Review: https://doi.org/10.5194/egusphere-2022-1045
Rescuing historical weather observations improves quantification of severe windstorm risks
Hawkins et al.

This study shows the improvement of simulations of a historical intense windstorm due to added rescued observations from the time. The authors compare a reanalysis with added observations and new data + improved data assimilation. It is shown that the spread decreases and distinct features appear in the ensemble mean with more observations. The study is really interesting, and it clearly shows the value of rescued weather data. I have only minor concerns, which should be discussed more thoroughly.

Major comments:

- I certainly agree that rescued observations are of great value for improving simulations of historic weather events. However, I feel the positive statements from comparisons and argumentations could be toned down a bit at times and more discussion is needed. For example, a short discussion about their trustworthiness, accuracy, error range of the rescued observations should be included.
- The discussion is purely based on the ensemble mean (except for the sting jet precursor). I wonder if the missing wind jets in Figure 5 and that you discuss in l. 164ff are rather due to looking at the mean and are actually present in individual ensemble members. With a large spread, the maximum wind speeds of smaller features, such as the cold conveyor belt jet or sting jet, probably differ in location and, hence, are weaker in the ensemble mean. Of course, Figure 5 shows an improvement nonetheless, however, the argumentation why that is changes and you should state that the features "are not present in the 20CRv3 mean" (l. 172).

Minor comments:

- Please consider using hPa instead of mb.
- How did you track the storm?
- Please be consistent with figure labels (e.g., "new data" vs. "new observations", colorbar labels).
- Some figures are not discussed to their full extent. You show interesting information in the figures, which are – sometimes – not even mentioned in the text (e.g., probabilities in Fig. 6)

- l. 72f: As you state, 960mb is an estimate, so the comparison of the pressure minima in simulations with this value should be put in relation and not seen as the absolute truth.
- l. 85ff: Please add 1-2 sentences with more information about the added data and especially the improved data assimilation to the main text. It would be good to at least have an idea about the improvements without having to read the Appendix, which should then be for readers with further interest.
- Figure 3 and elsewhere: Please consider putting the colorbar labels right next to the colorbar.
- l. 218: Please elaborate the "simpler grid point approach". Do you mean you simply make the tool independent of neighbouring grid points, hence you could use it on every grid point independently? Please discuss shortcomings of this approach.

- l. 221ff: When do you define a member to show precursors: Is this already the case for only one grid point? How do these percentages compare to the probabilities in Fig. 6?
- Figure 6: The difference between 20CRv3 and new data seems to be much smaller than new data and new data + improved assimilation. Can you comment on this? Could the improved assimilation be more important for the improvement than the new observations? However, this is not really the case in other figures.
- l. 257: “observed”: As in the caption, you should at least mention the HadUK-Grid, i.e., interpolated in-situ observations.
- Figure 8: Please consider another colour scheme. Furthermore, what is the reasoning behind the 16-84% range?