Nat. Hazards Earth Syst. Sci., 15, 769–788, 2015
www.nat-hazards-earth-syst-sci.net/15/769/2015/
doi:10.5194/nhess-15-769-2015

Natural Hazards and Earth System Sciences
Open Access

Comparison of storm damage functions and their performance

B. F. Prahl[1], **D. Rybski**[1], **O. Burghoff**[2], **and J. P. Kropp**[1,3]

[1]Potsdam Institute for Climate Impact Research (PIK), Potsdam, Germany
[2]German Insurance Association (GDV), Berlin, Germany
[3]Department of Earth and Environmental Sciences, University of Potsdam, Potsdam, Germany

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

Received: 15 July 2014 – Published in Nat. Hazards Earth Syst. Sci. Discuss.: 10 September 2014
Revised: 12 December 2014 – Accepted: 17 March 2015 – Published: 9 April 2015

Abstract. Winter storms are the most costly natural hazard for European residential property. We compare four distinct storm damage functions with respect to their forecast accuracy and variability, with particular regard to the most severe winter storms. The analysis focuses on daily loss estimates under differing spatial aggregation, ranging from district to country level. We discuss the broad and heavily skewed distribution of insured losses posing difficulties for both the calibration and the evaluation of damage functions. From theoretical considerations, we provide a synthesis between the frequently discussed cubic wind–damage relationship and recent studies that report much steeper damage functions for European winter storms. The performance of the storm loss models is evaluated for two sources of wind gust data, direct observations by the German Weather Service and ERA-Interim reanalysis data. While the choice of gust data has little impact on the evaluation of German storm loss, spatially resolved coefficients of variation reveal dependence between model and data choice. The comparison shows that the probabilistic models by Heneka et al. (2006) and Prahl et al. (2012) both provide accurate loss predictions for moderate to extreme losses, with generally small coefficients of variation. We favour the latter model in terms of model applicability. Application of the versatile deterministic model by Klawa and Ulbrich (2003) should be restricted to extreme loss, for which it shows the least bias and errors comparable to the probabilistic model by Prahl et al. (2012).

1 Introduction

As a major contribution to natural-hazard damages, windstorms are responsible for an average of 39 % of worldwide economic losses during 1980–2011 (Munich Re, 2013). Across Europe, losses from meteorological events are mainly caused by winter storms and comprise 68 % of total insured loss from natural catastrophes. The largest event so far, winter storm Daria in 1990, totalled USD 8.6 billion of insured loss in 2013 values (Swiss Re, 2014).

Recent climatological studies by Schwierz et al. (2010) and Held et al. (2013) indicate that the severity of winter storm-related loss is likely to increase markedly in the course of the 21st century. While there is no consensus on changes of winter storm frequency, a growing body of research supports a future increase in storm intensity (Feser et al., 2015). With this development in mind, it is questionable whether the anticipated damages will remain within the limits of insurability. Even though Held et al. (2013) come to a positive conclusion for the German insurance market, such analyses hinge on the choice and quality of the employed damage function.

A storm damage function describes the relation between the intensity of a storm and the typical monetary damage caused. While on the continental scale storm intensity can be best described by complex storm severity indices (Deroche et al., 2014; Roberts et al., 2014), local losses are ultimately caused by surface winds. As the magnitude of storm loss is highly sensitive to changes in wind speed, even small variations between potential damage functions could have severe implications for the reliability of loss estimates and their validity for economic and political decision making.

The work in hand tackles this issue by providing a model intercomparison of storm damage functions for the residential sector in the context of European winter storms.

In the discussion of storm damage functions, it is often assumed that loss should increase as the square or cube of the maximum wind (gust) speed. These presumptions originate from the following:

- the consideration of wind loads, which are approximately proportional to the exerted pressure and, hence, to the square of the wind speed (e.g. Simiu and Scanlan, 1996);
- the concept of proportionality between structural damage and the dissipation rate of the wind kinetic energy that scales with the third power of wind speed (recently: Emanuel, 2005; Powell and Reinhold, 2007; Kantha, 2008).

In particular, the notion of a cubic relationship is backed by empirical analysis of insurance records, which appear to exhibit cubic or quartic behaviour depending on the storms under scrutiny (Munich Re, 1993, 2001). However, recent literature provides evidence for a much stronger increase of insured storm loss with wind gust speed (Huang et al., 2001; Heneka and Ruck, 2008). For the insurance data set that we employ here, Prahl et al. (2012) found a power law with regionally varying exponents that approximately range between 8 and 12.

We reason that the apparent contradiction results from the negligence of a potential loss threshold due to insurance deductibles or similar economic effects. Thus, we schematically demonstrate the transition from very steep loss increase to a more modest cubic power-law.

The comparison of storm damage models is generally impeded by inconsistencies for reasons of (i) differing temporal or spatial resolution of meteorological data, (ii) deviating building codes and enforcement practices, and (iii) differing insurance policies and claims settlement practices (Walker, 2011).

In order to circumvent such inconsistencies, three recently developed damage functions (Klawa and Ulbrich, 2003; Heneka and Ruck, 2008; Prahl et al., 2012) are applied to a common data set of wind gusts and insurance loss data for Germany. These damage functions are complemented by a simple exponential model inspired by recent US hurricane loss models (Huang et al., 2001; Murnane and Elsner, 2012), yielding four mathematically distinct modelling approaches. For simple referencing, we assign the acronyms X and K to the deterministic exponential model and the model by Klawa and Ulbrich (2003), respectively. The probabilistic models by Prahl et al. (2012) and by Heneka and Ruck (2008) are referred to via the letters P and H, respectively.

The theoretical foundations and the implications of each model are discussed in order to mainstream terminology and conceptual structure of storm damage functions. Quantitative results are obtained from numerical estimation and allow a direct comparison of model performance under varied spatial aggregation, relating to either daily loss or particular major storms. During summer months, the employed loss data inseparably includes both wind and hail damages. Since the employed damage functions concern wind damage only, we limit the work in hand to days within the winter half-year (abbreviated as WH), comprising the months October through March.

We address the validation of countrywide loss estimates by applying a novel pairwise binomial test metric in conjunction with the relative metrics mean percentage error (MPE) and mean absolute percentage error (MAPE). Furthermore, a coefficient of variation is employed to assess the predictive uncertainty on district level at daily resolution.

The overall model estimation is based on annual cross validation, an iterative procedure for the sampling of the training data, safeguarding that loss estimates within any given year are obtained from independent training samples. We furthermore assess model robustness by employing a *jackknife* method for the systematic resampling of training data. Selectively excluding parts of the training sample, the jackknife method allows us to assess the dependence of model estimates on the training data. Probabilistic model results are obtained from a Monte Carlo simulation with a sample size of 1000.

In the following section, we give overviews of the employed wind gust and insurance data sets and of the model estimation procedure. In Sect. 3 a brief introduction of storm damage functions is followed by a detailed view on each of the compared models. The numerical modelling results are discussed in Sect. 4. In Sect. 5 we attempt a synthesis between a cubic wind–damage relation and the considerably steeper damage functions reported for German winter storms. The concluding synopsis and discussion of the theoretical and numerical aspects of the impact model intercomparison are given in Sect. 6.

2 Data and methods

2.1 Insurance data

In this work, the employed damage functions are calibrated against detailed insurance loss data obtained for storm damages to residential buildings. The German Insurance Association (GDV) provided loss data relating to the "comprehensive insurance on buildings" line of business resolved for 439 German administrative districts (as of 2006).

The data set comprises the magnitude of absolute losses and insured values as well as the number of claims for the years 1997 to 2007 on a daily basis. With its high spatiotemporal resolution and countrywide coverage, the GDV data set has been successfully applied for the calibration of different

Table 1. The three loss classes defined for the winter half-year. Given are the number of observations, the related quantiles, and the accumulated loss share for the period 1997 to 2007.

Loss class	Description	No.	Quantiles of daily losses	Loss share
I	Extreme	6	0.997–1.000	54.9 %
II	Large	34	0.980–0.997	23.4 %
III	Moderate	160	0.900–0.980	15.0 %

damage functions (e.g. Donat et al., 2011b; Prahl et al., 2012; Gerstengarbe et al., 2013).

In order to eliminate price effects and time-varying insurance market penetration, we consider relative figures for the amount of loss and claims throughout. The following definitions are applied:

- loss ratio (LR): the amount of insured loss per day and district, divided by the corresponding sum of insured value;
- claim ratio (CR): the number of affected insurance contracts per day and district, divided by the corresponding total number of insurance contracts.

These definitions are based on the assumption that insured buildings are randomly distributed in each district and are representative of the overall residential building stock. With data coverage of up to 13.4 million insured buildings and in excess of 90 % market coverage (GDV, 2013) we expect the assumptions to hold.

The highly skewed and heavy-tailed distribution of daily losses during the winter half-year is illustrated in Fig. 1. More than 50 % of total loss is recorded for the top 6 out of 2000 loss days. The shaded area in Fig. 1 highlights the upper 10 % of loss days, comprising in excess of 90 % of total loss. For economic relevance, our work focusses on this loss segment, with a sub-division into three distinct loss classes, as shown in Table 1.

The vast number of days exhibiting negligible insured loss appears to be due to a random scattering of small losses across time and districts. Supporting the attribution to noise, Prahl et al. (2012) found a direct proportionality between the magnitude of the temporally scattered losses and the number of insured contracts in a given district.

2.2 Wind gust data

Two sets of meteorological data were employed. The first set comprises daily maxima of the 3 s wind gust measured by the German weather service DWD[1] (Deutscher Wetterdienst). Applicable meteorological stations were selected according to the following criteria:

1. Missing values may not exceed 20 days for each year.
2. Average missing days per year may not exceed 10 for the period 1996 to 2008.
3. Stations should exclude mountainous stations above 1400 m a.s.l.

Based on the selection criteria, 85 meteorological stations were selected. Measurements obtained at anemometer heights other than 10 m were adjusted using the simple wind-profile power law

$$v(10) = \left(\frac{10}{h}\right)^{\lambda} v(h), \quad (1)$$

with wind velocity v, anemometer height h, and an exponent $\lambda = 1/7$ as discussed in Wan et al. (2010).

Inhomogeneities in meteorological times-series can be identified by finding an optimal solution to the multiple breakpoint problem. Standard methods are available, in particular, for finding inhomogeneities in monthly climatic time series (Venema et al., 2012). Application to daily time series is however subject to ongoing research (e.g. Wang, 2008; Mestre et al., 2011).

In the case of daily block maxima of climatic data, the relatively small change at the breakpoint as compared to the data's variance and the presence of long-term persistence adversely affect the capacity to identify breakpoints correctly. With a low signal-to-noise ratio, the presence of long-term correlation can lead to false identification of breakpoints (Rybski and Neumann, 2011; Bernaola-Galván et al., 2012).

We attempt to avoid over-detection by applying a conservative testing scheme based on multiple cross-comparison of neighbouring stations and the examination of metadata, e.g. about relocation of stations. The testing scheme employs the R implementation of the PMFred algorithm developed by Wang (2008) to identify potential breakpoints in time series of differences between daily gust maxima of any pair of meteorological stations. We reduced the skew of the distribution of gust speeds by applying a logarithmic transformation and hence improved the normality of the data, which constitutes a basic assumption of the PMFred algorithm.

To begin with, we chose a control group of 39 stations whose individual time series showed no significant inhomogeneities in the test algorithm. Subsequently, we paired each of the 85 stations with the 10 closest of the control group and performed the PMFred algorithm on the time series of their differences. If, within a 60 day window, at least three pairwise tests indicated a breakpoint that could be backed by metadata, the inhomogeneity was corrected. Furthermore, if all 10 pairwise comparisons suggested a significant and otherwise undocumented breakpoint it was also corrected. All corrections were performed using a quantile-matching algorithm (Wang et al., 2010).

Overall, we took a conservative stance on artificial manipulations of the raw time series and corrected only three sig-

[1] Data available at: http://www.dwd.de/webwerdis.

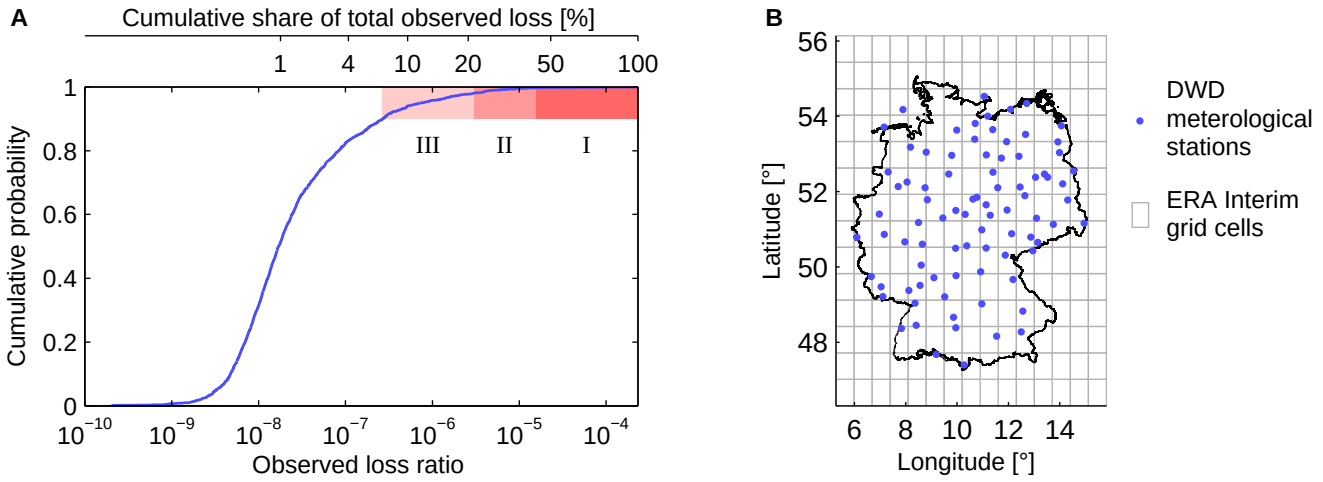

Figure 1. (a) shows the empirical cumulative distribution function of loss days in Germany during the winter half-year. The observations comprise 2000 loss days, which exhibit a steep increase of loss at the upper end of their distribution. The shaded area indicates the days within the upper 0.1 quantile, subdivided into the three loss classes defined in Table 1. The top scale shows the share of total loss that is accumulated for all losses smaller than or equal to a specific loss ratio. **(b)** shows the spatial distribution of the employed DWD stations and the ERA-Interim grid cell resolution.

nificant breakpoints in total, two of which were documented in metadata.

The second wind gust data set was obtained from the ERA-Interim reanalysis project[2] (Dee et al., 2011). We use the daily maxima of the 3-hourly values of the 10 m wind gust.

Both sets of wind gust data, DWD and ERA-Interim, require a downscaling to match the resolution of the insurance data. Prahl et al. (2012) demonstrate that wind gust observations from neighbouring meteorological stations provide sufficient information for the calibration of a storm damage function. Higher precision may be attained via the use of mesoscale climate models for the computation of detailed and physically valid wind fields from reanalysis or observational data (Heneka et al., 2006; Huttenlau and Stötter, 2011). As this is clearly beyond the scope of our work, we limit ourselves to a simple inverse-distance interpolation scheme applied to both DWD and ERA-Interim data sources. The wind field was interpolated at the centroids of each district, taking into account all locations (stations or grid points) within a certain radius of interaction. Employing leave-one-out cross validation, i.e. iteratively excluding each individual location from the interpolated data set, we calculated the average correlation between empirical and interpolated values at varying radii of interaction. The optimal radius of interaction was chosen as the value at which the average correlation reached its maximum. The estimated radii were 130 and 60 km for DWD and ERA-Interim, respectively.

[2]ERA-Interim data were obtained from: http://data-portal.ecmwf.int/data/d/interim_full_daily.

2.3 Model calibration

The analysis of daily insurance loss data of the winter half-year reveals an extremely broad and strongly skewed loss distribution. Relating loss and wind gust data, a pronounced heteroscedasticity is revealed (cf. Fig. 6 in Heneka and Ruck, 2008), with uncertainty resembling a log-normal error (Prahl et al., 2012). In conjunction with such pronounced heteroscedasticity, the scarcity of extreme events in the tail of the distribution may cause a bias of traditional regression methods, such as least squares, towards singular extremes present in the training data. While a data transformation, such as the logarithm, may reduce skew and heteroscedasticity, it would put stronger weight on smaller loss events and hence counteract the focus on extremes. In practice, potential data transformation and curve fitting methods are dependent on the specific damage model and are hence discussed in conjunction. Calibration issues that arise from the properties of the loss distribution are discussed alongside the mathematical model concepts in Appendix A.

2.4 Model estimation procedure

Since damage functions are typically employed as predictive models, it is of key importance how accurately they perform in practice. In addition to choosing the optimal model, there is the risk of overfitting to a training data which may not represent the high variability of weather extremes. In order to assess the predictive performance of the employed models, a k-fold cross validation scheme (Kohavi, 1995) is employed in conjunction with a jackknife procedure (Miller, 1974).

For annual cross validation, the 11-year data set is partitioned into annual subsamples. Iteratively, each individ-

ual subsample is retained for evaluation, while the model is trained in the 10 years remaining. This process ensures that each year is used exactly once for evaluation.

The employed cross validation enables out-of-sample prediction for each day and allows for the assessment of the model fit with regard to the range of frequently occurring losses.

However, for very scarce extreme events the evaluation of model robustness requires additional resampling of the training data. The resampling is performed via a jackknife procedure, where each individual annual subsample is excluded consecutively from the 10-year training sample.

For the joint analysis of deterministic and probabilistic models, two different schemes for loss aggregation are employed. Generally, we consider the daily district-wise loss estimates as independent random variables dependent only on the maximum gust speed. In the case of deterministic models, the model estimates are interpreted as expected values and were simply summed up over time or space. For the probabilistic models, we employed a Monte Carlo approach, where results of 1000 independent random realisations were aggregated. The expected value and distribution quantiles were then calculated from the distribution of Monte Carlo estimates at the desired level of aggregation.

2.5 Validation metrics

The broadness and skew of the loss distribution also play a role for the validation of model estimates, as they have significant impact on the applicability of evaluation metrics. Heteroscedastic dependence between prediction error and loss magnitude invalidates traditional moment-based metrics, such as R^2 or, equivalently, Pearson's ρ. In particular, very extreme events may attain the character of singularities and dominate absolute performance metrics. Alternatively, relative metrics such as mean percentage error (MPE) or mean absolute percentage error (MAPE) may be employed over well-defined loss ranges (Hyndman and Koehler, 2006). However, these metrics fail if predictions comprise both days with and days without loss, which is often the case for daily resolved data. Moreover, such zero values prevent the use of common transformations (e.g. power transformations such as Box–Cox Transformation, see Box and Cox, 1964) to increase the normality of the loss distribution required for most statistical metrics.

In order to eliminate the effects of scale of the loss distribution for model comparison, we propose a simple pairwise statistical test based on binomial statistics. The null-hypothesis is that both models have equal predictive skill and, hence, that their predictions are equally likely to be closest to the true observations. Successes (i.e. closer prediction) can be represented by independent Bernoulli trials with probability 0.5. In a one-tailed test, the binomial distribution then expresses the probability for a given success rate.

In order to apply the binomial test, the share of predictions where one or the other comes closer to the observation is estimated for each pairing of models. Significance is obtained from the binomial distribution with probability 0.5 and n independent trials, where n equals the total number of loss days for each loss class.

As the binomial test itself does not disclose why any specific model outperforms a competitor, we interpret the results of each model in conjunction with traditional relative metrics relating to a multiplicative error. For the employed data, Prahl et al. (2012) found a variability that is approximately symmetric on the log-scale, such that the assumption of a multiplicative error seems viable.

The employed multiplicative metrics are the mean absolute percentage error (MAPE, i.e. the mean of the moduli of deviations between model estimates and observations in percent) and the mean percentage error (MPE, i.e. the mean of the deviations between model estimates and observations in percent). While MAPE gives an estimate of the variability of model results, MPE provides an indication for systematic bias.

3 Storm-damage models

A damage function describes the relation between the intensity of a specific hazard and the typical monetary damage caused with respect to either a single structure (*microscale*) or a portfolio of structures (*macroscale*).

Microscale models can be empirical (i.e. statistically derived from data), engineering-based, or a mixture of both. On the macro scale, damages may be either aggregated from microscale models or obtained from statistical relationships based on empirical data (cf. Merz et al., 2010).

Due to the minimum resolution of our data (i.e. districts), our analysis is constrained to the macroscale models of the latter kind. Nonetheless, some of the damage functions under scrutiny contain assumptions on the nature of microscale damage. As there are no publicly available engineering-based models for our region of interest, only statistical models are considered.

For a general overview of modelling approaches, both statistical and engineering-based, we refer the reader to Walker (2011) and, with a focus on hurricane damage, to Pita et al. (2013). In the following, we present each of the four employed damage functions.

3.1 Generic exponential damage function [X]

The choice for an exponential damage function is motivated by empirical observation, showing quasi-linear increase of the logarithm of the loss ratio versus maximum wind (gust) speed over a wide range (e.g. Prettenthaler et al., 2012; Murnane and Elsner, 2012).

It is a non-physical damage function in the sense that it does not saturate with increasing wind gust speed and thus ignores an upper limit of physical damage. However, average loss levels reached during European winter storms typically range below or around a few tenths of a percent of insured value, such that loss saturation does not become an issue.

The damage function relates the loss ratio L to the exponential of the gust speed v,

$$L_X \propto e^{\mathcal{X}_1 v}. \quad (2)$$

The absolute gust speed is rescaled via a linear transformation governed by parameter $\mathcal{X}_1$. Primarily, the parameter reflects the particular vulnerability to wind damage. Additionally, rescaling of wind gust observations may be required for reasons such as

- variations of scale due to mismatches in altitude or location of the geographical reference of the gust data and the building portfolio
- loss being dependent on a differing wind predictor with approximate proportionality to the maximum gust speed
- systematic bias caused by the interpolation of wind gust data.

The exponential damage functions focuses on wind-dependent losses only. Typically, these are large losses within the upper tail of the loss distribution. For the employed insurance data, small losses that occurred at days with maximum gust speed beneath the 95th percentile show a predominantly random behaviour not captured by Eq. (2) and were hence neglected during calibration. This aspect is also seen exemplarily in Fig. 2, showing the independently trained damage function in the context of empirical loss data.

Further details about the calibration of the damage function are given in Sect. A1.

3.2 Probabilistic power law damage function [*P*]

In the literature, there are several proponents for power-law-based storm damage functions (e.g. Dorland et al., 1999; Nordhaus, 2010; Bouwer and Wouter Botzen, 2011.

For winter storms affecting Germany, Prahl et al. (2012) developed a macroscopic damage function based on the presumption of a power-law-based sigmoid curve. Considering the typical loss range of winter storms, the sigmoid curve can be approximated by a simple power law term. For the general case, their damage function comprises two key components. The first component describes the probability for the occurrence of damage within the portfolio, while the second component models the intensity of loss if a damage has occurred. In conjunction with the introduction of a noise constant, this two-part structure enables the modelling of the entire range of damages, thus not excluding information from the bulk of small losses that may provide additional support for the calibration of the damage function.

For an arbitrary district, Fig. 2 shows the curve fits for both components of the damage function as well as the resulting expected value for storm loss. The left-hand panel demonstrates that the predicted 95 % confidence bounds encompass the majority of loss observations and the right-hand panel shows how the probability of occurrence is inferred from the empirical occurrence rate (training data).

The model can be simplified for large wind gust speeds. In this case, the expected value of loss L is approximately proportional to the gust speed v raised to the power $\mathcal{P}_1$,

$$\mathbb{E}[L_P] \propto v^{\mathcal{P}_1}. \quad (3)$$

The exponent $\mathcal{P}_1$ is the key parameter and expresses the vulnerability of the building portfolio. Additional important parameters adjust the scale of the employed wind gust data and control the spread of the loss probability distribution (see Sect. A2 for details). Concerning the scale of the employed gust data, the observations may require a rescaling to relative values (cf. Sect. 3.1).

The original model published by Prahl et al. (2012) incorporates correlations between district losses caused by the same storm event. Due to the complexity of the employed modelling scheme, it was not feasible to include these correlations in this paper. However, the effect of correlations is perceived as minor to the overall performance of the damage function and their inclusion would lead primarily to a widening of confidence intervals.

Please refer to Sect. A2 for further details of the mathematical derivations and of the fitting procedure.

3.3 Cubic excess-over-threshold damage function [*K*]

Klawa and Ulbrich (2003) proposed a macroscopic damage function for German storm loss based on the hypothesis that storm damages grow with wind gust speed in excess of a specific threshold. The approach has since been applied to other European locations (e.g. Leckebusch et al., 2007; Etienne and Beniston, 2012; Cusack, 2013) and was recently refined to the scale of German districts by Donat et al. (2011b).

At the core of the damage function is the definition of a damage proxy D based on the regional wind gust speed v and its 98th percentile,

$$D = \begin{cases} \left(\frac{v - v_{98}}{v_{98}}\right)^3 & \text{if } v \geq v_{98} \\ 0 & \text{if } v < v_{98} \end{cases}. \quad (4)$$

The damage function is calibrated by performing a linear regression of loss observations against the damage proxy, thus involving two regression parameters (a scaling coefficient and an offset). In the upper limit, the damage function increases without bounds and hence ignores damage saturation at high gust speed.

The scaled damage proxy is shown exemplarily for an arbitrary district in Fig. 2. Since the additive offset parameter rather describes the bulk of loss that may occur below the

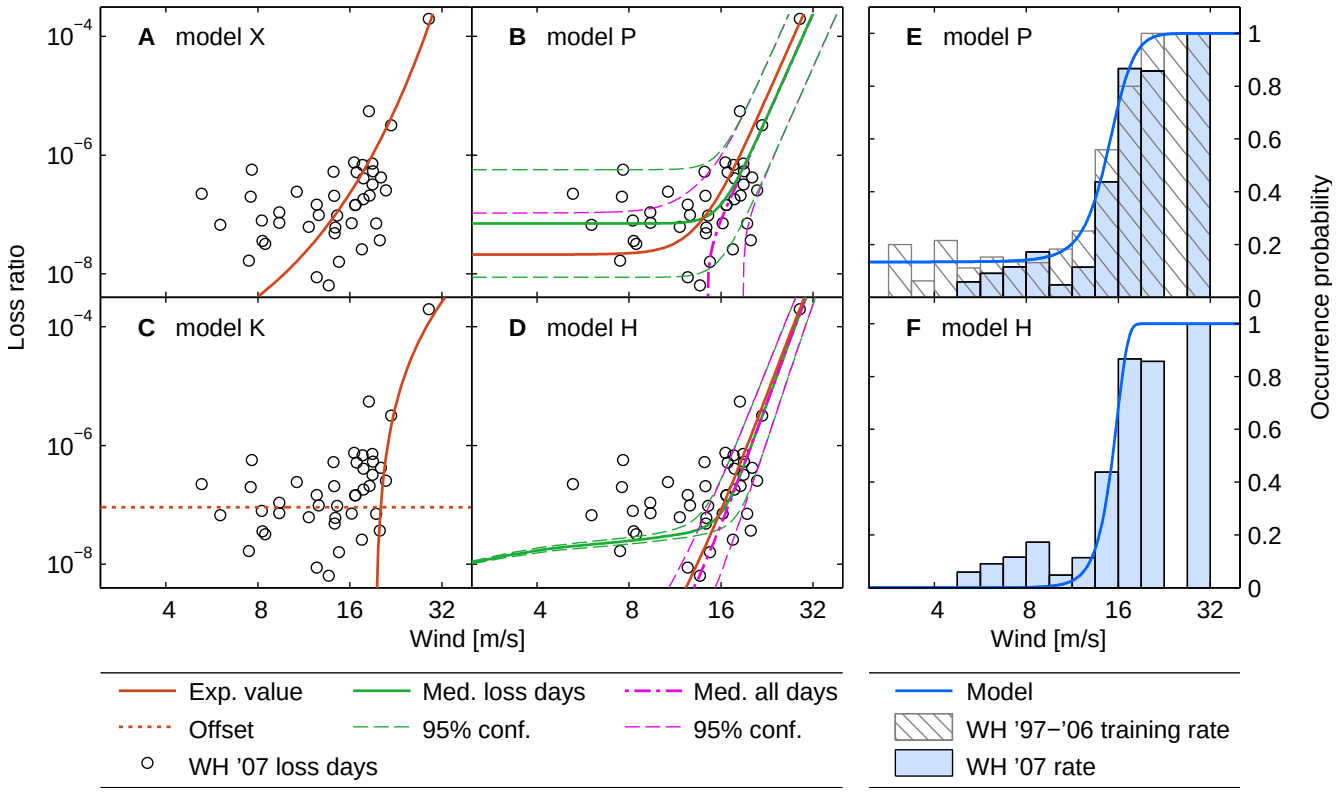

Figure 2. Example of model predictions for a single district obtained from DWD data for the training period 1997–2006 and set in contrast to year 2007 empirical data, all limited to the winter half-year. **(a)**–**(d)** show the expectation values of the loss ratio versus wind gust speed on a log–log scale, circles denote observed losses during 2007. For probabilistic models P and H, the median and 95 % confidence bounds are given. Additionally, we show for model P the median and confidence bounds of the curve fit to actual loss days and for model H an analogous but implied curve. **(e)** and **(f)** show the fitted and implied occurrence rate probability for models P and H, respectively. Year 2007 observed occurrence rates are indicated by blue bars. For model P, training data (shaded bars) is displayed as reference.

98th wind gust percentile, it is not directly attributable to any specific event and hence indicated via a dotted line in Fig. 2.

The employed wind gust percentile was empirically found by Klawa and Ulbrich (2003) and may be considered as a third parameter. Since the introduction of the European Standard EN 1991-1-4 describing the wind action on land structures, the 98th wind gust percentile has become a crucial factor for the reinforcement of buildings against wind damage. Even before its legal implementation during the first decade of the 21st century, it may be reasonable to presume an autonomous adaptation[3] to the wind climate and hence argue for the applicability of a wind percentile as a proxy for such adaptation.

The cubic relationship of the damage function has been repeatedly put into context with the advection of kinetic energy (Leckebusch et al., 2007; Pinto et al., 2007; Cusack, 2013). As a matter of fact, this line of reasoning is problematic due to the subtraction of the 98th percentile threshold, and hence the resulting damage function is inconsistent with the purely cubic dependence on gust speed. As a consequence, the gradient of the damage function is much steeper than that of a simple cubic gust relationship over the entire range of historical wind gust speeds. Only in the upper asymptotic limit, as the gust speed approaches infinity, does the damage function converge to the simple cubic dependence.

In Sect. A3, we demonstrate that on the basis of the employed data the increase of the loss curve for extreme winter storms is comparable to that of a power law with a steep exponent of approximately 10.

Although Klawa and Ulbrich (2003) developed their damage function for winter storms, the function can be applied to the entire loss range, in which case the regression offset parameter serves as baseline loss resulting from wind gusts beneath the defined percentile threshold. Figure 6 illustrates that there is a strong relation between loss and gusts below the 98th wind gust percentile, suggesting that the damage function could potentially utilise a lower wind percentile. Further mathematical details and the fitting procedure are described in Sect. A3.

[3]I.e. structures are reinforced to withstand frequent low-impact events, while adapting to the rare extremes may be too costly. A balance between the individually perceived (monetary) risk and tolerable adaptation cost is maintained.

3.4 Probabilistic claim-based damage function [*H*]

Heneka et al. (2006) put forward an integrated approach for modelling storm loss, combining a probabilistic description of affected buildings with a microscale damage relationship.

Within their theoretical framework, a building damage occurs if a critical wind gust speed, particular to that building, is exceeded. A continuous probability density function is employed to describe the probability of critical gust speeds within the overall building stock. For modelling purposes, Heneka et al. (2006) assumed a Gaussian distribution for critical gust speeds, which is non-physical in a sense as it yields finite probability for negative wind gust speeds. The claim ratio follows naturally as the cumulative distribution function of critical gust speeds, describing the fraction of buildings for which wind gust speed exceeds the critical threshold.

If an individual building i is affected, the damage D_i is assumed to rise as the square of the gust exceedance above threshold until complete destruction is reached at maximum exceedance level $\mathcal{H}_1$. Heneka et al. (2006) (see also Heneka and Ruck, 2008; Heneka and Hofherr, 2011) argue that the square term of their microscale damage relationship,

$$D_i = \left(\frac{v - v_c}{\mathcal{H}_1}\right)^2, \tag{5}$$

corresponds to proportionality between damage and wind force. Repeating the reasoning given in Sect. 3.3, we argue that such proportionality is violated due to the inclusion of the critical threshold v_c, which is inconsistent with the wind force being proportional to the square of the untranslated wind gust speed (e.g. Simiu and Scanlan, 1996).

In contrast to the other discussed damage functions, model fitting and loss estimation requires numerical integration, which makes the application of the damage function computationally more demanding. It was found that the model could not be reliably calibrated on loss data only, necessitating the use of additional data for the number of claims per region and day. Given the additional information from claims data, the damage function would be expected to perform as well or better than the competing models.

Due to its probabilistic description of the building stock, the damage function naturally incorporates an upper limit to the claim and loss ratio and may be applicable to a wide range of losses.

The model requires the calibration of four parameters, describing the wind gust speed at which half of the building stock is damaged and its associated standard deviation, the standard deviation of critical wind gust speeds, and the gust range over which building damages reach complete destruction. Further description of the mathematical details and the three-step calibration procedure is given in Sect. A4.

For an exemplary district, Fig. 2 shows the expected value and 95 % confidence bounds of the damage function. For better comparison with the probabilistic power law damage function, we further decomposed the damage function into the implied components for the occurrence probability and the loss intensity, both shown in Fig. 2.

4 Comparison results

Bringing together the four different models, the two wind gust data sources, and the modelling procedure (Sect. 2.4), model predictions were obtained for 2004 days (consisting of the winter halves of 11 years) and for each of the 439 administrative districts.

Due to the high level of detail, the presentation of results is focused on three distinct aggregation levels: (i) daily loss per district, (ii) daily countrywide losses, and (iii) countrywide losses caused by the six most severe storm events during their entire passage duration.

In case of models K and H, different setups for model calibration were possible (cf. Appendix A). For greater clarity, only those results that relate to the best-performing setup are reported, while additional results are provided in the Supplement.

The circumstances of comparing two deterministic and two probabilistic models require the choice of a common metric. The output of the deterministic models is hence considered equivalent to an expected value obtained from the probabilistic models and forms the basis of the model intercomparison.

4.1 Daily loss per district

While temporal or spatial aggregation generally leads to a convergence of model estimates and observations, strong variability is expected for daily storm loss estimates on the fine district scale.

On the basis of root-mean-square error, we define a coefficient of variation

$$\mathrm{CV_{RMSE}} = \frac{1}{\overline{x}}\left(\frac{1}{n}\sum_{i=1}^{n}\left(x_i - \hat{x}_i\right)^2\right)^{\frac{1}{2}}, \tag{6}$$

where, for n samples, x and $\hat{x}$ denote the observations and estimates of the expected value, respectively. Values are normalised to the mean of the observations $\overline{x}$.

Table 2 shows regional averages of $\mathrm{CV_{RMS}}$ for each of the four competing models. These results highlight the interdependence between model and wind gust choice. While model H mostly outperforms the competing models for DWD wind gust data, it appears less suited for ERA-Interim wind gust data, whose distribution properties are distinctly different from those of the DWD data. Of particular interest is the fact that, irrespective of the wind gust data source, model H performs best across southern Germany. With relatively complex terrain and less frequent storm events, this region poses the greatest challenge to the damage models, resulting in a wide spread of coefficient values between different models. In contrast, model K appears to be least reliable

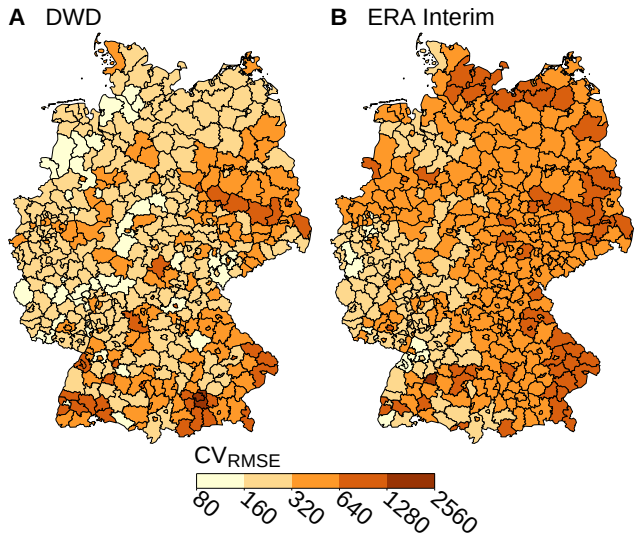

Figure 3. Coefficients of variation of the root mean squared error per district, evaluated for the entire 11-year modelling period. Depicted is the minimum value of CV_{RMSE} found for any of the four models. **(a)** and **(b)** show results obtained from DWD and ERA-Interim wind gust data, respectively.

Table 2. Spatial averages of the coefficient of variation (RMSE) for each model. For ease of comparison, values are sorted in ascending order. The respective model is indicated by the colour code. The spatial extent is defined by the four geographic regions (north, east, south, west) depicted in the map inset.

	North	East	South	West	All
DWD	231	342	436	228	331
	248	384	580	255	385
	266	403	911	290	508
	290	552	997	352	578
ERA Int.	356	327	417	286	376
	401	333	469	299	387
	515	342	738	305	458
	842	580	745	527	665

Model color code

X	H	P	K

in the south. While the exponential model X fares worst overall, it scores best for DWD wind gust data over northern Germany. It may be assumed that in this region the probability distribution of the DWD wind gust data are most favourable for the steep exponential model. Overall, models H and P show the least variation throughout. While model K performs well, with the exception of southern Germany, the exponential model consistently generates the largest amount of variation and, hence, modelling error.

Due to the fact that the district resolution exceeds the resolution of sampling points of the wind field, a strong influence of the choice of gust data are expected. Figure 3 shows a baseline CV_{RMSE} estimated as the minimum value found for any of the four competing models. The DWD-based values show relatively small variation across north-western Germany, while exhibiting stronger variation in southern Germany. In contrast to the DWD-based values, ERA-Interim-based CV_{RMSE} estimates show a marked increase of variation from east to west. The origin of this effect, however, remains unclear.

4.2 Countrywide daily loss

Our second appraisal of the model performance is based upon countrywide daily losses. The spatial aggregation has the beneficial effects of reducing loss variability and yielding a high number of otherwise spatially separated loss events.

Figure 4 shows the model predictions for the countrywide loss ratio plotted against the observations from insurance data. Focusing the initial examination onto results based on DWD wind gust observations (Fig. 4a), several important aspects are revealed.

First of all, the loss predictions from all models exhibit a very high variability in the range of few orders of magnitude. Since the variability cannot be significantly reduced by model choice, it may be a consequence of other aspects, such as the stochastic nature of the building damage, measurement error of gust speed, or the omission of further explanatory parameters. Secondly, the model variability appears nearly symmetric on the log-scale, indicating a strongly skewed distribution. In this case, expected values may be significantly lower than loss observations that fall into the upper tail of the uncertainty distribution.

Two models, K and P, show a lower bound for the expected value of predicted loss. In the case of K, this is a direct consequence of model design, which involves a constant baseline loss that accounts for any loss beneath the local 98th wind gust percentile. For model P, a similar lower bound exists, which reflects the expected value of the noise level present in the loss data at any wind gust speed.

When considering the binned loss ratios (black circles) in Fig. 4a, both models X and H exhibit an underestimation of small losses, which is more pronounced for model H. A comparison with Fig. 2 shows that this behaviour is in line with the rapid convergence to zero of the damage curve for model H. Unsurprisingly, model P shows good agreement of binned loss ratios over a wide range of loss due to the fact that this model is the only one specifically designed to match also the low and medium loss ranges. In comparison,

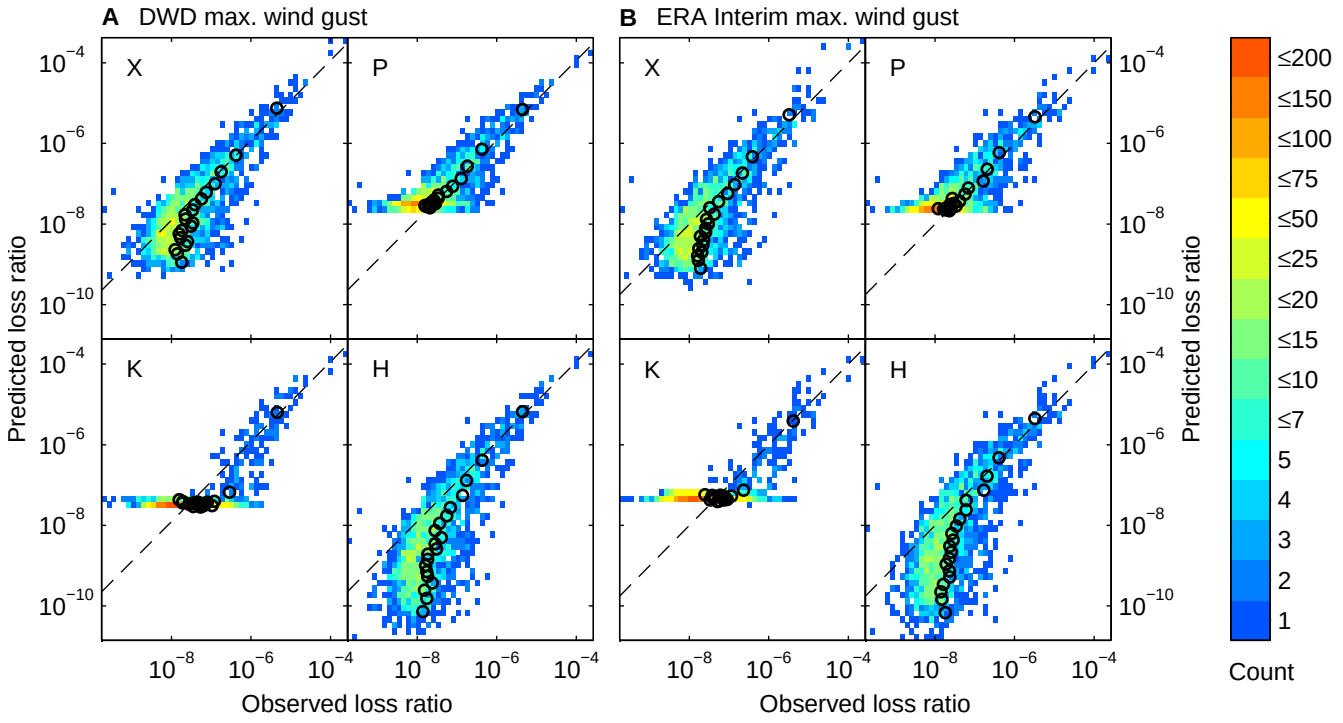

Figure 4. On country level, the predicted daily loss ratio (expected value) for each model is plotted versus observed losses using a double-logarithmic scale. **(a)** shows results based on DWD wind-gust data, ERA-Interim wind-gust data are used in **(b)**. The colours indicate the 2-D histogram count. 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.

model K maps a considerably larger fraction of losses onto its lower bound (baseline loss) and seems to underestimate losses especially in the region around 10^{-6}. This behaviour is a likely outcome of the wind gust threshold fixed to the 98th percentile. Losses near or below this threshold may be strongly underestimated, an effect that plays a larger role for small-scale storms than for extreme large-scale storm events.

For ERA-Interim-driven simulations, Fig. 4b shows a similar overall behaviour as for DWD wind gust data. Comparison indicates a stronger variability of model results for ERA-Interim. Likely causes for this effect are the reduced spatial resolution of ERA-Interim grid cells compared to the spatial distribution of DWD climate stations and the lack of precise geographical allocation of wind gust values attributable only to entire grid cells.

The similarity of results drawn from DWD and ERA-Interim wind gust data prevails for all further model results and we hence focus the subsequent discussion on DWD-based model estimates. The quality (performance) of wind gust data in the context of storm damages is beyond the scope of the work in hand. For special interest we provide results corresponding to ERA-Interim in the Supplement.

It is evident from an economic (or insurance) point of view that the performance for small and mid-range damages should be disregarded in case better performance is achieved for large loss events. In our further analysis we accommodate for this aspect by applying the loss categories defined in Table 1.

In order to compare model results over different loss ranges, we apply a simple scale-independent pairwise statistical test based on binomial statistics. For each pair of models, Table 3 provides the share of predictions where one or the other comes closer to the observation. Values with a statistical significance greater than 95 % are set in bold.

As the binomial test itself does not disclose why any specific model outperforms a competitor, we interpret the results of each model in conjunction with the MAPE (mean absolute percentage error) and the MPE (mean percentage error). Table 4 summarises the results both for MAPE and MPE.

For extreme losses in loss class I the binomial test gives prevalence to the model P, whose estimates exhibit the lowest MAPE. There appears to be indifference between models H and P, although MPE shows that model H tends to overestimate extreme losses, while model P shows a small downward bias. Model K exhibits the least bias and yields the lowest MPE.

Considering loss class II, all models show a strong tendency to overestimate large losses. Here, the smallest bias is produced by H with an MPE of 16 %. Results from P exhibit the least variability of the four models, so that the model can outperform the competitors in the binomial test.

Table 3. Results from a binomial test for prediction accuracy of the different models based on daily loss estimates calculated from DWD wind gust data. The model of each column is tested against each row of competing models and across loss classes (as defined in Table 1). Bold results indicate superiority of the tested model with statistical significance greater than 95 %.

Loss class	Tested against	Share of closest loss estimates in % (p value)			
		X	*P*	*K*	*H*
I	*X*	–	**83 (0.02)**	67 (0.11)	50 (0.34)
	P	17 (0.89)	–	17 (0.89)	67 (0.11)
	K	33 (0.66)	**83 (0.02)**	–	50 (0.34)
	H	50 (0.34)	33 (0.66)	50 (0.34)	–
II	*X*	–	**68 (0.01)**	24 (1.00)	62 (0.06)
	P	32 (0.97)	–	24 (1.00)	35 (0.94)
	K	**76 (0.00)**	**76 (0.00)**	–	**76 (0.00)**
	H	38 (0.89)	**65 (0.03)**	24 (1.00)	–
III	*X*	–	53 (0.24)	25 (1.00)	35 (1.00)
	P	48 (0.71)	–	35 (1.00)	47 (0.76)
	K	**75 (0.00)**	**65 (0.00)**	–	**71 (0.00)**
	H	**65 (0.00)**	53 (0.19)	29 (1.00)	–

Table 4. 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 Table 1) based on DWD wind gust data. Best values for each class are emphasized in bold.

Loss class	Model MAPE (MPE) both in %			
	X	*P*	*K*	*H*
I	56 (49)	**17** (−5)	27 **(–1)**	26 (11)
II	67 (27)	**51** (27)	79 (33)	55 **(16)**
III	**75 (6)**	97 (43)	85 (−51)	**75 (–6)**

In contrast, moderate losses in class III illustrate a completely different behaviour. The biggest change arises for K, which converts from significant overestimation to strong underestimation indicated by a negative bias of −51 %. While the upward bias of P increases for moderate losses, models X and H exhibit only small bias and generally the smallest MAPE.

All above metrics were based on model estimates obtained from DWD wind gusts (cf. Fig. 4a). Tables related to ERA-Interim wind gusts generally show the same tendencies and are given in the Supplement. There, we also provide an additional diagram showing results of the binomial test for small and minor losses below the 0.9 quantile.

4.3 Most severe storm events

Having so far considered only single loss days, Fig. 5 shows the aggregated loss ratios for the six most severe (in terms of loss) winter storms during the observation period. The daily loss estimates were accumulated for the entire passage duration of the respective cyclones, whose start and end dates are given in Table 5.

Table 5. Dates of the six most severe winter storms during the period 1997–2007 (Donat et al., 2011b).

Storm	Start date	End date
Anatol	2 Dec 1999	5 Dec 1999
Lothar	24 Dec 1999	27 Dec 1999
Jennifer	25 Jan 2002	30 Jan 2002
Anna	25 Feb 2002	1 Mar 2002
Jeanett	26 Oct 2002	29 Oct 2002
Kyrill	17 Jan 2007	19 Jan 2007

In addition to the expected value obtained from the full training sample, estimates of the expected value obtained from the jackknife resampling give an indication of the robustness of the model fit. A large spread of jackknife estimates, e.g. as seen for the model X, indicates a strong dependence on the training sample.

Robustness is of particular concern, since the short training period may not always contain very severe storms, and, hence, the storm damage function must reliably extrapolate beyond its support. Empirically, this aspect is illustrated most prominently for winter storms Jeanett and Kyrill, both affecting approximately the same geographical region.

In the case of model K, the outliers of the jackknife estimates for these storms relate to a training sample containing neither one as benchmark. It becomes apparent that the linear regression employed for model K straps the otherwise highly constrained damage function to the maximum level of losses present in the training sample.

With the exception of winter storm Lothar, model P exhibits the least spread of expected values. Even though there are no constraints on the exponent of the damage function as for model K, the model demonstrates robustness due to its larger support from the entire range of observed losses.

A similarly robust behaviour is shown by model H, albeit there appears to be some sensitivity to the training sample for winter storms Jeanett and Kyrill. In contrast to model P, the robustness of model H is likely to originate from the strong constraints imposed on the damage function by the choice of distribution function for the critical gust speed.

The least constrained model X appears not only to be sensitive to the training sample used, but also generates significant overestimation for the three most severe winter storms. Although a verdict may not be based on three events only, the exponential approach appears less reliable for extreme winter storms than the competing models.

Finally, Fig. 5 also shows the probability density contours for the probabilistic models P and H derived from Monte Carlo calculations, convolving all 10 jackknife model fits with 1000 realizations each. While a judgement on the ade-

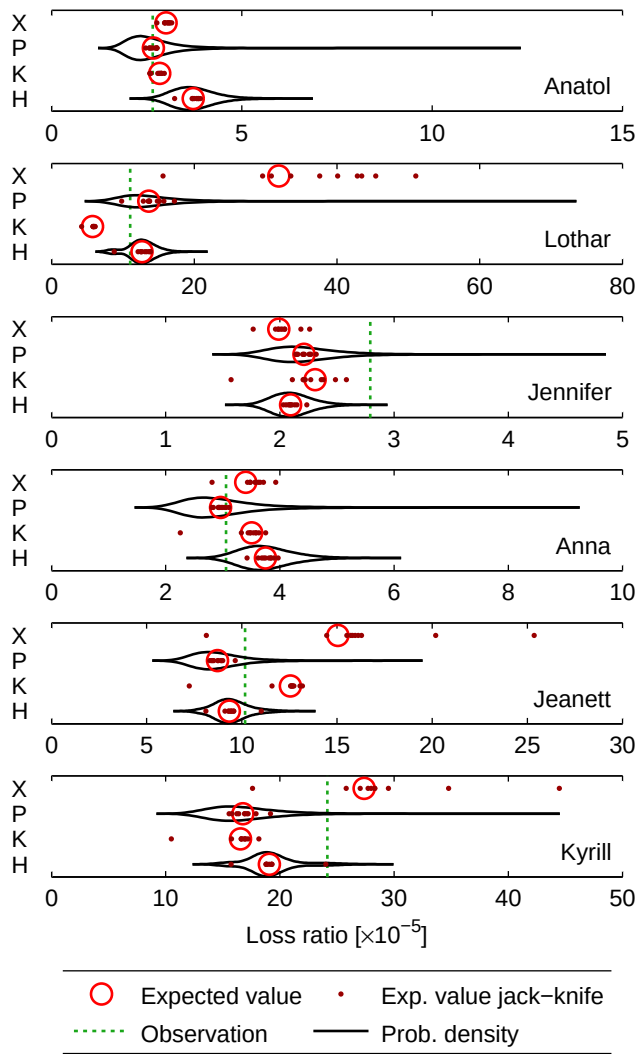

Figure 5. Model estimates for the six most severe winter storms in the period 1997–2007 based on DWD data. Red circles indicate the expected value obtained from models trained on the full 10-year data, while the red dots represent expected values from the 9-year resampled (jackknife) training periods. For models P and H, the black contours represent the probability distribution of predicted storm loss for the 10-year training data. Empirical insured loss is marked by green dashed lines.

quacy of the distributions cannot be made due to the scarcity of extreme events, some observations can be made. Model P, which assumes a log-normal uncertainty distribution with constant scale parameter generates heavily skewed loss distributions that by inspection seem too wide. In contrast, the less-skewed loss distribution produced by H appears more reasonable. In general, both models yield loss distributions that encompass empirical observations.

5 Towards a synthesis of storm damage functions

All of the four different damage functions discussed herein exhibit a loss increase that is much more rapid than a cubic power law derived from physical considerations about the kinetic energy of the wind mass. In this section, we propose a simple mechanism to reconcile the steep loss increase with a cubic power law. With our hypothesis we intend to expedite the discussion on the overall shape of the damage curve, since its behaviour beyond the support has strong implications for the extrapolation of loss.

Figure 6a shows the average loss increase obtained when superimposing data from all German districts. Visual comparison with the power law guiding lines suggests that both the LR and the CR curves increase significantly faster than the 3rd power of wind gust speed. Moreover, the average LR of affected buildings (i.e. those for which an insurance claim was filed) remains approximately constant over a wide range of wind gust speed. This implies a minimum loss threshold for damage compensation to be claimed. Such a threshold could be caused by insurance deductibles, but may also arise from small damages that either go unnoticed or are fixed autonomously.

We make the hypothesis that the steep loss increase that is observed from the GDV data may be a consequence of the presence of such a loss threshold. Mathematically, when applying a threshold T the expected loss ratio $\mathrm{LR_{all}}$ is given by

$$\mathrm{LR_{all}} = \int_{\mathrm{T}}^{\infty} L\, f_v(L)\mathrm{d}L, \tag{7}$$

where $f_v(L)$ denotes the probability distribution of the loss ratio L at gust speed v. The claim ratio CR follows from the respective cumulative distributive function, $F_v(L)$, as

$$\mathrm{CR} = 1 - F_v(T). \tag{8}$$

The loss ratio of affected buildings $\mathrm{LR_{affected}}$ is then simply given by

$$\mathrm{LR_{affected}} = \frac{\mathrm{LR_{all}}}{\mathrm{CR}}. \tag{9}$$

Assuming a log-normal uncertainty distribution, Fig. 6b illustrates the effect of a loss threshold on the expected LR obtained from a simple cubic loss–wind relationship. As a result, for low wind gust speed $\mathrm{LR_{affected}}$ remains close to the threshold value, while $\mathrm{LR_{all}}$ steeply increases. The noise level of the GDV data however entails a minimum loss level, approximately corresponding to a single damaged building per district portfolio.

To be consistent, both LR curves given in Fig. 6a must converge as gust speed increases. However, at these gust levels damages are unlikely to follow an idealized square or cubic relationship, especially with cascading effects in case

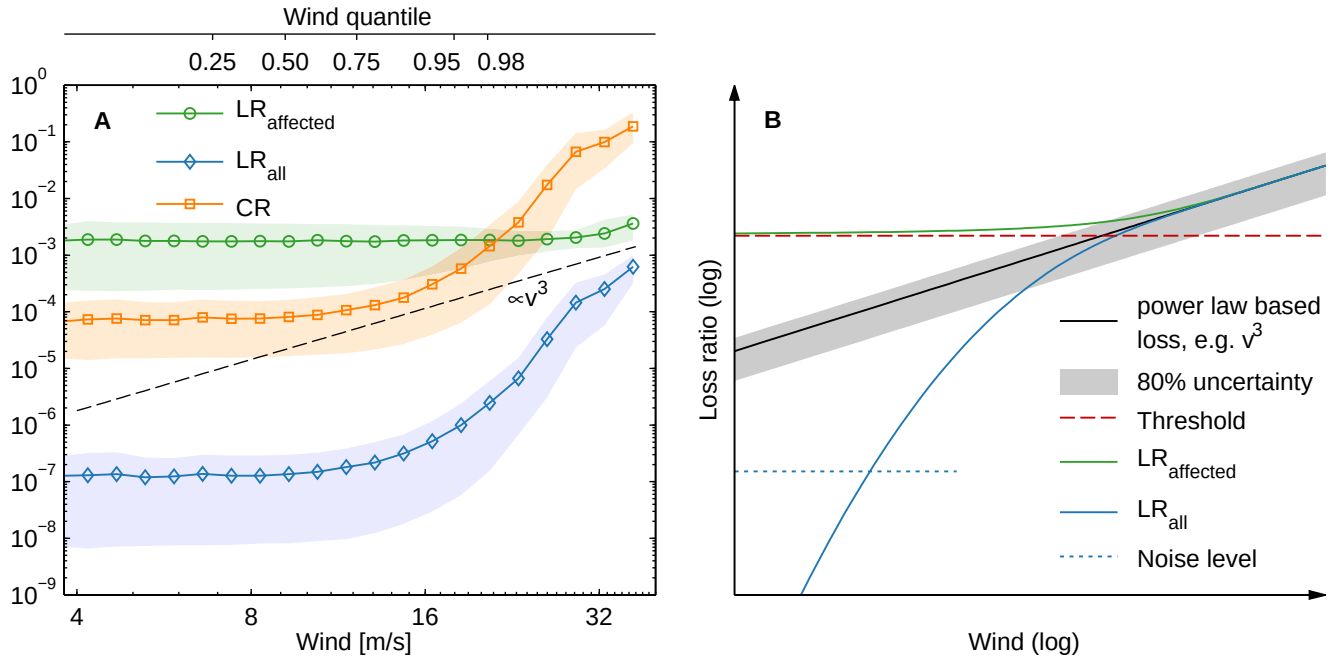

Figure 6. (a) shows the overall DWD gust dependence of the loss and claim ratio for all buildings and the loss ratio only of affected (i.e. damaged) buildings. Shown on a log–log scale, the solid curves represent expected values across all available districts and loss days, while the shaded areas indicate an 80 % uncertainty interval for observations. The dashed line provides a guide to the eye representing a power laws with exponent 3. The upper scale indicates the respective wind gust quantiles. **(b)** shows schematically the decomposition of the loss ratio of a cubic loss–wind relationship subject to a minimum loss threshold. With a lognormal uncertainty distribution, indicated by the shaded 80 % uncertainty bounds, a picture similar to **(a)** arises.

of a breach of the building envelope and additional damage caused by flying debris. Sparks and Bhinderwala (1994) show that at extreme wind speeds a minor fraction of overall loss is comprised by direct wind damage, while the majority of loss results from interior or non-wind damage that are not captured by the physical considerations above.

6 Discussion and concluding remarks

The non-linear processes behind wind and non-wind damage, as well as the effects of cascading failure of structural components, entail that reduced-form approaches as discussed here may only approximate the actual storm damage characteristics. In order to assess the robustness and quality of macroscale storm damage functions, we have analysed and compared the results of four different models applicable to the European winter storm season. As a growing body of climatological research indicates, an increase in future storm intensity (see e.g. the review article by Feser et al., 2015) could lead to the emergence of new hazard profiles. Conditional on the accurate reproduction of local wind characteristics, gust-based damage functions can provide a flexible tool to assess these changes.

Before we discuss the detailed results of the comparison, it is important to acknowledge the effect of deductibles on the shape of damage functions derived from insurance data. Care must be taken as to what extent physical damage concepts, such as a cubic wind–damage relationship, may be applied to insured storm loss. In this regard, all four compared damage functions exhibit a much stronger increase of loss, which is in good agreement with the GDV data employed herein. However, by introducing a simple loss threshold we could demonstrate how such a steep damage function for winter storm loss could be reconciled with a purely cubic wind–damage relationship. If, as climatological research suggests, future storm intensities increase beyond current levels, the overall shape of the damage function plays a crucial role for the extrapolation of future losses. With our threshold hypothesis we intend to expedite the discussion on the validity of damage functions beyond their original data support.

Storm-related insured losses generally exhibit a very broad distribution with a high dynamic range that spans several orders of magnitude. The loss distribution is highly skewed with very few extreme loss events dominating total annual loss. These two aspects pose severe difficulties for both the calibration and the evaluation of damage functions.

With a focus on the level of extreme losses, least-squares curve fitting has often been employed to calibrate damage curves to loss data. The combination of skewed loss distribution and heteroscedastic variance seen for the case of GDV data suggests a violation of the basic assumption for least-

Table 6. Ranking of the four damage functions according to their prediction quality, variability, and applicability.

Criteria	Rank	Model	Description
Extreme loss predictions (loss class I)	1.	*P*	least error, small bias
	1.	*K*	small error, least bias
	3.	*H*	slightly worse error, moderate positive bias
	4.	*X*	strong error and strong positive bias
Moderate to large loss predictions (loss classes III and II)	1.	*H*	good prediction, positive bias for ERA, smallest bias for DWD)
	2.	*P*	good prediction for large loss, positive bias
	3.	*X*	decent prediction for large loss, smallest error and least bias for moderate loss
	4.	*K*	reasonable prediction, strong bias flipping from negative to positive
Variability on district level	1.	*H*	best for DWD, overall good for ERA-Interim
	1.	*P*	very good for both gust data sources
	3.	*K*	better for ERA-Interim; best in north-eastern, worst in southern Germany
	4.	*X*	worst for DWD, large variability for ERA-Interim
Model applicability	1.	*K*	simple calibration, also on extreme losses only
	2.	*P*	requires data for all sizes of loss
	2.	*X*	requires large training data set
	4.	*H*	both number of claims and loss data required

squares fitting and potentially leads to biased results. Due to the high dynamic range even temporally or spatially aggregated loss figures, as used in the cubic excess-over-threshold damage function by Klawa and Ulbrich (2003) (model K), are subject to this effect as they are still dominated by extreme losses.

The optimal curve fitting procedure remains a matter of discussion. Relying on the assumption of general damage relation valid for a large range of losses, the probabilistic power law damage function by Prahl et al. (2012) (model P) puts equal weight on all data points. In contrast, the fitting procedure for the probabilistic claim-based damage function by Heneka and Ruck (2008) (model H) has given greater weight to extremes by using averages of binned losses. The comparison between model H and the simple exponential damage function (model X), both of which are calibrated in the same manner, shows that effective calibration relies on a combination of model constraints and curve fitting.

As was seen in Fig. 5, model H attains greater robustness against jackknife variations of the training sample due to the presumption of a specific claims distribution. Following a different philosophy, model P achieves robustness by rooting the damage function in the entire range of loss.

Transferability is one of the biggest challenges of empirical damage functions. All of the discussed damage functions require substantial calibration to loss data. On the one hand, Heneka and Hofherr (2011) applied their damage function to Germany by employing a static parametrisation originally obtained for the federal state of Baden-Württemberg. Donat et al. (2011a), on the other hand, assume the same vulnerability for nation-wide building stock. In both cases, spatial extension of the model comes at cost of blurring regional vulnerability.

From a practical point of view, model K is most easily calibrated since only a scaling of an otherwise robust raw damage term is required. More elaborate are the calibration procedures for models X and P, which both require detailed loss data. Mathematically, calibration of model H is most demanding and also requires additional data for the number of loss claims.

In order to assess the countrywide performance of the different models, a simple binomial test was devised. In conjunction with the more traditional metrics MAPE and MPE, it was shown that models H and P generally perform best, with some advantage for model P in the large loss class. Most interestingly, the behaviour for extreme losses is indecisive. Model P shows the least variability in terms of MAPE, while model K exhibits the least bias. In terms of the closest model predictions, the binomial test is indecisive between models H and P, whereas both are preferred to models K and X. A summary of the results is given in Table 6.

The applicability of model K appears to be focused on extreme losses. Its further behaviour turns from a positive bias for large losses into a strong negative bias for the moderate loss class. In Sect. A3 we showed that for extreme gust speeds, model K exhibits steepness similar to model P. However the model reaches a lower bound at the 98th wind gust percentile and hence appears to understate losses at speeds in the proximity of this threshold.

Overall, similar behaviour is found for ERA-Interim-based results which are given in the Supplement. A peculiar difference is that for the class of extreme loss days model K performs best in terms of deviation and bias, but fares worse when regarding the losses accumulated for the six largest storms. These contradictory findings can be explained by

the imprecise representation of major storms in ERA-Interim data, especially with regard to the temporal wind profile.

Generally, the obtained results were irrespective of either DWD or ERA-Interim wind gust data. Not surprisingly, ERA-Interim-based results showed greater variance than those based on direct wind gust observations. Interestingly, on district level, the estimated coefficients of variation reveal a marked increase of model variance from the west to the east of Germany.

Further analysis of the coefficient of variation emphasized the importance of the interplay between damage function and the particular wind gust distribution (from either DWD or ERA-Interim). Strong interdependence was seen for model H, performing best with DWD data, and for model K, which showed best results for ERA-Interim data. While model P showed low variability throughout and appeared most flexible to the different data sources, model X showed the greatest error variance overall.

It is worthwhile to note that the coefficient of variation indicates a strong level of *residual* error variance even for the best-performing model. The advantage of DWD over ERA-Interim gust data (cf. Fig. 3) suggests a strong influence of uncertainty in wind gust data. However, there are also a number of potential uncertainty sources connected to the employed insurance data. Uncertainties may arise from gradual damage accumulation masking the effect of individual storms, from incentives for insurance holders (e.g. deductibles), and from wealth levels that affect both building quality and insurance taken. While the employed data does not allow a stratification of losses along socioeconomic dimensions, our regional calibration implicitly accounts for spatial variations due to regionally differing vulnerability and wealth patterns. An altogether different situation would arise for models calibrated on a national scale, where such effects must be considered explicitly.

In our comparison, it would not be meaningful to draw a unique conclusion on the suitability of each model as the performance may crucially depend on the purpose for which it is applied. In the light of this limitation, the exponential modelling approach was found less adequate for the modelling of extremes. In contrast, model K showed its best results for extreme losses, albeit with a calibration procedure that appears less robust than those of the probabilistic models H and P.

Both probabilistic models provided good results over a wide range of loss (moderate to extreme), with their model differences being much smaller than the general variability of losses. On the regional level, they yielded smaller coefficients of variation than the two deterministic models. While models H and P exhibited comparable results, a slight preference could be given to model P in terms of robustness and applicability. With regard to the broadly skewed uncertainty of estimates, probabilistic models can give a better picture of potential loss and should generally be preferred. However, uncertainty estimates for extreme loss remain a concern and should be subject to further research.

Appendix A: Mathematical model description and calibration setup

A1 Generic exponential damage function [X]

The assumption of an exponential damage relationship is not uncommon in the related literature (Huang et al., 2001; Prettenthaler et al., 2012; Murnane and Elsner, 2012) and such models are characterized by a steeper increase than comparable power law models.

Mathematically, the damage function is comprised of a simple exponential term for the loss ratio,

$$L_X(v) = e^{\mathcal{X}_1(v-\mathcal{X}_2)}, \tag{A1}$$

where coefficient $\mathcal{X}_1$ re-scales the wind gust, and offset $\mathcal{X}_2$ adjusts the estimates of the exponential term to the observed loss figures.

Due to the high dynamic range of the loss data and their inherent heteroscedasticity, the damage function cannot be calibrated directly via least-squares. Similarly to the approach for model H, training data were truncated below the 95th wind gust percentile in order to discard the noisy lower end of the loss spectrum that would otherwise distort the damage function. Using gust speed, the remaining loss data were averaged in 10 equally spaced bins with a minimum of five losses each. Thus the relative weight of the few extremes compared to the abundance of small losses was increased. Finally, a logarithmic transformation of the loss averages was employed to reduce the dynamic range of loss and Eq. (A1) was fitted via least squares regression.

A2 Probabilistic power law damage function [P]

Prahl et al. (2012) advocate a probabilistic damage function based on a power law approximation to a more general sigmoid curve. The backbone of the damage function is given by the relationship for the median of the loss magnitude M (i.e. the loss ratio, given at least one loss claim),

$$\widetilde{M}_\mathrm{v} \approx \left(\frac{v}{\mathcal{P}_2}\right)^{\mathcal{P}_1} + \mathcal{P}_3, \tag{A2}$$

where in addition to the power law scaling $\mathcal{P}_2$ and exponent $\mathcal{P}_1$ a constant noise level $\mathcal{P}_3$ is included. Based on the observation that for given wind gust speed v the dispersion of insured losses approximately followed a log-normal distribution, LN(μ, σ), the stochastic loss magnitude is described as a random variable

$$M_\mathrm{v} \sim \mathcal{LN}\left(\ln\left(\widetilde{M}_\mathrm{v}\right), \mathcal{P}_4\right). \tag{A3}$$

The location parameter of the log-normal distribution is related to the median by $\mu = \ln(\widetilde{M}_\mathrm{v})$. The scale parameter $\sigma = \mathcal{P}_4$ describes both the variability due to imprecise gust observation and the aleatory uncertainty regarding the damage caused.

Complementary to the loss magnitude, the probability of loss occurrence (i.e. of receiving one or more loss claims) is given by the relationship

$$p(v) = 1 - \frac{\mathcal{P}_5}{1 + e^{\mathcal{P}_7(v-\mathcal{P}_6)}}. \tag{A4}$$

The turning point $\mathcal{P}_6$ relates to the transition from the noisy regime to the regime of physically driven damages. $\mathcal{P}_7$ determines the sharpness of the transition and $\mathcal{P}_5$ the noise level. Loss occurrence is described stochastically as a random variable

$$O_\mathrm{v} = \begin{cases} 1 & \text{if } P \le p(v) \\ 0 & \text{if } P > p(v) \end{cases}, \tag{A5}$$

where random variable P is drawn from the standard uniform distribution, $P \sim \mathcal{U}(0, 1)$.

In conjunction, loss occurrence and loss magnitude yield the stochastic expression for the loss ratio

$$L_\mathrm{P} = O_\mathrm{v} M_\mathrm{v}, \tag{A6}$$

with an expected value given by

$$\begin{aligned} \mathbb{E}[L_P]_\mathrm{v} &= \mathbb{E}[O_\mathrm{v}]\,\mathbb{E}[M_\mathrm{v}] \\ &= p(v) e^{\mu + \frac{\sigma^2}{2}} \\ &= p(v) e^{\frac{\mathcal{P}_4^2}{2}} \widetilde{M}_\mathrm{v}. \end{aligned} \tag{A7}$$

For high wind gust speeds $v \gg \mathcal{P}_6$, e.g. beyond the 95th percentile, the noise level becomes negligible and the expression for the expected value of loss simplifies to

$$\mathbb{E}[L_P]_{(v \gg \mathcal{P}_6)} \approx e^{\frac{\mathcal{P}_4^2}{2}} \left(\frac{v}{\mathcal{P}_2}\right)^{\mathcal{P}_1}. \tag{A8}$$

Equation (A8) demonstrates that for high wind gust speeds the expected value of the damage function is approximately proportional to the gust speed raised to the power $\mathcal{P}_1$.

Both components of the damage function are calibrated separately. The log-normally distributed loss magnitude is fitted via maximum likelihood to the empirical loss. A least-squares approach is used to fit the loss occurrence term against empirical occurrence rates derived from binned data, enforcing parameter constraints such that the loss occurrence probability is bound within the interval [0, 1].

A3 Cubic excess-over-threshold damage function [K]

Klawa and Ulbrich (2003) developed a simple storm damage function that was subsequently refined for regional application and calibrated to GDV data (Donat et al., 2011a, b). At the heart of the damage function is a cubed power law term as a proxy for storm damage,

$$D(v) = \begin{cases} \left(\frac{v - v_{98}}{v_{98}}\right)^3 & \text{if } v \geq v_{98} \\ 0 & \text{if } v < v_{98} \end{cases}. \quad \text{(A9)}$$

The damage function

$$L_K(v) = \mathcal{K}_1 D(v) + \mathcal{K}_2 \quad \text{(A10)}$$

is calibrated against loss data via linear regression, where constants $\mathcal{K}_1$ and $\mathcal{K}_2$ are the regression coefficients.

Keeping in mind the high dynamic range of loss claims with few dominating extreme losses, the linear regression implicitly puts a strong emphasis on extreme losses ensuring that these are closely matched (cf. Fig. 2).

The shape of the damage function is determined by the power law term, which is influenced only by the 98th wind gust percentile. We chose to determine the 98th percentile from the same training sample as used for calibration of the remaining parameters.

The value of this threshold is of particular interest, as it controls the shape and with it the steepness of the damage function. To clarify this statement, we relate the cubed power law term of the damage function with a tangent based on a simple power law without threshold. For every gust speed v, the tangency condition requires equality of the function values

$$c_1\left(\frac{v}{v_{98}} - 1\right)^3 = \left(\frac{v}{c_2}\right)^\gamma \quad \text{(A11)}$$

and equality of the first derivatives

$$\frac{3c_1}{v_{98}}\left(\frac{v}{v_{98}} - 1\right)^2 = \frac{\gamma}{c_2^\gamma} v^{\gamma-1}. \quad \text{(A12)}$$

Solving Eqs. (A11) and (A12) for the exponent γ yields the simple relationship

$$\gamma = 3\frac{v}{v_{98}}\left(\frac{v}{v_{98}} - 1\right)^{-1} \equiv \frac{3\eta}{\eta - 1}. \quad \text{(A13)}$$

Equation (A13) shows that the local steepness of the cubed excess-over-threshold term depends on the ratio $\eta = v/v_{98}$ of the gust speed to its 98th percentile. For the employed DWD data, the average ratio over all districts of the maximum measured gust speed to the 98th percentile is $\overline{\eta}_{\max} \approx 1.50$, implying that for extreme losses the damage curve increases approximately as a power law with exponent $\gamma \approx 9.0$. Repeating the calculation for ERA-Interim data, we estimated $\overline{\eta}_{\max} \approx 1.41$ and a local power law exponent $\gamma \approx 10.3$.

Hence, the steepness of the model is dependent on the wind gust data source, which may have a potential impact on the portability of the damage function. Additionally, the high local exponents around 10 indicate a similarity with other models that report exponents of a similar magnitude, e.g. Prahl et al. (2012). In physical terms, the two regression coefficients $\mathcal{K}_1$ and $\mathcal{K}_2$ are interpreted, respectively, as a scaling constant and a base loss for losses occurring at wind gusts beneath the threshold. As such, $\mathcal{K}_2$ must be constrained to be strictly non-negative.

For data-scarce applications, it may be opportune to resolve regional portfolio differences via population density as a proxy for (insured) value and obtain a global parametrisation via regression on the national level (e.g. Donat et al., 2011a). In contrast, the finely resolved loss data for our study allowed a local parametrisation and the simple summation of loss to the national level.

Finally, Donat et al. (2011a) perform the regression against annual loss aggregates, while Donat et al. (2011b) demonstrate calibration against a selected sample of the 34 most loss-intensive storm passages. We find that the former calibration method produces better results. However, for reference, results from both calibration methods are given in the Supplement.

A4 Probabilistic claim-based damage function [H]

Heneka et al. (2006) provide a theoretical framework for the modelling of storm loss. Their model was applied first to the federal state of Baden-Württemberg and subsequently to Germany (Heneka and Hofherr, 2011). Maintaining the key assumptions made by Heneka et al. (2006) as far as possible, the intercomparison was based on the following considerations for model design and calibration.

The fundamental concept of model H is the idea that buildings sustain damage only above a critical wind gust threshold v_c. The damage sustained by individual buildings is hence dependent on the specific value of the critical threshold and is formalized by a microscale damage relationship for the fractional damage g,

$$g(v, v_c) = \begin{cases} 0, & v < v_c \\ \left(\frac{v - v_c}{\mathcal{H}_1}\right)^2, & v_c \leq v \leq (v_c + \mathcal{H}_1), \\ 1, & v > (v_c + \mathcal{H}_1) \end{cases} \quad \text{(A14)}$$

reaching complete destruction at a wind gust increase of $\mathcal{H}_1$ above the critical threshold.

For a portfolio of buildings, each with individual critical threshold, a specific density distribution for v_c may be assumed or otherwise estimated. For simplicity, Heneka et al. (2006) idealized the density distribution of v_c by the density of the normal distribution $f(v_c, \mu_c, \mathcal{H}_2)$, with mean μ_c and standard deviation $\mathcal{H}_2$. It follows that the claim ratio $C_H(v)$, i.e. the relative share of affected buildings, is given by the integral

$$C_H(v) = \int_{-\infty}^{v} f(v_c, \mu_c, \mathcal{H}_2)\, \mathrm{d}v_c. \quad \text{(A15)}$$

Table A1. Comparison of the parameter values obtained for the federal state of Baden-Württemberg with those published by Heneka and Ruck (2008). Accordingly, relative wind gust speed was normalised by its 98th percentile. For easier comparison, the values in brackets are rescaled to match the published value of $\mathcal{H}_4$.

	Source	$\mathcal{H}_4$	$\mathcal{H}_3$	$\mathcal{H}_2$	$\mathcal{H}_1$
Absolute gust	Heneka and Ruck	50.5	2.5	7.8	70.0
	DWD	42.3	2.0	6.2	49.7
		(50.5)	(2.4)	(7.4)	(59.4)
	ERA-Interim	41.6	1.8	5.6	45.5
		(50.5)	(2.2)	(6.8)	(55.3)
Relative gust	Heneka and Ruck	1.31	0.04	0.20	1.85
	DWD	2.28	0.09	0.32	2.67
		(1.31)	(0.05)	(0.19)	(1.54)
	ERA-Interim	2.17	0.10	0.29	2.43
		(1.31)	(0.06)	(0.18)	(1.47)

The loss ratio $L_H(v)$ is then obtained by solving the convolution integral

$$L_H(v) = \int_{-\infty}^{v} g\,(v, v_c)\, f\,(v_c, \mu_c, \mathcal{H}_2)\, \mathrm{d}v_c, \tag{A16}$$

combining the density distribution of v_c with the microscale damage function $g(v, v_c)$.

Finally, uncertainty is introduced by assuming a Gaussian distribution $f(\mu_c, \mathcal{H}_4, \mathcal{H}_3)$ for the mean critical wind gust speed μ_c, with mean $\mathcal{H}_4$ and standard deviation $\mathcal{H}_3$. Putting all components together, we obtain an expression for the expected value of the loss ratio

$$\mathbb{E}[L_H] = \int_0^1 L f\,(\mu(L), \mathcal{H}_4, \mathcal{H}_3)\, \frac{\mathrm{d}\mu(L)}{\mathrm{d}L}\mathrm{d}L, \tag{A17}$$

where we define $\mu_c = \mu(L)$ as the inverse function of Eq. (A16) with respect to μ_c.

For calibration, Heneka et al. (2006) used least-squares fitting of claims and loss data that was pooled for the entire state of Baden-Württemberg. However, fitting the damage function to individual districts, it was found that least-squares curve fitting was yielding poor results due to frequent overfitting to the few number of "outlying" extreme events and the generally high dynamic range of the data. Furthermore, the model was developed on strong winds and could not deal with the noise present in the GDV data at low wind gust speeds.

For the work in hand, these problems were solved via a three-step fitting procedure. In order to exclude the effect of noise, data below the 95th wind gust percentile were discarded during the fitting procedure.

In the first step, Eq. (A15) was fitted to claims data. To overcome the problem of the high dynamic range, claims data were logarithmically transformed. To counteract the downside of the transformation, namely the increased weight of the abundant small damages as compared to the few extremes, the data were binned into 10 equally spaced bins, each containing a minimum of five data points. Using the method of least squares the curve was fitted to the mean values of each bin. In this step, we made the implicit assumption of a multiplicative error term, relating to a symmetric distribution around the mean of the log-transformed claims data (i.e. the geometric mean of the absolute numbers). This assumption is backed by actuarial practice for describing insurance damage claim distributions by log-symmetric distribution such as the log-normal distribution (Lawrence, 1988).

In the second step, the above described fitting procedure is used to calibrate Eq. (A16) to the loss ratio data.

Thirdly, the parameters of the normal distribution describing the random fluctuation of μ_c are determined via log-likelihood optimization based on loss data at full detail.

Due to the strong deviation from the original least-squares fitting employed by Heneka et al. (2006), it was necessary to validate the parameters obtained from the GDV data set. For this purpose, we pooled the GDV data for all districts in the state of Baden-Württemberg and compared the obtained model parameters against those values published by Heneka and Ruck (2008). The results presented in Table A1 show good agreement of the individual parameters across the different sources. As the wind gust data sources are not directly comparable, the parameters shown in brackets were rescaled according to $\mathcal{H}_4$. Regarding these values, only $\mathcal{H}_2$, which represents the wind gust range from beginning to total destruction, shows a significant difference of approximately $-15\,\%$ as compared to the original values.

While we report only those results that relate to the best performing model setup, results from applying the Baden-Württemberg calibration to entire Germany (similarly to Heneka and Hofherr, 2011) are included in the Supplement for special interest.

The Supplement related to this article is available online at doi:10.5194/nhess-15-769-2015-supplement.

Acknowledgements. We appreciate valuable discussions with U. Ulbrich and M. Boettle. We thank the German Insurance Association (GDV), the German Weather Service (DWD), and the European Centre for Medium-Range Weather Forecasts (ECMWF) for providing the data. This work was supported by the European Community's Seventh Framework Programme under Grant Agreement No. 308497 (Project RAMSES).

Edited by: I. Didenkulova
Reviewed by: three anonymous referees

References

Bernaola-Galván, P., Oliver, J., Hackenberg, M., Coronado, A., Ivanov, P., and Carpena, P.: Segmentation of time series with long-range fractal correlations, Eur. Phys. J. B, 85, 1–12, doi:10.1140/epjb/e2012-20969-5, 2012.

Bouwer, L. M. and Wouter Botzen, W. J.: How sensitive are US hurricane damages to climate? Comment on a paper by W. D. Nordhaus, Clim. Change Econ., 02, 1–7, doi:10.1142/S2010007811000188, 2011.

Box, G. E. and Cox, D. R.: An analysis of transformations, J. Roy. Stat. Soc. B, 26, 211–252, 1964.

Cusack, S.: A 101 year record of windstorms in the Netherlands, Climatic Change, 116, 693–704, doi:10.1007/s10584-012-0527-0, 2013.

Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C. M., van de Berg, L., Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J., Haimberger, L., Healy, S. B., Hersbach, H., Hólm, E. V., Isaksen, L., Kållberg, P., Köhler, M., Matricardi, M., McNally, A. P., Monge-Sanz, B. M., Morcrette, J.-J., Park, B.-K., Peubey, C., de Rosnay, P., Tavolato, C., Thépaut, J.-N., and Vitart, F.: The ERA-Interim reanalysis: configuration and performance of the data assimilation system, Q. J. Roy. Meteorol. Soc., 137, 553–597, doi:10.1002/qj.828, 2011.

Deroche, M.-S., Choux, M., Codron, F., and Yiou, P.: Three variables are better than one: detection of European winter windstorms causing important damages, Nat. Hazards Earth Syst. Sci., 14, 981–993, doi:10.5194/nhess-14-981-2014, 2014.

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, doi: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, doi:10.5194/nhess-11-2821-2011, 2011b.

Dorland, C., Tol, R. S. J., and Palutikof, J. P.: Vulnerability of the Netherlands and Northwest Europe to storm damage under climate change, Climatic Change, 43, 513–535, doi:10.1023/A:1005492126814, 1999.

Emanuel, K.: Increasing destructiveness of tropical cyclones over the past 30 years, Nature, 436, 686–688, doi:10.1038/nature03906, 2005.

Etienne, C. and Beniston, M.: Wind storm loss estimations in the Canton of Vaud (Western Switzerland), Nat. Hazards Earth Syst. Sci., 12, 3789–3798, doi:10.5194/nhess-12-3789-2012, 2012.

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

GDV: Naturgefahrenreport 2013, Gesamtverband der Deutschen Versicherungswirtschaft e.V., http://www.gdv.de/2013/10/naturgefahrenreport-2013-die-schaden-chronik-der-versicherer/ (last access: 25 August 2014), 2013.

Gerstengarbe, F.-W., Werner, P. C., Österle, H., and Burghoff, O.: Winter storm- and summer thunderstorm-related loss events with regard to climate change in Germany, Theor. Appl. Climatol., 1–10, doi:10.1007/s00704-013-0843-y, 2013.

Held, H., Gerstengarbe, F.-W., Pardowitz, T., Pinto, J. G., Ulbrich, U., Born, K., Donat, M. G., Karremann, M., Leckebusch, G. C., Ludwig, P., Nissen, K. M., Österle, H., Prahl, B. F., Werner, P. C., Befort, D. J., and Burghoff, O.: Projections of global warming-induced impacts on winter storm losses in the German private household sector, Climatic Change, 121, 195–207, doi:10.1007/s10584-013-0872-7, 2013.

Heneka, P. and Hofherr, T.: Probabilistic winter storm risk assessment for residential buildings in Germany, Nat. Hazards, 56, 815–831, doi:10.1007/s11069-010-9593-7, 2011.

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

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.

Huang, Z., Rosowsky, D. V., and Sparks, P. R.: Long-term hurricane risk assessment and expected damage to residential structures, Reliab. Eng. Syst. Saf., 74, 239–249, doi:10.1016/S0951-8320(01)00086-2, 2001.

Huttenlau, M. and Stötter, J.: The structural vulnerability in the framework of natural hazard risk analyses and the exemplary application for storm loss modelling in Tyrol (Austria), Nat. Hazards, 58, 705–729, doi:10.1007/s11069-011-9768-x, 2011.

Hyndman, R. J. and Koehler, A. B.: Another look at measures of forecast accuracy, Int. J. Forecasting, 22, 679–688, doi:10.1016/j.ijforecast.2006.03.001, 2006.

Kantha, L.: Tropical cyclone destructive potential by integrated kinetic energy, B. Am. Meteorol. Soc., 89, 219–221, doi:10.1175/BAMS-89-2-219, 2008.

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, doi:10.5194/nhess-3-725-2003, 2003.

Kohavi, R.: A study of cross-validation and bootstrap for accuracy estimation and model selection, in: Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1137–1145, 1995.

Lawrence, R. J.: Applications in economics and business, in: Lognormal Distributions: Theory and Applications, Marcel Dekker, New York, 229–266, 1988.

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, doi:10.1029/2006GL027663, 2007.

Merz, B., Kreibich, H., Schwarze, R., and Thieken, A.: Review article "Assessment of economic flood damage", Nat. Hazards Earth Syst. Sci., 10, 1697–1724, doi:10.5194/nhess-10-1697-2010, 2010.

Mestre, O., Gruber, C., Prieur, C., Caussinus, H., and Jourdain, S.: SPLIDHOM: a method for homogenization of daily temperature observations, J. Appl. Meteorol. Clim., 50, 2343–2358, doi:10.1175/2011JAMC2641.1, 2011.

Miller, R. G.: The jackknife – a review, Biometrika, 61, 1–15, doi:10.1093/biomet/61.1.1, 1974.

Munich Re: Winterstürme in Europa (II) – Schadenanalyse 1999 – Schadenspotentiale, Münchener Rückversicherungs-Gesellschaft, Munich, Germany, 1993.

Munich Re: Winterstürme in Europa – Schadenanalyse 1990 – Schadenspotentiale, Münchener Rückversicherungs-Gesellschaft, Munich, Germany, 2001.

Munich Re: Topics Geo – Natural Catastrophes 2012 – Analyses, Assessments, Positions, Münchener Rückversicherungs-Gesellschaft, Munich, Germany, 2013.

Murnane, R. J. and Elsner, J. B.: Maximum wind speeds and US hurricane losses, Geophys. Res. Lett., 39, L16707, doi:10.1029/2012GL052740, 2012.

Nordhaus, W. D.: The economics of hurricanes and implications of global warming, Clim. Change Econ., 01, 1–20, doi:10.1142/S2010007810000054, 2010.

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, doi:10.5194/nhess-7-165-2007, 2007.

Pita, G. L., Pinelli, J.-P., Gurley, K. R., and Hamid, S.: Hurricane vulnerability modeling: development and future trends, J. Wind Eng. Ind. Aerodyn., 114, 96–105, doi:10.1016/j.jweia.2012.12.004, 2013.

Powell, M. D. and Reinhold, T. A.: Tropical cyclone destructive potential by integrated kinetic energy, B. Am. Meteorol. Soc., 88, 513–526, doi:10.1175/BAMS-88-4-513, 2007.

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.

Prettenthaler, F., Albrecher, H., Koberl, J., and Kortschak, D.: Risk and insurability of storm damages to residential buildings in Austria, Geneva Pap. R. I. – Iss. P., 37, 340–364, doi:10.1057/gpp.2012.15, 2012.

Roberts, J. F., Champion, A. J., Dawkins, L. C., Hodges, K. I., Shaffrey, L. C., Stephenson, D. B., Stringer, M. A., Thornton, H. E., and Youngman, B. D.: The XWS open access catalogue of extreme European windstorms from 1979 to 2012, Nat. Hazards Earth Syst. Sci., 14, 2487–2501, doi:10.5194/nhess-14-2487-2014, 2014.

Rybski, D. and Neumann, J.: A review on the pettitt test, in: In Extremis, edited by: Kropp, J. and Schellnhuber, H.-J., Springer, Berlin, Heidelberg, 202–213, doi:10.1007/978-3-642-14863-7_10, 2011.

Schwierz, C., Köllner-Heck, P., Zenklusen Mutter, E., Bresch, D., Vidale, P.-L., Wild, M., and Schär, C.: Modelling European winter wind storm losses in current and future climate, Climatic Change, 101, 485–514, doi:10.1007/s10584-009-9712-1, 2010.

Simiu, E. and Scanlan, R. H.: Wind Effects on Structures: Fundamentals and Applications to Design, 3rd Edn., John Wiley, New York, 1996.

Sparks, P. R. and Bhinderwala, S. A.: Relationship between residential insurance losses and wind conditions in Hurricane Andrew, in: Hurricanes of 1992: Lessons Learned and Implications for the Future, ASCE, New York, USA, 111–124, 1994.

Swiss Re: sigma – Natural catastrophes and man-made disasters in 2013, Swiss Re Ltd, Zurich, Switzerland, 2014.

Venema, V. K. C., Mestre, O., Aguilar, E., Auer, I., Guijarro, J. A., Domonkos, P., Vertacnik, G., Szentimrey, T., Stepanek, P., Zahradnicek, P., Viarre, J., Müller-Westermeier, G., Lakatos, M., Williams, C. N., Menne, M. J., Lindau, R., Rasol, D., Rustemeier, E., Kolokythas, K., Marinova, T., Andresen, L., Acquaotta, F., Fratianni, S., Cheval, S., Klancar, M., Brunetti, M., Gruber, C., Prohom Duran, M., Likso, T., Esteban, P., and Brandsma, T.: Benchmarking homogenization algorithms for monthly data, Clim. Past, 8, 89–115, doi:10.5194/cp-8-89-2012, 2012.

Walker, G. R.: Modelling the vulnerability of buildings to wind – a review, Can. J. Civ. Eng., 38, 1031–1039, 2011.

Wan, H., Wang, X. L., and Swail, V. R.: Homogenization and trend analysis of Canadian near-surface wind speeds, J. Climate, 23, 1209–1225, doi:10.1175/2009JCLI3200.1, 2010.

Wang, X. L.: Accounting for autocorrelation in detecting mean shifts in climate data series using the penalized maximal *t* or *F* test, J. Appl. Meteorol. Clim., 47, 2423–2444, doi:10.1175/2008JAMC1741.1, 2008.

Wang, X. L., Chen, H., Wu, Y., Feng, Y., and Pu, Q.: New techniques for the detection and adjustment of shifts in daily precipitation data series, J. Appl. Meteorol. Clim., 49, 2416–2436, doi:10.1175/2010JAMC2376.1, 2010.