Review of "Insurance loss model vs meteorological loss index – How comparable are their loss estimates for European windstorms?" by Julia Moemken, Inovasita Alifdini, Alexandre M. Ramos, Alexandros Georgiadis, Aidan Brocklehurst, Lukas Braun and Joaquim G. Pinto

The authors describe a comparison between reanalysis-based Loss Index with losses obtained from the European Windstorm Model of Aon Impact Forecasting. In addition, two different reanalysis products (of diferent generations) are compared to assess the impact of the reanalysis on the actual Loss Index. The study addresses a very interesting subject and the ability of the authors to compare reanalysis-based Loss Index with the Aon Impact Forecasting model is an opportunity and very relevant for the community. However, I found the manuscript a bit difficult to follow at times but perhaps more importantly, I am concerned about the representation of the wind gusts that is used in the manuscript. The authors are right with their statement on line 137 that "the maximum daily wind gust speed is assumed to be the most relevant factor in these models". My most serious concern relates to this. A bit more detail is provided below. There are a few other issues the authors may want to look into.

My advice to the editor is to accept with major revisions.

Major concern:

line 105-110. There is a bit of concern on the parameterization of wind gusts from the reanalysis data. One approach is to use the local near-surface wind speed and its standard deviation in order to estimate the gust (like the Panofsky et al 1977 approach used in the manuscript). This approach makes use of similarity theory, and relates the gust to the friction velocity. The approach performs well in flat terrain, but is sensitive to the parametrization of the local roughness length. The accuracy of the estimated gusts relies heavily on the roughness map that is used, especially when the resolution of the NWP model increases and detailed information about the land-use (and the associated roughness lengths) is required. Errors in the supplied roughness lengths will directly influence the calculated gusts, which is a disadvantage of this approach.

My suggestion is to add a brief analysis where the Panofsky et al (1977) approach is compared to an alternative approach which is specifically suitable for use in a reanalysis product (van den Brink 2019). It links the 1-hour wind speeds at height (which is a standard output of the reanalysis) to 10m wind gusts. This comparison can be done over the entire domain or for specific storms.

To complete this analysis, actual observations of wind gusts should be combined to this assessment. Wind gust values for Europe can be obtained from the European Climate Assessment & Dataset at www.ecad.eu. If you have troubles finding the right data, simply contact ECAD staff.

This additional analysis assess the quality of the parameterization used for the wind gust calculations which is central to this study. The quality of the parameterization is therefore essential and requires a bit more scrutiny that the brief comments that is currently found in the manuscript.

Other aspects the authors may want to look at:

- section 3.1: fig. 2c shows that ERA5 generally has higher wind gust values than ERA-Interim, but the strongest differences are found over area with complex topography, like the Pyrennees, Alps, Norwegian coast and (perhaps) the Scottish highlands. The higher values over Europe are likely related to a mix of better physics and higher spatial resolution - as the authors correctly state. Could you explicitly state the spatial resolutions of ERA5 and ERA-Interim in the section where they are introduced? Now the resolution of ERA-Interim is mentioned on line 308 in the very last section of the manuscript. With a coarser resolution, complex topography will be much less well represented and peaks and vallayes will be less pronounced which directly affects the wind gust. Perhaps good to make this explicit in the discussion of the reanalysis.

- line 204-206. I am afraid that I fail to see why storm Irina is such an outlier. The Loss Index for ERA-Interim is for this storm much larger than for ERA5, but the storm footprint (fig. S1) does not really show a much larger region where the footrpint != 0 over the UK. ERA5 does show higher values (mostly because of the high resolution of ERA5 I guess). So, what do I fail to see in the explanation?

- Figure 3 and 4: it would be interesting to add an analysis where ERA5 is first regridded to the ERA-Interim resolution, and then the LI diagrams are made. This analysis gives a clue if it is the improved physics in ERA5 that makes the difference or that the increase in spatial resolution makes the difference. This would be nice to add to the Supplementary material. This analysis could then provide the basis for Section 5, bullet 1: I have not seen evidence that it is the resolution that makes ERA5 better than ERA-Interim (although this is likely).

- Section 5: bullet 1: the distribution of wind gusts may be shifted right in ERA5, but the footprint uses the 98th percentile - which is also shifted right. So this argument does not make sense.

- Figure S1: In the caption of the figure you write "Shown is the percentage of the maximum wind gust in 72 hours that exceed the 98th percentile of daily maximum wind gust." If you aim to show the outcome of equation 2, the this should be something like "Shown is the strength of the maximum wind gust in 72 hours as deviation from the 98th percentile and normalized with the 98th percentile."

REFERENCE

van den Brink, H. W. (2019). An effective parametrization of gust profiles during severe wind conditions. Environmental Research Communications, 2(1), 011001. DOI 10.1088/2515-7620/ab5777

NUTRevision for paper: nhess-2024-16: Insurance loss model vs meteorological loss index – How comparable are their loss estimates for European windstorms

Dear authors, I read your paper with curiosity. I think it is interesting for scientists to begin collaborating with the insurance sector, so that researchers and insurers can better understand the impacts of severe weather on different socio-economic sectors. I do think it is a timely topic, but I believe the manuscript requires quite some work to become clear and deliver its core messages effectively. Hence, my recommendation is a *Major Revision*, and I hope the comments in this document will be helpful. Good luck.

Major comments

In this section you can find comments in two categories: structural and data analysis. For the structure, I have the impression the paper could benefit from a clearer structure, with a better division between data descriptions and the methods, whereas for the analysis comments there are parts that remain unclear.

1) **Structural comments**

Introduction: The introduction requires some streamlining, since it intertwines motivating reasons to carry out such a study with lengthy descriptions of previous work. As a result, it is difficult to follow the storyline the authors wish to convey. For example, in L36-L51 you begin talking about the hazard, exposure, vulnerability framework, but this somehow becomes diluted in the rest of the paragraph. It might be helpful for readers to center the introduction about these three components of risk management using the risk propeller figure, so that the references to these multiple insurance companies and other articles are somehow anchored to this image. Then in L65-L68 the authors roughly describe the analysis that will be doing, which I find too detailed for an introduction, to then explain the paper structure, which jumps back to the general scope. Overall, I think this section requires streamlining and making sure the message the authors wish to convey is effectively delivered.

Data and methods: I would recommend re-structuring this section. While reading, there are parts mixing data description with the methods, which interrupt the flow. For example, L102-L114 describe the ERA5 data (and other generalities) right after the equations for LI are presented. Then in L116 the flow is recovered. Same goes for the description of PERILS in L154-L163. On the one hand, in the introduction the authors mention a hazard-exposure-vulnerability schema. On the other hand, I have the impression that the hazard

component is ERA5, the exposure is PERILS, and the vulnerability the data/curves from AON. So I would recommend restructuring this section in 2.1.1 - Hazard; 2.1.2 - Exposure; 2.1.3 - Vulnerability and then a 2.2 - Meteorological loss index and 2.3 - Catastrophe model that are thoroughly explained without data description intrusions.

Summary and discussion: I find this section long and I am not sure what the main conclusions of this work are. Is there any way of separating the "more technical" discussion part from the "more abstract" conclusions? Overall, I do not see the "take home message", or how does this relate with the two very concrete research questions posed in L60-L64. Also, how might the insurance sector be using the insights gained in this study?

2) Data analysis comments

2.1 - Meteorological loss index

In L86 the text say "Losses are proportional to the wind power or the wind kinetic energy flux….". Perhaps softening or extending this description might be useful for a generic reader to comprehend the meaning and implications of this.

In L87-L88 you mention that "…only the 2% of wind gusts….cause damage". I am missing here some elaboration about what are the damages that you have in mind. Are we talking infrastructural damage? Agricultural damage? To public or private assets? Is personal propriety included here? If this is one of the four assumptions in the paper, I would expect to have a solid description of what is the meaning of "damage" for the authors in this work.

In L90 you mention "In the case that no insurance data, population density can be used as a proxy for the exposure component". Indeed, but then does it mean that you are focused in damage in cities, hence, roads, agriculture, or forestry damages are out of the study? Also, how frequently do you bump into records that have no insurance data associated? I think this study could benefit from some extra clarity on how much insurance data is available, as long as its contents. This might help at assessing whether population density is a matching candidate for the insurance data or requires combining it with other layers (e.g. land use, urban tree, urban morphology).

Also, I wonder how the different spatial dimensions are accommodated in this analysis. For example, population density from CIESIN at 0.25deg is roughly 30km, but then how insurance data are aggregated? Per country? Per NUTS region? And how does this relate with the spatial resolution of ERA5, ERA5-Interim and the catastrophe model from Aon? I

believe it would be useful to have a section discussing the harmonization of the spatial dimension, so that it is clearer what the two models receive as input.

4.1 - Windstorm loss

Here in Figure 5 some results are visualized in the geographic space. In this figure I have two comments. First, the results are presented in a per-country basis, but the analysis seems to have been carried out on pixels much smaller than the country surface. I wonder if results can be presented using NUTS 2 regions or a spatial unit that is closer to the spatial dimension of the analysis. If results are aggregated for the sake of visualization, this would be understandable, but then I would expect a clearer description of the treatment of the spatial dimension throughout the manuscript. What is the resolution of the insurance data? How are all these harmonized? Second, the colorscale chosen in this figure might not be ideal to visually perceive differences. Perhaps a sequential colormap (with 3 colors) or a perceptually uniform sequential colormap (eg. Like viridis) might be a better choice to guide the reader to the differences you describe.

Also, I do not really understand how to interpret the Figures with the storm ranks. What helps the reader understand what is relevant?

Minor comments

- L43-L51: I think this might be a bit too detailed for an introduction, perhaps I would recommend streamlining this part.
- L44: "a direct view on the impacts" -> what type of impacts? Economic? Human?
- L50: "Actual loss reports… are usually not publicly available". Is this a limitation for this study, given the access to Aon data?
- L62: "In our study": I thought it was making reference to (Moemken et al., 2023) that is previously explained, so perhaps "in this study" reconnects with the paper at hand.
- L84: "Loss Index (LI)" has been previously defined
- L84: "…in the adaptation of Karremann et al (2014)" this line is written just as at the end of the introduction
- L131: The red line of the figure is not too visible on printed paper
- L168: Is this the right title for a section also containing storm rankings? Is there a more informative way of conveying the content of the section?

- L175: In Figure 2, I believe panel (c) is visually cluttered and it is hard to interpret the map with the overlying grid. Please, consider using something as “bivariate color scales” to represent 3 dimensions in a 2D space.
- L177: I understand that you have to cut somewhere, hence the choice of the 98th percentile, but I am lacking some extra motivation on why this is the right choice for this analysis. Can you elaborate? Is it common practice? What happens if you pick the 97th percentile?
- L178: “…is calculated for the winter half year”. I think it would be good to make explicit the reason of this choice
- L184: I am not familiar with the definition of “Core Europe”, but I would suggest coming up with a more neutral definition for a portion of the European continent. Is Central Europe and British Isles (CEBI) a reasonable option? Or study region/area? Is it correct to call it “Core Europe” after Brexit? I think it is important to strive for neutral terminology as much as possible.
- L205: “…Irina is easily…”: non-neutral language
- L209: “…on a paired Wilcoxon Signed-Rank test”. I am acquainted with Pearson, Spearman and Kendall indices (which I perceive as more common), hence I believe the choice of Wilcoxon could be better justified. What do you expect to obtain from its application? What are the benefits?
- L225: In Figure 4, I can see the names with a pair of integers in brackets. What are these? Are they part of the Wilcoxon test? If this is relevant information, I believe it should be elaborated somewhere. In the end, this also suffers from visual cluttering, so I think it is important to guide the reader on how they should interpret these plots.
- L226: Figure captions should be self-explanatory, so in my view using “Same as Figure X” is not an appropriate practice. Other figures also have this problem.
- L270: In the figure, is it okay that the axes bounds are not the same?
- L292: “…the wind gust distribution in ERA5 is slightly shifted towards higher values”. I can find this explanation in several places along the document, but then my comment would be, why don’t you show a couple of histograms illustrating the distribution of wind gust in your study period? It can go to the supplementary material if it does not belong here.
- L304: By inspecting Figure 8 (and its description) I am interpreting that the plain LI ERA5 index does a good job, and the improvements provided by Aon’s model might not be substantial. Perhaps if the analysis data was aggregated at a NUTS 2 region (provinces/subnational regions), more differences would pop up.
- L328: I found the explanation of the 98th percentile. Perhaps this could be moved up to the Data & Methods section.

- L340: Another mention to the tail of the distribution. I recommend adding this to the manuscript, since this seems to be a relevant piece of the discourse.