We thank the reviewer for his/her detailed and constructive comments, which helped to improve the manuscript and clarify key points. Below are point-to-point responses to each comment, including plans how to incorporate them in our manuscript.

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.

Answer: We agree with the reviewer that a more detailed analysis would be important to strengthen the scientific values of the manuscript and in particular to understand how comparable LI and the Aon IF model output are. We will include a detailed case study for recent storm Sabine (February 2020) in the updated manuscript, in which we compare LI and the Aon IF model output, and add aggregated market losses from the PERILS data as a benchmark. Figure RC2.1 shows both the normalized losses and the storm ranking at country level for storm Sabine for the three datasets. In the revised manuscript, we will replace Figure 5 with Figure RC2.1. Additionally, we will extend Figure 8 (Spearman rank correlation) by adding PERILS to the comparison (see Figure RC2.2).

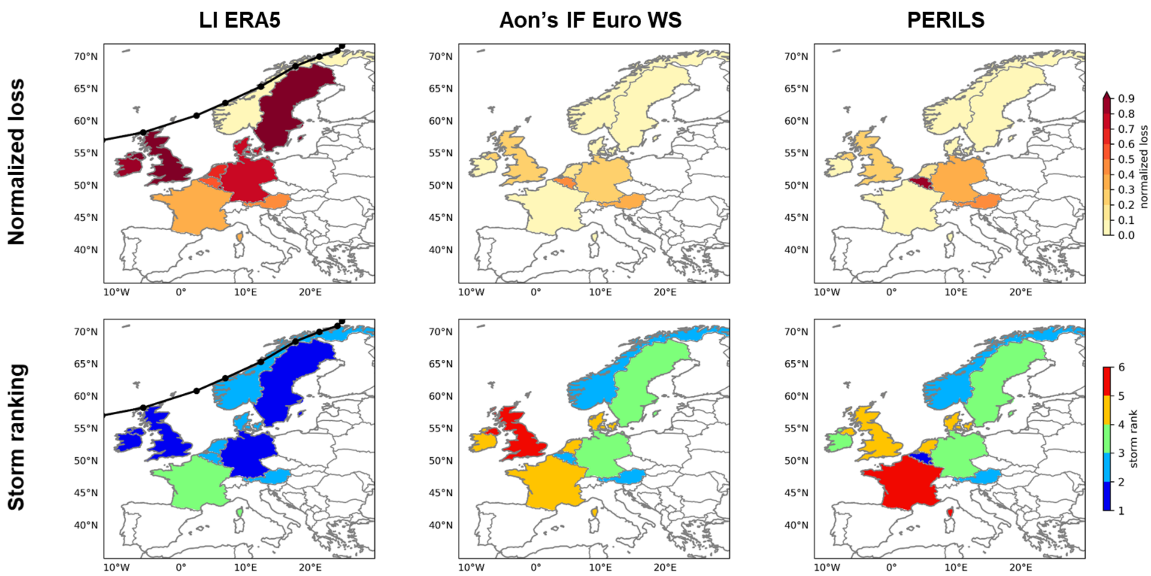

Figure RC2.1: *Normalized losses (upper row) and storm ranking (lower row) at country level for storm Sabine in February 2020. Losses are derived from LI ERA5 (left), Aon's IF Euro WS model (middle), and PERILS (right). The black line and dots in the left column denote the cyclone track derived from ERA5 using the tracking algorithm of Pinto et al. (2005). Losses are only shown for the 11 countries covered by Aon. The ranking is based on common storms.*

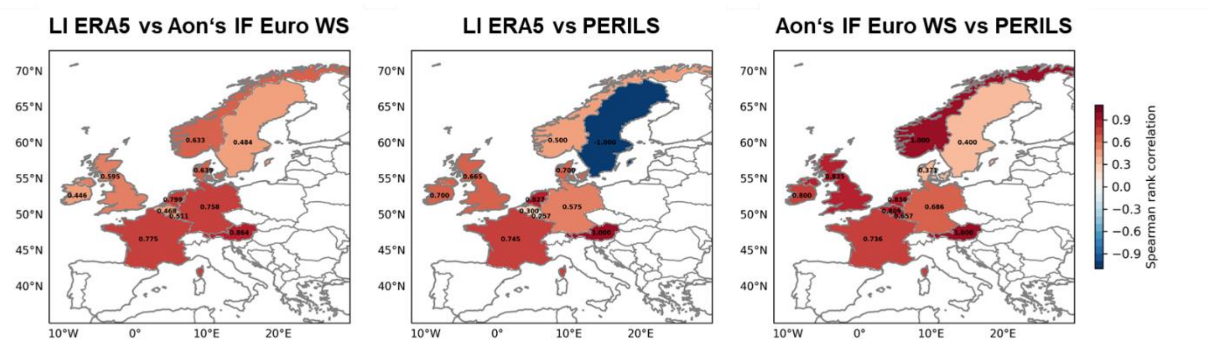

Figure RC2.2: *Spearman’s rank correlation coefficient at country level for LI ERA5 vs Aon’s IF Euro WS model (left), LI ERA5 vs PERILS (middle), and Aon’s IF Euro WS model vs PERILS (right). The ranking is based on common storms per country.*

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.

Answer: We agree with the reviewer in this point. We will provide a more detailed description (roughly 10 pages) of the Aon IF windstorm model, including a description of the hazard, vulnerability and exposure components, which will be included in the Supplementary. Based on this detailed description, we will also enhance the discussion on the methodological differences between the two approaches in the revised manuscript.

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.

Answer: We thank the reviewer for this comment. We will adapt the abstract so that it better matches the title of our paper.

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.

Answer: We will justify some of the statements by including actual 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

Answer: We started the introduction with a rather general statement to address a broad audience, including a more precise description in the following sentences. Therefore, we would like to keep the statement as it is.

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).

Answer: We appreciate this comment and the suggestion to focus more on wind-damage relations. We will shorten the paragraph on damage datasets and expand the part on damage functions.

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.

Answer: We agree with the reviewer that LI can depend on the size of the grid squares. However, we believe this effect to be negligible for our study, as we mainly focus on individual countries and Core Europe (which has no large expansion in north-south direction). Nevertheless, we will add a short note on this effect in the revised manuscript.

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)?

Answer: Yes, the reviewer is correct in this point. We will clarify it in the updated manuscript.

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.

Answer: We believe the reviewer may have misunderstood us here. We did not apply a wind gust parameterization ourselves. Instead, we use the officially published wind gust data from ERA5 and ERA-Interim, which is based on the parameterization approach by Panofsky et al. (1977) and Bechthold & Bidlot (2009). We will clarify this in the revised manuscript.

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.

Answer: We thank the reviewer for this comment. We will improve the structure of section 2 by clearly separating the data description from the methods.

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

Answer: The paragraph is intended as a general introduction to loss models. We would therefore like to keep it as it is.

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.

Answer: Please refer to our reply on the general comments above.

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

Answer: We will improve the structure of section 2 by clearly separating the data description from the methods.

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.

Answer: We thank the reviewer for this comment. Figure RC2.3 shows an extension of Figure 2, additionally including the 99.9th percentile for both datasets as well as the difference. The differences between ERA5 and ERA-Interim are larger for higher percentiles – suggesting a longer tail of the wind gust distribution for ERA5, which could result in overall higher LI values. We will replace Figure 2 by Figure RC2.3 and enhance the discussion in the revised manuscript.

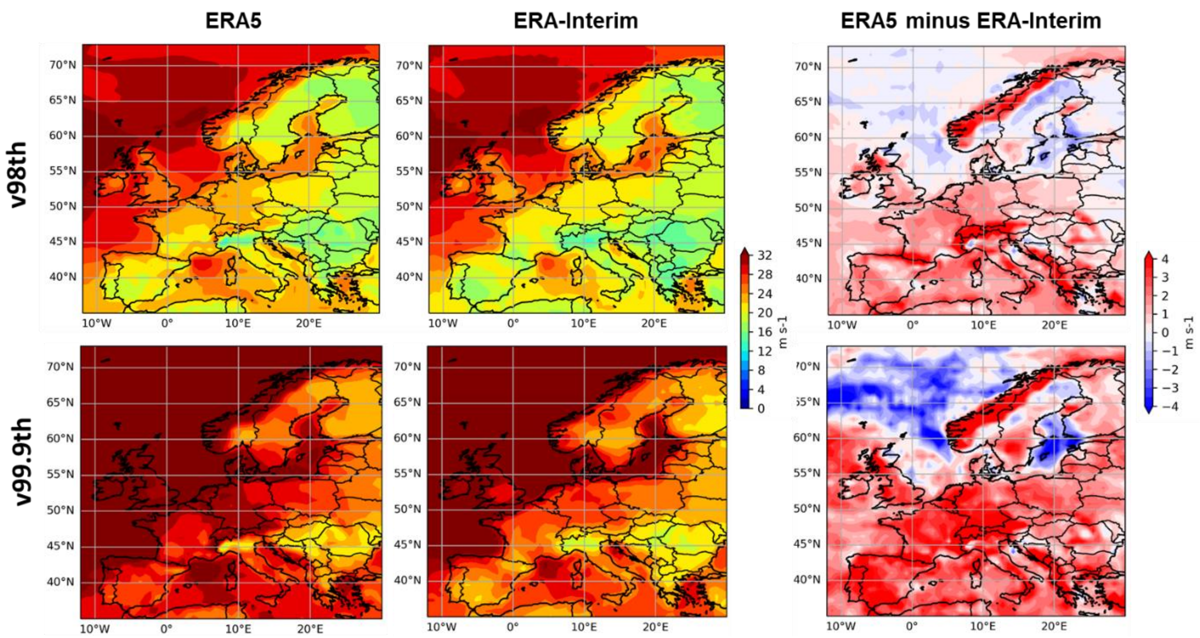

Figure RC2.3: *98th percentile (upper row) and 99.9th percentile (lower row) of daily maximum wind gust for the winter half year (October – March, ONDJFM) for the period 1979-2019.*

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?

Answer: The original (not normalized) loss values are shown in Supplementary Figure S2. We decided to keep the normalized values in the main paper in order to be consistent with Figure 6 in section 4.1.

We tested the scaling effect of the different spatial resolutions by re-gridding ERA5 to the ERA-Interim grid before calculating LI. Figure RC2.4 shows the LI for ERA5 re-gridded to the ERA-Interim resolution for different regions, compared to the original Figure S2. After re-gridding, LI ERA5 and LI ERA-Interim are in the same order of magnitude, while the overall behavior/order of storms does not change (as can be seen by the small changes in R^2). This confirms that the different resolution of the datasets is not decisive – and that differences may most likely result from differences in the wind gust distribution (see also previous comment). We will add the right part of Figure RC2.4 to the Supplementary and enhance the discussion in the revised manuscript.

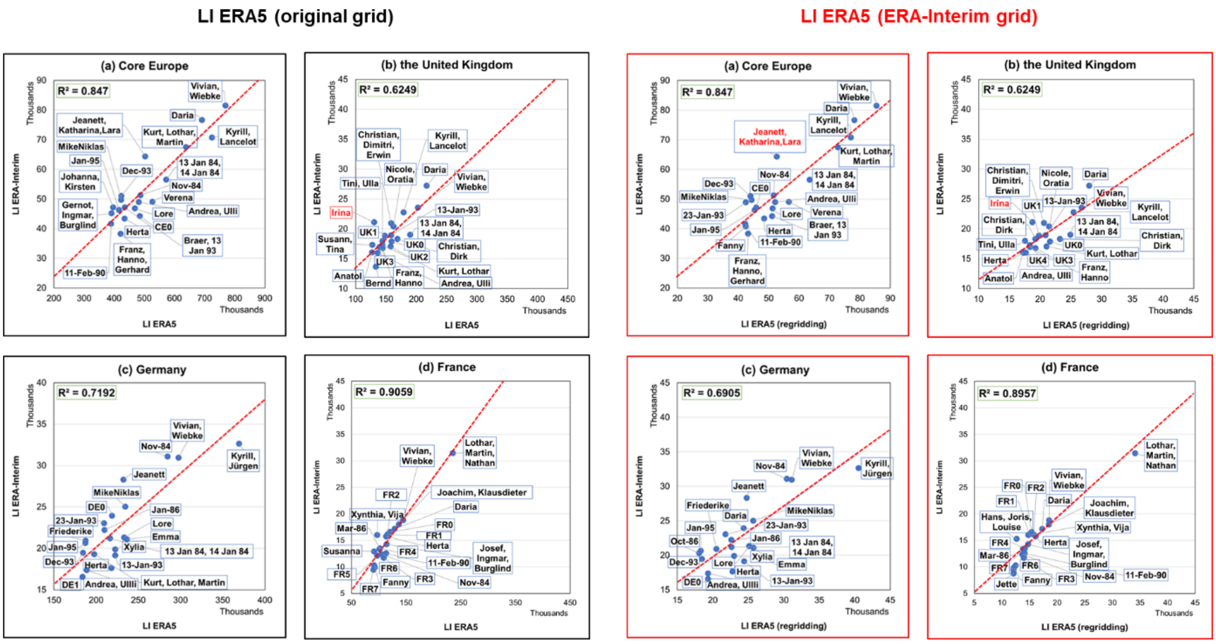

Figure RC2.4: Comparison of loss values (in thousands) based on LI ERA5 (x-axis) and LI ERA-Interim (y-axis) for the common 20 most extremes storms in the period 1979-2019. LI ERA5 is calculated from the original ERA5 gust data (left) and from the ERA5 gust data re-gridded to the ERA-Interim resolution (right).

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 R2 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.

Answer: We believe the reviewer may have misunderstood us here. We are not using Pearson correlation, but only the coefficient of determination R^2 that directly relates to the linear regression lines as shown in Figures 3, 4, 6 & 7. The larger R^2 is, the greater the linear relation between the two datasets. However, in order to avoid misunderstandings, we will merge Tables 1 and 2, and will additionally replace the R^2 with the one calculated based on Spearman's rank correlation coefficient (see Table RC2.1). Please also refer to one of the following comments.

Table RC2.1: *R^2 of Spearman's rank correlation coefficient between LI ERA5 and LI ERA-Interim (2nd column), LI ERA5 and Aon's IF Euro WS model (3rd column), LI ERA5 and PERILS (4th column), and Aon's IF Euro WS model and PERILS (last column). The number of common storms per country is given in brackets.*

	LI ERA5 vs LI ERA-Interim	LI ERA5 vs Aon's IF Euro WS	LI ERA5 vs PERILS	Aon's IF Euro WS vs PERILS
Core Europe	0.65 [20]	0.52 [23]	0.26 [17]	0.57 [19]
Austria	0.43 [20]	0.75 [15]	1.0 [4]	1.0 [4]
Belgium	0.62 [20]	0.22 [21]	0.09 [11]	0.66 [11]
Denmark	0.25 [20]	0.41 [15]	0.49 [5]	0.14 [6]
France	0.79 [20]	0.6 [17]	0.56 [10]	0.54 [11]
Germany	0.5 [20]	0.57 [23]	0.33 [15]	0.47 [15]
Ireland	0.37 [20]	0.2 [19]	0.49 [5]	0.64 [5]
Luxembourg	0.64 [20]	0.26 [15]	0.07 [6]	0.43 [6]
Netherlands	0.2 [20]	0.64 [21]	0.68 [11]	0.7 [11]
Norway	0.29 [20]	0.4 [9]	0.25 [3]	1.0 [3]
Sweden	0.51 [20]	0.23 [13]	1.0 [4]	0.16 [4]
United Kingdom	0.49 [20]	0.36 [20]	0.44 [13]	0.7 [13]

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.

Answer: As the whole study deals with windstorm losses, we use most extreme in the sense of loss. This does not necessarily mean extreme in terms of wind gust, core pressure minimum or other. We will state this clearly in the updated text.

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.

Answer: In general, the storm ranking depends on the total number of storms per dataset (Moemken et al., 2024). This is particularly relevant for the later comparison between LI and the Aon IF model. Therefore, we decided to first select the 20 common storms and rank only these events. However, in order not to lose the information from the original datasets, we will add the original rank of the events in Tables S2 and S3 in the revised manuscript.

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.

Answer: Please refer to one of the previous comments and Figure RC2.4.

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?

Answer: We are not completely sure that we understand this comment. Is the reviewer asking which correlation we use for the values in Table 1? In the current version, the values in Table 1 correspond to the coefficient of determination R^2 from the linear regression in Figures 3 and 4. Based on the other comments and in order to avoid misunderstandings, we will change this in the revised manuscript and use R^2 values from the Spearman's rank correlation (see Table RC2.1).

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 R2 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.

Answer: Our study is the first to compare a full insurance windstorm model (which is not available publicly) to a simplified meteorological loss index. Therefore, we decided to focus on a straightforward comparison of the two methods. With this aim, we decided to use the Spearman R and R^2 coefficients, which allows us to focus on simple but robust conclusions and to avoid a mix of different statistical methods.

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

Answer: Please refer to our reply on the comment regarding section 3.2

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.

Answer: We agree with the reviewer that there may be merit in the idea of investigating the different behavior of high and low loss events. However, as the sample size is already quite small (see Table RC2.1), we might have issues with splitting the data into even smaller samples. Therefore, we decided to analyze the sample as a whole.

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.

Answer: The LI method used in our study is well established. Based on the original approach developed for station data from Klawa & Ulbrich (2003), it was further developed by Pinto et al. (2012), and several formulations (also with other exponents) were tested. The same was done in other studies such as Prahl et al. (2015). All these studies agree that the performance of the different indices depends on the underlying event set. For some storm events, formulations with higher exponents seem to better suit to realistically estimate windstorm losses, while for other events, the cubic relationship provides results that are more realistic. In this sense, and based in our experience, no formulation clearly outperforms the others. Since our study is the first to compare a full insurance windstorm model (which is not available publicly) to a simplified meteorological loss index, we focus on a straightforward comparison of the two methods (as mentioned before). Therefore, in our opinion, the objective should not be to experiment with the LI formulation. Nevertheless, we will provide a more detailed discussion of the impact of the LI setup on the results in the revised manuscript.

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.

Answer: We agree that the original loss values of LI and the Aon IF model are not comparable. Based on the normalized losses, however, we do think we are able to draw conclusions such as the one referred to in the comment. Nevertheless, we will rephrase the conclusion to avoid confusion.

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.

Answer: We agree with the reviewer in this point. The Aon IF model is calibrated against the PERILS dataset, thereby it can be assumed as a representation of a “market-reality” for the purpose of this paper. Nevertheless, we decided to use none of the datasets as ground truth (Moemken et al., 2024). We will carefully go through section 5 and rephrase sentences where necessary.

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

Answer: Thanks for pointing this out. We will adjust the color scale in the revised manuscript.

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

Answer: We will make the summary more precise in the revised paper.

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?

Answer: We appreciate this suggestion. We did a sensitivity analysis for different time windows, e.g. calculating LI for 24-hour windows. Figure RC2.5 shows the results for the normalized losses and figure RC2.6 for the storm ranking, respectively. Overall, we find no systematic reduction in the differences between LI and the Aon IF model output when using 24 hours instead of 72 hours. For some storms and/or countries, the correlations increase with a shorter event definition (see e.g. Germany), while for others they decrease (see e.g. Core Europe). In addition, the number of common storms decreases when using 24-hour windows for the LI calculation (not shown). Therefore, we decided to keep the focus of our study on the 72-hour event definition. This has several advantages: First, we are able to capture the entire storm footprint; second, this is in line with the standard practice in insurance industry (the so-called 72-hour-clause); third, the correlations between LI and the Aon IF model are higher, especially for Core Europe. Nevertheless, we will include both figures in the Supplementary and expand the discussion in the revised manuscript.

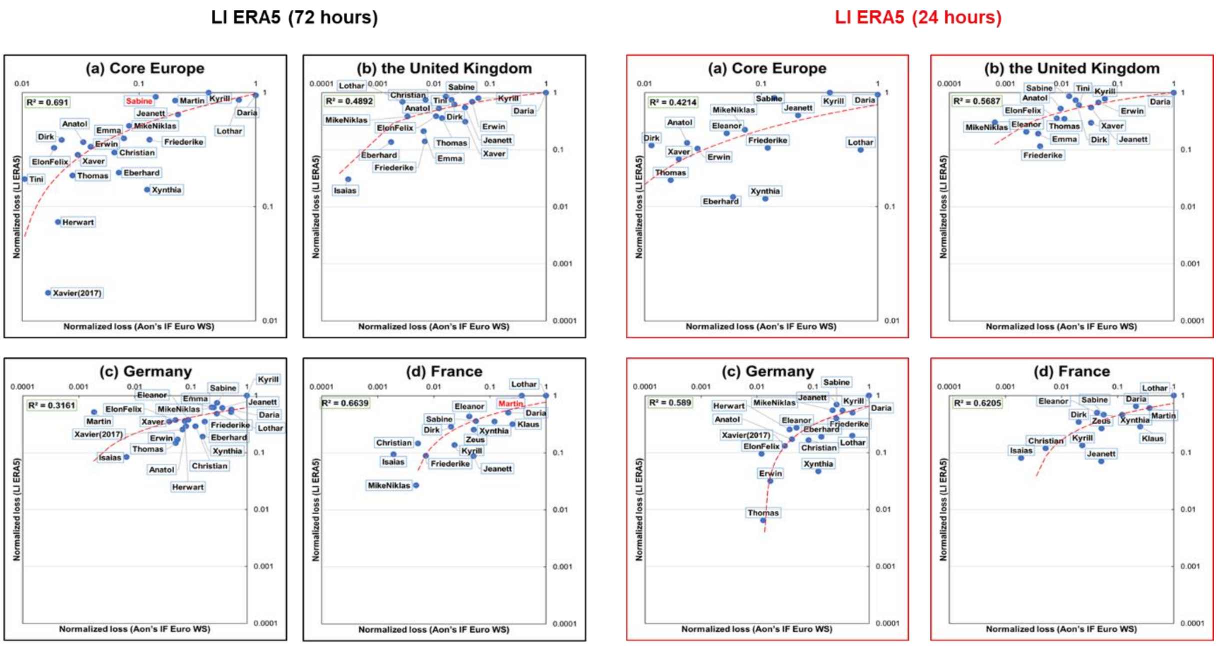

Figure RC2.5: *Comparison of normalized loss values between Aon’s IF Euro WS model (x-axis) and LI ERA5 (y-axis). Depicted are the common most extreme storms for the period 1990-2020 for (a) Core Europe, (b) the United Kingdom, (c) Germany, and (d) France. A logarithmic scale is used for the axes. The red dashed line denotes the logarithmic regression. The correlation between the datasets is given in the upper left corner (R^2 value). Outlier storms based on the IQR method are marked in red. LI ERA5 is calculated for 72-hour windows (left) and 24-hour windows (right).*

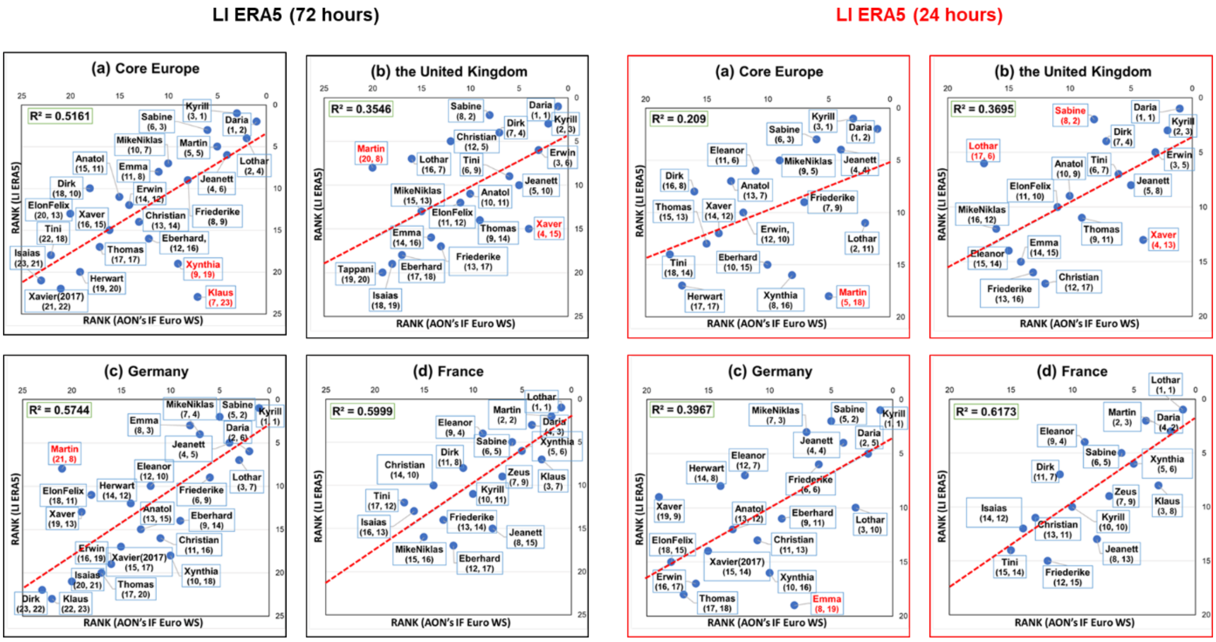

Figure RC2.6: *Same as Figure RC2.5, but for the comparison of storm ranks. The values in brackets indicate the rank (first value Aon's model, second value LI ERA5).*

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.

Answer: We agree with the reviewer that it can be complex to explain some of our conclusions from the analysis. The Aon IF model is calibrated against real loss data, using the PERILS data as the primary benchmark. Therefore, all storms considered in our analysis are calibrated/validated against a "market-reality". We will clarify this in the revised manuscript. Additionally, we will include the PERILS data in some of the analysis (see above).

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.

Answer: We agree with the reviewer and will reword the statement in the revised manuscript (see also previous comment).

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?

Answer: We agree with the reviewer in this point. We will reword the statement in the revised manuscript – taking into account the fact that we propose to use the Aon IF model output as a representation of a "market-reality" for the purpose of this study (see also the second last comment).

References

Bechtold, P. and Bidlot, J. R.: Parametrization of convective gusts, ECMWF Newsletter, 199, 15-18, https://doi.org/10.21957/kfr42kfp8c, 2009

Klawa, M. and Ulbrich, U.: A model for the estimation of storm losses and the identification of severe winter storms in Germany, Nat. Hazards Earth Syst. Sci., 3, 725-732, https://doi.org/10.5194/nhess-3-725-2003, 2003

Moemken, J., Messori, G., and Pinto, J. G.: Windstorm losses in Europe – What to gain from damage datasets, Weather and Climate Extremes 44, 100661, https://doi.org/10.1016/j.wace.2024.100661, 2024

Panofsky, H. A., Tennekes, H., Lenschow, D. H., and Wyngaard, J. C.: The characteristics of turbulent velocity components in the surface layer under convective conditions, Bound.-Lay. Meteorol., 11, 355-361, https://doi.org/10.1007/BF02186086, 1977

Pinto, J. G., Karremann, M., Born, K., Della-Marta, P., and Klawa, M.: Loss potentials associated with European windstorms under future climate conditions, Climate Res., 54, 1-20, https://doi.org/10.3354/cr01111, 2012

Prahl, B. F., Rybski, D., Burghoff, O., and Kropp, J. P.: Comparison of storm damage functions and their performance, Nat. Hazards Earth Syst. Sci., 15, 769-788, https://doi.org/10.5194/nhess-15-769-2015, 2015