Dear Editor and reviewer 1,

We thank the reviewer 1 for the helpful remarks. Our respective comments are in green.

Review of the paper "Reconstruction of wind and surge of the 1906 storm tide at the German North Sea Coast" by Meyer et al. (revised version)

I thank the authors for considering my remarks, most of which were satisfactory addressed. I want to make it clear that I am not questioning the expertise and competence of the people who produced the manually interpolated pressure fields. Probably, different forecasters would draw maps that are not too different from each other, and this would not be critical for the simulation of the 1906 event. I would just like to point out that the manual reconstruction cannot be considered as a proper reference (see my comment to No. 2, below), but rather one possible reconstruction.

My comments to the authors' replies follow. I only kept the relevant points and sentences of the authors' response document. Pages and lines are those of the "clean" text (not the marked-up version).

The paper can be published after taking into account these remarks.

(Note: At least from about page 14 the pages and lines mentioned in the answers do not match with the pdf.)

1) Manual reconstruction of the pressure field.

The first is that an objective
approach allows to estimate the reconstruction error, which is clearly not possible from a manual reconstruction.
Compared to an automatic interpolation method, the forecaster will keep in mind the air pressure distribution of the previous weather maps for deriving the air pressure tendencies in the manual analysis.
This can be used to improve the accurate location of the cyclone and the position of the isobars in areas
with poor data coverage. Incorrect air pressure values can be reliably detected by an experienced meteorologist and these values are then not used further in the analysis
Furthermore, the distance between the isobars is linked to the wind speed and the curvature of the isobars, which cannot be considered with automatic methods, but in a manual analysis. While the manual
analysis thus involves some degree of subjective assessment, the procedure is reproducible (as published in
Rosenhagen and Bork 2009) and the results are available and can be used by others for analysis or driving
models and comparison.
I also noted that, at least from the weather report in fig. 3, pressure used to be measured at different times between 7 and 8 CET, ...
The reports in Great Britain, Norway, Denmark, Belgium, the Netherlands, Sweden, Austria, and Germany
were made at 8 a.m. CET, in the other countries it was 7 a.m. local time or CET. The air

pressure data from
8 CET are therefore in the relevant area of the map. The different times were considered in the manual analyses.

I mentioned an "objective method" of interpolation (for instance objective analysis, optimal interpolation), which does not mean "automatic" because it requires the user to provide appropriate covariance functions, based on the analysis of the observations. The interpolation can be made simultaneously in space and in time, thus using the information coming from previous and following synoptic observations.

The reviewer is right, but the line of argumentation above still holds.

2) Page 13, lines 16-20. The text reads:
'Two inferences can be made: First, the similarity between the pressure charts suggests that the additional digitized data did not add value to the simulation of the wind fields compared to the used 20CRv3 reanalysis, because most of the new data are located in Central Europe. Second, in regions with a
high density of assimilated data, ensemble variability in the reanalysis is substantially smaller compared to areas where no such data exists. The similarity supports the reliability of the reanalysis for this event
adding some confidence in the available wind and pressure fields.'
What does 'similarity' mean? I understand that it is not a quantitative measure of how different the two
charts are, but rather that it means that they look more or less the same. Then the conclusions are totally subjective.
At least for the purpose of wind reconstruction, the authors conclude that the weather report data do not
seem to be useful. Therefore, considering the questionable way in which those data are used (point 1 above), it would be reasonable to drop everything connected with them.
Last but not least, I cannot agree that the reanalyses are reliable because they are similar to maps drawn
by hand!
We agree with the reviewer, that the comparison is subjective. We did this comparison to check if the patterns of the simulated isobars from the reanalyses are comparable with the weather situation, e.g. with the reconstruction. For long-ago historical storms, there may not be enough pressure data in key regions, so the storms may not be reconstructed in detail. For example:
Feuchter et al., 2013 investigated the 20CRv2 pressure data with the pressure data from a reconstruction
by Rosenhagen and Bork (2009) for the storm event 13.11.1872 in the Baltic Sea. In this case, a missing
pressure observation has changed the isobars for this storm in the reconstruction resulting in stronger gradients over the Bay of Lubeck compared with the reanalyses. Also, an investigation by Ed Hawkins et al.
2019 shows in a case study that added digitized data improved understanding of the severe storm from 1903 over the British Isles. In the 20CRv3-data these added data were not included and the reanalyses do
not simulate a severe storm. In his talk 'Improving Atmospheric Reconstructions For Historical Extreme
Events', Ed Hawkins shows the added value of such missing historical data for the severe

storm in 1903 (https://www.rmets.org/event/virtual-meeting-improving-atmospheric-reconstructions-historicalextreme- events)
For the severe storm 1906, there were data at the crucial regions already available and assimilated in the
reanalyse model. With the reconstruction, we can verify the pressure pattern with the reanalysis model data.
We addressed the points/concerns raised by the reviewer and revised these points in the chapters “1 Introduction” [page 3, lines 15-31 in the marked-up manuscript version] and “2.2. Reconstruction of sea level pressure and wind fields using the manual synoptic approach“ [page 7, lines 23-27 in the marked- up
manuscript version] to make them clearer.
Hawkins et al., 2019, 'Hourly weather observations from the Scottish Highlands (1883–1904) rescued by
volunteer citizen scientists', Geoscience Data, 6, 160, doi: 10.1002/gdj3.79

I fully agree that data archaeology is extremely valuable; the reanalyses themselves are produced using historical observations.
The piece of text I was referring to (now on page 13, lines 15-22) was extended but the authors did not answer the question about similarity, and the reanalyses are still considered reliable because they look similar to the manual reconstructions. The authors seem to consider the manual reconstruction as a reference but I think that scientific results should have an objective basis, when available.

I suggest to say that, because the maps look similar, the main features of the respective wind fields are consistent.

We changed the text accordingly. Clean text, on page 13, lines 21-22

Two inferences can be made: First, the similarity between the pressure charts suggests that the additional digitized data did not add value to the simulation of the wind fields compared to the used 20CRv3 reanalysis, because most of the new data are located in Central Europe and enough assimilation data in key regions were available for the reanalysis. Especially in the case of historical storm events, single missing pressure data can be crucial parameters for the course of isobars and thus incorrect wind speed can be calculated (Feuchter et al., 2013, Hawkins et al., 2019). Second, in regions with a high density
20 of assimilated data, ensemble variability in the reanalysis is substantially smaller compared to areas where no such data exists. The similarity between the pressure charts of the reconstruction and the reanalyses suggests that the main features of the respective wind are consistent.

3) 3.2 Ensemble simulations.

We split the figures to enhance the visibility. For the discussion it is important to make the most extreme water level visible. Therefore, we refrain from a synthetic assessment that will blur the individual results. [page 15- 16 in the marked-up manuscript version]

Ok, but the figures are still very busy, also because of the relatively large tidal signal. What about showing the de-tided curves and data?

We changed the colors of the ensemble members in grey. Only the maximum ensemble is highlighted. In addition, we included a table with maximum, median and standard deviation.

Clean text, page 14 - 16