Reply to referee 1

We thank the reviewer for reading our manuscript and giving thoughtful comments and suggestions. A detailed response to all comments is found below in blue.

The authors assess the performance of different storm damage functions, that model the relationship between wind speeds and insured losses, for Norway. They make use of district-level insurance loss data from Norway, that span the time period 1985 to 2020, combined with regional reanalysis/hindcasts of wind speeds. The study nicely assesses the relative benefits of the different (previously developed) storm loss models.

The manuscript is overall very well written and provides a very clear overview of the topic, and insightful explanations and discussions. I only have a few minor comments for the authors to take into account before publication, and congratulate on this nice piece of work.

Comment 1

Line 59/60: Also the Donat et al 2011b study (already cited in the manuscript) estimated the losses at district level in Germany, regionally adjusting the loss function by Klawa&Ulbrich

Thanks for bringing this to our attention. We will add this reference in the revised manuscript.

Comment 2

Line 114: maybe specify that these hindcasts are “retrospective” forecasts, to avoid any ambiguity

There does not seem to be a very clear separation between a hindcast or a retrospective (reforecast) in the literature. Both usually uses a reanalysis or a state-of-the-art analysis for its boundary conditions. However a reforecast sometimes is a forecast done in hindsight using the same initial conditions as the original forecast. As the terms retrospective forecast, reforecast and hindcasts seems to be more or less interchangeably we would prefer to keep the term 'hindcast' as it is the term used in the NORA3 reference paper for the wind data we use.

Comment 3

Line 130-132: I note that the authors follow here a different order of operations than some of the studies they reference for the loss functions they use. E.g. the Klawa approach first calculates a loss index (i.e. the cubic exceedance of the wind speed threshold), and applies the population density weighting afterwards when spatially aggregating the losses. This is different to what the authors are doing here, as they seem to apply the population weighting already to the daily wind speeds. It would be good to (i) be explicit about this variation in approaches, and (ii) discuss/demonstrate the effects of these different orders of operations.

The observation made by the reviewer points to an important methodological difference in our approach compared with previous studies. We chose to weight wind speeds with population first in order to have the same input for all damage functions fitting. That being said, it is also worthwhile to compare the proposed and the alternative methodology used in studies such as Pinto et al. (2007). As suggested by the reviewer, we have performed this additional analysis. The comparison of the damage functions and their predictive skill do not show any significant differences between both methodologies (Fig. R1). We have calculated the damage function (see eq. 2 in the manuscript) using both methodologies and found that there is a high correlation between the two (Fig. R1a). Upon calibrating the two damage functions with municipality level insurance losses using eq. 3, we observe that in the extreme loss class MAEs are highly correlated with each other (Fig. R1b). Moreover, we find that their magnitudes are similar, with about 91.6% of the municipalities having MAE differences within [-70 NOK, 70 NOK] (Fig. R1c). An independent sample t-test failed to conclude for any significant differences between the mean of MAEs from both methodologies. However, when not distinguishing loss classes, we find that the alternative method (such as in Pinto et al. 2007) has better skill in estimating the losses, although this result depends on the model evaluation metric used (not shown). In light of these results, we will add some sentences in the manuscript such as:

In section 2.4.2 (Cubic-excess over threshold model):

Here we weight the wind speeds with population and aggregate it to the municipality-level resolution such that it corresponds to the loss data resolution. However, other studies, such as Pinto et. al. (2007), weight the loss index and aggregate it to the district or national resolutions. As discussed later in the paper, these two methods do not give very different results.

In section 3.1 (Overview of the windstorms losses):

Unlike previous studies (e,g, Pinto et al. 2007), which weights spatially aggregated loss index devised from cubic exceedance of wind speed above sufficiently high threshold (computed as in eq. 2), we here weight the wind speed with population first and then aggregate it to a coarser resolution. We compare the Klawa damage function as in equation 2 obtained from the proposed methodology with the alternate methodology employed in Pinto et al. (2007). There is a high correlation between both damage functions/loss indices (Fig. R1a). Upon calibrating the two damage functions with municipality level insurance losses using eq. 3, we observe that in the extreme loss class MAEs are highly correlated with each other (Fig. R1b). Moreover, we find that their magnitudes are similar, with about 91.6% of the municipalities having MAE differences within [-70 NOK, 70 NOK] (Fig. R1c). An independent sample t-test *failed to conclude for any significant differences between the mean of MAEs from both methodologies. However, when not distinguishing loss classes, we find that the alternative method (such as in Pinto et al. 2007) has better skill in estimating the losses, although this result depends on the model evaluation metric used (not shown). Thus, there is no conclusive evidence for one of the methods exhibiting a higher predictive skill than the other.*

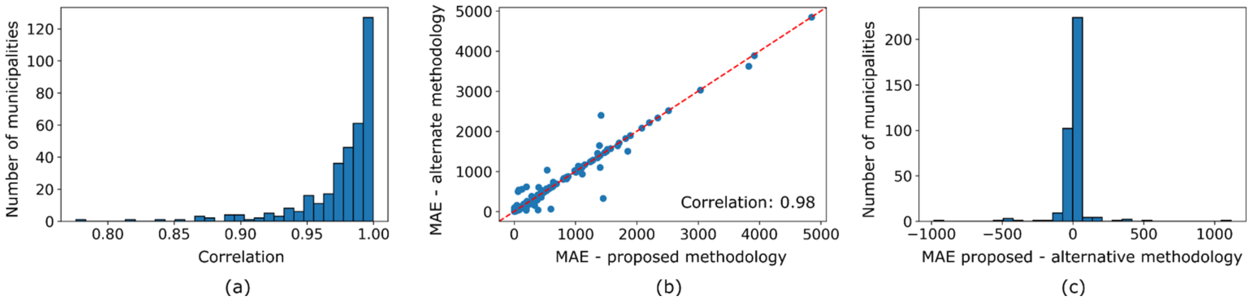

Fig R1: (a) Distribution of the Pearson correlation coefficient between the damage functions using the proposed (ours) and the alternate methodologies, (b) scatter plot of the MAEs of the extreme loss (losses above the 99.7th percentile) estimates for the proposed and the alternate methodologies using the testing data. Blue dots represents individual municipalities and the dashed red line represents the 1:1 line. (c) Distribution of MAE differences between the proposed and the alternate methodologies.

Pinto, J. G., Fröhlich, E. L., Leckebusch, G. C., and Ulbrich, U.: Changing European storm loss potentials under modified climate conditions according to ensemble simulations of the ECHAM5/MPI-OM1 GCM, Nat. Hazards Earth Syst. Sci., 7, 165–175, https://doi.org/10.5194/nhess-7-165-2007, 2007.

Comment 4

Line 146-150: I wonder how sensitive are the results to the specific choices of these testing and training samples?

The reviewer points to an important aspect of the damage functions: they are sensitive to the choice of the testing and training data (Prahl et. al. 2015). To quantify uncertainties involved in the choice of training data, we have performed a seven-fold cross-validation. The 36-year loss data is split into 7 groups in chronological order with each group containing five years of loss data (1985-1989, 1990-1994, 1995-1999, 2000-2004, 2005-2009, 2010-2014, 2015-2019) and the loss data of year 2020 is not included in any of the groups. Now taking each group as testing data, the damage functions are trained on the rest of the data. The predictive skills of the damage functions are evaluated on the testing data. The large spread in the model skill metrics (i.e., MAE and CV) indicates that the performance of damage functions is highly dependent on the choice of the training data (Fig. R2). For each model, the spread in the number of municipalities showing the smallest MAE (such as done in Table 1) remains relatively low across all loss classes, as defined in section 3.1 of the manuscript (Fig. R2 c, f, i). The black dots in Fig. R2 shows the results present in the manuscript (Table 1) obtained with another set of training and testing data. We notice that they often lie outside the interquartile range, especially when considering all loss days and the extreme loss days (top and bottom rows in Fig. R2), and are sometimes even outside the range of the seven-fold cross-validation analysis, emphasising again the strong dependence of the results to the chosen training and testing periods. In light of this new analysis, we will include these results in the manuscript and figure R2 in the supplement.

We will add some sentences in the manuscript on this topic, such as:

In section 2.6 (model evaluation section):
The damage functions are sensitive to extreme loss observations and the presence of few extreme events can heavily alter the damage functions' shape. Therefore, different training data sets may result in differing damage function fits. Cross-validation is an effective method to estimate the uncertainties involved in the choice of the testing and training data. We perform a seven-fold cross-validation by splitting data into seven with each set of testing data having five consecutive years of data. So, in the first fold the testing period is 1985-1989 and the training period is 1990-2020, in the second fold the testing period is 1990-1994 and the remaining years are in the training dataset, and so on.

In section 3.2:
The seven-fold cross-validation reveals that the parameters in the storm-damage functions obtained during the fitting step depend on the choice of the training dataset. Moreover, the loss estimates are also highly dependent on the choice of training dataset (see the range in the model evaluation metrics shown in Fig. R2 a,b,d,e,g,h). However, whatever the training dataset, the Klawa and exponential models still have the best skill in most of the municipalities (as also shown in Table 1) across the different loss classes (see Fig. R2 c,f,i).

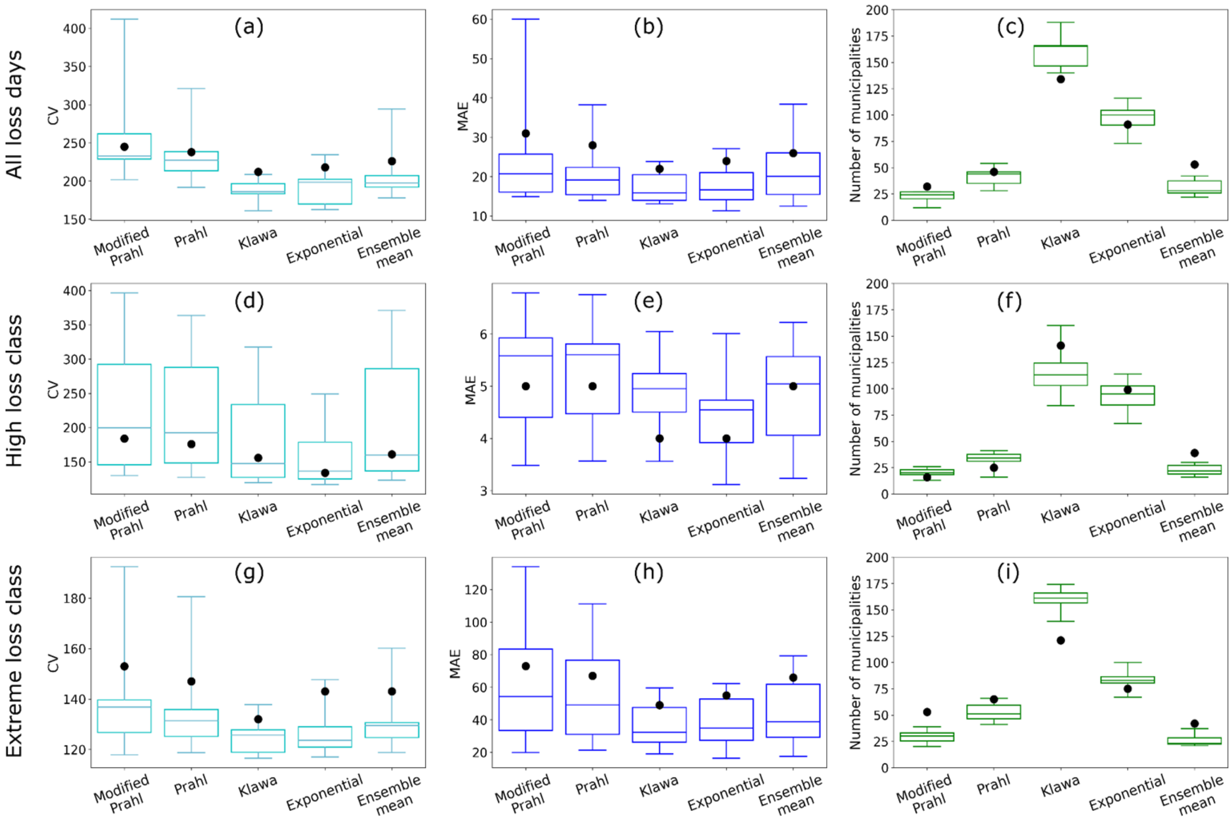

Figure R2: Distribution of model performance metrics from cross-validation (a) coefficient of variance (CV), (b) mean absolute error (MAE), (c) number of municipalities with smallest MAE for four damage functions and their ensemble mean for all loss days. (d), (e) and (f) same as (a), (b) and (c) but for the high loss class. (g), (h) and (i) same as (a), (b) and (c) but for the extreme loss class. The boxes represent the interquartile range, the horizontal line represents the median, the whiskers represent the minimum and maximum and the black dots represent the results from Table 1 in the manuscript.

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.

Comment 5

Line 170: not clear if this statement that 82% of losses are recorded above the 95th percentile is based on the Norwegian loss data analysed against NORA3 wind speeds? Also it is not clear if it refers to 82% of loss events (as count), or 82% of loss values?

Here, 82% corresponds to the sum of the losses over all municipalities in the training period. We will adjust the sentence in the manuscript.

Comment 6

Line 178: In my understanding Klawa developed the function for Germany-wide losses (not districts)?

Thanks for bringing this to our attention. We will replace '*German districts*' with *'Germany'* in line 178.

Comment 7

Line 179: As mentioned further up, Donat et al 2011b was first calibrating the Klawa function at district level for Germany

We will rephrase line 179 to:

Later, using the same insurance data, the damage function was calibrated by Donat et al. (2011b) for the German districts and by Pinto et al. (2012) for the affected areas of individual storm events.

Donat, M. G., Pardowitz, T., Leckebusch, G., Ulbrich, U., and Burghoff, O.: High-resolution refinement of a storm loss model and estimation of return periods of loss-intensive storms over Germany, Natural Hazards and Earth System Sciences, 11, 2821–2833, https://doi.org/10.5194/nhess-11-2821-2011, 2011b.

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

Comment 8

Line 182: You should specify over which time period and which seasons the percentile threshold was calculated (e.g. annual percentile, or percentile over the winter storm seasons ~October to March)?

We will rephrase line 182 to:

This damage function takes the third power of wind speeds above the 98th percentile of the wind speed determined using the whole study period (1980-2020) scaled by the same 98th percentile [...]

Comment 9

Line 197: It may be useful to clarify whether this fixed threshold of 9m/s applies to maximum gust or maximum wind speed from the NORA3 hindcast?

We will add two sentences such as:

To alleviate this, Karremann et al. (2014b) and Little et al. (2023) suggested a 9 m/s fixed threshold for wind speed causing damage in Norway. However, in our study, we do not need this 9 m/s threshold as we use the population-weighted averaged wind speeds, reducing the relative importance of grid cells with very low wind speeds and therefore avoiding the problem of very low 98th percentile.

Karremann, M. K., Pinto, J. G., Reyers, M., and Klawa, M.: Return periods of losses associated with European windstorm series in a changing climate, Environmental Research Letters, 9, 124 016, https://doi.org/10.1088/1748-9326/9/12/124016, 2014b.

Little, A. S., Priestley, M. D., and Catto, J. L.: Future increased risk from extratropical windstorms in northern Europe, Nature Communications, 14, 4434, https://doi.org/10.1038/s41467-023-40102-6, 2023.

Comment 10

Line 276: should “spread” better be “distribution”?

We agree that “distribution” is a better fit than “spread”. We will change it accordingly.

Comment 11

Line 286: I think it is not really data-scarce, I understand that you have good/complete data but these tell you that there are only very few losses?

With ‘data-scarcity’, we meant that there are only very few losses and the reviewers’ observation is correct in this context. We will replace ‘data-scarcity’ with ‘*rarity of loss days*’.

Comment 12

Line 320-325: Based on Table 1 it may be fair to say that seems to be the best function for most municipalities?

We agree with the reviewer and will add that, overall, our results suggest that the Klawa storm-damage function is the best model for a large share of municipalities (~37.6%).

Comment 13

Line 361: insert “is” after “This”

We will do this.

Comment 14

Line 390: remove the “ ‘ “ after “functions”

We will do this.