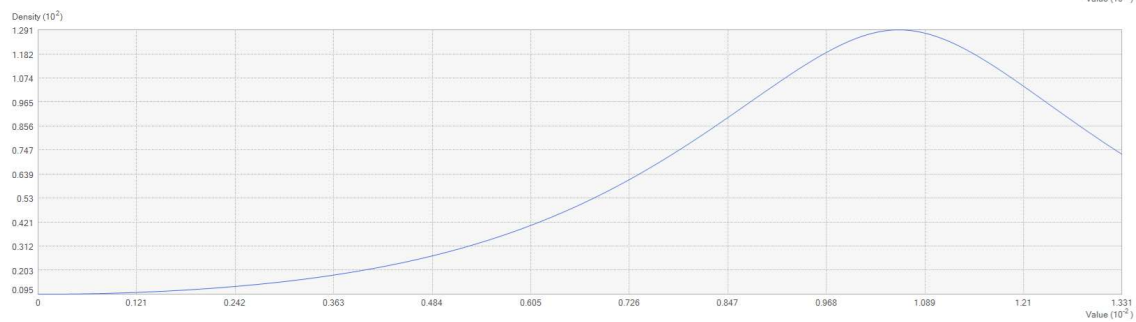
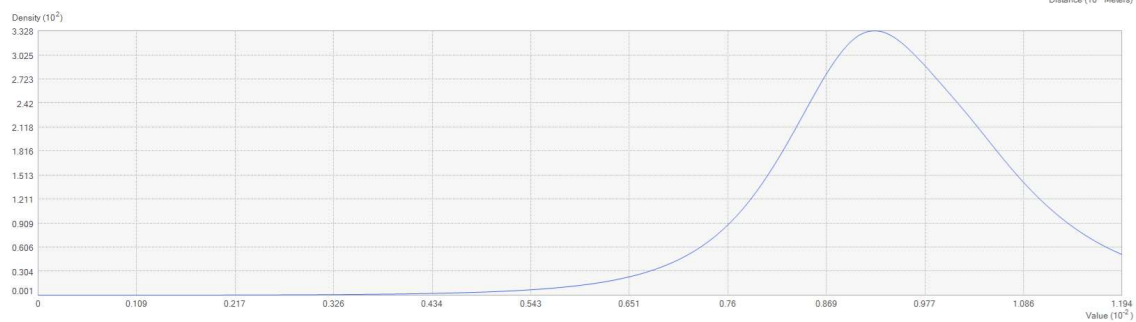
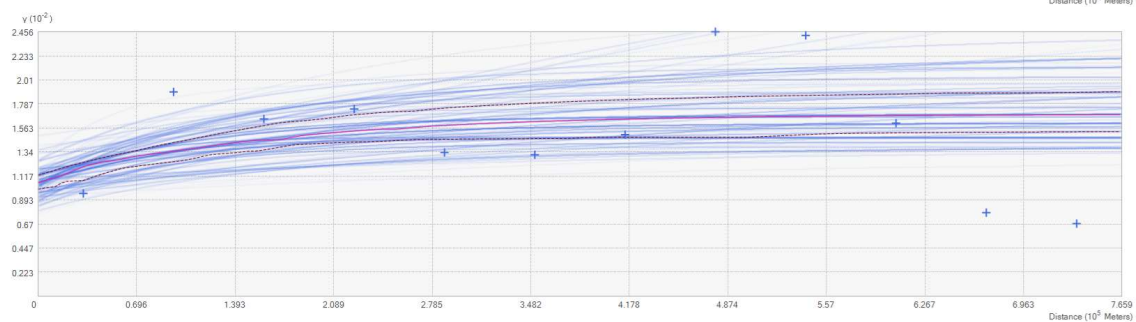
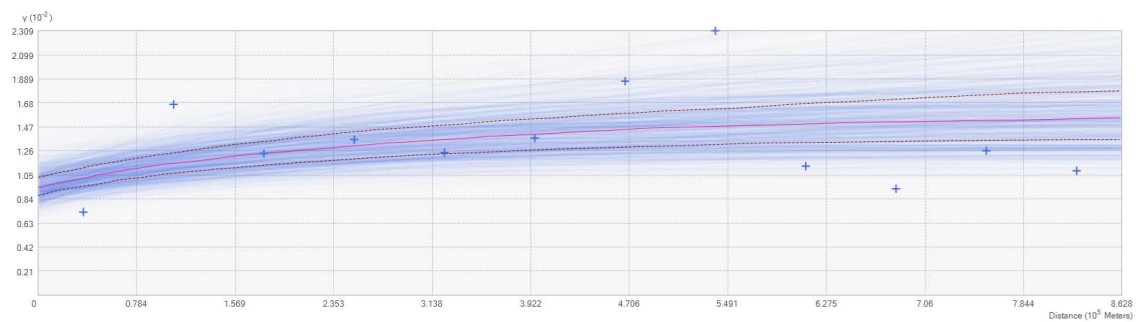
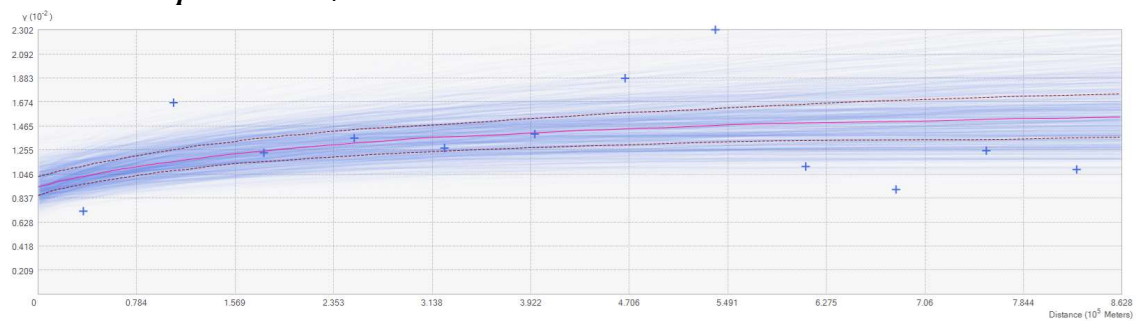
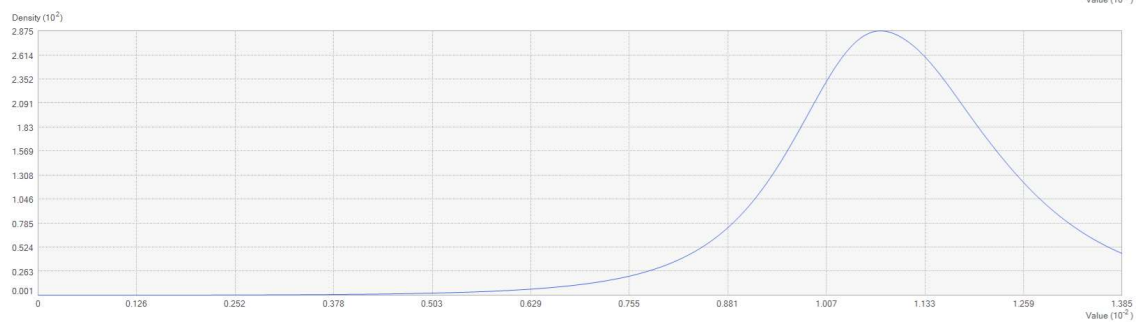
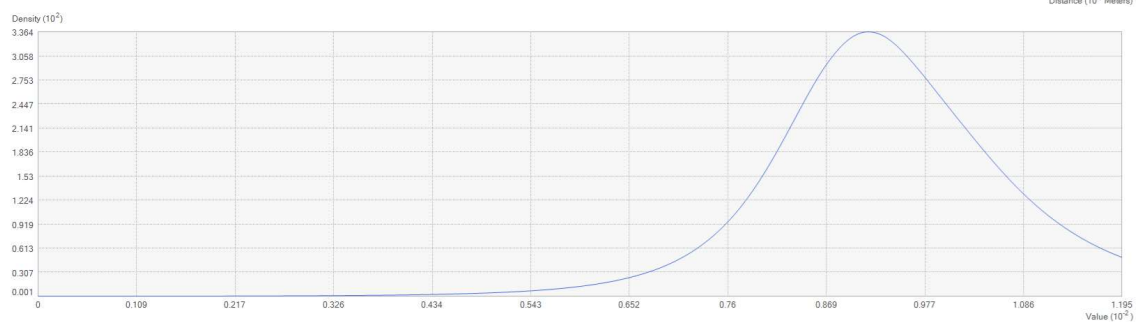
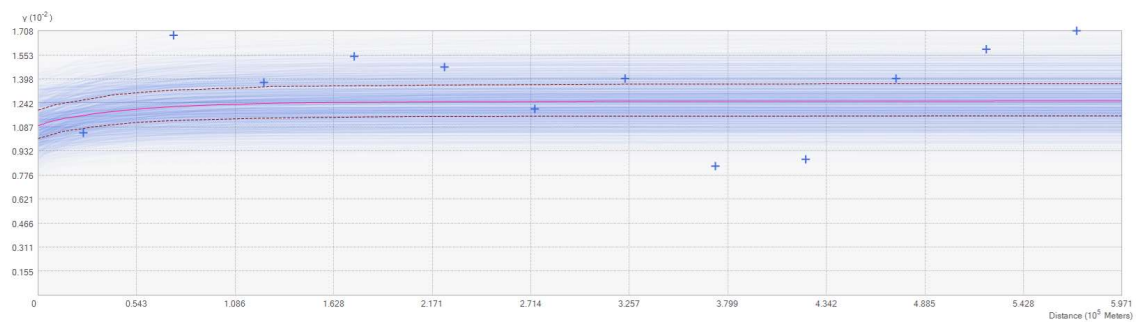


**Figure S1.** Distributions of the cross-validation statistics (estimated using kernel density) show the prediction regression scatterplot, the regression function, and the measured (blue line) and predicted (red line) values of nodularin ( $\mu\text{g l}^{-1}$ ). Values correspond to nodularin response to concentrations of sea surface (a) chlorophyll  $\text{mg m}^{-3}$  (a) nitrate  $\text{NO}_3 \mu\text{mol m}^{-3}$ , (c) phosphate  $\text{PO}_4 \mu\text{mol m}^{-3}$ , (d)  $\text{NO}_3:\text{PO}_4$  ratio, (e) salinity, (f) temperature  $^{\circ}\text{C}$  and (g) distance to shore (m). 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.

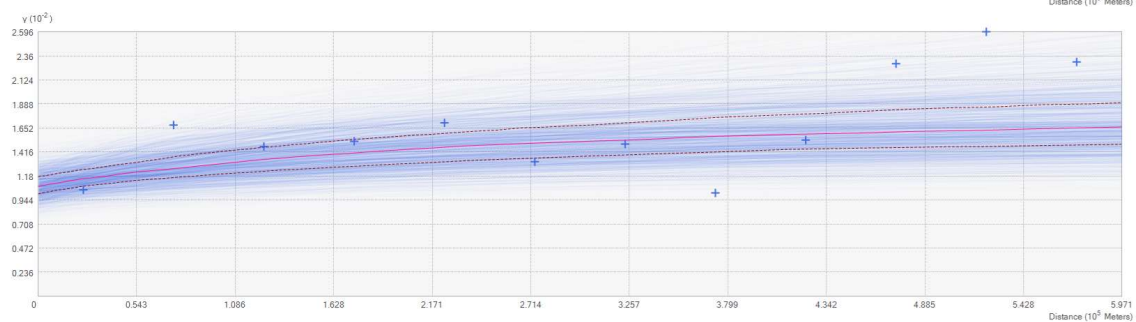
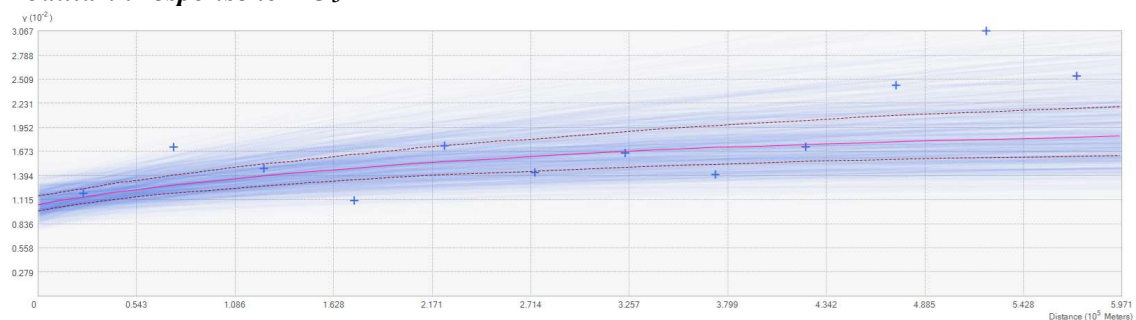


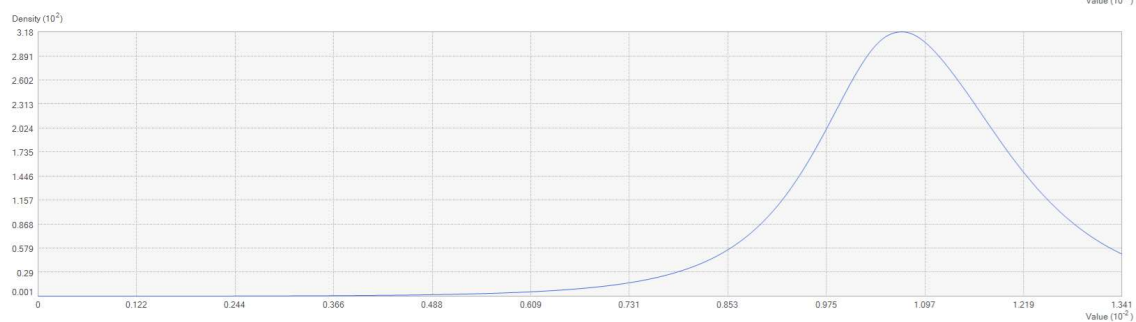
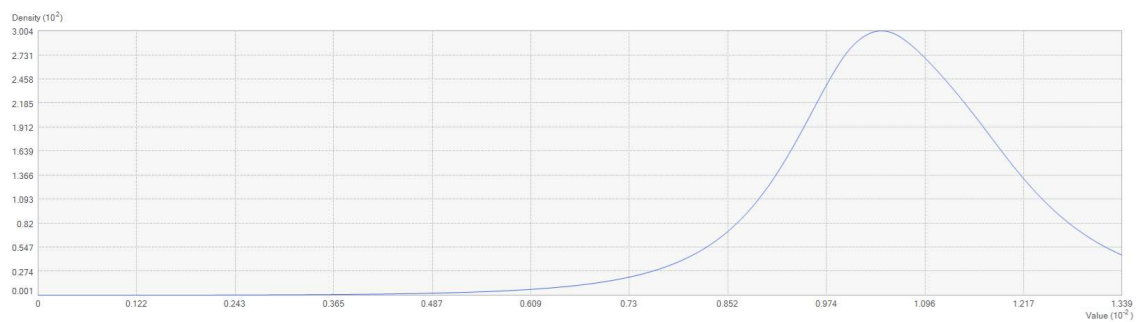
### ***Nodularin response to $PO_4$***



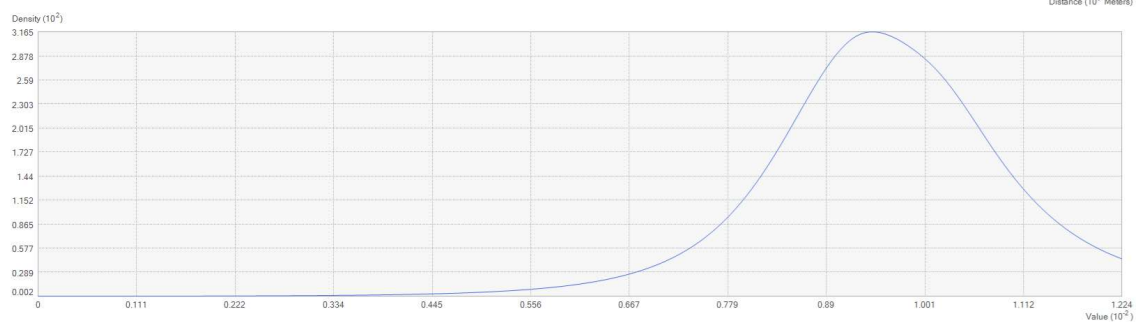
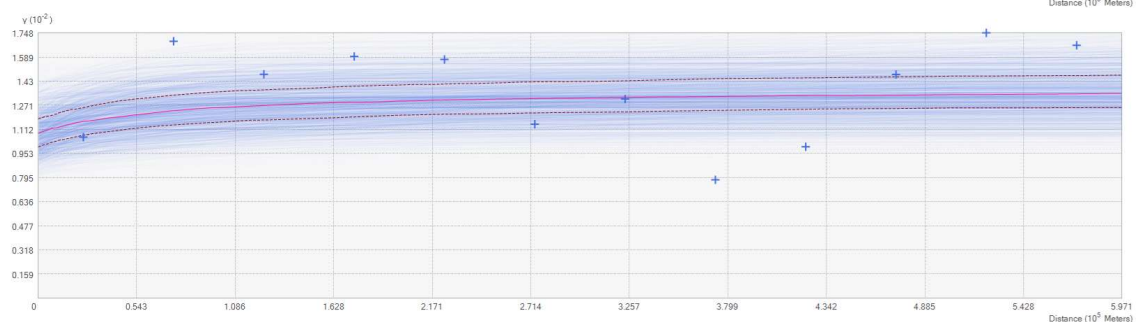
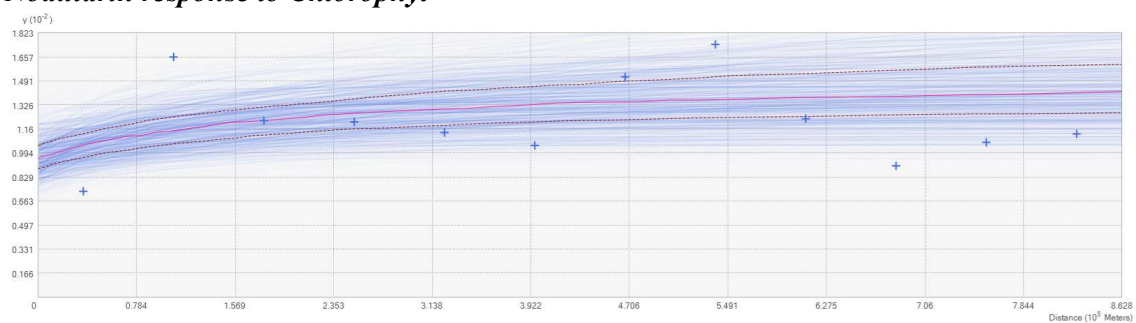


### *Nodularin response to $NO_3$*

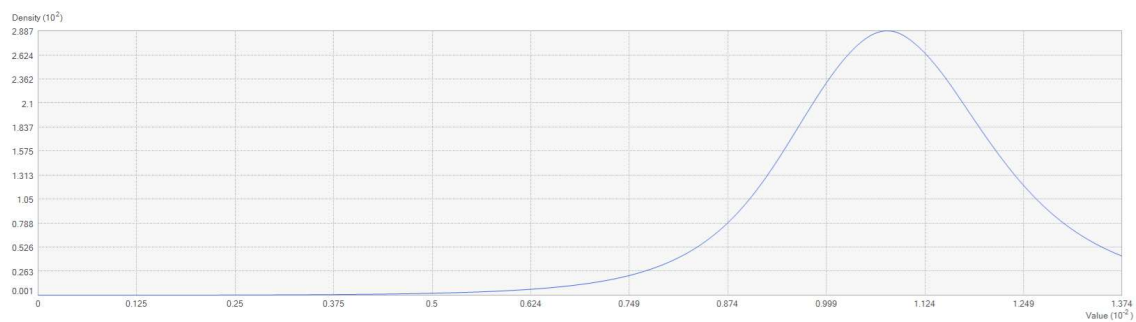




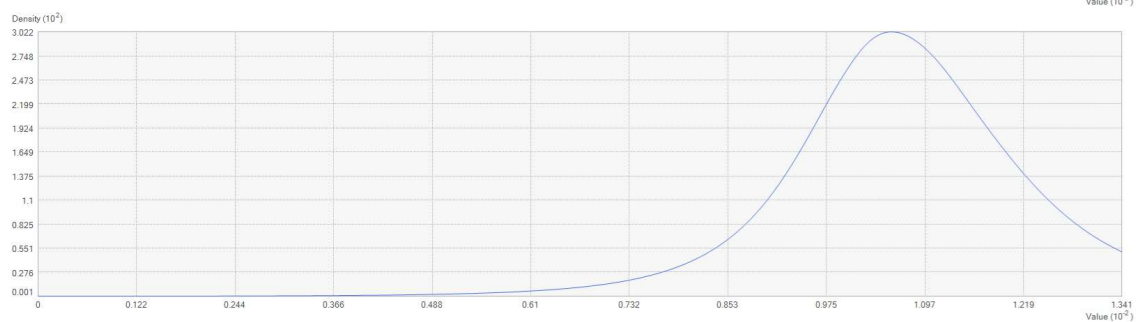
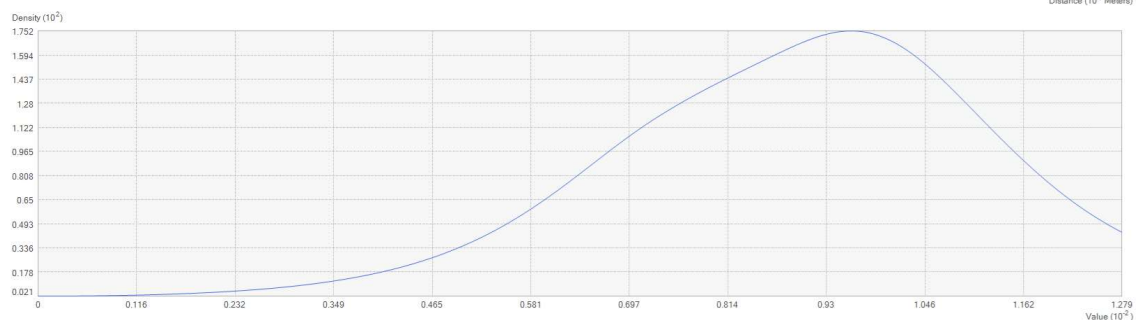
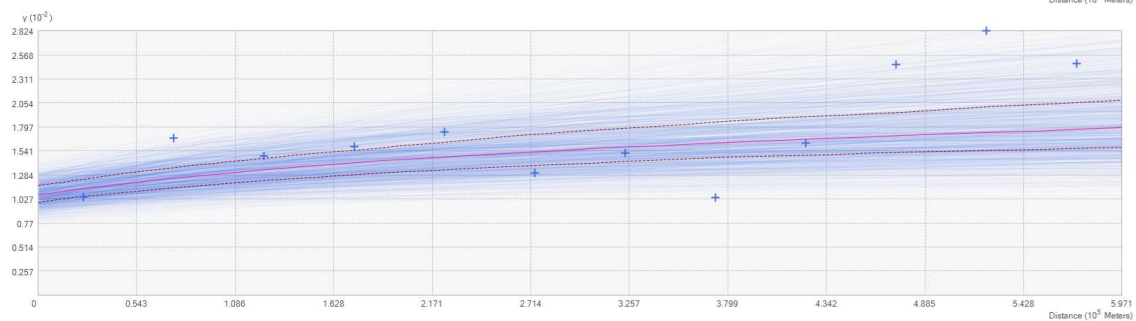
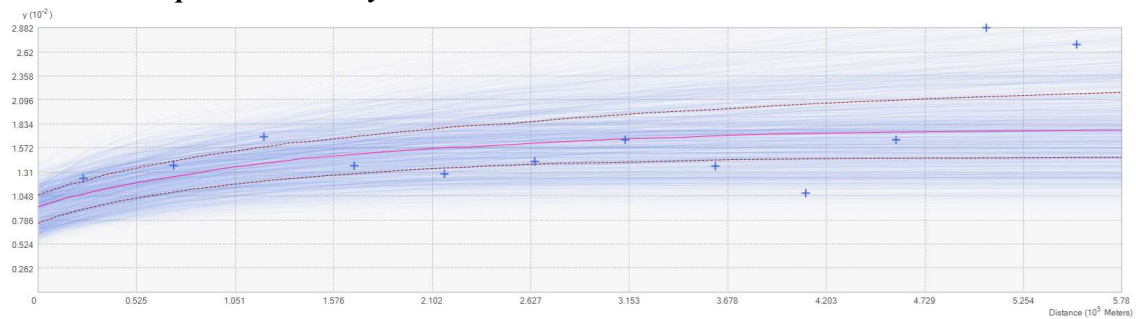
## *Nodularin response to Chlorophyll*



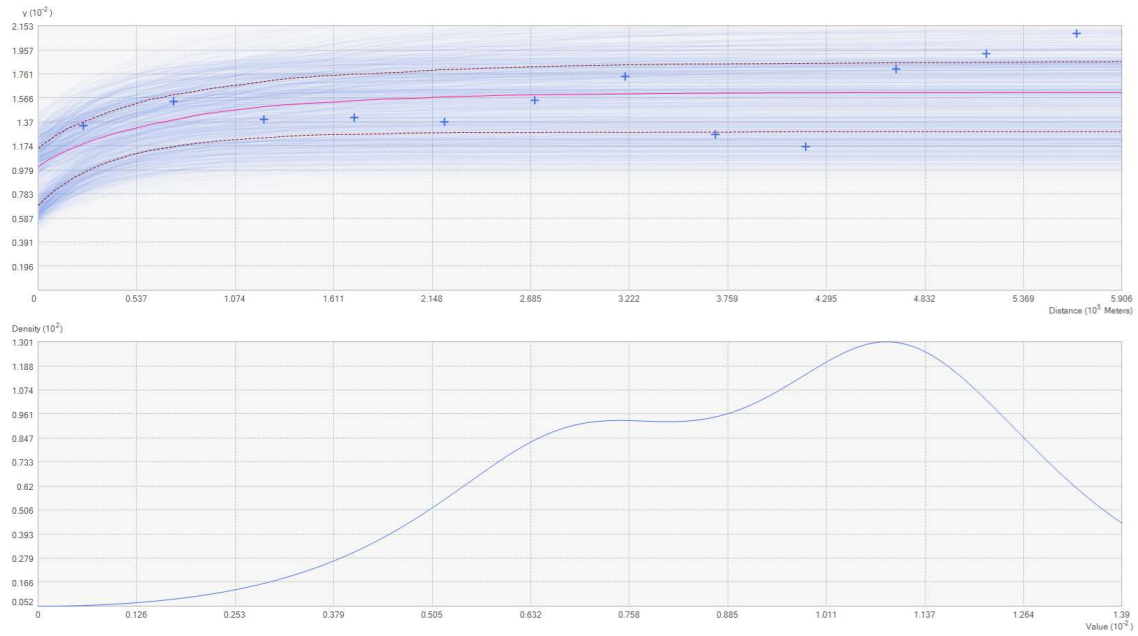




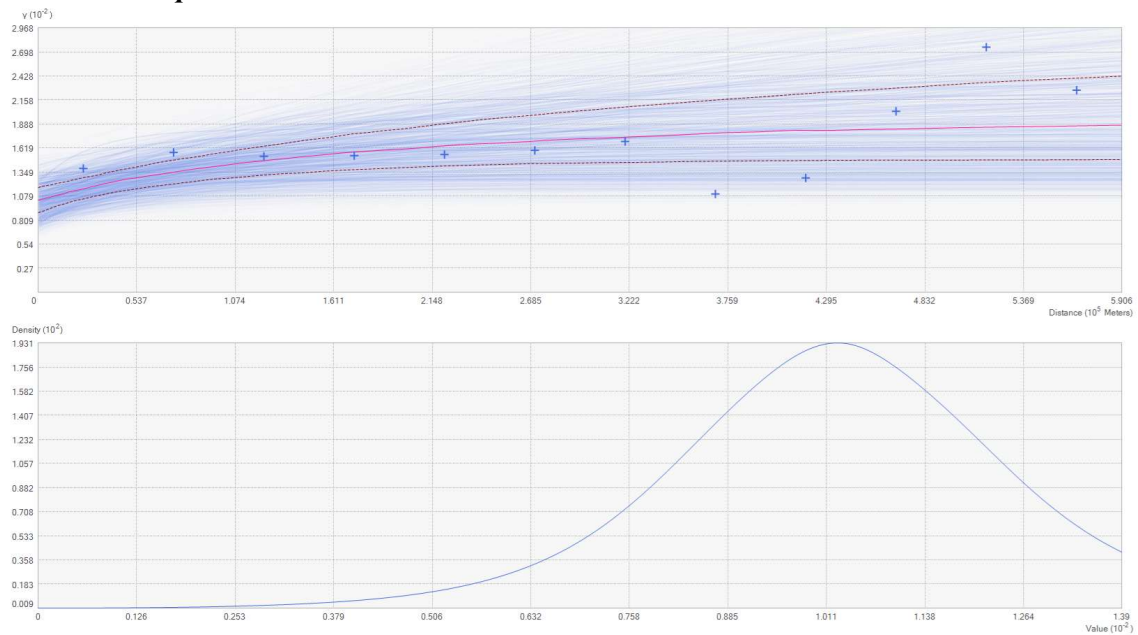
### ***Nodularin response to Salinity***



### ***Nodularin response to Temperature***

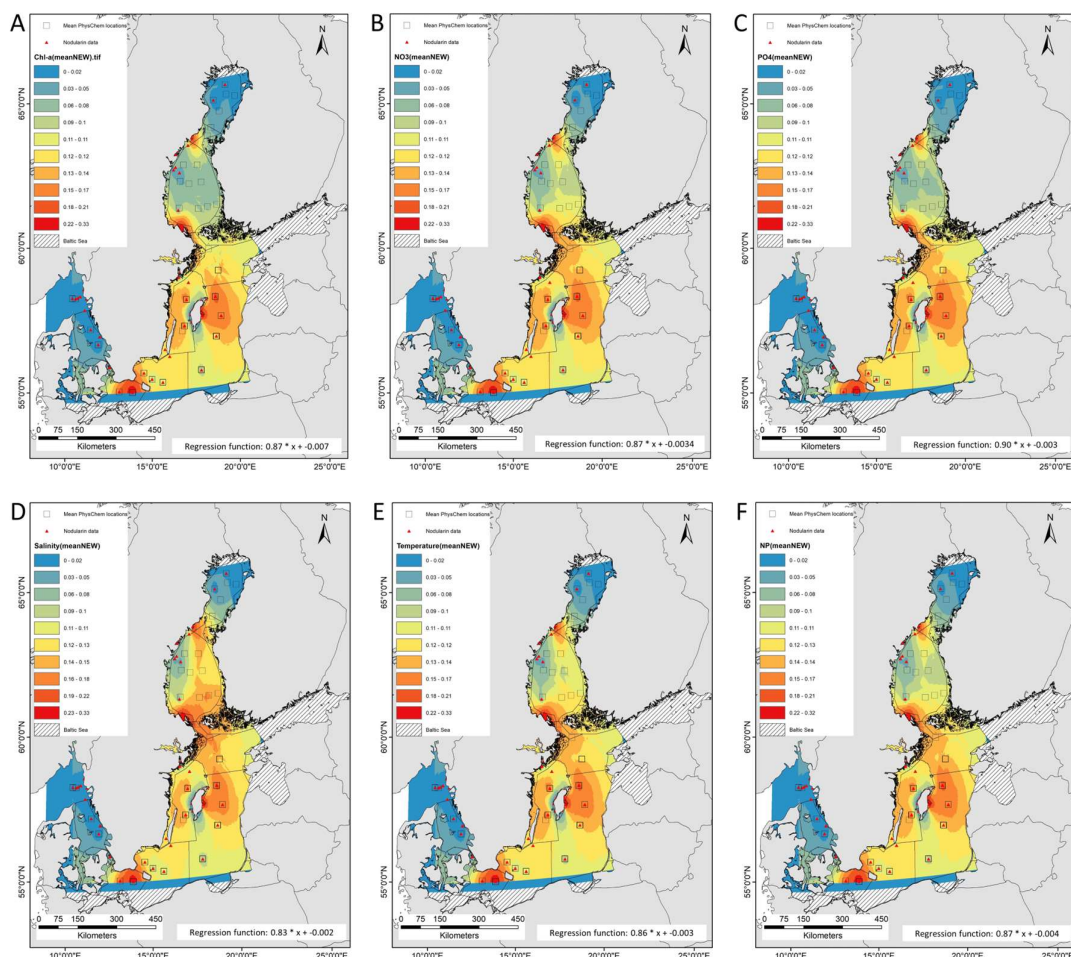


### ***Nodularin response to Distance to Shore***

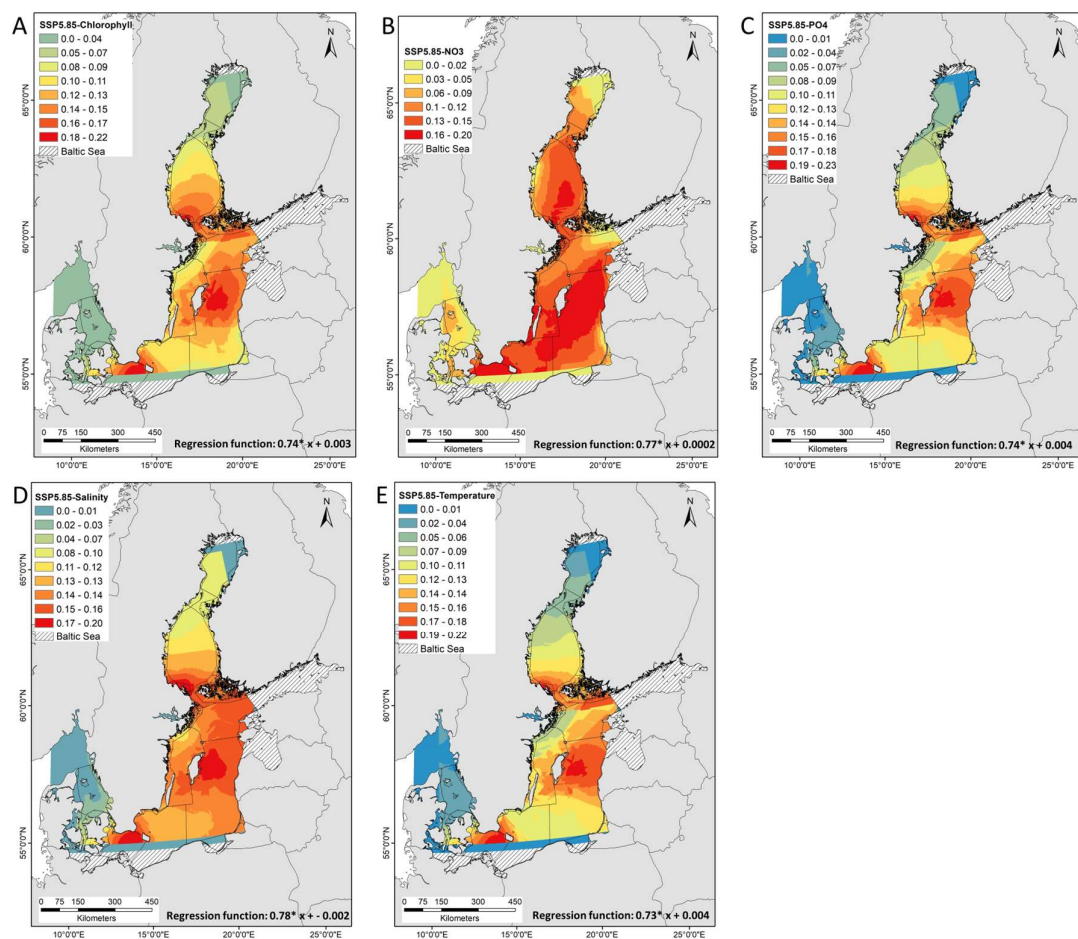


### ***Nodularin response to $\text{NO}_3\text{:PO}_4$***

**Figure S2.** Variogram (top two figures) and nugget (second two figures) analysis of the Kriging geostatistical interpolation illustrating the response of nodularin to current and future environmental predictors. The variogram and nugget figures correspond to the current (top) and future (below) environmental variables. The variogram measures spatial continuity and data variability, plotting variance against distance, defining the sill (total variance) and range (correlation distance). The nugget is the y-intercept, representing non-spatial variance caused by measurement errors or micro-scale variability, analyzed as a percentage of the sill.

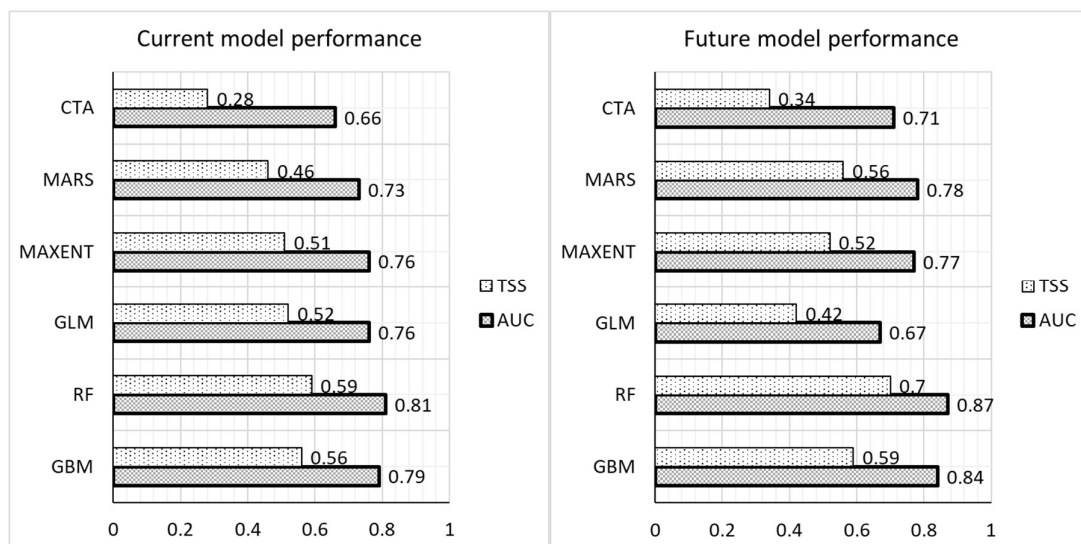


**Figure S3.** Co-kriging prediction of the model-predicted concentration and area distribution of nodularin ( $\mu\text{g l}^{-1}$ ) show area distribution and regression function of the concentration predicted at different sampling sites across the Baltic Sea. Predicted area distributions are based on mean concentration of (a) chlorophyll, (b)  $\text{NO}_3$ , (c)  $\text{PO}_4$ , (d) salinity, (e) temperature and (f)  $\text{NO}_3:\text{PO}_4$  ratio. Mean values were made in the National Swedish Marine Monitoring Program and retrieved from the Swedish National Oceanographic Data Centre at SMHI <https://shark.smhi.se>. Grading colors in key legend corresponds to predicted concentrations of nodularin ( $\mu\text{g l}^{-1}$ ). Geostatistical interpolations are performed using nodularin concentration as dependent variable and mean values as independents.

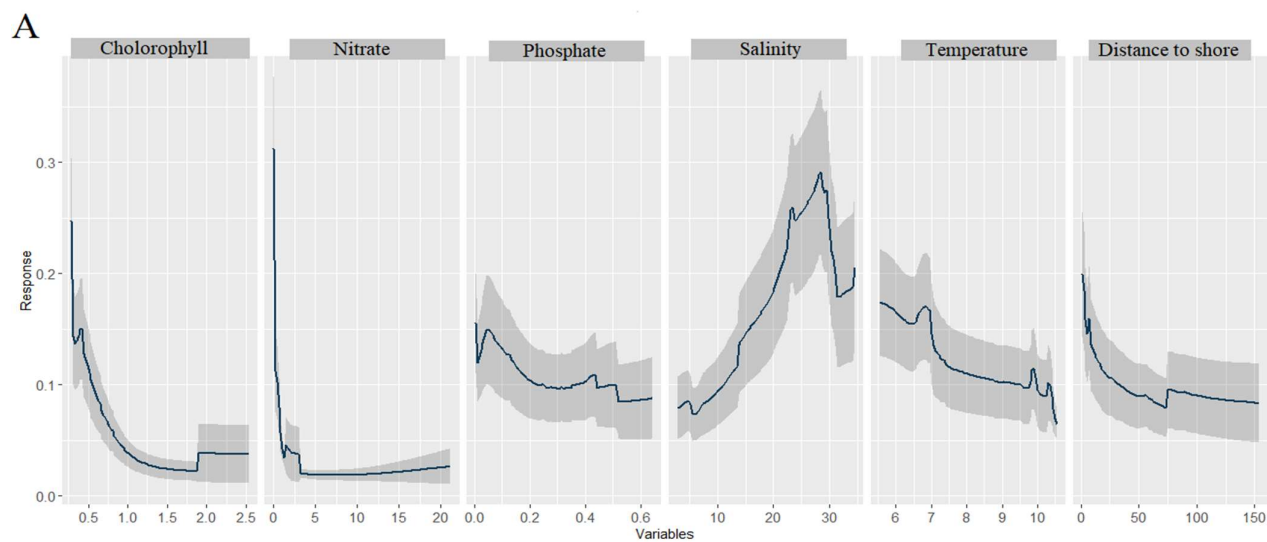


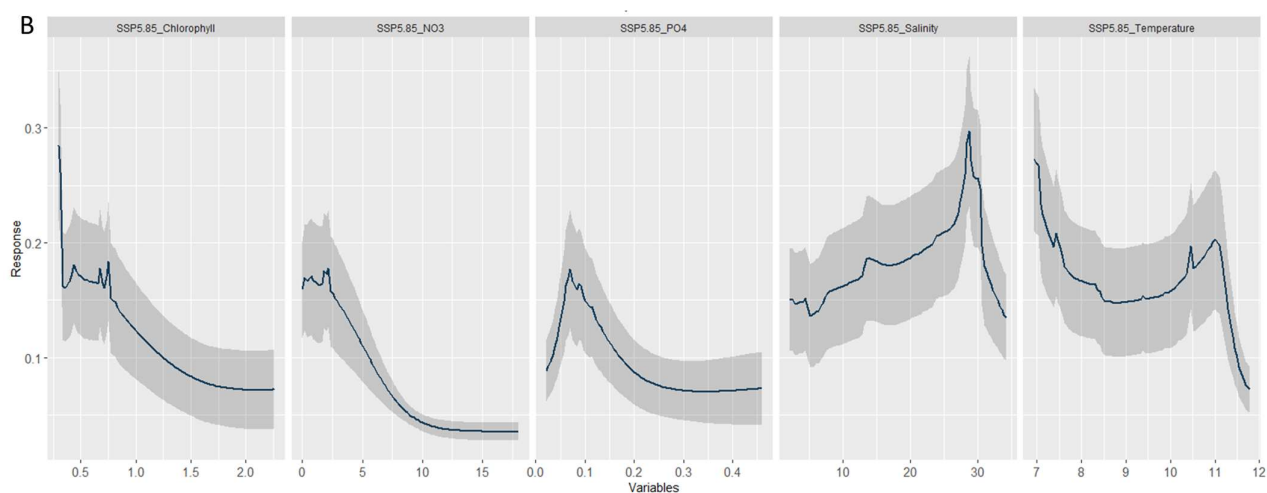
**Figure S4.** Empirical Bayesian kriging (EBK) regression prediction of the model-predicted concentration and area distribution of nodularin ( $\mu\text{g l}^{-1}$ ) (estimated using kernel density) show area distribution and regression function of the concentration predicted at different sampling sites across the Baltic Sea. Grading colors in key legend corresponds to predicted concentrations of nodularin ( $\mu\text{g l}^{-1}$ ). Predicted area distributions are based on projected concentration of (a) chlorophyll, (b)  $\text{NO}_3$ , (c)  $\text{PO}_4$ , (d) salinity and (e) temperature. Predicted area distributions and concentrations of nodularin are based on Shared Socioeconomic Pathway (SSP5.85) 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 variable and RCPs variables in raster GeoTIFF format as independents. Models are validated using 1000 simulations as cross-validation.



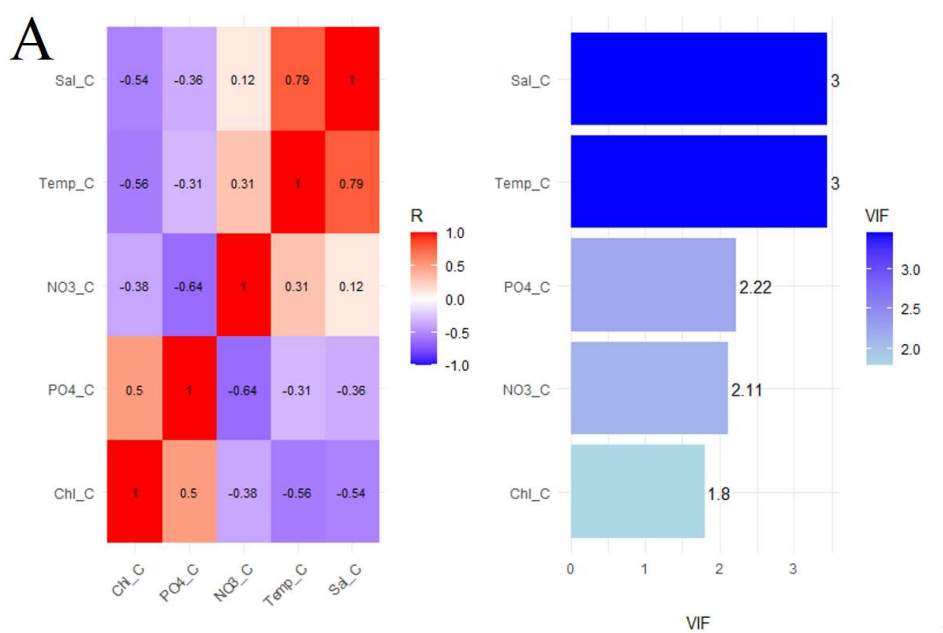


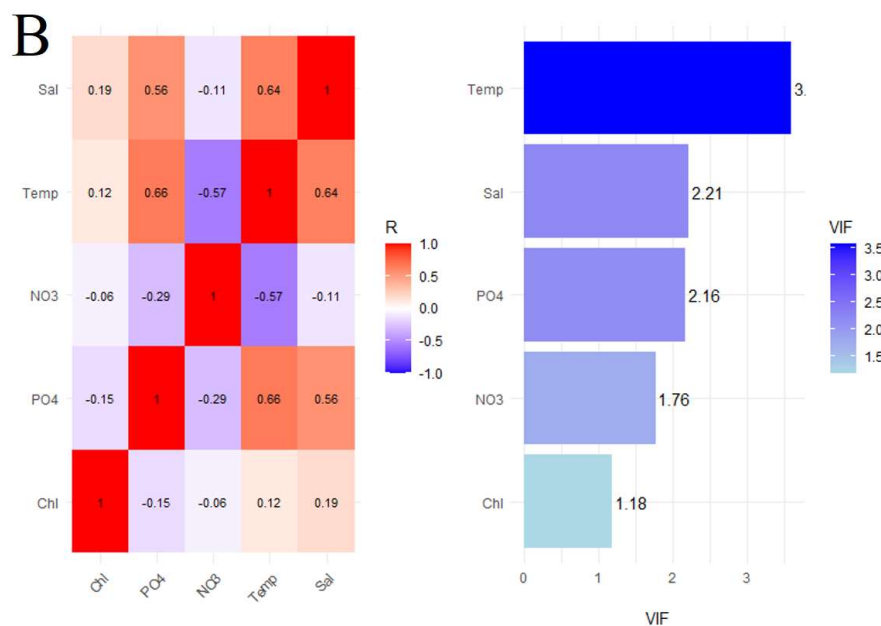
**Figure S5.** Model assessment shows the performance of the six machine-learning algorithms used in ensemble modeling. Algorithms are assessed using the area under curve AUC and the True skill statistics TSS.



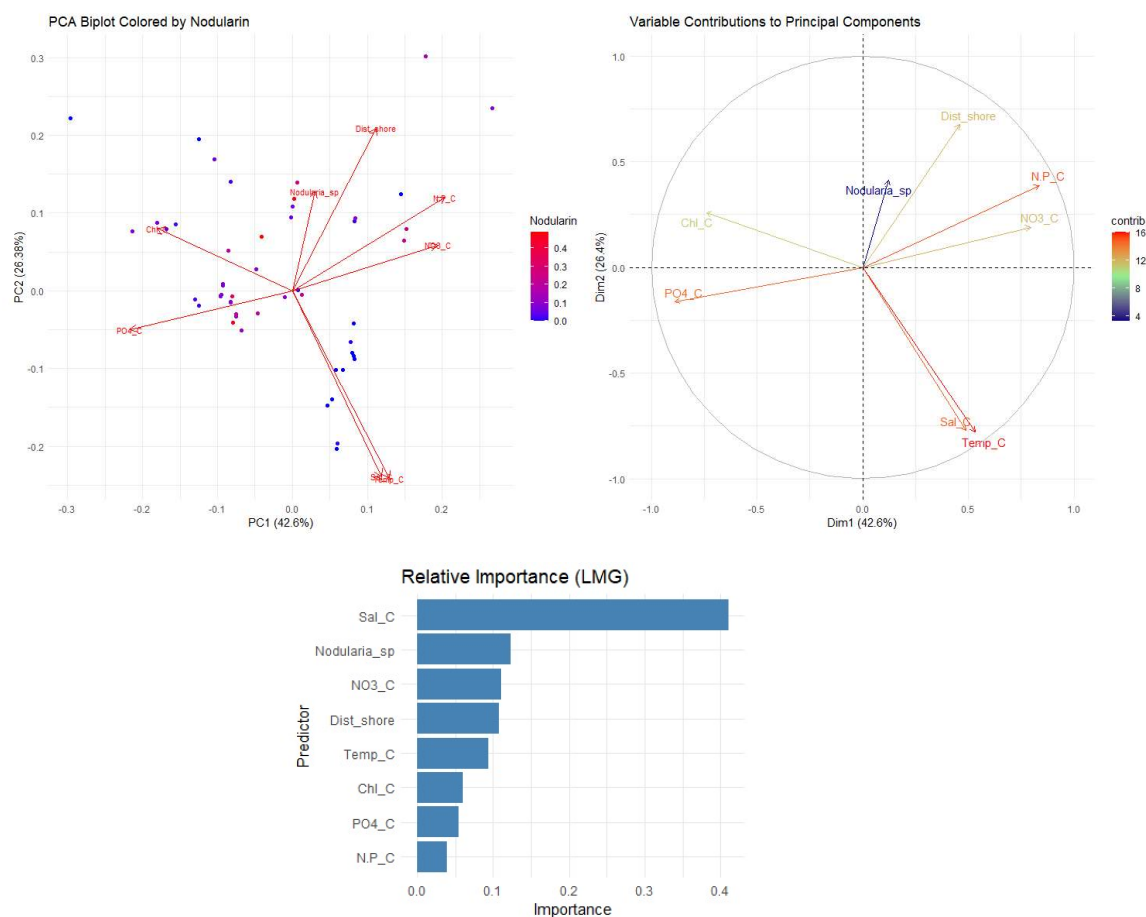


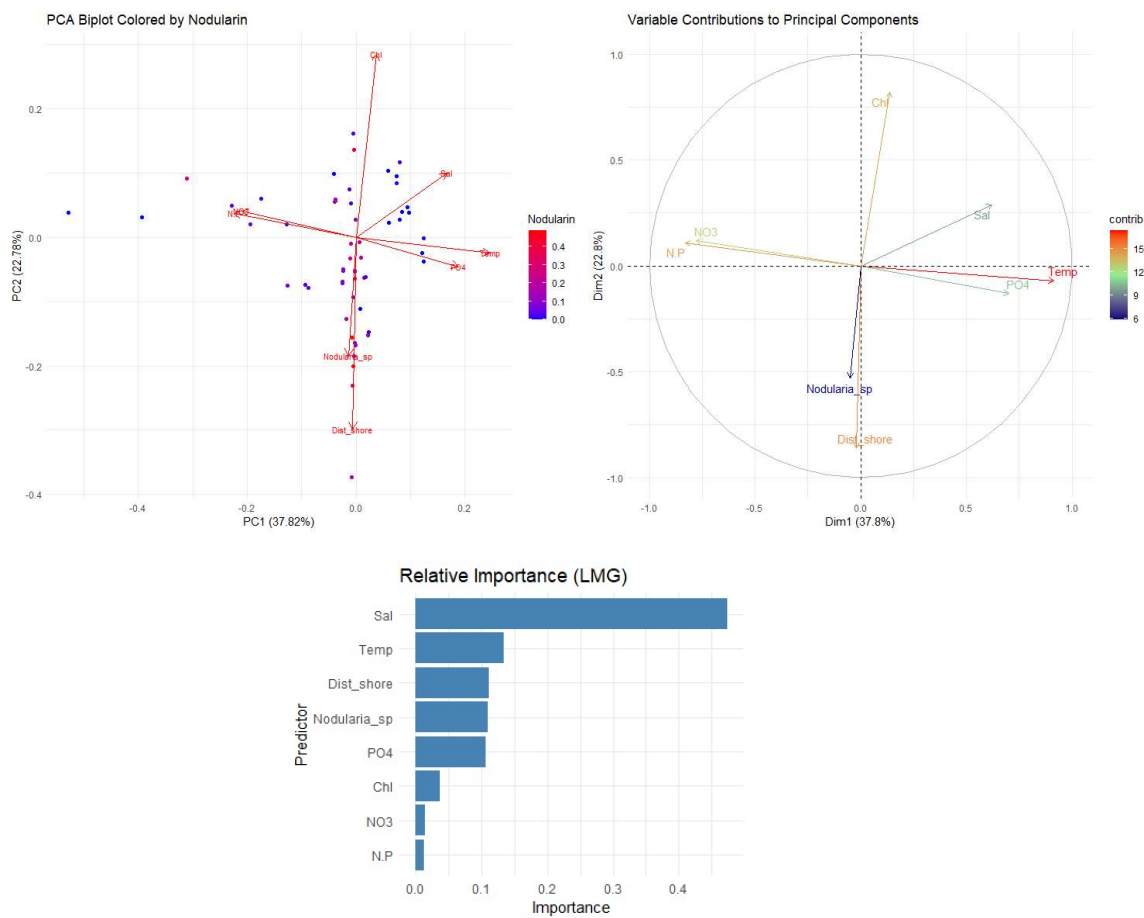
**Figure S6.** Response curves of ensemble learning modeling illustrate the effects of the predicted (a) current environmental variables and (b) future climate change scenarios SSP5.85 in the year 2100 on nodularin concentration ( $\mu\text{g l}^{-1}$ ). The response curve describes the performance of nodularin in response to a gradient of each predicted variable.





**Figure S7.** Test of multicollinearity using variance inflation factor VIF and correlation matrix between predictors downloaded from (A) Copernicus and (B) SMHI databases.





**Figure S8.** Contribution of variables of ERGOM from Copernicus (top) and NODC from SMHI (bottom) to nodularin concentration and relative variable importance (bar plots) as measured by the LMG method. Variables also include distance to shore and abundance of the *N. spumigena*.