AUTHORS' RESPONSE TO REFEREE #1

Research article:

Comparing an insurer's perspective on building damages with modelled damages from pan-European winter windstorm event sets: a case study from Zurich, Switzerland (Nat. Hazards Earth Syst. Sci. Discuss., https://doi.org/10.5194/nhess-2020-115, in review; submitted on 07 April 2020)

Authors:

Christoph Welker, Thomas Röösli, David N. Bresch

We thank the referee Dr. Alexandros Georgiadis for his comments, which have improved the quality of the manuscript.

The original comments from the referee are listed below directly followed by our responses in blue and italic and changes to the manuscript in blue and bold.

The main objective of the paper is to demonstrate the value of catastrophe modelling analysis in respect of estimating the frequency of high intensity storms, compared to a pure statistical analysis of the claims history from a portfolio that has a limited record of a few decades. Two catastrophe models with different vulnerabilities and exposures are used to calculate the losses, GVZ's proprietary model and the open source CLIMADA platform. Both models perform very well in calculating the losses of a numbers of historical storms (e.g. Vivian, Lothar, Burglind) so are clearly appropriate tools for the stated job. The selected hazard inputs include the: (i) WISC historical set of 75 years with 142 events and (ii) a probabilistic perturbation of the above event set, where every storm has 29 altered offsprings, thus the set is extended to have 4,260 storms covering 2,250 years. On the other hand, the insured claims dataset consists of about 40 years of losses that provides 18 storms which are available in the WISC historical set. Overall, I think that the presented work is of high quality: there are no obvious methodological errors and the findings are robust regarding the stated purpose, to complement claims-based risk assessment with a modelling approach. The conclusion that the return period of intense storms (like Lothar) cannot be determined sufficiently from a simple analysis of claims history is robust, but also well established in the Insurance industry. The proposed approach to produce a probabilistic event set by perturbing/expanding the WISC historical events, then calculate the losses using one or more damage models is technically correct and appropriate but it is not novel. Focusing on the results, I think that risk assessment at the tail will benefit from an attempt to build a more focused estimation of the uncertainty associated with the WISC probabilistic exceedance probability curves in Figure 2. The confidence interval based on the WISC historical set (CHF 19M to 33000M) is very conservative and negates much of the fundamental advantage of complementing risk assessment with probabilistic catastrophe modelling. I think that this is the major point to be addressed in the analysis, thus I would recommend publishing the article conditionally the authors provide a substantial response to this question (see below, bullet points: 2.a-c).

In the eyes of both referees, the uncertainty of the probabilistic event set "WISC probabilistic extension" should be discussed in more detail. Nonetheless, they have different opinions about the uncertainty estimations: While Referee #1 writes that "the confidence interval based on the WISC historical set [...] is very conservative and negates much of the fundamental advantage of complementing risk assessment with probabilistic catastrophe modeling", Referee #2 writes that "the authors correctly state in their discussion [...], the 'WISC probabilistic' dataset does not reduce uncertainty compared to 'WISC historic' because they're based on the same data".

The opposing ways of interpretation of both referees show that there are obviously different ways of interpretation about whether the uncertainty of risk assessment can be reduced by a probabilistic event set based on the same data. In a way, we represent the "conservative" way of interpretation in our paper, i.e. that the uncertainty cannot be reduced by statistical perturbation, and we would like to continue to support this way of interpretation. We will discuss this further below.

In this response, we will show that the illustration of uncertainty, as requested by Referee #1, partly ignores the parameter uncertainty and that is why the full uncertainty cannot be illustrated easily. We will mention the uncertainty more often as requested by Referee #2.

In the following, we would like to briefly clarify our way of interpretation of the uncertainty associated with historic and probabilistic event sets in general and in the case of this paper:

(1) Historic event sets:

Regarding the risk from rare events, an important source of uncertainty is the sampling uncertainty. In this paper, we illustrate the sampling uncertainty of both insurance claims data and modelled damages based on "WISC historic" by showing the 90-% confidence interval derived by resampling (see Fig. 2).

(2) Probabilistic event sets:

As Referee #1 summarises, a probabilistic event set can be generated by statistical perturbation and by using dynamical models. The sources of uncertainty are different for both approaches. In the following, we only want to discuss statistical perturbation, as this was the subject of the paper. We used statistical perturbation with two parameters with the aim of representing the distribution of pan-European windstorm severity. By doing this for the best-fit distribution, we transformed the sampling uncertainty of the severity of historic windstorm events into parameter uncertainty of our model. However, as our statistical approach does not add any additional information, the uncertainty is finally not reduced. In our opinion, only if the process of generating a probabilistic event set does include additional information one could argue in favour of the probabilistic event set reducing uncertainty.

As this interpretation and argumentation needs to be clarified in the manuscript, we will incorporate it at different points throughout the revised manuscript.

Also suggestions to further expand the work (beyond the scope of the current article) are available in the end of bullet point 1.

More specifically, I will address the following scientific question/issues:

(1) The proposed approach to produce a probabilistic event set by perturbing/expanding the WISC historical events is technically correct and appropriate given the scope of the analysis. Having said that, although acceptable, the approach is not novel. Several (re)insures have proprietary cat models that follow similar methodologies. A limited historical 'seeding' data-set (often based on reanalysis data, e.g. 20C_R, ERA-Int, ECMWF_R) is extended either by a statistical perturbation/resampling approach (e.g. Swiss Re) or extensive use of dynamical modelling (usually regional climate modelling-RCM) outputs (e.g. Weather Predict/Renaissance Re, Partner Re) to produce a realistic

probabilistic event set. The advantage of the latter is the physical consistency of each individual stochastic event due to the physics-based simulation of the RCM. Furthermore, the main catastrophe model vendors in the market (RMS, AIR, AON Impact forecasting and more) tend to provide probabilistic windstorm solutions based on outputs extracted for a variety of long global climate model (GCM) runs, calibrated (often fitted) against the available historical record. The advantage of this approach is that the simulation generates physically realistic storms that are not constrained by the attributes/parameters of the seeding historical windstorms.

As the referee rightly states, there are many different ways to assess the risks from European winter windstorms. We show two possible approaches in this paper, i.e. a methodology implemented in a proprietary model and one in an open source model, and discuss which uncertainties have to be considered with these two approaches. Furthermore, we check the reliability of the open source impact model CLIMADA with both GVZ's claims data and output from their proprietary damage model. Those kinds of proprietary data are usually not available for scientific publications.

The paper was not necessarily about showing a new methodology. In our view, the recent development of freely accessible data on windstorm footprints (WISC) in combination with an open source damage model (CLIMADA) opens up new opportunities for applied research and provides a straightforward entry point for insurance companies to model the risks associated with winter windstorms in Europe – thus providing an additional / alternative perspective compared to inhouse or commercial models (as listed by the referee above). The application example we give is something new because of the open source concept presented.

Such methodologies directly address the main limitation of the WISC probabilistic expansion approach used by the authors that results to almost identical AAD values in Tables 2 (1.4M CHF) and A1(1.1M-1.2M CHF) for the WISC historic and probabilistic sets. The probabilistic expansion adds very little further risk hazard information compared to the seeding historical set. A possible avenue for the authors to continue the current work would be to look into calibrating the WISC synthetic gusts distribution (in figure A1, lines 793-797) against the WISC historical event set to address the low gust speed intensity. Then repeat the loss calculation with the ‘enhanced’ WISC synthetic event set.

We thank the referee for his suggestion to calibrate the distribution of the event set "WISC synthetic" against “WISC historic”. However, we do not think that this would be successful in the case of “WISC synthetic” for the following reason.

The event set "WISC synthetic" contains wind gust footprints for around 23’000 synthetic windstorms: i.e., three sets of 7’660 events each. In each of the three sets a different approach was applied to carry out a calibration (see https://wisc.climate.copernicus.eu/wisc/#/help/products#eventset_section*), which however ultimately did not solve the problem of a generally lower severity of the synthetic windstorm events compared to the historic ones. It is possible, that such a calibration would be more successful, if applied to hourly wind gust data, before the aggregation to 72-hour events is done. This is analogous to the conclusion about correcting WISC wind gust data for higher altitudes in Marseille et al. (2017).*

We agree with the referee, that in general, a probabilistic event set originating from dynamical modelling could provide new information and would allow to reduce the uncertainty, which is the main limitation of our WISC probabilistic expansion approach. We think this is an important statement, that we would like to include as an outlook in the revised manuscript. We suggest to add the following sentence at L495:

“**In future studies, the information from dynamical models, which are run for many model years, would help to further reduce this uncertainty.**”

(2) The approach to expand the WISC historical events and determine the frequencies of the offspring probabilistic storms (GEV distribution fitted to the historical SSI values) has merit, and the concluding results in paragraphs 3.2 and 3.3, also provided in table 2, are realistic. I am not surprised the two WISC-based analyses reduce the calculated AAD value between 1.1 and 1.4M CHFs. Also, Lothar/Martin’s return period is (correctly) positioned at and above 75 yrs, potentially beyond 125 yrs. Considering the disproportional yet uncertain impact of the extreme event Lothar/Martin on the claims data analysis, the above results are plausible, yet the authors do not follow with a narrower estimation of the uncertainties. I understand why the authors prefer to retain the

confidence interval based on the WISC historical set (CHF 19M to 33,000M), yet this reduces somewhat the functionality of the probabilistic expansion model. It's main objective is to provide a tail view. Here are a few suggestions:

a. The 4,260 storms in the WISC probabilistic set provide the equivalent of 2,250 years of storm activity (based on the analysis assumptions). You may sample randomly the equivalent of 250 or 500 years of storms and build multiple exceedance frequency curves for each sample. A spaghetti plot of the 'secondary' exceedance frequency curves will enable a reviewed estimation of the uncertainty around the curve. Essentially the idea is not dissimilar to the re-sampling approach described in paragraph 2.4.3 for the Pareto Pricing.

b. Estimate multiple probabilistic extensions of the WISC historic event set with different initial assumptions including (but not limited to) fitting different extreme distributions (e.g. Weibull, Pareto), inclusion/exclusion of Lothar/Martin in the seeding WISC historic set to quantify the sensitivity of the methodology in the most extreme event in the set, for both damage models (GVZ & CLIMADA). This will produce an ensemble of exceedance frequency curves that can be visualized as a spaghetti plot.

c. A combination of the above two ideas can work as well.

We thank the referee for his suggestions. We have implemented all of them and discuss the results in the following. As a conclusion, we would still argue, that the yellow ribbon in Fig. 2 (i.e., the sampling uncertainty of the modelled damages based on "WISC historic") is the best illustration of the uncertainty for "WISC probabilistic extension". We will include this argumentation in the manuscript, alongside the arguments already provided in this response.

Following the referee's suggestion 2a and based on our data sample of total damages modelled based on the hazard event set "WISC probabilistic extension" and the GVZ damage model (red diamonds in Fig. 2), we sampled randomly the equivalent of 500 years of windstorms and built an exceedance frequency curve for each sample (number of samples = 1'000). Accordingly, the red shading in Fig. R1-1 shows the 90-% confidence interval as a result of the random resampling.

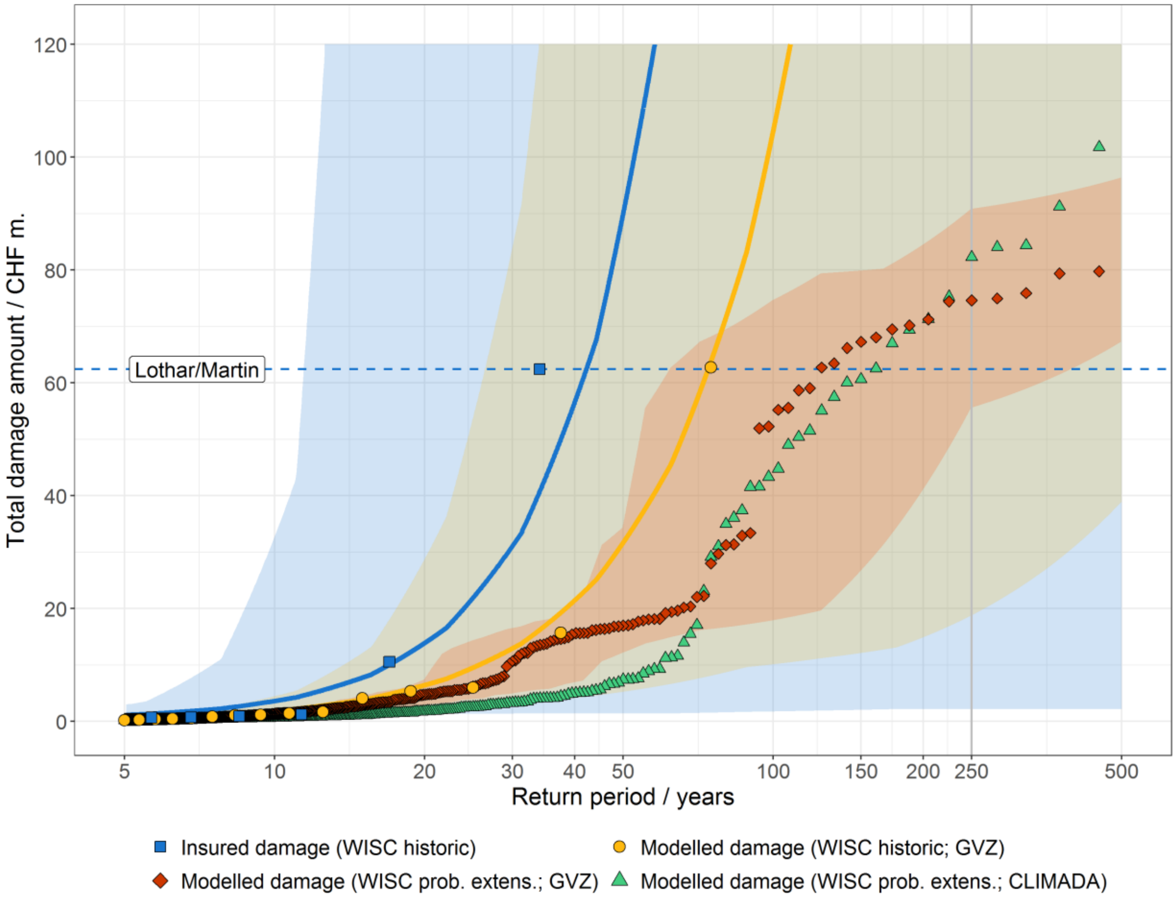

Figure R1-1: *Modified Fig. 2. New is the red shading, which shows the 90-% confidence interval for the modelled damages based on "WISC probabilistic extension" and the GVZ damage model computed by applying the referee's suggestion 2a.*

We are aware that the parameter uncertainty regarding the event set "WISC probabilistic extension" is important, especially in comparison with "WISC historic". However, in our opinion this source of uncertainty is not fully estimated and sufficiently illustrated with such a resampling methodology.

Following the referee's suggestion 2b, to include / exclude Lothar/Martin in the seeding, we tried a more systematic approach. We resampled (choice with replacement) the historic events (same number of events in each sample; choosing with replacement means some events are missing, whilst others are double). Then we created a probabilistic event set for each of these samples. The 90-% confidence interval is again given by the 5th and 95th percentiles of all samples. This is the best possible way we achieved to illustrate at least part of the uncertainty that originates from the fact that the best-estimate of the *distribution of the pan-European Storm Severity Index is unknown and thusly the parameters for the creation of the probabilistic sets can only be chosen with a certain degree of uncertainty. The uncertainty estimation up until a 30-year return period follows approximately the uncertainty estimation for "WISC historic"; at higher return periods the uncertainty estimation is levelling off, probably due to the limited ability of our probabilistic approach to create very different (e.g., much stronger) events from the seeding historic set. Therefore, we argue that the shown difference between the yellow ribbon and the red ribbon could be misleading.*

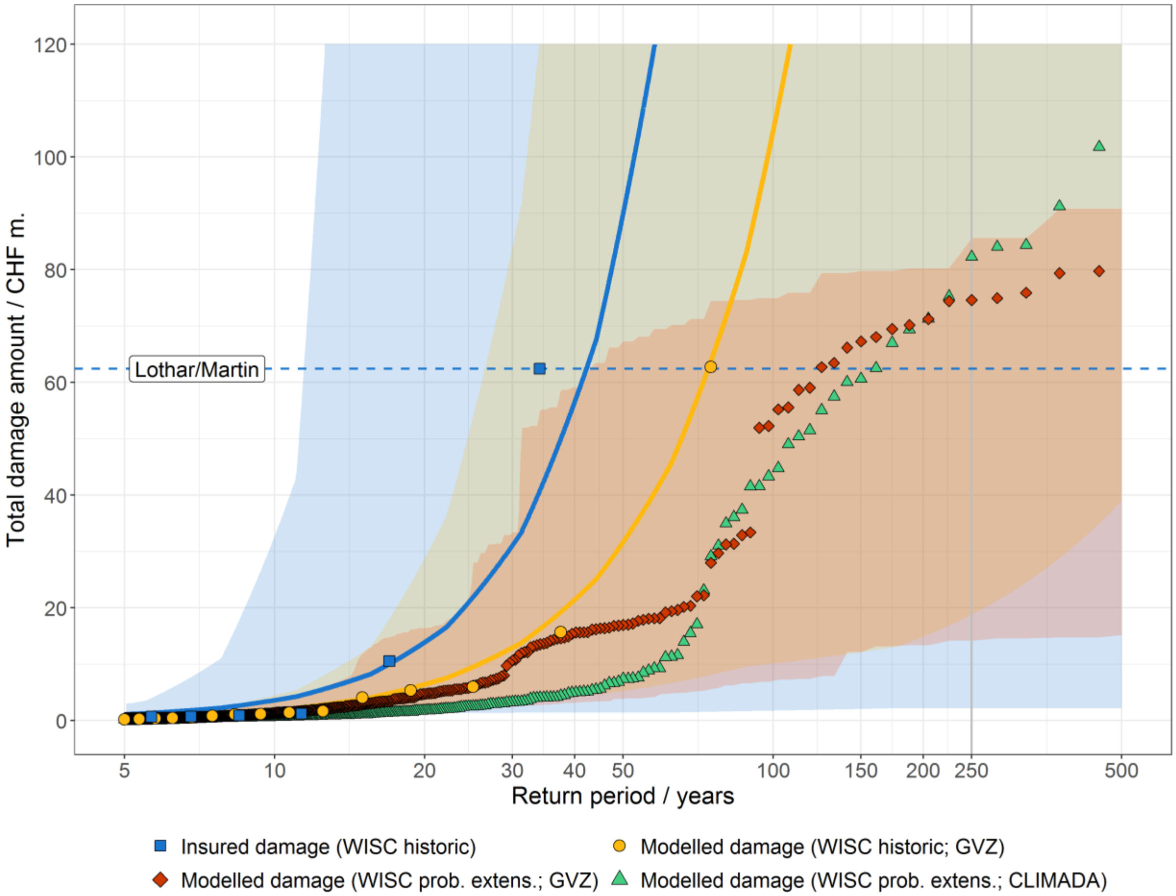

Figure R1-2: *Analogous to Fig. R1-1 but here the red shading shows the 90-% confidence interval for the modelled damages based on "WISC probabilistic extension" and the GVZ damage model computed by applying the referee's suggestion 2b.*

The results for the referee's suggestion 2c, which is a combination of his suggestions 2a and 2b, are given in Fig. R1-3. Firstly, we resampled (number of samples = 100) the historic events and then used these different historic samples to create an ensemble of probabilistic damage event sets (as suggested in 2b). Secondly, for each new *probabilistic damage event set, we sampled (number of samples = 20) randomly the equivalent of 500 years of windstorm events and built an exceedance frequency curve for each sample (as suggested in 2a). From this set of resampled and bootstrapped damage event sets (total number of samples = 2000), we then calculated the 90-% confidence interval. Whereas this combination provides a smooth illustration of the resampling uncertainty, it still suffers from the same problem as the illustration in Fig. R1-2. Therefore, we would still argue that the yellow ribbon in Fig. 2 is the best illustration of the uncertainty for "WISC probabilistic extension".*

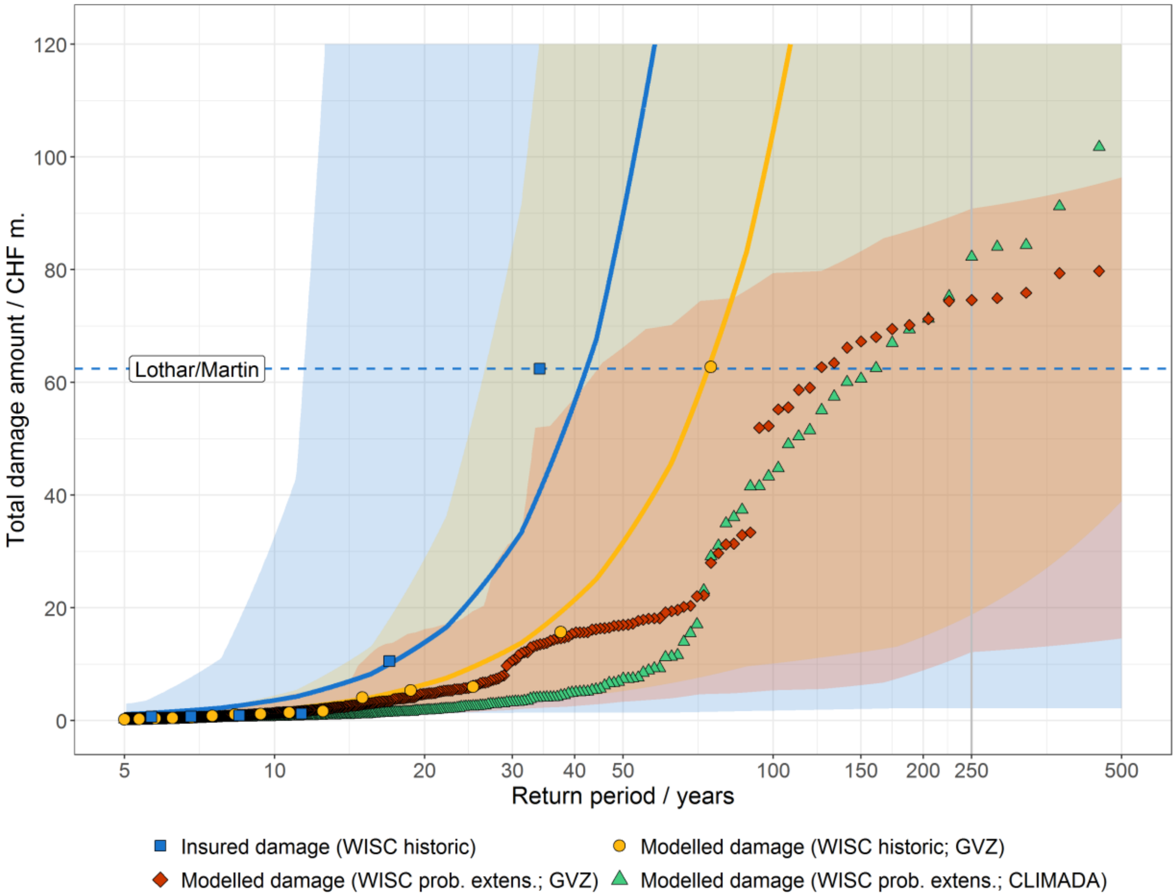

Figure R1-3: *Analogous to Fig. R1-1 but here the red shading shows the 90-% confidence interval for the modelled damages based on "WISC probabilistic extension" and the GVZ damage model computed by applying the referee's suggestion 2c.*

(3) One aspect which is underrepresented in the discussion is the role of the loss uncertainty due to the vulnerability (and exposure) components. GVZ's damage model has a stochastic component as seen in figure 4, also described in the text (lines 443 to 449), yet it is unclear whether the damage (given by the red bars in figure 4) informs the process of

building the exceedance frequency curve of the modeled damage based on the WISC probabilistic extension of figure 2. Please clarify.

The range of the modelled damages through the stochastic component in GVZ's damage model (represented by red bars in Fig. 4) is not directly included in the calculation of the exceedance probabilities in Fig. 2. Rather, we use the median of the damage range modelled for each event to calculate the exceedance probabilities.

Additionally, we suggest to include the uncertainty related to vulnerability and exposure in the following sentences at L510:

"**A disadvantage of the used vulnerability curve is that it does not implicitly provide a quantification of the uncertainty as a probabilistic vulnerability curve would (e.g., Heneka et al., 2006; Prahl et al., 2012). The quantification of the uncertainty of exposure and vulnerability information was generally omitted in this study to focus on the comparison of the claims and hazard datasets. But of course, for comparison of the presented risk numbers with other studies the uncertainty of the vulnerability and exposure information play a bigger role.**"

The two references used have also been included at L659 and L688:

"**Heneka, P., Hofherr, T., Ruck, B., and Kottmeier, C.: Winter storm risk of residential structures – model development and application to the German state of Baden-Württemberg, Nat. Hazards Earth Syst. Sci., 6, 721–733, doi:10.5194/nhess-6-721-2006, 2006.**"

"**Prahl, B. F., Rybski, D., Kropp, J. P., Burghoff, O., and Held, H.: Applying stochastic small-scale damage functions to German winter storms, Geophys. Res. Lett., 39, L06806, doi:10.1029/2012GL050961, 2012.**"

(4) The two modelling approaches (GVZ damage model & CLIMADA impact model) use different input exposures as described in lines 272 for GVZ's model and 303 for CLIMADA. Is it possible to get a feeling regarding the difference between the two input exposures (e.g. 10%, 50%)?

The GVZ damage model uses an exposure information (i.e., insured value of the buildings in the canton of Zurich) which sums up to approximately 480 billion CHF. The exposure used in the CLIMADA impact model sums up to 80 % of that value.

In this context, it is important to emphasise that differences in the total exposure values, compared to the GVZ damage model, were partially compensated by calibrating the damage functions in the CLIMADA impact model, in order be able to reproduce event damages comparable to those from the insurance claims database. We used publicly available exposure information in CLIMADA and not GVZ's proprietary portfolio information because of the open source concept presented in this paper. This way the presented methodology can be easily applied to other regions.

References used in this response

Heneka, P., Hofherr, T., Ruck, B., and Kottmeier, C.: Winter storm risk of residential structures – model development and application to the German state of Baden-Württemberg, Nat. Hazards Earth Syst. Sci., 6, 721–733, doi:10.5194/nhess-6-721-2006, 2006.

Marseille, G. J., Stoffelen, A., van den Brink, H., and Stepek, A.: WISC Bias Derivation and Uncertainty Assessment, https://wisc.climate.copernicus.eu/wisc/documents/shared/(C3S_441_Lot3_WISC_SC2-D3.3-CGI-RP-17-0071)%20(Final%20Bias%20Derivation)%20(v1.0).pdf*, 2017.*

Prahl, B. F., Rybski, D., Kropp, J. P., Burghoff, O., and Held, H.: Applying stochastic small-scale damage functions to German winter storms, Geophys. Res. Lett., 39, L06806, doi:10.1029/2012GL050961, 2012.