Dear Ricardo Trigo,

We would like to submit a revised version of our manuscript "*Insurance loss model vs meteorological loss index – How comparable are their loss estimates for European windstorms?*" to NHESS.

We thank the three reviewers and you as the editor for the detailed and valuable comments, which greatly helped to improve the manuscript. Below are point-to-point responses to each of the comments.

While revising the manuscript, we noticed that some storms were depicted incorrectly/inconsistently in Figures 3, 4, 7 and 8, e.g. the names for storms Jeanett and MikeNiklas were not consistent. We have corrected the respective figures in the revised manuscript. We would like to emphasize that this correction does not change the interpretation of the results in any way.

We look forward to your decision on our manuscript.

Yours sincerely,
Julia Mömken (on behalf of all co-authors)

Reviewer #1 (Gerard van der Schrier)

We thank Gerard van der Schrier for his comments and suggestions, which helped to improve the manuscript and to remove ambiguities/misunderstandings. Below are point-to-point responses to each comment, including references to the amended text in the revised manuscript.

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 assesses 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.

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). In the present study, we compare for the first time a simple meteorological index to the output of a full insurance windstorm model. Therefore, the assessment of the quality of wind gust parameterization itself is not the aim in our study. However, we would like to note that we have performed such assessments before in dedicated studies, e.g. by comparing different wind gust estimation methods (Born et al., 2012) and by comparing wind gust data from reanalysis to station observations (Seregina et al., 2014). Other authors (e.g. Minola et al., 2020) have also performed such analysis. We clarified this in the revised manuscript (lines 84-87) and included some discussion on the effect of wind gust parameterization on the differences between LI and the Aon IF model output (lines 396-398).

Other aspects

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.

Answer: The spatial resolution of the different datasets is now clearly stated in the Data section – in lines 82-84. We agree with the reviewer's comment that the effect of the different spatial resolutions can be important. Thus, we have decided to add some analysis (and corresponding discussion) with using ERA5 re-gridded to the ERA-Interim grid. See also detailed comment further below.

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 footprint = 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?

Answer: We thank the reviewer for this comment. The footprint for Irina is overall flatter in ERA5 compared to ERA-Interim. This is particularly the case for the UK, where the mean wind gust over land is 12.1 m/s for ERA5 and 24.6 m/s for ERA-Interim. Therefore, the LI for ERA-Interim is higher due to the cumulative effect (summation of v/v_{98}). Additionally, Figure R1 shows an extension of Figure S1 including another panel showing the ERA5 footprint re-gridded to the ERA-Interim resolution. This confirms the overall flatter structure of footprint. We agree this is an important point and we added more details (lines 227-230) and Figure R1 (as new Figure S1) in the revised manuscript.

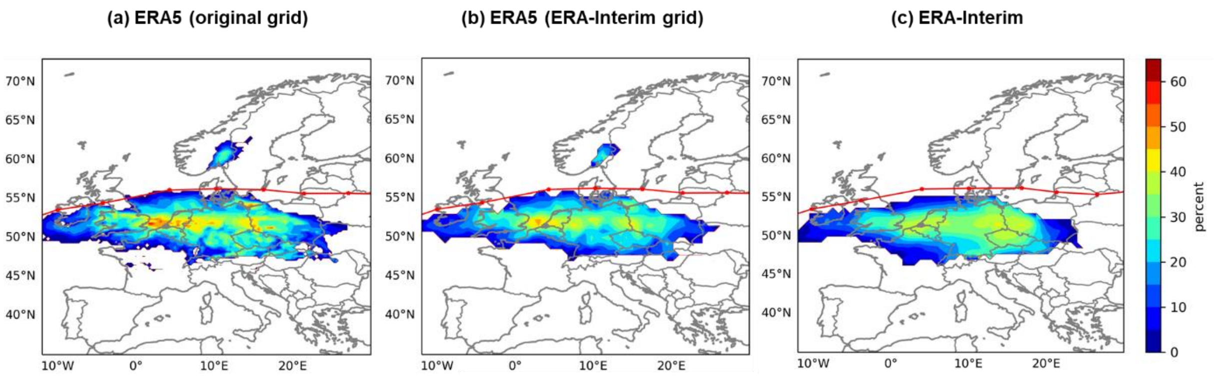

Figure R1: *Wind gust footprint for storm Irina in October 2002 based on ERA5 (a), ERA5 re-gridded to the ERA-Interim grid (b), and ERA-Interim (c). Shown is the largest exceedance (in percent) of the 98th percentile of daily maximum wind gust within 72 hours. The red line and dots denote the cyclone track derived from ERA5 (a, b) and ERA-Interim (c) using the tracking algorithm of Pinto et al. (2005).*

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).

Answer: We thank the reviewer for this comment. Figure R2 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 Figure R3). We added the right part of Figure R2 to the Supplementary (new Figure S3) and enhanced the discussion in the revised manuscript (lines 240-245).

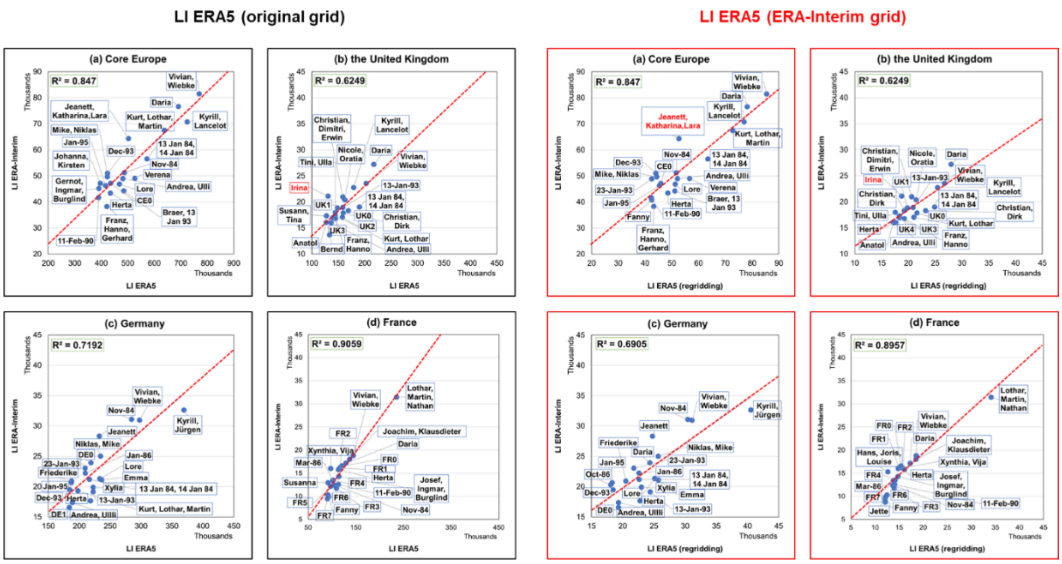

Figure R2: *Comparison of loss values (in thousands) based on LI ERA5 (x-axis) and LI ERA-Interim (y-axis) for the common 20 most extremes storms in the period 1979-2019. LI ERA5 is calculated from the original ERA5 gust data (left) and from the ERA5 gust data re-gridded to the ERA-Interim resolution (right).*

Section 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.

Answer: The reviewer is correct here – a shift in the wind gust distribution also implies a shift in the 98th percentile and therefore cannot explain the higher LI values for ERA5 alone. What we meant is that besides the overall shift in the distribution, the tail of the distribution is longer for ERA5, which then leads to higher LI values. To illustrate this, Figure R3 shows both the 98th and the 99.9th percentile for ERA5 and ERA-Interim as well as the difference. The figure clearly shows larger differences for higher (99.9th) percentiles, thereby confirming the longer tail of the distribution in ERA5. We apologize for the misunderstanding. We replaced Figure 2 by Figure R3 and clarified this point in the revised paper (lines 208-211 and 355-356).

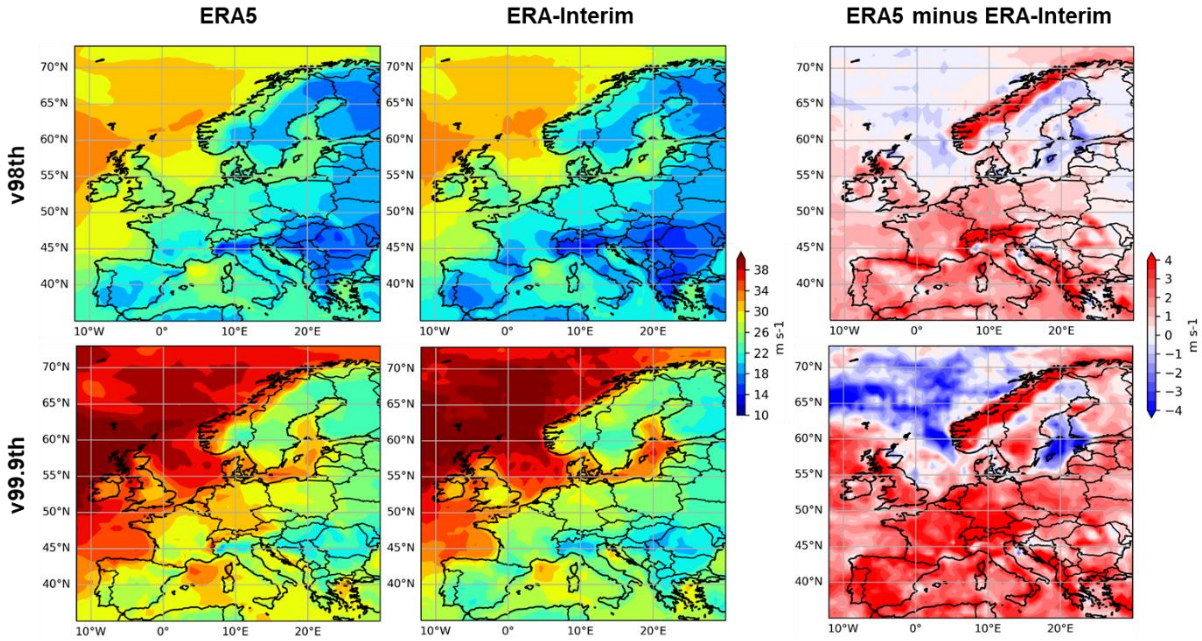

Figure RC3: *98th percentile (upper row) and 99.9th percentile (lower row) of daily maximum wind gust for the winter half year (October – March, ONDJFM) for the period 1979-2019.*

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.”

Answer: We apologize for the misleading figure caption. We changed it to “Shown is the largest exceedance (in percent) of the 98th percentile of daily maximum wind gust within 72 hours” in the revised manuscript.

Reviewer #2

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 references to the amended text in the revised 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 now include a detailed case study for recent storm Sabine (February 2020) in the updated manuscript (new section 5.1), in which we compare LI and the Aon IF model output, and add aggregated market losses from the PERILS data as a benchmark. Figure R4 shows both the normalized losses and the storm ranking at country level for storm Sabine for the three datasets. In the revised manuscript, we removed Figure 5 and added Figure R4 as new Figure 6. Additionally, we extended Figure 8 (Spearman's rank correlation; now Figure 9) by adding PERILS to the comparison (see Figure R5).

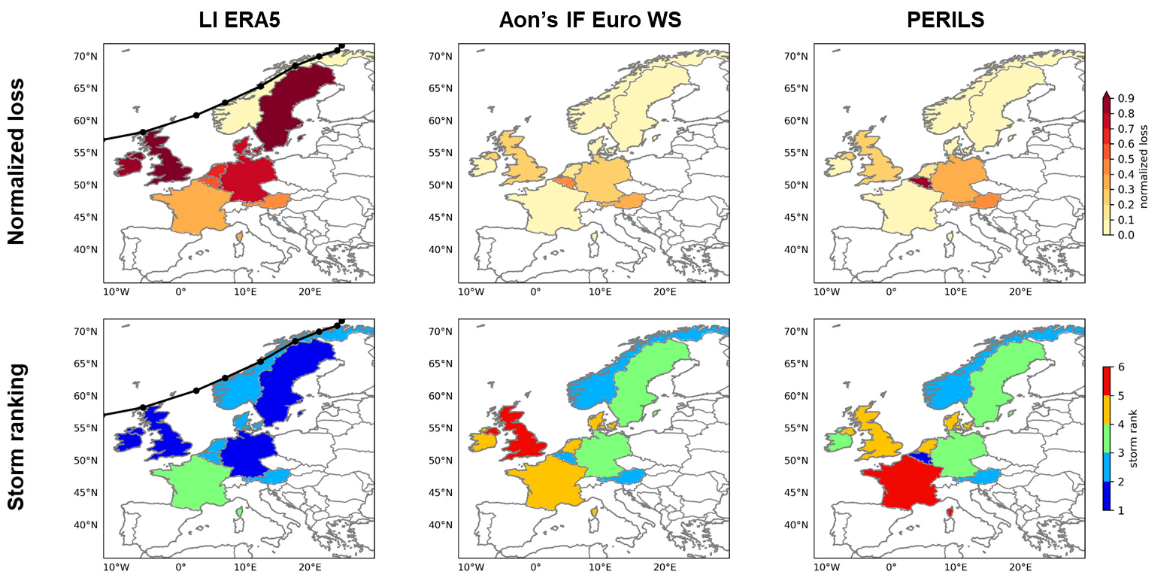

__Figure R4:__ Normalized losses (upper row) and storm ranking (lower row) at country level for storm Sabine in February 2020. Losses are derived from LI ERA5 (left), Aon's IF Euro WS model (middle), and PERILS (right). The black line and dots in the left column denote the cyclone track derived from 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.

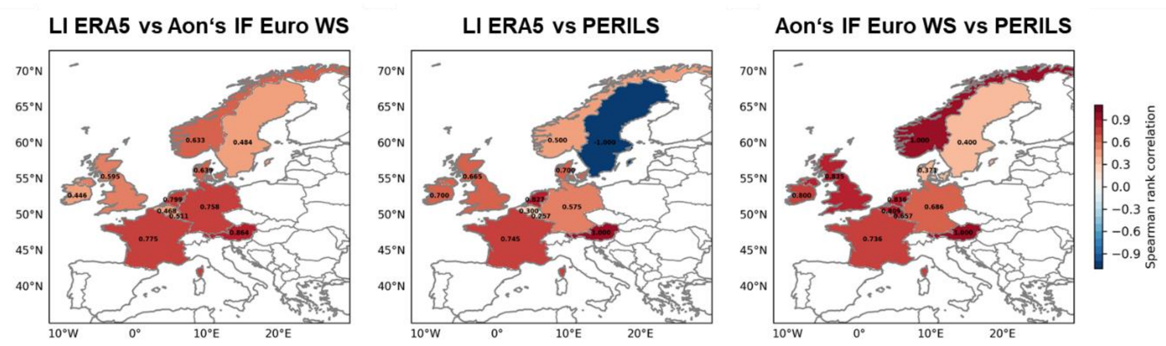

Figure R5: *Spearman's rank correlation coefficient at country level for LI ERA5 vs Aon's IF Euro WS model (left), LI ERA5 vs PERILS (middle), and Aon's IF Euro WS model vs PERILS (right). The ranking is based on common storms per country.*

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 now provide a more detailed description (roughly 15 pages) of the Aon IF windstorm model, including a description of the hazard, vulnerability and exposure components, which is included in Supplementary Part A. Based on this detailed description, we also enhanced the discussion on the methodological differences between the two approaches in the revised manuscript (lines 375-410).

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 adapted 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 justified 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 shortened the paragraph on damage datasets and expanded the part on damage functions (lines 40-53).

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).

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 clarified it in the updated manuscript (line 141).

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 clarified this in the revised manuscript (lines 84-87).

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 improved the structure of the revised manuscript by clearly separating data (new section 2) and methods (new section 3).

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 improved the structure of the revised manuscript by clearly separating data (new section 2) and methods (new section 3).

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 R3 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 replaced Figure 2 by Figure R3 and enhanced the discussion in the revised manuscript (lines 208-211).

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 7 in section 5.2.

We tested the scaling effect of the different spatial resolutions by re-gridding ERA5 to the ERA-Interim grid before calculating LI. Figure R2 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 added the right part of Figure R2 to the Supplementary (Figure S3) and enhanced the discussion in the revised manuscript (lines 240-245).

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 original 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 decided to merge Tables 1 and 2, and additionally replace the R^2 with the one calculated based on Spearman's rank correlation coefficient (see Table R1). Please also refer to one of the following comments.

Table R1: *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. This is now clearly stated in the updated text (lines 73-74).

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 added 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 R2.

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 previous 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 changed this in the revised manuscript and use R^2 values from the Spearman’s rank correlation (see Table R1).

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 R1), 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 now provide a more detailed discussion of the impact of the LI setup on the results in the revised manuscript (lines 399-407).

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 rephrased the conclusion to avoid confusion (line 359).

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 perspective for the purpose of this paper. Nevertheless, we decided to use none of the datasets as ground truth (Moemken et al., 2024). We carefully went through section 5 (now section 6) and rephrased sentences where necessary (e.g. lines 399-404 and 413-415).

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

Answer: Thanks for pointing this out. We adjusted 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 made the summary more precise in the revised paper (lines 354-364).

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 R6 shows the results for the normalized losses and figure R7 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 included both figures in the Supplementary (new Figures S5 and S6) and expanded the discussion in the revised manuscript (lines 316-322 and 378-384).

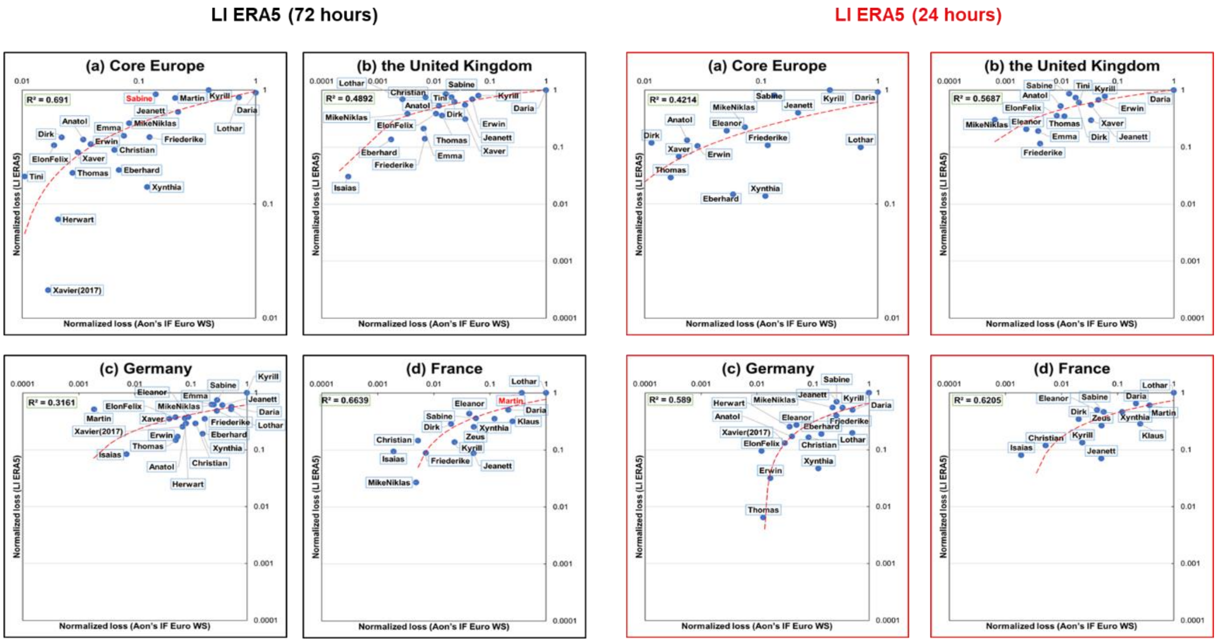

Figure R6: *Comparison of normalized loss values between Aon's IF Euro WS model (x-axis) and LI ERA5 (y-axis). Depicted are the common most extreme storms for the period 1990-2020 for (a) Core Europe, (b) the United Kingdom, (c) Germany, and (d) France. A logarithmic scale is used for the axes. The red 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).*

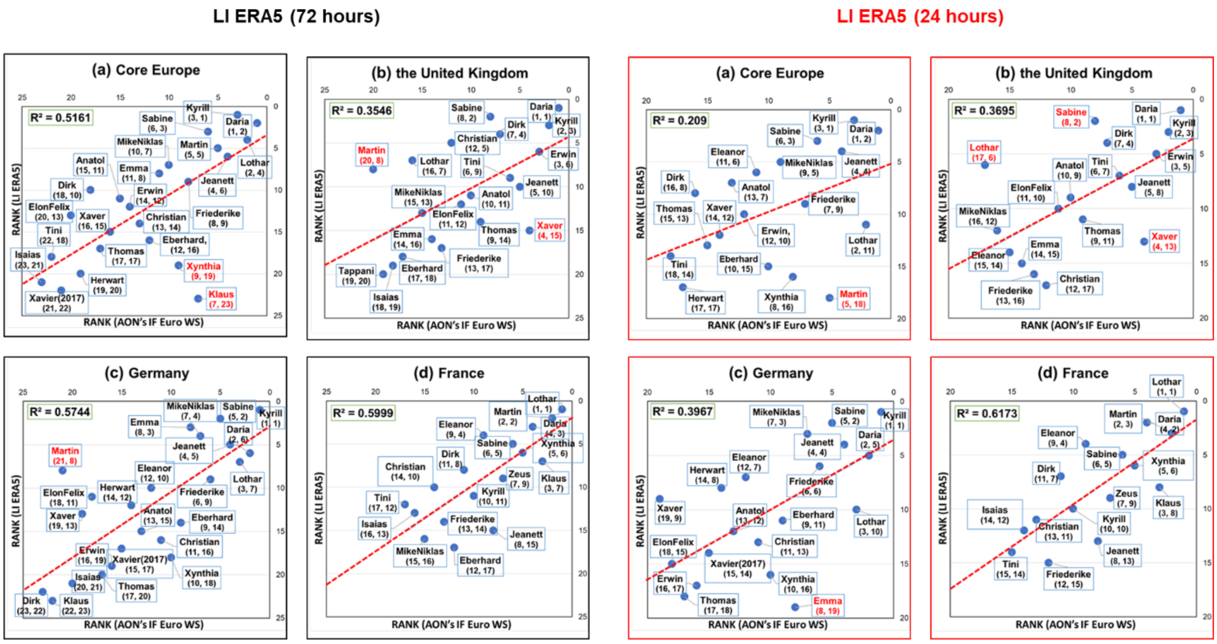

Figure R7: *Same as Figure R6, but for the comparison of storm ranks. The values in brackets indicate the rank (first value Aon's model, second value LI ERA5).*

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 perspective. We clarified this in the revised manuscript (lines 192-194). Additionally, we included 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 reworded 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 reworded the statement in the revised manuscript – considering the fact that we propose to use the Aon IF model output as a representation of a market perspective for the purpose of this study (see also the second last comment).

Reviewer #3

We thank the reviewer for his/her constructive comments, which helped to improve the manuscript and clarify important points. Below are point-to-point responses to each comment, including references to the amended text in the revised manuscript.

General comments

Given the high importance of natural catastrophes for society, especially in the light of climate change, this topic is clearly within the scope of NHESS. The paper is well-structured and is clearly written. The research questions are clearly posed in the final part of the introduction. The paper has the potential to shed light on the performance of a simple loss index which can be applied to both re-analysis data sets and climate model outputs and the dependency on e.g., spatial resolution. Furthermore, how well such a simple approach can be used to benchmark more refined commercial cat models. This can ultimately help to increase resilience by better understanding of past and future risk of European windstorms.

Overall, the analysis performed appear to be a bit simplistic, though, with mainly correlation and scatter plot analysis between the two meteorological loss indices and the losses from the AON cat model. Most of the analysis do not shed light on the cause of the differences and especially on the quality of the approaches. To answer the research question how comparable and how sensitive the approaches are, more in-depth and refined approaches are recommended.

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 now include a detailed case study for recent storm Sabine (February 2020) in the updated manuscript (new section 5.1), in which we compare LI and the Aon IF model output, and add aggregated market losses from the PERILS data as a benchmark. Figure R4 shows both the normalized losses and the storm ranking at country level for storm Sabine for the three datasets. In the revised manuscript, we removed Figure 5 and added Figure R4 as new Figure 6. Additionally, we extended Figure 8 (Spearman's rank correlation; now Figure 9) by adding PERILS to the comparison (see Figure R5).

The comparison between the transparent meteorological based indices and the AON insurance loss models is severely hampered by the fact that the AON model is basically a black box in this analysis. The differences between the approaches likely strongly depend on the vulnerability assumption/ damage functions applied. Here only generic information is given for the AON model in the paper. Therefore, no statements about quality of the approaches can be derived.

Answer: We agree with the reviewer in this point. We now provide a more detailed description (roughly 15 pages) of the Aon IF windstorm model, including a description of the hazard, vulnerability and exposure components, which is included in Supplementary Part A. Based on this detailed description, we also enhanced the discussion on the methodological differences between the two approaches in the revised manuscript (lines 375-410).

In my view the manuscript requires major work to expand the depth of analysis and make the conclusions more stringent.

Specific comments

Abstract: There are a number of qualitative statements such as "comparable storm ranks", "yet it is an effective index", etc, which should be underpinned with quantitative numbers/measures

Answer: We underpinned some of the statements by including actual numbers.

Introduction

1. Given the high interest of climate change, a short statement/references to recent trends of European windstorms should be added
   Answer: The manuscript is not about trends or decadal variability of windstorm activity in Europe. However, we added a statement that decadal and longer variability may be present in the datasets for the historical period (e.g. Feser et al., 2015) in the revised manuscript (lines 38-39).

2. The paper is mainly focusing on the wind-damage/loss relationships which is also one major conclusion from the comparison of the LI and the AON model output. However, loss datasets are not used in the current paper. Therefore, I suggest to either expand the scope of the paper and compare the loss estimates to loss datasets or to shorten the introduction in this

respect and more focus on wind-damage relationships in the literature

Answer: We thank the reviewer for pointing this out. We shortened the paragraph on damage datasets and expanded the part on damage functions (lines 40-53).

Section 2.1.

1. The LI equation (1) is summing up grid squares. One major reason for the differences between ERA5 and ERA-interim results are the different horizonal resolutions as stated by the authors later in the paper - which is kind of superficial. Suggest to exclude this effect in the analysis by appropriate measures.

   Answer: We tested the scaling effect of the different spatial resolutions by re-gridding ERA5 to the ERA-Interim grid before calculating LI. Figure R2 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 added the right part of Figure R2 to the Supplementary (Figure S3) and enhanced the discussion in the revised manuscript (lines 240-245).

2. Ln100: No rationale is given for the threshold chosen. Why 5 storms per season on average? How much % of historic losses of European windstorms are covered by this selection?

   Answer: From our experience, we typically have a maximum of 3-5 important windstorms per season (in terms of insurance losses and impact on the European market). Given that we are focusing on the top 20-25 storms for a period of 41 years most of the time, we do not think this pre-selection is a large constraint.

3. Ln109: see above, since the formula (1) is summing up every grid point, differences are to be expected for different horizonal resolutions. This effect should be normalized.

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

4. Ln 119: What would be the difference /effect if a 24-hour period is used, similar to the AON approach?

   Answer: We appreciate this suggestion. We did a sensitivity analysis for different time windows, e.g. calculating LI for 24-hour windows. Figure R6 shows the results for the normalized losses and figure R7 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 included both figures in the Supplementary (new Figures S5 and S6) and expanded the discussion in the revised manuscript (lines 316-322 and 378-384).

Section 2.2.

1. The AON model is only described rather high level. Especially the wind-damage relationship function is key in comparing the results with the LI index and derive meaningful conclusions. Without further details the conclusions will be rather qualitative and vague.
   Answer: Please refer to our reply on the general comments above.

2. Ln142: Commonly used damage functions assume either a power law or an exponential form. Please discuss why the v3 approach was chosen and discuss the strength and weaknesses of the approach compared to the "commonly used" ones.
   Answer: We thank the reviewer for this comment. We now provide a more detailed discussion of the impact of the LI setup on the results in the revised manuscript (lines 399-407).

3. Ln148: The hazard component consists of 26 historical events. What is the meteorological data used to define these events? In the table only date and name are given. How do the wind gust footprints compare between AON and ERA5?
   Answer: For historic modelling, footprints are built directly from weather station data. Inverse distance weighting is used to interpolate station values to the model grid at 7 km resolution. The gridded footprints are then implemented in the Aon IF model for loss calculation. Depending on the date and the geography of the specific event, a variety of data sources has been used to build the footprints – primary WMO station data, but also data from national meteorological services such as the British Meteorological Office (UKMO), the Danish Meteorological Institute (DMI), the Dutch Meteorological Institute (KNMI), the Irish Meteorological Service (Met Éireann), the Swedish Meteorological and Hydrological Institute (SMHI) the Finish Meteorological Institute (FMI), and the Norwegian Meteorological Institute (NMI). We now include a detailed description of the Aon IF model in Supplementary Part A and additionally added some information in the revised manuscript (lines 177-178).

4. Ln150: Was any downscaling performed to derive wind gust at higher horizontal resolution than native resolution of the ECHAM 5 global climate model?
   Answer: The stochastic model consists of 12,044 simulated storms. These have been extracted from the ECHAM5 Global Circulation Model. More information about the generation of this dataset is given in Karremann et al. (2014). The resolution of the pre-downscaled stochastic storms is 1.875° x 1.875°, approximately 200 x 200 km in the mid-latitudes. The stochastic storms are calibrated against a set of 124 historic storms taken from NCEP reanalysis data. A combination of dynamical downscaling with COSMO-CLM and statistical downscaling is used to produce the final high-resolution stochastic event set implemented in the Aon IF Euro WS model. We now include a detailed description of the Aon IF model in Supplementary Part A and additionally added some information in the revised manuscript (lines 178-182).

5. Ln164: See above, can more details be given about e.g. one damage curve used in the analysis?
Answer: The vulnerability functions within the Aon IF model estimate the likely damage to a risk at a given wind speed. The input to the vulnerability function is the gust speed and the output is the loss as a percentage of the total insured value (TIV), known as the damage ratio (DR). The vulnerability function is split into two components, the chance of loss (COL) and the conditional mean damage ratio (CMDR) which is the expected damage given that a loss is occurring. For a windstorm, the COL is low across most of the affected areas; if a postcode is hit by a 25 m/s gust, most of the buildings will not experience any loss.
The model uses damage matrices, wherein the hazard intensity is divided into bins, which are integer values of gust speed, and the estimated DR is divided into bins with each having a probability of being affected. Thereby, variation in the DR is considered. There is now a Figure illustrating this concept in the model documentation in Supplementary Part A.

Section 3.1: The 98th percentile acts quasi as a representation of building codes/standards for the LI approach. Does the AON model also have building code regions (where a similar wind speed would cause a different loss) implemented and if yes, how does the 98th "building code region" pattern compare to these?

Answer: Yes, there are 180 different vulnerability regions in Europe, where unknown building typology damage curves are based on the available building index. These vulnerability regions reflect differences in the engineering vulnerability due to the local building inventory.

Section 3.2

1. Table1: Correlation numbers for the loss seem to be mis-aligned with Figure 3, e.g. France 0.9059 vs 0.62
Answer: Thanks for pointing this out. Based on the comments by reviewer 2, we decided to merge Tables 1 and 2, and to replace the numbers with the R^2 values from the Spearman's rank correlation (see Table R1).

2. Ln205: "unlike ERA-interim, ERA5 shows a broader area of high wind guest, especially over the UK and Western continental Europe". Despite, the outlier shows very low LI ERA5 values?
Answer: The footprint for Irina is overall flatter in ERA5 compared to ERA-Interim. This is particularly the case for the UK, where the mean wind gust over land is 12.1 m/s for ERA5 and 24.6 m/s for ERA-Interim. Therefore, the LI for ERA-Interim is higher due to the cumulative effect (summation of v/v_{98}). Additionally, Figure R1 shows an extension of Figure S1 including another panel showing the ERA5 footprint re-gridded to the ERA-Interim resolution. This confirms the overall flatter structure of footprint. We agree this is an important point and we added more details (lines 227-230) and Figure R1 (as new Figure S1) in the revised manuscript.

3. Ln215: The obvious reason is the higher spatial resolution of ERA 5: as stated above: to make the analysis more revealing, it is suggested to aim to remove the grid square # dependency by normalization/scaling in the LI formula (1) and look at the remaining differences. Then also absolute values can be used in the analysis of Figure 3.

Answer: Please refer to our reply on the first comment in section 2.1.

4. Ln219: ranking of storms: Does this rank analysis really add significant value? At least some rationale for the differences should be given. Consider to instead translate the LI index values in monetary amounts by using one recent storm as reference loss/by normalizing it with e.g. PERILS loss estimate. This would be more tangible and can be also used to compare to AONs estimates (if available) in section 4.
   Answer: We think the analysis of the ranked storms is meaningful, as it actually demonstrates the strengths and usefulness of the LI methodology. Previous studies like Leckebusch et al. (2007) tried to translate loss from a meteorological index into monetary values, based on a very simple linear regression. However, the LI approach is always much simpler, as – unlike the Aon IF model – it does not have an exposure, vulnerability or financial component. Since our study is the first to compare a full insurance windstorm model (which is not available publicly) to a simplified meteorological loss index (which only considers the hazard component), we focus thus on a straightforward comparison of the two methods. Therefore, in our opinion, the objective should not be to translate the LI values in monetary values, but to compare the two methods while considering the Aon IF model output as a representation of real loss. Finally, we aim to extract meaningful conclusions about the similarities and differences, strengths and weaknesses.

Section 4: Table2: Suggest to add the number of storms for the ERA5 data set to allow for a comparison to the AON model

Answer: We are not sure if we understand this comment. The number of storms in ERA5 depends on our definition. As we are interested in extreme storms, we consider only events above a certain threshold, which corresponds to an average of five events per season. Therefore, the event set of ERA5 contains 205 events (=41 seasons*5events). For the comparison, we only focus on common events.

Section 4.1.

1. The core analysis of the paper (as also stated in the title) is to reveal the differences between the LI and AON model loss estimates which is mainly addressed in a scatter plot analysis. It is obvious that there are substantial differences between the two approaches. More emphasis should be put on revealing the reasons for the differences seen. Most likely this is due to the different vulnerability curve shape of the v3 approach and the AON model. Since the v3 approach was used in literature quite often in the past, it would be very beneficial to work out the limitations and suggest improvements based on the learnings of the comparison. However, this would likely mean to have more insights in the AON approach and to cross-check with real world loss numbers.
   Answer: Please refer to our replies to the general comments above. Additionally, we would like to mention that 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. Nevertheless, we now provide a more detailed discussion of the impact of the LI setup on the results in the revised manuscript (lines 399-407).

2. Figure 6 shows clearly the very different behavior of both approaches. Maybe a more in depth analysis for some few selected storms can help to shed more light on the differences, e.g. by looking into the loss contribution to the overall loss by wind speed.
   Answer: Please refer to our reply on the previous comment.

Section 5

1. "For all of Europe, LI values are higher for ERA5 than for ERA-interim". This is mostly an effect of different horizontal resolution. Suggest to remove and discuss the residual differences as stated above.
   Answer: Please refer to our reply on one of the previous comments.

2. "Compared to AON's IF model, LI ERA5 shows overall lower loss values". This statement is only true for normalized values but not for (more relevant) monetary values. For this (highly valuable) analysis, LI loss index values need to be translated in monetary loss values and compared to AON's output.
   Answer: Please refer to our comment on section 3.2

3. Ln296: "the AON model seems to better distinguish between high and moderate impact events": Without benchmark with real world loss numbers this statement appears quite subjective.
   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 perspective for the purpose of this paper. Nevertheless, we decided to use none of the datasets as ground truth (Moemken et al., 2024). We carefully went through section 5 (now section 6) and rephrased sentences where necessary (e.g. lines 399-404 and 413-415).

4. Ln 299: …, the catastrophe model shows a clear regional dependency of loss values. This regional dependence in less pronounced in LI ERA5. Suggest to discuss the reason for this behavior. Are AON footprints downscaled or have higher resolution?
   Answer: The hazard resolution in the Aon model is 7 km. Please also refer to the replies on the other comments regarding the Aon model.

5. Ln316: As stated above, what is the impact of using a 24h definition also for the LI approach?
   Answer: Please refer to our reply on one of the previous comments.

6. Ln333: “the LI index is missing a detailed damage component, thus struggles to capture the non-linear response of the buildings at the tail of the gust spectrum for high impact events”. The LI v3 approach is obviously non-linear, so the question is what stronger non-linearity would be more suitable. But foremost, from the analysis it is only clear how the wind-loss relationship compares to the unknown AON approach and it is difficult to draw conclusion on the quality of the approaches without comparison to reality.
   Answer: We agree with the reviewer and reworded the statement in the revised manuscript. 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 perspective. We clarified this in the revised manuscript (lines 192-194). Additionally, we included the PERILS data in some of the analysis (see above).

7. Ln343: “…it is suitable for estimating the impact…” Without more quantitative measure and real world comparisons (and given the large differences to the supposedly more sophisticated AON approach) it is hard to follow this conclusion.
   Answer: We reworded the statement in the revised manuscript (see also previous comments).

References

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

Born, K., Ludwig, P., and Pinto, J. G.: Wind gust estimation for Mid-European winter storms: towards a probabilistic view, Tellus A, 64, https://doi.org/10.3402/tellusa.v64i0.17471, 2012

Feser, F., Barcikowska, M., Krueger, O., Schenk, F., Weisse, R., and Xia, L.: Storminess over the North Atlantic and northwestern Europe – A review, Quart. J. Roy. Meteor. Soc., 141, 350-382, https://doi.org/10.1002/qj.2364, 2015

Karremann, M. K., Pinto, J. G., von Bomhard, P., and Klawa, M.: On the clustering of winter storm loss events over Germany, Nat. Hazards Earth Syst. Sci., 14,2041-2052, https://doi.org/10.5194/nhess-14-2041-2014, 2014

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

Leckebusch, G. C., Ulbrich, U., Fröhlich, L., and Pinto, J. G.: Property loss potentials for European midlatitude storms in a changing climate, Geophys. Res. Lett., 34, L05703, https://doi.org/10.1029/2006GL027633, 2007

Minola, L., Zhang, F., Azorin-Molina, C. et al. : Near-surface mean and gust wind speeds in ERA5 across Sweden: towards an improved gust parametrization, Clim. Dyn., 55, 887-907, https://doi.org/10.1007/s00382-020-05302-6, 2020

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

Seregina, L. S., Haas, R., Born, K., and Pinto, J. G.: Development of a wind gust model to estimate gust speeds and their return periods. Tellus A, 66, https://doi.org/10.3402/tellusa.v66.22905, 2014