Supplementary figures:

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.

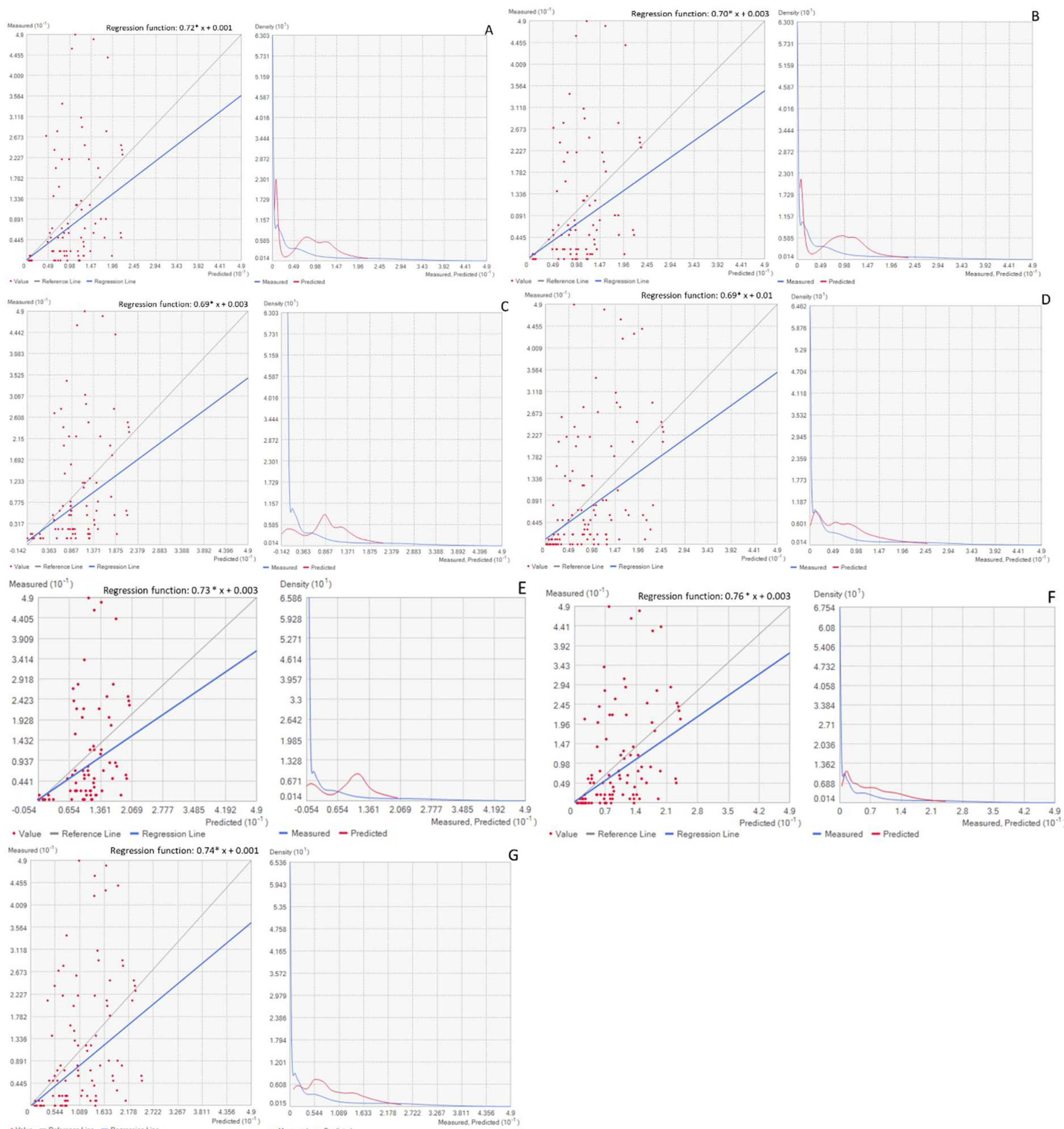


Figure S1. Distributions of the cross-validation statistics (estimated using kernel density) show the prediction regression scatterplot, the regression function, and the measured (blue line) and predicted (red line) values of nodularin (µg l^{-1}).Values correspond to nodularin response to concentrations of sea surface (a) chlorophyll mg m^{-3} (a) nitrate NO_3 µmol m^{-3}, (c) phosphate PO_4 µmol m^{-3}, (d) NO_3:PO_4 ratio, (e) salinity, (f) temperature °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.

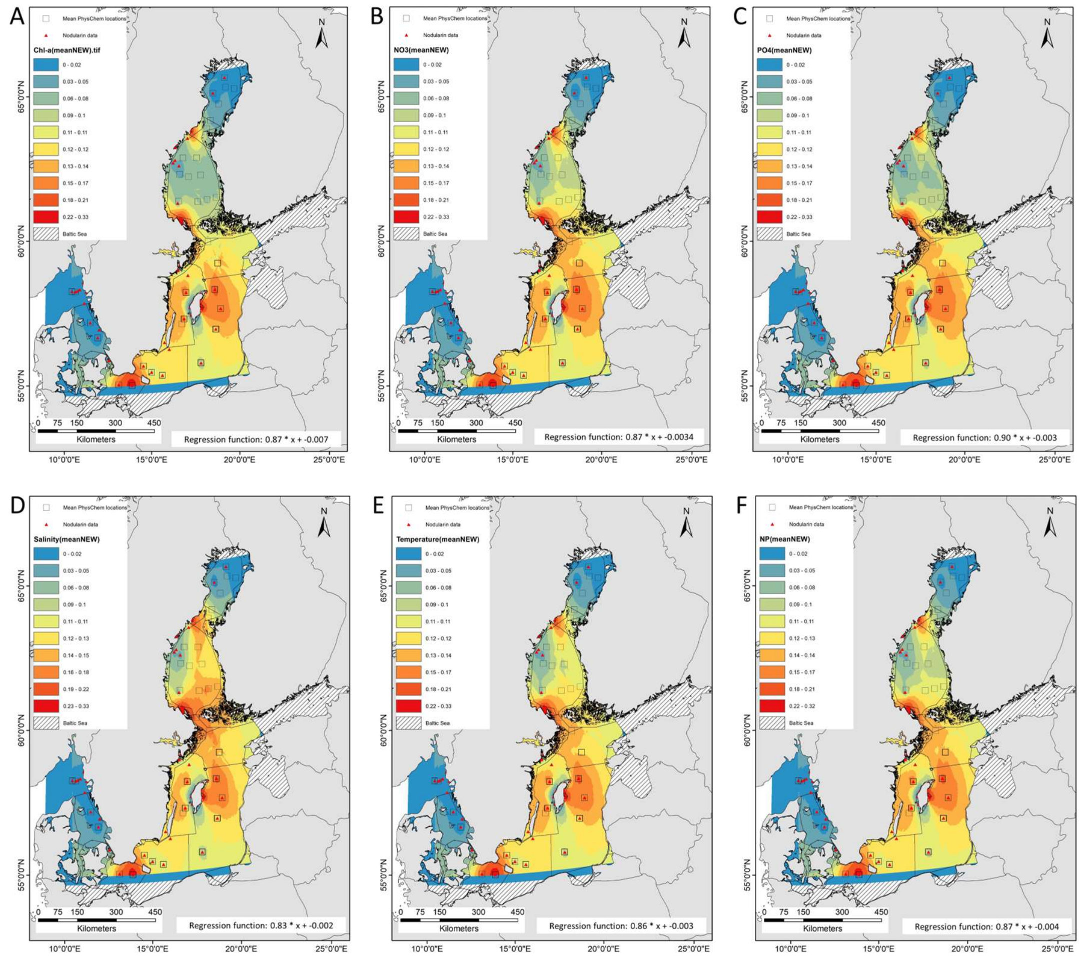


Figure S2. Co-kriging prediction of the model-predicted concentration and area distribution of nodularin (µg l^{-1}) show area distribution and regression function of the concentration predicted at different sampling sites across the Baltic Sea. Predicted area distributions are based on mean concentration of (a) chlorophyll, (b) NO_3, (c) PO_4, (d) salinity, (e) temperature and (f) NO_3: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 (µg l^{-1}). Geostatistical interpolations are performed using nodularin concentration as dependent variable and mean values as independents.

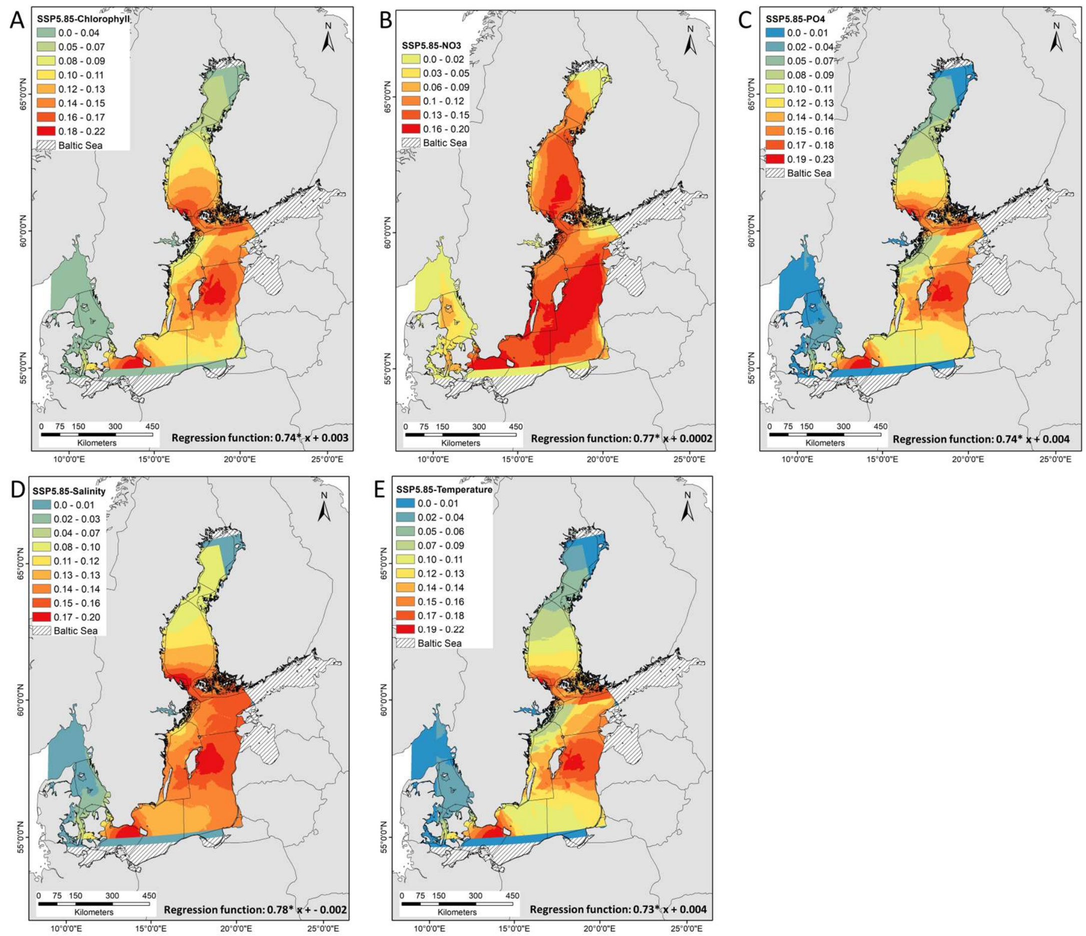


Figure S3. Empirical Bayesian kriging (EBK) regression prediction of the model-predicted concentration and area distribution of nodularin (µg l^{-1}) (estimated using kernel density) show area distribution and regression function of the concentration predicted at different sampling sites across the Baltic Sea. Grading colors in key legend corresponds to predicted concentrations of nodularin (µg l^{-1}). Predicted area distributions are based on projected concentration of (a) chlorophyll, (b) NO_3, (c) PO_4, (d) salinity and (e) temperature. Predicted area distributions 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 variable and RCPs variables in raster GeoTIFF format as independents. Models are validated using 1000 simulations as cross-validation.

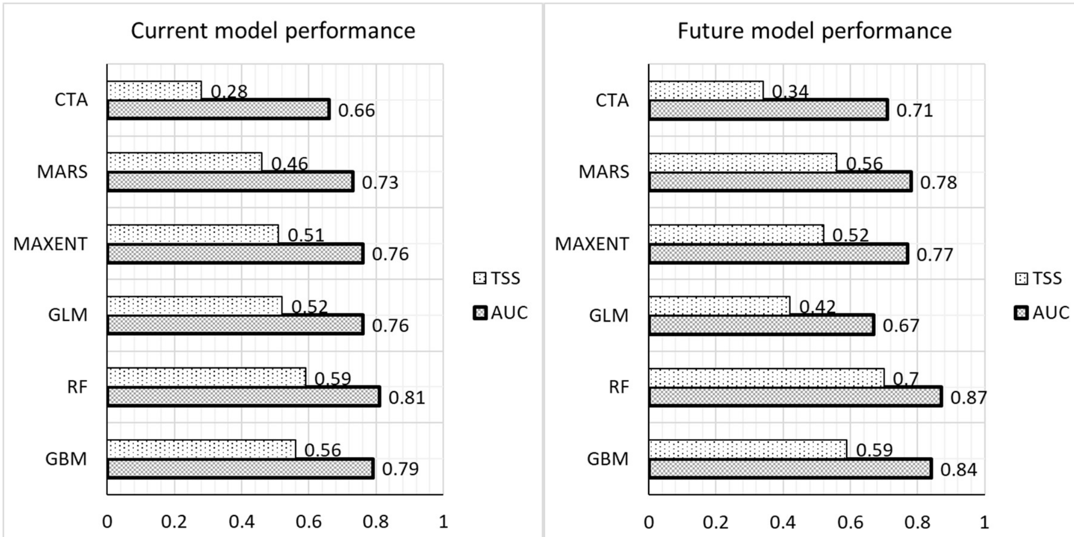
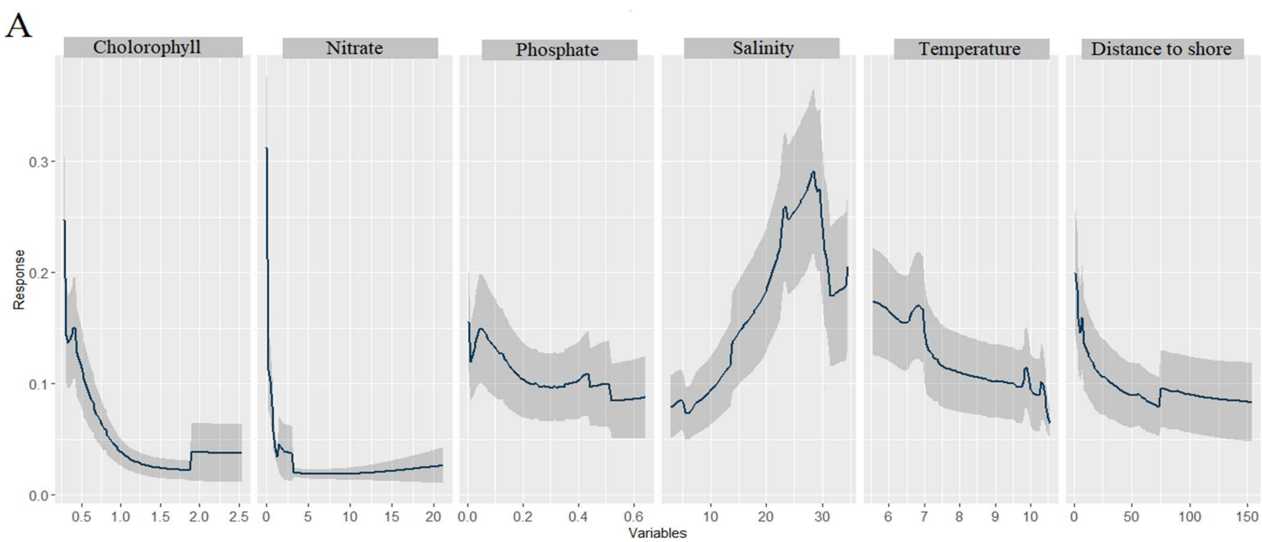


Figure S4. Model assessment shows the performance of the six machine-learning algorithms used in ensemble modeling. Algorithms are assessed using the area under curve AUC and the True skill statistics TSS.



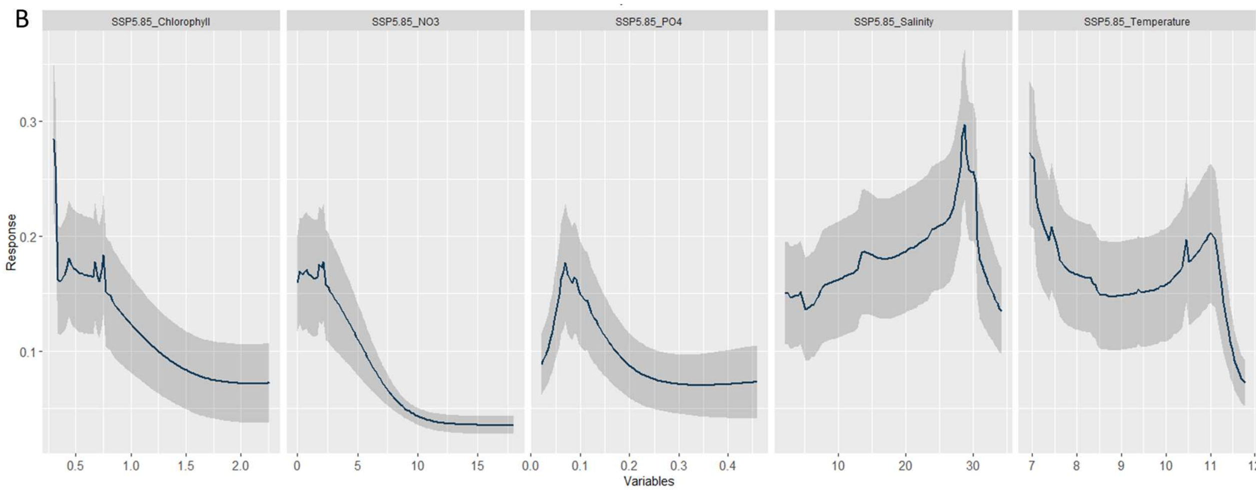


Figure S5. Response curves of ensemble learning modeling illustrate the effects of the predicted (a) current environmental variables and (b) future climate change scenarios SSP5.85 in the year 2100 on nodularin concentration (µg l^{-1}). The response curve describes the performance of nodularin in response to a gradient of each predicted variable.