Natural Hazards
and Earth System
Sciences

Open Access

Supplement of

Comparison of storm damage functions and their performance

B. F. Prahl et al.

Correspondence to: B. F. Prahl (corr@prahl.net)

1 Binomial test for model performance

While the results in Sec. 4.2 of the main paper focus on 10th percentile of the loss distribution, separated into the three defined loss classes, Fig. S1 shows the results of the simple binomial test based on a moving-window approach. Here, we apply a 2 and a 10 percentile window to the model estimates, which were ranked by descending order of the corresponding empirical loss. This approach enables a full view on the relative performance of the individual models also in medium or low loss ranges.

However, care must be taken when interpreting the behavior for small loss events, as the limitation of the different models play a deciding role. In particular the lower bounds of the models P and K, observed in Fig. 5 of the main paper, lead to a strong signal in Fig. S1. For example, pairwise comparisons with model K show an apparently strong performance of model K for a cumulative loss ratio between 10^{-8} and 10^{-7}. However, this effect corresponds to observed losses incidentally matching the lower bound of the model and, hence, does not indicate superior predictive skill. Considering the loss range indicated by the left-hand axis (cumulative loss ratio) it becomes clear that this effect may be rather seen as an artifact at an insignificant loss level.
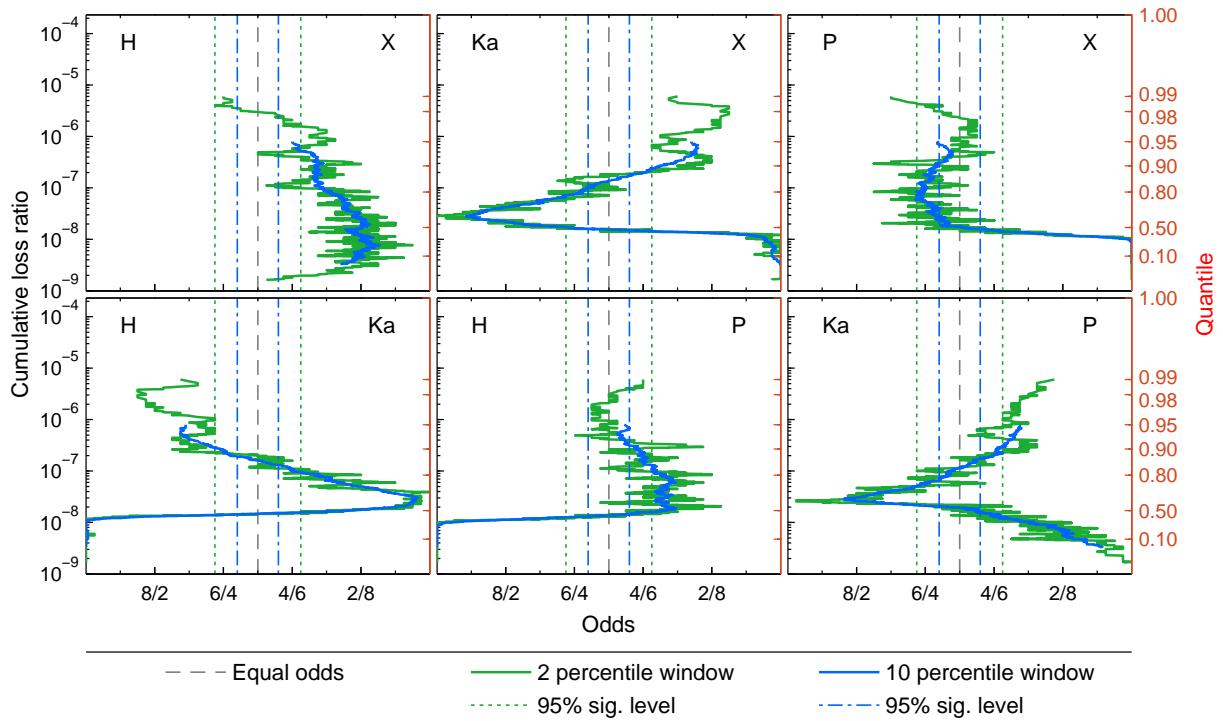
Fig. S1. The figure shows the odds (i.e. the ratio of the total hits – as being closest to the observation – of each of the models in the pairwise comparison) calculated for the simple pairwise binomial test on varying loss ranges. Estimates were ranked in descending order by the corresponding observed loss amount. Moving 2 and 10 percentile windows were used to estimate the pairwise odds. The results are significant if the odds exceed the 95% guidlines obtained from the binomial distribution.

2 Model results for different set-ups of models K and H

Based on their publication, different calibrations options are available for both the model K and the model H. In case of model K, Donat et al. (2011a) perform a regression against annual loss aggregates, while Donat et al. (2011b) demonstrate calibration against a selected sample of the 34 most loss-intensive storm passages. Figure S2 shows the results obtained after both calibrations, the annual [K] calibration employed in our main paper and the storm-based [K storm] calibration.

As outlined in the main paper, our approach to the calibration of model H differs from those originally published by Heneka and Ruck (2008). Whereas we applied a district-wise calibration, Heneka and Ruck (2008) calibrated their model against pooled data, merged from all post-code areas in the federal state of Baden-Württemberg. Following their approach, we have pooled all district data within Baden-Württemberg. In line with Heneka and Ruck (2008), the model was calibrated against absolute DWD wind gust data [H BW] and against relative DWD data [H BW rel] normalised to the 98th wind gust percentile. All results of the different calibrations are shown in Fig. S2.
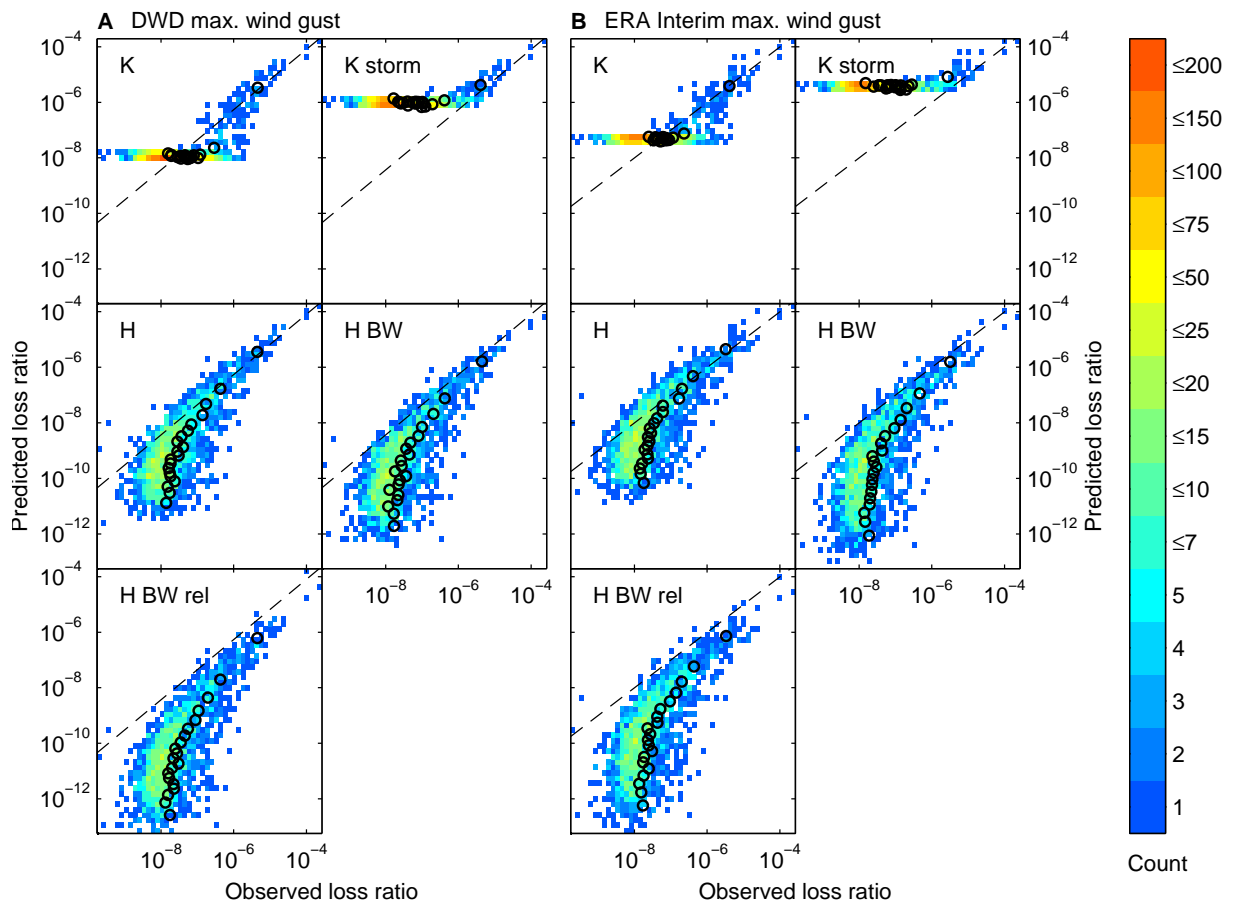
Fig. S2. On country level, the predicted daily loss ratio (expected value) for each model is plotted versus observed losses on double-logarithmic scale. Left-hand panel A shows results based on DWD wind gust data, ERA Interim wind gust data is used in panel B. The colors indicate the 2-D histogram count. The black circles represent (linear) averages of 100 losses each, binned by descending order of predicted loss. Black dashed lines have unity slope and indicate equality of observation and prediction.

3 Countrywide results of the binomial test and MPE/MAPE metrics obtained from ERA Interim data

The following Tables S1 and S2 show countrywide results obtained from ERA Interim wind gust data. They correspond to the DWD based Tables 3 and 4 in the main paper. The results based on ERA Interim are very similar to those obtained from DWD data, with the exception of loss class one, where model K appears to excel. The results should, however, be interpreted in conjunction with Fig. S4, which indicates that the loss accumulated over the entire passage duration of the 6 most severe winter storms (with peaks corresponding to the 6 most severe loss days in loss class I is not so well reproduced by model K.

Table S1. Results from a binomial test for prediction accuracy of the different models based on ERA Interim wind gust data. The model of each column is tested against each row of competing models and across loss classes (as defined in Tab. 1 of the main paper). Bold results indicate superiority of the tested model with statistical significance greater than 95%.

Loss class	Test vs.	Share of closest loss estimates in % (p-value)			
		X	P	K	H
I	X	–	**83 (0.02)**	**100 (0.00)**	67 (0.11)
	P	17 (0.89)	–	**83 (0.02)**	33 (0.66)
	K	0 (0.98)	17 (0.89)	–	50 (0.34)
	H	33 (0.66)	67 (0.11)	50 (0.34)	–
II	X	–	**79 (0.00)**	29 (0.99)	**71 (0.00)**
	P	21 (1.00)	–	26 (1.00)	18 (1.00)
	K	**71 (0.00)**	**74 (0.00)**	–	**71 (0.00)**
	H	29 (0.99)	**82 (0.00)**	29 (0.99)	–
III	X	–	53 (0.24)	34 (1.00)	34 (1.00)
	P	48 (0.71)	–	38 (1.00)	47 (0.76)
	K	**66 (0.00)**	**63 (0.00)**	–	**64 (0.00)**
	H	**66 (0.00)**	53 (0.19)	36 (1.00)	–

Table S2. Estimates of the mean absolute percentage error (MAPE) and mean percentage error (MPE) for each of the competing models and across loss classes (as defined in Tab. 1 of the main paper) based on ERA Interim wind gust data. Best values for each class are emphasized in bold.

Loss class	Model MAPE (MPE) both in %			
	X	P	K	H
I	250 (220)	62 (9)	**57 (0)**	91 (43)
II	133 (84)	**90 (45)**	108 (46)	102 (49)
III	**82 (16)**	95 (40)	85 (-51)	86 (19)

4 Results for the six most severe winter storms obtained from ERA Interim data

Figure S4 shows model estimates for the 6 most severe winter storms during the period under observation. The results are based on ERA Interim data and correspond to Fig. 6 of the main paper, which is based on DWD wind gust data. Overall, the ERA Interim based results show similar relative intervals than those obtained from DWD data. However, there appears to be a general bias in the wind gust data, e.g. shown by the pronounced overestimation for winterstorm *Anatol* and the underestimation of winterstorm *Anna*.
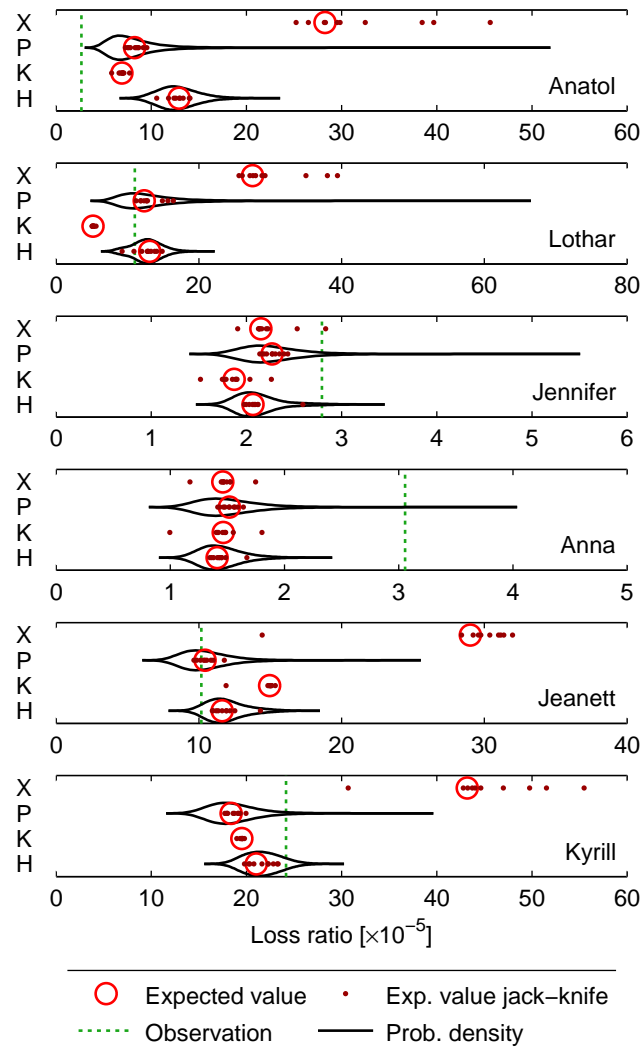

Fig. S3. Model estimates for the 6 most severe winter storms in the period 1997-2007 based on ERA Interim data. Red circles indicate the expected value obtained from models trained on full 10-year data, while the red dots represent expected values from the 9-year resampled (*jackknife*) training periods. For models P and H, the black contours reflect the probability distribution of predicted storm loss for the 10-year training data. Empirical insured loss is marked by green dashed lines.

References

Donat, M. G., Leckebusch, G. C., Wild, S., and Ulbrich, U.: Future changes in European winter storm losses and extreme wind speeds inferred from GCM and RCM multi-model simulations, Nat. Hazards Earth Syst. Sci., 11, 1351–1370, http://dx.doi.org/10.5194/nhess-11-1351-2011, 2011a.

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

Heneka, P. and Ruck, B.: A damage model for the assessment of storm damage to buildings, Eng. Struct., 30, 3603 – 3609, http://dx.doi.org/10.1016/j.engstruct.2008.06.005, 2008.