Response to reviewers RC1 and RC2

Manuscript number: egusphere-2025-3290
Manuscript title: Prediction of present and future spatial occurrence of cyanobacteria and the toxin nodularin in the Baltic Sea

Authors: Mohanad Abdelgadir, Bengt Karlson, Elin Dahlgren, and Malin Olofsson

Principal Investigator: Mohanad Abdelgadir
Preprint link: https://doi.org/10.5194/egusphere-2025-3290

RC1

Review of the manuscript “Prediction of present and future spatial occurrence of cyanobacteria and the toxin nodularin in the Baltic Sea” by Mohanad Abdelgadir, Bengt Karlson, Elin Dahlgren, Malin Olofsson

Summary: The authors use Empirical Bayesian Kriging (EBK) regression prediction, ensemble learning, and stacked species distribution modelling (SSDM) to predict and interpret the current and future area distribution of Nodularia sp. and the toxin nodularin across the Baltic Sea. The underlying data base consists of 139 observed samples, combined with numerical model data from various sources. Predictions for the future distribution of Nodularia sp. blooms and nodularin are based on projected climate change scenarios in the year 2100.

Major comments:

The subject of the study is of importance and general interest. The overall approach is interesting and all the work that went into this study is greatly appreciated.

Re: We greatly appreciate the effort made by the reviewer (RC1) on our manuscript.

Unfortunately, however, I found it very difficult to rate methods and results of the presented study because I got lost in many details while I still lack a description of important key aspects. For my feeling the authors try to do too much in just one publication and I lack a clear aim of all their experiments.

Re: The introduction part is now revised, and the study aim is now clarified. We added several paragraphs early in the introduction part clearly to state how the different analyses and methods fit together to address the study aim.

These paragraphs can now be read as follows:

“There are several factors that may limit the proper spatial estimation of nodularin occurrence across the Baltic Sea. Such factors include sampling difficulty due to spatial variability of nodularin concentrations and the large spatial scale of the area. In addition, seasonal variability of nodularin concentrations poses a sampling challenge due to the short period of time when the toxin nodularin is produced, mainly in the late summer. Moreover, limited, infrequent, and sparse monitoring, especially in the open sea, can lead to underreporting or missing peak toxin periods. The fact that nodularin 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 nodularin in the Baltic Sea. Furthermore, future prediction of nodularin occurrence under climate change is a modeling challenge, especially in the long-term time frame when the worst-case climate scenario is projected to occur in the Baltic Sea. For instance, climate scenario SSP5-8.5 for 2100 is critical for the Baltic Sea with projected surface temperature increases up to 3.2°C (Meier et al., 2022). This scenario is combined with increased runoff, accelerated oxygen depletion, decreased salinity, and worsened algal blooms (Friedland et al., 2012; Meier, 2006). All these factors, together with lack of proper modeling approach, make the current and future prediction of nodularin a methodological challenge..” see **lines 60-73**

“In the case of small dataset and clustered sampling points, kriging algorithm allows quantifying the spatial autocorrelation by determining the range or the distance within which points are correlated, and the sampling error or fine-scale variation called "nugget". Moreover, kriging assigns weights to neighboring points based on their spatial correlation, meaning that kriging firsthand considers the autocorrelation of points, not just their physical distance (Antonakos and Lambrakis, 2021; Chilès and Delfiner, 1999). By utilizing autocorrelation, kriging removes spatially correlated noise and creates a fine and less smeared kriged map. Kriging also ensures that the final residuals are not autocorrelated by modeling the main predictors and creating the best linear unbiased predictions (Chilès and Delfiner, 1999; Fotheringham and Rogerson, 2009).”, see ***lines 80-87***

“Using Ensemble Learning (multiple models instead of single model) is known to quantify uncertainty in long-term projections (Álvarez Fanjul et al., 2025; Eyring et al., 2024a, b). In addition, using climate change scenarios coupled with real-time monitoring data from satellite imagery altogether allows for long-term trends to be estimated (Eyring et al., 2024a, b). It is also shown that combining the in-situ data with real-time monitoring data can further improves the accuracy of models (Wang, 2023; Yang et al., 2016). This altogether could possibility reveals uncertainty in long-term projections regardless of how long the period is (Eyring et al., 2024a, b; Parker, 2013). Previous studies demonstrated the difficulty and the uncertainty of resolving species composition changes under impacts of global change, particularly species known to be sensitive to change (Thuiller, 2004; Thuiller et al., 2009, 2019). However, it has been shown that using ensemble modeling combined with Stacked Species Distribution Models SSDM (Schmitt et al., 2017) and climate scenarios, allows for predicting shifts in habitat suitability, species richness, and assemblage structure (Abdelgadir et al., 2025; Brandt et al., 2017; Khan and Verma, 2022; Moullec et al., 2022). Therefore, using multi-model approach may provide a robust method to mitigate the uncertainties of species forecasting..”, see ***lines 100-111***

“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. Using integrated results of the above methods, the study 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. This setup may act as proof of concept and can be applied when data is limited, and large sampling campaigns are challenging. .” See ***lines 120-126***

I thus strongly recommend to rather focus on one specific aspect. Potential candidates I could imagine are: (1) a comparison of different techniques for enhancing the output of climate models to resolve local cyanobacteria blooms and toxins (here I don’t regard the Baltic Sea as a perfect candidate because many important processes are not resolved in climate models). (2) a comparison of different methods to interpolate nodularin measurements in space (if the available data do not suffice considerations could be added on how many extra data would be required and where) or (3) suggestions how to implement nodularin into a prognostic Baltic Sea model which already contains cyanobacteria. This list is certainly not comprehensive.

Re: We do understand that this model setup can seem complex and unnecessarily complicated.

In this revised version, in the introduction part, we detailed how each method works with the aim of clarifying this setup and its benefits. Please refer to our pervious response (above). First we mention what factors make estimation and prediction of nodularin is a challenging task see **lines 60-73**. Secondly we describe which method and/approach can handle such factors see ***lines 80-87*** and ***lines 100-111***. We included short technical descriptions of the working procedure of each method and how each is processing the small dataset and how long-term future prediction can be achieved.

In this revision we also added a paragraph early in the materials and methods summarizing and simplifying the data processing and model setup.

This paragraph can be read as follows:

"2.1 Overview of data processing and model setup

"We compiled nodularin and Nodularia spumigena data collected across the Baltic Sea region and paired these observations with a suite of environmental variables derived from monitoring programs and high-resolution Copernicus model products. We harmonized all environmental and climate scenario datasets into a common spatial framework, extracted conditions for each sampling point, and prepared raster layers for geostatistical and machine learning analyses. We then applied empirical Bayesian kriging, co-kriging, and a multi-algorithm ensemble modeling workflow to predict present and future nodularin occurrence and N. spumigena distribution. Finally, we evaluated model performance and quantified environmental drivers using cross-validation, Bayesian regression, ANOVA/GLM, PCA, and LGM to assess how physical and biogeochemical conditions shape toxin patterns across the Baltic Sea. Geostatistical interpolation and ensemble learning modeling flow are illustrated in Fig 1." see ***lines 128-137***

As the study is designed now, I have several major concerns which in my eyes need to be addressed:

(1) There seem to be very few measurements of the toxin nodularin available (as I understood 139 observed sample were investigated while most of them are clustered at some coasts). Relating these few measurements to a multitude of environmental factors and their interactions will most likely lead to overfitting. While the authors state that they used part of the data for testing, I did not find clear evidence which could rule out this concern. I would expect something like a direct comparison plot of the best prediction versus observed nodularin for some independent test data.

Re: That's true. We have had 139 observations across the study area collected over one year during the period between June to September 2023. We also considered the small dataset, unsampled locations, sparse sampling locations and possible autocorrelation and clustering of points. For this purpose, we implemented a conceptual modeling setup utilizing two methods to resolve all these spatial issues related to the dataset. In this revised version, we detail these factors that make estimation and prediction of nodularin a challenging task see **lines 60-73**. Secondly we describe which method and/approach can handle such factors see ***lines 80-87*** and ***lines 100-111***. We included short technical descriptions of the working procedure of each method and how each is processing the small dataset and how long-term future prediction can be achieved.

Regarding the working procedure to overcome all these factors, we first described how ensemble learning procedure resolves the issue of unsampled locations by generating pseudo-absence data points and splitting the dataset into the training testing sets.

The first procedure for handling unsampled locations is now described in revised version and can be read as follows:

"Algorithms in ensemble learning can resolve the issues concerning autocorrelation, clustered samples, small datasets, and unsampled sites across the study area. Algorithms overcome the overfitting by generating what are called "pseudo-absence data points" (Hysen et al., 2022), which represent those unsampled sites across the study area and in a way compare the observed environment (represented by the presence) against what is available. ." see ***lines 112-115***

The second procedure for handling overfitting is now described in revised version and can be read as follows:

"Algorithms also avoid overfitting by dividing the data into training and testing sets, which allows the model to learn on one subset of the data (training set) and evaluate its performance on another subset (testing set). All these procedures together ensure that the model generalizes well to new data, making it more robust and reliable." see ***lines 115-118***

Here, the dataset is split into two segments: 70% training and 30% testing data, which allows the model to learn on one subset of the data (training set) and evaluate its performance on another subset (testing set). This ensures that the model generalizes well to new data, making it more robust and reliable. For more information, see below how the model was generated in R code:

```
#Current prediction
#Herem I add the pseudo-absence data to the background. I add 10000 pseudo-points as background using "bg" argument
dc <- sdmData(species~., train=newssp, predictors= env.c, bg=list(n=10000))
#let's fit the model by adding different algorithm methods, replication techniques and K-fold cross-validation
mc <- sdm(species ~ . , dc, methods=c('GBM','RF','GLM','MAXENT','MARS','CART'), replication=c("boot"),k=10,test.percent=30)

#Future prediction
#Here, I add the pseudo-absence data to the background. I add 10000 pseudo-points as background using "bg" argument
df <- sdmData(species~., train=newssp, predictors= env.f, bg=list(n=10000))
#let's fit the model by adding different algorithm methods, replication techniques and K-fold cross-validation
mf <- sdm(species ~ . , df, methods=c('GBM','RF','GLM','MAXENT','MARS','CART'), replication=c("boot"),k=10,test.percent=30)
```

Abbreviations: dc/df= data object (current c/future f), bg=background data object consists of 10000 randomly pseudoabsence points, newssp= data object contains occurrences, species= column of the data object contain the species (Nodularia sp.), methods= set of algorithms used, replication= method technique which is "bootstrapping", test.percent= percentage of data set used for testing which is 30%, k= 10-fold bootstrapping, mc/mf= model setting (current c/future f).

The third procedure for handling long-term future prediction is now described in revised version and can be read as follows:

""Using Ensemble Learning (multiple models instead of single model) is known to quantify uncertainty in long-term projections (Álvarez Fanjul et al., 2025; Eyring et al., 2024a, b). In addition, using climate change scenarios coupled with real-time monitoring data from satellite

imagery altogether allows for long-term trends to be estimated (Eyring et al., 2024a, b). It is also shown that combining the in-situ data with real-time monitoring data can further improves the accuracy of models (Wang, 2023; Yang et al., 2016). This altogether could possibility reveals uncertainty in long-term projections regardless of how long the period is (Eyring et al., 2024a, b; Parker, 2013). ***lines 100-105***

(2) For the occurrence of Nodularia Sp. biogeochemical model data of the ERGOM model are combined with the available observations. This approach might lead to inconsistencies and since this is the key aspect I would like to see at least some quality assessment of both, the numerically simulated and predicted Nodularia sp. blooms. Since Satellite data for Nodularia sp. are provided e.g. by SMHI, it would be fairly easy to show 2-dimensional plots of the extent of a particularly large and small bloom during the recent years as simulated (by the numerical model) and as predicted (by the methods of this study) in direct comparison to satellite observations. Additionally, the in-situ samples for Nodularia sp. could be plotted against the respective model data to ensure that these data sources can be combined without problems.

Re: We are fully aware of the available satellite data on the bloom and from SMHI, yet our study focus was mainly concerned about what drives toxin concentration (nodularin) at specific sampling sites and not detecting the bloom per se. The toxicity of blooms are not always related to bloom biomass and cannot therefore be directly inferred from satellites.

Following reviewer suggestion, we added a new 2-dimentional PCA plot and LMG analysis describing the independent effect of each model data (from Copernicus and SMHI) on nodularin concentration across all sampling sites. See supplementary **Figure S6 and Table S7.** The method is added and described in the statistical analysis **(lines 245-248)** and results **(lines 418-423)** and discussed in the discussion part **(lines 495-501)**.

Technically, our approach integrates Ensemble Learning (multiple models instead of a single model) that is known to quantify uncertainty in long-term projections. In addition, using climate change scenarios coupled with real-time monitoring data from satellite imagery altogether allows for long-term trends to be estimated. By combining the in-situ data, the nodularin concentration along with previous, further improves the accuracy of models. This altogether allows for accurate predictions for the next 75 years (2100 climate scenario) and reveals uncertainty in long-term projections regardless of how long the period is. We also considered having the same high-resolution scenarios for the current selected variables from all sources (near surface chlorophyll, nitrate, phosphate, salinity, and temperature) to ensure the data integrity and that the model is internally consistent. Along with integrity and accuracy, our study design ensures consistency that the transition from current data to forecasted data (future predictions) is seamless and logical, preventing abrupt, unrealistic shifts in the model's trajectory. This can clearly be observed in model accuracy and variable contribution, where our modeling approach captured the spatial changes in nodularin occurrence over the next 75 years with excellent performance, see ***lines 376-379 and Fig 7.***

(3) The authors then combine many different data sources from global to regional models which are almost for certain inconsistent and might make it very difficult to draw robust conclusions. E.g. global models do not resolve coastal upwelling and it is very unlikely that these models capture the complex salinity dynamics, nutrient inputs or sediment processes of the Baltic Sea. These aspects need at least attention.

Re: We fully agree about the challenge of combining data from different sources. However, our conceptual model setup aims to address this challenge. We decided on having the same high-resolution scenarios for the currently selected variables from all sources (near-surface chlorophyll, nitrate, phosphate, salinity, and temperature) to ensure the data integrity and that the model is internally consistent. Along with integrity and accuracy, our study design ensures consistency that the transition from current data to forecasted data (future predictions) is seamless and logical, preventing abrupt, unrealistic shifts in the model's trajectory. This can clearly be observed in model accuracy and variable contribution, where our modeling approach captured the spatial changes in nodularin occurrence over the next 75 years with excellent performance.

We added additional statements in the introduction part highlighting how our approach can resolve the inconsistency and increase the prediction performance, particularly over the next 75 years. These statements can be read as follows:

"Using Ensemble Learning (multiple models instead of single model) is known to quantify uncertainty in long-term projections (Álvarez Fanjul et al., 2025; Eyring et al., 2024a, b). In addition, using climate change scenarios coupled with real-time monitoring data from satellite imagery altogether allows for long-term trends to be estimated (Eyring et al., 2024a, b). It is also shown that combining the in-situ data with real-time monitoring data can further improves the accuracy of models (Wang, 2023; Yang et al., 2016). This altogether could possibility reveals uncertainty in long-term projections regardless of how long the period is (Eyring et al., 2024a, b; Parker, 2013). Previous studies demonstrated the difficulty and the uncertainty of resolving species composition changes under impacts of global change, particularly species known to be sensitive to change (Thuiller, 2004; Thuiller et al., 2009, 2019). However, it has been shown that using ensemble modeling combined with Stacked Species Distribution Models SSDM (Schmitt et al., 2017) and climate scenarios, allows for predicting shifts in habitat suitability, species richness, and assemblage structure (Abdelgadir et al., 2025; Brandt et al., 2017; Khan and Verma, 2022; Moullec et al., 2022). Therefore, using multi-model approach may provide a robust method to mitigate the uncertainties of species forecasting..", see ***lines 100-111***

We also clarified that the rational of choosing the climate scenario SSP5-85 for year 2100 and why this scenario is critical for the Baltic Sea. The statement can be read as follows:

"Furthermore, future prediction of nodularin occurrence under climate change is a modeling challenge, especially in the long-term time frame when the worst-case climate scenario is projected to occur in the Baltic Sea. For instance, climate scenario SSP5-8.5 for 2100 is critical for the Baltic Sea with projected surface temperature increases up to 3.2°C (Meier et al., 2022). This scenario is combined with increased runoff, accelerated oxygen depletion, decreased salinity, and worsened algal blooms (Friedland et al., 2012; Meier, 2006)."See ***lines 68-73***

(3) I did not find an independence test for the predictors. Then, I am surprised that distance to the shore has been used as predictor. I am well aware of a prominent study which uses this factor when investigating the onset of blooms – still blooms may then drift to the shore and frequently do so.

Re: 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. When combined with temperature, chlorophyl, nutrients, and salinity, having 'distance to shore' as a predictor tends to reflect the spatial, physical, and ecological conditions that promote nodularin expansion. This finding is clearly observed when ‘distance to shore’ has shown an interacting effect with other variables in promoting nodularin occurrence (See **Table 1A, Figure 3E, & Figure 4G**).

Independence test for each predictor was performed and provided using the distributions of the cross-validation statistics of kriging (estimated using kernel density). Those figures show the prediction regression scatterplot, the regression function, and the measured (blue line) and predicted (red line) values of nodularin (μg l-1). These figures (see **Figures S1, S2, and S3**) also contain the equation of the regression function. For ensemble learning we provided both current and future model performances for each algorithm quantified by area under curve (AUC) and the True Skill Statistic (TSS); see Figure S4. All predictions, either by kriging or ensemble, showed high prediction capability indicted by high values in regression function, AUC, TSS, Kappa, …etc. and all statistical metrics in the study.

In supplementary figures (see **Figure S5**) we also provided the response curve that demonstrates the independence effect of each predictor on nodularin concentration.

Following reviewer suggestion, we added a new 2-dimentional PCA plot and LMG analysis describing the independent effect of each model data (from Copernicus and SMHI) on nodularin concentration across all sampling sites. See supplementary **Figure S6 and Table S7.** The method is added and described in the statistical analysis **(lines 245-248)** and results **(lines 418-423)** and discussed in the discussion part **(lines 495-501)**.

(5) I did not find convincing evidence for reliable predictions for any of the many methods. I would like to see clear visual comparisons to independent test data.

Re: In supplementary figures, we provided the distributions of the cross-validation statistics of kriging (estimated using kernel density) showing the prediction regression scatterplot, the regression function, and the measured (blue line) and predicted (red line) values of nodularin (μg l^{-1}). These figures (see **Figures S1, S2, and S3**) also contain the equation of the regression function. For ensemble learning we provided both current and future model performances for each algorithm quantified by area under curve (AUC) and the True skill statistics (TSS), see **Figure S4**. All predictions, either by kriging or ensemble, showed high prediction capability indicted by high values in regression function, AUC, TSS, Kappa, …etc. and all statistical metrics in the study.

Following reviewer suggestion, we added a new 2-dimentional PCA plot and LMG analysis describing the independent effect of each model data (from Copernicus and SMHI) on nodularin concentration across all sampling sites. See supplementary **Figure S6 and Table S7.** The method is added and described in the statistical analysis **(lines 245-248)** and results **(lines 418-423)** and discussed in the discussion part **(lines 495-501)**.

(6) Even if the authors revise and illustrate robust relations to predictors for Nodularia Sp. bloom occurrence and nodularin under present climate conditions, it is still not at all guaranteed that these could be extrapolated to a much warmer climate (e.g. species composition and competitive advantages might well change). I do not at all recommend to base predictions or even recommendations for politics on such uncertain ground.

Re: We value your recommendation, yet, if carefully applied, our conceptual modeling setup can be tested for future prediction and potential implication for management/guidance.

Moreover, this setup can also be applied for testing larger datasets. This conceptual modeling approach could also help prioritize surveillance and implement earlier sampling efforts in areas predicted to have high cyanotoxin concentration. We have already highlighted this in the conclusion section; and can be read as: "*If used carefully, these approaches could help prioritize surveillance and implement earlier sampling efforts in areas predicted to have high and/or increasing cyanotoxin concentration.*" .see **lines 534-535.**

We are fully aware that species composition and toxin occurrence might well change over the long term under future scenarios and increased warming. In a way to resolve prediction problems associated with the long-term projection, our approach integrates the ensemble learning (multiple models instead of a single model) that is known to quantify uncertainty in long-term projections. In addition, using climate change scenarios coupled with real-time monitoring data from satellite imagery altogether allows for long-term trends to be estimated. By combining the in-situ data, there is nodularin concentration, further improving the accuracy of models. This altogether allows for accurate predictions for the next 75 years (2100 climate scenario) and reveals uncertainty in long-term projections regardless of how long the period is. We also have decided to have the same high-resolution scenarios for the currently selected variables from all sources (near-surface chlorophyll, nitrate, phosphate, salinity, and temperature) to ensure the data integrity and that the model is internally consistent. Along with integrity and accuracy, our study design ensures consistency so that the transition from current data to forecasted data (future predictions) is seamless and logical, preventing abrupt, unrealistic shifts in the model's trajectory.

In this revised version we clarified in a statement how the approach can resolve species composition and toxin occurrence changes over time and space. The statement can be read as follows:

"Previous studies demonstrated the difficulty and the uncertainty of resolving species composition changes under impacts of global change, particularly species known to be sensitive to change (Thuiller, 2004; Thuiller et al., 2009, 2019). However, it has been shown that using ensemble modeling combined with Stacked Species Distribution Models SSDM (Schmitt et al., 2017) and climate scenarios, allows for predicting shifts in habitat suitability, species richness, and assemblage structure (Abdelgadir et al., 2025; Brandt et al., 2017; Khan and Verma, 2022; Moullec et al., 2022). Therefore, using multi-model approach may provide a robust method to mitigate the uncertainties of species forecasting.." See ***lines 105-111***

Specific comments:

Ln 13: change "blooms often include toxin producing species" to " blooms can contain toxin producing stains,"

Re: Corrected to *" blooms can contain toxin producing strains,"* see ***lines 13-14***

Ln 14: change "climate change is expected to increase the occurrence of cyanobacterial blooms" to something like that "climate change may increase the occurrence of cyanobacterial blooms"

Re: Corrected to *"climate change may increase the occurrence of cyanobacterial blooms"* ***line 15***

Ln 17-18: The choice of each method should be motivated.

Re: The choice of each method, and why both methods were integrated are now motivated and were described in detail in the paragraphs below:

lines 60-73. Secondly we describe which method and/approach can handle such factors see ***lines 80-87*** and ***lines 100-111***

See **lines 60-73, *lines 80-87*, lines 120-131, *lines 100-111***and summarized in method section, see **lines 128-137**

Ln 29: I do not recommend any risk assessment or management decisions at current state.

Re: Risk assessment and management is now removed from the abstract and changed to "Our developed conceptual modeling approach is useful where toxicological data such as cyanotoxins are insufficient.", see **lines 29-30**

Ln 64: Empirical Bayesian Kriging (EBK) regression prediction depends heavily on the density and distribution of the underlying samples. In Figure 1 it appears that most of the few samples are clustered at some coasts. Also, I doubt that all samples were taken at the same time ad it is not clear which state of the system the Kriging refers to. It would be good if the time aspect was clarified and I would like to see a comparison to independent test data.

Re: That is true; kriging depends on the distribution of sampling points. However, kriging perfectly handles spatially and clustered points. In detail, applying semi-variogram geostatistical tool in Kriging algorithm allows quantifying the spatial autocorrelation by determining the range or the distance within which points are correlated, and the sampling error or fine-scale variation called "nugget". Moreover, kriging assigns weights to neighboring points based on their spatial correlation, meaning that kriging firsthand considers the autocorrelation of points, not just their physical distance. By utilizing autocorrelation, kriging removes spatially correlated noise and creates a fine and less smeared kriged map. Kriging also ensures that the final residuals are not autocorrelated by modeling the main predictors and creating the best linear unbiased predictions. Taking together, we believe that kriging is able not just to resolve the spatial autocorrelation in the dataset but also to create fine unbiased predictions. It is worth mentioning that the sampling period spanned from June to September 2023, and the predictors, chlorophyll, NO_3 , PO_4 ,salinity, and temperature, were downloaded from databases for the exact same time point as toxins were sampled for.

We added an additional paragraph like the above one clarifying how kriging algorithm resolves and handles spatially autocorrelated and clustered points. The paragraph can be read as follows:

"In the case of small dataset and clustered sampling points, kriging algorithm allows quantifying the spatial autocorrelation by determining the range or the distance within which points are correlated, and the sampling error or fine-scale variation called "nugget". Moreover, kriging assigns weights to neighboring points based on their spatial correlation, meaning that kriging firsthand considers the autocorrelation of points, not just their physical distance (Antonakos

and Lambrakis, 2021; Chilès and Delfiner, 1999). By utilizing autocorrelation, kriging removes spatially correlated noise and creates a fine and less smeared kriged map. Kriging also ensures that the final residuals are not autocorrelated by modeling the main predictors and creating the best linear unbiased predictions (Chilès and Delfiner, 1999; Fotheringham and Rogerson, 2009).", see ***lines 80-87***

Following reviewer suggestion, we added a new 2-dimentional PCA plot and LMG analysis describing the independent effect of each model data (from Copernicus and SMHI) on nodularin concentration across all sampling sites. See supplementary **Figure S6 and Table S7.** The method is added and described in the statistical analysis **(lines 245-248)** and results **(lines 418-423)** and discussed in the discussion part **(lines 495-501)**.

For the intended test of the data and predictor, please refer to my response to points #3 and #5 in your list of questions.

Ln 84: It does not become clear how these approaches could overcome the above-mentioned problems.

Re: Please read our responses to your previous questions on how Kriging and machine learning handling complex, and spatially autocorrelated dataset.

Statements can be read as follows:

"In the case of small dataset and clustered sampling points, kriging algorithm allows quantifying the spatial autocorrelation by determining the range or the distance within which points are correlated, and the sampling error or fine-scale variation called "nugget". Moreover, kriging assigns weights to neighboring points based on their spatial correlation, meaning that kriging firsthand considers the autocorrelation of points, not just their physical distance (Antonakos and Lambrakis, 2021; Chilès and Delfiner, 1999). By utilizing autocorrelation, kriging removes spatially correlated noise and creates a fine and less smeared kriged map. Kriging also ensures that the final residuals are not autocorrelated by modeling the main predictors and creating the best linear unbiased predictions (Chilès and Delfiner, 1999; Fotheringham and Rogerson, 2009).", see ***lines 80-87***

"Using Ensemble Learning (multiple models instead of single model) is known to quantify uncertainty in long-term projections (Álvarez Fanjul et al., 2025; Eyring et al., 2024a, b). In addition, using climate change scenarios coupled with real-time monitoring data from satellite imagery altogether allows for long-term trends to be estimated (Eyring et al., 2024a, b). It is also shown that combining the in-situ data with real-time monitoring data can further improves the accuracy of models (Wang, 2023; Yang et al., 2016). This altogether could possibility reveals uncertainty in long-term projections regardless of how long the period is (Eyring et al., 2024a, b; Parker, 2013). Previous studies demonstrated the difficulty and the uncertainty of resolving species composition changes under impacts of global change, particularly species known to be sensitive to change (Thuiller, 2004; Thuiller et al., 2009, 2019). However, it has been shown that using ensemble modeling combined with Stacked Species Distribution Models SSDM (Schmitt et al., 2017) and climate scenarios, allows for predicting shifts in habitat suitability, species richness, and assemblage structure (Abdelgadir et al., 2025; Brandt et al., 2017; Khan and

Verma, 2022; Moullec et al., 2022). Therefore, using multi-model approach may provide a robust method to mitigate the uncertainties of species forecasting..”, see ***lines 100-111***

Ln 87: Again, I do not at all recommend to base future predictions on such uncertain ground.

Re: See our responses to your question #6 *“(6) Even if the authors revise and illustrate robust relations to predictors for Nodularia Sp. bloom occurrence and nodularin under present climate conditions, it is still not at all guaranteed that these could be extrapolated to a much warmer climate (e.g. species composition and competitive advantages might well change). I do not at all recommend to base predictions or even recommendations for politics on such uncertain ground.”*

In this revised version we clarified in a statement how the approach can resolve species composition and toxin occurrence changes over time and space. The statement can be read as follows:

“Previous studies demonstrated the difficulty and the uncertainty of resolving species composition changes under impacts of global change, particularly species known to be sensitive to change (Thuiller, 2004; Thuiller et al., 2009, 2019). However, it has been shown that using ensemble modeling combined with Stacked Species Distribution Models SSDM (Schmitt et al., 2017) and climate scenarios, allows for predicting shifts in habitat suitability, species richness, and assemblage structure (Abdelgadir et al., 2025; Brandt et al., 2017; Khan and Verma, 2022; Moullec et al., 2022). Therefore, using multi-model approach may provide a robust method to mitigate the uncertainties of species forecasting..” See ***lines 105-111-132***

Ln 91: How many of the samples did contain nodularin? Where all samples taken during Nodularia bloom conditions? Where other variables, such as nutrients, salinity or temperature, measured as well?

Re: Yes, the data were collected during Nodularia bloom conditions (June to September). The predictors (nutrients, salinity or temperature) were downloaded from databases at same sampling sites at the same days of sampling as well. Nutrients, salinity or temperature were downloaded from Copernicus database from the same period of the sampling (June to September). This has been clarified in **lines 160-161** and **lines 175-176** respectively.

Ln 105ff: Blending the nutrients from the SMHI forecast with chlorophyll_a simulated by the ERGOM-model needs some good motivation because the respective fields will not be consistent.

Re: Yes, we fully agree, yet our study tends to resolve possible inconsistency by integrating two different approaches with different procedures. For how each method handles the data, please refer to my previous responses to your questions #3, #6, and the question related to Ln84. This will also be clarified in the revised version of the manuscript.

We fully agree about the challenge of combining data from different sources. However, our study addresses this challenge. We decided on having the same high-resolution scenarios for the currently selected variables from all sources (near-surface chlorophyll, nitrate, phosphate, salinity, and temperature) to ensure the data integrity and that the model is internally consistent. Along with integrity and accuracy, our study design ensures consistency that the transition from current data to forecasted data (future predictions) is seamless and logical, preventing abrupt, unrealistic shifts in the model's trajectory. This can clearly be observed in model accuracy and variable contribution, where our modeling approach captured the spatial changes in nodularin occurrence over the next 75 years with excellent performance.

We added additional statements in the introduction part highlighting how our approach can resolve the inconsistency and increase the prediction performance, particularly over the next 75 years. These statements can be read as follows:

"Using Ensemble Learning (multiple models instead of single model) is known to quantify uncertainty in long-term projections (Álvarez Fanjul et al., 2025; Eyring et al., 2024a, b). In addition, using climate change scenarios coupled with real-time monitoring data from satellite imagery altogether allows for long-term trends to be estimated (Eyring et al., 2024a, b). It is also shown that combining the in-situ data with real-time monitoring data can further improves the accuracy of models (Wang, 2023; Yang et al., 2016). This altogether could possibility reveals uncertainty in long-term projections regardless of how long the period is (Eyring et al., 2024a, b; Parker, 2013). Previous studies demonstrated the difficulty and the uncertainty of resolving species composition changes under impacts of global change, particularly species known to be sensitive to change (Thuiller, 2004; Thuiller et al., 2009, 2019). However, it has been shown that using ensemble modeling combined with Stacked Species Distribution Models SSDM (Schmitt et al., 2017) and climate scenarios, allows for predicting shifts in habitat suitability, species richness, and assemblage structure (Abdelgadir et al., 2025; Brandt et al., 2017; Khan and Verma, 2022; Moullec et al., 2022). Therefore, using multi-model approach may provide a robust method to mitigate the uncertainties of species forecasting..", see ***lines 100-111***

We also clarified that the rational of choosing the climate scenario SSP5-85 for year 2100 and why this scenario is critical for the Baltic Sea. The statement can be read as follows:

"Furthermore, future prediction of nodularin occurrence under climate change is a modeling challenge, especially in the long-term time frame when the worst-case climate scenario is projected to occur in the Baltic Sea. For instance, climate scenario SSP5-8.5 for 2100 is critical for the Baltic Sea with projected surface temperature increases up to 3.2°C (Meier et al., 2022). This scenario is combined with increased runoff, accelerated oxygen depletion, decreased salinity, and worsened algal blooms (Friedland et al., 2012; Meier, 2006)."See ***lines 68-72***

All new references (citation) were added and updated into the References section.

Ln 118: I could not find a meaningful comparison in the supplement.

Re: The supplement figures, **Figures S2 and S3**, illustrated that geographical estimations and maps produced by the kriging and ensemble-based models in this study are aligned despite the differences in approaches. The supplement was provided for the reader to compare, e.g., **Figure S2 with Figure 4 (line 322)** **and Figure 7** (**line 389**)in the main text. In other words, this explains how nodularin spatially expands in response to same set of predictors but from two different databases.

The comparison is now added in the results section and can be read as follows:

"3.3. Relative contribution of predictors to nodularin concentration
Variable contribution to nodularin concentration as measured by LMG method shows that abundance of N. spumigena had largest positive effect (with ERGOM: p = 0.0003, LMG%=20.0;

with NODC: $p < 0.05$, LMG%=28.5) followed by salinity (with ERGOM: $p < 0.001$, LMG%=15.2; with NODC: $p = 0.01$, LMG%=24.0), see Fig S6 and Table S7 for the rest effects of predictors. The coefficient of determination R^2 of both models, as calculated using LMG method, indicates the total proportion of variance in nodularin explained by all eight predictors jointly (ERGOM: $R^2 = 0.78$, NODC: $R^2 = 0.72$).” See **lines 418-23**

Furthermore, the comparison is now discussed in the discussion part as well, see ***lines 495-506***

Ln 120: It does not become clear how the predictions were “tested” based on the mentioned Copernicus data.

Re: Please see supplementary figures and tables. Multiple figures were presented to show how nodularin and Nodularia sp. respond to different predictors from Copernicus (Figure 3, 4 and 5) from ERGOM (in **Figure S2, S3, S4**). These figures were supported by statistical matrices and regression function equations. Following reviewer suggestion, we added a new 2-dimentional PCA plot and LMG analysis describing the independent effect of each model data (from Copernicus and SMHI) on nodularin concentration across all sampling sites. See supplementary **Figure S6 and Table S7.** The method is added and described in the statistical analysis **(lines 245-248)** and results **(lines 418-423)** and discussed in the discussion part **(lines 495-501)**.

Ln 123: I would be very interested to see 2-dimensional plots on predicted Nodularia Sp. and nodularin when using the climate models under present climate conditions in comparison the predictions based on the combination of numerical Baltic Sea models.

Re: Please refer to previous response to your question Ln 120. Check supplementary figures and tables. Multiple figures were presented to show how nodularin and Nodularia sp. respond to different predictors from Copernicus (Figure 3, 4 and 5) from ERGOM (in **Figure S2, S3, S4**). These figures were supported by statistical matrices and regression function equations.

Following reviewer suggestion, we added a new 2-dimentional plot PCA analysis describing the independent effect of each model data (from ERGOM and SMHI) on nodularin concentration across all sampling sites. See **Figure S6 and also Table S7**

Ln 185 Something went wrong here.

Re: Latin text (from journal template) is now removed in the revised manuscript, **line 252**. The sentence in the revision will start by “*Prediction of nodularin spatial* occurrence…”.

Ln 187ff: From here on I am somehow lost in many details, different methods (all with different design choices) and various feature selections while I am lacking a clear purpose of all the experiments.

Re: I assume you refer to analysis of Bayesian linear regression. To interpret the results and statistical terms of the Bayesian linear regression, please refer to Table 1. In brief, 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, and P(M|data) refers to the posterior probability of each model after seeing the data, while BFM compares each model to the average P(M|data) of the other models.

Citation: https://doi.org/10.5194/egusphere-2025-3290-RC1

##

RC2

The research topic is very interesting, and the model formulation is novel and conceptually strong.

Re: We greatly appreciate the effort made by the reviewer (RC2) on our manuscript.

The things I find problematic are the limited data points the model is based on. There are 139 points from 54 locations in 1 year. I think the model will benefit from atleast a multi-year data set. Is the data from previous years not available? Why was it limited to one year. I have serious doubts about conclusions drawn from a year's worth of data to interpolate results for 75 years from now.

Re: That's true. We have 139 observations across the study area, and no data is available from the previous year, unfortunately. The conceptual model setup was, however, intended to handle such limited dataset with consideration to large spatial extent of the study area and the moderate clustering of the points. Also, this is a proof-of-concept study where, when further data is available, it can be applied and strengthened further.

In this revised version, we clarified this approach and benefit of the model setup and that we ensure that it works even on small datasets. We added different statements clarifying how different methods resolve the issue of limited sampling size (i.e., number of observations) and that the overall approach managed to handle autocorrelation and moderate clustering of points. Those added paragraphs in the introduction part clearly state how the different analyses and methods fit together to address the study aim. These paragraphs can now be read as follows:

"There are several factors that may limit the proper spatial estimation of nodularin occurrence across the Baltic Sea. Such factors include sampling difficulty due to spatial variability of nodularin concentrations and the large spatial scale of the area. In addition, seasonal variability of nodularin concentrations poses a sampling challenge due to the short period of time when the toxin nodularin is produced, mainly in the late summer. Moreover, limited, infrequent, and sparse monitoring, especially in the open sea, can lead to underreporting or missing peak toxin periods. The fact that nodularin 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 nodularin in the Baltic Sea. Furthermore, future prediction of nodularin occurrence under climate change is a modeling challenge, especially in the long-term time frame when the worst-case climate scenario is projected to occur in the Baltic Sea. For instance, climate scenario SSP5-8.5 for 2100 is critical for the Baltic Sea with projected surface temperature increases up to 3.2°C (Meier et al., 2022). This scenario is combined with increased runoff, accelerated oxygen depletion, decreased salinity, and worsened algal blooms (Friedland et al., 2012; Meier, 2006). All these factors, together with lack of proper modeling approach, make the current and future prediction of nodularin a methodological challenge.." see **lines 60-73**

"In the case of small dataset and clustered sampling points, kriging algorithm allows quantifying the spatial autocorrelation by determining the range or the distance within which points are correlated, and the sampling error or fine-scale variation called "nugget". Moreover, kriging assigns weights to neighboring points based on their spatial correlation, meaning that kriging firsthand considers the autocorrelation of points, not just their physical distance (Antonakos

and Lambrakis, 2021; Chilès and Delfiner, 1999). By utilizing autocorrelation, kriging removes spatially correlated noise and creates a fine and less smeared kriged map. Kriging also ensures that the final residuals are not autocorrelated by modeling the main predictors and creating the best linear unbiased predictions (Chilès and Delfiner, 1999; Fotheringham and Rogerson, 2009).", see ***lines 80-87***

"Using Ensemble Learning (multiple models instead of single model) is known to quantify uncertainty in long-term projections (Álvarez Fanjul et al., 2025; Eyring et al., 2024a, b). In addition, using climate change scenarios coupled with real-time monitoring data from satellite imagery altogether allows for long-term trends to be estimated (Eyring et al., 2024a, b). It is also shown that combining the in-situ data with real-time monitoring data can further improves the accuracy of models (Wang, 2023; Yang et al., 2016). This altogether could possibility reveals uncertainty in long-term projections regardless of how long the period is (Eyring et al., 2024a, b; Parker, 2013). Previous studies demonstrated the difficulty and the uncertainty of resolving species composition changes under impacts of global change, particularly species known to be sensitive to change (Thuiller, 2004; Thuiller et al., 2009, 2019). However, it has been shown that using ensemble modeling combined with Stacked Species Distribution Models SSDM (Schmitt et al., 2017) and climate scenarios, allows for predicting shifts in habitat suitability, species richness, and assemblage structure (Abdelgadir et al., 2025; Brandt et al., 2017; Khan and Verma, 2022; Moullec et al., 2022). Therefore, using multi-model approach may provide a robust method to mitigate the uncertainties of species forecasting..", see ***lines 100-111***

In detail, see also my response to your next question: *"In general, kriging requires data that is spatially and temporally well distributed. The data here is pretty limited with 54 locations across an over 300,000 sq km area. What is the spatial correlation between data? How was the variogram like? Are the model results biased because of availability of more data in one section of the study area such as West and East Gotland Sea."*

The 139 data points represent moderate clustering according to Moran's I value = 0.22, and this is why we applied the variogram geostatistical tool implemented in the Kriging algorithm. In detail, applying the semi-variogram geostatistical tool in the Kriging algorithm allows quantifying the spatial autocorrelation by determining the range or the distance within which points are correlated and the sampling error or fine-scale variation called "nugget." Moreover, kriging assigns weights to neighboring points based on their spatial correlation, meaning that kriging firsthand considers the autocorrelation of points, not just their physical distance. By utilizing autocorrelation, kriging removes spatially correlated noise and creates a fine and less smeared kriged map. Kriging also ensures that the final residuals are not autocorrelated by modeling the main predictors and creating the best linear unbiased predictions. Taken together, we believe that kriging is able not just to resolve the spatial autocorrelation in the dataset but also to create fine, unbiased predictions.

For the second pathway of the analysis (i.e., the ensemble learning), and to overcome the overfitting issue, we generated what are so-called "pseudoabsence data points," which represent those unsampled sites across the study area. The general idea behind it is to generate points in the study area that will be used to compare the observed environment (represented by the presences) against what is available. Those points are NOT to be considered as absences and rather represent the available environment (see Barbet-Massin et al., 2012). Moreover, when learning dependence from data, to avoid overfitting, it is important to divide the data into the training set and the testing set. We first train our model on the training set, and then we use the data from the testing set to gauge the accuracy of the resulting model. Here, the dataset is split into two segments: 70% training and 30% testing data, which allows the model to learn on one subset of the data (training set) and evaluate its performance on another subset (testing set). This ensures that the model generalizes well to new data, making it more robust and reliable.

The current prediction period is too long a time frame to predict using modelling, especially for something so dynamic as algal blooms which are sensitive to a lot of factors. Why not use a shorter time frame like 2040 or 2050. Also, could the reason for similar results from all GCMs be because the model has some flaws that remain unaddressed.

Re: First, we selected this scenario (SSP5.85-2100), providing that this is the worst-case climate scenario with the highest emissions, severe global warming, and where temperatures are projected to rise significantly by the end of the century. First we added a sentence illustrating why this scenario is critical of the Baltic Sea. See below:

"For instance, climate scenario SSP5-8.5 for 2100 is critical for the Baltic Sea with projected surface temperature increases up to 3.2°C (Meier et al., 2022). This scenario is combined with increased runoff, accelerated oxygen depletion, decreased salinity, and worsened algal blooms (Friedland et al., 2012; Meier, 2006). All these factors, together with lack of proper modeling approach, make the current and future prediction of nodularin a methodological challenge.." See ***lines 70-73***

Secondly, we technically explained how ensemble learning handling such long period of time. This paragraph can be read:

"Using Ensemble Learning (multiple models instead of single model) is known to quantify uncertainty in long-term projections (Álvarez Fanjul et al., 2025; Eyring et al., 2024a, b). In addition, using climate change scenarios coupled with real-time monitoring data from satellite imagery altogether allows for long-term trends to be estimated (Eyring et al., 2024a, b). It is also shown that combining the in-situ data with real-time monitoring data can further improves the accuracy of models (Wang, 2023; Yang et al., 2016). This altogether could possibility reveals uncertainty in long-term projections regardless of how long the period is (Eyring et al., 2024a, b; Parker, 2013). Previous studies demonstrated the difficulty and the uncertainty of resolving species composition changes under impacts of global change, particularly species known to be sensitive to change (Thuiller, 2004; Thuiller et al., 2009, 2019). However, it has been shown that using ensemble modeling combined with Stacked Species Distribution Models SSDM (Schmitt et al., 2017) and climate scenarios, allows for predicting shifts in habitat suitability, species richness, and assemblage structure (Abdelgadir et al., 2025; Brandt et al., 2017; Khan and Verma, 2022; Moullec et al., 2022). Therefore, using multi-model approach may provide a robust method to mitigate the uncertainties of species forecasting..", see ***lines 100-111***

In general, kriging requires data that is spatially and temporally well distributed. The data here is pretty limited with 54 locations across an over 300,000 sq km area. What is the spatial correlation between data? How was the variogram like? Are the model results biased because of availability of more data in one section of the study area such as West and East Gotland Sea.

Re: The 139 data points represent moderate clustering according to Moran's I value = 0.22, and this is why we applied the variogram geostatistical tool implemented in the Kriging algorithm. In detail, applying this tool allows quantifying the spatial autocorrelation by determining the range or the distance within which points are correlated and the sampling error or fine-scale variation called "nugget." Moreover, kriging assigns weights to neighboring points based on their spatial correlation, meaning that kriging firsthand considers the autocorrelation of points, not just their physical distance. By utilizing autocorrelation, kriging removes spatially correlated noise and creates a fine and less smeared kriged map. Kriging also ensures that the final residuals are not autocorrelated by modeling the main predictors and creating the best linear unbiased predictions. Taken together, we believe that kriging is able not just to resolve the spatial autocorrelation in the dataset but also to create fine, unbiased predictions.

This procedure is now added to the introduction part and can be read as:

"In the case of small dataset and clustered sampling points, kriging algorithm allows quantifying the spatial autocorrelation by determining the range or the distance within which points are correlated, and the sampling error or fine-scale variation called "nugget". Moreover, kriging assigns weights to neighboring points based on their spatial correlation, meaning that kriging firsthand considers the autocorrelation of points, not just their physical distance (Antonakos and Lambrakis, 2021; Chilès and Delfiner, 1999). By utilizing autocorrelation, kriging removes spatially correlated noise and creates a fine and less smeared kriged map. Kriging also ensures that the final residuals are not autocorrelated by modeling the main predictors and creating the best linear unbiased predictions (Chilès and Delfiner, 1999; Fotheringham and Rogerson, 2009).", see ***lines 80-87***

For the second pathway of the analysis (i.e., the ensemble learning), and to overcome the overfitting issue, we generated what are so-called "pseudoabsence data points," which represent those unsampled sites across the study area. The general idea behind it is to generate points in the study area that will be used to compare the observed environment (represented by the presences) against what is available. Those points are NOT to be considered as absences and rather represent the available environment (see Barbet-Massin et al., 2012). Moreover, when learning dependence from data, to avoid overfitting, it is important to divide the data into the training set and the testing set. We first train our model on the training set, and then we use the data from the testing set to gauge the accuracy of the resulting model. Here, the dataset is split into two segments: 70% training and 30% testing data, which allows the model to learn on one subset of the data (training set) and evaluate its performance on another subset (testing set). This ensures that the model generalizes well to new data, making it more robust and reliable.

This procedure is now added to the revised version and is read as:

"Algorithms in ensemble learning can resolve the issues concerning autocorrelation, clustered samples, small datasets, and unsampled sites across the study area. Algorithms overcome the overfitting by generating what are called "pseudo-absence data points" (Hysen et al., 2022), which represent those unsampled sites across the study area and in a way compare the observed environment (represented by the presence) against what is available. Algorithms also avoid overfitting by dividing the data into training and testing sets, which allows the model to learn on one subset of the data (training set) and evaluate its performance on another subset (testing set). All these procedures together ensure that the model generalizes well to new data, making it more robust and reliable.." See ***lines 112-118***

I would have like to seen more about the limitations of the model structure over the results for 75 years which can vary substantially from the current modelled scenarios.

Re: The study conceptual modeling design aims to address all these limitations. For instance, using Ensemble Learning (multiple models instead of single model) is known to quantify uncertainty in long-term projections. In addition, using climate change scenarios coupled with real-time monitoring data from satellite imagery altogether allows for long-term trends to be estimated. By combining the in-situ data to previous further improves the accuracy of models. This altogether allows for accurate predictions for the next 75 years (2100 climate scenario) and reveals uncertainty in long-term projections regardless of how long the period is.

We added a paragraph demonstrating the procedure of handling long-term projections. This statement can be read as:

"Using Ensemble Learning (multiple models instead of single model) is known to quantify uncertainty in long-term projections (Álvarez Fanjul et al., 2025; Eyring et al., 2024a, b). In addition, using climate change scenarios coupled with real-time monitoring data from satellite imagery altogether allows for long-term trends to be estimated (Eyring et al., 2024a, b). It is also shown that combining the in-situ data with real-time monitoring data can further improves the accuracy of models (Wang, 2023; Yang et al., 2016). This altogether could possibility reveals uncertainty in long-term projections regardless of how long the period is (Eyring et al., 2024a, b; Parker, 2013). Previous studies demonstrated the difficulty and the uncertainty of resolving species composition changes under impacts of global change, particularly species known to be sensitive to change (Thuiller, 2004; Thuiller et al., 2009, 2019). However, it has been shown that using ensemble modeling combined with Stacked Species Distribution Models SSDM (Schmitt et al., 2017) and climate scenarios, allows for predicting shifts in habitat suitability, species richness, and assemblage structure (Abdelgadir et al., 2025; Brandt et al., 2017; Khan and Verma, 2022; Moullec et al., 2022). Therefore, using multi-model approach may provide a robust method to mitigate the uncertainties of species forecasting..", see ***lines 100-111***

Applying this model framework in other locations, needs context on strengths and weaknesses of the formulation so that it can be modified accordingly to meet the requirements of that problem.

Re: We fully agree. Regardless of the model's high accuracy and performance, we highlighted the limitations of the study considering the limited data points used and the shorter time of the sampling that extends for only one year.

This statement can be read:

"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. Moreover, the integrated model must be carefully inspected to ensure that it is still interpretable and computationally practical, providing that model success and performance are strongly reliant on the quality of the input data". See* ***lines 518-522***

Additionally, the results are presented in a confusing manner, there is too much back and forth between too many scenarios with numbers presented in both tables and text. I had a hard time following the narrative. I would recommend the question be stated clearly in the introduction and results focused around answering those particular questions. Maybe present the numbers in table and text in the narrative with numbers being brought up where it is critical for the narrative.

Re: The introduction and research question are now revised according to reviewer suggestion. The thread of the analysis is now following the narrative of the introduction. Since the study contains extensive statistical analysis, we included two supplementary materials data analysis (including figures + tables) and indicated within the main text. Major analysis was included within the main text as tables while other as in-text results (such as ANOVA). Also, a summary paragraph was included in the method section to guide the reader through the methods. This summary can be read as follows:

"2.1 Overview of data processing and model setup

"We compiled nodularin and Nodularia spumigena data collected across the Baltic Sea region and paired these observations with a suite of environmental variables derived from monitoring programs and high-resolution Copernicus model products. We harmonized all environmental and climate scenario datasets into a common spatial framework, extracted conditions for each sampling point, and prepared raster layers for geostatistical and machine learning analyses. We then applied empirical Bayesian kriging, co-kriging, and a multi-algorithm ensemble modeling workflow to predict present and future nodularin occurrence and N. spumigena distribution. Finally, we evaluated model performance and quantified environmental drivers using cross-validation, Bayesian regression, ANOVA/GLM, PCA, and LGM to assess how physical and biogeochemical conditions shape toxin patterns across the Baltic Sea. Geostatistical interpolation and ensemble learning modeling flow are illustrated in Fig 1." see **lines 128-137**

Could the reason for no nodularin in Bothnian and Southern Kattegat be the data not detecting any nodularin at the time of data collection. I have serious doubts about a model developed using such high accuracy developed using such limited data.

Re: Could be one reason. To clarify this point we added the following sentence in the introduction section:

Additionally, it has been shown that filamentous cyanobacteria over longer time scales not so common in Bothnian Bay (Olofsson et al., 2021) nor in Kattegat (Olofsson et al., 2020b). see ***lines 66-68***

With regards to the small dataset, please refer to our pervious response on this challenge. We added a new paragraph answering the reviewer doubts about the limited dataset and how the approach handles this issue.

This procedure is now added to the introduction part and can be read as:

"In the case of small dataset and clustered sampling points, kriging algorithm allows quantifying the spatial autocorrelation by determining the range or the distance within which points are correlated, and the sampling error or fine-scale variation called "nugget". Moreover, kriging assigns weights to neighboring points based on their spatial correlation, meaning that kriging firsthand considers the autocorrelation of points, not just their physical distance (Antonakos and Lambrakis, 2021; Chilès and Delfiner, 1999). By utilizing autocorrelation, kriging removes spatially correlated noise and creates a fine and less smeared kriged map. Kriging also ensures that the final residuals are not autocorrelated by modeling the main predictors and creating the

best linear unbiased predictions (Chilès and Delfiner, 1999; Fotheringham and Rogerson, 2009).", see ***lines 80-87***

Also see:

"Algorithms in ensemble learning can resolve the issues concerning autocorrelation, clustered samples, small datasets, and unsampled sites across the study area. Algorithms overcome the overfitting by generating what are called "pseudo-absence data points" (Hysen et al., 2022), which represent those unsampled sites across the study area and in a way compare the observed environment (represented by the presence) against what is available. ." see ***lines 133-136****.*

"Algorithms in ensemble learning can resolve the issues concerning autocorrelation, clustered samples, small datasets, and unsampled sites across the study area. Algorithms overcome the overfitting by generating what are called "pseudo-absence data points" (Hysen et al., 2022), which represent those unsampled sites across the study area and in a way compare the observed environment (represented by the presence) against what is available. Algorithms also avoid overfitting by dividing the data into training and testing sets, which allows the model to learn on one subset of the data (training set) and evaluate its performance on another subset (testing set). All these procedures together ensure that the model generalizes well to new data, making it more robust and reliable.." See ***lines 112-118***

If the paper discusses the causes for limited data, potential fallacies in its current formulation and the limitations of the results in its current state, I think it would be much more beneficial than just discussing how great the model performance is.

Re: The data limitation was addressed as cavities in the study; see **lines 518-522**

There are also some grammatical and structural corrections which can be fixed pretty easily. Attached below are specific locations I noticed.

Specific corrections:

Line 13: species. However,

Re: Corrected to "However, " **line 14.**

Line 75: This could be because the phenomenon....

Re: Corrected to "This could be because the phenomenon....", **lines 93-94 .**

Line 78 : ... identification of patterns ... ??

Re: Corrected to "identification of patterns", **line 96**

Line 80 : multiple variables

Re: Corrected to "multiple variables", **line 98**

Line 141: Would be better is this line is written differently. Starting a line with how is not recommended.

Re: Corrected and re-written as *"The cross-validation method 'leave-one-out resampling' was used to examine how interpolation model fits the data",* ***lines 210-211***

Figure 2. Good consolidation of information. Resolution needs to be improved.

Re: Resolution of Figure 2 (now is figure 1 in the revised version) is now improved, **line 156**

Line 185: Is it starting with latin ?

Re: Latin text (from journal template) is now removed in the revised manuscript, **line 253**. The sentence in the revision will start by *"Prediction of nodularin spatial* occurrence…".

Line 314: ….ensemble learning….

Re: Corrected to "ensemble learning", **line 395**.

Citation: https://doi.org/10.5194/egusphere-2025-3290-RC2