

Response to both referees for NHESS-2020-345

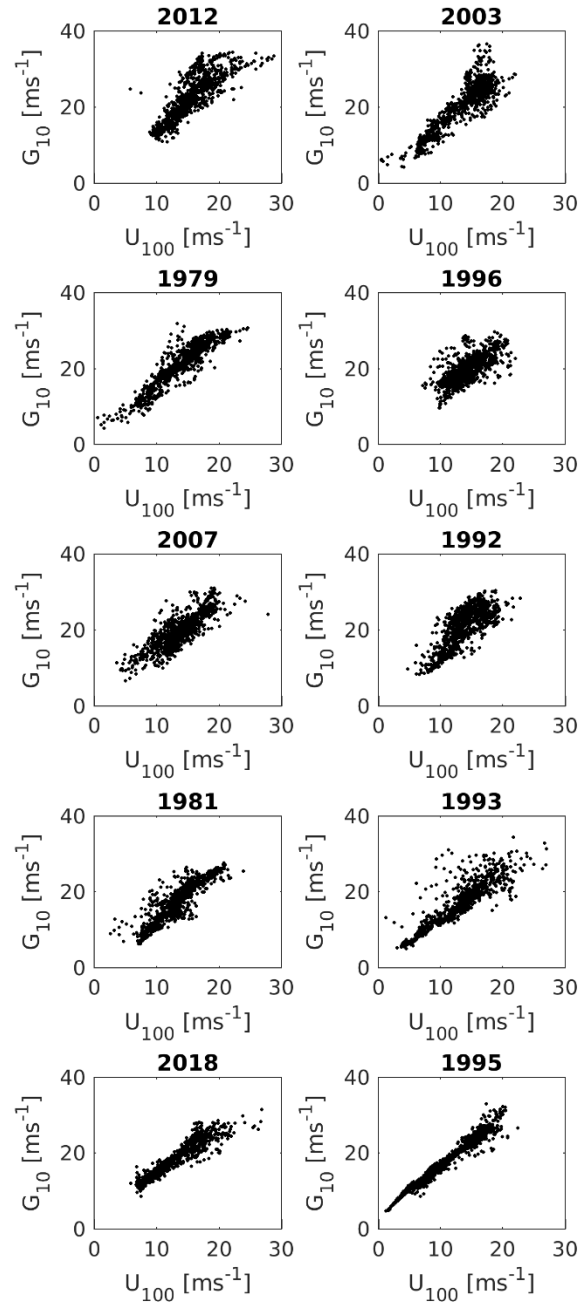
Intense windstorms in the Northeastern United States

Thank you both for your thoughtful comments on our work. Before going on to address your specific comments, we would like to address our use of (i) a wind speed threshold of the local 99.9th percentile (rather than 98%) and (ii) 100-m once-hourly sustained wind speeds rather than 10-m gusts. Regarding point (i) our research is seeking to identify truly exceptional windstorms. Use of a 98th percentile value is not very discriminatory (see details below). Regarding point (ii) while ERA5 like many reanalysis systems does generate wind gust estimates, when derived at a grid resolution of approximately 30 by 30 km they are systematically negatively biased relative to observations. Further 10-m wind speeds are strongly dictated by the roughness lengths used in the reanalysis models. Thus, we feel it is appropriate to use variables that are likely to be more presentative and simulated with higher fidelity. We hope this clarifies our perspectives and elaborates on materials within the submitted document that discuss the latter point. Quoting from the original submission; ‘However, it is important to acknowledge that wind parameters from any model do not fully reflect all scales of flow variability (Skamarock, 2004) and underestimate extreme wind speeds (Larsén et al., 2012), particularly in areas with high orographic complexity and/or varying surface roughness length. Here we use wind speeds at 100-m because flow at this height is less likely to be impacted by sub-grid scale heterogeneity in surface roughness length and uncertainties induced by unresolved sub-grid scale variability. Near-surface wind speeds are strongly coupled to wind speeds at 100-m (i.e., within the PBL) but wind speeds at 100-m are less strongly impacted by inaccuracies and/or uncertainty in surface roughness length (z_0) (Minola et al., 2020; Nelli et al., 2020). Applying an uncertainty of a factor of two to z_0 can lead to mean differences of up to 0.75 ms^{-1} for near-surface (40 to 150 m a.g.l.) wind speeds (Dörenkämper et al., 2020).’

In response to your comments regarding this matter, we have added new analysis of 10-m gusts and a comparison of our storm-ranking algorithm to one using a 98th percentile threshold. Related changes are as follows;

- 1) In section 2.1 we describe the data; ‘Wind gusts at a nominal height of 10 m are generated as a post-processing product from the ERA5 reanalysis product. Wind gusts estimates are derived from the sustained wind speed at 10 m along with a term representing shear stress and a convective term (Molina et al. 2020). They are also presented herein to provide a link to previous research on European windstorms that has focus on wind gusts. The association between these wind gust estimates and sustained wind speeds at 100 m are also presented. ‘
- 2) Wind gust values at 10 m for each storm are now briefly summarized in Table 2 and described in the associated text. ‘Maximum wind gusts at 10 m a.g.l. exceed the sustained wind speeds at 100 m a.g.l. at both the peak hour and over the entire windstorm. Maximum wind gusts from ERA5 for all windstorms are well above the U.S. National Weather Service ‘damaging winds’ threshold of 25.7 ms^{-1} (Trapp et al., 2006) (Table 2). The spatial correlation coefficient between 100-m sustained wind speeds and wind gusts at 10 m at t_p is > 0.68 for all storms and > 0.8 for 8 out of the 10 storms, indicating that the 100-m sustained wind speeds analyzed herein are strongly related to near-ground wind gusts in the ERA5 reanalysis.
- 3) The high degree of association between 10-m wind gusts and 100-m sustained wind speeds is now made explicit (see above.).

For the reviewers’ interest (and that of any readers who have sought out this document) in the following figure we show the relationship of 10-m gusts to 100-m hourly winds at the peak time of each windstorm.

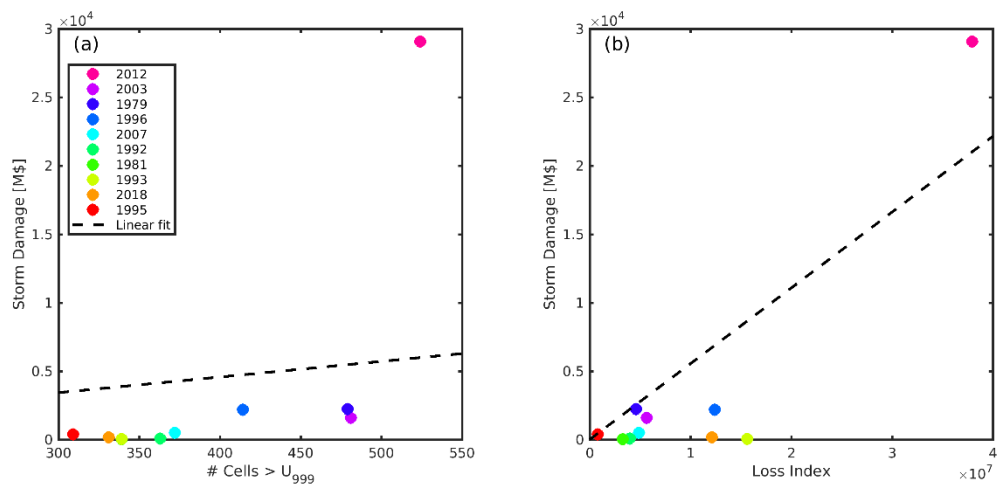


While retaining the 99.9th percentile wind speed threshold central to this work, we have presented information about use a the 98th percentile. See table 2 and the related text; ‘In previous work, the local 98th percentile value has been used to identify windstorms as it roughly corresponds to wind gusts at 10-m gusts that may cause property damage (Klawa and Ulbrich, 2003). Events with widespread exceedance of the 98th percentile threshold are common within the 40 years of ERA5 output. 139 events have sustained wind speeds in excess of their local 98th percentile in over half of all ERA5 grid cells over the Northeastern states. Herein, a higher threshold (99.9th percentile) is used to distinguish ten extraordinary windstorms. All ten also appear on the list of storms chosen using a 98th percentile threshold, with nine of the ten appearing in the top 50 (Table 2).’

At the reviewers request we have also included a Loss Index calculation based on Klawa and Ulbrich’s 2003 work. This is described in the methods section (2.7) and results are reported in section 3.4.

The following figure shows the relationship between storm damage and number of cells exceeding U_{999} and Loss Index. We have elected not to include this figure in the manuscript because it only confirms the fact the much of the variance in storm damage is not well represented by either metric. As with the figure

above, we have included this figure in our response for the reviewers' interest and for any future readers who seek out this document.



A tracked-changes version of the manuscript is included at the end of this response.

Responses to specific comments below are in **bold**.

Anonymous Referee #1

Received and published: 7 January 2021

This study presents an identification of severe windstorms over northeastern North America. Objective cyclone identification and wind speed exceedance methods are used to identify these phenomena and their dynamical features and impacts are put into the context of the wider climatology. The aims of the paper are good and I believe it could make an interesting publication that is appropriate for the journal and would provide a good piece of analysis to partner the numerous similar studies over Europe. However, I believe many of the methods used are inconsistent with these other studies and are not sufficiently justified or argued for in the text. All the methods used are slightly different from those established in the literature, such as tracking at 700 hPa instead of 850 hPa, analysing wind gusts at 100 metres instead of 10 metres, and using an extremely high exceedance threshold of the 99.9th percentile instead of the more commonly used 98th percentile. I would like to see some evidence for the choice of the methods. Furthermore, the introduction and framing of the paper (especially the first half) feels very incoherent and I believe requires significant re-structuring. In addition, the presentation of the figures feels clumsy and very difficult to interpret with numerous similarly styled lines and colours overlaid on similarly shaded fields, which needs rectifying. My full points can be found below. I therefore recommend that this study requires major corrections.

Please see our comments above regarding use of 10 m v 100 m and sustained wind speeds v wind gusts. Additionally, the reviewer mentions use of RV at 850 hPa in prior research over northern Europe for cyclone tracking. This level is not appropriate for use in North America because of the high terrain elevation that means the 850 hPa is frequently below the land surface. It is deemed preferable by the authors to use a level that is more typically realized in the atmosphere. Finally, we regret you had difficulty reading the figures, and have modified some as appropriate, and enlarged those that we could, to aid clarity.

Individual points

1. L43 – I feel the reference here to Fig. 1a should illustrate the two different types of cyclone tracks you are describing. As the figure is for wind gusts it in no way illustrates the differences in the genesis location

The Purpose of Figure 1a Is to explicitly communicate the locations of the Northeast states to readers less familiar with U.S. geography. The state names and abbreviations are used throughout the paper. We believe that changing the scale of this map to include cyclone genesis locations would make the state borders and labels difficult to interpret. The classification of storm tracks into; Alberta Clippers, Colorado Lows and tropical cyclones is consistent with past cyclone climatologies and is described at length in the introduction (lines 34 to 54)

2. L47-49. How is the influence of post-tropical cyclones in the USA consistent with Europe? Are less than 1% of cyclones also post-tropical in this region? Perhaps it is best to remove the 1% statement.

We regret the ack of clarity and have re-worded to; ‘Research on windstorm risk in Europe found that although less than 1% of cyclones that impact Northern Europe are post tropical cyclones, they tend to be associated with higher 10-m wind speeds (Sainsbury et al., 2020). Tropical cyclones are a major driver of extreme wind speeds along the U.S. eastern seaboard (Barthelmie et al. 2021) and events such as Hurricane Sandy have been associated with large geophysical hazards in the Northeast (Halverson and Rabenhorst, 2013;Lackmann, 2015).’

3. L61-63. Can a trend line or some evidence of the trend from the data in Fig. 1b also be included/quoted? Due to the large inter-annual variability of the data it is hard to tell from the figure you have presented that there is a positive trend. Furthermore, you quote the 98th percentile in the text (perhaps this refers to the results of Bronnimann et al., 2012), yet the figure in question (I assume Fig. 1b?) uses number of grid points exceeding the 99.9th percentile of U. Please clarify these differences,

or make the figure consistent with the text.

We believe the reviewer was confused by the reference to Figure 1 and Table 1 that emphasized/described the GEOGRAPHICAL region. We have removed those to avoid the possibility of confusion.

4. Section 1.3. From your introduction there is limited evidence of where this study fits within the established research and scientific literature. I think it could be framed better to give a clear narrative as to where the gaps in the literature are and how this study addresses those gaps.

We are very explicit in our objectives and links to past research have been clarified and extended in the revised version of the manuscript.

5. L123-124. I feel this statement is incorrect for the last 10-15 years with the introduction of more advanced and homogenous reanalysis products. Please make it clearer how ERA5 is beneficial compared to previous generation products such as ERA-Interim, etc.

We believe the real advantages of ERA5 are those we state in section 2.1. Assimilation of a massive suite of data; ‘The ERA5 reanalysis is derived using an unprecedented suite of assimilated in situ and remote sensing observations (Hersbach et al., 2020). ‘. Its relatively high demonstrable skill for wind speeds; ‘It exhibits relatively high fidelity for wind speeds (Kalverla et al., 2020;Olason, 2018;Kalverla et al., 2019;Pryor et al., 2020;Jourdir, 2020;Ramon et al., 2019)’. AND the relatively high resolution; ‘a spatial resolution of 0.25°×0.25°. This latter point – as the reviewer will know, compares very favorably with past global reanalysis products. Eg.; MERRA-2; 0.625°×0.5°, ERA-Interim (approximately 80 km).

6. I am a little confused as to why the 100 metre level is being used throughout this study as you are specifically interested in damaging windstorms. You have justified this in the text, however I feel it may be better to be consistent with previous studies, which generally use the 10 metre level. Does using 100 metres lead to drastically different results to using 10 metres? Furthermore, as you are using near-surface ground observations as a validation, surely it would make sense to also use the 10-metre product from ERA5 as there could be significant differences in the 100m and 10m distributions.

Please see above.

7. As above, doing the tracking at 700 hPa instead of 850 hPa is also confusing to me. Using the Hodges method tracking is mostly done at 850 hPa. Are there specific issues with representing the flow in the orographic regions? Are all cyclones identified at 700 hPa the same as those that would be identified at 850 hPa?

Please see above. In the text we note this justification briefly where we state; ‘RV values at 700 hPa are used rather than 850 hPa as in the XWS European analysis due to the presence of high elevation areas in U.S. cyclone source regions.’

8. I feel the methodology section can be condensed significantly. You introduce the tracking on line 148 and then describe it in detail, several pages later. It feels very incoherent and should be re-structured for conciseness.

We briefly introduce use of the ERA5 data for the cyclone tracking in section 2.1 and then present the methodology in section 2.5. We believe that is appropriate, but have edited the Data and methods section for brevity.

9. L210 – you use the 99.9th percentile, what is the justification for this? As you mentioned in the text your method is similar to numerous other storm severity assessments, however most use the 98th percentile. Due to the relative short extent of data from ERA5, are there not chances that the 99.9th percentile is exceptionally skewed by very large events?

This is an important point. We have amended the introduction to read; ‘This research is a part of the HyperFACETS project which uses a storyline-based analysis framework. Storylines are “physically self-consistent unfolding of past events, or of plausible future events or pathways” (Shepherd et al., 2018).

They provide a method of framing a research inquiry in terms of three elements: A geographic region, a historically important or notable event, and a set of process drivers for that event.’ To clarify we are indeed seeking to identify and characterize truly exceptional events.

10. L223 – is it possible that by using such a high threshold some spatially large events (p98<U<p99.9) are missed and events with a very small area of U>99.9 are counted instead? It may be that in some of these cases could the large, yet slightly lower intensity events have larger impacts than small scale high-wind events.

This is an important point. Use of any threshold inherently dictates the type of event that will be identified. As shown in Table 1 none of our events are confined to very small areas. Indeed, all of them cover at least a third of land grid cells. Nevertheless, we have added a discussion regarding the link between our events and their ranking using a different threshold (p98).

11. L225 – a 14 day restriction seems rather large. Have you tested this criteria to see if any high impact storms are excluded as a result of this threshold? Several studies have shown that more intense cyclones are more likely to cluster (e.g. Mailier et al., 2006), therefore this could be removing events from your set.

The 14-day exclusion window is used to ensure that all storms on the top ten list are independent of one another, and this window is reduced to two days when the serial correlation of storms is considered. There is one storm which would have been included in the top ten, were it not for this 14-day window. This is noted in the new text:

“A larger sample of 50 windstorms was also drawn from the 40-year time series to examine the serial dependence. This analysis reduces the 14-day exclusion window used in the identification of the top 10 windstorms to a 2-day window. One windstorm (on January 19th, 1996) is excluded by use of a 14-day separation window from the list of the top ten storms but is included if a 2-day exclusion period is used. It would have been ranked number ten.”

12. L235 – further references are needed here such as Vitolo et al. (2009), Mailier et al. (2006), Pinto et al. (2014).

We understand there have been other articles published on serial clustering of windstorms in Europe. Given our work is focused on North America and we already have 109 references, we think it is sufficient to refer to one of those studies.

13. L265 – do you have specific requirements for each of these cyclone classes (i.e. genesis location). If so please state this in the text.

We regret this lack of clarity. We have added the following text; ‘A cyclone is identified as an Alberta Clipper (AC) if the cyclone track originates over the North American continent north of 40°N, as a Colorado low (CL) if the track originates over the North American continent south of 40°N, and as a decaying tropical cyclone (TC) if the track originates south of 30°N over a water grid cell. The term nor’easters (NE) is applied if the cyclone retrogrades towards the coastline after moving offshore and/or is associated with strong northeasterly flow over the Northeastern states.’

14. Fig 2. I find the layout and organisation of this figure very messy. The legends should be moved outside of the panel boundaries and all text on the panels be made clearer as it is very hard to decipher any of the information.

We regret you had difficulty reading this figure. We have made it much clearer by moving the legends, removing damage estimates (since they are in the table) and increasing the size of the figure.

15. L317 – reference should be outside of brackets.

Typographic error corrected.

16. L348-349 – as discussed above. If your systems traverse your region of interest in ~72 hours, why the 14 day separation? Please clarify this.

We did not know before we did the analysis how long any individual storm would take to traverse the region. We wanted to set a sufficiently large window to ensure independence.

17. Fig 4 and throughout – is it worth referring to each storm by its ranking in the text and figures. I find it hard to keep track of which year is which through the text, this may be a simple way to avoid this.

We think that using the consistent color scale is effective in this regard but have added some information about ranking to the text also.

18. L396-406 – would the authors be able to illustrate these results in some way instead of just giving a description. Furthermore, commonly dispersion is calculated as counts per month, or counts per winter season, and not counts per calendar year. Vitolo et al. (2009) Fig. 15 demonstrates how dispersion (and surrounding uncertainty) commonly increases with aggregation period and as these storms you are interested in are only features of approx. 6 months of the year is it likely that this dispersion value is representative?

Possibly. Given the D value is small we prefer not to expand the discussion greatly.

19. Figs 6 and 7 – legibility of these figures is very difficult. Often green lines are plotted above an area of green (especially figure 7a) and also with text on the figures it makes it very difficult to distinguish and correctly identify features. I would recommend a redesign of these figures to aid legibility.

Thank you, the background colors in these figures (now combined into Figure 6) have been changed to reduce confusion, and the track outline color has been changed from white to black. The overall size of the figure has also been increased.

20. Fig 7b and table 3 – are the mslp units in hPa the anomaly from the background field? This needs to be made clearer as the magnitude of the values are confusing.

We have clarified by adding; ‘These are anomalies identified in the filtered fields. RV cyclone intensities are shown in units of 10^{-5} s^{-1} , and MSLP intensity estimates are given in hPa scaled by -1. These anomalies are relative to removal of the large-scale background for $n \leq 5$, where n is the total wavenumber in the spherical harmonic representation of the field.’

21. L412 – again, do you have requirements for these cyclone classes. It would be useful to define these earlier in the text instead of just approximating by eye if they are one class or the other. This is also applicable for L416 where you state a cyclone transitions to the NE class.

Please see response above.

22. L464-465 – please display this information somewhere in either a figure or table, with it being stated as is it feels unjustified. Also you state how all storms have $RP > 50$ years in some location, is this evident in figure 8 as the colorbar does not extend beyond 50? Perhaps to make this clearer the authors should extend the colorbar beyond 50 years and then highlight the regions which exceed 50 years.

The extent of the color scale in all panels of Figure 8 (Now Figure 7) has been increased to 60 years.

23. Figure 8 and table 3 – what is the uncertainty on the return period calculations? As with ERA5 there is only 40 years of data the uncertainty must be very large for the 100 year event. Please clarify and quantify this in the text and figures.

Quite. We have added this comment to methods; ‘Uncertainty intervals on the return period wind speeds are assigned using the 95% confidence intervals on the α and β parameters as derived using maximum likelihood estimation.’ And have added results to Table 3.

24. All figures – the visualisation and colour clashes at times makes for figures that are very

hard to interpret. Fig. 8 the red lines of states over the map is almost impossible to clearly distinguish. Fig 6 and 7 the track lines an intensity circles are hard to see as they overlap and also clash with the background colours. Figure 2 has legends overlapping figure space and also text on figures that cannot be read. I feel a redesign of these figures is required to accurately present the authors results.

All figures have been made larger to aid visibility. The red lines in Figure 8 (now Figure 7) have been changed to white to reduce visual clutter. Background colors in Figures 6 and 7 (now combined into Figure 6), have been changed to make the tracks, track labels, and intensity markers much more legible.

25. The storm severity metrics (following Klawns and Ulbrich, 2003) are introduced but in no way used in the analysis and I feel could be a strong contribution to the results. Performing an SSI-like calculation to compare ERA5 ranking to actual ranking would be a useful addition to the analysis. It would be good to compare the U₉₉₉ spatial extent and maximum wind speed to this storm loss metric.

At the reviewers request we have added a Loss Index based on their work in Europe (introduced in section 2.7) and results described in section 3.4. Our results suggest that for these ten windstorms the Loss Index is not very highly predictive of NOAA storm damages.

References

Mailier, P. J., Stephenson, D. B., Ferro, C. A., & Hodges, K. I. (2006). Serial clustering of extratropical cyclones. *Monthly weather review*, 134(8), 2224-2240.

Pinto, J. G., G3mara, I., Masato, G., Dacre, H. F., Woollings, T., & Caballero, R. (2014). Large-scale dynamics associated with clustering of extratropical cyclones affecting Western Europe. *Journal of Geophysical Research: Atmospheres*, 119(24), 13-704.

Vitolo, R., Stephenson, D. B., Cook, I. M., & Mitchell-Wallace, K. (2009). Serial clustering of intense European storms. *Meteorologische Zeitschrift*, 18(4), 411-424.

Anonymous Referee #2

Received and published: 15 January 2021

Review of “Windstorms in the Northeastern United States” by Frederick W. Letson, Rebecca J. Barthelmie, Kevin I. Hodges and Sara C. Pryor

The paper presents an analysis of the 10 largest windstorms over the northeastern United States during the past four decades. It combines metrics of extreme winds and precipitation based on both reanalyses and observations with tracks of the associated extratropical cyclones based on both pressure and vorticity. The time evolution of reanalysis and observational data qualitatively agrees during the storm events. The storms show typical tracks for the region but intensity about one magnitude higher than average. Several storms are associated with extreme damages over 1B\$ and their winds show long return periods above one century locally.

The regional focus on the northeastern United States complements earlier windstorm studies that focus on northwestern Europe mostly. In that sense, the paper is an important contribution to the windstorm community. However, it suffers from several short-comings that limit its actual impact. The paper tends to cover too many topics without a clear focus and to combine too many approaches applied in an unusual way. It would benefit from a well-defined scope, standard methods taken from the literature and a better structure altogether. General and specific comments are listed below to help improve the paper quality.

GENERAL COMMENTS

I. The scope of the paper is vague: it appears to be the 10 most intense (or largest?) windstorms over the 40 year period 1979–2018 but this is not clearly stated or motivated. Ten is not enough for a catalog and not significant to represent severe storms but likely too much for detailed case studies. The authors somehow need to choose between increasing the sample and focusing on a few particular storms.

Our work is being conducted within a project that centers on development and use of storylines. As our modified draft now reads; ‘This research is a part of the HyperFACETS project which uses a storyline-based analysis framework. Storylines are “physically self-consistent unfolding of past events, or of plausible future events or pathways” (Shepherd et al., 2018). They provide a method of framing a research inquiry in terms of three elements: A geographic region, a historically important or notable event, and a set of process drivers for that event.’

We hope this rewording helps to clarify. For the purposes of clarifying our purpose to the reviewer we note use of storylines facilitates interactions with stakeholders (i.e. they likely remember these extraordinary events) and also enables future work – e.g. using pseudo-global warming simulations to examine how such events might evolve in the future.

II. The methods are comprehensive (4 pages of description) but uncommon and not sufficiently explained or motivated: 100 m rather than 10 m winds, 99.9th rather than 98th percentile, intensity defined as instantaneous spatial extent of extreme wind rather than the footprint of, e.g., cubed winds above some threshold. This may result in the obtained ranking of storms not matching their damage.

Please see response above. We totally understand the reviewers (and editor) have conducted previous research over northern Europe that used a different methodology but feel confident ours is ‘fit for purpose’ for example it does identify windstorms that are present in the memories of our stakeholders.

III. Analyzing compound events is certainly meaningful but throughout the paper it is unclear whether the damages associated with the 10 selected storms are due to wind, precipitation, or both. This is somehow linked to the comment above, which applies to precipitation metrics as well.

We completely concur. In our new section 3.4 (added at the request of reviewer 1) we discuss this in more detail.

SPECIFIC COMMENTS

l. 1 The title should be more specific, e.g., referring to the most severe windstorms in the region

We have modified the title to be; Intense windstorms in the Northeastern United States

l. 14–15 “Alberta Clippers”, “Colorado Lows” and “Nor’easters” sound too specific for the abstract.

We think linking to the cyclone responsible for the windstorms is useful so, with apologies, invoke authors privilege.

l. 19 Why “those windstorms that occurred after the year 2000” only? What about the others?

The National Weather Service ASOS network of meteorological stations was subject to a series of notable upgrades that were concluded in 1999 (e.g. installation of sonic anemometers). Thus, verification of ERA5 winds and precipitation are most robust after the year 2000.

l. 28 In the midlatitudes?

Quite possibly, but as we exclusively cite studies from north America and Europe, we feel it is better to be specific.

l. 36–37 The second part of the sentence appears unnecessary.

We believe that the reviewer is referring to ‘It lies under a convergence zone of two prominent Northern Hemisphere cyclone tracks associated with cyclones that form or redevelop as a result of lee-cyclogenesis east of the Rocky Mountains’ We express it in this way because some weak cyclones do traverse the Rocky Mountains and re-intensify.

l. 43 What is to be seen on Fig. 1a?

We are notifying readers who may not know where Lake Superior is that they can see it in Figure 1a.

l. 48–49 I do not fully understand the sentence and the cited paper describes Europe only

We regret you did not understand this sentence. We have re-worded to read; ‘Research on windstorm risk in Europe found that although less than 1% of cyclones that impact Northern Europe are post tropical cyclones, they tend to be associated with higher 10-m wind speeds (Sainsbury et al., 2020). Tropical cyclones are a major driver of extreme wind speeds along the U.S. eastern seaboard (Barthelmie et al. 2021) and events such as Hurricane Sandy have been associated with large geophysical hazards in the Northeast (Halverson and Rabenhorst, 2013;Lackmann, 2015).’

l. 75–99 (Section 1.2) Many studies and numbers are cited but they lack focus on the topic of the paper: windstorms over the northeastern US

We regret you disagree with our choice of literature. Reviewer 1 suggested we add more literature about windstorms in Europe so it is indeed a difficult balance to strike. Compared to northern Europe there is little prior research on windstorms in the Northeast and we believe we have cited the majority of it.

l. 86 Repetition of l. 33–34

Quite correct, we apologize and have modified the text to avoid this duplication.

l. 101–102 This requires more details here and should somehow appear in the title and abstract

We believe the reviewer is asking us to add further details regarding ‘This research is inspired by and is conceptually analogous to development of the XWS (eXtreme WindStorms) catalogue of storm tracks and wind-gust footprints for 50 of the most extreme European winter windstorms (Roberts et al., 2014).’ We think that would not be appropriate to the abstract of our work.

l. 105 Most intense or severe?

Yes, this is an interesting point. We decided on intense because we cannot say they are the most severe given we use a relative threshold to absolute magnitude to define them.

l. 117–120 This paragraph is unclear and seems misplaced

We regret you felt so. We wanted to explain why we select extraordinary events.

l. 120 A description of each Section is expected here

It is not required by the journal style and checked five recently published articles in this journal (Nat. Hazards Earth Syst. Sci., 21, 587–605, Nat. Hazards Earth Syst. Sci., 21, 577–585, Nat. Hazards Earth Syst. Sci., 21, 607–627, Nat. Hazards Earth Syst. Sci., 21, 481–495, Nat. Hazards Earth Syst. Sci., 20, 3521–3549). Only one (Nat. Hazards Earth Syst. Sci., 21, 607–627) had such a statement so it appears to be at the discretion of the authors. We prefer not to include it as a matter of style since it uses space without conveying essential materials so for the time being, we do not include it.

l. 121ff (Section 2) The study period is unclear and inconsistent between subsections

We regret we are a little unclear on what the reviewer is referring to 1979-2018 is the time period we used exclusively to define the windstorms. Due to the presence of enhanced measurement system prior to 2000 the detailed evaluation is performed for a subset of the windstorms that occurred after 1999.

l. 123 Over long periods? It is not an issue for case studies

We have reworded to ‘Attempts to identify and characterize windstorms from a geophysical perspective and contextualize them in a climatological context have historically been hampered by limited data availability and/or quality from geospatially inhomogeneous observing networks.’

l. 128–129 Hourly or every 20 minutes?

We have reworded to; ‘Thus, herein we employ once-hourly wind speeds from the ERA5 reanalysis. The wind speeds are for a height of 100-m a.g.l. at the model time step of 20 minutes and a spatial resolution of $0.25^{\circ} \times 0.25^{\circ}$.’

l. 137–144 The motivation for using winds at 100 m agl is not fully convincing and slightly repetitive; this height is used for wind power mainly and may not reflect the surface impact of windstorms

We have rewritten to; ‘Here we use wind speeds at 100-m because flow at this height is less likely to be impacted by sub-grid scale heterogeneity in surface roughness length and uncertainties induced by unresolved sub-grid scale variability. Near-surface wind speeds are strongly coupled to wind speeds at 100-m (i.e. within the PBL) but wind speeds at 100-m are less strongly impacted by inaccuracies and/or uncertainty in surface roughness length (z_0) (Minola et al., 2020; Nelli et al., 2020). Applying an uncertainty of a factor of two to z_0 can lead to mean differences of up to 0.75 ms^{-1} for near-surface (40 to 150 m a.g.l.) wind speeds (Dörenkämper et al., 2020). Further, the scale of events we seek to characterize are regional rather than local scale, and are necessarily driven by winds aloft.’.

l. 145 Why 3-hourly rather than hourly as above?

Because its highly redundant. Most prior research has used 6-hourly (e.g XWS).

l. 155 Which ten storms?

Reworded to; ‘Precipitation intensity and hydrometeor class from ERA5 are used to identify to what degree each of the ten windstorms identified here are compound events.’

l. 168–171 I agree but this does not support the use of 100-m winds from ERA5

That is right. We believe from a theoretical perspective using 100-m a.g.l. wind speeds IS preferable (see discussion above). However, we have to be pragmatic about data availability for evaluation.

l. 175 Twice “product(s)”

Apologies. Corrected

l. 179–183 The two sentences sound repetitive

Quite. Corrected to; ‘In the current work, precipitation rates over the land areas of Northeastern states from RADAR and ASOS and ERA5 that are within 200 km of the 7 RADAR are averaged in time to match the hourly resolution of ERA5 precipitation and interpolated in space to the 0.25°×0.25° ERA5 grid. (Fig. 1c).’

l. 193–195 Verb missing?

No but the ‘to’ was extraneous. Apologies.

l. 194 What is the caveat here? The impact is indeed expected to depend on population density

Yes, our meaning was lost here. Reworded to; Damage and mortality estimates from this dataset provide an estimate of the impact of each windstorm, with the caveat that population density and hence the potential for loss of life and damage vary markedly between U.S. counties that also vary greatly in area (Fig. 1d).

l. 210–212 This approach reflects the severity rather than intensity (storm severity index)

See response above.

l. 212–215 The motivation for using the 99.9th percentile is unclear; although the approach is disputed by other authors, the choice of the 98th percentile arises from comparison with impact data in Klawa et Ulbrich (2003)

We hope we have clarified this matter in our other responses.

l. 215–217 Why?

We believe the reviewer is referring to; ‘While lower percentile thresholds have been used in previous work (Walz et al., 2017;Klawa and Ulbrich, 2003), use of the 99.9th percentile wind speed value is appropriate for identifying the truly extraordinary conditions we seek to characterize and is robust when applied to very long datasets with very large sample sizes.’

We really do seek to identify truly exceptional conditions and as we now show use of 98th percentile is not very discriminating.

l. 225 14 days appear as a very long period to separate events

Possibly but we wanted to ensure that Nor’easters’ were not double counted.

l. 227 if values are instantaneous then +/-48h makes 96h

Right but as we can see none last the full time window. The data used represent 97 hours.

1. 235–236 The focus on the 10 most intense (severe?) storms should be stated clearly and early in the paper

It is mentioned in the objectives explicitly.

1. 244 Reference without brackets

Corrected.

1. 259–266 Do the centroids of extreme ERA5 winds actually coincide with extratropical cyclones, in general and for the 10 most intense cases in particular?

No – this is shown in Figure 2.

1. 267–268 If I understand correctly the intensity is solely defined as the maximum instantaneous areal extent of extreme winds?

Yes.

1. 299–305 (Table 2) The windstorm category would be useful here (AC, CL, TC, NE)

We think it is more appropriate in Table 3.

1. 306–311 What are the actual most damaging extratropical cyclones for the considered period, and are they captured by the selection?

We think the reviewer is referring to Table 2. Certainly, we identify events that stand out in the damage record but in this region flooding linked to storm surge, and heavy precipitation also yield some high NOAA storm damage reports. Perhaps this is not what the reviewer is asking. We can say these ETC are exceptional relative to others with similar tracks see Figure 6 and associated text.

1. 313–320 An additional but crucial explanation may be that the used metric is simply not appropriate... In addition to different wind height and percentile, standard metrics are based on a footprint and of some function of winds (typically cubed) rather than on an instantaneous spatial extent.

Possibly – we have further elaborated on these matters in the new section 3.4 added at the request of the other reviewer.

1. 334–336 The strongest winds often occur to the south of the pressure center but not necessarily related with the cold front (see, e.g., the Hewson and Neu 2015 paper cited above)

Quite. We were too brief here we added an additional reference that talks more about some of the other mechanisms.

1. 340–345 (Figure 2) What is the reason for the large discrepancy between RV and MSLP tracks in some cases?

This response is combined with the response to the nest comment.

1. 350–351 It would be interesting to know why it is the case (1981 and 2003 storms)

We have added text to section 3.2 that addresses this matter; For most cyclones independent tracking of the center using MSLP and RV yields results that are highly consistent (Fig. 2). Nevertheless, some discrepancies exist in part due to the spectral field smoothing. Another possibility is if there is a strong background flow due to a strong pressure gradient, the vorticity can be offset relative to the pressure minimum (Sinclair, 1994).

l. 362 Hurricane Sandy was already introduced

Deleted.

l. 363–364 How many is “several” or “multiple” grid cells? 20 mm in 24 h is far from extreme. . .

Thank you. This text has been updated to read:

“Hurricane Sandy (windstorm during 2012) is associated with total 24-hour precipitation accumulation exceeding 100 mm in 0.5% of ERA5 grid cells, and exceeding 20 mm in 46% of ERA5 grid cells.”

l. 370 (Figure 4) The histograms do not emphasize extreme (i.e., potentially damaging) precipitation. And what is precipitation type 2?

Thunderstorm. We did not refer to it because it was not present for any of our windstorms, but have added it to the caption at your request.

l. 387–390 The underestimation of station observations may be explained by the spatial variability of precipitation but there is a factor ~ 2 between ERA5 and radar estimates

Quite. It is very interesting. We do not have an explanation but note this recent publication; Cucchi, M., Weedon, G. P., Amici, A., Bellouin, N., Lange, S., Müller Schmied, H., ... & Buontempo, C. (2020). WFDE5: bias-adjusted ERA5 reanalysis data for impact studies. *Earth System Science Data*, 12(3), 2097-2120.

l. 396–406 There is a confusion between clustering on yearly and daily time scales; I would expect the analysis focuses on the latter to distinguish between consecutive storms

We regret the confusion and have reworded.

l. 407ff (Section 3.2) This Section generally lacks structure

We have reorganized this section.

l. 408–412 How are these cyclones selected? Based on thresholds in MSLP or vorticity? The total number of extratropical cyclones likely exceeds 10 tracks per year by far

Right. we tested different limits and found this to illustrate the tracks well. Note this figure has been redrafted at the request of Reviewer 1.

l. 410 What are transitory cyclones?

We are using it in the context of transient weather systems, i.e. moving. Not, for example, the thermal low that dominates the desert Southwest.

l. 416–419 This was already stated above.

We believe the reviewer may be referring to ; ‘Consistent with past research employing other reanalysis data sets (Ulbrich et al., 2009), results from application of the cyclone detection and tracking algorithm to ERA5 output also indicate the U.S. Northeast exhibits a high frequency of transitory cyclones (Fig. 6). Also in accord with expectations, the tracks followed by the windstorms are generally characteristic of those dominant cyclone tracks, and derive from a mixture of intense nor’easters (NE), Alberta Clippers (AC), deep Colorado lows (CL), and decaying tropical cyclones (TC) (Table 3, Fig. 6).’ We think it is worth linking our work to past research and expectations since we have not yet describe the cyclone climatology.

l. 425–429 This belongs to the Introduction

We believe the reviewer MAY be referring to; ‘The Great Lakes are known to have a profound effect on passing cyclones during ice-free and generally unstable conditions that prevail during September to

November (Angel and Isard, 1997). Particularly during the early part of the cold-season, cyclones that cross the Great Lakes are frequently subject to acceleration and intensification via enhanced vertical heat flux and low-level moisture convergence due to the lake-land roughness contrast (Xiao et al. 2018). Cyclones such as Alberta Clippers that transit the Great Lakes during periods with substantial ice cover are subject to less alteration (Angel and Isard, 1997).’ We concur and have moved this text.

l. 436 The impact of the 2018 (not 2017) windstorm may be due to precipitation rather than wind mostly

Thank you for catching that typo. Yes, it is possible. Please see the new section 3.4.

l. 436–447 The impact is already discussed above (e.g., l. 418–424) and in other sections. Please reorganize

Done.

l. 450–460 (Figures 6–7) Same question as l. 340–345 (Figure 2)

See response above.

l. 463ff (Section 3.3) I am lost in the numerous details and wonder what to learn from this additional section

We believe the reviewer is referring to the return periods analysis. They earlier asked do our analysis approach identify real extremes. We believe this section is crucial to answering yes!

l. 496–502 This belongs to the Introduction

We regret we are unsure what text the reviewer is referring to.

l. 503 It is the first mention of the “largest” windstorms

Quite- we regret this typo and have corrected it.

l. 518–520 I do not agree on the accord with damage estimates

We have reworded to; The statistically significant correlation between the ERA5 windstorm intensity estimates and independent damage estimates provides further confidence in the fidelity of the windstorm catalogue presented herein.

l. 530–533 This is not convincing

We wonder if the reviewer is referring to; ‘These cyclones, however, exhibit considerably higher intensities (from both RV and MSLP perturbations) that are an order of magnitude higher than mean values sampled on those same tracks (Fig. 7). With the possible exception of Hurricane Sandy, these windstorms are largely differentiable from the cyclone climatology in terms of their intensification rather than the associated cyclone storm track.’ We have reworded to; ‘These cyclones, however, exhibit intensities (from both RV and MSLP perturbations) that are an order of magnitude higher than mean values sampled on those same tracks (Fig. 6). With the possible exception of Hurricane Sandy, these windstorms follow tracks that are not infrequent in the cyclone climatology.’

l. 560ff (References) Please . . . increase . . . spacing . . . between . . . lines

Done. Thank you.