Response to Reviewer Comment RC1

We sincerely thank Reviewer #1 for their constructive and insightful comments on our manuscript *CMIP6 Multi-model Assessment of Northeast Atlantic and German Bight Storm Activity*. The comments greatly helped us to improve the manuscript and clarify key points.

We respectfully acknowledge the reviewer's concern regarding the scope of the journal. However, we would like to point out that the editor, having considered the submission in light of the journal's aims and scope, deemed it suitable for peer review. We believe this reflects the editor's judgment that the manuscript aligns with the journal's thematic focus, which has recently also included studies of regional significance (e.g., https://esd.copernicus.org/articles/special_issue1088.html). In this context, we trust that decisions regarding scope remain within the editorial purview, while the peer review process can focus on the scientific quality, clarity, and contribution of the work.

In the following, we will give a point-by-point response to the reviewer's comments and describe how we plan to address the issues raised.

Main comments:

1 My main concern is that as presented, the study appears rather incremental. There are many studies examining model projections of North Atlantic storminess (as you summarise in your introduction), and your key conclusion of an overall reduction in storminess in future model projections but with an increase in the intensity of the most extreme storms, has been noted numerous times before. Please tweak the framing of your work to address this concern (particularly in the introduction) to better inform the reader exactly how this study aims to advance current understanding. Formulating one or two explicit research questions might help with this.

Response: We thank the reviewer for this comment. We agree that the framing of the manuscript can be improved to better convey the novelty of our approach. We will revise the introduction by formulating explicit research questions and emphasize our twofold novelty: (1) the combined use of CMIP6 multi-model-ensemble output and the 50-member MPI-GE to disentangle externally forced signals from internal variability, and (2) the inclusion of the pressure-based storm activity proxy. The number of CMIP6 models used in previous studies on storm activity changes has been limited, primarily due to the unavailability of key diagnostic variables across all models. To overcome this constraint, we apply the pressure-based proxy introduced by Schmidt and von Storch (1993), which enables the inclusion of a larger ensemble of 32 CMIP6 models. This broader model set allows for a more comprehensive assessment of projected changes and uncertainties in Northeast Atlantic storm activity under various anthropogenic forcing scenarios, as well as a direct comparison with observed pressure-based storm activity. These additions will clarify the contribution of our work to the ongoing debate on storm activity projections and their uncertainties.

2 To my mind, one key advance is the comparison of storminess between the climate models and the long-term dataset of direct observations, because the vast majority of climate model studies just compare against reanalysis products. However, this comparison is not mentioned in the abstract, and even a basic description of the observational dataset is omitted from the manuscript. I'd urge you to make more of this aspect in the text, and to extend the observational comparison to all relevant figures (e.g. 3, 4, 6, 7, 8) if possible.

Response: We thank the reviewer for highlighting this important aspect. We agree that this comparison deserves more emphasis. We will revise the abstract to explicitly mention the observational dataset. In the main text, we will introduce and describe the observational data in Section 2 and extend the comparison to all relevant figures, including Figs. 3, 4, 6, 7, and 8, where possible.

3 Your storminess diagnostics are annual in the sense that you don't subset the data to a particular season. However, I imagine most of the >95%ile geostrophic wind events happen in autumn/winter and so your projected future changes represent most closely the changes in these seasons. Given projected future

changes in storminess contain important seasonal variations, please add a discussion on this point to aid interpretation.

Response: We thank the reviewer for raising this point. We will add a discussion paragraph to Section 5 that clarifies the seasonality of the geostrophic storm activity metrics. We will state that while we use annual metrics for comparability and simplicity, the most extreme geostrophic wind events are indeed most frequent during late autumn, winter, and early spring. Earlier studies have shown that the geostrophic wind statistics of those seasons indeed closely resemble the annual statistic (e.g. Krieger et al., 2021). We will also add that the lack of seasonal disaggregation may mask more nuanced seasonal shifts in storm activity under climate change. Also, many studies do not focus on seasonal shifts or aspects, so that analyzing annual values improves comparability with other literature.

Other comments:

Abstract The last two sentences appear contradictory because you state "the upper percentiles of winds speeds from these directions decrease" and then "the most extreme storms may become stronger or more likely". I think the former is referring to the 95th percentile of the wind speeds whereas the latter is referring to more extreme percentiles. Please clarify.

Response: We thank the reviewer for bringing this inconsistency to our attention. Indeed, the former is referring to the 95th percentile of the wind speeds per wind direction, and the latter to the more extreme percentiles. We will rephrase the abstract to remove this confusing wording.

L25 and the following paragraphs Please be explicit about the seasonality of the projected changes in storminess presented in these papers. Some I know explicitly refer to winter only, and others I am not sure about.

Response: We will revise the introduction to specify which studies refer to seasonal/winter-only projections and which focus on annual statistics to provide better context for the reader.

L56 "upper wind speed percentiles" is unclear (I thought it meant upper-tropospheric wind speeds initially). Please clarify, e.g. "upper percentiles of near-surface wind speeds". Similar comment applies to L58.

Response: We apologize for the misleading wording and see the issue with the term "upper wind speed percentiles". We will adapt the wording as suggested.

L91 Is CMIP6 psl data daily means or instantaneous?

Response: The daily MSLP data consists of daily means, not instantaneous values. We will add this information to the data description.

L97 Just to be clear, do you standardise the annual 95th percentiles for each triangle separately, or average them together and then standardise?

Response: We standardize each triangle separately, and then average over the standardized timeseries of all triangles. We repeat this step for every member in the ensemble. We will add a clarifying statement to the respective section.

L115 I presume that the gradients are computed using the distances between the model grid points (which differ for each model), rather than the original station locations? Please specify.

Response: That is correct, all gradients are computed based on the locations of the respective model gridpoints. We will clarify this part to avoid confusion.

L133 The observed timeseries has not been introduced. Please add a description of it in section 2.

Response: We will add a description of the storm activity observations in the Methods and Data section, specifying the data sources, spatial coverage, time range, and method used for storm activity estimation.

L146 You claim that "the full pool of ensemble members can represent the variability present in the observations", but this is misleading and clearly must depend on the timescale examined. If I understand correctly, all the timeseries are independently standardised, so the interannual variability is by construction captured by the ensemble, at least during the period 1960-1990. What you show are ten year running means, so I assume your claim is something like "the full pool of ensemble members can represent the variability on decadal timescales". Please clarify.

Response: The reviewer is correct in their assumption. By showing the standard deviation of the full pooled ensemble and the observed storm activity as 10 year running means, we demonstrate that the decadal variability of observed storm activity is contained within the multi-model ensemble. While this is correct by construction for the reference period of 1961-1990, it also holds for periods after and before the reference period. We will rewrite this section to avoid further misconceptions and make clear that this figure does not show the interannual variability, but rather the variability on decadal and longer timescales.

Section 4 Several recent papers have highlighted deficiencies in the ability of climate models to simulate multi-decadal variability in the North Atlantic, and have questioned the reliability of model projections in this region as a result (e.g. see here, and references therein: Smith et al., 2025, https://doi.org/10.1038/s41558-025-02277-2). Given their importance, I'd urge you to extend your discussion to include reference to them and relate to the findings of your study.

Response: We thank the reviewer for bringing up the recent publication by Smith et al. The study is highly relevant to our discussion and we will thus expand our discussion section to reflect more on the points raised by Smith et al. (2025) and related papers.

L225 "increase" -> "increase in frequency"

Response: We will update our wording here.

L290 This paper presents a statistical methodology for assessing future changes from multi-model ensembles of differing sizes, which is very relevant to your suggestion: Zappa et al. (2013) A multimodel assessment of future projections of North Atlantic and European extratropical cyclones in the CMIP5 climate models. Journal of Climate, 26(16), pp.5846-5862.

Response: We appreciate the reviewer for pointing us to the study by Zappa et al. (2013). While we refer to this study in the introduction, we absolutely see the need to mention it in this paragraph as well. We will update this part of the discussion to include this study.

Fig 3 caption Please state the periods over which trends are computed (I assume the full experiment periods, but best to be precise).

Response: Yes, the trends are computed across the entire experiment length. We will add that to the figure caption.

Fig 4 caption Are daily geostrophic wind directions? Please clarify.

Response: Yes, the wind directions in Fig. 4 are daily means, as they are based on daily-mean MSLP input data. The wind directions in Fig. 6 are three-hourly. We will add the frequencies to the figure captions.

Fig 7 caption Repeated "the"

Response: We thank the reviewer for spotting this mistake. We will correct it in the revised version.

Fig 8 To what extent are the differences here statistically robust? Can you construct confidence intervals? (here and/or Fig 9)

Response: We thank the reviewer for bringing up the issue of missing statistical significance checks. In accordance with comments by Reviewer #2, we will add estimates of robustness to the manuscript and figures wherever necessary.

We, the authors, would like to thank Reviewer #1 again for their careful reading of our manuscript and for the constructive comments. We hope that our responses and proposed revisions clarified all outstanding points and look forward to further feedback.

With kind regards,

Daniel Krieger and Ralf Weisse