomass and percentage cover data of macrophytes was transformed into binary occurrence data (0—absence, 1—presence). Specifically, if a species was detected in a sample, it was recorded as a presence, irrespective of the species' quantity (cover or biomass). Conversely, if a species was not detected, it was recorded as an absence.

As the aim of the study was to address the geographical distribution of macrophytes, the temporal replicates from sampling sites, that were visited several times, were removed and only the data from the most recent sampling event were retained in the dataset. To eliminate spatial redundancy in the data, we applied a 25 m equirectangular grid to the study area. In cases where multiple observations fell within the same grid cell, we retained the data from the sampling site with the most recent year of data collection. If there were multiple sampling occasions in the most recent year, we randomly selected one sample for inclusion in the dataset. The application of these temporal and spatial filters resulted in data from 14,356 unique field sites (Fig. 1), with a total of 20 charophyte and angiosperm species. The species with a very low occurrence rates (< 100 sites) were excluded from the dataset to ensure comparable analytical results. Altogether 11 species were included in this study (Table 1).

Environmental variables

Altogether 14 environmental variables including topographical (depth, slope of seabed), hydrodynamic (wave exposure, currents, ice conditions), geological (seabed substrate), physico-chemical (temperature, salinity, transparency, nutrients), and biological (water chlorophyll concentration) variables were used in this study (Table 2, Appendix 1). Long-term mean values were used for variables with time series, because general spatial distribution patterns not temporal variation (seasonal, annual) of macrophytes was addressed in this study. All

Species	Clade	Environment	Salinity range
<i>Myriophyllum spicatum</i> L.	Angiospermae	B/F	0–10 ⁴⁴
<i>Potamogeton perfoliatus</i> L.	Angiospermae	B/F	0–12 ⁴⁵
<i>Ruppia maritima</i> L.	Angiospermae	B	0–55 (65) ⁴⁶
<i>Stuckenia pectinata</i> (L.) Börner	Angiospermae	B/F	0–19 ⁴⁷
<i>Zannichellia palustris</i> L.	Angiospermae	B/F	0–20 ⁴⁸
<i>Zostera marina</i> L.	Angiospermae	B	(2.5)5–35(40) ²⁴
<i>Chara aspera</i> Willd.	Charophyta	B/F	0–15 (20) ^{49,50}
<i>Chara baltica</i> (Hartman) Bruzelius	Charophyta	B	2–10 (18) ⁵¹
<i>Chara canescens</i> Loisel	Charophyta	B	1.5–21 ⁵²
<i>Chara connivens</i> Salzm. ex A. Braun	Charophyta	B/F	0–9 ²⁸
<i>Tolypella nidifica</i> (O.F.Müll.) Leonh.	Charophyta	B	(0)4–24 ⁵³

Table 1. List of the studied species. The environmental information (B—brackish water, F—fresh water) is given based on Estonian data^{55,56}.

Variable	Abbreviation	Source
Water depth (m)	depth	1
Mean water depth in 2 km radius (m)	depth_2km	1
Slope of seabed (°)	slope	1
Slope of seabed in 2 km radius (°)	slope_2km	1
Salinity (PSU): long-term (2002–2008) mean	salinity	2,3
Wave exposure based on simplified wave model (m ² s ⁻¹): based on long-term (2003–2019) mean wind speeds and directions	wave	4
Current velocity (m s ⁻¹): long-term (2005–2019) mean	current	5
Water temperature in warm season (°C): long-term (2005–2019) mean in May–October	temp_warm	5
Water temperature in cold season (°C): long-term (2005–2019) mean in November–April	temp_cold	5
Concentration of nitrates (mmol m ⁻³): long-term (2005–2019) mean values in winter (December–February)	nitrate	6
Concentration of phosphates (mmol m ⁻³): long-term (2005–2019) mean values in winter (December–February)	phosphate	6
Chlorophyll a content of sea surface (mg m ⁻³): long-term (2005–2019) mean	chl	6
Proportion of ice cover (0...1): long-term (2000–2016) mean	ice	7
Proportion of soft sediment (0...1)	sediment	2

Table 2. Environmental variables. See Appendix 1 for schematic overview maps of the general spatial patterns of the variables. Sources: 1—Bathymetric data by Estonian Maritime Administration (georeferenced depth raster with 10 m pixel size). 2—Databases of the Estonian Marine Institute, University of Tartu. 3—COHERENS ocean circulation model⁵⁷. 4—Simplified wave model based of fetch and wind data^{58,59}. 5—Copernicus Marine Environment Monitoring Service's Baltic Sea Physics Reanalysis product BALTICSEA_REANALYSIS_PHY_003_011 (https://resources.marine.copernicus.eu/product-detail/BALTICSEA_REANALYSIS_PHY_003_011). 6—Copernicus Marine Environment Monitoring Service's Baltic Sea Biogeochemical Reanalysis product BALTICSEA_REANALYSIS_BIO_003_012 (https://resources.marine.copernicus.eu/product-detail/BALTICSEA_REANALYSIS_BIO_003_012). 7—Ice data developed by the Marine Systems Institute, Tallinn University of Technology⁶⁰.

environmental variables were available as georeferenced raster layers in GeoTIFF format. The extents of raster layers fully covered the study area. Raster cell size was 10 m for depth and seabed slope and 50 m for all the other variables. The environmental variables were selected based on previous knowledge on potential relationships with the distribution of the studied macrophytes, e.g.^{45,55}, and data availability.

Data analysis

R programming language version 4.2.2⁶¹ in the development environment *RStudio*⁶² was used for data preparation, analysis, and modeling. In addition to the base *R* functionality, the following packages were used: *raster*⁶³ for handling spatial raster data, *sf*⁶⁴ for handling spatial vector data, *randomForest*⁶⁵ for fitting random forest models, *caret*⁶⁶ for creating folds for cross-validation and for calculating Kappa coefficients, *PresenceAbsence*⁶⁷ for converting probability of occurrence to presence-absence, *tidyverse*⁶⁸ for tabular data manipulation and plotting. *ArcMap* software⁶⁹ was used to generate the map of the study area.

Species distribution modeling (SDM) framework⁵³ was used to generate seamless distribution maps of the studied macrophyte species. Random forest (RF) was chosen as the modeling method because it offers many advantages over alternative methods: a) RF has proven to be one of the most accurate methods in SDM; b) RF boasts a minimal number of user-settable parameters, with default settings consistently producing reliable results; c) random selection of predictor variables at each node mitigates the risk of overfitting; d) RF excels at directly handling the modeling of binary, continuous, and nominal variables^{65,70–79}. RF is a machine learning method that generates a large number of regression trees, each calibrated on a bootstrap sample of the original data⁵⁸. Each node is split using a subset of randomly selected predictors and the tree is grown to the largest possible extent without pruning. RF can handle both regression and classification type of modeling. For predicting a value of a new data point, the data are run through each of the tree in the forest and each tree provides a value. The model prediction is then calculated as the average value over the predictions of all the trees in the forest in regression model and by majority vote in classification model.

The species presence-absence data from the 14,356 field sampling sites (Fig. 1; see the section Biological sampling) was used as a response variable in RF models. We employed both regression-type and classification-type RF models as candidate models for each species. The motivation for fitting both regression and classification models was to identify the optimal RF model type based on their performance characteristics. The values of the 14 environmental variables (Table 1; see section Environmental variables) served as predictor variables. A separate model was built for each macrophyte species. The models were then used to generate spatial distribution prediction for each species. Using the trained RF models, species' predictions were calculated for each data point in the prediction dataset ($n = 3,685,338$) covering the whole Estonian marine area from the coastline to the outer border of the exclusive economic zone (36,800 km²) with 100 m rectangular equispaced grid (see modeling area in Fig. 1). The values of all environmental variables (predictor variables *sensu* predictive modeling) were available for each point in the prediction dataset. Given the binary nature of the input data (0 for absence, 1 for presence), the RF regression-type model generated probability estimates of occurrence as predictions, falling within the range of zero to one (encompassing any decimal value between zero and one). Conversely, in the classification-type RF model, the prediction outcome was a binary classification, with values of zero and one. The predictions of regression-type models (probabilities of occurrence) of each species were separately converted to binary presence-absence classes (1—presence, 0—absence) using the sensitivity–specificity difference minimizer method that has been found to produce more accurate results compared to other methods^{80–82}. The conversion of regression-type RF outcome to binary presence-absence was done in order to estimate the distribution area of the species and to compare the predictive performance with that of classification-type RF outcome.

In RF model fitting, the function *randomForest* in package *randomForest* was used. Two parameters must be set in *randomForest*: the number of predictor variables to be randomly selected at each node (*mtry*) and the number of trees in a forest (*ntree*). The *mtry* parameter was retained at its default values: one-third of the number of predictor variables for regression-type RF and the square root of the number of predictor variables for classification-type RF models, in accordance with the recommendations provided by Liaw and Wiener⁶⁵. *ntree* was set to 1000 as already 500 trees usually yield stable results⁶⁵. In all models, we verified the adequacy of using 1000 trees to achieve stable results by employing the *plot* function within the *randomForest* package. The importance of predictor variables was calculated using the permutation test inside the *randomForest* function by specifying *importance = TRUE* and *nPerm = 10*. All other arguments in the *randomForest* function were left as default values. Further information on the *R* code is shown in Appendix 2.

The predictive performance of the models was evaluated using tenfold cross-validation (CV)^{83,84}. Using CV, area under the receiver operating curve (AUC)⁸⁵ and Cohen's kappa^{86,87} were calculated for regression-type RF models and only Cohen's kappa for classification-type models. For calculating kappa in regression-type models, the probability of occurrence was converted into binary classes (1—presence, 0—absence) using the sensitivity–specificity difference minimizer method⁸². Stratified random sampling without replacement was used to generate data folds for each species using the presence-absence of a given species as strata to ensure equal proportions of presences in all folds⁸³. Using ten folds resulted in a division of 90% data for model training and 10% of data for validation at each iteration. Tenfold cross-validation was done only to estimate the predictive performance of models. Final models for generating the spatial predictions were trained using 100% of the data. Mean AUC and kappa values over ten folds were used as the estimate of predictive performance. AUC values vary between 0.5 and 1. AUC value of 0.5 indicates that a model performs the same as random chance while the value of 1 indicates perfect model prediction^{85,88}. There is no universal evaluation criteria of AUC values between 0.5 and 1, but often the values 0.5–0.6 are considered “poor”, 0.6–0.7 “fair”, 0.7–0.8 “good”, 0.8–0.9 “very good”, 0.9–1 “excellent”^{74,89}. Following Landis and Koch⁸⁷, the Cohen's kappa coefficient value below zero

indicates “poor”, 0–0.2 “slight”, 0.21–0.4 “fair”, 0.41–0.6 “moderate”, 0.61–0.8 “substantial” and 0.81–1 “almost perfect” strength of agreement.

Relationships between species richness (number of species) and the environmental variables were depicted using generalized additive models (GAM)⁹⁰ with cubic regression splines. The degrees of freedom were limited to five in GAMs to avoid overfitting. Separate GAM fits were produced for each species richness–environmental variable pair for plotting purposes.

Two different methods were applied to assess the distribution overlap among the studied species: Schoener’s D index and proportion of co-occurrence in sampling sites. Species occurrences in the field data was used as the input for both methods. Schoener’s D can vary between zero and one: zero indicates no distribution overlap while one indicates full distribution overlap (i.e. identical distributions) between species. While there are several metrics available for estimating distribution overlap, Schoener’s D was chosen due to its good performance, simplicity, and long history of use^{91,92}. While Schoener’s D enables to generate a single value for estimating the spatial overlap of each pair of the studied species, it cannot give full insight into the overlap patterns because a distribution overlap between two species is seldom symmetrical and the sizes of distribution areas vary. For example, the Schoener’s D value of 0.5 may mean that the distribution area of a species A is fully (100%) inside the distribution area of a species B while only 50% of the distribution of the species B coincides with that of A. This is possible given that the distribution area of B is twice the size of A. To account for the asymmetry, for each species, percentage overlap of distribution with all the other species was also calculated. Permutation test⁹³ with 100,000 permutations was conducted for each pair of species to assess whether their co-occurrence, as measured by the Schoener’s D value, deviated significantly from what would be expected by random chance⁹⁴. By randomly shuffling the field sites and calculating the Schoener’s D based on the 100,000 iterations, the so called null distribution⁹⁵ of D values for each pair of species was produced. The permutationally generated null distribution of D values was subsequently compared to the observed D value. The null hypothesis of random co-occurrence was rejected when the observed D value fell outside a predefined percentile range (e.g., 2.5% to 97.5% with a p-value threshold set at 0.05), signifying a substantial departure from random chance⁹³. To account for multiple comparisons, the p-values were corrected using the Benjamini–Hochberg method⁹⁶. R code for the permutation test is shown in Appendix 2. Kruskal–Wallis test⁹⁷ was used to assess statistical differences of overlap both within and between the species groups based on the pair-wise D values with the following group levels: within charophytes, within angiosperms, between charophytes and angiosperms.

Research involving human and/or animal participants

The study did not involve animal or human participants.

Results

Based on the Cohen’s kappa coefficients of tenfold cross-validation, the average prediction accuracy of regression-type RF models was marginally higher than that of classification-type models (Table 3) and subsequently the regression-type RF models were selected to produce the maps of species distribution. The AUC values of all species were over 0.89 indicating high prediction accuracy while the model performance measured using kappa coefficients indicated mainly fair to moderate agreement between predicted and observed species occurrences (Table 3). Notably low Cohen’s kappa coefficient of 0.15 was found in *T. nidifica*. Depth was the most influential environmental variable in the regression-type RF models, followed by proportion of soft sediment and slope of seabed (Appendix 3).

Species	Regression RF		Classification RF
	AUC	kappa	kappa
<i>C. aspera</i>	0.963	0.529	0.498
<i>C. baltica</i>	0.943	0.274	0.213
<i>C. canescens</i>	0.95	0.36	0.288
<i>C. connivens</i>	0.947	0.275	0.236
<i>T. nidifica</i>	0.894	0.156	0.127
<i>M. spicatum</i>	0.948	0.507	0.469
<i>P. perfoliatus</i>	0.955	0.462	0.424
<i>R. maritima</i>	0.926	0.348	0.309
<i>S. pectinata</i>	0.948	0.645	0.632
<i>Z. palustris</i>	0.901	0.323	0.285
<i>Z. marina</i>	0.943	0.427	0.351
Average	0.938	0.391	0.348

Table 3. Results of tenfold cross-validation of regression-type and classification-type RF models. Average values of the statistics over all species are shown in the last row. AUC—area under the receiver operating curve, kappa—Cohen’s kappa coefficient.

The species richness of charophytes sharply decreased with increasing depth while the richness of angiosperms exhibited the highest values between 0 to 2 m. However, generally, the species richness of both angiosperms and charophytes followed similar patterns along the environmental gradients (Fig. 2).

Based on the modeled species distributions, *S. pectinata* had by far the largest distribution area with 1125 km². *C. aspera* ranked the second with 451 km², all other species had distribution area below 400 km². *C. baltica* had the smallest area with 26 km² (Fig. 3). The majority of the species were spread along the whole coastline while *C. baltica* and *C. connivens* were mainly restricted to western Estonia. Unlike the other species, the presence of *Z. marina* in the Gulf of Riga was limited. Compared to angiosperms, the distribution of charophytes was limited to shallower areas (except for *T. nidifica*) and sheltered bays (Fig. 4).

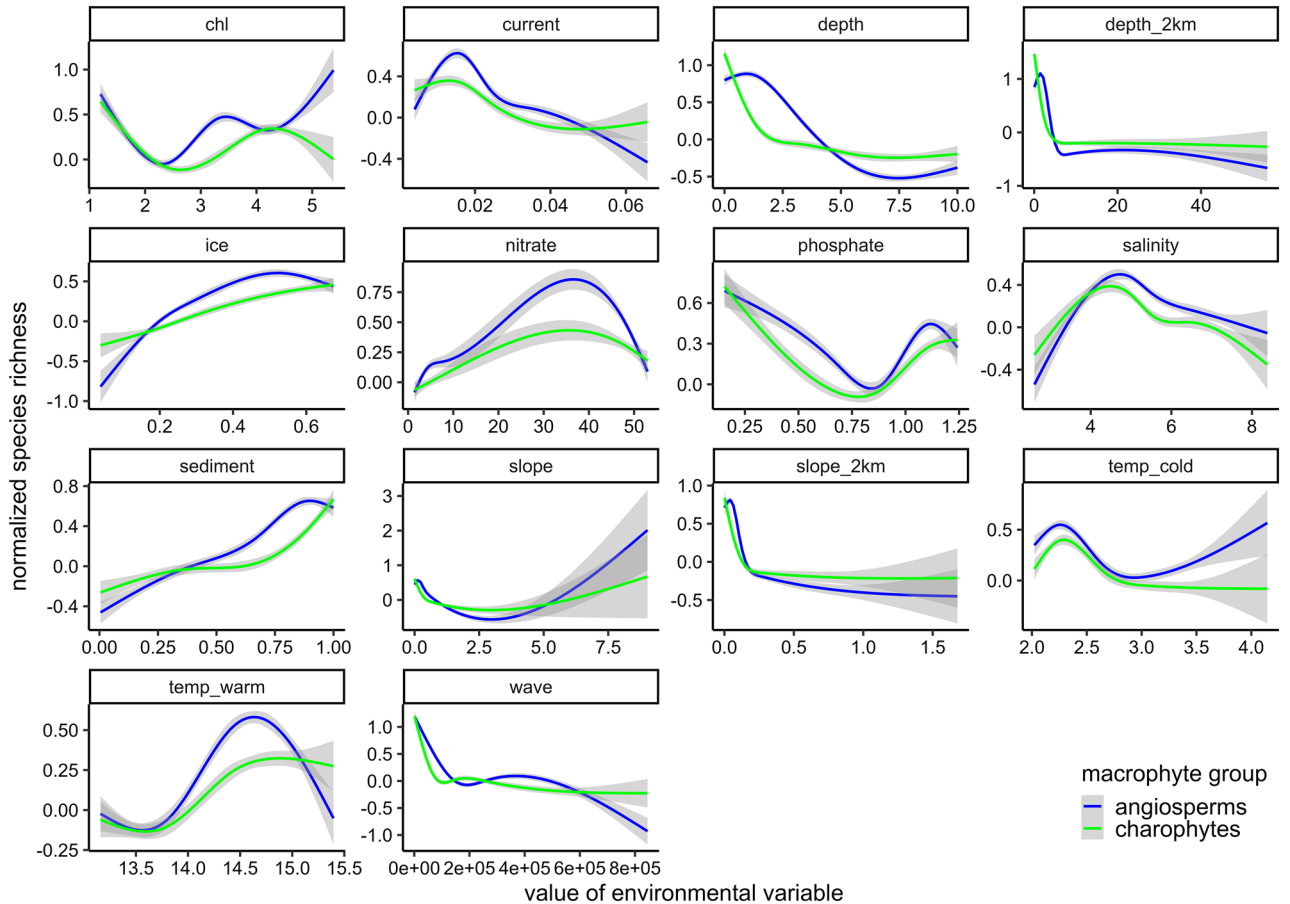


Figure 2. Variability of the species richness of angiosperms and charophytes along the environmental gradients as fit by GAMs. 95% confidence interval is shown in gray. Data from sampling sites with water depth exceeding 10 m is excluded.

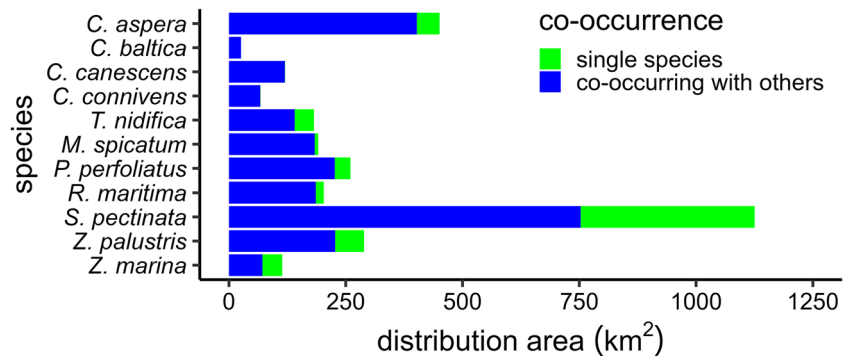


Figure 3. Distribution area of the studied species based on distribution modeling. Colors indicate the distribution area of a species occurring alone (single species) or together with at least one other studied species.

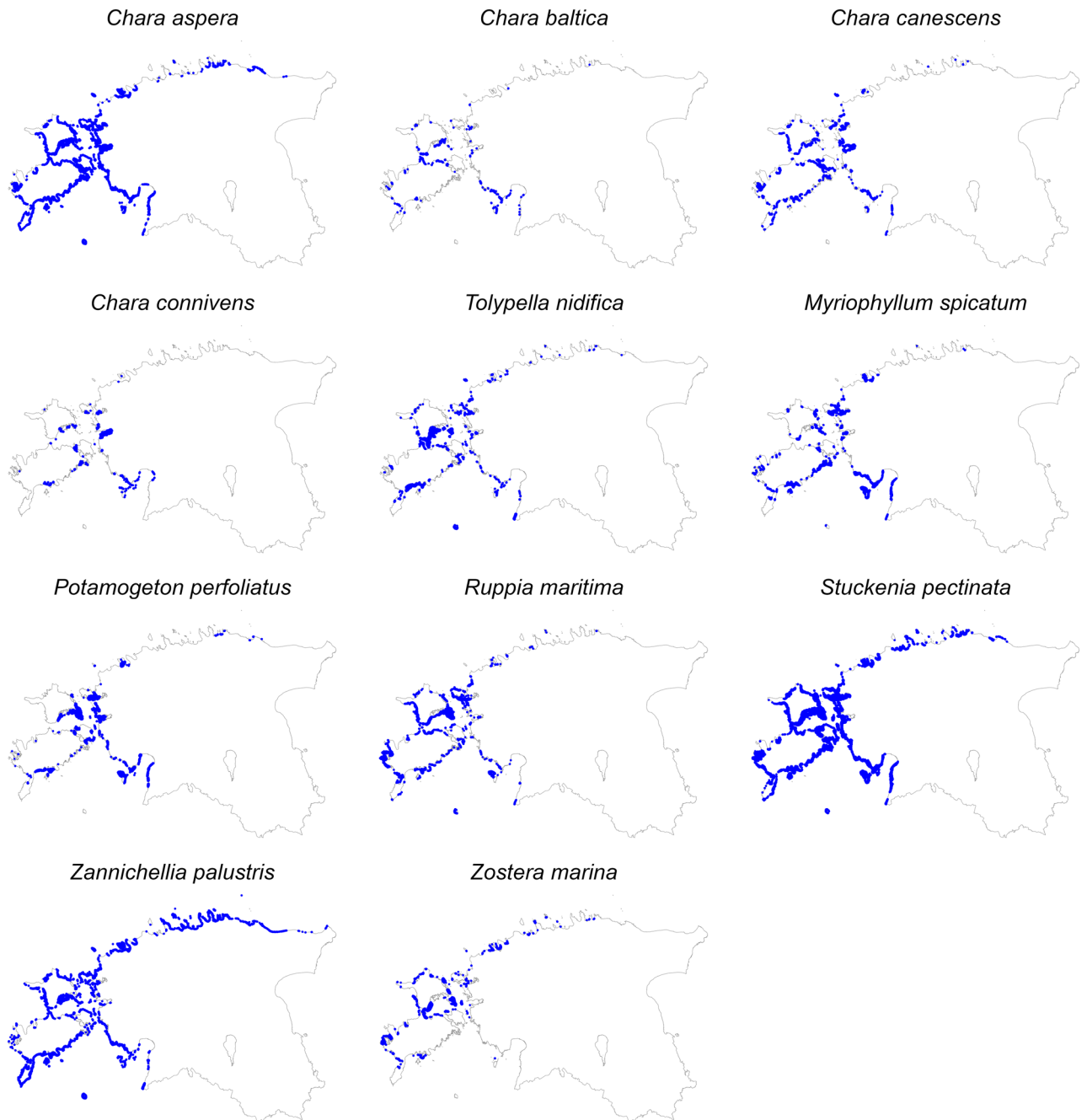


Figure 4. Modeled distributions of the studied species. Blue color represents the occurrence of a species. The size of the blue symbols of occurrence is increased compared to the original 100×100 m pixel size to improve readability. R programming language version 4.2.2 (<https://www.r-project.org/>) in the development environment RStudio (<https://posit.co/products/open-source/rstudio/>) was used to generate this map.

SDM predictions showed that 29% of the modeling area was occupied by co-habitation of charophytes and angiosperms (i.e., at least one charophyte and one angiosperm present in a grid cell). 61% of the modeling area was occupied exclusively by angiosperms (i.e. at least one angiosperm present and no charophytes) while only 10% was occupied exclusively by charophytes (i.e. at least one charophyte present and no angiosperms).

Based on the field data, the highest distribution overlap was between *C. aspera* and *C. canescens* (Schoener's $D = 0.39$) followed by *C. aspera* and *Z. palustris* (Schoener's $D = 0.35$). *C. aspera* also had the largest mean overlap with all the other species. *Z. marina* had the smallest overlap with the other species (Fig. 5). Based on permutation tests, it was found that 41 out of the 55 unique pairs of species exhibited D values that were statistically different ($p < 0.05$) from what would be expected by random chance. The group-wise mean D values showed statistically significant differences among the species groups (Kruskal–Wallis test, $p < 0.05$). Specifically, the mean D value within charophyte species was the highest at 0.24, followed by a mean D value of 0.16 within angiosperms, and the mean D value between charophytes and angiosperms was the lowest at 0.11.

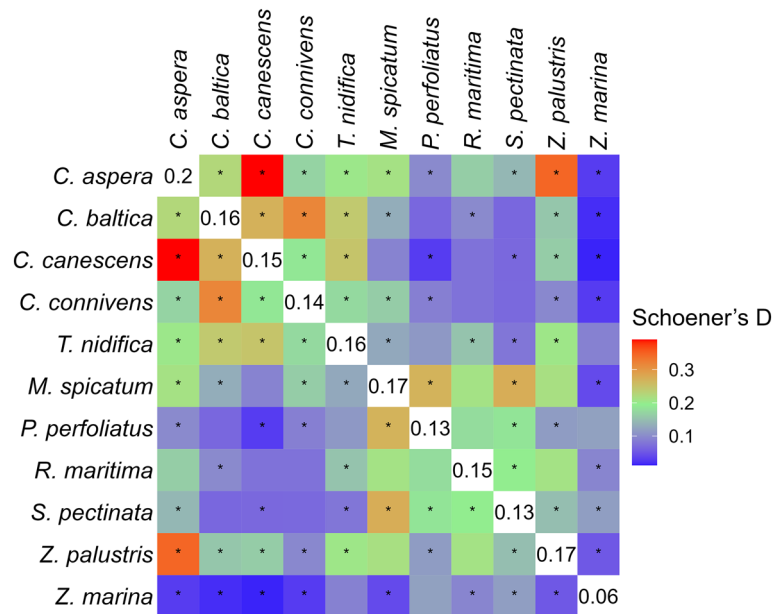


Figure 5. Occurrence overlap between the studied species based on field observations. Higher Schoener's D value indicates larger overlap. The numbers in the diagonal represent average D values of the species. Asterisks indicate statistically significant D values ($p \leq 0.05$, corrected using Benjamini–Hochberg method for multiple comparisons). A statistically significant value indicates that the observed D value differs from what would be expected by random chance.

Charophytes *C. baltica*, *C. canescens*, and *C. connivens* almost exclusively (>99% of their distribution area) co-occurred with other species (Fig. 3). Most frequently the studied species had 2–3 co-occurring species (Fig. 6). Notable deviations from this pattern were *S. pectinata*, *Z. palustris* and *Z. marina* that most frequently had none or only one co-occurring species. The mean number of co-occurring species was the highest in *C. baltica*. The proportions of co-occurring in field sites were most evenly distributed in *S. pectinata* and *T. nidifica* while *S. pectinata* itself was the most common co-occurring species for all the other species (Fig. 6). *Z. marina* had the largest share of single-species occurrence (Fig. 3) and coexistence with charophytes was almost lacking (Figs. 5, 6).

Discussion

The level of co-occurrence of the studied species varied along the environmental gradients (Fig. 2). This means that the potential rate of competition between the species depends on the environmental conditions. The highest rate of competition of the studied macrophytes can be expected in shallow depth, low wave exposure, soft seabed sediments and salinity around five (Fig. 2). According to previous studies, the distribution of charophytes and angiosperms is mainly controlled by depth that reflects light availability as well as hydrodynamic activity^{19,98,99}. Likewise, depth was the most influential environmental variable in the distribution models of the studied macrophytes in this study. In the brackish Baltic Sea region, soft sediments are inhabited by charophytes and angiosperms of fresh, brackish, and marine water origin. Living close to their physiological limits amplifies the effect of abiotic environmental gradients and even small changes in salinity can cause significant changes in species' distributions^{100,101}. Most of the studied species can tolerate wider salinity range than the range present in the study area. However, *Z. marina*, *T. nidifica*, and *C. baltica* are more common in areas with salinity over 5. Eelgrass *Z. marina* also differs from the other species by preferring greater water depth^{45,102}.

The most wide-spread soft-bottom species in the coastal waters of Estonia, northern Baltic Sea, was sago pondweed (*S. pectinata*) with an estimated distribution area of about 1100 km². The species is globally widespread and is also one of the most common species in the brackish Baltic Sea region²⁵. *S. pectinata* is of freshwater origin and can tolerate salinity fluxes up to 19⁴⁷. The species is regarded as tolerant to eutrophication¹⁰³ and has relatively wide niche space⁴⁵. Angiosperms *Z. palustris*, *M. spicatum*, and *P. perfoliatus* and charophyte *C. aspera* had also large distribution area of about 300 km² (Fig. 3, 4). Although regarded as relatively common species in the Estonian coastal waters and with relatively wide niche space⁴⁵, the charophyte *C. baltica* was estimated to have the smallest distribution area of less than 50 km². The salinity tolerance of *C. baltica* is probably the key factor: the species prefers salinity over two¹⁰⁴ and does not penetrate the low-salinity estuaries which are common habitats for all the other studied charophyte species. *Z. palustris* and *Z. marina* have been shown to have the widest niche space followed by *C. aspera*, *T. nidifica*, and *S. pectinata* while *R. maritima*, *P. perfoliatus*, *M. spicatum*, and *C. connivens* had narrow niche space⁴⁵. This shows that neither the numerous findings nor the large niche space directly translate into large distribution area in a given regional scale. Generally, realized distributions of generalist species tend to be wider than those of specialist species but there is a lot of variability and it cannot be

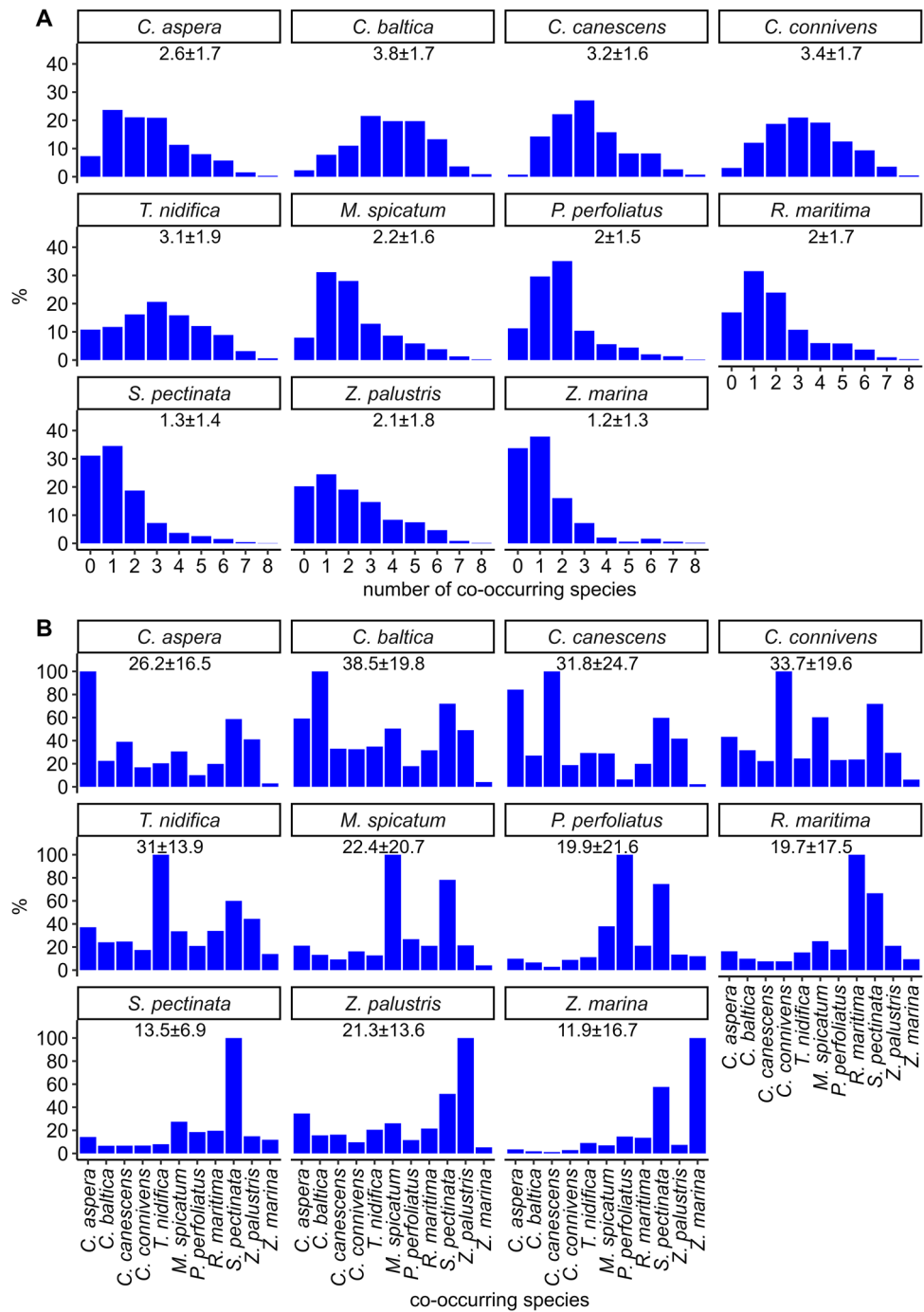


Figure 6. Proportions of occurrences with different number of co-occurring species (A) and proportions of occurrence overlap between the studied species (B) based on species' field observations. The numeric values in each panel represent mean value \pm standard deviation. For uniform x-axis among all plots, the focus species is also included with 100% value in panel B.

considered as a strict rule^{41,46,105,106}. Several earlier studies e.g. ^{46,107–109}, support our findings about discrepancy between niche width and distribution area.

The aspect of abundance (biomass, percentage coverage, shoot density or other measure) was not addressed in this study but it must be noted that in addition to distribution area, the abundance is an important feature when estimating the status and viability of a species and need to be taken into account in marine spatial planning and nature conservation activities. In regard to variation in plant abundance, it has been shown in terrestrial environment (especially in agriculture) that an optimal density is required for healthy plants. It has also been shown in eelgrass that positive interactions among shoots in the early stage of colonization are more important than competitive processes¹¹⁰: increased plant density has positive effects on shoot growth, areal expansion of

patches and rate of increase of shoot density. In the low-salinity areas of the Baltic Sea eelgrass reproduces asexually through growth of rhizomes that creates patches and meadows. In a multispecies angiosperm community, *Z. marina* has been shown to be more resistant to shading, but at the same time the monocultures of eelgrass showed faster recovery after disturbance¹¹¹. However, the number of studies on interspecific interactions of submerged soft-bottom species in the Baltic Sea region is limited¹¹¹ and no information exists for stress-tolerance of e.g., *T. nidifica* in mono- or multispecies communities. The revealed co-occurrence patterns of the studied species provide an important insight for the future targeted research on interspecific relationships.

The level of co-occurrence was higher inside the charophytes species group than inside the angiosperms species group. While the charophytes commonly co-exist with both charophytes and angiosperms, the angiosperms were more commonly associated with other angiosperms rather than charophytes. Moreover, three species of the *Chara* genus (*C. baltica*, *C. canescens*, *C. connivens*) showed very low proportion (less than 4%) of single-species occurrences while four angiosperm species out of six formed single-species communities in 17–34% of their field occurrence sites. Contrasting to our findings in the brackish water, formation of single-species meadows is noted to be characteristic of charophytes in freshwater¹¹². *T. nidifica* differed from the other charophytes by exhibiting higher proportion of single-species occurrences and co-occurrence with *Z. marina* (Fig. 6). *T. nidifica* is found in low coverage both within charophyte and vascular plant communities, however, it is also often found as a single specimen on bare sand. It is a brackish water species and grows mostly in salinities over four, but can tolerate large range of salinity^{21,113}. In Estonia the species is found growing in depth from 0.1 to 9.5 m, but it is most common in shallower area with depth range 1–3 m⁴⁵.

The highest level of co-occurrence, as measured by the Schoener's D, was observed between *C. aspera* and *C. canescens* (Fig. 5). However, the overlap was not symmetrical: 84% of the field occurrences of *C. canescens* overlapped with that of *C. aspera* but only 39% of the occurrences of *C. aspera* coincided with that of *C. canescens* (Fig. 6). The distribution area of *C. aspera* was more than twice the size of that of *C. canescens*. While both species are widespread in the Baltic Sea, it has been shown that the difference between them is that *C. aspera* often forms monospecific stands or dominant communities with association of other species while *C. canescens* mainly occurs as single plants or patches inside the stands of *C. aspera* or other charophytes²¹.

Z. marina had the lowest co-occurrence with all other species; the overlap was particularly low with charophytes (Fig. 5). The rare co-occurrence of *Z. marina* and the *Chara* genus of charophytes is an anticipated result, but in this study, the pattern was quantified using extensive database of field observations together with permutation tests. Eelgrass is a marine species and cannot tolerate salinity lower than five. In Estonian coastal waters, *Z. marina* commonly forms single-species communities in a depth range of 2–5 m and overall, it has been found growing in depths from 0.1 to 8.6 m. Our results also clearly indicated that compared to the other studied species, *Z. marina* had the largest proportion of single-species occurrences. When *Z. marina* grew with other species, then it had the highest proportion of overlap (58%) with the species with the largest distribution area—*S. pectinata*. While the overlap with the species of genus *Chara* was very low (<4%), the occurrence of *Z. marina* overlapped with that of the charophyte *T. nidifica* by almost 10% (Fig. 6).

Potamogeton perfoliatus was another angiosperm species that had notably low occurrence overlap with the species of *Chara* genus. While in the case of *Z. marina* and *Chara* spp., the lack of overlap can be clearly explained by variations in depth, salinity and wave exposure, the segregation of *P. perfoliatus* and *Chara* spp. is not that straightforward. It seems that for *P. perfoliatus*, the sediment composition has the most important role—the distribution of *P. perfoliatus* is limited to soft sediments with the share of sand around 70–100% while the share of sand can be much smaller in case of the species of *Chara* genus⁴⁵.

It has been previously shown that among the studied species, the charophyte *C. connivens* and angiosperm *M. spicatum* grow in the most similar environmental conditions⁴⁵. Although there is significant overlap in the distribution area of these species in the NE Baltic Sea region, *M. spicatum* had more than two-fold larger distribution area than that of *C. connivens* because *M. spicatum* also inhabits more exposed coasts. However, both species occur more commonly together with *S. pectinata* than with each other. This indicates that similar niches do not necessarily translate into highest distribution overlap but instead, the highest distribution overlap exists with the most widespread species. Theoretically, there are two extreme scenarios in the niche–distribution interplay^{114–116}: if competitive interactions strongly determine community composition, then the species with similar niches should be spatially segregated; if community composition is not strongly driven by competition, then the species with similar niches should exhibit high spatial overlap. Based on empirical evidence of different organism groups e.g.^{114,117}, and references therein, the relationships between niche overlap and spatial overlap can take various forms. Based on the results of this study and previous studies²², frequent co-occurrence of the studied species existed and thus it is more likely that the competition is not very important structuring force of communities of soft-bottom macrophytes in the northern Baltic Sea. However, it must be considered that spatial and temporal scales play an important role in the outcome of distribution studies. Closely related species or species with similar niches rely on mutually exclusive distribution at small spatial scales due to competition but co-occur in larger spatial scales due to similar broad-scale environmental preferences^{51,118,119}. This means that distribution studies should address multiple spatial scales (from microhabitat to regional level) in order to fully understand the coexistence and competition patterns between species. In this study, the environmental variables were available in 10–50 m raster layers. We lacked the small-scale (sub-meter to below 10 m) environmental data were direct competition between species actually takes place. Such high resolution environmental data is lacking for large marine areas due to obvious practical reasons. The coexistence patterns of soft bottom macrophytes in small spatial scales remain to be elucidated in dedicated future field studies.

Species' potential geographic distribution areas often differ from their occupied distribution areas due to biotic interactions¹¹⁷. The complex interplay between plants' physiological tolerance limits, availability of environmental niche and biotic interactions (competition, herbivory, pathogens etc.) is further modified by stochastic disturbance events (e.g., severe storms) and anthropogenic pressures (e.g., eutrophication, human-induced climate

change). The complexity is further increased by concurrent effects and interactions acting at different spatial and temporal scales. Thus, it must be taken into account that results of SDMs only provide simplified insights into geographic aspect of the species coexistence but cannot reveal the full complexity of the system.

Conclusions

Analyzing a comprehensive dataset of field samples of soft substrate macrophytes, we found that the observed co-occurrence patterns among sympatric species significantly diverged from what would be expected by random chance. The analysis showed that species exhibited a higher tendency to co-occur within their taxonomic groups (charophytes and angiosperms) than between these groups. When considering previous studies of niche positions and niche sizes of the studied species, it can be concluded that large niche space does not always directly translate into large distribution area in a given regional scale. In future studies multiple spatial scales (from microhabitat to regional level) as well as temporal scales should be addressed to fully understand the coexistence and competition patterns between sympatric macrophyte species.

Data availability

Enquiries about data availability should be directed to Kristjan Herkül (kristjan.herkul@ut.ee).

Code availability

Custom R code, deemed central to the conclusions, is added in Appendix 2.

Received: 4 March 2023; Accepted: 9 November 2023

Published online: 16 November 2023

References

- Mcarthur, M. A. *et al.* On the use of abiotic surrogates to describe marine benthic biodiversity. *Estuar. Coast. Shelf Sci.* **88**, 21–32 (2010).
- Kahlert, M. *et al.* Gaps in current Baltic Sea environmental monitoring—Science versus management perspectives. *Mar. Pollut. Bull.* **160**, 111669 (2020).
- Robinson, L. M. *et al.* Pushing the limits in marine species distribution modelling: Lessons from the land present challenges and opportunities. *Glob. Ecol. Biogeogr.* **20**, 789–802 (2011).
- Robinson, N. M., Nelson, W. A., Costello, M. J., Sutherland, J. E. & Lundquist, C. J. A systematic review of marine-based Species Distribution Models (SDMs) with recommendations for best practice. *Front. Mar. Sci.* **4**, 1–11 (2017).
- Melo-Merino, S. M., Reyes-Bonilla, H. & Lira-Noriega, A. Ecological niche models and species distribution models in marine environments: A literature review and spatial analysis of evidence. *Ecol. Modell.* **415**, 108837 (2020).
- Secretariat of the Convention on Biological Diversity. *Global Biodiversity Outlook 2* (2006).
- OSPAR. *Convention for the Protection of the Marine Environment of the North-East Atlantic (OSPAR Convention)*. *Yearbook of International Cooperation on Environment and Development 1998–99* 122–123 (1992). <https://doi.org/10.4324/9781315066547-30>.
- HELCOM. Convention on the Protection of the Marine Environment of the Baltic Sea Area. (1992).
- Houde, E., Kenny, A., Zhou, S. EBM in highly impacted coasts and estuaries. *The Sea* **16**. <https://doi.org/10.1016/c2017-0-00731-0> (2014).
- Elliott, M. *et al.* “And DPSIR begat DAPSI(W)R(M)!” - A unifying framework for marine environmental management. *Mar. Pollut. Bull.* **118**, 27–40 (2017).
- Kenny, A. J. *et al.* Assessing cumulative human activities, pressures, and impacts on North Sea benthic habitats using a biological traits approach. *ICES J. Mar. Sci.* **75**, 1080–1092 (2018).
- Troell, M. *et al.* Regime shifts and ecosystem services in Swedish coastal soft bottom habitats. *Ecol. Soc.* **10**, 30 (2005).
- Helmuth, B., Mieszkowska, N., Moore, P. & Hawkins, S. J. Living on the edge of two changing worlds: Forecasting the responses of rocky intertidal ecosystems to climate change. *Annu. Rev. Ecol. Syst.* **37**, 373–404 (2006).
- Defeo, O. *et al.* Threats to sandy beach ecosystems: A review. *Estuar. Coast. Shelf Sci.* **81**, 1–12 (2009).
- Vinagre, C. *et al.* Ecological traps in shallow coastal waters—Potential effect of heat-waves in tropical and temperate organisms. *PLoS One* **13**, e0192700 (2018).
- Chen, L. *et al.* Long-term changes of marine subtidal benthic communities in North East Asia (Yellow and Japan seas) in a global change context: A review. *Aquat. Conserv. Mar. Freshw. Ecosyst.* **30**, 1451–1475 (2020).
- Schneider, S. C., García, A., Martín-Closas, C. & Chivas, A. R. The role of charophytes (Charales) in past and present environments: An overview. *Aquat. Bot.* **120**, 2–6 (2015).
- Mtswana Nordlund, L., Koch, E. W., Barbier, E. B. & Creed, J. C. Seagrass ecosystem services and their variability across genera and geographical regions. *PLoS One* **11**, e0163091 (2016).
- Kautsky, H., Martin, G. & Snoeijs-Leijonmalm, P. The phytobenthic zone in *Biological Oceanography of the Baltic Sea* (eds. Snoeijs-Leijonmalm, P., Schubert, H. & Radziejewska, T.) 387–455 (Springer Netherlands, 2017). https://doi.org/10.1007/978-94-007-0668-2_11.
- HELCOM. HELCOM Checklist 2.0 of Baltic Sea Macrospecies. *Balt. Sea Environ. Proc.* **174**, 1–76 (2020).
- Schubert, H. & Blindow, I. *Charophytes of the Baltic sea* (Koeltz Scientific Books, 2004).
- Torn, K., Martin, G., Kuk, H. & Trei, T. Distribution of charophyte species in Estonian coastal water (NE Baltic Sea). *Sci. Mar.* **68**, 129–136 (2004).
- Boström, C. *et al.* Distribution, structure and function of Nordic eelgrass (*Zostera marina*) ecosystems: Implications for coastal management and conservation. *Aquat. Conserv. Mar. Freshw. Ecosyst.* **24**, 410–434 (2014).
- Den Hartog, C. *The sea-grasses of the world* (North-Holland, 1970).
- GBIF Secretariat. GBIF Backbone Taxonomy, Checklist dataset. <https://doi.org/10.15468/39omei> (2023).
- Kotilainen, A. *et al.* *Threatened habitat types in Finland 2018: the Baltic Sea*. (2020).
- Thibaut, T., Pinedo, S., Torras, X. & Ballesteros, E. Long-term decline of the populations of Fucales (*Cystoseira* spp. and *Sargassum* spp.) in the Albères coast (France, North-western Mediterranean). *Mar. Pollut. Bull.* **50**, 1472–1489 (2005).
- Orth, R. J. *et al.* A global crisis for seagrass ecosystems. *Bioscience* **56**, 987–996 (2006).
- Zhang, Y. *et al.* Global loss of aquatic vegetation in lakes. *Earth-Sci. Rev.* **173**, 259–265 (2017).
- Botrel, M. & Maranger, R. Global historical trends and drivers of submerged aquatic vegetation quantities in lakes. *Glob. Chang. Biol.* **29**, 2493–2509 (2023).
- Becker, R. *Gefährdung und Schutz von Characeen. Armleuchteralgen: Die Characeen Deutschlands* (Springer, 2016).

32. Waycott, M. *et al.* Accelerating loss of seagrasses across the globe threatens coastal ecosystems. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 12377–12381 (2009).
33. Green, A. E., Unsworth, R. K. F., Chadwick, M. A. & Jones, P. J. S. Historical analysis exposes catastrophic seagrass loss for the United Kingdom. *Front. Plant Sci.* **12**, 261 (2021).
34. Rey-Boissezon, A. & Auderset Joye, D. Habitat requirements of charophytes-Evidence of species discrimination through distribution analysis. *Aquat. Bot.* **120**, 84–91 (2015).
35. Sleith, R. S., Wehr, J. D. & Karol, K. G. Untangling climate and water chemistry to predict changes in freshwater macrophyte distributions. *Ecol. Evol.* **8**, 2802–2811 (2018).
36. Brzozowski, M., Pelechaty, M. & Bogawski, P. A winner or a loser in climate change? Modelling the past, current, and future potential distributions of a rare charophyte species. *Glob. Ecol. Conserv.* **34**, e02038 (2022).
37. Costion, C. M. *et al.* Will tropical mountaintop plant species survive climate change? Identifying key knowledge gaps using species distribution modelling in Australia. *Biol. Conserv.* **191**, 322–330 (2015).
38. Bogawski, P. *et al.* Current and future potential distributions of three *Dracaena* Vand. ex L. species under two contrasting climate change scenarios in Africa. *Ecol. Evol.* **9**, 6833–6848 (2019).
39. Hutchinson, G. E. *A treatise on limnology* (Wiley, 1957).
40. Wiens, J. J. & Graham, C. H. Niche conservatism: Integrating evolution, ecology, and conservation biology. *Annu. Rev. Ecol. Evol. Syst.* **36**, 519–539 (2005).
41. Clavel, J., Julliard, R. & Devictor, V. Worldwide decline of specialist species: Toward a global functional homogenization? *Front. Ecol. Environ.* **9**, 222–228 (2011).
42. Snoeijs, P. Marine and brackish waters. *Acta Phytogeogr. Suec.* **84**, 187–212 (1999).
43. Frankovich, T. A., Morrison, D. & Fourqurean, J. W. Benthic macrophyte distribution and abundance in estuarine mangrove lakes and estuaries: relationships to environmental variables. *Estuar. Coasts* **34**, 20–31 (2011).
44. Bučas, M. *et al.* How much can the occurrence and coverage of charophytes in an estuarine lagoon (Curonian Lagoon) be explained by environmental factors?. *Estuar. Coast. Shelf Sci.* **216**, 128–138 (2019).
45. Herkül, K., Torn, K. & Möller, T. The environmental niche separation between charophytes and angiosperms in the northern Baltic Sea. *Bot. Lett.* **165**, 1–13 (2018).
46. Bastrup-Spohr, L., Iversen, L. L., Borum, J. & Sand-Jensen, K. Niche specialization and functional traits regulate the rarity of charophytes in the Nordic countries. *Aquat. Conserv. Mar. Freshw. Ecosyst.* **25**, 469–481 (2015).
47. Verhoeven, M. R., Glisson, W. J. & Larkin, D. J. Niche models differentiate potential impacts of two aquatic invasive plant species on native macrophytes. *Diversity* **12**, 162 (2020).
48. Boschilia, S. M., Oliveira, E. F. & Thomaz, S. M. Do aquatic macrophytes co-occur randomly? An analysis of null models in a tropical floodplain. *Oecologia* **156**, 203–214 (2008).
49. Azzella, M. M., Bresciani, M., Nizzoli, D. & Bolpagni, R. Aquatic vegetation in deep lakes: Macrophyte co-occurrence patterns and environmental determinants. *J. Limnol.* **76**, 97–108 (2017).
50. Chmara, R., Szmaja, J. & Ulrich, W. Patterns of abundance and co-occurrence in aquatic plant communities. *Ecol. Res.* **28**, 387–395 (2013).
51. Thomaz, S. M. & Michelan, T. S. Associations between a highly invasive species and native macrophytes differ across spatial scales. *Biol. Invasions* **13**, 1881–1891 (2011).
52. Silveira, M. J. & Thomaz, S. M. Interspecific associations between *Hydrilla verticillata* and three dominant native genera of submerged macrophytes are taxa dependent. *Aquat. Sci.* **81**, 1–8 (2019).
53. Elith, J. & Leathwick, J. R. Species distribution models: Ecological explanation and prediction across space and time. *Annu. Rev. Ecol. Evol. Syst.* **40**, 677–697 (2009).
54. Kotta, I., Herkül, K., Paalme, T., Püss, T. & Kotta, J. Contribution of scale-dependent environmental variability on the biomass patterns of drift algae and associated invertebrates in the Gulf of Riga, northern Baltic Sea. *J. Mar. Syst.* **74**, S116–S123. <https://doi.org/10.1016/j.jmarsys.2008.03.030> (2008).
55. Torn, K., Kovtun-Kante, A., Herkül, K., Martin, G. & Mäemets, H. Distribution and predictive occurrence model of charophytes in Estonian waters. *Aquat. Bot.* <https://doi.org/10.1016/j.aquabot.2014.05.005> (2014).
56. Kukkk, T., Kull, T., Luuk, O., Saar, P. & Mesipuu, M. *Eesti taimede levikuatlas 2020 [Atlas of the Estonian flora 2020]*. (Pärandkoosluste Kaitse Ühing, 2020).
57. Bendtsen, J., Gustafsson, K. E., Söderkvist, J. & Hansen, J. L. S. Ventilation of bottom water in the North Sea-Baltic Sea transition zone. *J. Mar. Syst.* **75**, 138–149 (2009).
58. Isæus, M. Factors structuring *Fucus* communities at open and complex coastlines in the Baltic Sea. (2004).
59. van der Meijjs, F. & Isæus, M. Wave exposure calculations for the Gulf of Finland. *AquaBiota Rep.* **2020**(13), 1–26 (2020).
60. Uiboupin, R. & Pärn, O. *Mereala planeeringu alusuuring: jääolude analüüs ja kaartide koostamine [Basic research for marine planning: analysis of ice conditions and production of ice maps]*. (2018).
61. R Core Team. *R: A language and environment for statistical computing*. (R Foundation for Statistical Computing, 2022).
62. RStudio Team. *RStudio: integrated development environment for R*. (RStudio, (Posit Software, PBC, 2023).
63. Hijmans, R. J. raster: Geographic Data Analysis and Modeling. R package version 3.6–20. <https://cran.r-project.org/package=raster> (2023).
64. Pebesma, E. J. Simple features for R: Standardized support for spatial vector data. *R J.* **10**, 439 (2018).
65. Liaw, A. & Wiener, M. Classification and Regression by randomForest. *R News* **2**(3), 18–22 (2002).
66. Kuhn, M. Building predictive models in R Using the caret Package. *J. Stat. Softw.* **28**, 1–26 (2008).
67. Freeman, E. A. & Moisen, G. PresenceAbsence: An R package for presence absence analysis. *J. Stat. Softw.* **23**, 31 (2008).
68. Wickham, H. *et al.* Welcome to the Tidyverse. *J. Open Source Softw.* **4**, 1686 (2019).
69. ESRI. ArcGIS Desktop: version 10.8.1. <http://esri.com> (2020).
70. González-Irusta, J. M. *et al.* Comparing species distribution models: A case study of four deep sea urchin species. *Hydrobiologia* **745**, 43–57 (2015).
71. Maglietta, R. *et al.* Environmental variables and machine learning models to predict cetacean abundance in the Central-eastern Mediterranean Sea. *Sci. Rep.* **13**, 2600 (2023).
72. Lindegarth, M. *et al.* Testing the potential for predictive modeling and mapping and extending its use as a tool for evaluating management scenarios and economic valuation in the Baltic Sea (PREHAB). *Ambio* **43**, 82–93 (2014).
73. Bucas, M. *et al.* Empirical modelling of benthic species distribution, abundance, and diversity in the Baltic Sea: evaluating the scope for predictive mapping using different modelling approaches. *ICES J. Mar. Sci.* **70**, 1233–1243 (2013).
74. Ren-Yan, D., Xiao-Quan, K., Min-Yi, H., Wei-Yi, F. & Zhi-Gao, W. The predictive performance and stability of six species distribution models. *PLoS ONE* **9**, e112764 (2014).
75. Zhao, Z., Xiao, N., Shen, M. & Li, J. Comparison between optimized MaxEnt and random forest modeling in predicting potential distribution: A case study with *Quasipaa boulengeri* in China. *Sci. Total Environ.* **842**, 156867 (2022).
76. Adhikari, P., Lee, Y. H., Poudel, A., Hong, S. H. & Park, Y. S. Global spatial distribution of *Chromolaena odorata* habitat under climate change: random forest modeling of one of the 100 worst invasive alien species. *Sci. Rep.* **13**, 1–13 (2023).
77. Parviainen, M., Marmion, M., Luoto, M., Thuiller, W. & Heikkinen, R. K. Using summed individual species models and state-of-the-art modelling techniques to identify threatened plant species hotspots. *Biol. Conserv.* **142**, 2501–2509 (2009).

78. Marmion, M., Parviainen, M., Luoto, M., Heikkinen, R. K. & Thuiller, W. Evaluation of consensus methods in predictive species distribution modelling. *Divers. Distrib.* **15**, 59–69 (2009).
79. Zhang, L. *et al.* The use of classification and regression algorithms using the random forests method with presence-only data to model species' distribution. *MethodsX* **6**, 2281–2292 (2019).
80. Liu, C., Berry, P. M., Dawson, T. P. & Pearson, R. G. Selecting thresholds of occurrence in the prediction of species distributions. *Ecography* **28**, 385–393 (2005).
81. Jiménez-Valverde, A. & Lobo, J. M. Threshold criteria for conversion of probability of species presence to either-or presence-absence. *Acta Oecol.* **31**, 361–369 (2007).
82. Manel, S., Williams, H. C. & Ormerod, S. J. Evaluating presence-absence models in ecology: the need to account for prevalence. *J. Appl. Ecol.* **38**, 921–931 (2001).
83. Kohavi, R. A study of cross-validation and bootstrap for accuracy estimation and model selection. In *Proceedings of the 14th International Joint Conference on Artificial Intelligence*. 1137–1145 (Montreal, Canada, 1995).
84. Alpaydin, E. *Introduction to Machine Learning* (MIT Press, 2020).
85. Fielding, A. H. & Bell, J. F. A review of methods for the assessment of prediction errors in conservation presence/absence models. *Environ. Conserv.* **24**, 38–49 (1997).
86. Cohen, J. A coefficient of agreement for nominal scales. *Educ. Psychol. Meas.* **20**, 37–46 (1960).
87. Landis, J. R. & Koch, G. G. The measurement of observer agreement for categorical data. *Biometrics* **159–174** (1977).
88. Carter, J. V., Pan, J., Rai, S. N. & Galandiuk, S. ROC-ing along: Evaluation and interpretation of receiver operating characteristic curves. *Surgery* **159**, 1638–1645 (2016).
89. Swets, J. A. Measuring the accuracy of diagnostic systems. *Science* **240**, 1285–1293 (1988).
90. Wood, S. N. *Generalized Additive Models: An Introduction with R* (CRC Press, 2017).
91. Warren, D. L., Glor, R. E. & Turelli, M. Environmental niche equivalency versus conservatism: quantitative approaches to niche evolution. *Evolution* **62**, 2868–2883 (2008).
92. Rödder, D. & Engler, J. O. Quantitative metrics of overlaps in Grinnellian niches: Advances and possible drawbacks. *Glob. Ecol. Biogeogr.* **20**, 915–927 (2011).
93. Moore, D. S. *Introduction to the Practice of Statistics* (WH Freeman, 2009).
94. Schluter, D. A variance test for detecting species associations, with some example applications. *Ecology* **65**, 998–1005 (1984).
95. Gotelli, N. J. Null model analysis of species co-occurrence patterns. *Ecology* **81**, 2606–2621 (2000).
96. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B* **57**, 289–300 (1995).
97. Kruskal, W. H. & Wallis, W. A. Use of ranks in one-criterion variance analysis. *J. Am. Stat. Assoc.* **47**, 583–621 (1952).
98. Schwarz, A.-M., de Winton, M. & Hawes, I. Species-specific depth zonation in New Zealand charophytes as a function of light availability. *Aquat. Bot.* **72**, 209–217 (2002).
99. Hill, N. A., Lucieer, V., Barrett, N. S., Anderson, T. J. & Williams, S. B. Filling the gaps: Predicting the distribution of temperate reef biota using high resolution biological and acoustic data. *Estuar. Coast. Shelf Sci.* <https://doi.org/10.1016/j.ecss.2014.05.019> (2014).
100. Jonsson, P. R. *et al.* High climate velocity and population fragmentation may constrain climate-driven range shift of the key habitat former *Fucus vesiculosus*. *Divers. Distrib.* **24**, 892–905, <https://doi.org/10.1111/ddi.12733> (2018).
101. Torn, K., Peterson, A. & Herkül, K. Predicting the impact of climate change on the distribution of the key habitat-forming species in the NE Baltic Sea. *J. Coast. Res.* **95**, 177–181 (2020).
102. Krause-Jensen, D. & Carstensen, J. *Light Requirements of Marine Rooted Macrophytes* (The Danish Environmental Protection Agency, 2018).
103. Bakker, E. S. *et al.* Effect of macrophyte community composition and nutrient enrichment on plant biomass and algal blooms. *Basic Appl. Ecol.* **11**, 432–439 (2010).
104. Schubert, H., Blindow, I. & van de Weyer, K. *Chara baltica* in Arbeitsgruppe Characeen Deutschlands. in *Armeleuchteralgen: Die Characeen Deutschlands [Charophyceae: Charophytes in Germany]* (Springer Spektrum, 2016).
105. Devictor, V., Julliard, R. & Jiguet, F. Distribution of specialist and generalist species along spatial gradients of habitat disturbance and fragmentation. *Oikos* **117**, 507–514 (2008).
106. Slatyer, R. A., Hirst, M. & Sexton, J. P. Niche breadth predicts geographical range size: A general ecological pattern. *Ecol. Lett.* **16**, 1104–1114 (2013).
107. Eterovick, P. C. & Barros, I. S. Niche occupancy in south-eastern Brazilian tadpole communities in montane-meadow streams. *J. Trop. Ecol.* **19**, 439–448 (2003).
108. Kotta, J., Möller, T., Orav-Kotta, H. & Pärnoja, M. Realized niche width of a brackish water submerged aquatic vegetation under current environmental conditions and projected influences of climate change. *Mar. Environ. Res.* **102**, 88–101 (2014).
109. Pagel, J. *et al.* Mismatches between demographic niches and geographic distributions are strongest in poorly dispersed and highly persistent plant species. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 3663–3669 (2020).
110. Worm, B. & Reusch, T. B. H. Do nutrient availability and plant density limit seagrass colonization in the Baltic Sea? *Mar. Ecol. Prog. Ser.* **200**, 159–166 (2000).
111. Gustafsson, C. & Boström, C. Influence of neighboring plants on shading stress resistance and recovery of eelgrass, *Zostera marina* L. *PLoS ONE* **8**, e64064 (2013).
112. Jakubas, E. & Gąbka, M. Significance of current velocity gradients for distribution patterns of charophytes versus mosses and vascular plant communities in a lowland stream. *Oceanol. Hydrobiol. Stud.* **44**, 139–150 (2015).
113. Hamann, U. & Becker, R. *Tolypella nidifica* in Arbeitsgruppe Characeen Deutschlands. in *Armeleuchteralgen: Die Characeen Deutschlands [Charophyceae: Charophytes in Germany]* (Springer Spektrum, 2016).
114. Hofer, U., Bersier, L. F. & Borcard, D. Relating niche and spatial overlap at the community level. *Oikos* **106**, 366–376 (2004).
115. Geange, S. W., Pledger, S., Burns, K. C. & Shima, J. S. A unified analysis of niche overlap incorporating data of different types. *Methods Ecol. Evol.* **2**, 175–184 (2011).
116. Pascual-Rico, R. *et al.* Ecological niche overlap between co-occurring native and exotic ungulates: Insights for a conservation conflict. *Biol. Invasions* **22**, 2497–2508 (2020).
117. Anderson, R. P. *et al.* *Ecological Niches and Geographic Distributions* (Princeton University Press, 2011).
118. Procheş, Ş., Wilson, J. R. U., Richardson, D. M. & Rejmánek, M. Searching for phylogenetic pattern in biological invasions. *Glob. Ecol. Biogeogr.* **17**, 5–10 (2008).
119. Cadotte, M. W. *et al.* Phylogenetic patterns differ for native and exotic plant communities across a richness gradient in Northern California. *Divers. Distrib.* **16**, 892–901 (2010).

Acknowledgements

This manuscript is a result of Blue4All project, funded by the European Union under the Horizon Europe Programme, “HORIZON-MISS-2021-OCEAN-02-01” Theme, Grant Agreement no. 101094014.

Author contributions

K.H. contributed to the study concept, design, data preparation, data analysis, interpretation of the results, and manuscript preparation. K.T., T.M.-R., contributed to the study concept and design, data preparation, interpretation of the results and manuscript preparation. G.M. contributed to the manuscript preparation.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-023-47176-8>.

Correspondence and requests for materials should be addressed to K.H.

Reprints and permissions information is available at www.nature.com/reprints.

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



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023