

Review of ‘Rescuing historical weather observations improves quantification of severe windstorm risks’ by Hawkins et al.

The authors present an overview of Storm Ulysses and through the use of historical data records are able to produce new reconstructions of the storm using historical reanalysis. This new reconstruction results in a more intense storm and the authors highlight the major attraction of rescuing historical data for the purpose of expanding the record of historical storms, especially those of high intensity. In general, this manuscript is very well written and the authors clearly demonstrate the benefits of this sort of analysis and utilising historical written records for improving our understanding of historical storms. I only have several minor points of issue with this manuscript and after being addressed I would have no issues recommending this be accepted for publication. My main issue is around the framing of this manuscript and I believe slightly more background/justification for the analysis is required. This, and my other minor comments are detailed below.

Major comment

1. I believe a broader introduction is required and motivation of performing these re-runs is required. This is noted clearly in the abstract that a transformation of our understanding of historical variability is possible, but the authors do not note this until L141-146. No real aim or scientific purpose of the study is given until the results are discussed and this is something that needs to be rectified.

Minor comments

1. For all figures I would recommend labelling of panels via a, b, c, d, etc. This would make it easier to know which panels are being referred to rather than ‘bottom row’, etc.
2. L125, please include figure reference.
3. L125, Fig 4. How does the assimilation density of 20CRv3 compare to the assimilation performed by the authors using the new data? Will this affect the ranking that you have done in Fig. 4 (d-f)? By improving the density of Ulysses this may mean it’s winds are not representative of the 1950-2015 reanalysis and so the ranking may not be correct. Please clarify this.
4. Fig 4 (and throughout). It may be useful to show the plots of new data (and improved DA) as difference plots to highlight exactly where the re-imagined storms have strengthened relative to 20CRv3.
5. L221-224, how do the quoted percentages of ensemble members (49% and 22%) relate to the probabilities quoted in Fig 6. These values are different and I find it hard to understand why or how the authors have computed them to be different. This needs clarifying.
6. L226-228, this feels like repetition of two paragraphs prior. Please consider re-phrasing.
7. L231-232, Fig 6, it would be good to also quote the windspeeds of the non-precursor members. Furthermore, are there any statistical differences in the distribution of windspeeds simulated between the precursor and non-precursor members? If not it needs to be stated that even though the ensemble mean is higher, there is no statistical increase in simulated windspeed with sting jet precursors.

8. L316-319, Fig 8, Is the gauge data used in Fig, 8 a point estimate? If so I would not expect the output of the coastal surge model to match that of a point estimate as it has resolution of 12km. It may just be that the coarse nature of the reanalysis is unable to simulate such wave heights. This section is stated as if the storm is still not simulated to the correct strength is the driving factor of this, whereas it should be restated (in my opinion) that the difference in resolution of the two datasets is the leading driver of the difference and that an underestimation in intensity may be another reason why.