Response to Referee Comment 3: “Concurrent modes of climate variability linked to spatially compounding wind and precipitation extremes in the Northern Hemisphere”

Comments to the authors

Overview:

This paper focuses on compound wind and precipitation (CWP) extremes, aiming to identify the drivers behind the occurrence of these events in the Northern Hemisphere. Climate model simulations from the Community Earth System Model are used with reanalysis data (ERA5) providing a “sense check”. A few key climate variable modes are considered (ENSO, AMV, NAO & PNA). The individual effects of these events are found to follow existing literature, e.g. NAO+ increasing CWP extremes in Northern Europe. Concurrent phases of variability modes are considered with specific regional effects discussed. The NAO- & ENSO+ combination increased the likelihood of CWP extremes in eight regions. This motivated exploring spatially compounding extremes, where a positive trend between the number of anomalous variability modes and the number of regions was identified. Physical mechanisms for the statistical relationships were then discussed. This paper concludes ENSO is the most influential mode of variability for CWP extremes in the Northern Hemisphere.

Compound events are an area of current interest and this manuscript will appeal to the community. It is suitable for this NHESS special issue and I therefore recommend its publication subject to the changes outlined below. I would therefore appreciate the author’s response on the comments below.

Response:

We would like to thank the referee for their positive comments and detailed feedback. All the comments and our point-by-point responses are given below.

Comments:

General comments:

As this study covers a large region and many combinations of variability modes, the presentation of results is important. The paper has a wide scope which at times means detail on specific regions is lacking. Choosing two or three regions or one teleconnection index to focus on gives this study more impact.

Response:

Thanks for this comment. We agree that the presentation of results is important, particularly for a broad study like ours that covers many regions and four variability modes and their combinations. The primary aim of this study is to explore the broad patterns of spatially compounding CWP extremes across the Northern Hemisphere. We understand the referee’s point that choosing a reduced number of regions and teleconnection indexes would allow us to provide valuable insights for the specific modes and regions selected. However, we think that, by not choosing to focus on specific regions or teleconnection index, our study can provide more broad information. In particular, by providing the most significant results for multiple regions and combinations of modes, we envision the study providing a broad overview of the influence of combinations of modes on CWP extremes. In this direction, we hope this overview will guide and motivate future, more targeted investigations into regions or modes of particular interest.

Comments:
While the standard of written English is fine, the language used makes this paper difficult to read at times. There are some very long sentences which could be split up or multiple sentences which may be more readable as a bullet pointed list.

Response:
Thanks for this comment. We recognize the importance of clear and concise language in scientific communication, particularly for a study with a broad scope and technical complexity. We worked on the paper to improve readability. While we aimed for a formal and detailed style, we split up long sentences to improve readability.

Comments:
Redrafting Section 3 will make the paper more readable and therefore accessible to the wider scientific community. Figures are meant to help convey information simply, Figures 3, 4 & 5 are complex. The authors should only include combinations of variability modes discussed in the text with the full figures available in the supplementary material.

Response:
We thank the reviewer for pointing this out. While accessibility and understandable figures are of great importance to us, we feel the message Figures 3, 4 and 5 convey is quite nuanced. To simplify them further, we would need to compromise on the clarity and preciseness of our message, which we feel would be detrimental for the paper. Although we agree simplified figures may more easily provide a first level of information to the reader, the current figures provide a crucial overview of the relationships between variability modes and CWP extremes and associated spatially compounding extremes, which is essential for understanding the global patterns we aim to highlight. Including all combinations allows readers to evaluate the relationships beyond the specific examples discussed in the text and, for example, distill those that are relevant to their region of interest. This broader context is also critical for readers interested in the full spectrum of variability mode interactions. For these arguments, we have decided to keep the figures in their current form. However, to ease their interpretation, we have improved and refined the text in Section 3 relating to the mentioned figures to better guide readers through the analyses, emphasizing the most critical patterns and combinations while leaving room for individual exploration of the full dataset. We also emphasized in the text that a selection of the various mode combinations is presented:

L258 of the article originally submitted: “**In the following, we focus on describing the effects of a selection of mode combinations and regions in Figs. 3-5. To maintain clarity and conciseness, we do not discuss all regions and mode combinations in the text, and readers can explore specific regional effects directly in the figures.**”

Comments:
The choices of percentile thresholds are arbitrary. The results of this study would hold more weight if a sensitivity analysis on these had been conducted. e.g. 98th percentile of daily precipitation seems low as this data is zero inflated.

Response:

Thanks for this comment. Although using different variables (that is, wind gusts instead of wind speed), some studies considered the local 98th percentile to investigate precipitation and wind extremes (Martius et al., 2016). Also, Klawa and Ulbrich (2003) show that the local 98th wind percentile is a damage-relevant wind threshold for wind gusts. In our study, daily data for the December-January-February months are used. It represents 90 days by season. By choosing the 98th percentile as a threshold, the expected number of exceedances per season for wind and precipitation in isolation is equal to 90*0.02 ≈ 2 events per season, which can be considered sufficient to analyse co-occurrences of wind and precipitation values above these thresholds. Choosing a percentile higher than the 98th would allow us to focus on more extreme events (e.g., Zhang et al., 2011), but it would reduce the sample size for the analyses. Note that for model evaluation, we use the 95th percentiles, to ensure a sufficiently large sample size given the shorter record length of reanalysis data. Although CWP events exceeding the 98th percentile of wind and precipitation can be considered moderate extremes, we think that they can still be considered impact-relevant. Choosing a local and impact-relevant threshold for wind and precipitation extremes is difficult, especially for precipitation for which incorporating effects such as surface runoff, snow melt and landslides would be needed (e.g., Williams, 1978) and is thus out of the scope of this study.

To expand on these trade-off issues between sufficiently large sample sizes and sufficiently extreme events, we suggest adding the following explanations (in blue) to the text:

L110 of the article initially submitted: “We use the 98th percentile of wind and precipitation over the 1950–2019 period for the main analysis based on data from the CESM model. **Percentile-based thresholds are frequently used to investigate climate extremes (e.g., Zhang et al., 2011, Martius et al., 2016). Following Klawa and Ulbrich (2003) and Martius et al (2016), we chose the 98th percentile, which is a compromise to capture the most extreme events in the CESM simulations while ensuring a sufficiently large sample size for robust statistical analysis.** For model evaluation, **which involves both the CESM model and ERA5 reanalyses (Figs. S1-S5 of the Supplement only)**, we use the 95th percentiles over the 1950-2019 period -- such a lower threshold allows for a more robust evaluation. **The reason for this is that,** given the ERA5’s limited period, **extremes in the reanalysis data set are more scarce and associated statistics for very extreme events are largely affected by sampling uncertainty (Bevacqua et al., 2021b). Selecting a slightly lower threshold allows us to reduce this sampling uncertainty and thus improve confidence in assessing the model's ability to simulate extremes (e.g., Bevacqua et al., 2021b, Kelder et al., 2022, Fischer et al., 2023)**.”

References:

- Zhang, X. B., L. Alexander, G. C. Hegerl, P. Jones, A. K. Tank, T. C. Peterson, B. Trewin, and F. W. Zwiers (2011), Indices for monitoring changes in extremes based on daily temperature and precipitation data, *Wires Clim. Change*, 2, 851–870.
- Klawa, M., and U. Ulbrich (2003), A model for the estimation of storm losses and the identification of severe winter storms in Germany, *Nat. Hazard Earth Syst. Sci*, 3, 725–732.

- Williams, G. P. (1978), Bank-full discharge of rivers, *Water Resour. Res.*, 14(6), 1141–1154, doi:10.1029/WR014i006p01141.

Following the suggestion of the referee, we produced a sensitivity analysis by considering the 99th and 99.5th percentile to define seasonal counts of CWP extremes. New Figures S16-S19 have been added to the Supplement and display the results we obtained for:

- Metric 1: the influence of individual and concurrent variability modes on regional wintertime CWP frequency (Figs. S16 and S17, same results as those presented in Fig. 3 but consider the 99th and the 99.5th percentile, respectively).

- Metrics 2 and 3: the influence of variability modes on spatially CWP extremes (Figs. S18 and S19, same results as those presented in Fig. 6 but consider the 99th and the 99.5th percentile, respectively).

Regarding Metric 1, by increasing the percentile, some differences can be observed (Figs. 3, S16, and S17). Although the effect of the combinations on CWP extremes remains generally consistent in magnitude across percentiles (not shown), increasing the threshold generally limits the test procedure to identifying significant combinations. Significant effects were detected in 20 regions when using the 98th percentile (Fig. 3). Increasing the percentile to the 99th (Fig. S16) and 99.5th (Fig. S17) led to detect significant effects in 17 and 11 of these regions, respectively. Such a systematic reduction in the number of regions when considering higher thresholds aligns with the fact that higher thresholds lead to more seasons without CWP events, making it more difficult to detect a significant signal for Metric 1. Still, the results obtained for the three different thresholds (98th, 99th, and 99.5th percentiles) are fairly consistent. In particular, (1) despite increasing the threshold generally limits the test procedure identifying all combinations that were significant at lower thresholds, the combinations detected at higher thresholds are consistently included among those identified at lower thresholds. Note that, in line with what was stated above, the differences might be due to increased sampling uncertainty associated with higher thresholds rather than differences in the involved physical mechanisms. Furthermore, (2) the magnitude of the effects of the combinations that were detected as significant at lower thresholds but not at higher thresholds are generally consistent across thresholds (not shown).

Regarding Metrics 2 and 3, we also observe some differences depending on the threshold (Figs. 6, S18 and S19). However, the main conclusions of our study are not changed: for the 99th and 99.5th percentiles, combinations of variability modes have a significant effect on the total number of affected regions (Figs. S18a and S19a), along with an amplified effect relative to their underlying mode sub-combinations, with ENSO+ being the predominant mode phase (see "+" sign). For the population affected, the influence of variability modes is primarily driven by ENSO- (see the '−' sign in Figs. S18b and S19b), consistent with the findings of the main study based on the 98th percentile (see Fig. 6b).

We added some sentences in the Discussion:

“We analysed event counts aggregated over winter and at the scale of predefined SREX regions, given that high counts of compound extremes at these scales are expected to have negative effects on society. **While the 98th percentile has been used in this study to**

focus on extremes and is relatively well-established in the literature (e.g., Klawa et al., 2003, Martius et al., 2016), other higher thresholds could have been chosen to consider more intense extreme events (e.g., Liu et al., 2013, Schar et al., 2016, Camuffo et al., 2020). Figs. S16-S19 show results from a sensitivity analysis on the influence of variability modes on regional CWP extremes (Metric 1) and spatially compounding events (Metrics 2 and 3) with the 99th and 99.5th percentiles used as thresholds. Although there are some variations in the results compared to those for the 98th percentile, the main conclusions drawn across the different thresholds are broadly consistent for all Metrics. The magnitude of the effects of the combinations are generally consistent across thresholds, and the combinations detected at higher thresholds are generally included among those identified at lower thresholds (Figs. S16-S19). Such slight differences may be due to larger sampling uncertainty for higher thresholds limiting the ability to detect significant effects for higher thresholds rather than different physical mechanisms involved for different thresholds. While the sensitivity analyses broadly indicate the robustness of most of our findings, possible relevant differences across thresholds highlight the importance of identifying impact-relevant thresholds, though this task is challenging (Williams, 1978, Bloomfield et al., 2023). In addition, the selected SREX regions may not reflect the natural spatial patterns of variation of CWP extremes, potentially occurring at a more localized scale or span across multiple regions.”

Comments:

Daily precipitation is not always proportional to any resulting impact – the authors should acknowledge the complexity of the precipitation-flood relationship. For more on this see Bloomfield et al. (2023) [https://doi.org/10.1016/j.wace.2023.100550]. While compound wind-precipitation events cause large impacts, they are rare (e.g. Fig. 2 from Jones et al. (2024) [https://doi.org/10.1002/wea.4573]). Considering these extremes in isolation gives the complete picture of a compound hazard.

Response:

Thanks for this comment. We agree that the relationship between daily precipitation and its impacts, particularly flooding, is highly complex and not necessarily proportional. This complexity arises from numerous factors, such as antecedent soil moisture conditions, land use, and drainage capacity, which influence the translation of precipitation into flooding. We will acknowledge this in the Discussion section and cite Bloomfield et al. (2023) to provide additional context.

Regarding the rarity of compound wind-precipitation events, as noted by Jones et al. (2024), we agree that their infrequent nature does not diminish their potential for significant societal and environmental impacts. Our focus on these events aims to understand the drivers and spatial relationships of compound wind-precipitation extremes. While considering precipitation and wind extremes in isolation may offer valuable insights, our study aims to focus on compound events explicitly.

We added some sentences in the Discussion:

“**While the sensitivity analyses broadly indicate the robustness of most of our findings, possible relevant differences across thresholds highlight the importance of**

identifying impact-relevant thresholds, though this task is challenging (Williams, 1978, Bloomfield et al., 2023)."

Comments:
You have cited Manning et al. (2024) to highlight extratropical cyclones as drivers of CWP events, but Manning et al. (2024) notes CWP events can be driven by precipitation extremes.

Response:
Thanks for this comment. Indeed, Manning states that the expected increase in precipitation due to the influence of climate change will make compound wind and precipitation extremes more likely, and that they will be produced by extratropical cyclones. We really don't see a contradiction here, but we understand the importance of your statement. This paper (from Owen et al., https://www.sciencedirect.com/science/article/pii/S2212094721000384#sec4) is more adequate to support our point, therefore we added this reference to support the statement.

Comments:
Specific comments:
L2: Change "agricultural crops" to "crops"

Response:
Thanks for this comment. We changed the text accordingly.

Comments:
L6: Remove NAO & PNA abbreviations, they are not used in rest of abstract.

Response:
Thanks for this comment. However, for consistency, we kept all the abbreviations in the Abstract, which are then used in the rest of the study.

Comments:
L13: Remove "For example" here, the reader knows you're giving them an example.

Response:
Thanks for this comment. However, we think that "for example" is important in the structure of the sentence.

Comments:
L17-22: Split into two sentences and rejig. Define compound events first, then highlight their importance from this IPCC report.

Response:
Thanks for this comment. We split the sentence into two as follows: "**Compound weather and climate events, defined as the combination of multiple drivers and/or hazards that contribute to societal or environmental risk, often cause more severe impacts than the respective single hazards (Zscheischler et al., 2018). The Intergovernmental Panel on Climate Change (IPCC) Special Report on Managing the Risks of Extreme Events**

and Disasters to Advance Climate Change Adaptation (SREX) highlighted the importance of studying compound events to improve modeling and risk estimation of weather impacts (IPCC, 2012)".

Comments:
L51: Useful to describe what the deviation from mean NAO conditions is, how does it affect frequency & intensity of events?

Response:
Thanks for this comment. We changed the text (in blue) as follows:

L51: "During extreme phases of the PNA and NAO, the intensity and location of storms and moisture transport deviate from mean conditions over the Pacific-North American region (e.g., Wallace et al, 1981, Xie et al., 2020) and the Euro-Atlantic region (e.g., Hurrell et al., 2003, Lodise et al., 2022), respectively. **While positive NAO phases intensify westerly winds and shift the North Atlantic storm track toward the northeast, leading to increased storm frequency and intensity over Northern Europe, negative NAO phases weaken the westerlies and amplify storm activity in the Mediterranean region (e.g., Hurrell and Deser, 2010).**"

Comments:
L74: Specify which months the winter season covers.

Response:
Thanks for this comment. We now precise in the Abstract and Introduction which months the winter season covers in our study.

Comments:
L75: Change "effective" to "influential"

Response:
This sentence is not part of the Manuscript after incorporating the changes from the other reviewers.

Comments:
L84: Make the rationale behind the choice of these regions clearer. These shapes cut across country boundaries, making this study less applicable to the insurance industry.

Response:
Thanks for this comment. We modified the text as it follows (in blue):

"We examine the influence of four variability modes on CWP extremes across 25 selected regions in the Northern Hemisphere defined in the SREX (Iturbide et al., 2020, see Fig. 1). **We chose these regions as they are standard reference in IPCC reports, as they encompass areas with relatively homogeneous climatic characteristics (Iturbide et al., 2020). While using these regions does not enable an explicit analysis of dependencies between local-scale CWP extremes and modes of variability, it allows for complementing IPCC assessments.**"

Comments:
L96: Why did you choose to begin with 1959? ERA5 covers from 1940 so matching the same period as CESM makes sense.

Response:
Thanks for this comment. When we started the project in 2022, ERA5 1940-1958 was not available. As a result, it was not possible to incorporate this data during the first steps of our project. While ERA5 data for 1940-1958 is now available, its quality and reliability for this period remain questionable due to sparse observational input, as acknowledged by the Copernicus Climate Change Service (https://www.ecmwf.int/en/newsletter/175/news/era5-reanalysis-now-available-1940). For this reason, we opt not to use it, as it would potentially compromise the robustness of our analysis and it would require an important computation effort.

Comments:
L96: "Singh et al. (2021)" reference doesn't make sense here? As far as I can tell, Singh et al. (2021) doesn't use ERA5?

Response:
Thanks for this comment. We removed the reference.

Comments:
L110: The 95th percentile of daily data considers 1114 days in this period (1959-2019) to be extreme. Yet the 98th percentile over 1950-2019 only considers 511 extreme days. Surely a higher threshold of ERA5 data is required for these periods to be comparable?

Response:
We want to thank the reviewer for this comment, as some clarifications are needed. In the study, we consider two percentile-based thresholds to determine CWP extremes: 98th percentile of wind and precipitation for the main analysis (for CESM) to ensure a sufficiently large sample size of simulated events while assessing extremes sufficiently extreme, and the 95th percentile for model evaluation only (both for CESM and ERA5; Figs. S1-S5 of the Supplement) to ensure a sufficiently large sample size in the reanalysis data which is only one realization over a shorter record. Therefore, thresholds are identical when comparing occurrences of CWP extremes in CESM simulations and ERA5 data. We suggest to provide the following clarifications (in blue) to the text:

L110 of the article initially submitted: "We use the 98th percentile of wind and precipitation over the 1950–2019 period for the main analysis based on data from the CESM model. **Percentile-based thresholds are frequently used to investigate climate extremes (e.g., Zhang et al., 2011, Martius et al., 2016). Following Klawa and Ulbrich (2003) and Martius et al (2016), we chose the 98th percentile, which is a compromise to capture the most extreme events in the CESM simulations while ensuring a sufficiently large sample size for robust statistical analysis.** For model evaluation, **which involves both the CESM model and ERA5 reanalyses (Figs. S1-S5 of the Supplement only)**, we use the 95th percentiles over the 1950-2019 period -- such a lower threshold allows for a more robust evaluation. **The reason for this is that,** given the ERA5's limited period, **extremes in the reanalysis data set are more scarce and associated statistics for very extreme events are largely affected by sampling uncertainty (Bevacqua et al., 2021b). Selecting a slightly lower threshold allows us to reduce this sampling uncertainty and thus improve confidence in assessing the model's ability to simulate extremes (e.g., Bevacqua et al., 2021b, Kelder et al., 2022, Fischer et al., 2023)**."

Comments:
L115: Include rationale for weighting by cosine of latitude.

Response:
Thanks for this comment. The weighting by the cosine of latitude is applied to account for the spherical geometry of the Earth. Without this correction, grid cells closer to the poles, which cover smaller physical areas, would be overrepresented in the analysis compared to those near the equator, which cover larger areas. This approach ensures that regional averages are spatially representative, reflecting the actual physical extent of each grid cell.

By applying this weighting, we maintain consistency with standard practices in climate and atmospheric sciences, ensuring that the metrics derived are not artificially biased by the unequal spatial resolution inherent to a latitude-longitude grid system. This correction is particularly important in studies like ours that involve large-scale regional analyses across diverse latitudes.

We included the rationale as follows:

L115: Wintertime CWP counts are averaged by region over landmasses, weighted by the cosine of latitude **to prevent overrepresentation of grid cells closer to the poles.**

Comments:
L151: Change "That is, in this study, we do not…" to "This study does not".

Response:
Thanks for this comment. We changed the text accordingly.

Comments:
L154: Remove ", in principle,"

Response:
We changed the text accordingly.

Comments:
L162: The 280 year return period seems to be an arbitrary choice. Sensitivity analysis on this threshold would be of interest.

Response:
Thanks for this comment. The choice on the return period is required to ensure that the combinations of variability modes analyzed had a sufficiently large sample size for robust statistical assessment while focusing on relatively rare, impactful events. It was decided to

consider samples of minimum size 10 years, which given the yearly resolution of the aggregated data implies to consider combinations occurring more than 10 years in our 2800-year dataset. Given the length of our 2800-year dataset, this implies exploring mode combinations with a maximum return period of 280 years. Changing a different return period as a threshold would only change the combinations of modes displayed in the figure, but importantly, the effects presented for the combinations illustrated in the submitted paper would not change.

Comments:
L176: Mismatched bracket after “subsection 2.2.3”.

Response:
Thanks for this comment. The mismatched bracket was deleted.

Comments:
L180: A 10% significance level seems high, 5% (or even 1%) level is much more standard practice.

Response:
Thanks for this comment. The choice of a 10% significance level was intentional to balance the detection of meaningful effects while avoiding false negatives. With a limited sample size (which is the case when we compare distributions in the study), a lower threshold might be too stringent to detect meaningful effects, especially at a 5 or 1% significance level. Choosing a larger significance level is aligned with the exploratory nature of our work, allowing us to shed light on potential effects of modes on CWP extremes.

We suggest to add the following sentence:
L180 of the article initially submitted: “Specifically, for a given CWP metric, we test whether the ratio of the average of the metric associated with a given set of phases of interest (e.g., NAO+ENSO-, set as the numerator) to the average of the metric under neutral conditions (set a denominator) is larger than one at significance level α = 0.10 based on one-sided tests. **Compared to a lower significance level, our chosen level allows the detection of significant effects of modes of variability while reducing false negatives in the context of small sample sizes.**”

Comments:
L185: How many times is “several times”? State this in the text.

Response:
Thanks for this comment. Here “several times” is used to describe the concept behind permutation testing, therefore we did not deem it necessary to give the exact number of permutations. Depending on the case, we used a different number of permutations: (m=100,000 for the analysis of the three metrics; m=100 when applied to the grid cell level for Figs. 2 and 7). We will change the text in this way to avoid the confusion.

L185: “By repeating this procedure ~~several times~~, we can then define a confidence interval for the ratio and a critical region for test rejection.”

Comments:
L199-200: Change 100.000 to 100,000

Response:
Thanks for this comment. We changed the text accordingly.

Comments:
L224: A significant body of literature exists linking extreme windstorms to strong winds (favourable conditions for CWP events). Here I would at least cite:
- Mailier et al. (2006) https://doi.org/10.1175/MWR3160.1
- Priestley et al. (2024): https://doi.org/10.5194/nhess-23-3845-2023

Response:
Thanks for this comment. We no longer discuss the findings in the Results section. References are thus not needed here. The reference for Priestley et al., 2024 has been however added to the Introduction.

Comments:
L312: I'd make this sentence clearer, "generally covers most of the time" is very ambiguous.

Response:
We agree with the reviewer. We suggest deleting "most of the time" in the text.

Comments:
L391: Change "Europa" to "Europe".

Response:
Thanks for this comment. We changed it.

Comments:
L430: Is this not driven by atmospheric circulation patterns?

Response:
Thanks for this comment. While we agree with the referee, this sentence has been removed from the text.

Comments:
L480: Change "found" to "estimated"

Response:
Thanks for this comment. We suggest changing the text as follows (in blue):

L479-480: "By repeating this procedure among different modes, we ~~found~~ estimated a wide range of return periods for the different mode combinations (Fig. 3)."

Comments:
L484: A natural next step would be repeating this study for the southern hemisphere.

Response:

Thanks for this comment. The current study focuses on the Northern Hemisphere due to its dense population and economic significance. Indeed, we acknowledge that extending this study to the Southern Hemisphere would be a valuable next step, considering the relevant variability modes for that Hemisphere and meteorological season. We now mention it in the Discussion.