Referee 2

The manuscript deals with a important topic, which is the decadal variability of windstorms affecting Europe, and the different perspectives of meteorology vs. insurance. While the idea is good, the execution has unfortunately many shortcomings, both in terms of the assumptions and analysis, and the reasoning and conclusions are very speculative on some parts. Therefore, I must suggest the rejection of the manuscript in its present form. However, I think that a strongly reworked version could be an important contribution to NHESS. Below you can find a list of the major caveat of this study, and I hope these help the author to reformulate the manuscript

- The author thanks Referee 2 for providing very clear and informative comments in their feedback, which led to substantial improvements in the revised manuscript.
- The manuscript is thoroughly revised: descriptions of wind and loss datasets are more complete; methods to build storm footprints of wind, and their conversion to loss have been re-written and the manuscript was re-structured for clarity; analyses of results include a fuller presentation and discussion of all evidence and uncertainties. This helps clarify how conclusions are connected to the work reported in the manuscript.
- Despite all these changes, the core of the work remains the same, on developing a loss record and using it to assess the climate indices predicted by forecasting systems. The findings are similar too, and useful for both climate researchers and the insurance industry.

Major points

1) datasets, line 75ff - it is not understandable to me why the AVERAGE between two different reanalysis is taken here. This will flatten the fields without necessity. Please do two separate analysis, one for each Reanalysis.

- The NCEP reanalyses are not used in the revised manuscript, and no averaging of different reanalyses values is applied.
- For background, the referee may wish an explanation of why it was used in the original manuscript. The averaging of reanalyses was used in conjunction with the proposed new climate index. This index depended on tropical heights in the 1950s and 1960s when observational density was low hence reanalyses values are more susceptible to model biases. Based on findings in Bengtsson et al. (2004), and a recommendation in Thorne and Vose (2010) to use ensembles to reduce non-meteorological trends caused by assimilation of inhomogeneous observed data, model-averaging was applied. However, the revised manuscript has its focus shifted onto fully describing the data, methods and analysis used to make the loss dataset, and the part on climate indices is much reduced. The revised manuscript suggests the development of a new climate index, with a stronger connection to losses in the 21st century, as a candidate for further work.

To look into the relationship between the large-scale patterns / pressure gradients and storms, a sub-monthly time scale would be preferable (e.g. Fink et al., 2009, doi:10.5194/nhess-9-405-2009)

- Fink et al. (2009) did a forensic analysis of the mechanisms causing extreme gusts during storm Kyrill, and Browning (2004) performed a similar investigation of storm 87J. They both looked at small space and time scales to identify the process causing extreme gusts in different areas. These articles are held in very high regard by the author.
- However, the main aims of the manuscript – to create a timeseries of losses then measure how well the winter climate indices (used in decadal forecasts) correspond to the losses – are different from the aims of Browning, and Fink et al. Though there is one aspect of the revised manuscript which

overlaps, namely the finding that near-surface winds have a different long-term trend from gusts. The manuscript contains a possible explanation that gusts have a different mix of processes from winds, due to their different timescales. In this context, both articles (Fink et al. and Browning) are a rich source of information on gust mechanisms, and referenced in the manuscript.

2) same section - it not clear to me why the wind gust variable from ERA5 was not used, but rather the 10m winds, as it is the former that is responsible for the damage.

- The author agrees *observed* gust values are the better predictors of damage. However, during development, it was found that storm losses based on ERA5 modelled gusts validate poorly. ERA5 gusts are based on low-level winds with a gust parameterization added (section 3.10.4 of the IFS documentation https://www.ecmwf.int/en/elibrary/79697-ifs-documentation-cy41r2-part-iv-physical-processes), and it was found that storm damage based on ERA5 near-surface winds (no gust model scheme) validated better to observed damage totals.
- The first paragraph of Section 2.2 in the revised manuscript discusses this point.

3) same section - The version perils data used by the author give extremely limited information and are heterogeneous in space an time, thus hampering the analysis. I would recommend to use the commercial version of the data, to which the author should have access to.

- The author has no access to the commercial product with PERILS detailed losses.
- The publicly available data contains information appropriate for this study of Europe-wide losses. The commercial license from PERILS provides access to more refined data (by cresta, and line of business) and not used in this study.
- PERILS issue estimates of full event losses to 12 countries which together experience the majority of insured losses, and they use homogeneous data collection and methods since their inception in 2009. They are widely regarded as high-quality estimates of windstorm losses over the past 14 years. The revised manuscript includes a new description of the quality of PERILS loss estimates.

4) data processing - this section is badly written and lacks an lot of details, and thus the methodology is not understandable. For example, I do not unterstand how "monthly data is processed into storm seasons", and the quantification of the storms and their impacts must be done at the sub-daily scale, optimally 3h for ERA5.

- The manuscript was revised to include more information on the data and methods to make footprints of wind and loss. It also has a new structure which is intended to provide a clearer description of the data and methods. For example, the steps to create storm wind and loss footprints are gathered together into the new Section 3.
- More specifically, the revision may now be clearer on how storm winds are based on hourly data from ERA5, and damage impacts follow standard practice and estimated using event-peak winds.

5) 3.1. basic method. The metholody by Klawa and Ulbrich is well established and has been used by a large number of publications since. However, it is not clear for which area the data is calculated, or many other details

- The author agrees the original version lacked full details of data and methods. The revised Section 2 includes more details on data and its processing, including a map of the studied area in Figure 1, and the new Section 3 contains more details on the method
- The aim of the revision is that a person with some experience can replicate the methods and get the same results.

6) same section - Given that it is not clear for which area the index was calculated, and the PERILS data is only available as a single value for a subset of countries affected by a storm, the reasoning regarding Figure 1 is not understandable. A lot is section 3.2. is quite speculative. In particular, the conclusion in line 161 is not justified.

- The first paragraph of Section 2 in the revised manuscript displays the study area of 12 countries used in this manuscript, and mentions how PERILS loss estimates cover the same 12 countries, which incur the vast majority of European windstorm losses covered by the insurance industry.
- The data and analysis leading to the conclusion in line 161 of the original manuscript is now described in much more detail, because the conclusion is key to the development of the new loss dataset. Sections 4.1 and 4.2 of the revised version contain a fuller discussion of all available evidence, and why it was concluded that reanalyses winds are the likeliest cause of a high bias in the initial loss dataset for the most recent period.

7) section 3.3. it is not understandable for me how the loss variability of a single, small country like the Netherlands can be used to make assumptions for such a large area. Small countries have either a full hit or not hit (and thus a steeper loss curve), while larger countries like Germany or France have often partial hits, leading to a flatter curve (e.g. Karremann et al. 2014, doi:10.1088/1748-9326/9/12/124016) Thus I totally disagree with the statement in lines 170-171.

- This comment refers to a method to adjust the long-term trend in the initial loss timeseries towards what is observed, and is replaced by a new method in the revised manuscript
- The original manuscript used a trend correction applied directly to losses, and partly based on Dutch loss history. This is now replaced by a trend correction applied to ERA5 winds, and based on observed gust trends. The difference between ERA5 and observed gust trends are described in section 4.2, while section 4.3 describes how ERA5 winds are modified to contain the observed gust trends.
- This means the statement in lines 170-171 are not relevant to the revised method.
- Though the Dutch loss history plays no part in the new trend correction, there may be a misunderstanding which is worth clarifying. The author fully agrees that the loss curve is steeper for a spatially smaller country. However, the original manuscript was referring to variations in long-term *average* losses, and the spatial scales of anomalies in *average* loss are larger than the Netherlands and contain a lot of exposure from Belgium, southeast England and North Rhine-Westphalia. Nevertheless, the use of Dutch losses to calibrate loss trends is replaced in the revised manuscript with a method regarded as better, because (a) it treats the likeliest cause of the trend bias in the initial version of the dataset (the ERA5 winds) and (b) the new calibration is based on a much bigger dataset spanning many decades and countries (GSOD weather station gusts).

8) section 4.1 the variability of the storm activity over Central Europe is partially associated with the mentioned large-scale patters, but is best associated with a eastern shifted NAO pattern (see e.g. Fink et al., 2009). Regarding the NAO pattern I think the author is overstating the results, because a correlation of 0.60 indicates an explained variance of 36% only.

- The manuscript assesses standard climate indices, and one of these is the Scandinavian Pattern (SCA) which represents pressure gradients over Europe, and it is found to be the climate index with closest association to storm losses, in agreement with the reviewer. (Note though that its improvement relative to NAO-PC is quite marginal.)
- The correlation values are different in the revised manuscript due to changes in filtering suggested by Referee 1, and the language used in discussion is modified too.

9) I strongly disagree with section 4.2. - to achieve a better skill for storm impacts over Europe, a more specific index should be chosen, like a shifted NAO index, and not an even larger-scale pattern.

- The original section 4.2 is completely removed from the revised manuscript. It was found that a detailed description of data and methods to make the initial loss timeseries, and the deeper analysis required to correct its recent multidecadal trend, followed by its comparison to standard climate indices, amounted to a significantly longer manuscript than the original.
- The reason why current climate indices fail to simulate the steepness of loss declines from 1990s to the 2010s remains an open question, and the revised manuscript includes a suggestion that this would be a good topic for further research.

10) same section - the author is assuming a linear relationship between winds above the boundary layer and surface gusts, which is a oversimplification as the main factor is actually the turbulence in the boundary layer (see e.g. Born et al., 2012, doi:10.3402/tellusa.v64i0.17471). Thus is particularly true for when strong convective is enbedded in the cold fronts of the storms, see again Kyrill, Fink et al., 2009).

- As discussed above in reply to point (9), the revised manuscript does not propose a new climate index, and instead recommends it for future work, and Section 4.2 of the original manuscript is removed.
- Though this topic is beyond the scope of the revised manuscript, the referee may be interested in published work describing processes that extend above the boundary layer and cause some of the strongest gusts in extreme events. For example, Fink et al. and other studies have linked the peak storm gusts at locations to deep convection reaching far above the boundary layer, and modellers consider their near-surface gusts to be produced both by vertical momentum of the downdraft being deflected at the surface, and the transport of upper-level momentum to the surface.

11) Conclusions - given that the analysis has many weaknesses, the conclusions are unfortunately highly speculative.

- The author appreciates the referee's very clear views on the manuscript.
- Several changes have been made in response to their comments: (i) the descriptions of wind and (especially) loss data are more complete, (ii) more information and data are included (especially on loss trending, and observed gusts), and (iii) there is a more detailed analysis of all evidence before drawing conclusions.
- As a result, the manuscript has grown in length and the proposal for a new climate index based on hemispheric geostrophic winds is removed. Instead, the development of a climate index more tightly coupled to recent low levels of loss is suggested for further work.