The paper by Lin and coauthors deals with marine heatwaves in the North Sea. The contents and quality of the presentation need improvement. I have some doubts on the methodology and clarity of the diagnosed mechanisms. The Authors should consider the comments provided below.

Response: Thank you very much. We carefully considered each of your comments and suggestions, and we would like to respond to them in the following content.

I have also noticed this paper https://www.researchsquare.com/article/rs-6503093/v1 from the same authors, but with a different first author. Contents are similar but not identical. Could you please clarify the situation?

Response: Thank you for raising this point. The manuscript you referred to is a preprint that was previously submitted to another journal and subsequently posted online. That submission was unfortunately not accepted for publication. We would like to clarify that the first author of the preprint is also the first author of the present manuscript. The corresponding author's name appeared first in the header of that preprint, but in the manuscript the order of authors is consistent in both versions.

The present manuscript in Ocean Science represents a substantially improved and revised version of that earlier work. In preparing the current submission to Ocean Science, we have reanalyzed the results, clarified the scientific focus, revised the methodology, and reorganized the manuscript to better address the scope of Ocean Science. While the two manuscripts share a common scientific background, the current version contains significant modifications and improvements and is not under consideration elsewhere.

General comments

While there may be implication for prediction, mentioning forecasting in the abstract can be misleading. I suggest to clarify that is a diagnostic analysis, instead.

Response: Thank you for this helpful comment. We agree that our study is based on diagnostic and mechanistic analyses rather than forecasting. To avoid potential misunderstanding, we will revise the abstract by clarifying the diagnostic nature of the study and by removing wording that could be interpreted as implying direct forecasting.

"variabilities" in the plural form sounds strange. Suggest using singular or another noun, e.g. "modes" depending on what you want to convey.

Response: We thank the reviewer for pointing this out. We agree that the plural form "variabilities" is confusing in this context. We will revise the manuscript by replacing "climate variabilities" with "climate modes" to better reflect the physical meaning.

Nonlinear quantities are computed for both atmospheric and ocean reanalysis data; can you estimate the error made with this compared to using daily data?

Response: Thank you for this important comment. We acknowledge that computing nonlinear terms from monthly mean reanalysis fields may differ from first calculating these terms at daily resolution and then averaging, and that high-frequency variability can contribute to nonlinear processes. Our analysis focuses on seasonal to interannual variability, for which monthly-mean fields are commonly used. At these timescales, high-frequency synoptic fluctuations tend to partially cancel when averaged, and their net contribution to the low-frequency nonlinear terms is generally smaller than that of the resolved mean and low-frequency anomalies. As a result, our estimates should be interpreted as diagnostics of low-frequency, large-scale contributions, rather than an exact closure of the instantaneous nonlinear budget.

While using daily data could provide a more complete quantification of high-frequency nonlinear effects, we will prove in the revised version that this does not affect the main conclusions regarding the dominant processes at seasonal to interannual timescales.

I am not sure to understand your use of K-means in this work. In other works (e.g. Vogt et al. 2022 10.3389/fclim.2022.847995, Wong et al. 2024 https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2023AV001059) the procedure is meant to identify groups of data points, while in your case two regions (Fig. 1) are identified.

Response: Thank you for this comment. We apologize that the explanation of the K-means analysis was not sufficiently clear in the original manuscript. In this study, K-means clustering is applied to the temporal evolution of MHW cumulative intensity (MHWCI) at each spatial grid point, rather than to individual MHW events or time steps. Each grid point is therefore represented by its MHWCI time series, and grid points that exhibit similar temporal variability are grouped by the clustering. As a result, the clusters naturally emerge as contiguous spatial regions (Fig. 1). We will revise the manuscript to better clarify this. Further validation of clustering robustness is provided in Appendix B. This analysis results in two well-separated and spatially coherent regions in the North Sea, broadly corresponding to the southern and northern subregions, indicating distinct temporal variability characteristics of MHWCI.

And from Fig. 3, I have the impression this just means seasonality. How can these two domains be related to MHW occurrence?

Response: Thank you for this comment. We agree that Fig. 3 highlights clear seasonal contrasts between the two domains. The K-means clustering is not applied to individual MHW events or occurrence number, but to the time series of MHWCI at each grid point. MHWCI integrates the occurrence, duration, and intensity of MHWs, and its seasonal evolution therefore provides a meaningful representation of MHW activity. By clustering grid points with similar seasonal and interannual variability of MHWCI, the analysis identifies regions that share common temporal characteristics of MHW activity. In this sense, the two domains are directly related to MHW occurrence through their distinct seasonal modulation of MHWCI, rather than being defined by climatological seasonality alone. We will add explanations in the revised version to clarify this.

From line 201 and following I understand MHWs may happen predominantly in one or the other region depending on the season, but more explanation is due.

Response: Thank you for this comment. We agree that additional explanation is needed. In the revised version, we will add an explanatory paragraph to explicitly link the seasonal diagnosis shown in Fig. 2b with the subsequent analysis strategy.

Various climate indices are shown; sources for the associated data is missing.

Response: We thank the reviewer for pointing this out. We will revise the manuscript to explicitly describe the data sources and calculation methods of all climate indices used in this study. A new subsection (Sect. 2.4) will be added to the Methods section, detailing the definitions, computational approaches, and data sources for the NAO, EAP, and AMV indices.

The discussion on the mechanisms is quite poor, as results for a much larger domain is presented, so local mechanisms are not discussed. More details should be provided on the stratification procedure;

Response: We thank the reviewer for the constructive comments. Regarding the discussion of mechanisms, the local processes were investigated through a detailed upper-ocean heat budget analysis conducted exclusively within the North Sea. However, we acknowledge that the current manuscript places stronger emphasis on large-scale circulation patterns and teleconnections, which may obscure the local nature of the diagnosed mechanisms. In the revision, we plan to reorganize the Results and Discussion sections to more explicitly highlight the heat budget analysis as the key representation of local physical mechanisms.

With respect to stratification, we understand the reviewer's comment as referring to vertical ocean stratification. Although stratification effects are implicitly included through the mixed layer depth (MLD) in the heat budget framework, this aspect was not sufficiently discussed in the current version. We plan to explicitly address the role of stratification and clarify how differences in MLD modulate the relative importance of surface heat fluxes and oceanic advection between the two clusters.

are you using fixed or dynamic thresholds? This should be consistent across indices, e.g. if they are all standardized already.

Response: We thank the reviewer for this helpful comment. We clarify that fixed thresholds are used throughout the analysis, but they are applied differently depending on the type of variable.

(1) Climate indices: The large-scale climate indices (NAO, EAP, and AMV) are obtained directly from publicly available datasets. These indices are anomaly-based indices derived from atmospheric or oceanic fields, and we use them in their original form without applying additional standardization. Positive and negative phases shown in Fig. 2 and Fig. 9 are therefore defined using a fixed threshold at zero, with values greater (less) than zero indicating positive (negative) phases.

(2) Cluster-specific MHWCI time series: The cluster-specific MHWCI time series are constructed in this study and are explicitly standardized to zero mean and unit variance. Periods of enhanced MHWCI are defined using a fixed threshold of +1 standard deviation, corresponding to values greater than 1 in the standardized time series. This threshold is applied consistently in all composite and regression analyses.

We will revise the Methods and Results sections to explicitly clarify these definitions and to avoid any ambiguity regarding the thresholds used.

The level of discussion when presenting figures (which are often hard to read) is insufficient, and the various domains used (e.g., larger in Fig. 10) complicates comparisons.

Response: We thank the reviewer for pointing out issues related to figure presentation and clarity. We agree that, in several cases, the discussion accompanying the figures does not sufficiently guide the reader through what is shown in the plots. In the revised manuscript, we plan to revise the Results section so that each figure is first described in a clear and descriptive manner (i.e., spatial patterns, temporal changes, and contrasts), before moving on to physical interpretations and discussion.

Regarding the use of different spatial domains (e.g., the larger domain in Fig. 10), we acknowledge that this may complicate direct comparison with other figures. To address this, we will explicitly highlight the key comparison regions by adding boxes and annotations in the figure and clarify in the captions and text. This will help guide the reader's attention and improve consistency across figures.

I would suggest to first analyse non-MHW variables and understand how they change over time and due to teleconnections, and then focus on the influence of these indices on MHWs.

Response: We appreciate this suggestion and agree that the current presentation may give the impression that the causal chain is not clearly structured. While our analyses of atmospheric and oceanic variables were conducted to diagnose the mechanisms underlying MHW variability, the manuscript currently introduces these results in direct connection with MHWs.

In the revision, we plan to improve the narrative flow by first more clearly describing the evolution of non-MHW variables (e.g., circulation, heat fluxes, mixed layer depth) and their modulation by large-scale climate modes, and then link these changes to the observed MHW responses. This restructuring will better reflect the physical causality and improve readability without altering the core analyses.

I feel like the summary of Fig. 11 is actually not aiding interpretation, as new concepts and indices are added. This should be simplified.

Response: We agree that the current summary figure (Fig. 11) introduces additional concepts and climate indices, which may complicate the interpretation. In the revised manuscript, we will simplify Fig. 11 so that it functions purely as a conceptual summary of the key mechanisms.

Appendix A seems unnecessary, as this has likely been done by the producers in more detail.

Response: We agree that the validation of the SST product has been extensively documented by the data producers. Following this recommendation, we will remove Appendix A from the manuscript. The validation will be mentioned in the Data section as supplementary information.

The work needs careful proofreading, as there are errors in the titles of most figures and references. Only some examples are given below.

Response: Thank you. We will carefully proofread the entire manuscript and correct typographical errors and inconsistencies in figure titles, references, and text throughout the paper.

Comments by line

l21 why speaking about "prediction"?

Response: We agree that the reference to "prediction" in the abstract is confusing. As the present study is based on diagnostic and mechanistic analyses rather than forecasting, we will revise the abstract to clarify its diagnostic nature and remove confusing wording.

l69 same comment as for the title

Response: We will revise the manuscript at line 69, replacing "climate variabilities" with "climate modes".

Fig. 1 typo in panel b title

Response: The typo in the title of Fig. 1b will be corrected from "MHWTI" to "MHWCI".

l88/184 as far as I understand, cumulative intensity is a per-event quantity, and events are discrete. How do you compute the Fourier spectrum then? Is it just SST?

Response: We thank the reviewer for this comment. We clarify that the Fourier spectrum is not computed from individual MHW events. While cumulative intensity is defined at the event level, we construct a continuous time series by aggregating the cumulative intensity of all MHW events occurring within each month. MHWCI is set to zero when no marine heatwaves occur during a given month. This results in a monthly MHWCI time series, which is then used for the spectral analysis shown in Fig. 1b. The analysis is therefore based on monthly MHWCI, not on SST directly. Clarification will be provided in the revision.

l102 The data record starts in 1940. Are you using 1982-2021 as for the SSTs? Please clarify

Response: The satellite-based OSTIA SST product used to identify MHW is available only from 1982 onwards. Although the atmospheric datasets (e.g. ERA5) extend back to 1940, our analysis is restricted to the common period 1982-2021, consistent with the SST record.

l104 A reference and some more details, both on ERA5 and ORAS5, should be given

Response: We thank the reviewer for this comment. We will revise the manuscript to include appropriate references and additional details for both ERA5 and ORAS5 in the Data section, including a brief description of the datasets and their temporal and spatial resolutions.

l113 MHWCI definition ain't very clear to me. What happens say if an event starts in Jan 27 and ends on Feb 10? Which month gets the CI? Please specify also the baseline period used.

Response: Thank you for this comment. We constructed monthly MHWCI by aggregating event cumulative intensity within each calendar month. Specifically, monthly MHWCI is obtained by summing the cumulative intensity of all events whose onset date falls in that month. Therefore, an event starting on 27 January and ending on 10 February is assigned to January in our monthly aggregation. We additionally tested the sensitivity of the k-means classification to the event-to-month assignment by assigning events to their termination month instead of their initiation month. The resulting spatial clustering remains very similar to the original classification. The main difference is a slight seasonal shift in the relative contributions within each cluster: in Cluster 1, which is primarily winter-dominated, the contribution from spring becomes somewhat larger, while in Cluster 2, which is primarily summer-dominated, the contribution from autumn increases. Importantly, however, the dominant seasons associated with MHW occurrence remain winter and summer, respectively. Therefore, this sensitivity does not affect the main physical interpretation or conclusions of the study.

Furthermore, the MHW threshold (90th percentile) is computed from the detrended SST time series using the baseline period 1982-2021. Explanations will be provided in the revision.

l154 state the source of this version of the heat tendency equation. Information on the number of levels in ORAS5 should be added, to clarify which is the accuracy of the MLD estimated as such eq3 I imagine these equation is also calculated with monthly mean data?

Response: Thank you for the comment. The mixed-layer heat tendency equation (Eq. 3) follows the standard formulation commonly used in previous studies of upper-ocean heat budgets (Liu et al., 2014; Tan et al., 2016). We will add the relevant references in the manuscript.

ORAS5 provides ocean variables on 75 vertical levels (level spacing increasing from 1 m at the surface to 200 m in the deep ocean), allowing a physically consistent estimation of mixed-layer depth and associated heat budget terms at seasonal to interannual timescales. All heat budget terms are computed using monthly mean data. We will clarify these points in the Methods section.

l189 so you apply K-means clustering over the whole year or individual seasons?

Response: The K-means clustering is not applied separately to individual seasons, nor directly to the raw monthly time series. Instead, it is performed using seasonal variability characteristics derived from the full monthly record. Specifically, for each grid point, we compute the interannual variability (standard deviation) of MHWCI anomalies separately for DJF, MAM, JJA, and SON (with DJF treated across calendar years). These four seasonal variability measures form a feature vector that is used as input to the K-means clustering. Therefore, the clustering is based on information from all seasons, while explicitly accounting for their distinct variability characteristics.

Fig. 2 typos in titles, and no units in Fig c and d; how are time series normalized?

Response: We thank the reviewer for pointing this out.
(1) We will correct the typos in the titles of Fig. 2c-d.
(2) We will clarify in the figure labels and captions that the time series shown in Fig. 2c and d are standardized MHWCI, which are therefore dimensionless.
(3) The time series are normalized using z-score standardization, defined as

$$z(t) = \frac{x(t) - \mu}{\sigma},$$

where μ and σ denote the mean and standard deviation of the original time series, respectively. This procedure results in dimensionless time series with zero mean and unit variance.

l247 anomalies from what? Is this is stratified according to some indices? How?

Response: We thank the reviewer for this comment. The geopotential height anomalies shown in Fig. 4 are defined relative to the long-term monthly climatology. They are obtained using composite analysis during periods of enhanced MHWCI, defined as months when the standardized cluster-specific MHWCI time series exceeds +1 standard deviation. We will clarify the anomaly reference state and the composite criteria in the revised manuscript.

l266 with units m^2/s^2, this is geopotential (not height)

Response: We thank the reviewer for pointing this out. We agree that the variable shown has units of $m^2 s^{-2}$ and therefore represents geopotential rather than geopotential height. We will revise the text and figure labels throughout the manuscript accordingly to ensure consistent terminology.

l269 why this value? Why not using some significance threshold?

Response: We thank the reviewer for this comment. The composite anomalies shown are tested for statistical significance at the 90% confidence level. For the wave activity flux vectors, in addition to significance testing, we apply a display threshold of 0.1 $m^2 s^{-2}$ in the figure to emphasize dynamically meaningful flux patterns. Similar display thresholds have been used in previous studies of atmospheric wave activity flux diagnostics (Hou et al., 2023). Explanation will be provided in the revision.

Fig 7 cannot be understood. Why are arrows colored? They are hard to see. What are the black arrows on the right side? Missing labels on axes

Response: We agree that Fig. 7 currently contains too many elements, which make it difficult to interpret. In the revised version, we will simplify the presentation of the vector fields by representing current speed through arrow length rather than color, while using a uniform color for all vectors. In addition, we will increase arrow size and spacing to further enhance readability. The black arrows on the right-hand side indicate the direction of the meridional overturning circulation. We agree that this was not sufficiently clear in the current version and will clarify this explicitly in both the caption and the main text. In addition, we will add missing axis labels and ensure that all plotted elements are clearly defined.

l277 is this some sort on average on some sub-domain?

Response: The contributions shown in Fig. 6 are computed as spatial averages over the Cluster 1 region during periods of enhanced MHWCI. We will clarify this in the revised manuscript.

l318 "telecommunications"?

Response: The term “telecommunications” was an error and will be corrected to “teleconnections” in the revised manuscript.

l340 "predominated"

Response: We thank the reviewer for pointing this out. The term “predominated” will be replaced with “dominated” in the revised manuscript.

l342 and otherwise? Some climatological maps should be presented.

Response: We understand that the concern refers to the description of mixed layer depth (MLD), where only regional mean values are reported, and more generally to the representation of stratification in the manuscript. In the revised manuscript, we will include climatological maps of MLD and discuss stratification together with MLD, linking this discussion to the heat budget analysis and to the contrasting summer and winter mechanisms identified in the two clusters.

Fig. 9 why do arrows look quite different? And what is the box in the maps?

Response: The differences in the WAF vectors among the three panels reflect physically distinct wave propagation patterns associated with different AMV phases and background climate states, as all fields are consistently regressed onto the IPO index. We will clarify this point in the figure caption. In addition, the rectangular box shown in Fig. 9 indicates the North Sea region, which is the focus of the present study. This will be stated in the caption.

l400 is it because some versions of the AMO index are influenced by trends?

Response: We thank the reviewer for this insightful question. We agree that some versions of the AMO index can be influenced by long-term trends. However, the AMO index used in Mohamed et al. (2023) is based on the NOAA index definition in which the North Atlantic SST time series is detrended and thus represents multidecadal internal variability rather than the long-term warming signal.

In our study, we detect marine heatwaves using detrended SST. Therefore, the relationships discussed here are unlikely to arise from shared long-term trends but instead reflect the influence of low-frequency Atlantic variability and associated atmosphere-ocean processes.

l425 why the West Med now?

Response: We agree that the reference to the western Mediterranean is not essential in this context and may appear abrupt. We will remove this part of the sentence in the revised manuscript to improve the focus and clarity of the discussion.

l662 incomplete citation, also l681, l700...

Response: We thank the reviewer for pointing this out. We will carefully check the entire reference list and correct all incomplete citations, including those at lines 662, 681, and 700.

Fig. B1 what are the units of the ordinate?

Response: The ordinate in Fig. B1 represents the within-cluster sum of squared correlation distances (WCSS), which is a dimensionless clustering metric. We will clarify this in the figure caption.

Reference used in the response

Hou, J., Fang, Z., & Geng, X. (2023). Recent Strengthening of the ENSO Influence on the Early Winter East Atlantic Pattern. *Atmosphere*, *14*(12), 1809. https://www.mdpi.com/2073-4433/14/12/1809

Liu, Q. Y., Wang, D., Wang, X., Shu, Y., Xie, Q., Chen, J., Liu, Q. Y., Wang, D., Wang, X., & Shu, Y. (2014). Thermal variations in the South China Sea associated with the eastern and central Pacific El Nino events and their mechanisms. *Journal of Geophysical Research Oceans*, *119*(12), 8955-8972.

Tan, W., Wang, X., Wang, W., Wang, C., & Zuo, J. (2016). Different Responses of Sea Surface Temperature in the South China Sea to Various El Niño Events during Boreal Autumn. *Journal of Climate*, *29*(3), 1127-1142.