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, Joaquim G. Pinto

The paper compares estimated windstorm losses using the meteorological Loss Index (LI) with losses obtained from the European Windstorm Model of Aon Impact Forecasting. The paper has a twofold focus. First, how sensitive the loss estimates of the meteorological Loss Index (LI) is to the meteorological input data using two different reanalysis products. Secondly, comparison of windstorm loss estimates from the meteorological Loss Index estimates and an insurance loss model using the Aon's Impact Forecasting model. They conclude that the loss distribution in the LI estimate is not steep enough and accordingly the tail of the loss index distribution is too short, leading to an underestimation of high impact windstorms compared to the insurance catastrophe model. Despite of its differences, compared to the Aon model, the authors conclude that the meteorological Loss Index model is suitable for estimating the impacts and rank events.

General comment

It is clearly within the scope of NHESSS and is written in a clear and well-structured way. The research questions are clearly outlined in the final part of the introduction. The paper has the potential to shed light on the differences between a rather simple, but well documented open access approach and a more refined proprietary commercial product. This is very welcome contribution that could inform the community on the differences between approaches pursued by the academic community and private sector.

I find the analysis to be somewhat superficial and with few exceptions it consists of correlation analysis and scatter plots. The correlation analysis is rather hard to interpret and to properly answer the research question "How comparable are windstorm loss estimates from this meteorological index and an insurance loss model?" a much more multi-faceted approach is needed.

Aon's Impact Forecasting model is a commercial windstorm model and is treated as a black box. From a scientific point of view this seriously hampers both the depth of the analysis and the information value of the conclusions. Questions like why the two approaches differ can only be answered with statements concerning the shape of the final loss distribution and there is no comparison against reality which prevents any statements on quality.

I believe the manuscript requires major work before it can be accepted and recommend a resubmission or a major revision.

Specific comments

- Abstract: Research question two (comparison between the models) which from the title of the paper is the most important only have three sentences in the abstract.

- Abstract: It would be beneficial if the qualitative statement such as "shows comparable storm ranks", "yet it is an effective index", "higher LI values for ERA5 than for ERA-Interim" are justified by some numbers.

- Introduction: The introduction starts out with a rather weak statement "In Central and Western Europe, windstorms are among the major natural hazards". I am sure a more precise description of the importance can be found. For example in https://wmo.int/publication-series/atlas-of-mortality-and-economic-losses-from-weather-climate-and-water-related-hazards-1970-2021

- Introduction: There is a rather lengthy paragraph on loss datasets, but very little on different wind-damage functions. Given that the paper is on estimation of loss and particularly the steepness of the loss with increasing winds, I suggest that more emphasis is put on the describing the literature on wind-damage relations and less on the damage datasets (which was reviewed in the authors previous paper).

- Section 2:1: The LI equation is summing up grid squares without any grid square area scaling. A storm centered at 45N will have grid squares that are around 40% larger than one centered at 60N. Thus, equally sized storms will have rather different number of grid squares, making the LI values not directly comparable. This should be mentioned in the text as a possible weakness of the LI when used over regions spanning a large area in the north-south direction.

- Section 2:1: Only events with LI values above a certain threshold are kept. The authors note that they are only interested in extreme storm events and they use a threshold "which corresponds to the selection of an average of five events per season". Does this mean that if you have 41 years of data there will be 41*5=205 events (assuming that there is only one season in this study, ONDJFM)?

- Section 2.1: The post-processing of wind gusts based hourly data is difficult and needs a more detailed description. I wonder if this step is really necessary? Post-processed wind gusts based on hourly data tend to be strongly correlated to the hourly winds and have similar distribution shape. If this is the case the $(v/v\_98)^3$ ratio in the Loss Index (LI) should be almost the same if the daily max hourly wind is used instead of the wind gust data. If this is the case, there is no need to introduce a questionable wind gust parameterization, the $(v/v\_98)^3$ ratio could be calculated directly from the daily max hourly wind speed.

- Section 2.1: Make a separate section for the input data/information. (ERA, population density data, storm names etc.). It should not be described in the Meteorological loss index section.

- Section 2.2: The first paragraph is not about the Aon model or the Perils data. Consider removing it.

- Section 2.2: The Aon model is not described in much detail. Is there no more information available on the model? The paper would benefit strongly from a more detailed description of the Aon model.

- Section 2.2: Consider making a separate data section where the PERILS data can be described.

- Section 3.1: Section 3 starts with focusing is on the comparison between ERA5 and ERA-Interim 98th percentile wind gust. The 98th percentile is the loss-no loss threshold which is only marginally interesting. As the loss is increasing with $(v/v\_98)^3$ a more revealing analysis would be the difference between for example $(v\_99.7/v\_98)^3$ (99.7 percentile is the once-a-year value) or $(v\_max/v\_98)^3$. I think this would better describe how the differences between ERA5 and ERA-Interim would affect the loss calculations.

- Section 3.2: It is rather unfortunate that the losses is max-min normalized when ERA5 and ERA-Interim are compared. Is it not more revealing to do a grid square area scaling to get rid of the grid square dependence in the LI method and then look at the remaining differences in the Loss Index distributions instead of showing the rescaled versions?

- Section 3.2: The storm losses and storm ranking comparison is done using Pearson and rank correlations. For the Pearson correlation to be informative the data need to be normally distributed. It is pretty clear that the loss data is heavily skewed and the Pearson correlation becomes pretty meaningless or at least very hard to interpret. The R^2 will not be the variance the two datasets have in common. Either the data must be transformed to become normally distributed (Box-Cox or similar) or other measures of similarity should be applied. It also seems that the authors have used the paired Wilcoxon Signed-Rank Test on the Pearson correlations. But as the Pearson correlation assumes normal distributed data, the t-test would be the appropriate test.

- Section 3.2: The authors note that they are "ranking for the 20 common most extreme storms", but it is not necessarily the most extreme storms, but the one with largest loss values.

- Section 3.2: How is the "20 common most extreme storms" found and how will this affect the ranking analysis? If a common storm is no. 18 in one dataset and no. 24 in another, but picked as one of the 20 common most extreme storms, will they be reranked (between 1 and 20) and the ranking analysis based on the reranked values?

If they are based on reranked values, the results may not be representative for the original datasets.

- Section 3.2: The authors state that the LI values based on ERA5 are approximately 10 times larger than for ERA-Interim due to smaller grid squares. To highlight possible other reasons, the LI values could be scaled with grid square area, to investigate if any systematic differences that was not due to the obvious grid square size dependence.

- Section 3.2: Storm loss rankings are based on shared variance of the ranked scores (rank correlation squared). It is not mentioned in the text, but I guess this is based on the Spearman rank correlation?

- Section 3.2: One might wonder if the correlation analysis really is the way is to go for showing differences in the loss estimates? I have a hard time understanding what a R^2 of 0.5 really means. According to the paired Wilcoxon Signed-Rank Test correlations squared down to 0.27 and rank correlations squared down to 0.18 does not indicate significant differences between LI ERA5 and LI ERA-Interim. Given the low number of events and possibly a few events that are very different in the two sets, there might be other methods than correlation analysis that are more useful and easier to understand. Maybe categorizing the losses, Kendal tau, Rank Biased Overlap, Goodman and Kruskal's gamma, precision etc. or other methods can be considered.

- Section 4.1: The same comments on the use of Pearson correlation etc. mentioned in section 3.2 applies on this analysis.

- Section 4.1: I find it rather hard to get an overview of the differences between the LI and Aon estimates only based on the scatter plots and correlations. A possibility is to categorize the result (for example low, medium, high losses) and do a more in-depth analysis of the differences within each category. A table summarizing the national losses as low, medium and high losses based on the model's normalized losses and then show statistics of the proportion of events where the models agree on the same category in each country or some other summarizing statistics beyond correlations would help.

- Section 4.1: Figure 6 clearly shows how the LI estimates are smaller than the Aon estimates for high loss events. But it is not clear how much smaller. The LI estimates is proportional to $(v/v\_98)^3$ but experimenting with other exponents $(v/v\_98)^n$ with $n>3$ in the LI equation would tell us how much the cubic assumption in the LI formulation has to be adjusted for the results to be in line with the Aon results for high loss values. This will inform the reader about the level of adjustment needed for the LI formulation to approach the Aon model results.

- Section 5: The conclusion "Compared to Aon's IF Euro WS model, LI ERA5 shows overall lower loss values" cannot be drawn from the analysis. The loss values are not comparable. Aon models' monetary loss and the LI is just a loss index. The max-min scaling, rescales the values, but the underlying original values are still not comparable.

- Section 5: The conclusion “the Aon model seems to better distinguish between high and moderate impact events” is not justified by the analysis. As the models are not compared to reality we do not know if the Aon model does a “better” job in distinguishing the events, we just know that it separates the loss values between the different events more than the LI estimates.

- Figure 8: Adjust colour scale to better distinguish the different values.

- Section 5: The summarizing list of findings is rather unprecise. Wording such as “comparable behaviour”, “slightly shifted”, “ranks are comparable” are not very informative.

- Section 5: The authors mention the 72-hour event definition in LI as a possible source of the differences between LI and Aon. Could this be investigated by changing the 72-hour event definition in LI to a 24 hours?

- Section 5+Abstract: I fail to see those conclusions on quality such as “… the loss distribution in LI is not steep enough …” can be justified from the current analysis. What is shown is that it is less steep than the loss distribution of the Aon model. There is no comparison against reality in the paper so we cannot know if it steep enough or not. Recent windstorms like Ciarán/Emir (2023), that is not used for calibration in Aon’s model could have been used to shed light on the quality of the loss estimates.

- Section 5: The authors states that the “LI index is missing a detailed damage component, thus it struggles to capture the non-linear response of the buildings at the tail of the gust spectrum for the high impact events.” Is not non-linearity what the cubic relation in the LI expression is trying to achieve? The way I see it, is that it is not the lack of non-linearity, but that the non-linearity is less strong than in the Aon model.

- Section 5: A main conclusion for the LI estimates is that “Although it cannot be used to price a storm (due to the missing vulnerability information), it is suitable for estimating the impacts and rank events.” It is not clear why it is judged as “suitable”. What was the benchmark for the suitability conclusion. What was needed for the estimates to be judged as unsuitable and how do we know it is suitable when it is not compared to reality?