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 plans how to incorporate them in our 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 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 reference. Figure RC3.1 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 RC3.1. Additionally, we will extend Figure 8 (Spearman rank correlation) by adding PERILS to the comparison (see Figure RC3.2).

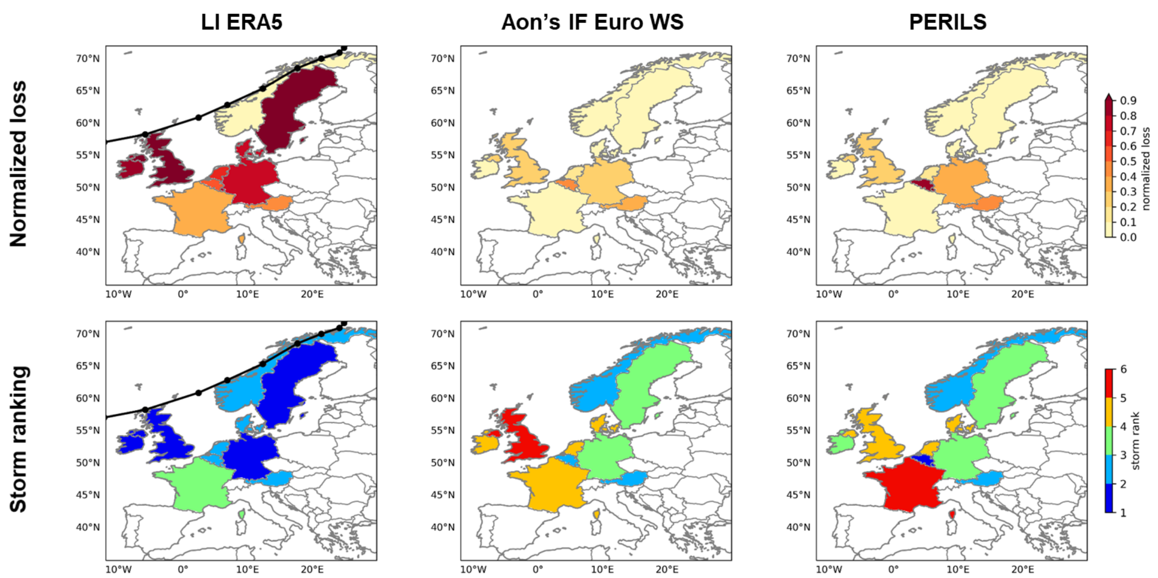

Figure RC3.1: *Normalized losses (upper row) and storm ranking (lower row) at country level for storm Sabine in February 2020. Losses are derived from LI ERA5 (left), Aon's IF Euro WS model (middle), and PERILS (right). The* *black line and dots in the left column denote the cyclone track derived from 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.*

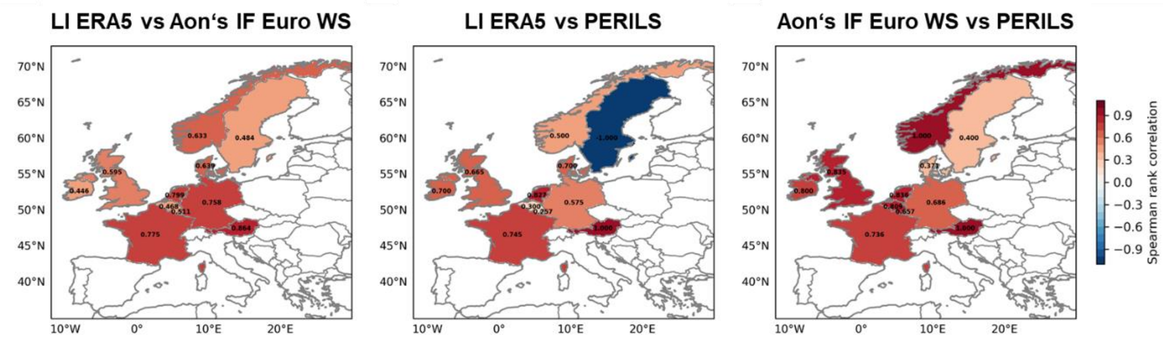

Figure RC3.2: *Spearman's rank correlation coefficient at country level for LI ERA 5 vs Aon's IF Euro WS model (left), LI ERA5 vs PERILS (middle), and Aon's IF Euro WS model vs PERILS (right). The ranking is based on common storms per country.*

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 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.

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 will underpin 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 will add a statement that decadal and longer variability may be present in the datasets for the historical period (e.g. Feser et al., 2015).

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 will shorten the paragraph on damage datasets and expand the part on damage functions.

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 RC3.3 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 RC3.3 to the Supplementary and enhance the discussion in the revised manuscript.

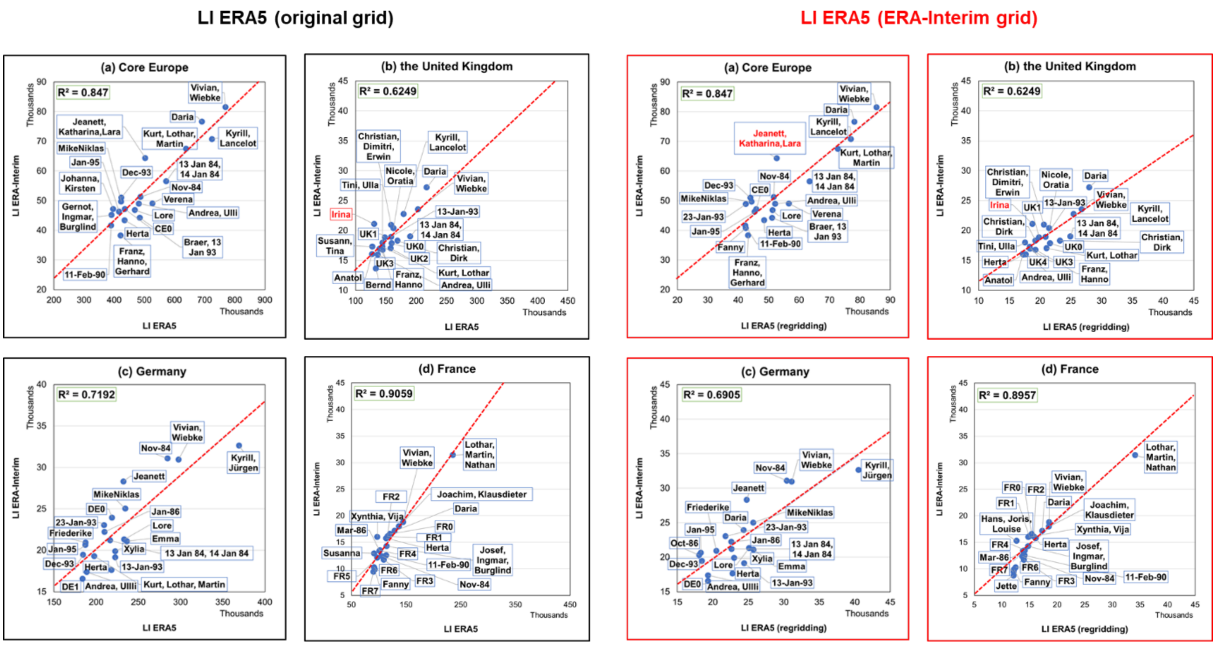

Figure RC3.3: *Comparison of loss values (in thousands) based on LI ERA5 (x-axis) and LI ERA-Interim (y-axis) for the common 20 most extremes storms in the period 1979-2019. LI ERA5 is calculated from the original ERA5 gust data (left) and from the ERA5 gust data re-gridded to the ERA-Interim resolution (right).*

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 RC3.4 shows the results for the normalized losses and figure RC3.5 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.

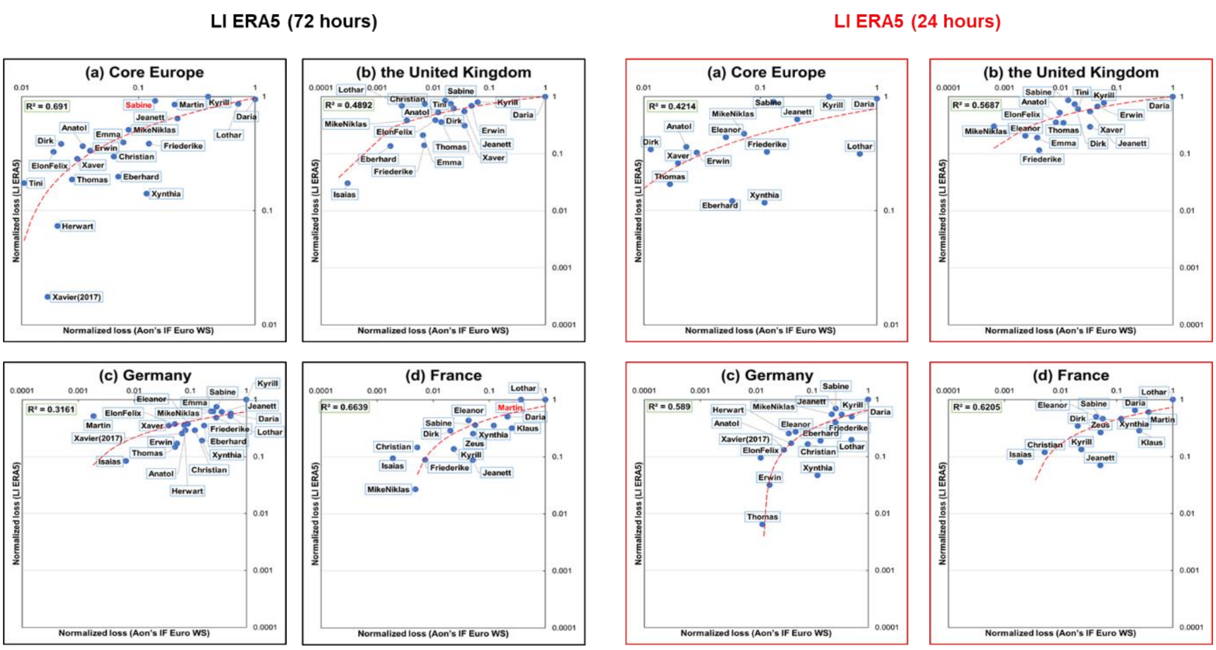

Figure RC3.4: *Comparison of normalized loss values between Aon's IF Euro WS model (x-axis) and LI ERA5 (y-axis). Depicted are the common most extreme storms for the period 1990-2020 for (a) Core Europe, (b) the United Kingdom, (c) Germany, and (d) France. A logarithmic scale is used for the axes. The red dashed line denotes the logarithmic regression. The correlation between the datasets is given in the upper left corner (R2 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).*

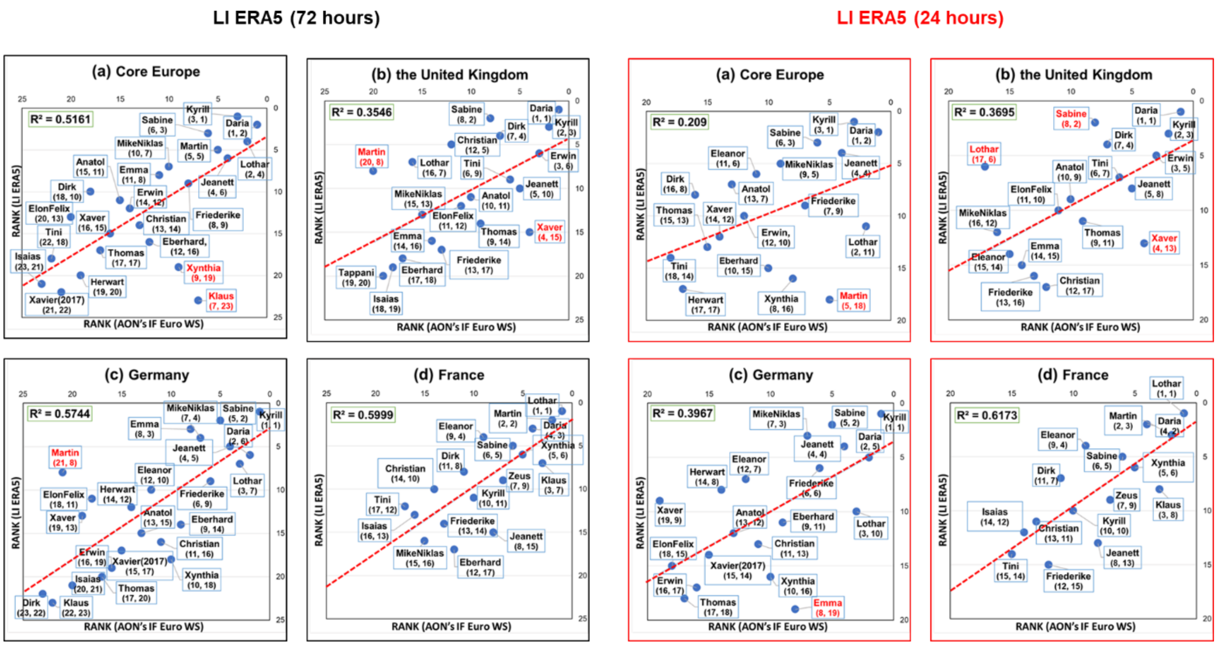

Figure RC3.5: Same as Figure RC3.4, but for the comparison of storm ranks. The values in brackets indicate the rank (first value Aon's model, second value LI ERA5).

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 will provide a more detailed discussion of the impact of the LI setup on the results in the revised manuscript.

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 will provide a detailed description in the revised Supplementary.

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. A detailed description will be made available in the Supplementary of the revised manuscript.

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 will be a Figure illustrating this concept in the model documentation in the Supplementary of the revised manuscript.

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 RC3.1).

Table RC3.1: *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]

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 RC3.6 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 will add more details and Figure RC3.6 in the revised manuscript.

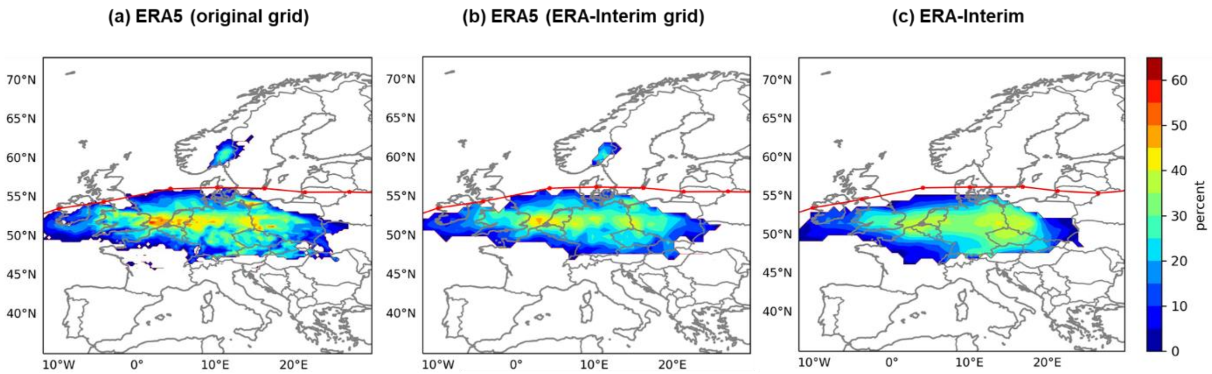

Figure RC3.6: *Wind gust footprint for storm Irina in October 2002 based on ERA5 (a), ERA5 re-gridded to the ERA-Interim grid (b), and ERA-Interim (c). Shown is the largest exceedance (in percent) of the 98th percentile of* *daily maximum wind gust within 72 hours. The red line and dots denote the cyclone track derived from ERA5 (a, b) and ERA-Interim (c) using the tracking algorithm of Pinto et al. (2005).*

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 economic 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 will provide a more detailed discussion of the impact of the LI setup on the results in the revised manuscript.

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-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.

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 will reword 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-reality”. We will clarify this in the revised manuscript. Additionally, we will include 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 will reword the statement in the revised manuscript (see also previous comments).

References

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

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

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