We thank the reviewers for their time and very constructive suggestions, which helped us to further improve the manuscript. Below, we present the reviewers' comments, our responses, and the corresponding revisions to the manuscript. The reviewers' comments are in black, our responses and the previous version of the manuscript are in blue, and the changes in the manuscript are in purple.

Reviewer #1

L81-86: I would suggest moving this addition ('Previous work suggests…' onwards) to earlier in the paragraph, right after the sentence citing Bulter and Polvani (2011) on L77, for improved readability.

Response: We shifted the last few lines after the sentence citing Bulter and Polvani (2011).

Previous version: L75-86: In addition, tropical Pacific SST variability influences the Arctic polar vortex and Northern Hemisphere circulation in a highly nonlinear way, with both El Niño and La Niña winters exhibiting enhanced SSW frequency relative to neutral conditions (Butler and Polvani, 2011). ENSO teleconnections to the North Atlantic–European region also evolve seasonally: early-winter impacts are mainly tropospheric and project more strongly onto the East Atlantic Pattern (Thornton et al., 2022), while the stratospheric pathway becomes increasingly important later in winter, favouring a positive NAO projection (Ineson and Scaife, 2009). This seasonal evolution and multiplicity of pathways make it more difficult to identify the remote SST precursor of extreme winds in the North Sea, as the linkage is likely to occur through several pathways and be strongly nonlinear. Previous work suggests that the apparent enhancement of SSW frequency during both El Niño and La Niña phases in reanalyses such as ERA5 may be influenced by the limited length of the observational record (Polvani et al., 2017). Using large hindcast ensembles, (Ineson et al., 2024) show that this behaviour can arise by chance, and that with a larger sample size, a more linear ENSO–polar vortex relationship may emerge. This would be more consistent with the established impact of ENSO on boreal winter stratospheric circulation.

New version: L75-86: In addition, tropical Pacific SST variability influences the Arctic polar vortex and Northern Hemisphere circulation in a highly nonlinear way, with both El Niño and La Niña winters exhibiting enhanced SSW frequency relative to neutral conditions (Butler and Polvani, 2011). Previous work suggests that the apparent enhancement of SSW frequency during both El Niño and La Niña phases in reanalyses such as ERA5 may be influenced by the limited length of the observational record (Polvani et al., 2017). Using large hindcast ensembles, Ineson et al. (2024 show that this behaviour can arise by chance, and that with a larger sample size, a more linear ENSO–polar vortex relationship may emerge. This would be more consistent with the established impact of ENSO on boreal winter stratospheric circulation. ENSO teleconnections to the North Atlantic–European region also evolve seasonally: early-winter impacts are mainly tropospheric and project more strongly onto the East Atlantic Pattern (Thornton et al., 2022), while the stratospheric pathway becomes increasingly important later in winter, favouring a positive NAO signal (Ineson and Scaife, 2009). This seasonal evolution and multiplicity of pathways make it more difficult to identify the remote SST precursor of extreme winds in the North Sea, as the linkage is likely to occur through several pathways and be strongly nonlinear.

L128-135: It is possibly worth noting I think that the zonally symmetric component of the stratosphere circulation ought to be less relevant for the three SNAPSI cases looked at in Hitchcock et al. (2022), in the context of studying the effects of strengthened polar vortex conditions. That is because weak/SSW events can have greatly contrasting vortex geometries/zonal asymmetries.
That's not to dispute the authors' inference based on their results, but a strong, stable polar vortex does tend to be more zonally symmetric. So, it may be worth the authors revisiting/revising the text here.

Response: We changed this paragraph. Reviewer #2 mentioned (**L133-134:**) that imposing zonally symmetric changes in the polar stratosphere necessarily results in zonally symmetric impacts on the troposphere, which is not always the case, as tropospheric responses can be highly asymmetric due to stationary waves and regional dynamics. That is why we removed that sentence from the manuscript.

Previous version: L131-135: "The SNAPSI simulations are limited to three polar vortex anomaly cases. Later in our study, we argue that, on average, the zonally symmetric component of the polar vortex is unlikely to be responsible for North Sea storminess, as the statistical spatial coherence length of seasonal storminess is highly regional, restricted to the North Sea. A general impact of the zonally symmetric component of the polar vortex would imply coherent circumpolar storminess. This may, however, not be the case for the three cases investigated in the SNAPSI simulations."

New version: L127-133: "The SNAPSI simulations are limited to three polar vortex anomaly cases. While the zonally symmetric component may be particularly relevant for these cases, the limited sample size hampers a broader assessment of its role. Later in our study, we argue that, on average, the zonally symmetric component of the polar vortex is unlikely to be responsible for North Sea storminess, as the statistical spatial coherence length of seasonal storminess is highly regional, restricted to the North Sea. This lack of circumpolar spatial coherence suggests that the frequency of North Sea extremes is modulated by regional dynamics or asymmetric vortex configurations, such as displacement vs split events, rather than solely by the average intensity of the zonally symmetric circumpolar jet."

Fig. 1c: Does it really make sense to show this correlation map for the whole globe? Only the NH mid to high latitudes are mentioned and the area of significance, in connection with North Sea storminess, would appear only to be confined to the North Atlantic region.
Therefore, it is perhaps worth the authors refining the plot to just the North Atlantic sector in the final version.

Response: We replaced the previous figure (Fig. 1c) with this new figure of only the North Atlantic section. Reviewer #2 suggested that we should do a statistical significance test for Fig. 1b, 1c and 3, so the figure was changed due to the comment (**Figure 3**).

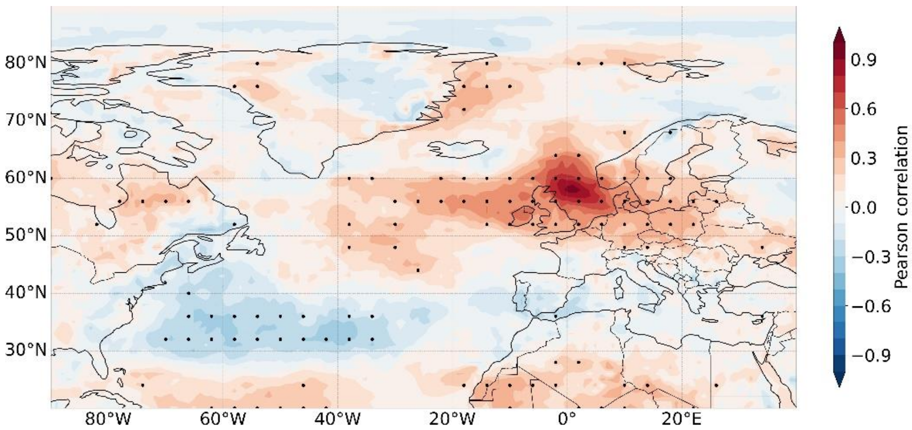

Figure 1. b) Correlation between January storm counts (1940–2024) at each North Sea grid point and the storm count time series at the reference location (1°E, 58°N: Northern North Sea); c) Same as panel b, but extended to all grid cells in the North Atlantic. In Figs. 1b and 1c, we calculated the statistical significance using a Monte Carlo permutation test (100 iterations). Correlations exceeding the 95% confidence level ($p < 0.05$) were identified as statistically significant and marked with black dots.

Previous version: L165-170: "In addition, this figure shows a similar correlation map but extended to all other ERA5 grid cells (Fig. 1c). This third panel indicates that the spatial coherence scale of storminess is regionally limited. Only a region in the western Atlantic is weakly negatively correlated with North Sea storm counts, whereas other mid and high-latitude regions around the polar vortex are not correlated with the North Sea. This leads to the assumption that the possible predictors of North
Sea storminess should be regional in scale, discounting the role that remote large-scale sea-surface temperature patterns could play on average."

New version: L163-168: "In addition, this figure shows a similar correlation map but extended to all other ERA5 grid cells in the Northern Atlantic (Fig. 1c). This third panel indicates that the spatial coherence scale of storminess is regionally limited. Only a region in the western Atlantic is weakly negatively correlated with North Sea storm counts, whereas other mid and high-latitude regions around the polar vortex are not correlated with the North Sea. This leads to the assumption that the possible predictors of North Sea storminess should be regional in scale, discounting the role that remote large-scale sea-surface temperature patterns could play on average."

L243: '…corresponding ensemble of simulations…' → As only one simulation is performed for control and perturbed individually, should the word 'ensemble' be changed?
Or perhaps at least change to '…corresponding ensemble of years…', so it's clear that initial conditions weren't tweaked and multiple forecast experiments were generated for a given season.

Response: We modified this paragraph to omit the word "ensemble".

Previous version: L241-244: "These identified patterns of average anomalies in the Decembers preceding Januaries with the highest frequency of North Sea stormy days were used to modify the initial conditions in an ensemble of simulations with the atmospheric emulator ACE2, starting on December 1st and running for two months. The output was then compared with a corresponding ensemble of simulations started with observed initial conditions on December 1st. The results are described in the next sections."

New version: L250-253: "These identified patterns of average anomalies in the Decembers preceding Januaries with the highest frequency of North Sea stormy days were used to modify the initial conditions in simulations with the atmospheric emulator ACE2, starting on December 1st and running for two months. The output was then compared with corresponding years, starting with observed initial conditions on December 1st. The results are described in the next sections."

Fig. 8: Although I think its good the authors show the signal for November initialisation start dates (Fig. 8); to illustrate the lack of signal between an artificially strengthened polar vortex and North Sea wind speed increase 7-60 days into the forecast, this result is not central to the main story in my view. I strongly suggest the authors move it to the Appendix for the final version.

Response: We moved the signals for the November initialisation dates (Fig. 8) to the appendix.

Fig. 9: Maybe the authors could add quantification of the difference in 90th percentile of the PDF distributions, perhaps as a percentage value (perturbed – control/control)*100, due to the strengthened polar vortex?
Also of consideration might be calculation of risk ratios of the upper tail end shift (e.g., using Eq. 4 in Zhang and Wang, 2019)? It appears that for some years, the probability of days exceeding control 90th percentile might perhaps be doubled or even trebled, so calculating such metric would provide a clear, quantitative indication of the risk enhancement in surface extreme winds.

Response: We added the difference at the 90th percentile of the PDF distributions as a percentage in Fig. 9 and added a bit more detail in the caption. We also added the mentioned Risk Ratio to Table 2.

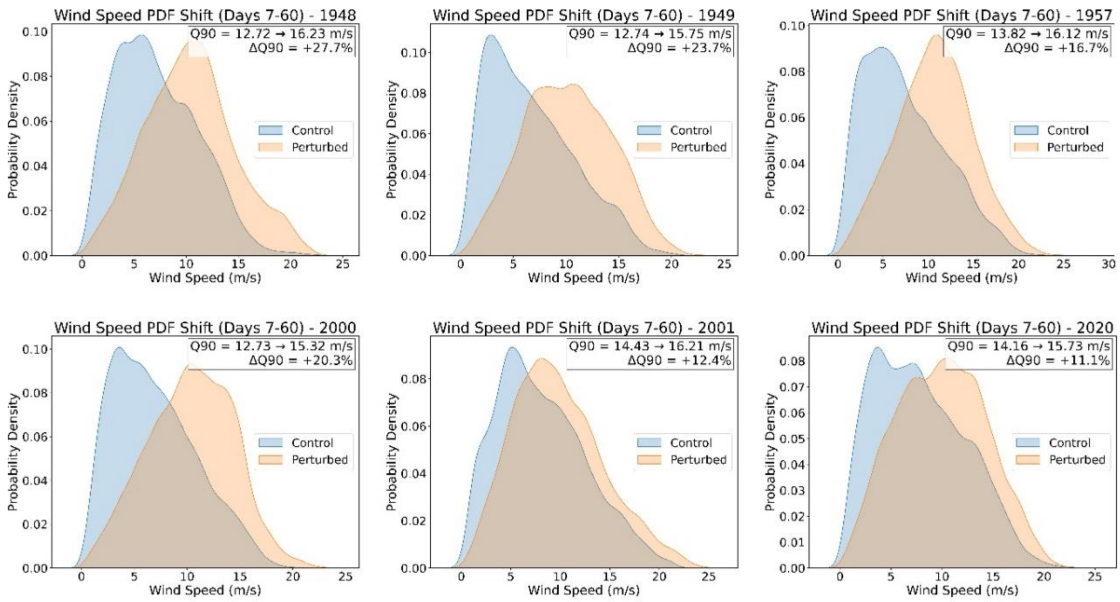

Figure 9. PDFs of daily wind speed over the North Sea (days 7–60) for control and perturbed simulations, aggregated across multiple grid cells (Longitude: 4°W to 10°E, Latitude: 51°N to 61°N). The perturbed experiment consistently shows a rightward shift and a more populated upper tail. The figure also includes quantifications of the difference at the 90th percentile of the PDF distributions as a percentage.

Year	Event Risk Ratio (RR1)	Control Q90 Risk Ratio (RR2)
1948	4.09	2.78
1949	3.77	2.85
1957	4.15	2.27
2000	4.08	3.00
2001	3.71	1.60
2020	4.04	1.80

Table 2. Comparison of risk ratios for 90th percentile wind speed events. The first column uses the climatological 90th percentile threshold (*Event Risk Ratio*), whereas the second column uses the 90th percentile threshold derived from the control simulation (*Control Q90 Risk Ratio*. The fixed-threshold risk ratio is defined as $RR1 = P_{\text{pert}}/P_{\text{ctrl}}$, where $P_{\text{ctrl}} = 0.10$ corresponds to the exceedance probability of the control 90th percentile threshold. The control-Q90 risk ratio is calculated as $RR2 = P_{\text{pert}}(X > Q90_{\text{ctrl}})/P_{\text{ctrl}}(X > Q90_{\text{ctrl}})$) (Zhang and Wang, 2019).

Previous version: L374-377: "Table 2 summarises the regional mean changes, showing an increase in extreme event frequency across all six analysed years, with values ranging from approximately 3% to 4%. Together, the PDF and exceedance analyses demonstrate that a strengthened polar vortex leads to both higher wind speeds overall and a systematic increase in the likelihood of extreme surface wind events over the North Sea."

New version: L383-390: "Table 2 summarises the changes in the likelihood of 90th percentile wind speed events over the North Sea. Both risk ratio metrics indicate a systematic increase in extreme wind events under a strengthened polar vortex. The Event Risk Ratio (RR1), based on a fixed climatological 90th percentile threshold, ranges from 3.71 to 4.15, indicating that extreme wind speed events become approximately four times more likely in the perturbed simulations. The Control Q90 Risk Ratio (RR2), computed using the 90th percentile threshold derived from each control simulation, ranges from 1.60 to 3.00 and likewise demonstrates a substantial increase in exceedance probability. Together, the PDF and exceedance analyses demonstrate that a strengthened polar vortex leads to both higher wind speeds overall and a systematic increase in the likelihood of extreme surface wind events over the North Sea."

Table 2: I suggest moving this until after Fig. 10, to avoid any confusion by the reader that the analysis pertains to the new Fig. 9 analysis.

Response: Thanks to the Reviewer for pointing it out. Table 2 has been shifted after Fig. 10.

Fig. 11: Would it make more sense to display the years in chronological order? Or better yet perhaps, order them as a function of increasing Storm Index?

Response: We changed the x-axis to chronological order.

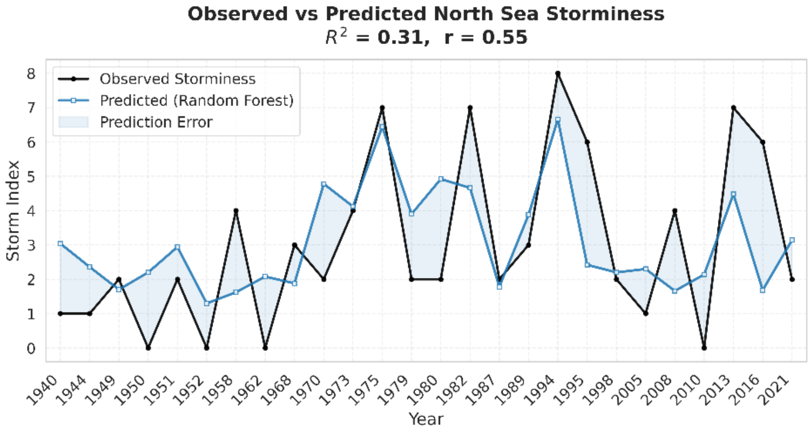

Figure 11. Comparison between observed and Random Forest predicted North Sea storminess (1940–2024). Time series of observed (black-test set) and predicted (blue) storm indices.

Table 3: The R2 values still do not appear to correspond to the r values shown (e.g., Nov-Dec is 0.12 to 0.1 for R2 but 0.31 to 0.52 for r?). Also, R2 by definition can only ever be positive! Requires fixing.

Response: We would like to clarify that the R^2values in Table 3 are not calculated as the square of the correlation coefficient (r^2). While $R^2 = r^2$for a simple linear regression under specific conditions, the R^2metric used in our study is the coefficient of determination,

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

where y_i are the observed values, $\hat{y}_i$ are the predicted values, and $\bar{y}$ is the mean of the observations. R^2 > 0 indicates that the model explains a fraction of the variance in the target variable and performs better than a prediction based solely on the mean observation. R^2 = 1 corresponds to a perfect prediction, while R^2 = 0 indicates performance equivalent to using the observed mean as a predictor. R^2 < 1 occurs when the residual sum of squares exceeds the total sum of squares, implying that the model performs worse than the mean-value predictor.
Therefore, R^2 is not constrained to be positive and is not necessarily related to the squared correlation coefficient. As a result, the values of R^2 and r shown in Table 3 are not expected to satisfy the relationship $R^2 = r^2$.
To avoid confusion, we have revised the manuscript to explicitly state the definition of R^2, explain the interpretation of positive and negative values, and include the corresponding formula in the table caption.

Previous version: "**Table 3.** Summary of the Random Forest model performance metrics for different predictor–predictand month pairs. The first month is the predictor month, and the second month is the predictand month. The correlations are calculated between predictions and target in a reserved independent data set not used for the training of the RF methodology, in a similar fashion as for the December-January case."

New version: "**Table 3.** Summary of the Random Forest model performance metrics for different predictor-predictand month pairs. The first month is the predictor month, and the second month is the predictand month. The correlations are calculated between predictions and the target in a reserved independent data set not used for the training of the RF methodology, in a similar fashion to the December–January case. The coefficient of determination is computed as

$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$,

where y_i are the observed values, $\hat{y}_i$ are the predicted values, and $\bar{y}$ is the mean of the observations. R^2 > 0 indicates that the model explains a fraction of the variance in the target variable and performs better than a prediction based solely on the mean observation. R^2 = 1 corresponds to a perfect prediction, while R^2 = 0 indicates performance equivalent to using the observed mean as a predictor. R^2 < 1 occurs when the residual sum of squares exceeds the total sum of squares, implying that the model performs worse than the mean-value predictor.

L403: '…when the polar vortex is weak, the atmosphere becomes less active and stormier…' → Worth double checking this is correct. From the abstract in Zhang et al. (2024), the suggestion is the opposite as 'synoptic-scale eddies tend

to be suppressed' following weak polar vortex events.
Unless the authors mean to say, 'less stormy' and it is perhaps just poorly worded?

Response: We changed the word "stormier" to "less stormy". Thanks to the Reviewer for pointing it out.

Technical Corrections
L174: '…grid near Hull…' → '…grid cell near Hull…'
L175: '…The second grid is chosen…' → '…The second grid cell is chosen…'
L186: '…basis grid cell.' → '…basin grid cell.'?

Table 1 Caption: 'None of these values is statistically significant.' → 'None of these values are statistically significant.'

Response: All of them have been corrected. In L186, we meant to say "basis grid cell," but we understood that the word might not have been the best choice. We replaced the word "basis grid cell" with "chosen grid cell".

Reviewer #2

L23: The authors use the phrase "normal westward zonal flow." Please clarify if you mean "typical westerlies" (which flow eastward and I think it is what the authors mean) or if you are referring to an actual westward (easterly) flow, as the current phrasing is confusing.

Response: We think that the Reviewer meant in L63. We replaced the sentence.

Previous version: L63-64: "The normal westward zonal flow of the lower stratospheric winds within the polar vortex can be disrupted by Sudden Stratospheric Warming (SSW) events."

New version: L63-64: "The typical westerlies in the lower stratosphere within the polar vortex can be disrupted by Sudden Stratospheric Warming (SSW) events."

L79: I suggest removing the word "positive" when referring to the "NAO projection." The literature demonstrates that different ENSO phases can lead to different NAO phases. Using a more neutral term like "NAO response" or "NAO signal" would be more accurate.

Response: We replaced the sentence.

Previous version: L77-80: "ENSO teleconnections to the North Atlantic–European region also evolve seasonally: early-winter impacts are mainly tropospheric and project more strongly onto the East Atlantic Pattern (Thornton et al., 2022), while the stratospheric pathway becomes increasingly important later in winter, favouring a positive NAO projection (Ineson and Scaife, 2009)."

New version: L81-84: "ENSO teleconnections to the North Atlantic–European region also evolve seasonally: early-winter impacts are mainly tropospheric and project more strongly onto the East Atlantic Pattern (Thornton et al., 2022), while the stratospheric pathway becomes increasingly important later in winter, favouring a positive NAO signal (Ineson and Scaife, 2009)."

L81-86: I suggest moving these new sentences so they appear before the discussion on the seasonal evolution of the ENSO teleconnection to the Euro-Atlantic sector.
This would maintain a more logical narrative flow, keeping the general discussion about SSW occurrence during different ENSO phases together before diving into intraseasonal variations of the ENSO teleconnection.

Response: Reviewer #1 also suggested that. We rectified the paragraph (**L81-86:**).

L105-107: I recommend removing this sentence. The specific definitions of stormy days and seasons would fit better in Section 2. This detailed information is not necessary for understanding the subsequent text regarding the approaches used to explore the predictability of stormy winter seasons in the North Sea.

Response: We removed the sentence.

Previous version: L105-107: "In our study, we employ two approaches to explore the predictability of stormy winter seasons in the North Sea. We define stormy days as days in which the daily mean wind exceeds a high percentile of the climatological probability distribution. Stormy seasons are those with a higher frequency of stormy days (see section 2). One prediction methodology used here is Random Forest, a machine learning method…"

New version: L104-105: "In our study, stormy seasons are those with a higher frequency of stormy days (see section 2). One prediction methodology used here is Random Forest, a machine learning method…"

L124-126: As noted in my previous review, the citation of Mouallem et al. (2025) remains unclear. That study focuses primarily on the generation of SSWs within an idealized model, rather than the downward propagation of the stratospheric signal to the troposphere. As currently written, the text implies that in that study stratospheric circulation changes are impossed to assess their tropospheric impacts. Please revise this section to accurately reflect their methodology and findings.

Response: We thank the Reviewer for pointing it out**.** We revised the manuscript to remove the misleading description of Mouallem et al. (2025).

Previous version: L124-128: "This approach follows a similar concept to a recent idealised modelling study by Mouallem et al. (2025), in which controlled changes are applied to the stratospheric circulation to isolate how such perturbations affect the troposphere, without attempting to reproduce the full evolution of SSW events. To trigger SSWs, their model is perturbed using an idealised, moving topographic forcing that generates vertically propagating planetary waves."

New version: L121-124: "This approach follows a similar concept of utilising a minimalistic modelling framework, as in a recent study by Mouallem et al. (2025). While our study imposes stratospheric initial conditions to assess their downward influence, Mouallem et al. (2025) focused on the upward initiation of SSW events."

L133-134: Do the authors mean to imply that imposing zonally symmetric changes in the polar stratosphere necessarily results in zonally symmetric impacts on the troposphere? This is not always the case, as tropospheric responses can be highly asymmetric due to stationary waves and regional dynamics. Please clarify or qualify this statement.

Response: Reviewer #1 also had issues with this statement. We made changes to this part to satisfy both of the reviewers.

Previous version: L132-135: "The SNAPSI simulations are limited to three polar vortex anomaly cases. Later in our study, we argue that, on average, the zonally symmetric component of the polar vortex is unlikely to be responsible for North Sea storminess, as the statistical spatial coherence length of seasonal storminess is highly regional, restricted to the North Sea. A general impact of the zonally symmetric component of the polar vortex would imply coherent circumpolar storminess. This may, however, not be the case for the three cases investigated in the SNAPSI simulations."

New version: L127-133: "The SNAPSI simulations are limited to three polar vortex anomaly cases. While the zonally symmetric component may be particularly relevant for these cases, the limited sample size hampers a broader assessment of its role. Later in our study, we argue that, on average, the zonally symmetric component of the polar vortex is unlikely to be responsible for North Sea storminess, as the statistical spatial coherence length of seasonal storminess is highly regional, restricted to the North Sea. This lack of circumpolar spatial coherence suggests that the frequency of North Sea extremes is modulated by regional dynamics or asymmetric vortex configurations, such as displacement vs split events, rather than solely by the average intensity of the zonally symmetric circumpolar jet."

L152: Please include meridional wind at 70hPa as it is used later (e.g. Figure 6).

Response: We added meridional wind at 70 hPa.

L171: Please add Figure before “1a”.
Response: We added it.

L194: There is a discrepancy between the text and Figure 3. The manuscript mentions that three scenarios were analyzed, including Jan-Feb, but only two are plotted. Please update Figure 3 to include the Jan-Feb analysis, or provide a justification in the text for its omission.

Response: We mistakenly added the Jan-Feb scenario in the text. We thank the Reviewer for catching the error. We removed the Jan-Feb scenario from the text.

Previous version: L193-194: “We also examined the correlation between global SST and storminess in the North Sea across three scenarios: January–January, December–January, and January–February (Fig. 3).”

New version: L194-195: “We also examined the correlation between global SST and storminess in the North Sea across two scenarios: January–January and December–January (Fig. 3).”

Figure 3: Please add information about statistical significance of results. Same in Figure 1.

Response: We calculated the statistical significance using a Monte Carlo permutation test (100 iterations), in which the storm time series for one grid cell in the North Sea was randomly shuffled while the other variables (i.e., SST and the global number of storms) were held fixed. Correlations exceeding the 95% confidence level ($p < 0.05$) were identified as statistically significant and marked with dots.
In response to Reviewer #1 (**Fig. 1c:**), Figure 1c was revised to focus on the North Atlantic instead of the global domain. The updated figure also incorporates the suggestions from Reviewer #2. Therefore, the previous version of Figure 1c has been replaced with a revised version that addresses both reviewers’ comments.

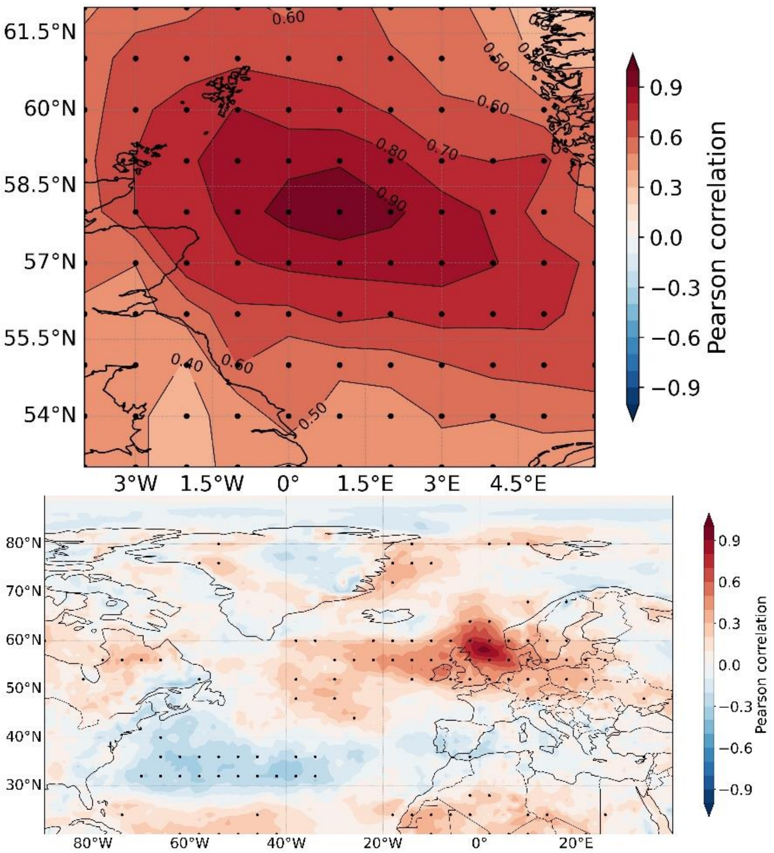

Figure 1. b) Correlation between January storm counts (1940–2024) at each North Sea grid point and the storm count time series at the reference location (1°E, 58°N: Northern North Sea); c) Same as panel b, but extended to all grid cells

in the North Atlantic. In Fig. 1b and 1c, we calculated the statistical significance using a Monte Carlo permutation test (100 iterations). Correlations exceeding the 95% confidence level ($p < 0.05$) were identified as statistically significant and marked with black dots.

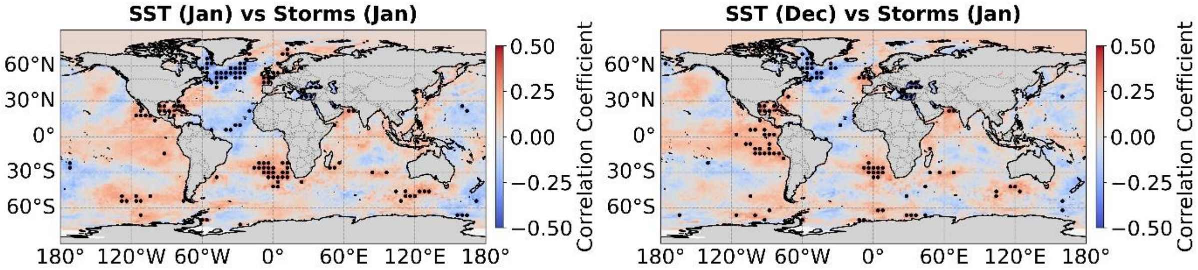

Figure 3. Global correlation maps between sea surface temperature (SST) and North Sea storminess for two cases: simultaneous January–January and 1-month lagged December–January. Statistical significance was assessed using a Monte Carlo permutation test (100 iterations). Correlations exceeding the 95% confidence level ($p < 0.05$) are considered statistically significant and are indicated by black dots.

L233-235: While I acknowledge the authors' previous response regarding the use of meridional wind for the analysis with ACE2, I still strongly encourage showing geopotential height in this subsection. Geopotential height maps are much more effective than meridional or zonal wind plots for providing a clear idea of the polar vortex structure preceding stormy or calm Januaries.

Response: We added the geopotential height at 200 hPa composite maps for three of the chosen grid cells.

Newly added figures in the results and appendix:

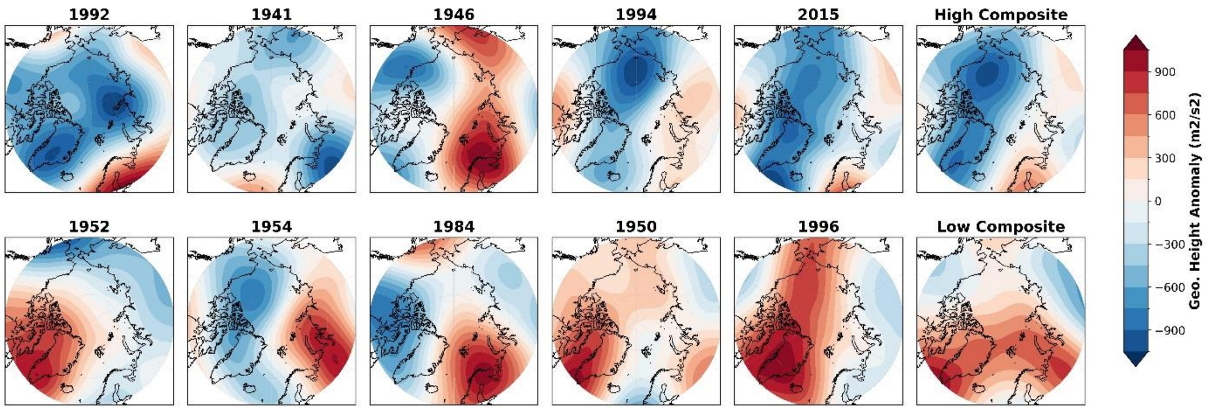

Figure 7. Geopotential height anomalies at 200 hPa in the Decembers preceding Januaries characterised by a high (upper row) and zero (lower row) number of stormy days in the Northern North Sea grid cell

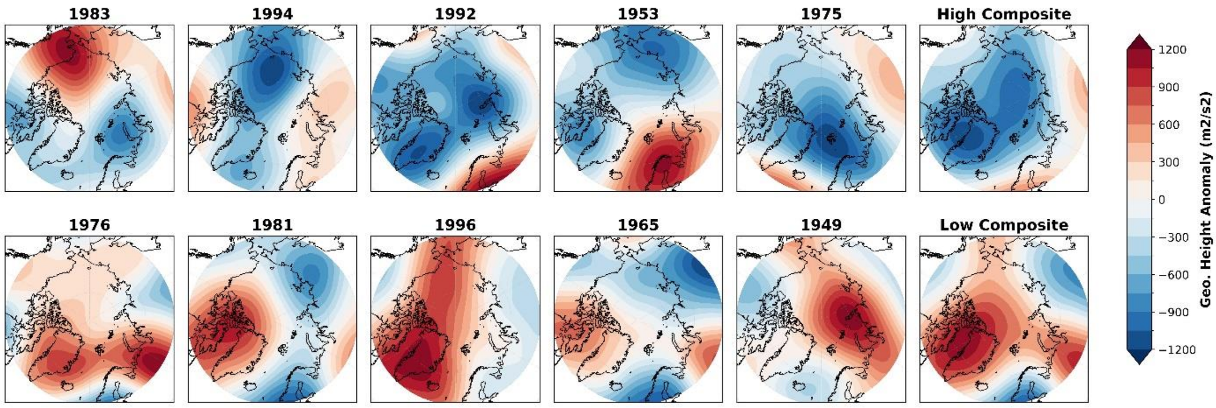

Figure A7. Geopotential height anomalies at 200 hPa in the Decembers preceding Januaries characterised by a high (upper row) and zero (lower row) number of stormy days in the UK Coast grid cell

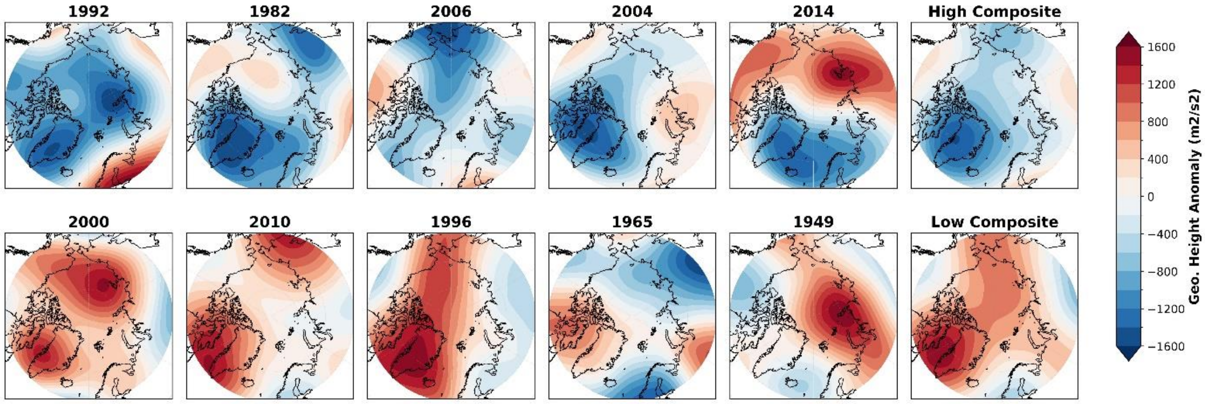

Figure A8. Geopotential height anomalies at 200 hPa in the Decembers preceding Januaries characterised by a high (upper row) and zero (lower row) number of stormy days in the German coast grid cell

New paragraph added for the geopotential height at 200 hPa:
L238-245: "A similar composite analysis was done for geopotential height anomalies at 200 hPa during the Decembers preceding the five stormiest and five calmest Januaries. The results show a clear contrast between the two groups (Fig. 7). The stormiest Januaries are generally preceded by negative geopotential height anomalies over the Arctic and northern North Atlantic, whereas the calmest Januaries are associated with positive anomalies in the same regions. That means that the stormy Januaries are preceded by lower-than-normal geopotential heights, indicative of a stronger polar vortex and enhanced large-scale circulation. In contrast, calm Januaries are associated with higher geopotential heights, suggesting a weaker polar vortex and more blocked circulation patterns. These results indicate that upper-tropospheric circulation anomalies in December may serve as a useful precursor to January storm activity in the North Sea."

L395: Please update the description of the new figure 14 layout.
Response: The description has been updated.

Previous version: L394-395: "To interpret these physically, the PCs' loading maps are essential (Fig. 14). They reveal coherent Arctic-wide structures in geopotential height at 200 hPa (top row), zonal wind (middle row) and temperature at 70 hPa (bottom row)."

New version: L407-408: "To interpret these physically, the PCs' loading maps are essential (Fig. 14). They reveal coherent Arctic-wide structures in temperature at 70 hPa (first column), zonal wind at 70 hPa (middle column) and geopotential height at 200 hPa (third column)."

L422: Please replace "formation" with "seasonal cycle".
Response: We replaced it.

We added a paragraph in the Methods section to explain why we used Pearson's correlation rather than another method that might be more suitable for a discrete dataset, such as the storm count dataset.

Newly added paragraph in Methods section L175-177: "Throughout the paper, we used Pearson's correlation instead of the Martingale difference or Spearman's correlation, despite the discrete nature of the storm count dataset. As we use correlation analysis only for illustrative purposes to show a rough link to the number of storms in January, we chose Pearson's correlation because it is widely used and readily interpretable."
Newly added reference:
Zhang, J. and Wang, F.: Changes in the Risk of Extreme Climate Events over East Asia at Different Global Warming Levels, Water, 11,
https://doi.org/https://doi.org/10.3390/w11122535, 2019