https://doi.org/10.5194/egusphere-2025-3290
Preprint. Discussion started: 11 August 2025

Prediction of present and future spatial occurrence of cyanobacteria and the toxin nodularin in the Baltic Sea

Mohanad Abdelgadir[1], Bengt Karlson[2], Elin Dahlgren[3], Malin Olofsson[1]

[1]Department of Aquatic Sciences and Assessment, Swedish University of Agricultural Sciences, Box 7050, 750 07, Uppsala, Sweden
[2]Department of Research and Development, Oceanography, Swedish Meteorological and Hydrological Institute, SMHI, Västra Frölunda, Sweden
[3]Department of Aquatic Resources, Swedish University of Agricultural Sciences, Stångholmsvägen 2, SE-178 93 Drottningholm, Sweden

Correspondence to: Mohanad Abdelgadir (mohanad.mohammed@slu.se)

Abstract. Blooms of filamentous cyanobacteria are recurrent phenomena in the brackish Baltic Sea. These blooms often include toxin producing species, however, predicting and modeling the toxins spatial distribution poses great challenges. In addition, projected rising temperature due to climate change is expected to increase the occurrence of cyanobacterial blooms, making it vital to understand the distribution of the blooms and the associated cyanotoxins across ecosystems. Herein, we integrated measured concentration of the cyanotoxin nodularin, abundance of the toxin producer *Nodularia spumigena*, and environmental variables using Empirical Bayesian Kriging (EBK) regression prediction, ensemble learning, and stacked species distribution modeling (SSDM). This setup was used to predict and interpret the current and future area distribution of *N. spumigena* and nodularin across the Baltic Sea. Predictions were based on results from biogeochemical models describing current and projected future concentrations of near surface chlorophyll, nitrate, phosphate, salinity, and temperature along with nitrate-to-phosphate ratio and a geographical variable of distance to shore. Prediction for the future distribution was performed using projected climate change scenarios in the year 2100. Findings show that the predicted area distribution of nodularin is determined by concentrations and interaction effects of salinity, temperature, phosphate, nitrate to phosphate ratio, and distance to shore, and is associated with the predicted area distribution of *N. spumigena*. Predicted site distribution shows increased nodularin occurrences in the Eastern and Western Gotland Sea, the Northern Baltic Proper, southern parts of the Bothnian Sea, and in the Arkona basin. By the year 2100, area distribution of nodularin is predicted to increase in the northern part of the Eastern Gotland Sea, Northern Baltic Proper, Åland Sea, southern parts of the Bothnian Sea, Arkona Basin, and slightly into the Bothnia Bay in response to projected climate change scenarios. Our developed modeling approach is useful for risk assessment and management of cyanotoxins where toxicological data are insufficient.

https://doi.org/10.5194/egusphere-2025-3290
Preprint. Discussion started: 11 August 2025

1 Introduction

Blooms of filamentous cyanobacteria occur regularly in the Baltic Sea during summer (Kahru et al., 2020; Kahru and Elmgren, 2014; Karlson et al., 2022), mainly from end of June until August, and are dominated by the diazotrophic filamentous cyanobacterial species *Nodularia spumigena*, *Aphanizomenon flos-aquae*, and *Dolichospermum* spp. (Carlsson and Rita, 2019; Klawonn et al., 2016; Olofsson et al., 2020). The toxicity of these blooms is mainly attributed to the hepatotoxin nodularin, produced by *N. spumigena.* The chemical structure and toxicity of nodularin are similar to microcystins (Lundholm et al., 2009). The toxicity of cyanobacterial blooms is known to cause economic losses to the surrounding societies of the Baltic Sea (Jonasson et al., 2008) with negative effects on the food chain dynamics on several trophic levels (Karjalainen et al., 2007). How cyanobacteria and their toxins respond to the ongoing and future climate change in the Baltic Sea is, however, not well known. Therefore, an understanding of what influences and predicts the spatial occurrence of cyanobacteria and related toxins is needed for efficient management of the Baltic Sea.

The spatial and temporal distribution of cyanobacteria in the Baltic Sea area has been investigated using several different approaches. Monitoring data based on microscope analyses (Karlson et al., 2022; Olofsson et al., 2020), satellite remote sensing of ocean color where high reflectance at wavelengths in the red part of the spectrum is indicative of near surface accumulations of filamentous cyanobacteria (Hansson and Hakansson, 2007; Kahru and Elmgren, 2014; Karlson et al., 2022), and several modelling approaches have been applied to describe the blooms (Hense et al., 2013; Hieronymus et al., 2021; Löptien and Dietze, 2022; Munkes et al., 2021). From these, several hypotheses have developed for what could influence the expansion and increase the occurrence of filamentous cyanobacterial blooms. For instance, availability of inorganic nitrogen and phosphorus (Andersen et al., 2020; Carpenter, 2005; Lu et al., 2019; Paerl et al., 2018; Wurtsbaugh et al., 2019; Yang et al., 2008), the ratio of nitrogen to phosphorus (Burford et al., 2023; Pliński et al., 2007), chlorophyll-a concentration (Budakoti, 2024), salinity (Lehtimäki et al., 1994; Moisander et al., 2002; Silveira and Odebrecht, 2019), temperature (Stal, 2009; Walls et al., 2018), and also a combination of salinity and temperature (Olofsson et al., 2020). In addition to promoting cyanobacterial growth, environmental drivers such as temperature, light, and nutrient availability can also modulate the production of cyanotoxins. For instance, nodularin production in Baltic Sea species has been shown to vary with nitrogen availability and physiological state (Lehtimaki et al., 1997). Similarly, oxidative stress conditions have been linked to microcystin release, potentially as part of a programmed cell death response (Ross et al., 2006). These findings suggest that bloom toxicity is not solely a function of biomass, but also of stress-related cellular responses. Despite the available knowledge on cyanobacterial blooms, little is known regarding what factors control spatial occurrence of *N. spumigena* and the production of nodularin across the Baltic Sea. Maps of bloom intensity cannot be assumed to translate to cyanotoxin distribution. The fact that cyanotoxin cannot be directly measured by satellite and remote sensing along with lack of standardized modeling approaches poses great challenges to make predictions about current and future occurrence of cyanotoxins in aquatic ecosystems.

https://doi.org/10.5194/egusphere-2025-3290
Preprint. Discussion started: 11 August 2025

To mitigate this methodological gap we integrated the results of a geostatistical interpolation method, the Empirical Bayesian Kriging (EBK) regression prediction (Gribov and Krivoruchko, 2020), and ensemble learning algorithms (Dietterich, 2000), to predict and interpret the spatial distribution of nodularin occurrence across the Baltic Sea. Kriging algorithms in EBK regression prediction are advanced geostatistical prediction methods and known as best linear unbiased estimators producing robust estimates at unsampled locations (Goovaerts, 1997; Krivoruchko and Gribov, 2019). EBK regression prediction methods are also known to generate robust and better accuracy than other kriging techniques both for small datasets and even when data is locally moderately non-stationary (Gribov and Krivoruchko, 2020; Krivoruchko, 2012). The relationships between explanatory and dependent variables could alter in several locations; yet the EBK regression prediction is able to accurately simulate these regional variations and takes regional influences into consideration (Deutsch and Journel, 1992; Gribov and Krivoruchko, 2020; Krivoruchko and Gribov, 2019; Pyrcz and Deutsch, 2014).

There are also cases where the spatial estimation of continuous variables could be a challenging task when EBK regression prediction does not determine the relevance of the explanatory variables that affect the predictive variables or the independent variables that are highly correlated with the response variables (Olea, 1999; Wang et al., 2019). This could be due to the phenomenon being sampled, may be produced by nonlinear processes, and the data may exhibit complex multivariate features, non-Gaussianity, and non-stationarity. In such a situation, the power of machine-learning algorithms in ensemble modeling allows identity of the patterns in the complex datasets and to make estimation and prediction with no requirement for rigid statistical assumptions, such as stationarity and linearity (Ghannam and Techtmann, 2021; Sathya and Abraham, 2013). Machine-learning algorithms also allow the inclusion of multivariable, coping with missing values, and being able to reveal the hidden relationships between predictors (Abdelgadir et al., 2023; Beery et al., 2021; Jiang et al., 2022; Thompson et al., 2019). In addition, stacked species distribution modeling SSDM (Schmitt et al., 2017) is an advanced approach combining several ensemble outputs to model species assemblages into one forecast.

Herein we integrated the results of the EBK regression prediction, ensemble learning algorithms and SSDM, using 139 nodularin concentration measurements, abundance of the cyanobacteria *N. spumigena*, and model-based raster layers on environmental and geographical variables to predict and interpret the spatial occurrence of nodularin across the Baltic Sea. The overall aim was to understand what factors drive the spatial occurrence of nodularin, and how the spatial occurrence of this cyanotoxin will be affected by the projected climate change scenarios in the year 2100 across the Baltic Sea.

2 Materials and methods

2.1 Data collection

Data was compiled for 139 samples in 54 locations across the Baltic Sea, the Kattegat, and the Skagerrak, sampled June to September 2023 **(Fig. 1)**. The dataset comprised locations in longitude and latitude with corresponding nodularin in biomass (µg l^{-1}), collected onto 47 mm GFC filters and analyzed using enzyme-linked immunoassay (ELISA) for

https://doi.org/10.5194/egusphere-2025-3290
Preprint. Discussion started: 11 August 2025

microcystins/nodularins from Gold Standard Diagnostics, following the manufacturer's instructions; and the abundance of *Nodularia spumigena* (units l^{-1}) based on Olenina et al., (2006).

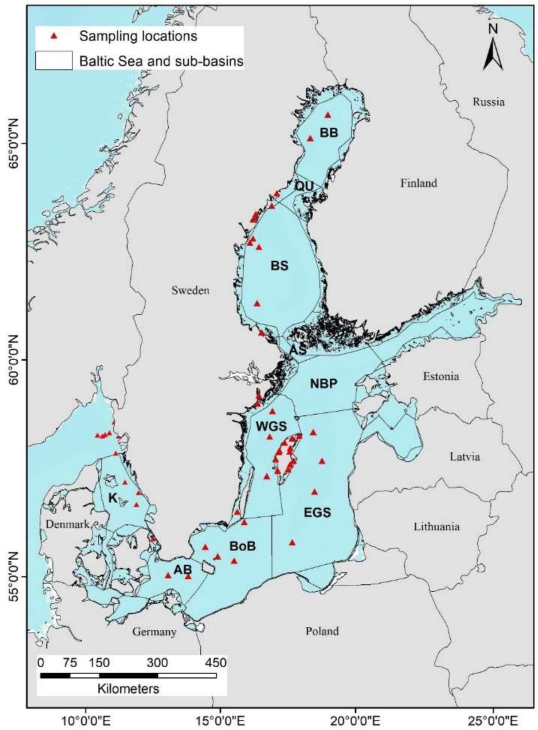


Figure 1: Map of the Baltic Sea indicating sampling locations from where nodularin was analyzed. Sub-basins described in this study are abbreviated on the map as follows: BB= Bothnian Bay, QU= The Quark, BS= Bothnian Sea, ÅS= Åland Sea, NBP= Northern Baltic Proper, WGS= Western Gotland Sea, EGS= Eastern Gotland Sea, BoB= Bornholm Basin, AB= Arkona Basin, and K= Kattegat

2.2 Environmental parameters

Prediction and geostatistical interpolation of spatial occurrence of nodularin was investigated using different environmental and geographical variables. In this study we used data from samples collected in monitoring programs and results from modelled concentrations of chlorophyll-a in surface water (mg m^{-3}) from the Baltic Sea Biogeochemistry Analysis and Forecast (ERGOM) https://doi.org/10.48670/moi-00009 (Neumann et al., 2021). Concentration of inorganic dissolved nutrients: phosphate PO_4, nitrate NO_3 (µmol m^{-3}) originate from results of a Baltic Sea Biogeochemistry Reanalysis by SMHI https://doi.org/10.48670/moi-00012. Salinity and sea surface temperature (°C) are based on Global Ocean Physics Analysis and Forecast https://doi.org/10.48670/moi-00016. Salinity was derived from the Baltic Sea Physics Reanalysis by SMHI https://doi.org/10.48670/moi-00013. The data was downloaded from E.U. Copernicus Marine Service Information

https://doi.org/10.5194/egusphere-2025-3290
Preprint. Discussion started: 11 August 2025

(European Union-Copernicus Marine Service, 2016, 2018b, a, 2019). Downloaded modelled data with a horizontal resolution of 1 nm (nautical mile) and vertical depth layers varying between 1-24 meters were limited to the period of sampling from June to September 2023 at a depth range of 0.5 to 10 meters. Data was downloaded as NetCDF-4 format and transformed to ESRI raster grid file format 'GeoTIFF' in QGIS Desktop 3.34.13 (QGIS Development Team, 2022).

The model data from Copernicus were compared with data from measurements made in the National Swedish Marine Monitoring Program, including mean values of chlorophyll, NO_3, PO_4, salinity, and temperature, downloaded from the Swedish National Oceanographic Data Centre (NODC) at SMHI https://shark.smhi.se. A comparison of the derived satellite remote sensing data or modelled data and the observations in the monitoring program are presented in the results and supplementary data.

Furthermore, prediction of nodularin concentration and *N. spumigena* spatial distribution modeling were tested using results from modelled data available at Copernicus, i.e., mean sea surface temperature, chlorophyll, salinity, PO_4, NO_3, distance to shore (m), and future climate change scenarios developed under the Shared Socioeconomic Pathway (SSP) scenarios of future climate change. Data layers reflecting future conditions following the Paris Agreement of reduced greenhouse gas emissions, to the "fossil-fueled development" SSP5-8.5 scenario of high emissions and low challenges to adaptation (Assis et al., 2024). Raster layers were derived from Atmosphere-Ocean General Circulation Model (AOGCM) as raster grids at 2.5 arc-minute spatial resolution, approximately 5 km^2 grid cell sizes at the equator corresponding to future greenhouse gas concentrations in the year 2100. All environmental data and climate change scenarios were downloaded as raster GeoTIFF from Bio-ORACLE project v3.0 (Assis et al., 2024) using 'sdmpredictors' package (Bosch and Fernandez, 2022) in RStudio (RStudio Team, 2020). All raster layers were cropped to the spatial extent of the Baltic Sea (latitude/longitude: min 10, max 30/min 53, max 66) and coordinate reference system (CRS) EPSG: 5845-SWEREF99 TM.

Environmental data corresponding to each sampling location (XY coordinates) and time point were extracted in RStudio equipped with R v.4.4.0 (R Core Team, 2021) using the function implemented in the packages 'raster' (Etten, 2012; Hijmans & Etten, 2012),'sp' (Bivand et al., 2013; Pebesma and Bivand, 2005) and 'tidyverse' (Wickham et al., 2019). Data holding all information in 'csv' format was exported to ESRI Shapefile 'shp' in QGIS to be used for the analysis.

2.3 Interpolation and ensemble modeling flow

Geostatistical interpolation and ensemble learning modeling flow are illustrated in **Fig. 2.**

2.3.1 Geostatistical interpolation analysis

Analysis and prediction of nodularin occurrence was performed using the Empirical Bayesian Kriging (EBK) Regression Prediction based on Gaussian process regression (see Oliver & Webster, 1990). How well the interpolation model fits the data was examined using the cross-validation method 'leave-one-out resampling' (Dubrule, 1983; Zhang and Wang, 2010). In brief, regression prediction was performed using nodularin concentration as dependent variable and the environmental variables (in raster format) as independent. The EBK regression predictions were validated using 1000 simulations as cross-

https://doi.org/10.5194/egusphere-2025-3290
Preprint. Discussion started: 11 August 2025

validation. Prediction for nodularin occurrence in response to *N. spumigena* abundance was performed using CoKriging (Goovaerts, 1997, 1998). Furthermore, spatial interpolation of nodularin in response to the mean values measured by National Swedish Marine Monitoring to was performed using co-kriging by applying nearest neighbor as proximal interpolation method. Geostatistical interpolation analysis was performed in QGIS and by Esri. ArcGIS® and ArcMap™.

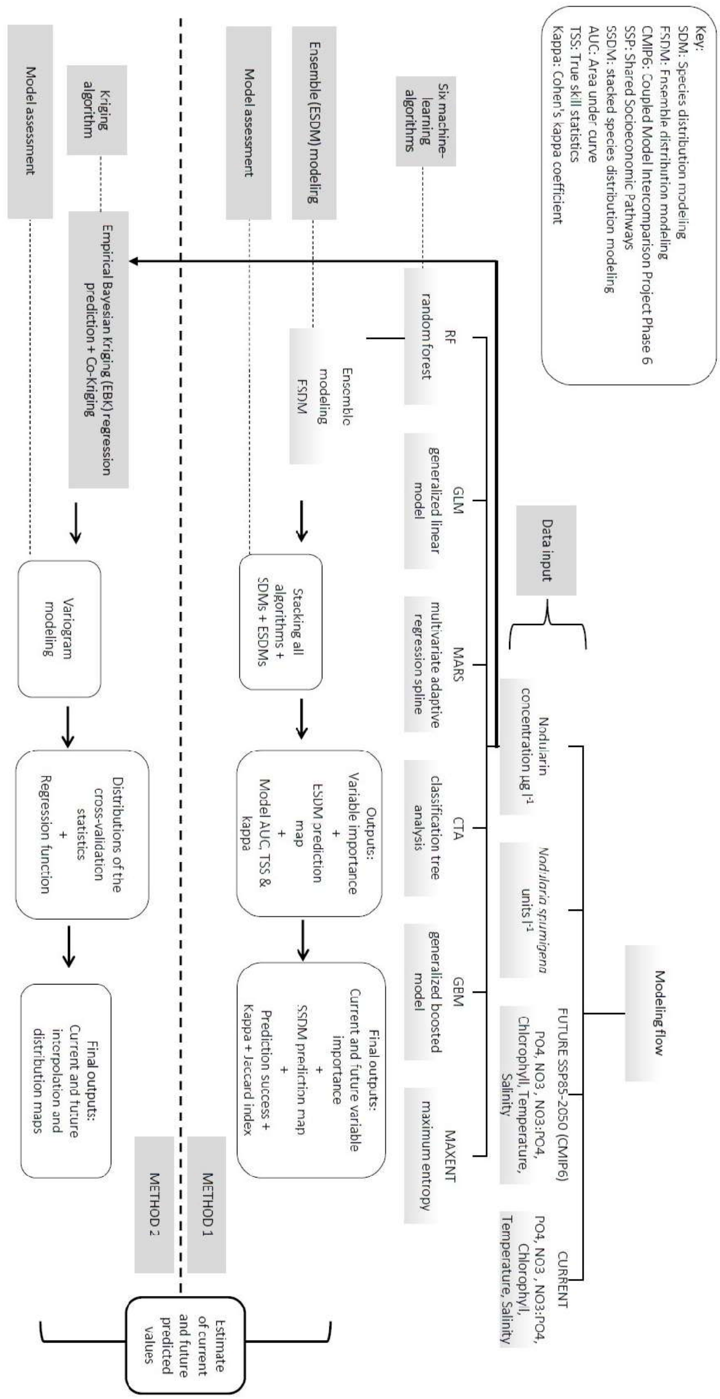
2.3.2 Ensemble prediction of nodularin occurrence and *Nodularia spumigena* distribution

Prediction for current and future occurrence of nodularin was performed using Ensemble learning method (Araújo and New, 2007), and modeling occurrence of *N. spumigena* performed using Stacked Species Distribution Modeling SSDM (Schmitt et al., 2017). Models were built using six different machine-learning algorithms including Classification Tree Analysis CTA (Steinberg, 2009), Maximum-Entropy learning MAXENT (Phillips et al., 2006), Multivariate Adaptive Regression Spline MARS (Friedman, 1991), Random Forest RF (Breiman, 2001), Generalized Boosting regression Model GBM (Friedman, 2002), and Generalized Linear Model GLM (Guisan et al., 2002).

In brief, models were built by creating two data objects with the 'sdmData' function in the "sdm" package (Dietterich, 2000; Naimi and Araújo, 2016). The first data object for nodularin ensemble prediction was created by having the occurrence as presence "1" and absence "0", measured concentrations and all environmental variables. Second data object for the species ensemble and stacked distribution modeling was created by having all occurrence records of *N. spumigena*, environmental variables, and 10,000 pseudoabsence datapoints randomly generated within the spatial extent of the study area as a background (Barbet-Massin et al., 2012). The two prediction models were tested with 10-fold bootstrapping (Harrell et al., 2005; Lima et al., 2019) as the replication technique and validated in a repeated split-samples procedure, i.e., 70% of the occurrence dataset was used for model training and the remaining 30% as test data repeated over 10-fold cross validation (Roberts et al., 2017).

Models were assessed for prediction success, omission rate (Franklin, 2010; Phillips et al., 2006), accuracy and performance based on Area Under the Curve AUC (Hanley and McNeil, 1982), True Skill Statistics TSS (Monserud and Leemans, 1992), Cohen's KAPPA coefficient (Allouche et al., 2006), and Jaccard coefficient (Jaccard, 1912). Predicted variables of importance and percentage of contribution were assessed using AUC, Pearson Correlation coefficient (Benesty et al., 2009) and Jackknife test (Efron, 1982).

https://doi.org/10.5194/egusphere-2025-3290
Preprint. Discussion started: 11 August 2025



171

172 **Figure 2: A diagram showing the modeling flow used to predict the present and future spatial occurrence of**
173 **cyanobacteria and the toxin nodularin in the Baltic Sea.**

https://doi.org/10.5194/egusphere-2025-3290
Preprint. Discussion started: 11 August 2025

2.4 Statistical analysis

The effect of both environmental variables and abundance of *N. spumigena* on the predicted nodularin concentrations were quantified using Bayesian linear regression (see Clyde et al., 2011), analysis of variance ANOVA and general linear model GLM. The effect and interaction effects between predictors on nodularin concentration were assessed using Bayes factor BF_{10}, p-value (at $p<0.05$), coefficient of determination R^2 and Pearson correlation coefficient. Assessment of the EBK regression prediction models was performed using cross-validation statistics with a regression function calculated using a robust regression procedure. Statistical analyses were performed in RStudio, and open software jamovi v.2.3.21 (The jamovi project, 2023) and JASP v.0.19.3 (JASP Team, 2025).

3 Results

3.1 Results of geostatistical interpolation analysis

In pharetra massa dictum gravida scelerisque. Sed vitae purus eget purus tincidunt accumsan ut at magna. Prediction of nodularin spatial occurrence was first tested using environmental predictors from the European Copernicus dataset of Baltic Sea Biogeochemistry Analysis and Forecast (ERGOM). Results of the Bayesian linear regression show that there were significant interaction effects between concentrations of salinity, temperature and distance to shore (*$BF_{10} > 1$, R^2=0.23, regression function: 0.70* x + 0.004*), and between PO_4, salinity, temperature, and distance to shore (*$BF_{10} > 1$, R^2=0.23, regression function:0.70 * x + 0.008)* and the abundance of *Nodularia spumigena* (*$BF_{10} > 1$, R^2=0.1*) on predicted spatial occurrence of nodularin (**Table 1a&b, Fig 3 & Table S1).** This finding is also supported by the regression function of the EBK regression prediction (*distance to shore: 0.74* x + 0.001; salinity: 0.73 * x + 0.003; temperature: 0.76 * x + 0.003*), see **Fig 4 & Fig S1**. Analysis of ordinary and multiple linear regression shows that there was significant positive linear regression of temperature, and distance to shore (*$p < 0.05$*), and negative with salinity (*$p < 0.001$*), (*linear regression R^2=0.3, multiple R^2=0.2*). Analysis of variance ANOVA shows that there was significant effect of salinity (*ANOVA, F=11.11, $p< 0.001$*), PO_4 (*ANOVA, F=6.6, $p< 0.01$*), NO_3:PO_4 ratio(*ANOVA, F=5.03, $p< 0.05$*), distance to shore (*ANOVA, F=4.73, $p< 0.05$*), temperature (*ANOVA, F=3.93, $p< 0.05$*) with interaction effects between chlorophyll, temperature (*ANOVA, F=7.27, $p< 0.01$*), NO_3, temperature (*ANOVA, F=6.70, $p< 0.05$*), and between chlorophyll and distance-to-shore (*ANOVA, F=4.59, $p< 0.05$*), **Fig 3 &Table S2&3a&b.**

Table 1: First panel (a) shows the best 10 models for Bayesian linear regression analysis between nodularin concentration as dependent and environmental variables and *Nodularia spumigena* abundances as predictors, and second panel (b) posterior summaries of coefficients. The Bayes factor BF_{10} is a ratio which quantifies evidence in favor of an effect (represented by "1") versus no effect (represented by "0"). If $BF_{10} > 1$ indicates evidence in favor of an effect. $0 < BF_{10} < 1$ indicates evidence in favor of no effect. The P(M) indicates that the prior probabilities of the other models are equal, P(M|data) refers to the posterior probability of each model after seeing the data while BF_M compares each model to the average P(M|data) of the other models. For all Bayesian linear regression models and effects refer to **Table S1.**

https://doi.org/10.5194/egusphere-2025-3290
Preprint. Discussion started: 11 August 2025

(A)						
Models	P(M)	P(M|data)	BF_M	BF_{10}	R^2	Predictors
Null model	0.01	0.00	0.01	1.00	0.00	Environmental variables
Salinity + Temperature + Dist-to-shore	0.01	0.16	24.00	1682.35	0.24	
PO_4 + Salinity + Temperature + Dist-to-shore	0.01	0.10	13.67	1028.65	0.25	
Salinity + Temperature + NP + Dist-to-shore	0.01	0.06	7.79	611.65	0.24	
NO_3 + Salinity + Temperature + Dist-to-shore	0.01	0.05	6.46	512.13	0.24	
Salinity + Chlorophyll + Temperature + Dist-to-shore	0.01	0.05	6.38	506.52	0.24	
Salinity + NP + Dist-to-shore	0.01	0.04	5.82	463.69	0.21	
PO_4 + Salinity + Temperature + NP + Dist-to-shore	0.01	0.04	5.65	450.61	0.25	
PO_4 + Salinity + Chlorophyll + Temperature + Dist-to-shore	0.01	0.04	5.00	401.24	0.25	
NO_3 + PO_4 + Salinity + Temperature + Dist-to-shore	0.01	0.03	4.57	367.78	0.25	
NO_3 + Salinity + Temperature + NP + Dist-to-shore	0.01	0.03	3.51	284.87	0.25	
Null model	0.50	0.11	0.12	1.00	0.00	Cyanobacteria
Nodularia spumigena	0.50	0.89	8.30	8.30	0.06	

(B)								
						95% Credible Interval		
Coefficient	Mean	SD	P(incl)	P(incl|data)	BFinclusion	Lower	Upper	
Intercept	0.08	0.01	1.00	1.00	1.00	0.06	0.10	Environmental variables
Salinity	-0.01	0.00	0.50	0.97	34.17	-0.02	-0.01	
Temperature	0.02	0.01	0.50	0.71	2.50	0.00	0.03	
Dist-to-shore	0.00	0.00	0.50	0.90	9.43	0.00	0.00	
Intercept	0.09	0.01	1.00	1.00	1.00	0.07	0.11	Cyanobacteria
Nodularia spumigena	0.01	0.00	0.50	0.89	8.30	0.00	0.02	

https://doi.org/10.5194/egusphere-2025-3290
Preprint. Discussion started: 11 August 2025

Intercept	0.09	0.01	1.00	1.00	1.00	0.07	0.11	

Geostatistical interpolation based on the interaction effects between predictors indicated area increases of nodularin into the Eastern Gotland Sea, Northern Baltic Proper, Bornholm and Arkona basin (**Fig 3a, b, c & e**) Western Gotland Sea **Fig 3a&b** as result of interactions between salinity, temperature, distance to shore, PO_4, and chlorophyll. There is an area increase of nodularin into the Bothnian Sea in response to interaction between salinity, temperature, and chlorophyll with distance to shore (**Fig 3a&e)**. The northern parts of the Baltic Sea at Bothnian Bay and southern parts at Kattegat predicted low or no increase in nodularin.

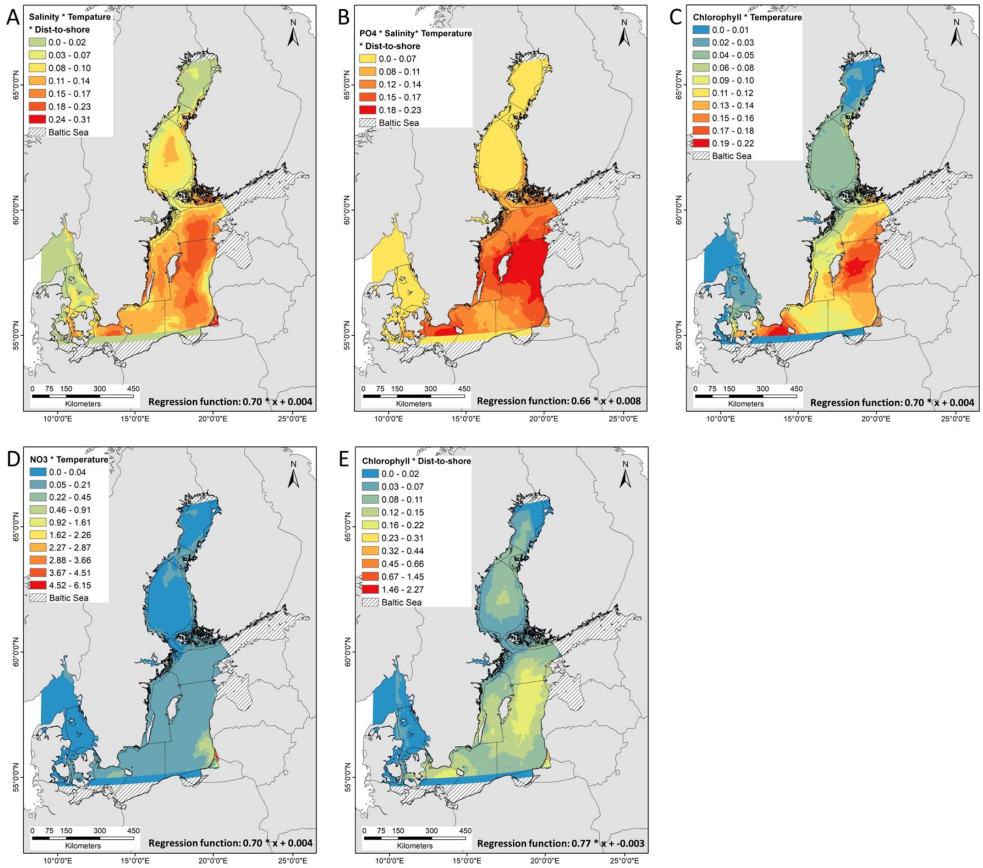
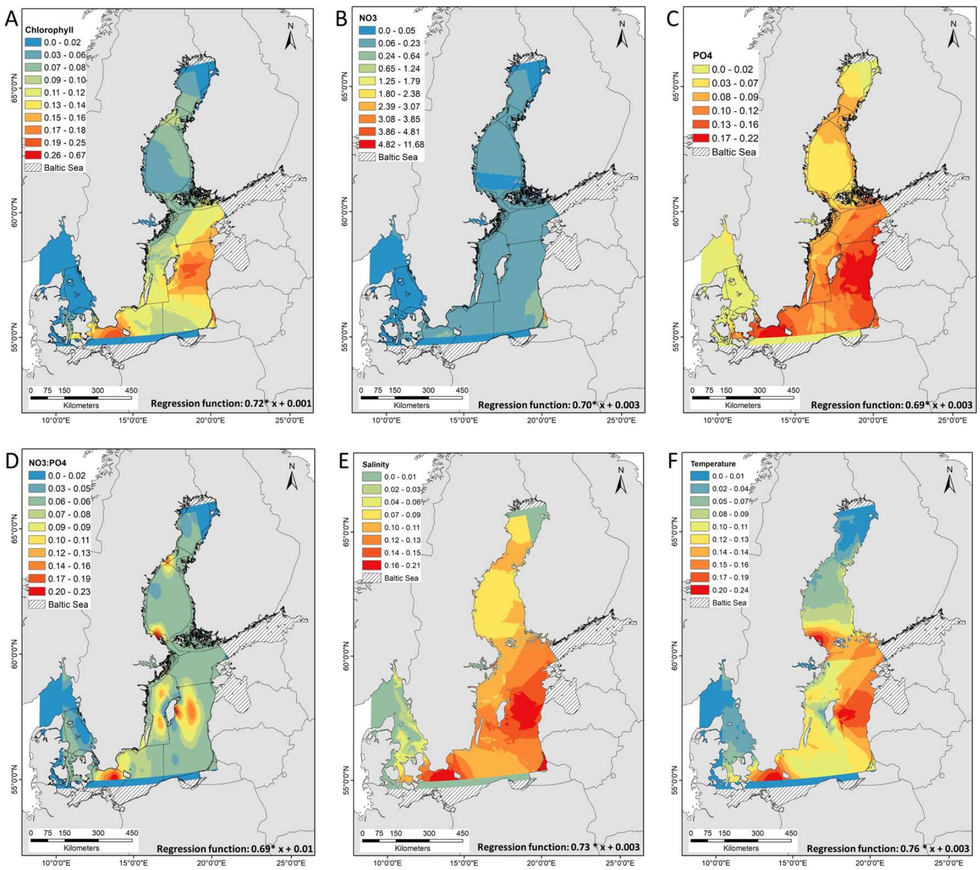


Figure 3: Empirical Bayesian kriging (EBK) regression prediction of the model-predicted concentration and area distribution of nodularin (ug l^{-1}) at different sampling sites across the Baltic Sea. Predicted area distribution and concentrations of nodularin are based on the coefficient of determination R^2 of Bayesian linear regression in (Table S1) and significant interaction effects of ANOVA at p<0.05 (Table S3) between (a) salinity, temperature and distance to shore (b) PO_4, salinity, temperature and distance to shore (c) chlorophyll and temperature, (d) NO_3 and temperature and between (e) chlorophyll and distance to shore. The

https://doi.org/10.5194/egusphere-2025-3290
Preprint. Discussion started: 11 August 2025

regression function of kriging describes the relationship between nodularin concentration and interaction effect between independence variables. Grading colors in key legend corresponds to predicted concentrations of nodularin (ug l^{-1}). Significant interaction effects in (a) and (b) were quantified using Bayesian linear regression while (c), (d) and (e) were assessed in ANOVA. Geostatistical interpolations are performed using nodularin concentration as dependent variable and environmental variables in raster GeoTIFF format as independents. Environmental data was retrieved from E.U. Copernicus Marine Service Information during the period of sampling from June to September 2023 at a depth range of 0.5 to 10 meters.

Predicted sites and area distribution show that there are area increases of nodularin occurrence into Eastern Gotland Sea and Northern Baltic Proper in response to PO_4, salinity temperature, and in response to distance to shore **(Fig. 4c, e, f & g)**; and into Bothnian Sea and Åland Sea in response to Chlorophyll, NO_3:PO_4 ratio, and temperature **(Fig. 4a, d, f)**. There are increases in area distribution of nodularin in Arkona Basin and Bornholm Basin in response to chlorophyll, PO_4, salinity, NO_3:PO_4 ratio, temperature and in response to distance to shore **(Fig. 4a, c, d, e, f &g)**. Despite measured occurrences of nodularin, Kattegat and some parts of the Bothnian bay are predicted to have no or low area distribution of nodularin.



https://doi.org/10.5194/egusphere-2025-3290
Preprint. Discussion started: 11 August 2025

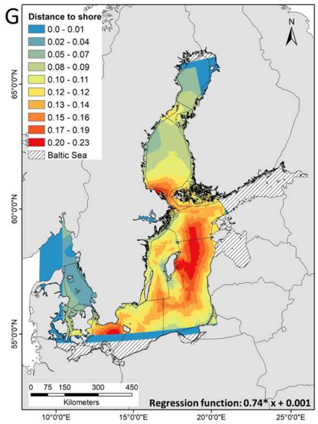


Figure 4: Empirical Bayesian kriging (EBK) regression and co-kriging prediction of the model-predicted concentration and area distribution of nodularin (µg l^{-1}) at different sampling sites across the Baltic Sea. Predicted area distribution and concentrations of nodularin are based on sea surface concentrations of (a) chlorophyll mg m^{-3} (b) nitrate NO_3 µmole m^{-3}, (c) phosphate PO_4 µmole m^{-3}, (d) NO_3:PO_4 ratio, (e) salinity, (f) temperature °C and (g) distance to shore (m). The regression function of kriging describes the relationship between nodularin concentration and each independent variable. Grading colors in key legend corresponds to predicted concentrations of nodularin (µg l^{-1}). Geostatistical interpolations are performed using nodularin concentration as dependent variable and environmental variables in raster GeoTIFF format as independents. Environmental data was retrieved from E.U. Copernicus Marine Service Information during the period of sampling from June to September 2023 at a depth range of 0.5 to 10 meters.

There was a conformity in area distribution of nodularin in response to measured abundances of *N. spumigena.* **(Fig. 5)**. Predicted area distribution shows a slight increase in nodularin occurrence into Western Gotland Sea, Eastern Gotland Sea, Arkona Basin and smaller areas into southern and northern parts of Bothnian Sea. Predicted area distribution of nodularin was significantly driven by abundances of *N. spumigena.* (*ANOVA*, $F=8.51$, $p< 0.001$, $BF_{10} > 1$, $R^2=0.1$) see **Tables 1& S3b**. Predicted area distribution in response to measured abundances of *N. spumigena* show increases in occurrence of nodularin into Northern Baltic Proper and Eastern Gotland Sea, see **Fig. 5** and regression function.

https://doi.org/10.5194/egusphere-2025-3290
Preprint. Discussion started: 11 August 2025

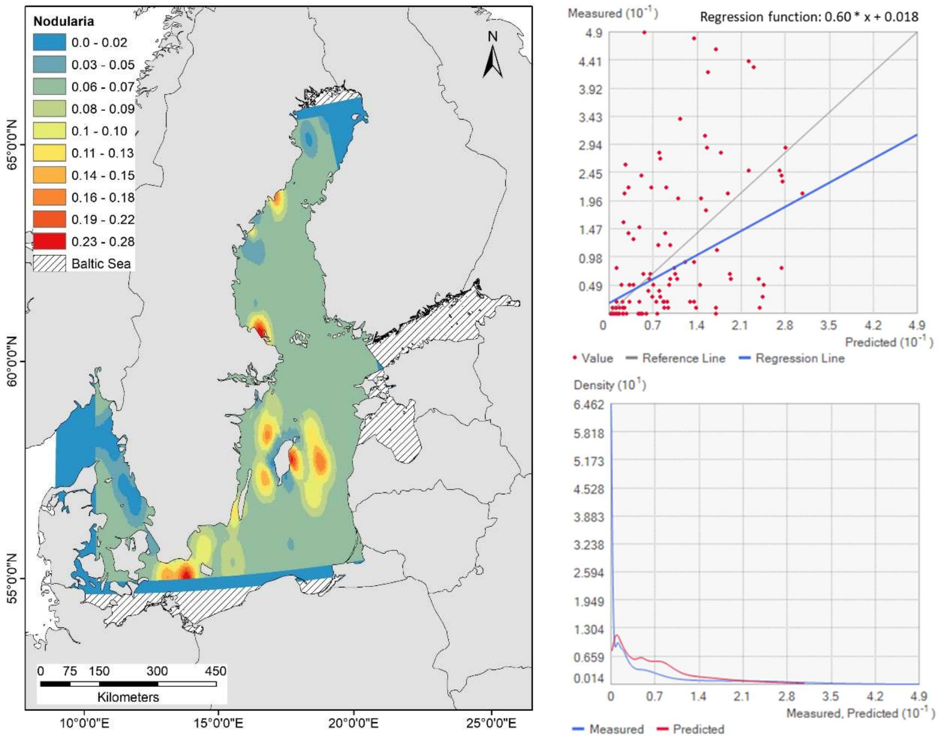


Figure 5: Spatial prediction map of Co-kriging interpolation (left) and distributions of the cross-validation statistics (right), estimated using kernel density, show the prediction regression scatterplot, the regression function of the concentration and predicted area distribution of nodularin (µg l^{-1}) at different sampling sites across the Baltic Sea. Grading colors in key legend corresponds to predicted concentrations of nodularin (µg l^{-1}). The blue and red lines correspond to the measured and predicted values of nodularin (µg l^{-1}). Predicted area distribution and concentrations of nodularin are based on the measured abundances of *Nodularia spumigena.*

Prediction and geostatistical interpolation of nodularin based on measured mean values obtained from Swedish National Oceanographic Data Centre (NODC) at SMHI shows that there were significant interaction effects between temperature and distance-to-shore and between salinity, PO_4, temperature and distance-to-shore on predicted spatial occurrence of nodularin ($BF_{10} > 1$, $R^2=0.1$), **Table S4.** Prediction of occurrence of nodularin based on mean values measured by Swedish National Oceanographic Data Centre (NODC) showed that they are occurring at the Eastern and Western Gotland Sea, Northern Baltic Proper, Bothnian Sea, The Quark and Arkona basin. Predicted area distribution of nodularin based on NODC values showed similar distribution patterns as the prediction made with full location dataset using Copernicus values **(see Fig S2)**.

Future predictions of nodularin occurrence based on projected climate change scenarios (SSP5-8.5) in year 2100 show that there are areas increases of occurrence into the Eastern Gotland Sea, Northern Baltic Proper, the Bothnian Sea, the Quark,

https://doi.org/10.5194/egusphere-2025-3290
Preprint. Discussion started: 11 August 2025

and Arkona Basin in response to chlorophyll, NO_3, PO_4, salinity and temperature (*regression function*: *0.73 * x + 0.005),* **(Fig. 6)**. Bayesian linear regression shows that there is area increase of nodularin in response to NO_3 (*ANOVA,* $F=25.33$, $p< 0.0001$, $BF_{10} > 1$, $R^2=0.25$,), **Fig. S3, Table S5a**. There was a significant linear regression revealed by multiple linear regression (Multiple $R = 0.5$, Multiple $R^2=0.2$, 95% CI [5, 127]) with negative coefficient observed in NO_3 ($p< 0.05$), see **Table S5b.**

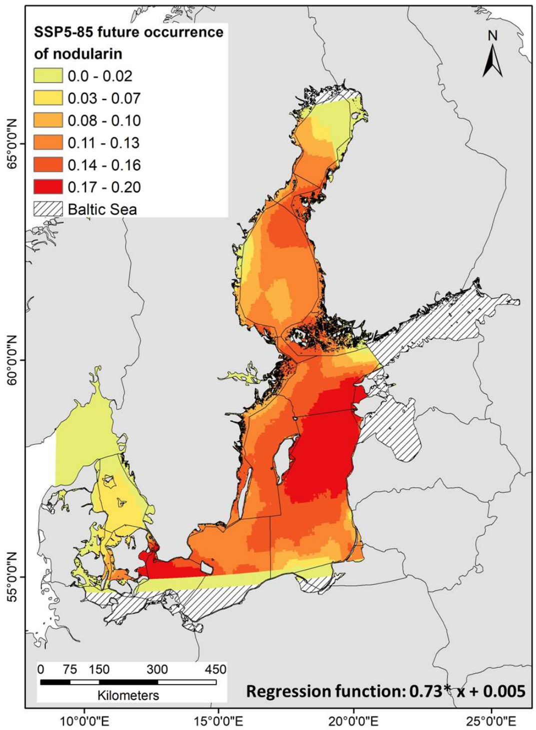


Figure 6: Empirical Bayesian kriging (EBK) regression prediction of the model-predicted concentration and area distribution of nodularin (µg l^{-1}) (estimated using kernel density) show area distribution and regression function of the concentration predicted at different sampling sites across the Baltic Sea. Grading colors in key legend corresponds to predicted concentrations of nodularin (µg l^{-1}). Predicted area distributions are based on projected future concentrations of chlorophyll, NO_3, PO_4, salinity and temperature. Predicted area distribution and concentrations of nodularin are based on Shared Socioeconomic Pathway (SSP5-8.5) scenarios of future climate change corresponding to future greenhouse gas concentrations in the year 2100. Climate change scenarios were downloaded as raster GeoTIFF from Bio-ORACLE project v3.0. Geostatistical interpolations are performed using nodularin concentration as dependent variables and SSP5-8.5 variables in raster GeoTIFF format as independents. Models are validated using 1000 simulations as cross-validation.

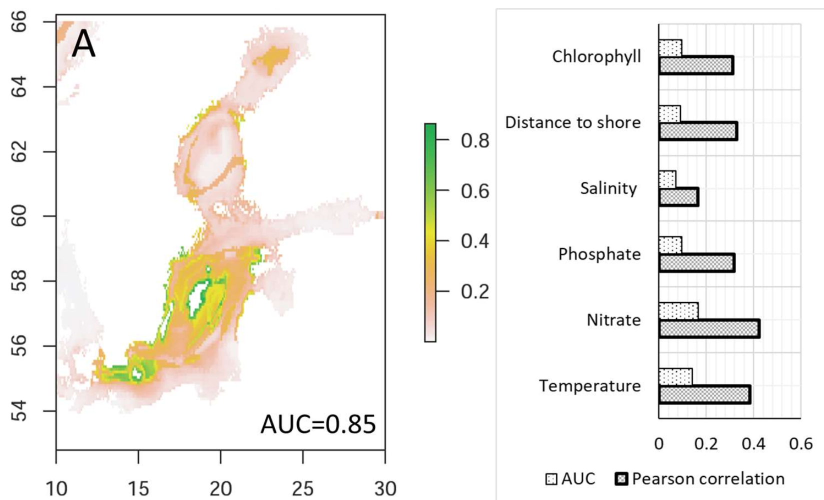
3.2 Ensemble prediction of nodularin occurrence and *Nodularia spumigena* distribution

After removing collinear environmental layers with Pearson correlation coefficient ≥ 0.7, that is NO_3:PO_4, six predictors were used in the ensemble modeling. Since we applied pseudoabsence points as background, data points in areas predicted

https://doi.org/10.5194/egusphere-2025-3290
Preprint. Discussion started: 11 August 2025

by Kriging with no nodularin occurrence, particularly on the west coast, were removed from the ensemble modeling to avoid overestimation. The overall ensemble model performance was excellent *(Current: Threshold=0.35, AUC=0.84, proportion of correct prediction= 0.72; Future: Threshold=0.40, AUC=0.80, proportion of correct prediction= 0.71),* see **Table S6**. The average performance of the six algorithms was excellent (*Current: AUC=0.75±0.05, TSS= 0.48±0.11; Future: AUC=0.77±0.07, TSS= 0.52±0.12*), in which RF, GBM, MAXENT, GLM and MARS had highest AUC values (between 0.73-0.87) indicating excellent performance, **Fig. S4**. Response curves of ensemble learning modeling illustrating the effects of the predicted variables and projected climate change scenarios on nodularin occurrence can be found in **Fig. S5.**

Variables predicted with highest contribution to current nodularin occurrence were NO_3 (*AUC=0.2, Pearson coefficient=0.4*), temperature (*AUC=0.1, Pearson coefficient=0.4*), , Distance to shore (*AUC=0.1, Pearson coefficient=0.4*), PO_4 (*AUC=0.1, Pearson coefficient=0.3*), and salinity (*AUC=0.1, Pearson coefficient=0.2*), see bar chart in **Fig. 7a**. Predicted climate change scenarios with highest contribution were NO_3 (AUC= 0.2, Pearson coefficient = 0.4), salinity (*AUC=0.1, Pearson coefficient=0.3),* and temperature *(AUC=0.1, Pearson coefficient=0.3*), see bar chart in **Fig. 7b.**



https://doi.org/10.5194/egusphere-2025-3290
Preprint. Discussion started: 11 August 2025

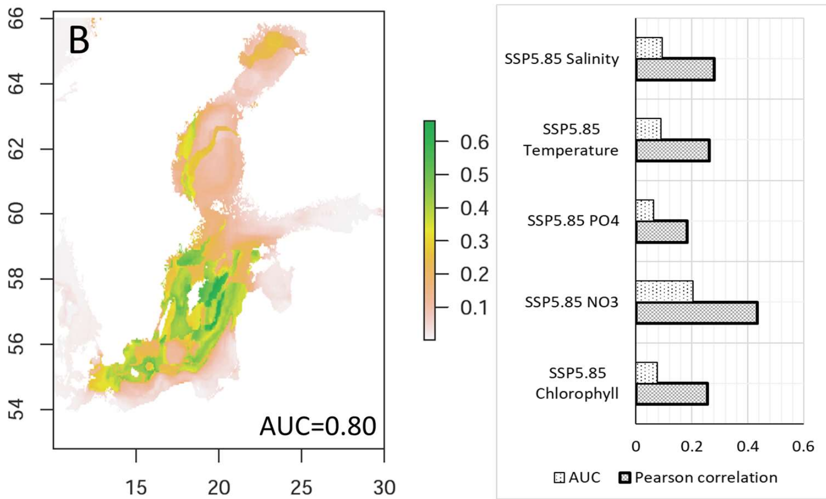
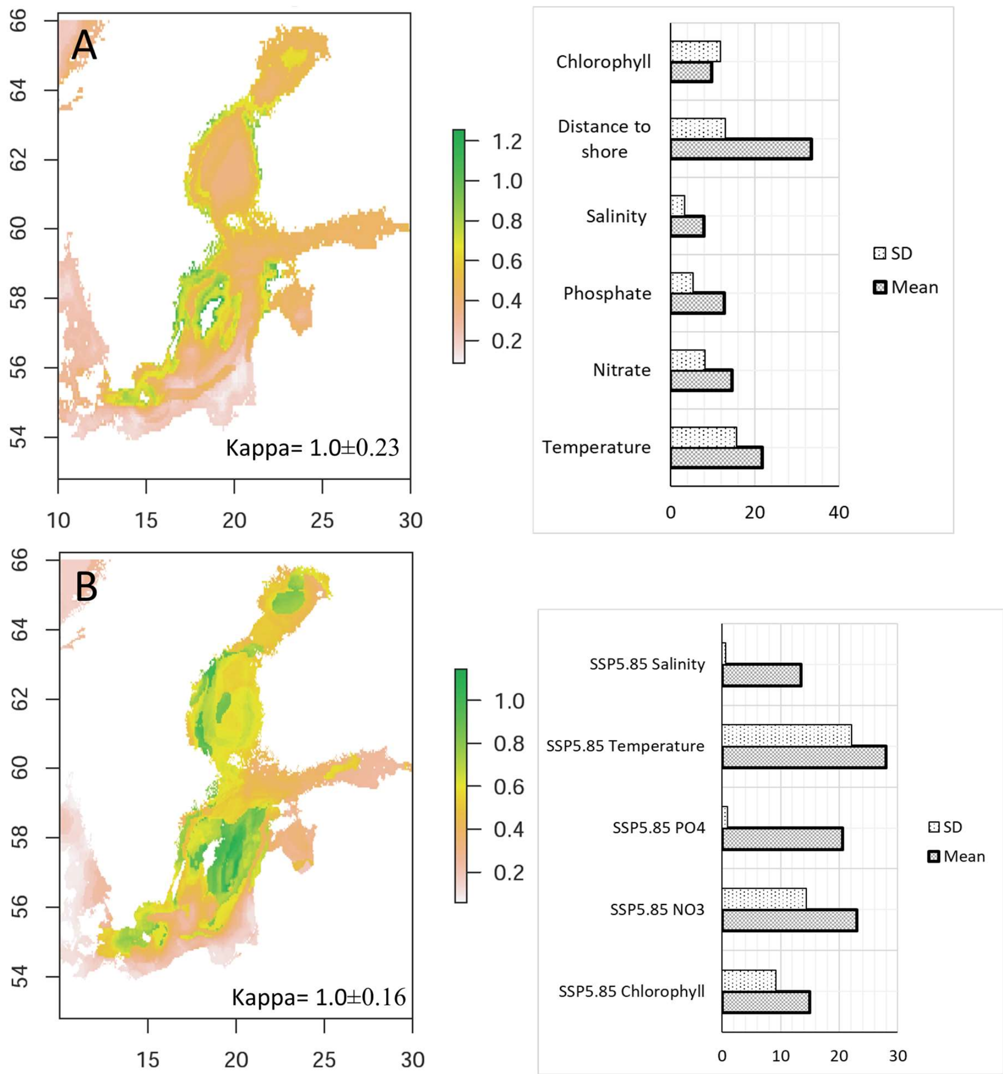


Figure 7: Binary maps show the ensemble learning prediction for nodularin occurrence in response to (a) current environmental variables and (b) future climate change scenarios SSP5-8.5 in 2100. The bar charts show variable contributions to predicted distribution as revealed by the ensemble model. The scale bar from 0.1-0.8 shows predicted suitable areas where dark green colors correspond to areas predicted to have high nodularin occurrence. Models are assessed using area under curve AUC and Pearson correlation coefficient. AUC values close to 1.0 indicate excellent model performance.

There was conformity in predicted area distribution of nodularin between Ensemble learning and EBK regression prediction method. There is an increase in nodularin occurrence predicted into the Eastern Gotland Sea, Arkona Basin and Kattegat in response to the six environmental variables (**Figs. 6a and 7a)**. Area increases of nodularin occurrence in response to future scenarios in 2100 are predicted into the Eastern, Western Gotland Sea, Northern Baltic Proper, Arkona Basin, Kattegat and southern parts of Bothnian Bay in the Quark area **Figs. 6b and 7b**.

There is also a conformity in the predicted area increase of nodularin and predicted distribution for *N. spumigena*. Predicted areas revealed by SSDM are the coastal sites along the Eastern Gotland Sea, and Northern Baltic Proper, along with Western Gotland Sea, Arkona Basin and Kattegat (*Kappa=1.0±0.23*), **Figs 8a**. Predicted area distribution of *N. spumigena* in response to future climate change scenarios are Eastern, Western Gotland Sea, Northern Baltic Proper, Arkona Basin, Kattegat and southern parts of Bothnian Bay in the Quark area (*Kappa=1.0±0.16*), **Figs 8b.**

https://doi.org/10.5194/egusphere-2025-3290
Preprint. Discussion started: 11 August 2025



330 **Figure 8: Binary maps show the predicted distribution of *Nodularia spumigena* estimated by stacked species distribution modeling SSDM in response to (a) current environmental variables and (b) projected climate change scenarios SSP5-8.5 in the year 2100. The bar charts show variable contributions to predicted distribution as revealed by SSDM in Mean±SD. The scale bar from 0.2 - 1.2 unit represents the degree of suitability in which warmer colors correspond to highly predicted suitable areas. Model performances are assessed using Cohen's kappa as Mean±SD, in which Kappa values close to 1.0 indicate excellent model performance.**

https://doi.org/10.5194/egusphere-2025-3290
Preprint. Discussion started: 11 August 2025

4 Discussion

Blooms of toxic cyanobacteria are a global issue, and it's challenging to make predictions of toxin occurrence as the connections to organism abundance are not straightforward. Therefore, understanding bloom formation and toxicity of cyanobacterial blooms in the Baltic Sea is difficult using observational data alone. Here, models can be very effective in increasing spatial and temporal resolution and the interconnection between abundance and toxicity. We integrated and compared the results of two different modeling approaches to understand and interpret the current and future spatial expansion of nodularin occurrence across the Baltic Sea, and our findings show conformities between the approaches. Although there are no expectations on substantial variations in degree of accuracy between geostatistical and ML-based predictions, there has in recent years been a general move from geostatistics to machine learning (Veronesi and Schillaci, 2019). According to the same study, the number of documents mentioning ML algorithms-based predictions has rapidly increased, surpassing kriging-based algorithms in the last decade. Our findings suggest the efficiency of both kriging and ML algorithms given that the overall geographical estimations and maps produced by the kriging and ensemble-based models in this study are aligned despite the differences of approaches.

Several abiotic variables, such as PO_4, NO_3, sea surface temperature, and salinity, are known to directly or indirectly affect and control the abundance of cyanobacteria (Almroth-Rosell et al., 2016; Andersen et al., 2020; Karlson et al., 2010; Lips and Lips, 2008; Lu et al., 2019; Olofsson et al., 2020; Stal, 2009; Unger et al., 2013; Wurtsbaugh et al., 2019; Yang et al., 2008), and therefore we also included many environmental factors. Considering the importance of abiotic variables in modeling cyanobacterial abundance, the corresponding general ocean circulation models GCMs require adequate horizontal and vertical resolution and trustworthy boundary conditions. An earlier study by Reissmann et al., (2009) showed that variations in GCMs are likely to affect the simulation of cyanobacteria (see also Munkes et al., 2021). Therefore, we employed different GCMs to reduce possible errors that could arise from abiotic variables of different ocean models. Our simulation showed that there were conformities in predicted distribution of nodularin despite the different GCMs used in this study. Although each one of the combined approaches has different working procedures, each model effectively yielded similar performance observed in the predicted distribution of nodularin occurrence across the Baltic Sea. This finding indicates that both kriging and ML-based algorithms, the ensemble, can capture the differences in abiotic variables of different GCMs.

Using our model-setup we were able to locate areas with high predicted nodularin concentrations in response to current environmental variables and projected future changes of the same variables. These areas were the Eastern and Western Gotland Sea, the Northern Baltic Proper, southern parts of the Bothnian Sea, and the Arkona basin, corresponding to the effects and interactions between temperature, PO_4, NO_3:PO_4, salinity, and distance to shore. The effects of temperature, PO_4 and NO_3:PO_4 are aligned with previous studies by Deng et al., (2022); Lürling et al., (2017), and Olofsson et al., (2016), documenting that water temperature increase is associated with higher nutrient loads and phosphate release from sediments, all of which could have an impact on cyanobacterial biomass increase. Rising temperature is also associated with terrestrial runoff of organic matter but also causes a decrease in salinity in the northern Baltic Proper, northern parts of the Bothnian Sea, Bothnian Bay, and southward to the Arkona basin (see Andersson et al., 2013, 2015). Moreover, the increase in the abundance of cyanobacteria may also be associated with basin-specific decreases in salinity (Kuosa et al., 2017; Olofsson et al., 2020; Suikkanen et al., 2007), which could highlight the area increase of nodularin predicted in the Arkona basin.

Distance to shore, surprisingly, appears to have significant positive direct and interaction effects together with other variables on the predicted area increase of nodularin. Our findings show that predicted sites with high nodularin concentration are located further offshore. Considering the differences in biogeochemical characteristics between coastal and offshore waters (Vigouroux et al., 2021) and that offshore waters are generally more nutrient-depleted than waters closer to the coast (see Löptien & Dietze, 2022), our results, in contrast to common assumption, suggest that predicted occurrence of nodularin is not only driven by nutrient concentrations like PO_4 and NO_3. However, the predicted occurrence of nodularin in nutrient-depleted areas could be attributed to the limitation of model-based projection of nutrient load scenarios, which do not differentiate between dissolved inorganic phosphorus stored by cyanobacteria and dissolved organic phosphorus. The fact that cyanobacteria can utilize dissolved organic phosphorus under nutrient-depleted conditions (see Caille et al., 2024; Rabouille et al., 2022) could explain the observed and predicted occurrence of nodularin. Our model finding suggests the importance of including distance to shore as a geographical variable in future modeling of cyanobacteria and cyanotoxin, particularly in nutrient-depleted areas across the Baltic Sea.

Results of Bayesian linear regression show that there were variations in the interaction effects between different variables of ERGOM and NODC on the predicted occurrence of nodularin, with a relatively low coefficient of determination $R^2 \leq 0.25$. This statistical observation can be attributed to the differences in resolutions of variables measured by GCM of different origins (see e.g., Dietze & Löptien, 2016; Löptien & Dietze, 2022; Meier et al., 2011; Meier & Kauker, 2003; Munkes et al., 2021; Schrum & Backhaus, 2002). However, these statistical observations are not great enough to offset the spatial regression observed in the regression function $R^2 \geq 0.7$ and the detrimental and significant effects of other variables with $p <$ 0.001, 0.01 and 0.05 observed, e.g., in salinity, temperature, PO_4, distance to shore, and their interaction effects. Despite the statistical results of Bayesian linear regression, our model finding suggests that the spatial regression function produced by the kriging algorithm is useful to interpret the drivers of the predicted occurrences of nodularin.

Future increased occurrence of nodularin was predicted for the Eastern Gotland Sea, Northern Baltic Proper, the Bothnian Sea, the Quark, and Arkona Basin in response to overall projected changes in environmental variables, particularly the NO_3 concentration and temperature. Earlier studies demonstrated that cyanobacterial species, such as *Nodularia spumigena* and *Aphanizomenon* spp., will benefit from climate change due to increased stratification by higher temperatures and that the growth and nitrogen fixation are favored by rising temperatures (Karlberg and Wulff, 2013; Munkes et al., 2021; Paerl and Otten, 2013; Visser et al., 2016). Furthermore, it has been shown that the effects of climate change under different scenarios are smaller than the effects of considered nutrient load changes (Saraiva et al., 2019). Taken together, our model findings suggest that reducing nutrient load could lead to improved future environmental conditions in the Baltic Sea, which in turn could reduce future area distribution of nodularin.

Given the great divergence and large spread between findings of other models projecting cyanobacterial bloom (e.g., Hense et al., 2013; Meier et al., 2012, 2019; Munkes et al., 2021; Neumann, 2010; Saraiva et al., 2019), the variation in GCMs, and that cyanotoxins cannot be directly measured by satellite and remote sensing, our modeling approach that combines geostatistical and ML algorithms could help mitigate this methodological gap. Model prediction of both EBK regression and ensemble learning provided useful and conformable spatial estimation of nodularin occurrence, suggesting that combining results of geostatistical interpolation and ML-based modeling could serve as a promising future approach for cyanotoxin investigation. This approach could also help mitigate future expansion of nodularin by detecting and predicting possible hotspot occurrences that could arise due to different future climatic scenarios across the Baltic Sea. Despite the model's accuracy and performance, care should be taken when interpreting the model findings, given that there was limited data availability that covers both nodularin concentration and abundance of *N. spumigena* in corresponding samples and that the sampling period of the study was conducted only during the summer and in one year.

5 Conclusion

We underscore the need to integrate geostatistical interpolation and machine-learning-based modeling to monitor and mitigate spatial expansion of nodularin across the Baltic Sea and to enhance decision-making strategies in this area. Our models can be useful tools to generate spatial estimates of nodularin occurrence at unsampled locations, and we show evidence that the predicted area distribution of nodularin is a result of several interacting environmental variables together with a geographical factor of distance to shore that could further help explain increased expansion into different areas in the future. Furthermore, modeling using approaches presented in this study could be crucial for effective management applied for nutrient reduction and predicting future climate change impacts, as cyanobacteria can exacerbate eutrophication, affect water quality, and pose health risks. If used carefully, these approaches could help prioritize surveillance and implement earlier sampling efforts in areas predicted to have high cyanotoxin concentration.

Author contributions

MA: Conceptualization (equal), Methodology (lead), Formal analysis (lead), Software (lead), Visualization (lead); Writing-original draft (lead), Writing-review & editing (equal). **BK**: Conceptualization (equal), Funding acquisition, Writing-review & editing (equal). **ED**: Conceptualization (equal), Funding acquisition, Writing-review & editing (equal). **MO**: Data curation (lead), Investigation (lead), Project administration (lead), Resources (lead), Supervision, Funding acquisition, Writing-review & editing (equal).

Acknowledgement

https://doi.org/10.5194/egusphere-2025-3290
Preprint. Discussion started: 11 August 2025

This work was funded by the Swedish Board of Agriculture and European Maritime, Fisheries and Aquaculture Fund (grants nr. 2022-4385 and 2023-1281). In addition, the work was supported by the Swedish Research Council Formas (2022-02117). We thank Ana Cekol and Anabella Aguilera at the Swedish University of Agricultural Sciences for analyzing the toxin samples. Sampling was conducted in connection with the National Swedish Marine Monitoring Program; by the Swedish Meteorological and Hydrological Institute in the offshore Baltic Proper, Umeå University (Umeå Marine Research Centre) in the Gulf of Bothnia, and Stockholm University (Department of Ecology, Environment and Plant Sciences).

References

Abdelgadir, M., Alharbi, R., AlRashidi, M., Alatawi, A. S., Sjöling, S., and Dinnétz, P.: Distribution of denitrifiers predicted by correlative niche modeling of changing environmental conditions and future climatic scenarios across the Baltic Sea, Ecological Informatics, 78, 102346, https://doi.org/10.1016/j.ecoinf.2023.102346, 2023.

Allouche, O., Tsoar, A., and Kadmon, R.: Assessing the accuracy of species distribution models: prevalence, kappa and the true skill statistic (TSS): Assessing the accuracy of distribution models, Journal of Applied Ecology, 43, 1223–1232, https://doi.org/10.1111/j.1365-2664.2006.01214.x, 2006.

Almroth-Rosell, E., Edman, M., Eilola, K., Meier, H. E. M., and Sahlberg, J.: Modelling nutrient retention in the coastal zone of an eutrophic sea, Biogeosciences, 13, 5753–5769, https://doi.org/10.5194/bg-13-5753-2016, 2016.

Andersen, I. M., Williamson, T. J., González, M. J., and Vanni, M. J.: Nitrate, ammonium, and phosphorus drive seasonal nutrient limitation of chlorophytes, cyanobacteria, and diatoms in a hyper-eutrophic reservoir, Limnol Oceanogr, 65, 962–978, https://doi.org/10.1002/lno.11363, 2020.

Andersson, A., Jurgensone, I., Rowe, O. F., Simonelli, P., Bignert, A., Lundberg, E., and Karlsson, J.: Can Humic Water Discharge Counteract Eutrophication in Coastal Waters?, PLoS ONE, 8, e61293, https://doi.org/10.1371/journal.pone.0061293, 2013.

Andersson, A., Höglander, H., Karlsson, C., and Huseby, S.: Key role of phosphorus and nitrogen in regulating cyanobacterial community composition in the northern Baltic Sea, Estuarine, Coastal and Shelf Science, 164, 161–171, https://doi.org/10.1016/j.ecss.2015.07.013, 2015.

Araújo, M. and New, M.: Ensemble forecasting of species distributions, Trends in Ecology & Evolution, 22, 42–47, https://doi.org/10.1016/j.tree.2006.09.010, 2007.

Assis, J., Fernández Bejarano, S. J., Salazar, V. W., Schepers, L., Gouvêa, L., Fragkopoulou, E., Leclercq, F., Vanhoorne, B., Tyberghein, L., Serrão, E. A., Verbruggen, H., and De Clerck, O.: Bio-ORACLE v3.0. Pushing marine data layers to the CMIP6 Earth System Models of climate change research, Global Ecol Biogeogr, 33, e13813, https://doi.org/10.1111/geb.13813, 2024.

Barbet-Massin, M., Jiguet, F., Albert, C. H., and Thuiller, W.: Selecting pseudo-absences for species distribution models: how, where and how many?: How to use pseudo-absences in niche modelling?, Methods in Ecology and Evolution, 3, 327–338, https://doi.org/10.1111/j.2041-210X.2011.00172.x, 2012.

Beery, S., Cole, E., Parker, J., Perona, P., and Winner, K.: Species Distribution Modeling for Machine Learning Practitioners: A Review, in: ACM SIGCAS Conference on Computing and Sustainable Societies (COMPASS), COMPASS '21: ACM SIGCAS Conference on Computing and Sustainable Societies, Virtual Event Australia, 329–348, https://doi.org/10.1145/3460112.3471966, 2021.

Benesty, J., Chen, J., Huang, Y., and Cohen, I.: Pearson Correlation Coefficient, in: Noise Reduction in Speech Processing, vol. 2, Springer Berlin Heidelberg, Berlin, Heidelberg, 1–4, https://doi.org/10.1007/978-3-642-00296-0_5, 2009.

Bivand, R. S., Pebesma, E., and Gomez-Rubio, V.: Applied spatial data analysis with R, Second edition, Springer, NY, 2013.

Bosch, S. and Fernandez, S.: sdmpredictors: Species Distribution Modelling Predictor Datasets, 2022.

Breiman, L.: Random Forests, Machine Learning, 45, 5–32, https://doi.org/10.1023/A:1010933404324, 2001.

Budakoti, S.: Examining the impact of water quality and meteorological drivers on primary productivity in the Baltic Sea, Marine Pollution Bulletin, 209, 117266, https://doi.org/10.1016/j.marpolbul.2024.117266, 2024.

Burford, M. A., Willis, A., Xiao, M., Prentice, M. J., and Hamilton, D. P.: Understanding the relationship between nutrient availability and freshwater cyanobacterial growth and abundance, Inland Waters, 13, 143–152, https://doi.org/10.1080/20442041.2023.2204050, 2023.

Caille, C., Duhamel, S., Latifi, A., and Rabouille, S.: Adaptive Responses of Cyanobacteria to Phosphate Limitation: A Focus on Marine Diazotrophs, Environmental Microbiology, 26, e70023, https://doi.org/10.1111/1462-2920.70023, 2024.

Carlsson, P. and Rita, D.: Sedimentation of Nodularia spumigena and distribution of nodularin in the food web during transport of a cyanobacterial bloom from the Baltic Sea to the Kattegat, Harmful Algae, 86, 74–83, https://doi.org/10.1016/j.hal.2019.05.005, 2019.

Carpenter, S. R.: Eutrophication of aquatic ecosystems: Bistability and soil phosphorus, Proceedings of the National Academy of Sciences, 102, 10002–10005, https://doi.org/10.1073/pnas.0503959102, 2005.

Clyde, M. A., Ghosh, J., and Littman, M. L.: Bayesian Adaptive Sampling for Variable Selection and Model Averaging, Journal of Computational and Graphical Statistics, 20, 80–101, https://doi.org/10.1198/jcgs.2010.09049, 2011.

Deng, L., Cheung, S., Kang, C., Liu, K., Xia, X., and Liu, H.: Elevated temperature relieves phosphorus limitation of marine unicellular diazotrophic cyanobacteria, Limnology & Oceanography, 67, 122–134, https://doi.org/10.1002/lno.11980, 2022.

Deutsch, C. V. and Journel, A. G.: GSLIB: geostatistical software library and user's guide, Oxford University Press, New York, 1992.

Dietterich, T. G.: Ensemble Methods in Machine Learning, in: Multiple Classifier Systems, vol. 1857, Springer Berlin Heidelberg, Berlin, Heidelberg, 1–15, https://doi.org/10.1007/3-540-45014-9_1, 2000.

Dietze, H. and Löptien, U.: Effects of surface current–wind interaction in an eddy-rich general ocean circulation simulation of the Baltic Sea, Ocean Sci., 12, 977–986, https://doi.org/10.5194/os-12-977-2016, 2016.

https://doi.org/10.5194/egusphere-2025-3290
Preprint. Discussion started: 11 August 2025

Dubrule, O.: Cross validation of kriging in a unique neighborhood, Mathematical Geology, 15, 687–699, https://doi.org/10.1007/BF01033232, 1983.

Efron, B.: The Jackknife, the Bootstrap and Other Resampling Plans, Society for Industrial and Applied Mathematics, https://doi.org/10.1137/1.9781611970319, 1982.

Etten, R. J. H. & J. van: raster: Geographic analysis and modeling with raster data, 2012.

European Union-Copernicus Marine Service: Global Ocean 1/12° Physics Analysis and Forecast updated Daily, https://doi.org/10.48670/MOI-00016, 2016.

European Union-Copernicus Marine Service: Baltic Sea Biogeochemistry Reanalysis, https://doi.org/10.48670/MOI-00012, 2018a.

European Union-Copernicus Marine Service: Baltic Sea Physics Reanalysis, https://doi.org/10.48670/MOI-00013, 2018b.

European Union-Copernicus Marine Service: Global Ocean Biogeochemistry Analysis and Forecast, https://doi.org/10.48670/MOI-00015, 2019.

Franklin, J.: Mapping Species Distributions: Spatial Inference and Prediction, 1st ed., Cambridge University Press, https://doi.org/10.1017/CBO9780511810602, 2010.

Friedman, J. H.: Multivariate Adaptive Regression Splines, Ann. Statist., 19, https://doi.org/10.1214/aos/1176347963, 1991.

Friedman, J. H.: Stochastic gradient boosting, Computational Statistics & Data Analysis, 38, 367–378, https://doi.org/10.1016/S0167-9473(01)00065-2, 2002.

Ghannam, R. B. and Techtmann, S. M.: Machine learning applications in microbial ecology, human microbiome studies, and environmental monitoring, Computational and Structural Biotechnology Journal, 19, 1092–1107, https://doi.org/10.1016/j.csbj.2021.01.028, 2021.

Goovaerts, P.: Geostatistics for Natural Resources Evaluation, Oxford University PressNew York, NY, https://doi.org/10.1093/oso/9780195115383.001.0001, 1997.

Goovaerts, P.: Ordinary Cokriging Revisited, Mathematical Geology, 30, 21–42, https://doi.org/10.1023/A:1021757104135, 1998.

Gribov, A. and Krivoruchko, K.: Empirical Bayesian kriging implementation and usage, Science of The Total Environment, 722, 137290, https://doi.org/10.1016/j.scitotenv.2020.137290, 2020.

Guisan, A., Edwards, T. C., and Hastie, T.: Generalized linear and generalized additive models in studies of species distributions: setting the scene, Ecological Modelling, 157, 89–100, https://doi.org/10.1016/S0304-3800(02)00204-1, 2002.

Hanley, J. A. and McNeil, B. J.: The meaning and use of the area under a receiver operating characteristic (ROC) curve., Radiology, 143, 29–36, https://doi.org/10.1148/radiology.143.1.7063747, 1982.

Hansson, M. and Hakansson, B.: The Baltic Algae Watch System - a remote sensing application for monitoring cyanobacterial blooms in the Baltic Sea, J. Appl. Remote Sens, 1, 011507, https://doi.org/10.1117/1.2834769, 2007.

Harrell, F. E., Lee, K. L., and Mark, D. B.: Prognostic/Clinical Prediction Models: Multivariable Prognostic Models: Issues in Developing Models, Evaluating Assumptions and Adequacy, and Measuring and Reducing Errors, in: Tutorials in

Biostatistics, edited by: D'Agostino, R. B., John Wiley & Sons, Ltd, Chichester, UK, 223–249, https://doi.org/10.1002/0470023678.ch2b(i), 2005.

Hense, I., Meier, H. E. M., and Sonntag, S.: Projected climate change impact on Baltic Sea cyanobacteria: Climate change impact on cyanobacteria, Climatic Change, 119, 391–406, https://doi.org/10.1007/s10584-013-0702-y, 2013.

Hieronymus, J., Eilola, K., Olofsson, M., Hense, I., Meier, H. E. M., and Almroth-Rosell, E.: Modeling cyanobacteria life cycle dynamics and historical nitrogen fixation in the Baltic Proper, Biogeosciences, 18, 6213–6227, https://doi.org/10.5194/bg-18-6213-2021, 2021.

Hijmans & Etten: raster: Geographic analysis and modeling with raster data, 2012.

Jaccard, P.: The distribution of the flora in the alpine zone., New Phytol, 11, 37–50, https://doi.org/10.1111/j.1469-8137.1912.tb05611.x, 1912.

JASP Team: JASP (Version 0.19.3)[Computer software], 2025.

Jiang, Y., Luo, J., Huang, D., Liu, Y., and Li, D.: Machine Learning Advances in Microbiology: A Review of Methods and Applications, Front. Microbiol., 13, 925454, https://doi.org/10.3389/fmicb.2022.925454, 2022.

Jonasson, S., Vintila, S., Sivonen, K., and El-Shehawy, R.: Expression of the nodularin synthetase genes in the Baltic Sea bloom-former cyanobacterium Nodularia spumigena strain AV1: Expression of nda genes in Nodularia spumigena, FEMS Microbiology Ecology, 65, 31–39, https://doi.org/10.1111/j.1574-6941.2008.00499.x, 2008.

Kahru, M. and Elmgren, R.: Multidecadal time series of satellite-detected accumulations of cyanobacteria in the Baltic Sea, Biogeosciences, 11, 3619–3633, https://doi.org/10.5194/bg-11-3619-2014, 2014.

Kahru, M., Elmgren, R., Kaiser, J., Wasmund, N., and Savchuk, O.: Cyanobacterial blooms in the Baltic Sea: Correlations with environmental factors, Harmful Algae, 92, 101739, https://doi.org/10.1016/j.hal.2019.101739, 2020.

Karjalainen, M., Engström-Öst, J., Korpinen, S., Peltonen, H., Pääkkönen, J.-P., Rönkkönen, S., Suikkanen, S., and Viitasalo, M.: Ecosystem Consequences of Cyanobacteria in the Northern Baltic Sea, AMBIO: A Journal of the Human Environment, 36, 195–202, https://doi.org/10.1579/0044-7447(2007)36[195:ECOCIT]2.0.CO;2, 2007.

Karlberg, M. and Wulff, A.: Impact of temperature and species interaction on filamentous cyanobacteria may be more important than salinity and increased pCO2 levels, Mar Biol, 160, 2063–2072, https://doi.org/10.1007/s00227-012-2078-3, 2013.

Karlson, B., Eilola, K., and Hansson, M.: Cyanobacterial blooms in the Baltic Sea: Correlating bloom observations with environmental conditions, Proc 13th Int Conf on Harmful Algae, 247–252, 2010.

Karlson, B., Arneborg, L., Johansson, J., Linders, J., Liu, Y., and Olofsson, M.: A suggested climate service for cyanobacteria blooms in the Baltic Sea – Comparing three monitoring methods, Harmful Algae, 118, 102291, https://doi.org/10.1016/j.hal.2022.102291, 2022.

Klawonn, I., Nahar, N., Walve, J., Andersson, B., Olofsson, M., Svedén, J. B., Littmann, S., Whitehouse, M. J., Kuypers, M. M. M., and Ploug, H.: Cell-specific nitrogen- and carbon-fixation of cyanobacteria in a temperate marine system (Baltic Sea), Environmental Microbiology, 18, 4596–4609, https://doi.org/10.1111/1462-2920.13557, 2016.

https://doi.org/10.5194/egusphere-2025-3290
Preprint. Discussion started: 11 August 2025

Krivoruchko, K.: Empirical bayesian kriging implemented in ArcGIS geostatistical analyst, ArcUser, 15, 6–10, 2012.

Krivoruchko, K. and Gribov, A.: Evaluation of empirical Bayesian kriging, Spatial Statistics, 32, 100368, https://doi.org/10.1016/j.spasta.2019.100368, 2019.

Kuosa, H., Fleming-Lehtinen, V., Lehtinen, S., Lehtiniemi, M., Nygård, H., Raateoja, M., Raitaniemi, J., Tuimala, J., Uusitalo, L., and Suikkanen, S.: A retrospective view of the development of the Gulf of Bothnia ecosystem, Journal of Marine Systems, 167, 78–92, https://doi.org/10.1016/j.jmarsys.2016.11.020, 2017.

Lehtimäki, J., Sivonen, K., Luukkainen, R., and Niemelä, S. I.: The effects of incubation time, temperature, light, salinity, and phosphorus on growth and hepatotoxin production by Nodularia strains, archiv_hydrobiologie, 130, 269–282, https://doi.org/10.1127/archiv-hydrobiol/130/1994/269, 1994.

Lehtimaki, J., Moisander, P., Sivonen, K., and Kononen, K.: Growth, nitrogen fixation, and nodularin production by two Baltic Sea cyanobacteria, Applied and environmental microbiology, 63, 1647–1656, 1997.

Lima, T. A., Beuchle, R., Langner, A., Grecchi, R. C., Griess, V. C., and Achard, F.: Comparing Sentinel-2 MSI and Landsat 8 OLI Imagery for Monitoring Selective Logging in the Brazilian Amazon, Remote Sensing, 11, 961, https://doi.org/10.3390/rs11080961, 2019.

Lips, I. and Lips, U.: Abiotic factors influencing cyanobacterial bloom development in the Gulf of Finland (Baltic Sea), Hydrobiologia, 614, 133–140, https://doi.org/10.1007/s10750-008-9449-2, 2008.

Löptien, U. and Dietze, H.: Retracing cyanobacteria blooms in the Baltic Sea, Sci Rep, 12, 10873, https://doi.org/10.1038/s41598-022-14880-w, 2022.

Lu, J., Zhu, B., Struewing, I., Xu, N., and Duan, S.: Nitrogen–phosphorus-associated metabolic activities during the development of a cyanobacterial bloom revealed by metatranscriptomics, Sci Rep, 9, 2480, https://doi.org/10.1038/s41598-019-38481-2, 2019.

Lundholm, N., Bernard, C., Churro, C., Escalera, L., Fraga, S., Hoppenrath, M., Iwataki, M., Larsen, J., Mertens, K., Moestrup, Ø., Murray, S., Salas, R., Tillmann, U., and Zingone, A.: IOC-UNESCO Taxonomic Reference List of Harmful Micro Algae. Accessed at https://www.marinespecies.org/hab on yyyy-mm-dd, https://doi.org/10.14284/362, 2009.

Lürling, M., Van Oosterhout, F., and Faassen, E.: Eutrophication and Warming Boost Cyanobacterial Biomass and Microcystins, Toxins, 9, 64, https://doi.org/10.3390/toxins9020064, 2017.

Meier, H. E. M. and Kauker, F.: Modeling decadal variability of the Baltic Sea: 2. Role of freshwater inflow and large-scale atmospheric circulation for salinity, J. Geophys. Res., 108, 2003JC001799, https://doi.org/10.1029/2003JC001799, 2003.

Meier, H. E. M., Höglund, A., Döscher, R., Andersson, H., Löptien, U., and Kjellström, E.: Quality assessment of atmospheric surface fields over the Baltic Sea from an ensemble of regional climate model simulations with respect to ocean dynamics, Oceanologia, 53, 193–227, https://doi.org/10.5697/oc.53-1-TI.193, 2011.

Meier, H. E. M., Müller-Karulis, B., Andersson, H. C., Dieterich, C., Eilola, K., Gustafsson, B. G., Höglund, A., Hordoir, R., Kuznetsov, I., Neumann, T., Ranjbar, Z., Savchuk, O. P., and Schimanke, S.: Impact of Climate Change on Ecological

https://doi.org/10.5194/egusphere-2025-3290
Preprint. Discussion started: 11 August 2025

Quality Indicators and Biogeochemical Fluxes in the Baltic Sea: A Multi-Model Ensemble Study, AMBIO, 41, 558–573, https://doi.org/10.1007/s13280-012-0320-3, 2012.

Meier, H. E. M., Dieterich, C., Eilola, K., Gröger, M., Höglund, A., Radtke, H., Saraiva, S., and Wåhlström, I.: Future projections of record-breaking sea surface temperature and cyanobacteria bloom events in the Baltic Sea, Ambio, 48, 1362–1376, https://doi.org/10.1007/s13280-019-01235-5, 2019.

Moisander, P. H., McClinton, E., and Paerl, H. W.: Salinity Effects on Growth, Photosynthetic Parameters, and Nitrogenase Activity in Estuarine Planktonic Cyanobacteria, Microb Ecol, 43, 432–442, https://doi.org/10.1007/s00248-001-1044-2, 2002.

Monserud, R. A. and Leemans, R.: Comparing global vegetation maps with the Kappa statistic, Ecological Modelling, 62, 275–293, https://doi.org/10.1016/0304-3800(92)90003-W, 1992.

Munkes, B., Löptien, U., and Dietze, H.: Cyanobacteria blooms in the Baltic Sea: a review of models and facts, Biogeosciences, 18, 2347–2378, https://doi.org/10.5194/bg-18-2347-2021, 2021.

Naimi, B. and Araújo, M. B.: sdm: a reproducible and extensible R platform for species distribution modelling, Ecography, 39, 368–375, https://doi.org/10.1111/ecog.01881, 2016.

Neumann, T.: Climate-change effects on the Baltic Sea ecosystem: A model study, Journal of Marine Systems, 81, 213–224, https://doi.org/10.1016/j.jmarsys.2009.12.001, 2010.

Neumann, T., Koponen, S., Attila, J., Brockmann, C., Kallio, K., Kervinen, M., Mazeran, C., Müller, D., Philipson, P., Thulin, S., Väkevä, S., and Ylöstalo, P.: Optical model for the Baltic Sea with an explicit CDOM state variable: a case study with Model ERGOM (version 1.2), Geosci. Model Dev., 14, 5049–5062, https://doi.org/10.5194/gmd-14-5049-2021, 2021.

Olea, R. A.: Geostatistics for engineers and earth scientists, Kluwer Academic, Boston, 303 pp., 1999.

Olenina, I., Hajdu, S., Edler, L., Andersson, A., Wasmund, N., Busch, S., Göbel, J., Gromisz, S., Huseby, S., Huttunen, M., Jaanus, A., Kokkonen, P., Jurgensone, I., and Niemkiewicz, E.: Biovolumes and size-classes of phytoplankton in the Baltic Sea, HELCOM Balt. Sea Environ. Proc., 106, 2006.

Oliver, M. A. and Webster, R.: Kriging: a method of interpolation for geographical information systems, International journal of geographical information systems, 4, 313–332, https://doi.org/10.1080/02693799008941549, 1990.

Olofsson, M., Egardt, J., Singh, A., and Ploug, H.: Inorganic phosphorus enrichments in Baltic Sea water have large effects on growth, carbon fixation, and N2 fixation by Nodularia spumigena, Aquat. Microb. Ecol., 77, 111–123, https://doi.org/10.3354/ame01795, 2016.

Olofsson, M., Suikkanen, S., Kobos, J., Wasmund, N., and Karlson, B.: Basin-specific changes in filamentous cyanobacteria community composition across four decades in the Baltic Sea, Harmful Algae, 91, 101685, https://doi.org/10.1016/j.hal.2019.101685, 2020.

Paerl, H. W. and Otten, T. G.: Harmful Cyanobacterial Blooms: Causes, Consequences, and Controls, Microb Ecol, 65, 995–1010, https://doi.org/10.1007/s00248-012-0159-y, 2013.

https://doi.org/10.5194/egusphere-2025-3290
Preprint. Discussion started: 11 August 2025

Paerl, H. W., Otten, T. G., and Kudela, R.: Mitigating the Expansion of Harmful Algal Blooms Across the Freshwater-to-Marine Continuum, Environmental Science & Technology, 52, 5519–5529, https://doi.org/10.1021/acs.est.7b05950, 2018.

Pebesma, E. J. and Bivand, R.: Classes and methods for spatial data in R, R News, 5, 9–13, 2005.

Phillips, S. J., Anderson, R. P., and Schapire, R. E.: Maximum entropy modeling of species geographic distributions, Ecological Modelling, 190, 231–259, https://doi.org/10.1016/j.ecolmodel.2005.03.026, 2006.

Pliński, M., Mazur-Marzec, H., Jóźwiak, T., and Kobos, J.: The potential causes of cyanobacterial blooms in Baltic Sea estuaries, Oceanological and Hydrobiological Studies, 36, 134–137, https://doi.org/10.2478/v10009-007-0001-x, 2007.

Pyrcz, M. and Deutsch, C. V.: Geostatistical reservoir modeling, Second edition., New York, New York : Oxford University Press, Oxford, 433 pp., 2014.

QGIS Development Team: QGIS Geographic Information System, Open Source Geospatial Foundation, 2022.

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

Rabouille, S., Tournier, L., Duhamel, S., Claquin, P., Crispi, O., Talec, A., Landolfi, A., and Oschlies, A.: Organic Phosphorus Scavenging Supports Efficient Growth of Diazotrophic Cyanobacteria Under Phosphate Depletion, Front. Microbiol., 13, 848647, https://doi.org/10.3389/fmicb.2022.848647, 2022.

Reissmann, J. H., Burchard, H., Feistel, R., Hagen, E., Lass, H. U., Mohrholz, V., Nausch, G., Umlauf, L., and Wieczorek, G.: Vertical mixing in the Baltic Sea and consequences for eutrophication – A review, Progress in Oceanography, 82, 47–80, https://doi.org/10.1016/j.pocean.2007.10.004, 2009.

Roberts, D. R., Bahn, V., Ciuti, S., Boyce, M. S., Elith, J., Guillera-Arroita, G., Hauenstein, S., Lahoz-Monfort, J. J., Schröder, B., Thuiller, W., Warton, D. I., Wintle, B. A., Hartig, F., and Dormann, C. F.: Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure, Ecography, 40, 913–929, https://doi.org/10.1111/ecog.02881, 2017.

Ross, C., Santiago-Vázquez, L., and Paul, V.: Toxin release in response to oxidative stress and programmed cell death in the cyanobacterium Microcystis aeruginosa, Aquatic Toxicology, 78, 66–73, https://doi.org/10.1016/j.aquatox.2006.02.007, 2006.

RStudio Team: RStudio: Integrated Development Environment for R, RStudio, PBC., Boston, MA, 2020.

Saraiva, S., Markus Meier, H. E., Andersson, H., Höglund, A., Dieterich, C., Gröger, M., Hordoir, R., and Eilola, K.: Baltic Sea ecosystem response to various nutrient load scenarios in present and future climates, Clim Dyn, 52, 3369–3387, https://doi.org/10.1007/s00382-018-4330-0, 2019.

Sathya, R. and Abraham, A.: Comparison of Supervised and Unsupervised Learning Algorithms for Pattern Classification, IJARAI, 2, https://doi.org/10.14569/IJARAI.2013.020206, 2013.

Schmitt, S., Pouteau, R., Justeau, D., Boissieu, F., and Birnbaum, P.: ssdm: An r package to predict distribution of species richness and composition based on stacked species distribution models, Methods Ecol Evol, 8, 1795–1803, https://doi.org/10.1111/2041-210X.12841, 2017.

Schrum, C. and Backhaus, J. O.: Sensitivity of atmosphere-ocean heat exchange and heat content in the North Sea and the Baltic Sea: atmosphere-ocean heat exchange and heat content, Tellus A, 51, 526–549, https://doi.org/10.1034/j.1600-0870.1992.00006.x, 2002.

Silveira, S. B. and Odebrecht, C.: Effects of Salinity and Temperature on the Growth, Toxin Production, and Akinete Germination of the Cyanobacterium Nodularia spumigena, Front. Mar. Sci., 6, 339, https://doi.org/10.3389/fmars.2019.00339, 2019.

Stal, L. J.: Is the distribution of nitrogen-fixing cyanobacteria in the oceans related to temperature?. Environmental Microbiology, 11, 1632–1645, https://doi.org/10.1111/j.1758-2229.2009.00016.x, 2009.

Steinberg, D.: Chapter 10 CART: Classification and Regression Trees, 2009.

Suikkanen, S., Laamanen, M., and Huttunen, M.: Long-term changes in summer phytoplankton communities of the open northern Baltic Sea, Estuarine, Coastal and Shelf Science, 71, 580–592, https://doi.org/10.1016/j.ecss.2006.09.004, 2007.

The jamovi project: The jamovi project (2023). jamovi (Version 2.3) [Computer Software]. Retrieved from https://www.jamovi.org, 2023.

Thompson, J., Johansen, R., Dunbar, J., and Munsky, B.: Machine learning to predict microbial community functions: An analysis of dissolved organic carbon from litter decomposition, PLoS ONE, 14, e0215502, https://doi.org/10.1371/journal.pone.0215502, 2019.

Unger, J., Endres, S., Wannicke, N., Engel, A., Voss, M., Nausch, G., and Nausch, M.: Response of Nodularia spumigena to pCO_2 – Part 3: Turnover of phosphorus compounds, Biogeosciences, 10, 1483–1499, https://doi.org/10.5194/bg-10-1483-2013, 2013.

Veronesi, F. and Schillaci, C.: Comparison between geostatistical and machine learning models as predictors of topsoil organic carbon with a focus on local uncertainty estimation, Ecological Indicators, 101, 1032–1044, https://doi.org/10.1016/j.ecolind.2019.02.026, 2019.

Vigouroux, G., Kari, E., Beltrán-Abaunza, J. M., Uotila, P., Yuan, D., and Destouni, G.: Trend correlations for coastal eutrophication and its main local and whole-sea drivers – Application to the Baltic Sea, Science of The Total Environment, 779, 146367, https://doi.org/10.1016/j.scitotenv.2021.146367, 2021.

Visser, P. M., Verspagen, J. M. H., Sandrini, G., Stal, L. J., Matthijs, H. C. P., Davis, T. W., Paerl, H. W., and Huisman, J.: How rising CO2 and global warming may stimulate harmful cyanobacterial blooms, Harmful Algae, 54, 145–159, https://doi.org/10.1016/j.hal.2015.12.006, 2016.

Walls, J. T., Wyatt, K. H., Doll, J. C., Rubenstein, E. M., and Rober, A. R.: Hot and toxic: Temperature regulates microcystin release from cyanobacteria, Science of The Total Environment, 610–611, 786–795, https://doi.org/10.1016/j.scitotenv.2017.08.149, 2018.

Wang, H., Guan, Y., and Reich, B.: Nearest-Neighbor Neural Networks for Geostatistics, in: 2019 International Conference on Data Mining Workshops (ICDMW), 2019 International Conference on Data Mining Workshops (ICDMW), Beijing, China, 196–205, https://doi.org/10.1109/ICDMW.2019.00038, 2019.

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T., Miller, E., Bache, S., Müller, K., Ooms, J., Robinson, D., Seidel, D., Spinu, V., Takahashi, K., Vaughan, D., Wilke, C., Woo, K., and Yutani, H.: Welcome to the Tidyverse, JOSS, 4, 1686, https://doi.org/10.21105/joss.01686, 2019.

Wurtsbaugh, W. A., Paerl, H. W., and Dodds, W. K.: Nutrients, eutrophication and harmful algal blooms along the freshwater to marine continuum, WIREs Water, 6, https://doi.org/10.1002/wat2.1373, 2019.

Yang, X., Wu, X., Hao, H., and He, Z.: Mechanisms and assessment of water eutrophication, J. Zhejiang Univ. Sci. B, 9, 197–209, https://doi.org/10.1631/jzus.B0710626, 2008.

Zhang, H. and Wang, Y.: Kriging and cross-validation for massive spatial data, Environmetrics, 21, 290–304, https://doi.org/10.1002/env.1023, 2010.