**Natural Hazards
and Earth
System Sciences**

European extra-tropical storm damage risk from a multi-model ensemble of dynamically-downscaled global climate models

M. R. Haylock

PartnerRe, Bellerivestrasse 36, 8034 Zurich, Switzerland

Abstract. Uncertainty in the return levels of insured loss from European wind storms was quantified using storms derived from twenty-two 25 km regional climate model runs driven by either the ERA40 reanalyses or one of four coupled atmosphere-ocean global climate models. Storms were identified using a model-dependent storm severity index based on daily maximum 10 m wind speed. The wind speed from each model was calibrated to a set of 7 km historical storm wind fields using the 70 storms with the highest severity index in the period 1961–2000, employing a two stage calibration methodology. First, the 25 km daily maximum wind speed was downscaled to the 7 km historical model grid using the 7 km surface roughness length and orography, also adopting an empirical gust parameterisation. Secondly, downscaled wind gusts were statistically scaled to the historical storms to match the geographically-dependent cumulative distribution function of wind gust speed.

The calibrated wind fields were run through an operational catastrophe reinsurance risk model to determine the return level of loss to a European population density-derived property portfolio. The risk model produced a 50-yr return level of loss of between 0.025 % and 0.056 % of the total insured value of the portfolio.

1 Introduction

European winter storms have the potential to cause large damage to property and the environment as well as loss of life. From 1970 to 2008, of the 40 global natural and man-made catastrophes with the highest insurance losses (hereafter referred to as "loss"), seven were European winter storms causing 11.2 % of the total losses (Enz et al., 2009).

For example, storm Lothar in 1999 passed close to the centre of Paris and, with 110 European fatalities, caused the highest number of deaths (Enz et al., 2009). In 2005, storm Erwin passed through northern Denmark before felling an estimated 75 million m^3 of trees in southern Swedish forests (Guy Carpenter, 2005). This figure was close to the average annual production of total Swedish forests and much of which was not adequately priced by reinsurers at the time due to a scarcity of similar events in recent history.

While long storm catalogues exist spanning recent centuries (e.g. Lamb and Frydendahl, 1991), such documents are generally an incomplete selection of earlier storms based on limited available documentation with possibly no instrumental observation. Also, their generally descriptive nature prohibits detailed quantitative analysis. Therefore it is only since the era of global reanalyses, with their uninterrupted six-hourly data and modern measurement techniques that we are able to gain a more complete picture of storm activity over Europe. With generally only a few severe damaging storms a year, quantitative risk assessment using just the 50 yr period of global reanalyses is an uncertain process, especially since government regulation increasingly requires insurance companies to protect their balance sheets up to losses with an expected return period of 200 yr or above.

Reinsurance companies, usually the holders of financial risk from catastrophes, most often quantify storm risk to an insurance portfolio (hereafter referred to as "portfolio") using catastrophe models. Catastrophe models determine loss to a portfolio from a catalogue of storms by applying a transfer function to convert the wind speed at each location to loss. The transfer functions (often referred to as "vulnerability functions"), are determined by a combination of analysis of past insurance claims and engineering considerations. To overcome the limitations of a short historical record, catastrophe models often extend the storm catalogue using artificially-generated storms. Traditionally, the most common approach to generating artificial storms has been to

statistically perturb the wind field intensity, shape or location of historical storms. Peer-reviewed literature exists for tropical storm track perturbation (Hall and Jewson, 2008), but similar methods have also been used for European windstorm modelling by commercial catastrophe model providers, for example previous generations of the European wind model by Risk Management Solutions (www.rms.com) and current versions of the model by EQECAT (www.eqecat.com). While methodologically relatively straightforward and computationally undemanding, there is a large degree of subjectivity as to the limits of the perturbations and how to assign a return period to such events.

There is a growing interest in the use of dynamical climate models for generating physically-realistic storms differing from what has been observed. The detailed wind fields demanded by catastrophe models require that high resolution dynamical models are employed. Until recently, it has proven computationally unfeasible to generate the required millennium-scale datasets of regional climate models (RCMs) nested in global climate models (GCMs). Two published studies have performed experiments with shorter RCM runs coupled to loss models. Schwierz et al. (2010) coupled a Europe loss model with 30-yr time slices of wind data from two RCMs nested in two different GCMs to assess possible changes to loss under an A2 emission scenario (Nakicenovic et al., 1996). Under the constraint of working with just several decades of data, they found a region-dependent increase of up to 44 % in the annual expected loss and up to 104 % in the expected 100-yr loss, with the largest losses in the central European latitudes. The loss was based on a Europe-wide "market portfolio" – a portfolio representative of the entire European insurance market. Heneka et al. (2006) resolved wind fields at 1 km resolution using the Karlsruher Atmospheric Mesoscale Model for 30 severe storms between 1971 and 2000 to assess damage risk to the German state of Baden-Württemberg. They created loss exceedance probability curves at the community level with an annual expected loss to the region of 15 million Euro.

Three recent studies have investigated the use of coarser resolution GCMs for estimating loss from European winter storms. Of the three studies, two have used climate change GCM experiments coupled with loss models. Leckebusch et al. (2007) regressed loss data with maximum 10 m wind speed using four GCMs for present and future climate conditions. They found an expected increase in loss of up to 37 % for the late 21st Century compared with present day. Pinto et al. (2007) applied a similar technique to estimate storm losses from a 3-member ensemble of a single GCM forced by two emission scenarios. The loss model was calibrated using wind data from the ERA40 analyses with German loss data. They found increases in the mean storm losses in northern and western central Europe, the highest being a 49 % increase in Germany, and decreases in the south. Della-Marta et al. (2010) took a different approach by examining whether an ensemble of seasonal forecast model runs from the European Centre for Medium-Range Weather Forecasts (ECMWF) could be used to minimise the statistical uncertainty of estimates of higher return period storm losses. They found that the forecast model runs, which could be considered an ergodic sample when the first month of the ensemble forecasts was discarded, provided a greatly extended storm set compared with history. After calibrating the wind fields using statistical scaling with those of the SwissRe historical storm catalogue and coupling the wind fields with a simplified SwissRe loss model, they were able to reproduce the loss-return period curve (hereafter referred to as the "loss curve") of the historical storms but with greatly reduced statistical uncertainty.

Other studies have used simplified indices of loss in place of loss models. Notably, a recent study by Donat et al. (2010) used a loss index, based on the cubed exceedance of maximum wind speed above a quantile threshold, to explore the benefits of dynamical downscaling in storm loss estimation. Comparing the loss index from ERA40-driven RCMs (a subset of the present study) against historical loss data from the German market, they concluded that boundary-forced dynamical downscaling provided superior loss estimates to using the ERA40 data. Importantly, they also concluded that the performance of the ensemble mean was as good as the performance of the best RCM.

The present analysis extends on the methodologies of the above studies in order to generate losses using wind fields from a large multi-model set of storms coupled to a commercially-operational loss model. The aim is to quantify uncertainty in the return levels of insured loss by using a multi-model storm set derived from high resolution dynamical models. Wind fields were derived from 22 model runs comprising 13 different 25 km RCMs forced at the boundary by either ERA40 or one of four coupled ocean-atmosphere GCMs. RCM wind speeds for each model were first downscaled to 7 km using 7 km surface roughness length and orography data then statistically scaled to a set of 7 km modelled historical storms (Sect. 3). The resulting 22 calibrated storm catalogues were coupled to a commercially operational loss model for estimating the return levels of loss to a Europe-wide population density-derived portfolio (Sect. 4). When averaged across the 22 climate models, the return levels of loss show close agreement with the historical storm set. The spread between the models provides an estimate of the sampling uncertainty arising from the limited size of the historical storm catalogue.

2   Data

Throughout this paper, we refer to the 72-h storm "footprint". This is the maximum wind speed at each location over a 72-h period. Seventy-two hours was chosen as reinsurance contracts commonly define a single event as damages occurring within a 72 h period and it is enough time for a storm

to completely pass over the western European region. This creates the possibility that a single 72-h footprint may have resulted from more than one extra-tropical cyclone crossing the domain. Choosing a shorter period raises the risk that a single extratropical cyclone could generate more than one storm event, thereby violating the independence requirement of the extreme value analyses undertaken in this study (Palutikof et al., 1999).

2.1 The historical storm set

The original European windstorm catastrophe model of PartnerRe, made operational in 2006, used a set of 100 historical storms modelled with a high-resolution numerical weather prediction (NWP) model, which assimilated available observations over a 72-h period for each storm. The model was the Swiss implementation of the Lokal Modell of the Consortium for Small Scale Modelling (COSMO), a non-hydrostatic model with 45 vertical layers run on a rotated latitude-longitude grid at a horizontal resolution of 1/16 degree (≈7 km) Rockel et al. (2008). The initial set of 100 storms spanned the period 1957–2002, the period corresponding to the ERA40 reanalyses (Uppala et al., 2005) which provided the initial and lateral boundary data. The storms were chosen subjectively through knowledge of large historical insurance losses and qualitative examination of weather maps. An additional nine storms were added that occurred between 2002 and 2009. Storm footprints were calculated from 10 m maximum wind gust, which uses a turbulent gust parameterisation based on observations and derived from just the lowest model level mean wind speed. Convective gusts are parameterised using the method of Schulz and Heise (2003) a slight adaption of the method proposed by Nakamura et al. (1996).

In 2007 the historical storm set was extended to overcome the subjectivity of storm selection in the initial 100 storms. Objective selection of the spatially largest and most intense storms was undertaken using six-hourly ERA40 data and these storms were dynamically downscaled to 7 km using the COSMO model. To find the largest storms, overlapping 72-h storm footprints were calculated starting every six hours. From these a storm severity index (SSI) was derived based on the latitude-weighted cubed exceedance above a threshold of the 72-h storm footprint over land (Eq. 1).

$$\text{SSI} = \sum_i \cos\theta_i \left[\max(0, \text{wind}_i - \text{threshold})\right]^3 \tag{1}$$

At each grid point i with latitude θ_i wind$_i$ is the maximum 10 m wind speed over the 72 h taken from six-hourly instantaneous values. While using instantaneous values created the possibility of aliasing of the maximum wind field, whereby a storm front moved more than one grid box in six hours, this was preferable to using the six-hourly maximum 10 m gust, which is problematic in areas of high surface roughness (Della-Marta et al., 2009).

A threshold of $11\,\text{m s}^{-1}$ was used, this being the mean 90th percentile of the six-hourly wind footprints, or the average of the value that was exceeded by ten percent of the grid points for each storm footprint. Many similar SSIs have been used previously in the literature (Della-Marta et al., 2009; Donat et al., 2010; Leckebusch et al., 2008) as well as in defining financial contracts based on parameters of observed or modelled wind speed (usually referred to as "Catastrophe Bonds"). Our SSI is similar to that of Leckebusch et al. (2008) who used cubed wind speeds normalised by a local quantile. The underlying assumption of their approach is that a particular wind speed would be expected to cause more damage in areas with lower mean wind. We have not found evidence of this variation within a country in our own analysis of insurance loss data, but our damage transfer functions (Sect. 4) do vary between countries. Nevertheless, we did not test other indices, but since we are using the SSI only to find the largest storms, variations in the index are not so critical.

The largest storms were identified using the non-overlapping 72-h periods with the highest SSI and without any overlap with the original 100 PartnerRe historical storms. Budget and time constraints restricted the dynamical downscaling to an additional 22 storms. The resulting set of 131 storms comprises the PartnerRe historical storm catalogue, which includes objective selection of the 70 largest storms in the ERA40 period based on their SSI. This historical storm set is used in the present study to calibrate storms from other models.

2.2 The RCM data set

Regional climate models nested in ERA40 reanalyses and coupled GCMs form the basis of this study (Table 1). Daily model data were provided by the European Union-funded ENSEMBLES project (Christensen et al., 2008; http://ensembles-eu.metoffice.com). Transient RCM-GCM model runs were available from 1950 up to the middle or end of the 21st Century, depending on the model (Christensen et al., 2009); however, data were used up to the year 2008 to compare with the historical storm set. All GCMs use SRES A1B scenario emissions after the year 2000, except for the ECHAM5-forced C4I RCM which uses the A2 scenario. Although the 10 m maximum wind gust from the historical storm set is used for calibration of RCM output, the daily maximum 10 m mean wind speed from the RCMs was used, since maximum wind gust was available for only half of the models. Instead, a consistent offline gust parameterisation was applied to all models (Sect. 3.2). An additional model run by the University of Castilla-La Mancha (UCLM) in Spain was not used after it was discovered that the data saved as the maximum daily wind speed were in fact instantaneous values (E. Sanchez, personal communication, 2009).

The output of all models, except the data provided by Recherche en Prévision Numérique (RPN), was provided on the native model grid, which varied by model. RPN provided

Table 1. Period of regional climate model runs with their driving global model.

	Institution	RCM	GCM				
			ERA40	HadCM3	ECHAM5	ARPEGE	BCM
C4I	Community Climate Change Consortium for Ireland	RCA3	1961–2000	1951–2008	1951–2008		
CHMI	The Czech Hydrometeorological Institute	ALADIN	1961–2000				
CNRM	Centre national de la recherche scientifique	RM4.5	1958–2001			1950–2008	
DMI	Danish Meteorological Institute	HIRHAM5	1958–2002			1951–2008	
ETHZ	Swiss Federal Institute of Technology, Zurich	CLM	1961–2000	1951–2008			
KNMI	Royal Netherlands Meteorological Institute	RACMO2	1958–2002		1950–2008		
METNO	Norwegian Meteorological Institute	HIRHAM	1961–2000				1951–2008
METO	UK MetOffice	HadRM3Q3	1959–2002				
METO	UK MetOffice	HadRM3Q16	1959–2002				
METO	UK MetOffice	HadRM3Q0	1959–2002	1951–2008			
MPI	Max Planck Institut für Meteorologie	REMO	1961–2000		1951–2008		
RPN	Recherche en Prévision Numérique	GEMLAM	1961–2000				
SMHI	Swedish Meteorological and Hydrological Institute	RCA	1961–2000				

their data interpolated to a 0.25 degree Cartesian latitude-longitude grid. The models run by Centre National de la Recherche Scientifique (CNRM) and the Czech Hydrometeorological Institute (CHMI) employed a 25 km Lambert Conformal grid and all other models used a 0.22 degree rotated-pole grid with the North Pole at 39.25° N, 162° W.

2.3 The insurance portfolio

The input data for catastrophe risk models are typically a table of insured values (hereafter referred to as "values") for each of many locations. Locations usually coincide with regional administrative boundaries (e.g. postcodes), but geographic resolution varies by country. In this study a sample Europe-wide portfolio was created based on population density for the following countries: Germany, United Kingdom, France, Netherlands, Norway, Sweden, Ireland, Belgium, Denmark, Luxembourg, Switzerland, and Austria. Gridded population density on a 2.5 arc-minute grid (GPWv3, 2005) was mapped onto the locations supported by the PartnerRe European wind model using the nearest neighbour to create the geographic distribution of value (Fig. 1). The units are $1000 \times$ no. of people km^{-2}, but since all losses are reported here as percent of the total portfolio value, the absolute magnitude of the values is irrelevant. The portfolio contains no insurance deductible (the first part of the loss paid by the policy holder), so all losses reported are the gross or "ground-up" loss.

3 Model calibration

Climate models vary in their physical representation of relevant processes. For the thirteen ERA40-forced RCMs, both the median and the 90th percentile of maximum daily wind speed over land varied by almost a factor of two amongst the RCMs for the period 1961–2000. Therefore calibration of
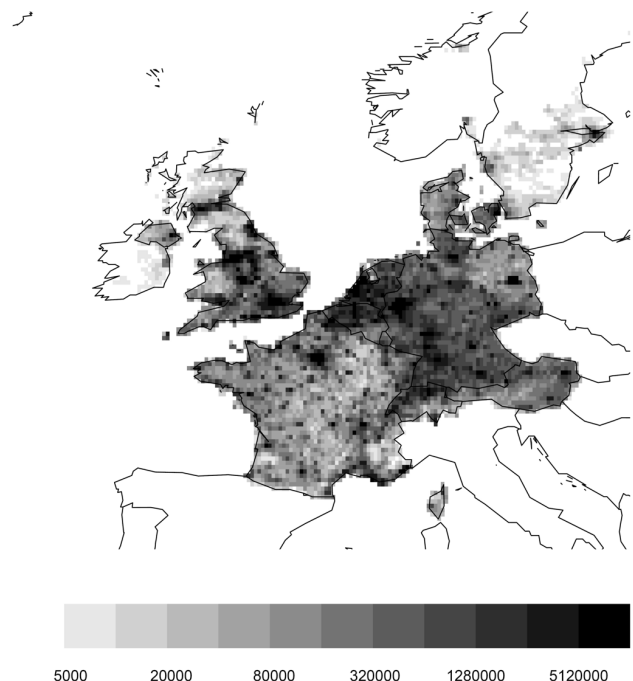

Fig. 1. The geographic distribution of insured value in the population density-derived portfolio. Units are people $\mathrm{km}^{-2} \times 1000$.

the RCM data was required. A two-stage calibration process was employed: roughness length and orography-dependent downscaling of the RCM footprints to the same grid as the historical storm set (Sect. 3.2); then statistical scaling of the footprints (Sect. 3.3). Model calibration resulted in the RCM wind speeds being statistically comparable to the historical set, from which our damage transfer functions were derived. Note that we are not interested in calibrating the models against observations, since our damage transfer functions were derived from wind fields of the historical storm

set and not from observations. If the historical storm set contains a wind speed bias, such as under representing the most extreme wind speeds, then this bias needs to be carried through to the RCMs. A fundamental assumption of our study is that climate models with possibly differing storm climatologies can be calibrated. Such an approach follows from Della Marta et al. (2010) (Sect. 1) with the additional step of using the objectively-selected largest 70 storms in the candidate (RCM) and target (historical) distributions of the calibration.

3.1 Identifying the largest storms

Statistical scaling relationships (Sect. 3.3) were defined using the base period 1961–2000 for which all models and historical storms were available. Since the historical storm set captured the 70 storms with the highest SSI (Sect. 2.1 – hereafter referred to as "largest storms"), our calibration was performed by matching the wind speed distribution (Sect. 3.3) of the largest 70 storms in each model over the base period with those from the historical set.

The largest 70 storms for each RCM and the historical storm set in the base period were first identified. To do this, a SSI was calculated for the 131 historical storms and for each RCM for each running four-day period in the base period. Since only daily RCM data were available, we used four day maximum wind speed for the RCM storm footprints in order to capture 72-h storms spanning four calendar days. The storm identification method was identical to as discussed earlier for ERA40 (Eq. 1), but a model-dependent threshold was used. For the historical set, the threshold used was the mean 90th percentile of the 131 storm footprints. For the RCMs the 131 highest non-overlapping (separated by at least four days) values of the 90th percentile of the storm footprint were averaged, maintaining consistency with our historical set methodology. Using this model-dependent threshold, the SSI for each RCM and four-day period was calculated to identify the largest 70 storms in the base period.

3.2 Downscaling

For the largest 70 storms in each RCM over the base period, the 25 km storm footprints were downscaled to the 7 km rotated-pole grid of the historical storms. A schematic of this downscaling process is presented in Fig. 2. Surface roughness length data (z_0) were not available for many of the RCMs. Therefore the 7 km effective surface roughness length from the COSMO model was used, which incorporates both land use and orographic form drag (Wood and Mason, 1993). The 7 km roughness length data were first averaged into grid boxes representative of the 25 km grid for each model by assigning each 7 km grid point to its closest 25 km grid point. With this upscaled roughness length the wind speed at a blending height of 100 m (U_b) was calculated from the wind speed (U_{ref}) at the 10 m reference height
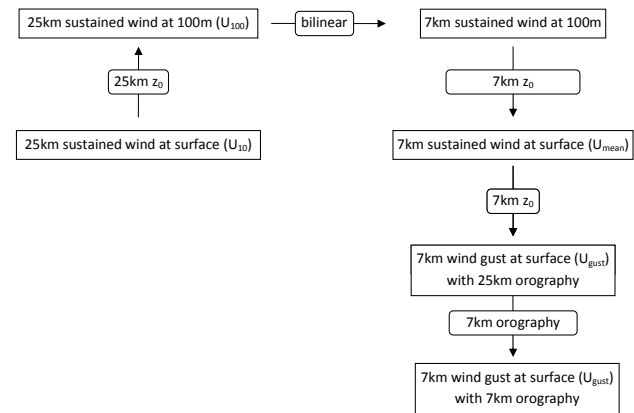

Fig. 2. Schematic of the process of downscaling of the RCM 25 km mean wind speed to 7 km wind gust speed.

(z_{ref}) assuming a logarithmic wind profile (Eq. 2) (Wieringa, 1976).

$$U_b = U_{ref} \log\left(\frac{z_b}{z_0}\right) / \log\left(\frac{z_{ref}}{z_0}\right) \tag{2}$$

Although this equation was derived for mean wind speeds, it is applied in this study using the maximum daily mean wind speed. Blending heights used in the literature for exposure correction vary. Wieringa (1976) suggests Eq. (2) is appropriate for blending heights between 60 m and 100 m in high wind speeds over rough homogenous terrain. McNaughton and Jarvis (1984) used 100 m and de Rooy and Kok (2004) used 140 m, but suggest that, although exact heights are not critical because vertical gradients are small at these heights, higher values are more appropriate for near-neutral and unstable conditions. In the present study 100 m was used, being the middle value of the above studies. The 100 m footprint (U_b) was then linearly interpolated to the 7 km grid, from which the 7 km footprint at 10 m was calculated using the 7 km roughness length and the inverse of Eq. (2). An empirical gust model (Eq. 3) (Wieringa, 1973) was applied to convert the maximum daily wind speed to maximum 3-s (t_{gust}) wind gust speed using a length scale (L) of 990 m, recommended by Wieringa (1973) as the most appropriate for the data he analysed.

$$U_{gust} = U_{mean}\left\{1 + \left(\left(1.42 + 0.301 \times \log\left(\frac{L}{U_{mean} \times t_{gust}} - 4\right)\right) / \log\left(\frac{z_{ref}}{z_0}\right)\right)\right\} \tag{3}$$

A comprehensive discussion of this gust model and comparison with that of Beljaars (1987) can be found in Verkaik (2000).

Downscaling from 25 km to 7 km can also utilise additional detail provided by the finer resolution 7 km orography. A linear regression-based downscaling was applied, using the mean vertical gradient of horizontal wind speed

$(0.0077 \, \text{m s}^{-1} \, \text{m}^{-1})$. Since a vertical wind profile was not available in the 7 km historical wind fields, the difference in 10 m wind gust speed in adjacent grid points with respect to their difference in surface elevation was used. This gradient was applied to the difference in the elevation between the coarser 25 km RCM elevation and the 7 km elevation.

3.3 Statistical scaling

Statistical scaling of the downscaled 7 km RCM wind gust speeds to the historical storm set was done individually for each model by applying a quantile-dependent scaling factor to match the statistical distributions of wind gust speed (hereafter referred to as just wind speed) of the 70 largest storms (Eq. 4) (Della-Marta et al., 2010).

$$\text{wind}_{\text{scaled}} = F_{\text{hist}}^{-1} \, F_{\text{RCM}} \, (\text{wind}_{\text{unscaled}}) \tag{4}$$

Here, F_{hist} and F_{RCM} are the empirical cumulative distribution function (ECDF) for the combined 70 largest historical and RCM storm footprints respectively. ECDFs and scaling factors were calculated for probability increments of 0.0001. ECDFs were calculated according to Eq. (5).

$$F(t) = \frac{\text{number of elements in the sample} \le t}{n} = \frac{1}{n} \sum_{i=1}^{n} 1(x_i \le t) \tag{5}$$

The scaling adopted a moving geographic window to remove any geographic bias in the models. For each 10 by 10 block of grid points (70 by 70 km), scaling factors were calculated using all points in the 70 by 70 grid point (490 by 490 km) box centred on the target box. Due to the large difference in the distribution of wind over land and sea, scaling was performed independently for land and sea.

Scaling the wind footprints as described above led to storms with lower variability in intensity than those in the historical storm set (Fig. 3a). The SSI of the largest rescaled RCM storms was generally not as high as the largest historical storms and the index did not decrease as quickly with regards to the storm rank. There are several plausible explanations for this observation. Firstly, the coarse resolution parent GCMs may not be adequately resolving the most intense cyclones, which can therefore not be adequately downscaled with the RCMs. Wernli et al. (2002) analysed the dynamics of storm Lothar using the T319L60-resolution mesoscale model simulations of the European Centre for Medium-Range Weather Forecasts and found them of insufficient temporal and spatial resolution to study Lothar's rapid intensification. However, this is an unlikely cause for the SSI differences since the 7 km COSMO model used for the historical storm set was also driven by ERA40, although it also assimilated observations within the model domain. Secondly, even if the GCM were providing cyclones of sufficient intensity for downscaling, the 25 km RCMs may still not be able to generate as intense cyclones as the higher resolution 7 km historical model. Finally, the physical parameterisation of wind gust in the historical wind fields would be expected
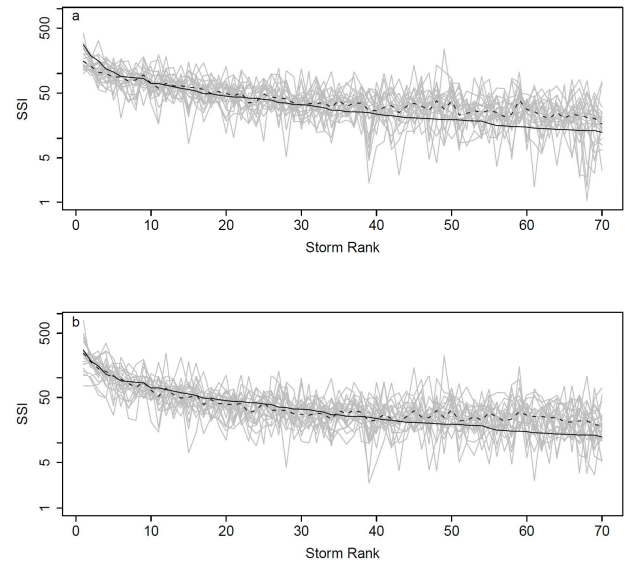

Fig. 3. Storm severity Index (SSI) for **(a)** variance-unadjusted and **(b)** variance-adjusted wind fields. The solid line is for the historical storms. Grey lines are the 22 individual RCMs with their median (dashed).

to behave differently to the empirical model derived from the maximum mean wind (Eq. 3). The SSI comparison (Fig. 3a) reveals that, whatever the cause, the statistical scaling tended to raise the mean storm intensity more than the extreme tail, evident in the lower SSI for the most intense storms (rank 1–5) and higher SSI for storms with rank 40 and greater.

The intensity of the largest storms was therefore increased by scaling the standard deviation σ_i of the individual storm footprints by an amount that depended on their relative SSI rank i (Eq. 6), before applying the global CDF scaling (Eq. 4).

$$\sigma_i' = \sigma_i \frac{e^{-ai}}{e^{-bi}}, \text{where } i \text{ is the storm rank} \tag{6}$$

Exponential functions e^{-ai} and e^{-bi} were fit to a measure of variance of wind gust speed of the individual historic and RCM footprints respectively, as a function of the rank i. The variance measure used was a robust coefficient of variation (RCOV – the median absolute deviation from the median divided by the median) for each storm footprint. This results in the highest wind speeds being intensified for the storms with the largest SSI. A robust measure of variance was used because the distribution of wind for each storm footprint was usually very positively skewed. Similar to the SSI (Fig. 3a) the RCOV was, for the largest storms, generally higher for the historical storms than the RCMs. After applying this correction and the global CDF rescaling, the resulting SSI matches much more closely the historical set for the most intense storms, but still yields storms with too high SSI for lower ranked storms (Fig. 3b).
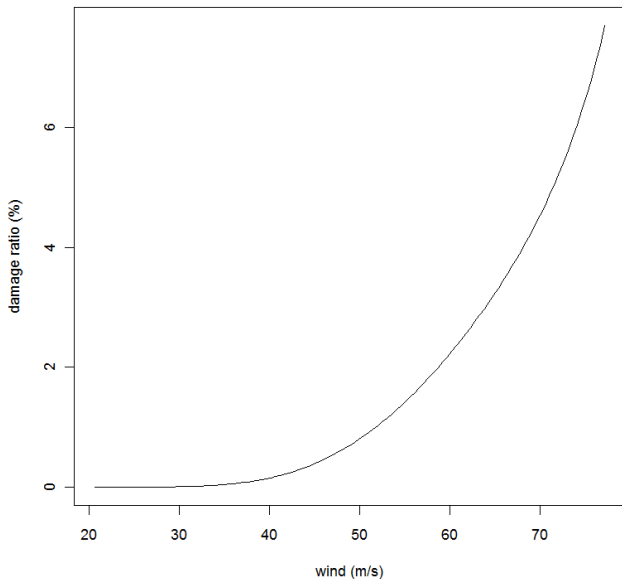
Fig. 4. Value-weighted mean transfer function to map wind speed to damage ratio.

Fig. 5. All-Europe losses for the RCM storm sets using the largest 250 storms from each RCM. Historical losses are given by black points with some important historical storms represented by letters. GPDs fitted to the individual RCMs are given by the grey lines with their mean (dotted line). The solid black line is the GPD fit through the historical losses, with 95 % confidence limits (dashed).

4 The RCM storm sets and their damages

The scaling coefficients, calculated using the largest 70 storms in the base period 1961–2000, were applied to the largest 250 storms in the full model period (Table 1). The largest 250 storms were used so as to include the largest 4–5 storms per year; a balance between computational time and ensuring that the largest damage-producing storms were not missed.

While the scaling leads to the result that no wind speeds in the rescaled storms within the base period will be greater than has been observed in history, there is still the possibility of having storms with higher SSI (and damages), as this is dependent on both wind speed and area. Also the rescaling can produce higher wind speeds than has been observed in history for those models with their highest wind speeds occurring outside the base period. An alternative method of calibration that permits the possibility of higher wind speeds than has been observed in history would be to fit extreme value distributions to the highest wind speeds in the RCM and historical storm sets and then quantile match the wind speeds with reference to these distributions. Della-Marta et al. (2010) used such an approach to calibrate indices of storms between two models. It would be worth considering here, but the geographically-dependent calibration used here would require a robust automatic threshold detection algorithm for peak-over-threshold extreme value distributions.

Since the wind speed of each RCM was independently calibrated to the historical storm set, the set of 250 storms from each RCM is considered as an alternative historical storm set to assess uncertainty in storm climatology.
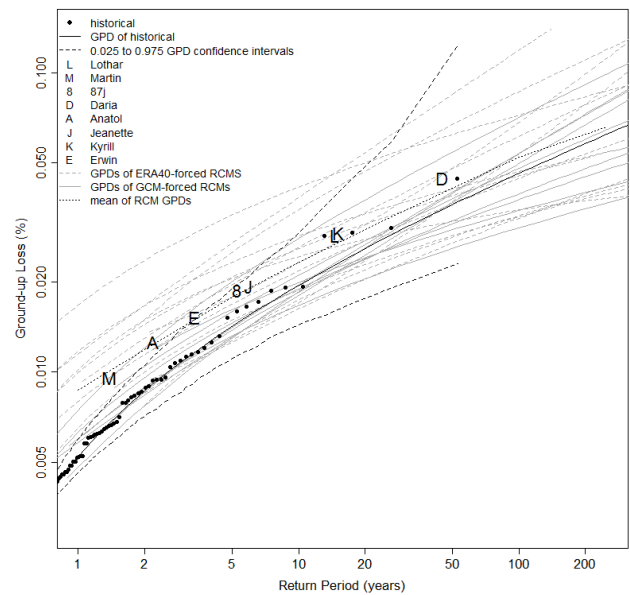
The 22 RCM storm sets were implemented in the PartnerRe CatFocus® European wind loss model and the model was run with the European population density-derived property portfolio (Sect. 2.3). PartnerRe's loss model is conceptually simple, taking as input the insured value at each location (e.g. postcode) and, for each storm in the set, finding the wind speed at the nearest grid point to each location, converting this to a damage ratio then summing the damages across all locations. The loss model maps wind speed to damage ratio using region-dependent transfer functions. A typical transfer function is given in Fig. 4, which is a value-weighted sum of the regional transfer functions across the entire portfolio.

A generalised Pareto distribution (GPD) was fitted to the historical storm losses and the losses from each of the RCM storm sets. Thresholds for the GPD were selected by minimising the Anderson-Darling (AD) goodness-of-fit test (Choulakian and Stephens, 2001). The model-dependent maximum allowable threshold was set to the 25th highest loss in each model. Choulakian and Stephens (2001) show that the asymptotic critical values of the AD test can be used for significance testing when there are at least 25 exceedances above the GPD threshold. With the optimum threshold, fitting GPDs to the historical and RCM storm sets losses produced good fits. For the historical and 21 of the 22 RCM storm sets, the AD test accepted the null hypothesis of the losses being derived from the GPD at a probability $p > 0.5$ (Choulakian and Stephens, 2001) with the
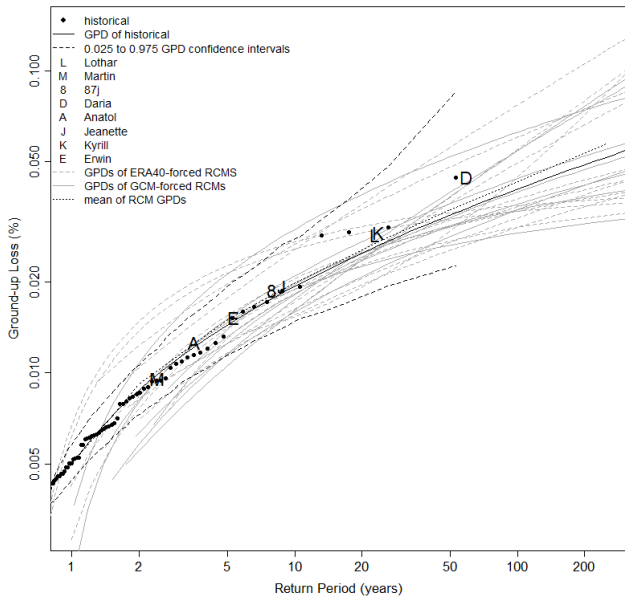
Fig. 6. As for Fig. 5 but using only the 70 largest storms from each RCM.

final RCM storm set significant at $p > 0.38$. Choulakian and Stephens (2001) suggest a threshold $p > 0.1$ for an acceptable fit and provide critical values up to $p > 0.5$. If fewer than 25 exceedences were permitted above the GPD threshold, the model-dependent threshold could be raised to give a closer fit to the tail with a lower AD value, but significance could not be easily assessed as critical values would change.

The GPD shape and scale parameters were estimated using the method of L-moments (Hosking, 1990). Della-Marta et al. (2010) found that estimation using L-moments resulted in lower biases in parameter estimates for small sample sizes than using maximum likelihood estimates.

Figure 5 shows the return period of modelled loss for the historical and 22 RCM storm sets. Confidence limits in this figure were determined by drawing random samples of the length of the historical storm set from a GPD with the fitted shape and scale parameters. The confidence limits are therefore not meant to represent the parametric uncertainty in GPD fitting, but rather the goodness of fit of the RCM storm set losses to the GPD fitted to the historical storm set. Confidence limits for the individual RCM GPDs have not been included so as not to clutter the diagram, but would be comparable in width to the historical set limits (dashed lines) since they are derived from a similar length storm set. Figure 5 also includes the individual historical storms. The empirical return period RP_i years for a storm with a loss of rank i is given by Eq. (7).

$$RP_i = \frac{n+1}{\lambda i}, \text{ where } \lambda = \frac{n}{\text{years}} \qquad (7)$$

Here the empirical distribution function of Coles (2001) is used. Makkonen (2008) argues that, despite a long history of proposed alternative functions, the formulation in Eq. (7) is the only correct one.

Return levels of modelled loss (Fig. 5) for the mean of the RCM storm sets (dotted line) are generally higher than the historical storm set (solid black line) at all return periods, especially lower periods. Ideally our RCM calibration should produce a loss curve that agrees with the historical storm set, at least for low return periods where uncertainty is lower in the empirical return period assigned to the historical storms. The higher return level in the RCM storms indicates possible inadequacies in our assumptions or methodology. Firstly, our assumption that the historical storm set contains the largest 70 historical storms in the base period was based on an SSI derived from coarse scale ERA40 data. If we had 7 km storm footprints for the complete historical period (not possible due to computational constraints), it is likely that the 70 storms with the highest SSI would not match exactly the 70 selected from the coarse scale data. Non-linear wind indices, such as our SSI, are very sensitive to small changes in the wind field at the region of highest wind speeds. Secondly, the largest 70 storms were used to derive scaling relationships and these were applied to the largest 250 storms. Despite applying an exponentially decaying scaling of the wind field variance based on their SSI rank (Sect. 3.3), intermediate storms are being scaled such that their losses are too high. The higher SSI in the rescaled RCM wind fields compared with the lower order historical storms (Fig. 3b – rank 45 and higher) suggests this is likely a problem. Therefore, to be able to use the largest 250 storms from the RCMs, more historical storms are needed against which to calibrate. Alternatively, another method to preprocess the RCM wind fields to better match the historical SSI is necessary. Thirdly, the higher losses in the rescaled RCMs may be indicative of wind fields that are simply too different in physical character to the historical fields, having been derived from an empirical gust parameterisation applied to daily maximum mean wind speed compared with the more physically based turbulent and convective parameterisations in the COSMO model. While the lower order moments of the 25 km RCM maximum mean wind were pre-processed to be closer to the historical 7 km maximum wind gust, the downscaled fields may still be too spatially broad to give return levels of loss identical to the historical set. Again, the loss is a non-linear (cubic polynomial in log-log space) function of wind speed and very sensitive to the magnitude and shape of the wind field.

Taking only the largest 70 storms in the RCM storm sets brings the return period curve of the mean of the RCM storm sets much closer to the historical (Fig. 6). The mean RCM curve lies well within the confidence bounds of the historical set for all return periods. This suggests that our second concern above, that we can't use the largest 250 storms from the RCMs when we have only the largest 70 historical storms for calibration, is an important conclusion.
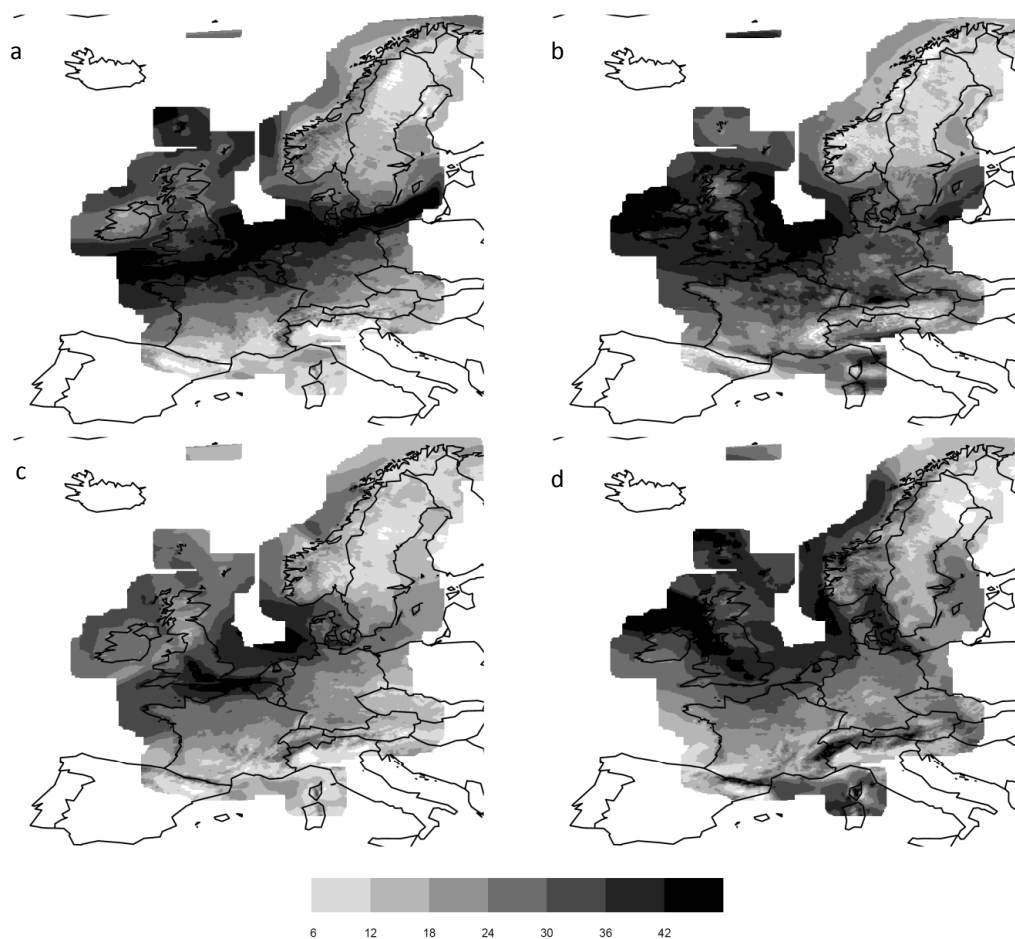
Fig. 7. Wind footprints for the RCM storms with the four highest losses (m s^{-1}). **(a)** SMHI RCA ERA40 1990-01-23, **(b)** RPN GEMLAM ERA40 1990-02-26, **(c)** METO HadRM3Q0 HadCM3Q0 2003-11-14 and **(d)** METO HadRM3Q0 HadCM3Q0 2008-03-03.
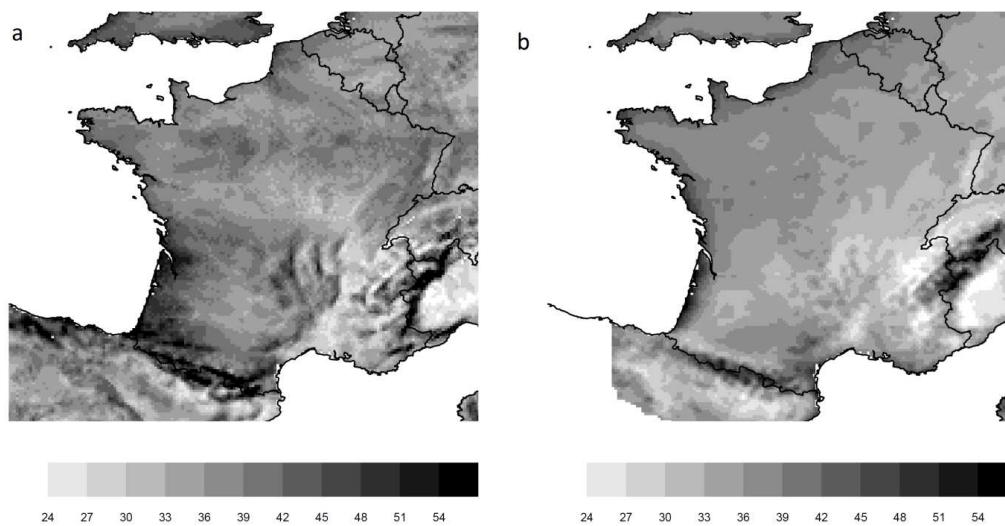

Fig. 8. Circa 50-yr return level of wind gust (m s^{-1}) for the **(a)** historical (rank 1 grid-point wind speed) and **(b)** RCM (mean rank 1 grid-point wind speed) storm sets.

Table 2. Modelled damage ratios (% of total value) to the European population density-derived portfolio using the RCM and historical storm sets.

years	Historical	min(RCM)	mean(RCM)	max(RCM)
1	0.005	0.003	0.005	0.007
2	0.009	0.006	0.009	0.015
5	0.014	0.011	0.015	0.023
10	0.019	0.015	0.020	0.030
25	0.027	0.021	0.028	0.042
50	0.033	0.025	0.035	0.056
100	0.041	0.029	0.043	0.077
250	0.052	0.032	0.057	0.117

The footprints of the RCM storms with the four highest losses are shown in Fig. 7. The largest two storms, driven by the ERA40 reanalyses, have historical counterparts: Daria in 1990 (also the largest historical loss – Fig. 5) and Wiebke, also in 1990. The 3rd and 4th highest losses are from GCM-forced models and therefore do not have historical origins. Both these artificial storms incur large losses due to their high impact over London and southeast England.

Using the RCM storm sets with the largest 70 storms, expected losses to our population density-derived portfolio at various return periods are presented (Table 2). The advantage of using the RCM storm sets for risk assessment is illustrated in Fig. 8, which shows the empirical circa 50-yr return level of wind speed for the historical and RCM storm sets over France. This was calculated as the maximum wind at each grid point for the historical set and the mean of the RCM maximum grid-point wind speeds. For the historical set (Fig. 8a), the east-west lineation reflects the tracks of individual storms, with the 1999 storm Lothar across the north of the country and the 1999 and 2009 storms Martin and Klaus in the southwest. The RCM mean (Fig. 8b) return levels are smoother due to the 20-fold increase in the number of storms sampled.

5   Conclusions

We have demonstrated here an end-to-end study that has used 22 RCM runs to produce statistically-homogenous sets of storms useful for quantifying storm risk to an insurance portfolio. Like many such studies, subtle decisions at the front end of the process (model calibration) can have large impacts at the outcome (risk assessment). This illustrates the great benefit brought by application-driven analyses.

The greatest challenge in this multi-model study was to calibrate the individual models to a common base. It was critical that this step required a point of reference: that it was objectively determined from an analysis of ERA40 wind data that the high resolution historical storm set contained the 70 storms with the highest SSI in the calibration period

1960–2000. Therefore, a key assumption of the calibration is that wind scaling factors were defined based on matching the 70 largest storms in each model over the base period with the historical storm set. Without such a point of reference, matching two storm sets becomes a more subjective process.

By identifying and ranking storms in the RCMs and historical storm set using a model-dependent SSI, wind scaling factors were calculated using the largest 70 storms in the 40-yr base period 1961–2000 for which all models were available. These scaling factors were applied to the largest 250 storms in each RCM for the period 1950–2008. When run through the PartnerRe CatFocus® European wind catastrophe model with a Europe-wide population density-derived property portfolio, the mean of the RCM storm sets gave loss return levels that were higher than for the historical set. We suspect this was primarily due to limitations of extrapolating the SSI-dependent scaling relationships developed with the largest 70 storms out to the largest 250 storms, but could also be partly due to our historical storm set lacking some of the largest storms. Selecting the largest 70 storms in each model gave a much closer match in return levels between the historical and mean of the RCM storm sets.

While the calibration has yielded individual RCM storm sets whose mean Europe-wide loss curve is close to a GPD fitted to the historical set losses, we would not expect this to be the case for smaller regional insurance portfolios. The benefit of using the RCM storm sets for risk assessment becomes much more apparent for smaller portfolios where sampling uncertainty from a 50-yr historical storm set is much higher. Such portfolios may have only seen a couple of large losses in the last 50 yr, or even been lucky enough to miss out on direct storm hits entirely.

Edited by: U. Ulbrich
Reviewed by: M. K. Kunz and two other anonymous referees

References

Beljaars, A. C. M.: The influence of sampling and filtering on measured wind gusts, J. Atmos. Ocean Tech., 4, 613–626, doi:10.1175/1520-0426(1987)004<0613:TIOSAF>2.0.CO;2, 1987.

Choulakian, V. and Stephens, M. A.: Goodness-of-fit tests for the Generalized Pareto Distribution, Technometrics, 43, 478–484, doi:10.1198/00401700152672573, 2001.

Christensen, J. H., Boberg, F., Christensen, O. B., and Lucas-Picher, P.: On the need for bias correction of regional climate change projections of temperature and precipitation, Geophys. Res. Lett., 35, L20709, doi:10.1029/2008GL035694, 2008.

Christensen, J. H., Rummukainen, M., and Lenderink, G.: Formulation of very-high-resolution regional climate model ensembles for Europe, in: ENSEMBLES: Climate Change and its Im-

pacts: Summary of research and results from the ENSEMBLES project., edited by: van der Linden, P. and Mitchell, J. F. B., Met Office Hadley Centre, Exeter, UK, 47–58, 2009.

Coles, S.: An introduction to statistical modeling of extreme values, Springer series in statistics, Springer, London, 2001.

de Rooy, W. C. and Kok, K.: A combined physical-statistical approach for the downscaling of model wind speed, Weather Forecast., 19, 485–495, 10.1175/1520-0434(2004)019¡0485:ACPAFT¿2.0.CO;2, 2004.

Della-Marta, P. M., Mathis, H., Frei, C., Liniger, M. A., Kleinn, J., and Appenzeller, C.: The return period of wind storms over Europe, Int. J. Climatol., 29, 437–457, doi:10.1002/joc.1794, 2009.

Della-Marta, P. M., Liniger, M. A., Appenzeller, C., Bresch, D. N., Koellner-Heck, P., and Muccione, V.: Improved estimates of the European winter wind storm climate and the risk of reinsurance loss using climate model data, J. Appl. Meteorol. Clim., 49, 2092–2020, doi:10.1175/2010JAMC2133.1, 2010.

Donat, M. G., Leckebusch, G., Wild, S., and Ulbrich, U.: Benefits and limitations of regional multi-model ensembles for storm loss estimations, Clim. Res., 44, 211–225, doi:10.3354/cr00891, 2010.

Enz, R., Zimmerli, P., and Schwartz, S.: Natural catastrophes and man-made disasters in 2008, Swiss Reinsurance Company Limited, Zurich, 2/2009, 41, 2009.

GPWv3: Gridded Population of the World Version 3, Socioeconomic Data and Applications Center (SEDAC), Columbia University, Palisades, NY, 2005.

Guy Carpenter: Windstorm Erwin/Gudrun – January 2005, Guy Carpenter & Company, New York, 2005.

Hall, T. M. and Jewson, S.: Comparison of Local and Basinwide Methods for Risk Assessment of Tropical Cyclone Landfall, J. Appl. Meteorol. Clim., 47, 361–367, doi:10.1175/2007JAMC1720.1, 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.

Hosking, J. R. M.: L-moments: analysis and estimation of distributions using linear combinations of order statistics, J. R. Stat. Soc., 52B, 105–124, 1990.

Lamb, H. H. and Frydendahl, K.: Historic storms of the North Sea, British Isles and Northwest Europe, edited by: Frydendahl, K., Cambridge University Press, Cambridge, UK, 204 pp., 1991.

Leckebusch, G., Ulbrich, U., Froelich, L., and Pinto, J. G.: Property loss potentials for European mid-latitude storms in a changing climate, Geophys. Res. Lett., 34, L05703, doi:10.1029/2006GL027663, 2007.

Leckebusch, G., Renggli, D., and Ulbrich, U.: Development and application of an objective storm severity measure for the northeast Atlantic region, Meteorol. Z., 17, 575–587, doi:10.1127/0941-2948/2008/0323, 2008.

Makkonen, L.: Bringing closure to the plotting position controversy, Commun. Stat. Theory, 37, 460–467, doi:10.1080/03610920701653094, 2008.

McNaughton, K. G. and Jarvis, P. G.: Using the Penman-Monteith equation predictively, Agr. Water. Manage., 8, 263–278, doi:10.1016/0378-3774(84)90057-X, 1984.

Nakamura, K., Kershaw, R., and Gait, N.: Prediction of near-surface gusts generated by deep convection, Meteorol. Appl., 3, 157–167, doi:10.1002/met.5060030206, 1996.

Nakicenovic, N., Gruebler, A., Ishitani, H., Johansson, T., Marland, G., Moreira, J. R., and Rogner, H. H.: Energy Primer, in: Climate Change 1995, Impacts, Adaptation and Mitigation of Climate Change: Scientific-Technical Analyses, edited by: Watson, R. T., Zinyowera, M. C., and Moss, R. H., Cambridge University Press, Cambridge, UK, 1996.

Palutikof, J. P., Brabson, B. B., Lister, D. H., and Adcock, S. T.: A review of methods to calculate extreme wind speeds, Meteorol. Appl., 6, 119–132, doi:10.1017/S1350482799001103, 1999.

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.

Rockel, B., Will, A., and Hense, A.: The regional climate model COSMO-CLM (CCLM), Meteorol. Z., 17, 347–348, doi:10.1127/0941-2948/2008/0309, 2008.

Schulz, J. P. and Heise, E.: A new scheme for diagnosing near-surface convective gusts, COSMO Newsletter, 3, 221–225, 2003.

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

Uppala, S. M., Kallberg, P. W., Simmons, A. J., Andrae, U., Bechtold, V. D., Fiorino, M., Gibson, J. K., Haseler, J., Hernandez, A., Kelly, G. A., Li, X., Onogi, K., Saarinen, S., Sokka, N., Allan, R. P., Andersson, E., Arpe, K., Balmaseda, M. A., Beljaars, A. C. M., Van De Berg, L., Bidlot, J., Bormann, N., Caires, S., Chevallier, F., Dethof, A., Dragosavac, M., Fisher, M., Fuentes, M., Hagemann, S., Holm, E., Hoskins, B. J., Isaksen, L., Janssen, P. A. E. M., Jenne, R., McNally, A. P., Mahfouf, J. F., Morcrette, J. J., Rayner, N. A., Saunders, R. W., Simon, P., Sterl, A., Trenberth, K. E., Untch, A., Vasiljevic, D., Viterbo, P., and Woollen, J.: The ERA-40 re-analysis, Q. J. Roy. Meteor. Soc., 131, 2961–3012, doi:10.1256/qj.04.176, 2005.

Verkaik, J. W.: Evaluation of two gustiness models for exposure correction calculations, J. Appl. Meteorol., 39, 1613–1626, doi:10.1175/1520-0450(2000)039<1613:EOTGMF>2.0.CO;2, 2000.

Wernli, H., Dirren, S., Liniger, M. A., and Zillig, M.: Dynamical aspects of the life-cycle of the winter storm "Lothar" (24–26 December 1999), Q. J. Roy. Meteor. Soc., 128, 405–428, doi:10.1256/003590002321042036, 2002.

Wieringa, J.: Gust factors over open water and built-up country, Bound.-Lay. Meteorol., 3, 424–441, doi:10.1007/BF01034986, 1973.

Wieringa, J.: An objective exposure correction method for average wind speeds measured at a sheltered location, Q. J. Roy. Meteor. Soc., 102, 241–253, doi:10.1002/qj.49710243119, 1976.

Wood, N. and Mason, P. J.: The pressure force induced by neutral turbulent flow over hills, Q. J. Roy. Meteor. Soc., 119, 1233–1297, doi:10.1002/qj.49711951402, 1993.