Dear Thomas Gorgues,

The authors wish to thank Dr. Thomas Gorgues for his helpful comments regarding our manuscript "A novel multispecies approach for the detection of regime shifts in a plankton community - a case study in the North Sea". We will respond to comments below.

General comments

The overall quality of the manuscript is very good. The novelty of this piece of research relies on the robustness of the model to detect regime shifts and its validation through the study case in the North Sea. The manuscript is well written and the sections are well structured. The conclusion is supported by the results and analyses that were performed.

My major concern is the issue about the scale up of the results to the ecosystem level. As it is now, the analyses performed does not enable the scale up since only the plankton community (phytoplankton and zooplankton) was assessed. I have highlighted this issue in several places throughout the manuscript.

- Thank you for pointing this out. We have made the use of language throughout the manuscript much more careful.
- Specifically, all points where before we described the RST model as capable of detecting ecosystem wide regime shifts to regime shifts in the plankton community.

Specific comments

Besides the specific comments embedded in the attached pdf I want to highlight the following:

Regarding the Introduction:

In my opinion, the manner of beginning this section is not helpful. See comment in the attached pdf.

The term "planktonic ecosystem" needs a better explanation. See comment in the attached pdf.

- Thank you for these comments about the Introduction. The beginning of the manuscript is much more broad now, and less to the introduction of tipping points ecological regime shifts. Regime shifts are then defined in the following paragraph.
- See below for an example of this:
- "The global ocean plays a critical role in regulating climate on Earth through heat and carbon redistribution. Global mean ocean temperature has increased since the Industrial Revolution, potentially inducing extreme events (IPCC, 2019). Changes in the oceanic environment associated with long-term temperature increase include more frequent marine heatwaves (Oliver et al., 2018, 2019), decreased pH (Doney et al., 2009), loss of oxygen concentration (Heinze et al., 2021; IPCC, 2019), and changes to large scale circulation patterns such as the Atlantic Meridional Overturning Circulation (AMOC) (Johnson et al., 2020; Robson et al., 2014), salinity (Curry et al., 2009; Skliris et al., 2014), and stratification stability in the upper ocean (Hallegraeff, 2010; Sharples et al., 2006; Wells et al., 2020). The projected rate of ocean warming, acidification, and oxygen concentration is strongly dependent on the rate of greenhouse gas emissions (Kwiatkowski et al., 2020; Schmidtko et al., 2017). In 2017, signatories to the Paris Climate Agreement agreed therefore to limit global temperature increase to less than 2 °C by the end of the 21st Century (Fox-Kemper et al., 2021). At current emission rates, it is likely that we will exceed these limits by 2050 (Fox-Kemper et al., 2021), which will potentially lead to the crossing of planetary boundaries, tipping points and ecological regime shifts (Heinze et al.,

2021; Rocha, 2022)."

- We have removed all instances of "planktonic ecosystem" and instead used "plankton community".

Regarding the Methods:

Abundance data is missing the units. This is crucial for any cross-comparison. See comment in the attached pdf.

Regarding the Validation of the model:

In the majority of the figures, axis labels are incorrect or missing. See comments in the attached pdf.

- We apologize for this. All units have been added to figures now.
- We have also added axes titles and more coherent figure captions.

Regarding the Discussion:

I suggest to avoid the use of references to figures and tables. See comments in the attached pdf.

- We have removed figure and table references from the Discussion, apart from when comparing specific results from a figure directly to previous studies. See below text for details:
- "In Fig. 7, abrupt changes to the abundance of two phytoplankton species can be observed but were not accompanied by an increase in the regime shift probability time series above the critical gradient. This is similar to Beaugrand et al. (2014), where results suggested a regime shift near the beginning of their study period which could not be confirmed."
- "Preserving more of the dataset's original structure appears to allow the RST model to identify regime shifts with a low false positive rate, relative to other methods (Fig. **??**) (Haines et al., 2024; Rudnick and Davis, 2003)."

Please see comment in line 437-439 about the mismatch between the model results and changes in single plankton species.

- Thank you, this was an interesting question.
- We have removed a lot of this part of the Discussion and rewritten it. This was partly because of other problems commented on by Reviewer 2 and partly because a lot of what was written originally was outside the scope of this manuscript.
- We believe the replacement text will satisfy the original question:
- "The ability to calculate risk of regime shifts in the community by analyzing all species, instead of those assumed to be most important, exhibits the importance of the RST model. [...] Large and dramatic changes in species abundance can occur without an accompanying change in regime shift likelihood, as observed in *C. finmarchicus* abundance in the entire North Sea, or either *Ceratium* spp. or PCI in Area D1. The RST model is designed to predict regime shift likelihood for all species that are input, and changes to single species do not always affect the entire community (Fauchald et al., 2011; Djeghri et al., 2023). This is an important difference from other regime shift detection algorithms used in the North Sea (Beaugrand et al., 2014; Bedford et al., 2020b,a). Output from the original $regimeshifts$ function was multiplied by the mean abundance of species for that month, resulting in the RST model estimating regime shift likelihood for all species in a particular database rather than simply measuring the number of species exhibiting

abrupt changes. In this way, the RST model was kept appropriate for non-expert users while scaling the proportion that each species contributes to the percentage likelihood of a regime shift. Applying more specific weights to keystone species such as *C. finmarchicus* or *C. helgolandicus* may be an interesting avenue for future research."

Please see comment in line 473-474 about the detection of changes in phytoplankton and zooplankton species.

- Thanks for this comment. There is an interesting conversation to be had about how abrupt shifts in phytoplankton and zooplankton interact with each other (see Review 2 and the reply concerning whether bottom-up control is an appropriate assumption)
- Our original manuscript stated that bottom-up control of the North Sea ecosystem was assumed, and a window length of 24 months was therefore programmed into the RST model. We have rewritten this part of the Methods to justify our choice, and added Supplementary material showing the effects of different window lengths that the total likelihood of a regime shift occurring can be impacted by abrupt changes in both zooplankton and phytoplankton.

The rest of my specific comments are embedded in the attached pdf.

Technical corrections

I suggested a few number of technical corrections. These are embedded in the attached pdf.

- Thank you for these specific comments and annotations. We believe we have amended the manuscript based on these concerns.

Thank you again for taking the time to review our manuscript. The manuscript is much improved now.

The authors: Paul Dees, Friederike Fröb, Beatriz Arellano-Nava, David G. Johns, and Christoph Heinze

Dear Referee 2,

Thank you for your very comprehensive review of our manuscript "A novel multispecies approach for the detection of ecosystem regime shifts - a case study in the North Sea". We would like to respond to each of your comments.

A. General Comments

This manuscript presents a novel methodology, the Regime Shift Tool (RST) model, for detecting ecosystem regime shifts, with an application to planktonic time series from the North Sea using Continuous Plankton Recorder (CPR) data. The study addresses a critical gap in the detection of regime shifts, particularly within open-ocean systems, by synthesizing multiple plankton time series into a composite likelihood metric.

The manuscript introduces an innovative and potentially valuable methodological advancement for the detection of ecosystem regime shifts. It is generally well-organized, and the research question is both timely and significant. The RST model represents a promising addition to the suite of available analytical tools for detecting complex ecological transitions.

Nevertheless, substantial revisions are required to address methodological limitations, improve the clarity of the exposition, and ensure the robustness of the interpretation of results. Several aspects concerning the methodology, results presentation, and interpretation necessitate major improvements to enhance analytical rigor and overall coherence. I therefore recommend that the manuscript be considered for publication only after major revisions have been satisfactorily completed.

To strengthen the manuscript, the authors should: (i) expand the literature review to include a detailed comparison with existing regime shift detection methods; (ii) provide comprehensive justifications for all key modeling decisions, supported by sensitivity analyses; (iii) enhance the description of the RST model to improve clarity and reproducibility; (iv) improve the quality and contextualization of figures; (v) deepen the discussion of the ecological mechanisms underlying the observed regime shifts and their management implications; and (vi) moderate claims regarding predictive capabilities unless further robust evidence is provided.

- (i) The literature review has undergone a significant rewrite. In particular we have emphasized how our regime shift identification tool differs from previous methods, and the advantages and limitations of using the new tool.
- (ii) We have justified and explained in more thorough depth the key modeling decisions made, and where appropriate explained what would happen if different choices were made.
- (iii) The steps of how the Boulton and Lenton (2019) method was developed into our method has been explained more carefully. We hope there is greater clarity and reproducibility now.
- (iv) Captions and figures have been improved.
- (v) The discussion has been written to be more focused, though as pointed out later this manuscript was not written to be a management paper. What we seek to do in this paper has been explained more carefully.
- (vi) The language describing the method and what it can do has been made more careful and conservative.

B. Specific Comments

Introduction

The introduction successfully contextualizes the challenges associated with detecting ecosystem regime shifts and highlights the critical role of long-term ecological datasets, such as the Continuous Plankton Recorder (CPR) records, in addressing these challenges. Nevertheless, the review of existing regime shift detection methodologies remains insufficiently comprehensive. A more detailed and critical examination of prior approaches, particularly those developed by Beaugrand et al. (2014) and others, is warranted. The authors should explicitly articulate the comparative advantages and potential limitations of the RST model relative to these established methods, thereby more clearly situating their contribution within the broader methodological landscape.

- Thank you for these helpful comments.
- We have added thorough descriptions of what prior approaches did to identify regime shifts, and contrasted this approach to our own.
- We describe advantages of using the new method described here (more automation than previous methods, an idea of how much trust can be put in the method by providing an estimate of type I and II error production, suitability for non-expert users, future analysis may be able to derive drivers more easily) as well as limitations of the method (over simplification of multiple time series).

Methods

In the section on CPR data processing, the rationale for applying a logarithmic transformation (Equation 1) requires further elaboration. While data normalization is a standard practice, the authors should clarify the specific motivations for its use in this context and discuss any potential implications for the study's findings. Furthermore, it would be beneficial to understand if the authors explored alternative transformations and why the logarithmic transformation was ultimately chosen. Similarly, the description of seasonality removal (Equation 2) is overly succinct. Greater detail is needed regarding the methodologies employed, including any filtering techniques or assumptions made. Importantly, the potential influence of pre-processing steps on the detection of regime shifts, which may themselves exhibit seasonal patterns, should be explicitly addressed. It is also important to consider the limitations of the CPR data itself. The authors should discuss potential biases inherent in the data, such as spatial coverage, sampling frequency, and the possibility of damage to plankton during the sampling process.

- We have added a paragraph describing the advantages and limitations of using CPR data. We have clarified why logarithmic transformations were applied and the implications of this.
- "Abundance data concerning phytoplankton and zooplankton from 1958 until 2019 were obtained from the CPR (see Reid et al. (2003) for detailed methodology). Abundance data from the CPR for the over 200 phytoplankton species and 80 zooplankton species are given in captured organisms per 10 nautical miles of towing (Reid et al., 2003). Each length of CPR silk corresponding to 10 nautical miles is one sample. Removing the seasonality trends of these data was accomplished by converting abundances into monthly anomalies for each year using Eq. 1 and Eq. 2.
  Because the CPR survey is conducted by the use of ships of opportunity, data from the CPR are highly variable in spatial and temporal resolution (Reid et al., 2003; Richardson et al., 2006). Sampling devices are towed behind ships at an average depth of 7 m, but this is also variable. There are some groups of plankton which cannot be collected by the CPR survey, primarily large zooplankton (Marques et al., 2024) and some smaller species of phytoplankton (Richardson et al., 2006; Mengs et al., 2024). However, the distance covered and the period of operation provided by the CPR survey is unmatched by other plankton datasets, and is thus a highly valuable source of ecological data (Beaugrand,

2004; Beaugrand et al., 2014; Djeghri et al., 2023)."

In the section on CPR data processing, the rationale for applying a logarithmic transformation (Equation 1) requires further elaboration. While data normalization is a standard practice, the authors should clarify the specific motivations for its use in this context and discuss any potential implications for the study's findings. Furthermore, it would be beneficial to understand if the authors explored alternative transformations and why the logarithmic transformation was ultimately chosen. Similarly, the description of seasonality removal (Equation 2) is overly succinct. Greater detail is needed regarding the methodologies employed, including any filtering techniques or assumptions made. Importantly, the potential influence of pre-processing steps on the detection of regime shifts, which may themselves exhibit seasonal patterns, should be explicitly addressed.

- We have now elaborated on the rationale for applying the logarithmic transformation in the revised manuscript. As now stated in the methods section, the log-transformation was applied to reduce strong seasonal contrasts in plankton abundance, particularly during bloom periods, and to stabilize variance across the time series. This approach is widely used in plankton studies where data often span several orders of magnitude and exhibit right-skewed distributions. We considered alternative transformations (e.g., square root), but the logarithmic transformation was ultimately chosen for its effectiveness in reducing skewness and preserving ecological interpretability in terms of proportional changes.
- "The logarithmic transformation was applied to reduce the strong seasonal contrasts in abundance, particularly during bloom periods, and to stabilize variance across time series. This approach is commonly used for plankton data, where values can span several orders of magnitude (Djeghri et al., 2023; Beaugrand, 2014; Bedford et al., 2020a). Logarithmic transformation necessitates the additional manipulation of adding half the limit of quantification to avoid calculating logs of zero. Changing original data in two ways instead of one, such as in square root transformations, is not ideal but logarithmic transformations were chosen because of the exponential difference in seasonal range observed in plankton abundance data, and due to its use in previous studies (Beaugrand, 2014; Dees et al., 2017)."
- The use of equations 1 and 2 have been fully explained in the text of the article.
- "Mean abundance at each time step was calculated using Eq. 2, which also removed the seasonality signal from abundance time series. [...] The regime shift detection algorithm implemented here follows the method described by Boulton and Lenton (2019). Boulton and Lenton (2019)'s method was designed for use on detrended data, which required the removal of a seasonal signal from abundance time series (see Eq. 2). Converting raw data into logarithmically transformed anomalous means using Eq. 1-2 is a necessary step before detecting abrupt shifts, as described in the following section. There is a risk that important features of the dataset, such as maxima and minima, are removed with this transformation. This is likely to result in more conservative estimates of abrupt changes in abundance. The largest driving force of plankton abundance throughout the year in temperate oceans is seasonality, but this was not the primary concern of this study. This model may not be suitable for identification of regime shifts exhibiting as changes to phenology, or the timing of phytoplankton blooms."

The explanation of the RST model, based on Boulton and Lenton (2019) and its subsequent adaptation by Arellano-Nava et al. (2022), is currently insufficient for readers unfamiliar with these references. A more thorough, self-contained description of the algorithmic steps within the manuscript would greatly improve accessibility. In addition, it is crucial to ascertain whether the linear regressions employed in the Boulton and Lenton (2019) method were checked for *underlying assumptions such as normality and homoscedasticity.*

- We appreciate the reviewer highlighting the lack of clarity in our description of the method. We agree that it was not thoroughly explained and have now expanded the text to better convey the underlying idea and algorithmic steps. Regarding the suggestion about checking for assumptions of normality or homoscedasticity, we have clarified that these are not required in this context, as the linear regressions are used solely to estimate local rates of change rather than for statistical inference.
- See for example these lines from the amended document: "The core idea behind this approach [(Boulton and Lenton, 2019)] is to detect anomalous rates of change in a time series, based on the premise that regime shifts correspond to unusually steep increases or decreases in the gradient over short periods. The algorithm begins by dividing the time series into fixed-length, non-overlapping segments. Within each segment, a simple linear regression is applied to estimate the slope, representing the local rate of change. These slopes are then compared across all segments. To identify anomalous trends, the algorithm calculates the median slope and the median absolute deviation (MAD). Slopes that deviate from the median by more than three MADs are flagged as anomalous. In this context, linear regression is used solely to quantify rates of change; since no statistical inference is involved, assumptions such as normality and homoscedasticity are not required (Schmidt and Finan, 2018)."

The selection of a 24-month rolling window for calculating the weighted probability of regime shifts is a critical modeling decision that requires more rigorous justification. A sensitivity analysis examining the influence of varying window lengths would strengthen the methodological foundation. It is also important to address whether the authors considered the potential impact of autocorrelation within the time series when determining the rolling window size, as autocorrelation can significantly affect the results of such analyses.

- We thank the reviewer for pointing out that the choice of a 24-month rolling window needed to be explained. We have explained this choice in the text, and added a series of figures where the rolling window was increased from 6 months to 60 months in the Supplemental Material.
- "A rolling mean with a window length of 24 months was applied to the time series of abrupt shift likelihood, for each species. The length of this rolling window was chosen because changes in the phytoplankton community are assumed to influence the zooplankton community (Capuzzo et al., 2018; Marques et al., 2024) and we assumed that two annual cycles would be long enough to capture this influence. After making this assumption, a series of figures were made to test the effect of iteratively changing the number of months in the rolling window. The code provided here allows the user to explore alternative rolling window lengths, and we encourage future studies to do so (see Supplemental Material)."

Equation 5, which integrates phytoplankton and zooplankton indicators, is predicated on an assumption of bottom-up control. Although the authors acknowledge this simplification, a more extensive discussion of its ecological plausibility, and the potential impact of alternative mechanisms such as top-down regulation or lateral environmental forcing, is necessary. It would also be valuable to discuss how the model could be adapted in future research to incorporate top-down or mixed control scenarios.

- We appreciate the reminder to fully explain the use of equations. We have explained the use of equation 5 and given some context around the potential for bottom-up or waist-wasp controls.

- "The product $P_{phy}(t)P_{zoo}(t)$ captures co-occurring abrupt changes in both phytoplankton and zooplankton, reflecting potential bottom-up propagation of regime shifts. Doubling this interaction term emphasizes the influence of phytoplankton variability on zooplankton dynamics within a 24-month window. There is some controversy concerning the relative strength of bottom-up (Di Pane et al., 2022) and wasp-waist controls (Fauchald et al., 2011) on zooplankton in the North Sea. However, we here assume a change in phytoplankton could induce sustained changes in zooplankton, which allowed us to quantify $\mathrm{P_{RS}}(t)$."

The decision to adopt a 20% threshold for regime shift identification also demands greater justification. A sensitivity analysis evaluating the effects of alternative thresholds on regime shift detection outcomes would provide essential context for interpreting the results.

- Thank you for this suggestion. We have justified the use of 20% as a critical gradient by referring to the validation test comparing the use of 18% and 20%. When 20% is used as a critical gradient, fewer false positives are generated.

Regarding model validation, the description of the simulated datasets generated using Equations 6 and 7 could be improved. The authors should explicitly discuss the extent to which these simulations reflect the complexity of real-world plankton dynamics. Moreover, the rationale for parameter choices, notably $\alpha = 0.99$ and $\sigma = 0.2$, should be thoroughly explained.

- Equations 6 and 7 have been explained in more detail, with greater explanation of the parameter choices made.
- "[...] σ is a constant standard deviation, set to $\sigma = 0.2$ [...]. Standard deviation for species in which abrupt shifts were induced was larger than when shifts were not induced. The original reason for adding this effect was because it has been shown variation is increased before regime shifts caused by critical slowing down (Scheffer et al., 2001; Scheffer, 2009). It became clear that our interest should not be limited to regime shifts preceded by critical slowing down, but increasing variation for simulated species experiencing abrupt shifts resulted in increase range and a greater chance of their time series appearing in graphs (Fig. 2 - 7)".

Furthermore, it is essential to acknowledge the inherent limitations of any model. The authors should more explicitly state the limitations of the RST model, discussing its potential applicability and the types of regime shifts it might not be able to detect. The model's sensitivity to noise is also a concern. While Type I and II errors are discussed, a more thorough noise sensitivity analysis would bolster confidence in the model's robustness.

- Thank you for highlighting this important part of the study. We have discussed important limitations of this modeling approach in the Methods sections and throughout the study, although the word "limitation" was not always used. We have further explained more of the limitations and the implications of this on how the model can be interpreted.
- We have made the limitations of using CPR data more clear under the subtitle "2.2 Ecological Data", noting that not all plankton species can be captured in CPR surveys and the variable coverage over time.
- We have noted that the use of the 24-month rolling window and the 20% critical gradient threshold may impact on whether regime shifts are detected. We have tried to sow the effects of varying these levels and recommended future users of the RST model thoughtfully explore using other choices for these options.
- In the beginning of the Discussion, we note how is impossible using the RST model alone

to determine whether ecological regime shifts have occurred due to advection or a changing physical environment.

- "Distinguishing regime shifts in North Sea plankton from changes in abundance due to advection is a limitation of this approach. Future work could address this limitation by incorporating spatial patterns from hydrodynamic models, helping to separate genuine ecological shifts from transport-driven variability. A step towards this would be to divide the North Sea into hydrodynamically appropriate regions van Leeuwen et al. (2015), but this is beyond the scope of the present study."
- We have changed the phrase "potential issue" in the Discussion, when commenting on when the earliest regime shifts can be detected by the RST model to "limitation".
- We have made more clear that a limitation of the model is validation tests show that at least 14 years between regime shifts is needed for the model to distinguish between different events.

Results

The presentation of the results is generally effective; however, the clarity and informativeness of the figures could be substantially improved. All axes should be clearly labeled with appropriate units, and figure captions should include sufficient contextual detail to aid interpretation. Additionally, annotating Figures 2–7 and 9–13 with the timing of known environmental events—such as major storms, temperature anomalies, or changes in fishing pressure—would provide valuable context and enhance the interpretability of observed patterns.

- Axes labels and units have been added to all figures which lacked them before. Greater detail has been given to figure captions. Adding timing of major environmental events would turn this study into a much deeper and more complex analysis. We have not discussed drivers of regime shifts in concrete terms; this study describes a new method of detecting regime shifts and discusses how it can aid our collective understanding of ecological changes. Future studies can use the detection method to compare regime shifts with environmental events, but this is not the focus of the present study.

The discussion of Type I and Type II errors constitutes a notable strength of the manuscript. However, the authors should more explicitly relate these error estimates to the confidence with which regime shifts are identified. Establishing a clearer linkage between the quantified uncertainty and the interpretation of regime shift periods would enhance the robustness and transparency of the study's conclusions.

- Thank you for noting this. We have more explicitly linked using the results of comparing Type I and Type II errors under different critical gradients with our choice to use a critical gradient of 20% when identifying regime shifts in CPR data.

Discussion

The discussion provides a reasonable interpretation of the study's results and acknowledges key limitations. However, several areas require further development to enhance the depth, rigor, and critical analysis of the findings.

First, a more detailed comparison with previous studies on regime shifts in the North Sea is necessary. Although prior research is mentioned, the authors should explicitly discuss the degree of concordance or divergence between their findings and earlier reports. Specifically, the comparison should assess how the timing and magnitude of regime shifts identified by the RST model align with or differ from those detected using established methods. Furthermore, the discussion *should articulate the novel contributions of the RST model beyond merely analyzing a greater number of species, clarifying the specific insights it offers into ecosystem dynamics.*

- Thank you for pointing this out. We have added substantially to the discussion, and explicitly compared regime shifts identified by the RST model to previous methods.
- We have discussed how, as the RST produces a single time series of regime shift likelihood, output from the model can be compared more easily to potential physical drivers of regime shifts.
- "A single time series of regime shift likelihood may lead to the future development of automated detection algorithms. Comparison of regime shift likelihood to potential drivers is less challenging to implement than more complex multivariate models."

The ecological mechanisms underpinning the observed regime shifts also warrant deeper examination. While changes in plankton abundances are identified as potential drivers, a more nuanced exploration of environmental forcing factors (e.g., temperature variability, nutrient availability), biotic interactions (e.g., competition, predation), and other relevant processes is needed. In particular, the roles of climate change and anthropogenic impacts should be more thoroughly integrated into the discussion. Although these drivers are introduced earlier in the manuscript, a stronger and more explicit connection between them and the identified regime shifts is necessary to contextualize the findings.

- This is an interesting observation, but outside the scope of the present manuscript. While we agree that the combining the results of our model with different potential driving mechanisms would be interesting, the main focus of our manuscript was to introduce a new method, throroughly test and validate it, and apply it to existing data to test its applicability. Discussing potential drivers is out of scope for our manuscript and is planned for future work.

Moreover, the discussion should more explicitly address the practical implications of applying the RST model in ecosystem management contexts. Potential applications, benefits, and limitations should be thoroughly considered. For instance, the authors could discuss how the model might inform fisheries management, conservation strategies, or adaptive management frameworks. Addressing the challenges inherent in translating model outputs into actionable management advice would also enhance the practical relevance of the study.

- Thank you for this comment. We agree that the RST may be useful for ecosystem management contexts in the future, this is not the present focus of the manuscript.
- We have added a suggestion that regime shift identification may be used in management situations, as the RST tool was designed for non expert use. We think this is sufficient for this manuscript where we seek to introduce a new method for identifying regime shifts and discuss results of the model.
- "If this method is used to predict regime shifts in future scenarios or management situations, [...]."

The influence of advection processes on plankton dynamics, and the associated difficulty of distinguishing these from genuine regime shifts, is acknowledged but remains underdeveloped. A more detailed treatment of this limitation, including suggestions for future methodological improvements, would strengthen the manuscript. For example, incorporating hydrodynamic modeling or spatial analyses into the RST framework could offer promising avenues for future research.

- This is a very welcome comment, thank you. We have further developed discussion of the difference between regime shifts caused by advection in the North Sea and those caused by changes to the physical environment.

- "Distinguishing regime shifts in North Sea plankton from changes in abundance due to advection is a limitation of this approach. Future work could address this limitation by incorporating spatial patterns from hydrodynamic models, helping to separate genuine ecological shifts from transport-driven variability. A step towards this would be to divide the North Sea into hydrodynamically appropriate regions van Leeuwen et al. (2015), but this is beyond the scope of the present study."

The manuscript's discussion of the predictive potential of the RST model, although intriguing, is currently under-supported by the presented evidence. The authors are encouraged to temper claims regarding predictive capability or to provide more substantial empirical support. A critical evaluation of the model's current limitations in forecasting regime shifts, including its sensitivity to noise and the risk of false positives or negatives, should be included. In addition, outlining specific future research directions to validate and enhance the model's predictive accuracy would be valuable.

- Thank you for this reminder to be more careful with language used. There is some apparent predictive power of the model, but this is due to the fixed length windows used in the original Boulton and Lenton (2019) method rather than anactual predictive ability.
- We have removed the Discussion heading "5.2 Apparent early warning for imminent regime shifts" and combined the parts not removed with the following heading which was renamed "RST model output".
- We have added a number of suggestions for how future research could use the RST model:
- "The code provided here allows the user to explore alternative rolling window lengths, and we encourage future studies to do so."
- "Not all plankton populations are bottom-up controlled. Regime shifts may sometimes be caused by removal of predators (Eklöf et al. 2020), and we encourage future studies to explore modifying the weights described here."
- "Applying more specific weights to keystone species such as *C. finmarchicus* or *C. helgolandicus* may be an interesting avenue for future research."
- We believe the manuscript is easier to read and understand now.

Furthermore, the limitations of the RST model itself should be discussed in greater depth. Rather than simply acknowledging that "all models are simplifications," the authors should critically evaluate the contexts in which the RST model may be less applicable, the types of regime shifts it may fail to detect, and the extent to which its performance depends on data quality. Specific limitations arising from the choice of a 24-month rolling window and the assumption of bottom-up control should be examined. Additionally, the statistical underpinnings of the model, including assumptions embedded in the Boulton and Lenton method and the potential influence of autocorrelation, should be critically assessed.

- As mentioned under the Methods section, we have more explicitly stated the limitations of the model, and of the data used.
- We have developed the end of the Discussion and the Conclusion to state particular problems the RST model may have if used.
- From the Discussion: "Decreasing the value of the critical gradient used to identify regime shifts in this study will likely decrease false positive, but validation tests performed in this study indicate doing so will increase the rate of false positives. Having a conservative

estimate of regime shift likelihood should increase the certainty in identified regime shifts. [...] We therefore advise RST model users to thoughtfully explore the use of different window lengths and critical gradients."

- From the Conclusion: "Although there are several advantages to condensing an estimate of regime shift likelihood into a single time series, limitations of this approach have also been discussed. These limitations include needing approximately 14 years between abrupt changes for the algorithm to detect regime shifts, and the inability to distinguish between apparent regime shifts caused by advection."

In summary, while the discussion provides a solid foundation, it requires substantial expansion to offer a more comprehensive, critical, and contextually grounded interpretation of the results. A stronger emphasis on limitations, broader ecological and management contexts, and practical applicability would significantly enhance the scientific contribution of the manuscript.

Thank you so much for a very thorough review. We believe the manuscript is more clear now, and the uses and limitations of the RST model have been made more obvious.

The authors: Paul Dees, Friederike Fröb, Beatriz Arellano-Nava, David G. Johns, and Christoph Heinze

*References

Beaugrand, G.: The North Sea regime shift: evidence, causes, mechanisms and consequences, Proceedings of Oceanography, 60, 245–262, https://doi.org/10.1016/j.dsr2.2008.12.022, 2004.

Beaugrand, G.: Pelagic ecosystems and climate change, Global Environmental Change, pp. 141–150, 2014.

Beaugrand, G., Harlay, X., and Edwards, M.: Detecting plankton shifts in the North Sea: a new abrupt ecosystem shift between 1996 and 2003, Marine Ecology Progress Series, 502, 82–104, https://doi.org/10.3354/meps10693, 2014.

Bedford, J., Ostle, C., Johns, D. G., Atkinson, A., Best, M., Bresnan, E., Machairopoulou, M., Graves, C. A., Devlin, M., Milligan, A., Pitois, S., Mellor, A., Tett, P., and McQuatters-Gollop, A.: Lifeform indicators reveal large-scale shifts in plankton across the North-West European shelf, Global Change Biology, 26, 3482–3497, https://doi.org/10.1111/gcb.15066, 2020a.

Bedford, J., Ostle, C., Johns, D. G., Budria, A., and McQuatters-Gollop, A.: The influence of temporal scale selection on pelagic habitat biodiversity indicators, Ecological Indicators, 114, 106 311, https://doi.org/10.1016/j.ecolind.2020.106311, 2020b.

Boulton, C. and Lenton, T.: A new method for detecting abrupt shifts in time series [version 1; peer review: 2 approved with reservations], F1000Research, 8, https://doi.org/10.12688/f1000research.19310.1, 2019.

Capuzzo, E., Lynam, C. P., Barry, J., Stephens, D., Forster, R. M., Greenwood, N., McQuatters-Gollop, A., Silva, T., van Leeuwen, S. M., and Engelhard, G. H.: A decline in primary production in the North Sea over 25 years, associated with reductions in zooplankton abundance and fish stock recruitment, Global change biology, 24, e352–e364, https://doi.org/10.1111/gcb.13916, 2018.

Curry, R., Dickson, B., and Yashayaev, I.: A change in the freshwater balance of the Atlantic Ocean over the past four decades, Nature, 426, 826–829, https://doi.org/10.1038/nature02206, 2009.

Dees, P., Bresnan, E., Dale, A. C., Edwards, M., Johns, D., Mouat, B., Whyte, C., and Davidson, K.: Harmful algal blooms in the Eastern North Atlantic ocean, Proceedings of the National Academy of Sciences, 114, E9763–E9764, https://doi.org/10.1073/pnas.1715499114, 2017.

Di Pane, J., Wiltshire, K. H., McLean, M., Boersma, M., and Meunier, C. L.: Environmentally induced functional shifts in phytoplankton and their potential consequences for ecosystem functioning, Global Change Biology, 28, 2804–2819, https://doi.org/10.1111/gcb.16098, 2022.

Djeghri, N., Boyé, A., Ostle, C., and Hélaouët, P.: Reinterpreting two regime shifts in North Sea plankton communities through the lens of functional traits, Global Ecology and Biogeography, 32, 962–975, https://doi.org/10.1111/geb.13659, 2023.

Doney, S. C., Fabry, V. J., Feely, R. A., and Kleypas, J. A.: Ocean Acidification: The Other CO2 Problem, Annual Review Of Marine Science, 1, 169–192, https://doi.org/10.1146/annurev.marine.010908.163834, 2009.

Fauchald, P., Skov, H., Skern-Mauritzen, M., Johns, D., and Tveraa, T.: Wasp-Waist Interactions in the North Sea Ecosystem, PLoS ONE, 6, https://doi.org/10.1371/journal.pone.0022729, 2011.

Fox-Kemper, B., Hewitt, H., Xiao, C., Aalgeirsdóttir, G., Drijfhout, S., Edwards, T., Golledge, N., Hemer, M., Kopp, R., Krinner, G., Mix, A., Notz, D., Nowicki, S., Nurhati, I., Ruiz, L., Sallée, J.-B., Slangen, A., and Yu, Y.: Ocean, Cryosphere and Sea Level Change, In Climate

Change 2021: The Physical Science Basis. Contribution of Working Group I to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change, pp. 1211—-1362, https://doi.org/10.1017/9781009157896.011, 2021.

Haines, H., Planque, B., and Buttay, L.: Poor performance of regime shift detection methods in marine ecosystems, ICES Journal of Marine Science, pp. 1–14, https://doi.org/10.1093/icesjms/fsae103, 2024.

Hallegraeff, G. M.: Ocean Climate Change, Phytoplankton Community Responses, And Harmful Algal Blooms: A Formidable Predictive Challenge, Journal of Phycology, 46, 220–235, https://doi.org/10.1111/j.1529-8817.2010.00815.x, 2010.

Heinze, C., Blenckner, T., Martins, H., Rusiecka, D., Döscher, R., Gehlen, M., Gruber, N., Holland, E., Øystein Hov, Joos, F., Matthews, J. B. R., Rødven, R., and Wilson, S.: The quiet crossing of ocean tipping points, Proceedings of the National Academy of Science, 118, 1–9, https://doi.org/10.1073/pnas.2008478118, 2021.

IPCC: IPCC Special Report on the Ocean and Cryosphere in a Changing Climate, Cambridge University Press, Cambridge, UK and New York, NY, USA, https://doi.org/10.1017/9781009157964, 2019.

Johnson, C., Inall, M., Gary, S., and Cunningham, S.: Significance of Climate Indices to Benthic Conditions Across the Northern North Atlantic and Adjacent Shelf Seas, Frontiers in Marine Science, 7, https://doi.org/10.3389/fmars.2020.00002, 2020.

Kwiatkowski, L., Torres, O., Bopp, L., Aumont, O., Chamberlain, M., Christian, J. R., Dunne, J. P., Gehlen, M., Ilyina, T., John, J. G., Lenton, A., Li, H., Lovenduski, N. S., Orr, J. C., Palmieri, J., Santana-Falcón, Y., Schwinger, J., Séférian, R., Stock, C. A., Tagliabue, A., Takano, Y., Tjiputra, J., Toyama, K., Tsujino, H., Watanabe, M., Yamamoto, A., Yool, A., and Ziehn, T.: Twenty-first century ocean warming, acidification, deoxygenation, and upper-ocean nutrient and primary production decline from CMIP6 model projections, Biogeosciences, 17, 3439–3470, https://doi.org/10.5194/bg-17-3439-2020, 2020.

Marques, R., Otto, S., Di Pane, J., Boersma, M., Meunier, C., Wiltshire, K. H., Möllmann, C., and Renz, J.: Response of the meso-and macro-zooplankton community to long-term environmental changes in the southern North Sea, ICES Journal of Marine Science, 81, 526–539, https://doi.org/10.1093/icesjms/fsad121, 2024.

Mengs, G., Stern, R. F., Clarke, J. L., Faith, M., and Medlin, L. K.: Mapping Selected Emergent Marine Toxin-Producing Organisms Using Historical Samples with Two Methods (Biosensors and Real-Time PCR): A Comparison of Resolution, Applied Microbiology, 4, 312–328, https://doi.org/10.3390/applmicrobiol4010021, 2024.

Oliver, E. C., Donat, M. G., Burrows, M. T., Moore, P. J., Smale, D. A., Alexander, L. V., Benthuysen, J. A., Feng, M., Sen Gupta, A., Hobday, A. J., et al.: Longer and more frequent marine heatwaves over the past century, Nature communications, 9, 1–12, https://doi.org/10.1038/s41467-018-03732-9, 2018.

Oliver, E. C., Burrows, M. T., Donat, M. G., Sen Gupta, A., Alexander, L. V., Perkins-Kirkpatrick, S. E., Benthuysen, J. A., Hobday, A. J., Holbrook, N. J., Moore, P. J., et al.: Projected marine heatwaves in the 21st century and the potential for ecological impact, Frontiers in Marine Science, 6, 1–12, https://doi.org/10.3389/fmars.2019.00734, 2019.

Reid, P., Colebrook, J., Matthews, J., Aiken, J., and Team, C. P. R.: The continuous plankton recorder: concepts and history, from plankton indicator to undulating recorders, Progress in Oceanography, 58, 117–173, https://doi.org/10.1016/j.pocean.2003.08.002, 2003.

Richardson, A., Walne, A., John, A., Jonas, T., Lindley, J., Sims, D., Stevens, D., and Witt, M.: Using continuous plankton recorder data, Progress in Oceanography, 68, 27–74, https://doi.org/10.1016/j.pocean.2005.09.011, 2006.

Robson, J., Sutton, R., and Smith, D.: Decadal predictions of the cooling and freshening of the North Atlantic in the 1960s and the role of ocean circulation, Climate Dynamics, 42, 2353–2365, https://doi.org/10.1007/s00382-014-2115-7, 2014.

Rocha, J. C.: Ecosystems are showing symptoms of resilience loss, Environmental Research Letters, 17, https://doi.org/10.1088/1748-9326/ac73a8, 2022.

Rudnick, D. L. and Davis, R. E.: Red noise and regime shifts, Deep-Sea Research 1, 50, 691–699, https://doi.org/10.1016/S0967-0637(03)00053-0, 2003.

Scheffer, M.: Critical transitions in nature and society, Princeton University Press, Princeton, NJ, 2009.

Scheffer, M., Carpenter, S., Foley, J. A., Folke, C., and Walker, B.: Catastrophic shifts in ecosystems, Nature, 413, 591–596, https://doi.org/10.1038/35098000, 2001.

Schmidt, A. F. and Finan, C.: Linear regression and the normality assumption, Journal of Clinical Epidemiology, 98, 146–151, https://doi.org/10.1016/j.jclinepi.2017.12.006, 2018.

Schmidtko, S., Stramma, L., and Visbeck, M.: Decline in global oceanic oxygen content during the past five decades, Nature, 542, 335–339, https://doi.org/10.1038/nature21399, 2017.

Sharples, J., Ross, O. N., Scott, B. E., Greenstreet, S. P., and Fraser, H.: Inter-annual variability in the timing of stratification and the spring bloom in the North-western North Sea, Continental Shelf Research, 26, 733–751, https://doi.org/10.1016/j.csr.2006.01.011, 2006.

Skliris, N., Marsh, R., Josey, S. A., Good, S. A., Liu, C., and Allan, R. P.: Salinity changes in the World Ocean since 1950 in relation to changing surface freshwater fluxes, Climate Dynamics, 43, 709–736, https://doi.org/10.1007/s00382-014-2131-7, 2014.

van Leeuwen, S., Tett, P., Mills, D., and van der Molen, J.: Stratified and nonstratified areas in the North Sea: Long-term variability and biological and policy implications, Journal of Geophysical Research: Oceans, 120, 4670–4686, https://doi.org/10.1002/2014JC010485, 2015.

Wells, M. L., Karlson, B., Wulff, A., Kudela, R., Trick, C., Asnaghi, V., Berdalet, E., Cochlan, W., Davidson, K., Rijcke, M. D., Dutkiewicz, S., Hallegraeff, G., Flynn, K. J., Legrand, C., Paerl, H., Silke, J., Suikkanen, S., Thompson, P., and Trainer, V. L.: Future HAB science: Directions and challenges in a changing climate, Harmful Algae, 91, 101 632, https://doi.org/10.1016/j.hal.2019.101632, 2020.