

Improved Estimates of the European Winter Windstorm Climate and the Risk of Reinsurance Loss Using Climate Model Data

PAUL M. DELLA-MARTA, MARK A. LINIGER, AND CHRISTOF APPENZELLER

Federal Office for Meteorology and Climatology, MeteoSwiss, Zurich, Switzerland

DAVID N. BRESCH

Swiss Reinsurance Company, Zurich, Switzerland

PAMELA KÖLLNER-HECK

Swiss Reinsurance Company, Zurich, and Federal Office for the Environment, Bern, Switzerland

VERUSKA MUCCIONE

Swiss Reinsurance Company, and Myclimate, Zurich, Switzerland

(Manuscript received 6 October 2008, in final form 22 March 2010)

ABSTRACT

Current estimates of the European windstorm climate and their associated losses are often hampered by either relatively short, coarse resolution or inhomogeneous datasets. This study tries to overcome some of these shortcomings by estimating the European windstorm climate using dynamical seasonal-to-decadal (s2d) climate forecasts from the European Centre for Medium-Range Weather Forecasts (ECMWF). The current s2d models have limited predictive skill of European storminess, making the ensemble forecasts ergodic samples on which to build pseudoclimates of 310–396 yr in length. Extended winter (October–April) windstorm climatologies are created using scalar extreme wind indices considering only data above a high threshold. The method identifies up to 2363 windstorms in s2d data and up to 380 windstorms in the 40-yr ECMWF Re-Analysis (ERA-40). Classical extreme value analysis (EVA) techniques are used to determine the windstorm climatologies. Differences between the ERA-40 and s2d windstorm climatologies require the application of calibration techniques to result in meaningful comparisons. Using a combined dynamical–statistical sampling technique, the largest influence on ERA-40 return period (RP) uncertainties is the sampling variability associated with only 45 seasons of storms. However, both maximum likelihood (ML) and L-moments (LM) methods of fitting a generalized Pareto distribution result in biased parameters and biased RP at sample sizes typically obtained from 45 seasons of reanalysis data. The authors correct the bias in the ML and LM methods and find that the ML-based ERA-40 climatology overestimates the RP of windstorms with RPs between 10 and 300 yr and underestimates the RP of windstorms with RPs greater than 300 yr. A 50-yr event in ERA-40 is approximately a 40-yr event after bias correction. Biases in the LM method result in higher RPs after bias correction although they are small when compared with those of the ML method. The climatologies are linked to the Swiss Reinsurance Company (Swiss Re) European windstorm loss model. New estimates of the risk of loss are compared with those from historical and stochastically generated windstorm fields used by Swiss Re. The resulting loss-frequency relationship matches well with the two independently modeled estimates and clearly demonstrates the added value by using alternative data and methods, as proposed in this study, to estimate the RP of high RP losses.

Corresponding author address: Dr. Paul M. Della-Marta, Partner Reinsurance Company, 36 Bellerivestrasse, Zurich 8034, Switzerland.
E-mail: paul.della-marta@partnerre.com

DOI: 10.1175/2010JAMC2133.1

© 2010 American Meteorological Society

1. Introduction

Extreme winds associated with midlatitude cyclones represent a major loss potential for reinsurance companies (Swiss Re 2000). To quantify this risk many insurance companies use a combination of historical analyses of windstorms and stochastic and/or dynamical models to generate hypothetical windstorms. Generally, all data sources, whether they are based on historical observations or dynamical models, have various advantages and disadvantages. Historical datasets are relatively short, have coarse spatiotemporal resolution, or suffer from inhomogeneities, whereas dynamical models have biases associated with their physics and parameterizations and their use is often a trade-off between spatiotemporal resolution and integration length.

Reinsurance companies are especially interested in the frequency (or return period) of very rare extreme wind events on continental and synoptic scales and their associated losses. The primary influence on the accuracy of the estimation of the frequency of rare events is the number of observations of such events. Della-Marta et al. (2009) used the 40-yr European Centre for Medium-Range Weather Forecasting (ECMWF) reanalysis (ERA-40) to estimate the return period (RP) of extreme winds at continental and regional scales. They found that the uncertainties associated with long RPs (approximately 30 yr) are in the range between -60% and $+200\%$ of the RP estimate. This places limitations on the use of these data for reinsurance risk analysis applications.

The main aim of this study is to reduce the uncertainty associated with estimating the RP of very rare windstorm events. Previous studies investigating the frequency of extreme winds or extreme cyclones in the Atlantic–European domain use a number of different data and methodologies depending on the aim of the study (see Della-Marta et al. 2009). Generally all such observation-based studies are limited by the length and/or resolution of the data used. Improvements have been made using statistical resampling techniques (e.g., Dukes and Palutikof 1995; Palutikof et al. 1999), allowing more accurate inferences about the nature of the parent distribution. These studies suffer from incomplete knowledge of the dynamics that lead to such extremes. Alternatively, long ensemble integrations of dynamical coupled atmosphere–ocean general circulation models (AOGCMs) can be used as a dynamical sample to estimate the statistics of extreme events. Applications include the estimation of extreme synoptic-scale winds (van den Brink et al. 2004a), coupling of extreme atmospheric and ocean conditions with hydrological models (van den Brink et al. 2004b, 2005), and the statistics of extreme temperatures

(Vannitsem 2007). Broadly speaking, these types of data are referred to as seasonal-to-decadal (s2d) data [see Schwierz et al. (2006) for an overview].

The second major aim of this study is to investigate the feasibility of coupling s2d windstorm climatologies with the propriety loss model of the Swiss Reinsurance Company (Swiss Re) to gain a better understanding of high RP losses to a Swiss Re portfolio. In this respect, this study is similar to Schwierz et al. (2009), who investigated windstorm-related losses in expected future climate scenarios. Other studies that look at current and future extreme wind-related damage are Klawa and Ulbrich (2003), Heneka et al. (2006), Leckebusch et al. (2007), and Pinto et al. (2007). However, these studies are either regionally focused or use limited sample size datasets or reanalyses as the reference current windstorm climate.

It is perhaps worth briefly explaining why we are interested in using dynamical models (e.g., s2d) for the estimation of extreme value statistics. Observations of the climate system represent a sample of its dynamics whose elements interact on different time scales. The fact that some of these time scales exceed the length of our observations (e.g., ocean processes) suggests that observation-based extreme value statistics probably do not sample the entire range of possible values. While statistical resampling helps to overcome the observation length limitations it cannot reproduce properties of dynamical systems such as bifurcations, fixed points, etc. A key assumption of classical extreme value theory is that the distribution function is continuous and differentiable (Gumbel 1958). Balakrishnan et al. (1995) show that this is not valid for simple dynamical systems, which display deterministic chaos. Moreover, Vannitsem (2007) shows that the statistics of temperature extremes in a simplified general circulation model (GCM) only weakly converge (as a function of integration time) to a stable extreme value distribution.

We follow a similar approach to van den Brink et al. (2004b, 2005) in this study by using AOGCMs that are run in seasonal forecast mode (Anderson et al. 2003; Palmer et al. 2004; Anderson 2008). These models are initialized with perturbed initial conditions based on observations at the beginning of each season and then integrated freely with boundary conditions thereafter (such as anthropogenic forcings; Liniger et al. 2007). At the point in forecast lead time where the skill of forecasts is close to zero the ensemble members can be treated as independent samples of the model climate system (i.e., ergodic samples). A possible drawback of using such models to estimate the current (and future) climate is that the results are likely to be model dependent because of many model uncertainties (see, e.g.,

TABLE 1. An overview of the datasets used in this study.

Dataset	Resolution (spherical harmonics)	Output time resolution (h)	No. of hindcast years (year range)	Combination of forecasts used to create an ONDJFMA season	Total number of ensemble members
SYS3	T159	12	26 (1981–2006)	September (1) and November (2) forecasts: 1111222	316 (11 for each forecast, 41 for 2006)
SYS2	T95	12	20 (1987–2006)	September (1) and November (2) forecasts: 1111222	310 (5 between 1987 and 2000; 40 between 2001 and 2006)
EC-DEM	T95	24	44 (1958–2001)	August (1) and November (2) forecasts: 1111222	396 (9 for each forecast)
ERA-40	T159	6	45 (1957–2001)	—	—

Kharin and Zwiers 2000; Schwierz et al. 2006). These uncertainties need to be addressed when we use the Swiss Re loss model, which relies on observations for its calibration.

Below we elaborate the specific aims of this study:

- 1) to quantify the predictability of storminess over the western European domain in the current generation of s2d climate forecasts from the ECMWF;
- 2) to assess the usefulness of s2d data as a surrogate (pseudo) climate to estimate the frequency and intensity of windstorms over Europe and to quantify the amount to which ERA-40 may under- or overestimate the severity and frequency of windstorms;
- 3) to apply the new storm climate to a user application model, namely, the Swiss Re windstorm loss model, to compare the loss-frequency estimates based on the Swiss Re and s2d windstorm climatologies.

We start with an overview of the data and models used in this study and the calculation of extreme wind indices (EWI) that will form the basis of estimating the windstorm climate. The following sections describe the methodology and the main results, including an intercomparison of the extreme wind climatologies and modeled loss estimates, followed by some discussion and conclusions.

2. Data and model description

In this section we describe the datasets used in this study and the necessary background details about the Swiss Re loss model.

a. S2d model data and ERA-40 reanalysis data

In this study we use the ECMWF seasonal climate forecasts, which are also known as s2d forecasts (e.g., Anderson 2008), and ERA-40 (Uppala et al. 2005). The 10-m-level wind gust speed is a useful parameter for assessing wind extremes at the surface that are responsible for damage to infrastructure and the loss of human life. A previous study by Della-Marta et al. (2009) demonstrated that the wind gust parameterization in ERA-40 produces

unrealistic wind gust estimates in the areas of either steep orographic gradients or coastal regions of Europe. The same parameterization has been used in earlier versions of the ECMWF s2d forecasts. The latest version of the ECMWF s2d forecast model (system 3; ECMWF 2007; Anderson et al. 2007) has an improved wind gust parameterization and therefore could be used in this study; however, to maximize the use and transferability of analysis methods, we use geostrophic wind speed (GW) at the 850-hPa level. By using GW we lose both information about real surface winds affected by, for example, surface roughness, and possible intensification associated with mesoscale frontal features. On the other hand we gain a wind climatology that is independent of the various parameterizations and one that focuses model differences on those associated with large-scale circulation. The latter represents an advantage for model intercomparisons.

We use three different sets of s2d hindcasts from the ECMWF (an overview is given in Table 1). The latest s2d model is known as system 3 (SYS3) and is based on the physics documented in ECMWF (2007). Other datasets are system 2 (SYS2) and the ECMWF model used in the Development of a European Multimodel Ensemble System for Seasonal-to-Interannual Prediction (DEMETER) project and the first phase of the European Union Sixth Framework Programme (EU FP6) Ensemble-Based Predictions of Climate Changes and their Impacts (ENSEMBLES) project (EC-DEM; Palmer et al. 2004) both based on similar physics (ECMWF 2003). GW is calculated from the model geopotential height (GPH) at 850 hPa (Holton 2004). Our analysis concentrates on the extended winter season, October–April, and we focus our results on the eastern North Atlantic and western Europe (35°–73°N, 15°W–30°E) domain. Figures 1a,b present the domain and example wind fields as represented in the Swiss Re dataset and ERA-40 GW data, respectively. Note that the October–April season of s2d data is formed using a composite of data from two different forecast initialization dates. Rationale for this is found in the predictability assessment presented in section 3a.

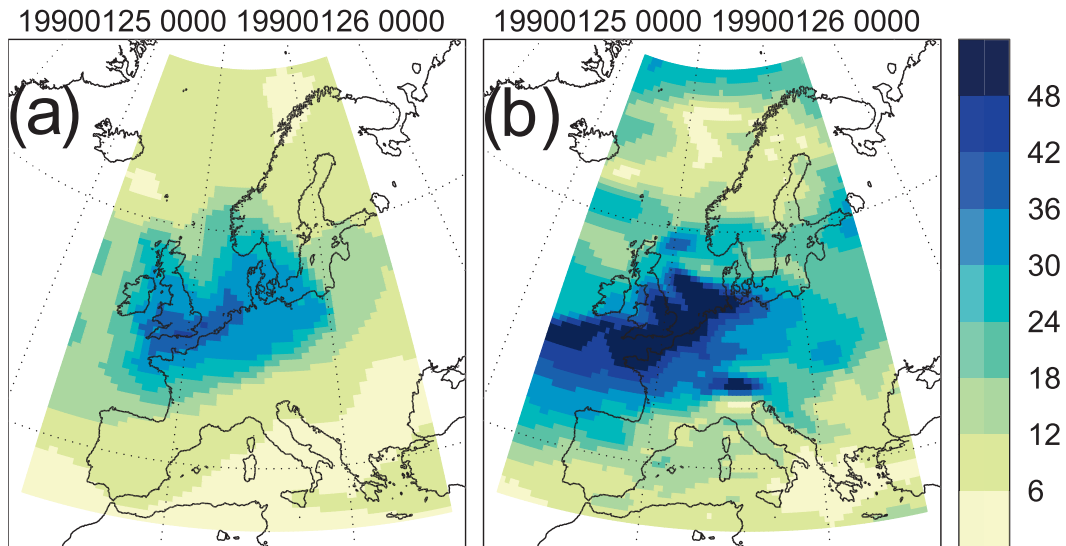


FIG. 1. Example wind fields (m s^{-1} ; legend on the right-hand side) from the Swiss Re operational windstorm hazard dataset for Daria, which occurred between 0000 UTC 25 Jan and 0000 UTC 26 Jan 1990. (a) The maximum wind gust at the 100-m level over the lifetime of the storm at each grid point. The resolution of the gridded dataset is 0.5° . (b) The maximum ERA-40 850-hPa GW. See sections 2a and 2c for more details.

b. Extreme wind indices

Della-Marta et al. (2009) used a number of different scalar indices (time series, known as EWI) to summarize a windstorm's magnitude and spatial extent over a given domain. Three of the five indices used in the study (Mean, Q95, and Sw3q90) are based on absolute measures of the wind speed, whereas two (Sfq95 and Sfq95q99) are based on wind speeds relative to local climatology (see appendix B for their exact definition). They found that the EWIs—Mean, Sfq95, and Sfq95q99—result in similar RP estimates for individual storms in ERA-40 and recommend that these indices be used when more weight to local winds relative to their climatology is needed and to find storms affecting regions not subject to regular extreme wind events. The EWIs Sw3q90 and Q95 result in similar RP estimates for individual storms. These indices should be used when more weight to the absolute magnitude of a windstorm is needed, regardless of the local wind climatology. Throughout the remainder of the paper we will concentrate on presenting results of the Sw3q90 index, which is the sum of wind speed cubed above the extended winter season 90th percentile, and the Sfq95q99 index, which is the sum of the fraction of extreme wind divided by the length of the distribution tail at each grid point. See appendix B for a complete specification of these indices and Della-Marta et al. (2009) for more details, including an intercomparison of the RPs of historical storms based on these indices.

c. Swiss Re windstorm data

Swiss Re bases its estimate of European windstorm risk on a dataset consisting of around 150 historical windstorm fields from 1947 to the present (SR Hist) and a derived, stochastically generated windstorm field dataset (SR Stoch). Each field represents the maximum wind gust at 10-m level over the lifetime of the storm. An example is shown in Fig. 1b. The creation of a historical wind field is a multistep process starting with the digitization of surface air pressure and the position of frontal features from weather charts. Quasigeostrophic principles are used to create a first guess of the sustained wind field. Frontal gust correction factors are then calculated based on the linear relationship between the quasigeostrophic wind speeds and measured gust wind speed at the positions of the digitized fronts. Then a third and final stage is the correction of the wind field by adding a bilinearly interpolated grid of observed 10-m wind gusts values in the regions where the storm had its highest impacts. The SR Hist dataset has undergone many iterations to improve its quality, from specific corrections in Alpine regions to country-specific validation/corrections with in situ wind measurements. The process of developing such a dataset is labor intensive and has been focused on mapping wind fields that have affected Swiss Re's business. For this reason, the dataset represents a subset of the extreme wind climate over Europe and is not necessarily a homogeneous sample of the storminess over the larger domain considered here (see Della-Marta et al. 2009 for more discussion). To

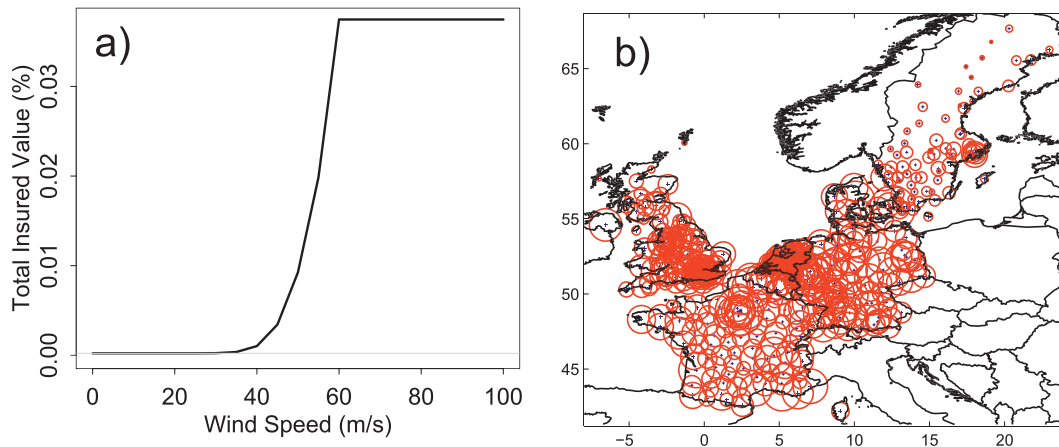


FIG. 2. (a) The catXos vulnerability curve and (b) the value distribution of the Swiss Re portfolio used in this study. Panel (a) shows the expected loss (percent of the total insured value, y axis) as a function of wind speed intensity (m s^{-1} , x axis). In (b) the small blue crosses denote the location of the asset in the portfolio, and the size of the circle represents the relative value of the asset. See section 2d for more details.

overcome these shortcomings Swiss Re has created a stochastically generated windstorm dataset, using the historical fields as a basis. Each historical windstorm field has been resampled by perturbing its intensity and position within physically constrained boundaries. The position of each historical storm is perturbed either by a rigid motion or a rotation with the bounds of $\pm 50\text{--}100$ km and $\pm 5^\circ$ (along the major axis of storm movement), respectively. Each position perturbation is considered equally likely as the original historical wind field on which it is based; however, the wind intensity of the historical wind fields are altered subject to the constraint of the observed frequency–intensity relationship derived from a storm severity index (Lamb and Fyendahl 1991; similar the Sw3q90 EWI). In other words, each intensity perturbed storm is assigned a frequency depending on its observed frequency at a given intensity. Frequencies for the most extreme perturbed windstorms, where there are few observed storm severity index values, are assigned by using an extreme value distribution fitted to the observed index data. This ensures consistency in the frequency–intensity in SR Hist and SR Stoch datasets. The SR Stoch dataset represents around 3000 years of synthetic windstorm climate and contains 10 010 events. See Schwiertz et al. (2009) for more details.

d. Swiss Re loss model

The proprietary loss model of the Swiss Re is based on four modules (Zimmerli 2003; Schwiertz et al. 2009): (i) the hazard module is essentially the SR Hist and SR Stoch datasets described above (Fig. 1b). (ii) The vulnerability module models a measure of damage (loss) as a function of the intensity of an event. A vulnerability curve (see Fig. 2a) is derived from loss data or based on

engineering considerations for each risk type (e.g., residential, commercial, agricultural). (iii) The value distribution describes the amount (total sum insured), location, and risk type of the insured values (see Fig. 2b). (iv) The insurance conditions comprise all the information on the specific details of particular contracts and determine the proportion of the loss that is insured. In this study (as also used in Schwiertz et al. 2009) a slightly simplified version of the operational Swiss Re model (called catXos) is employed. It uses an average vulnerability curve for all risk types and locations (Fig. 2a). The western European property insurance portfolio (Fig. 2b) is chosen to be representative of the western European market.

3. Methodology

In the following section we describe the most important details of methods used throughout the paper. It starts with the methodology associated with and the results of assessing the prediction skill of seasonal western European storminess using s2d models. It continues with an overview of the storm selection process and extreme value analysis (EVA) techniques, followed by methods of calibration applied to storm fields or an EWI. All analyses of uncertainty are presented as a 95% confidence interval (CI) unless otherwise stated.

a. Prediction skill for winter storminess

We start with investigating the large-scale bias of s2d models relative to ERA-40 for the mean extended winter season GPH at 850 hPa. Figure 3a shows the mean pattern of GPH in ERA-40 and Figs. 3b,c,d show the bias of the s2d models relative to ERA-40. SYS3 GPH

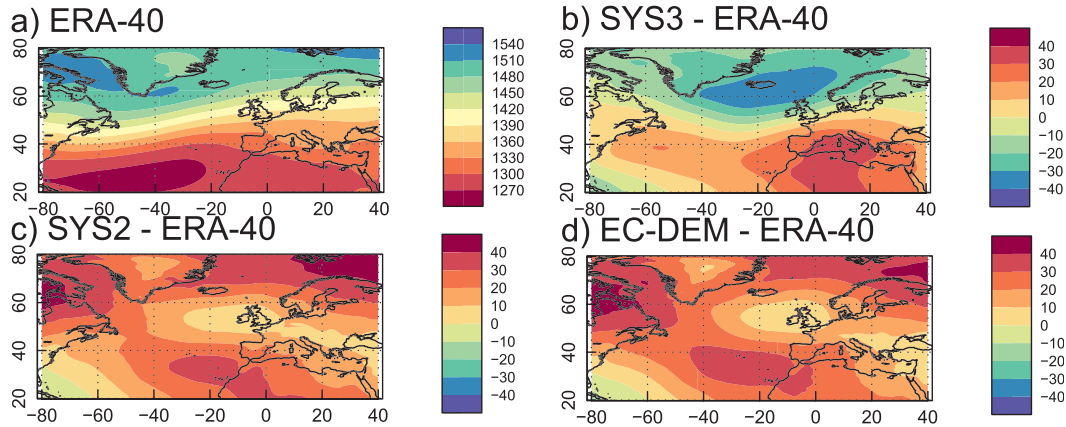


FIG. 3. The long-term extended winter mean GPH at 850 hPa (m) over the North Atlantic and European domain. (a) ERA-40 and the difference with s2d data: (b) SYS3, (c) SYS2, (d) EC-DEM.

differences display an enhanced westerly flow compared to ERA-40 (Fig. 3b), with a deeper Icelandic low and a stronger Azores high, whereas both SYS2 and EC-DEM display a general positive GPH bias over the North Atlantic region compared to ERA-40. Biases in trends are not considered here but can be assumed to be of minor relevance for GPH in the extratropics (Liniger et al. 2007).

The skill in predicting extended winter season storminess is assessed in terms of evaluating the Q95 index (appendix B) from s2d model data (specified in Table 1) against that from the ERA-40 data. Two skill metrics are used: (i) correlation skill score, which correlates the ERA-40 Q95 with the ensemble mean s2d Q95 for the months November–April (May) [NDJFMA(M)] and (ii) debiased ranked probability skill score (RPSSd; Weigel et al. 2007), which measures the probabilistic predictability of quantile forecasts (here we use terciles). The skill metrics are computed from monthly mean Q95 values. Sampling uncertainty is assessed using a bootstrap with replacement 1000 times.

Figure 4 shows the results of each skill metric applied to each s2d model. A notable feature is significant positive skill for the first month of each of the models' forecasts (except EC-DEM RPSSd; Fig. 4f). Using either skill metric shows that the predictability of the first month of the EWJ increases with model development. All models show little or no skill after November. Generally the widths of bootstrapped CIs are smaller for EC-DEM skill metrics than the other models because of the larger sample (Table 1, fourth column). If both the correlation skill and the RPSSd skill are not significantly different from zero, there is no predictive skill and the ensemble members are statistically independent (Weigel et al. 2008b). Therefore it is justified that these data are used as a pseudoclimate once the first month of each forecast is removed (see Table 1).

b. Windstorm selection and extreme value analysis

The EWJ time series are the basis of our storm selection procedure. From these series it is possible to identify and quantify the intensity of many of the well-documented windstorms that have affected this area over the ERA-40 period (Della-Marta et al. 2009). We use the peaks over threshold approach (e.g., Davison and Smith 1990). Their distribution is modeled with a generalized Pareto distribution (GPD) whose parameters are fitted using either maximum likelihood (ML) or L-moments (LM; Hosking 1990), and we assume a Poisson occurrence for the events exceeding the threshold (see also Palutikof et al. 1999 and Della-Marta et al. 2009). The independent peaks over threshold series is obtained using a standard runs declustering (Coles 2001; Stephenson and Gilleland 2006) where the minimum separation between exceedances of the threshold is set to 3 days. Threshold selection and model evaluation follow the typical graphical diagnostic methods outlined in Coles (2001) and an objective goodness-of-fit (GOF) test based on the Anderson–Darling statistic detailed in Choulakian and Stephens (2001). The declustered peaks over threshold series (DPOT) of each EWJ represent the occurrence and intensity of windstorms over the western European domain.

To introduce notation that will be used throughout the paper we rewrite the relevant formulae from Coles (2001) and Palutikof et al. (1999). The GPD can be written in terms of a generic variable w as

$$G(w) = 1 - \left[1 + \frac{\xi}{\sigma}(w - u) \right]^{-1/\xi}, \quad (1)$$

conditional on $w > u$ and $\xi \neq 0$ where u is the selected threshold. The GPD is characterized by two parameters:

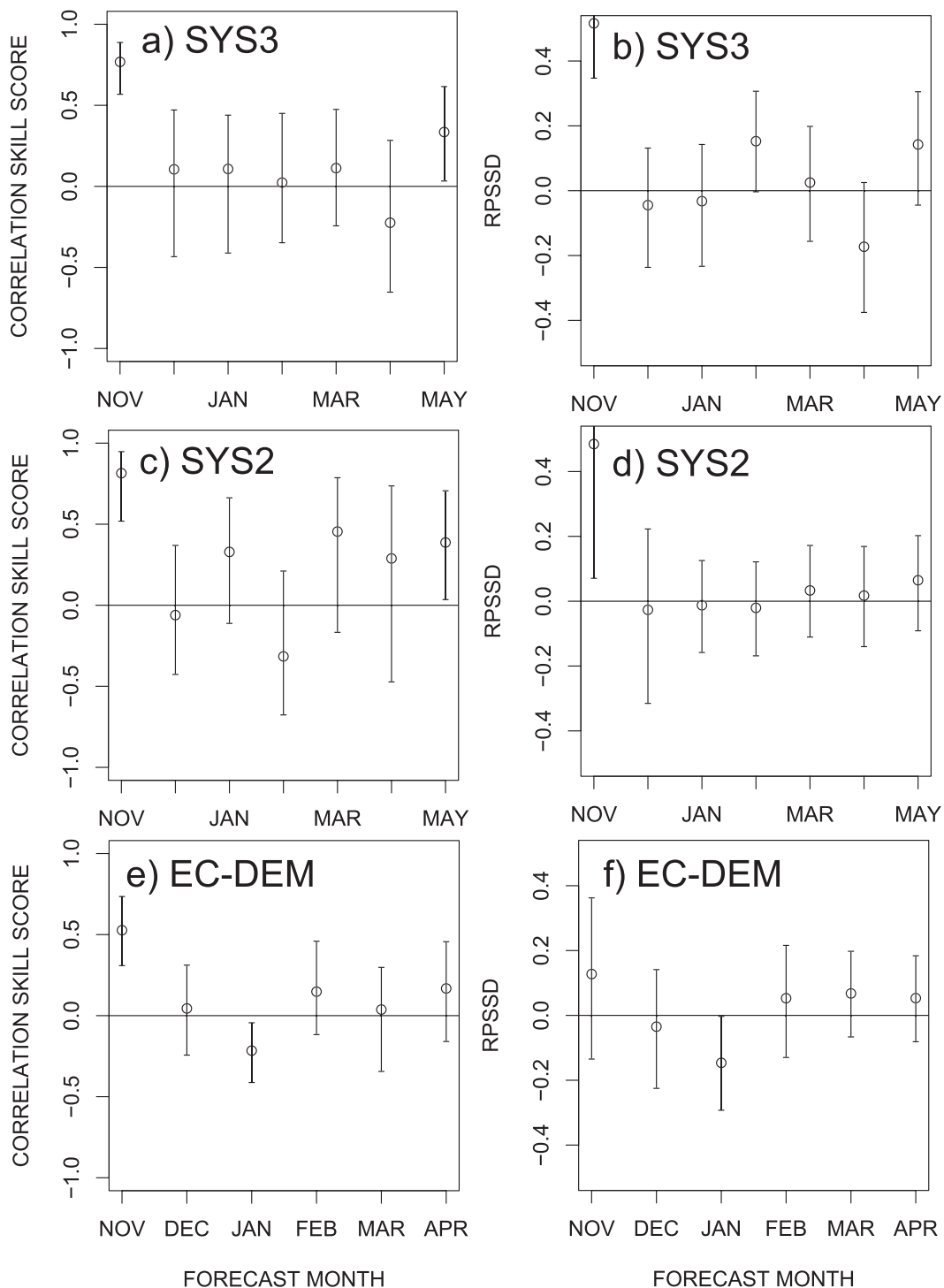


FIG. 4. Prediction skill for winter storminess in ECMWF s2d forecast data using the monthly average EWI (Q95) over the eastern North Atlantic and western Europe. (a),(c),(e) The deterministic skill measured by the correlation between the monthly mean ensemble mean Q95 of the labeled s2d hindcasts and ERA-40. (b),(d),(f) The probabilistic skill of Q95 terciles assessed in terms of the RPSSd skill score (Weigel et al. 2007). The number of hindcast years and the number of ensemble members for each of the hindcasts are given in Table 1. The error bars on all plots represent the CI based on a bootstrapping technique; see section 3a.

ξ the shape parameter and σ the scale parameter. If $\xi > 0$ then the maximum of the GPD is unbounded, whereas if $\xi < 0$ the tail has a finite extent, if $\xi = 0$ then the GPD reduces to the exponential distribution and is unbounded in the limit $\xi \rightarrow 0$.

Equation (1) can be rewritten in terms of probabilities, which leads to the calculation of the N -year return level (RL) w_N , which is exceeded on average once every N years (the RP) and is given by

$$w_N = u + \frac{\sigma}{\xi} [1 - (\lambda N)^{-\xi}], \quad (2)$$

where λ is the mean number of threshold exceedances per extended winter season. When the ML method is used we estimate the GPD fit uncertainty using the profile log-likelihood method recommended by Coles (2001) where accurate information about the uncertainty of derived parameters (such as the RP) are needed. Where the LM method is used we adopt a parametric resampling-based approach to estimating the uncertainty (Frei et al. 2006; Della-Marta et al. 2009). The declustering method behaves differently for different temporal resolutions of the EWI time series. For example, the EC-DEM dataset has a temporal resolution 24 h, whereas ERA-40 has a time resolution of 6 h. Since the difference in time resolution is of the same order as a typical windstorm development and translation time across the domain, it is probable that a certain amount of aliasing of the windstorm signal is present in the EC-DEM EWI relative to ERA-40 EWI (see Della-Marta et al. 2009). To compare windstorm climatologies the time resolution of ERA-40 EWI is degraded to match that of the s2d dataset with which we compare it (Table 1); we refer to this as aliasing.

c. Calibration

There are two reasons why we use calibration techniques in this study: 1) each s2d model has circulation biases with respect to ERA-40 (Fig. 3) that are also present in s2d extreme wind climatologies (after calibration the s2d extreme wind climatologies are referred to as $C_{s2d_to_ERA}$); 2) to derive meaningful loss estimates comparable to real market losses. CatXos is calibrated for use with SR Hist, therefore ERA-40 and s2d windstorm fields must be calibrated to this dataset (referred to as $C_{ERA_to_SR}$, and $C_{s2d_to_SR}$, respectively).

We apply the techniques to data of the form $w(t)$, which represents the scalar extreme wind index, for example, $Q95(t)$ ($C_{s2d_to_ERA}$), or data in the form of a wind field, $w(x, y, t_s) = \max\{w(x, y, t): t \in \{t_s, \dots, t_e\}\}$,

where t_s and t_e denote the start and end times of the windstorm, respectively. Note that $C_{s2d_to_ERA}$, $C_{ERA_to_SR}$, and $C_{s2d_to_SR}$ are performed on a wind field, where all grid points' wind speeds are pooled together. $C_{s2d_to_ERA}$ is also applied to EWI and uses all values of the time series. For each of the calibration steps, $C_{s2d_to_ERA}$, $C_{ERA_to_SR}$, and $C_{s2d_to_SR}$, there are three different methods used to perform the calibration. The percentile-based calibration approach (PERC), whether it is applied to EWI or wind fields, is given by the general form

$$\begin{aligned} C_{s2d_to_RA} &= S(F_{C_{s2d_to_ERA}}^{-1}) = F_{ERA}^{-1}(p) - F_{s2d}^{-1}(p) \\ C_{ERA_to_SR} &= S(F_{C_{ERA_to_SR}}^{-1}) = F_{SR}^{-1}(p) - F_{ERA}^{-1}(p) \\ C_{s2d_to_R} &= S(F_{C_{s2d_to_ERA}}^{-1}) + S(F_{C_{ERA_to_SR}}^{-1}), \end{aligned} \quad (3)$$

where p are probabilities at which the quantile functions are evaluated and $S(\cdot)$ represents a cubic spline function. In the case of calibrating wind fields, the quantile function is calculated empirically from $F^{-1}(p) = \min\{w: p \leq F_*(w)\}$, where $F_*(w)$ is the (latitude weighted) cumulative distribution function where $p \in \{0.01, 0.02, \dots, 0.99, 0.991, 0.992, \dots, 0.999\}$. In the case of calibrating EWI each quantile function is approximated using a piecewise approach. The quantile function for index values below the chosen GPD threshold is calculated empirically, using a sequence of $p \in \{0.01, 0.02, \dots, u\}$, whereas the quantile function above the threshold is taken from fitting a GPD of the form in Eq. (1) to the DPOT EWI values using the same methodology given in section 3b. Essentially, the distribution functions of each data source are matched (Della-Marta and Wanner 2006); however, the added benefit is that we use an extreme value distribution to model the extremely high EWI values, reducing the possibility of sampling errors and allowing the estimation of extreme quantiles for which there exist few or no observations. Therefore this approach also removes the possibility for a different windstorm climate (different from ERA-40) to be obtained.

The other two calibration methods are special cases of the PERC method detailed above and adjust all data based on the difference in one quantile in the distribution. For example, choosing p in Eq. (3) with a single number such as 0.95 calibrates the differences in the 95th percentile from each set dataset (95PERC). The p could also represent the probability which corresponds to the mean of the data (MEAN). Such approaches allow varying magnitude differences in the climatologies to be maintained.

CALIBRATION SAMPLING ERROR ESTIMATION

Each of the calibration curves [Eq. (3)] is subject to sampling error. We assess the uncertainty using a bootstrap technique (Efron and Gong 1983) combined with kernel smoothers (to improve efficiency). For $C_{s2d_to_SR}$ the technique consists of resampling with replacement the windstorm fields $w(x, y, t_s)$ 200 times. All 200 storm fields are combined to a single vector before the probability density function is estimated using a kernel smoother. We use a Gaussian kernel (R Development Core Team 2005), which gives good representations of the densities without over smoothing (not shown).

d. Sampling error estimation

Given that the ERA-40 dataset provides only 45 seasons compared with around 300 seasons from the s2d data (Table 1, column 6), we test to what extent the uncertainty in the ERA-40 storm RP estimates is due to sampling uncertainty using two different approaches. The first approach calculates the rate of convergence of GPD parameters as a function of the number of seasons used to fit the GPD (similar to Vannitsem 2007). For example, a random ordered sample of the total number of seasons from the DPOT series is taken (approximating chaotic interannual variability). Then a GPD is fit successively to a gradually increasing number of seasons. This is repeated 1000 times (resampling with replacement) to visualize the rate of convergence to the s2d climatology. We repeat this general procedure using pseudorandom numbers conditioned by the GPD parameters and the threshold exceedance frequency λ fitted to each EW. The random draws from the fitted EW GPD represent scalar metrics of windstorm fields and allow us to test hypotheses concerning possible biases in the ML or LM method of fitting the GPD. We compare the magnitude of possible biases in parameters to the magnitude of the interquartile range (IQR) according to $R = 100.0 \|\text{bias}\|/\text{IQR}$ to determine their relative importance to sampling variability.

The second approach focuses on 45-season samples and two different types of uncertainty: (i) sampling uncertainty, assessed by taking random 45-season samples of the DPOT series from s2d and looking at the range of fitted GPDs, and (ii) fit uncertainty, the uncertainty associated with fitting the GPD to each of the 45-season samples. These two uncertainties are not mutually exclusive, however; the sampling uncertainty can be thought of as dynamical uncertainty associated with having only 45 seasons of data, whereas the fit uncertainty can be thought of as the derived statistical uncertainty associated with having only 45 seasons of data. The sampling uncertainty calculation consists of taking 1000 different

sets of 45 seasons of the DPOT series (resampling with replacement). The fit uncertainty calculation uses the same samples as above; however, we fit a GPD to each and calculate the median of the differences between the fit and the upper and lower bounds of the CI. The median fit uncertainty series are then added back to the best-fit ERA-40 GPD.

e. Bias correction of the estimated climatologies

Both ML and LM are known to result in biased parameter estimates when used on limited samples and for various ranges of possible parameter settings; generally the biases being highest in magnitude for small samples (e.g., Cox and Snell 1968; Pandey et al. 2003). These biases can be corrected using resampling and bootstrap techniques (e.g., Zhang and Stephens 2009; Pandey et al. 2003) or directly using analytical techniques (Cox and Snell 1968). Using a general method to correct statistical model parameter biases (Cox and Snell 1968), Giles and Feng (2009) derive a first-order analytical bias correction formula using the log-likelihood equation of the GPD and show that the analytical approach is more efficient than a bootstrap approach. However, at the time of writing their correction scheme was not entirely applicable to our ML-fitted GPD parameters since many of our GPDs have a negative shape parameter where their scheme does not work efficiently. Therefore we have simply corrected the shape and scale parameters by bias estimated using the dynamical–statistical resampling technique.

4. Results

The section starts with a summary of the storm selection process and extreme wind climatologies, followed by the main results, the comparison of estimates of the European windstorm climate, and the associated losses to a Swiss Re portfolio.

a. Windstorm identification, extreme value analysis, and climatology

Figure 5 summarizes the storm selection process. The first step 1) calculate the EW series, 2) decluster the peaks over threshold series (Fig. 5a), 3) fit the GPD (Fig. 5b), and 4) assess the quality of the fit using various diagnostics such as a quantile–quantile (qq) plot (Fig. 5c) combined with objective GOF tests (Fig. 5d). Figure 5a shows the effectiveness of the Sw3q90 in measuring the difference in magnitude of extreme winds and the effectiveness of the declustering method in identifying major windstorm events. Well-known storms such as Daria, which occurred on 24 January 1990, are identified

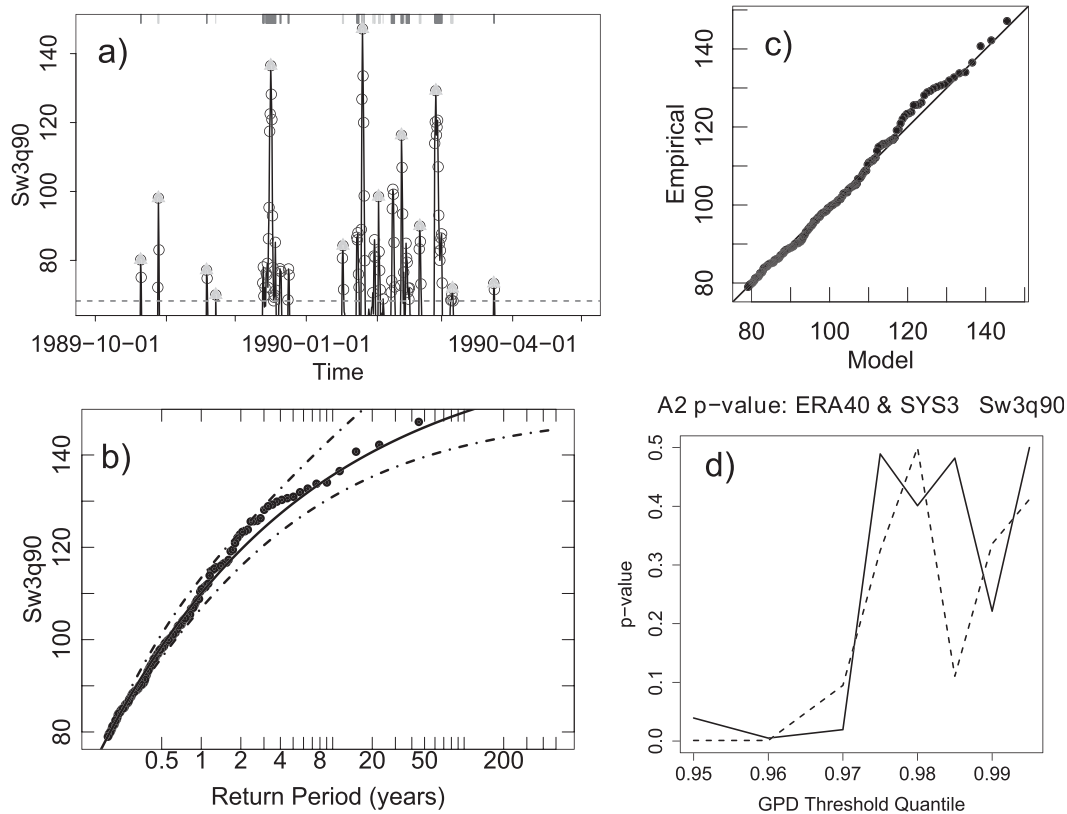


FIG. 5. A summary of the windstorm selection process. (a) The Sw3q90 DPOI for the 1989/90 extended winter season. The thin black line is the Sw3q90 index. The circles indicate values of the index that exceed the 97.5th percentile threshold (dashed black line). Gray filled triangles show the maximum value of the index within each cluster. Membership of DPOIs (circles) to a particular cluster is denoted by alternating light and dark gray bands on the top margin of the plot. (b) The RP (yr) and RL of the GPD fit to the DPOI series [black line, Eq. (2)]. The dash-dotted lines show the upper and lower bounds of the CI of both RL and RP calculated using profile log-likelihood. (c) Plot of qq of the fitted GPD to ERA-40. (d) GOF p value for ERA-40 (solid) and SYS3 (dashed). The black dots in (b) and (c) represent the maxima of the DPOI series that correspond to the gray triangles in (a).

as a strong peak in the Sw3q90 series (seventh gray triangle from the left of Fig. 5a), whereas the storms Vivian and Wiebke, which were separated by only a short amount of time (25 and 27 February 1990, respectively) are identified as one cluster maxima in the DPOI series. The latter is a limitation of the scalar index approach to separate storms that occur close in time; however, the statistics of GPD fit do not seem to be overly affected as discussed in Della-Marta et al. (2009). The GPD fit (Fig. 5b) to ERA-40 Sw3q90 is bounded, that is, a negative shape parameter. Other indices generally display less negative shape parameters; however, all shape parameters are less than zero and remain so for all tested thresholds (see below). The qq plot (Fig. 5c) indicates that the GPD fit is reasonable apart from a slight departure from the GPD distribution in the middle tail. The GOF test (Choulakian and Stephens 2001) suggests that the chosen GPD threshold (97.5% quantile) results in a good GPD fit (Fig. 5d, solid black line).

According to Choulakian and Stephens (2001), p values greater than 0.1 indicate a sufficient GOF to the GPD distribution. For comparison, GOF results for the SYS3 Sw3q90 climatology are shown as a dashed line in Fig. 5d. The 97.5% threshold was also chosen for this data-index combination because of the jump in GOF at the 97.5% quantile. Not all data-index combinations resulted in the same threshold being chosen for each dataset (see Table 2), and sometimes the GOF statistic did not increase in a general monotonic way with increasing threshold, which would be expected if the index data were asymptotically Pareto distributed. We tested the sensitivity of our results for each data, EWI, fitting method, and threshold combination (using the 95%, 96%, 97%, 97.5%, 98%, 99%, and 99.5% quantiles). The LM method results in generally higher GOF p values than the ML method, indicating a better GOF using LM. Sf95q99 scores consistently high GOF p values across all thresholds and all datasets, indicating

TABLE 2. The threshold above which a GPD was fitted to the DPOT of each EWI (percent quantile) for each dataset; in parentheses is the number of windstorms identified above the threshold.

EWI	ERA-40	SYS3	SYS2	EC-DEM
Mean	95.0 (263)	95.0 (1661)	98.0 (846)	97.0 (1334)
Q95	95.0 (380)	95.0 (2331)	98.0 (1258)	99.5 (367)
Sw3q90	97.5 (232)	97.5 (1516)	97.0 (1621)	97.5 (1475)
Sfq95	95.0 (332)	95.0 (2042)	95.0 (1995)	95.0 (2230)
Sfq95q99	95.0 (361)	95.0 (2240)	95.0 (2199)	95.0 (2363)

that the normalizing effect of this index makes it a stable choice for comparisons between datasets. The three magnitude-based indices—Mean, Q95, and Sw3q90—all display GOF at various thresholds (Table 2), which are generally higher for SYS2 and EC-DEM, Sw3q90 being the exception where the threshold for all datasets was set close to the 97.5% quantile. The index Sfq95 displays high GOF fitted to ERA-40 but low GOF to SYS2 and SYS3. Depending on the EWI, dataset, and threshold, the storm selection method finds between 263 and 380 windstorms in ERA-40 and between 367 and 2363 storms in the s2d data (Table 2).

A comparison of the ERA-40 windstorm climatology with the s2d windstorm climatologies reveals large differences in the EWI absolute values (Fig. 6) and their corresponding model parameter values (Tables 2 and 3) for the Mean, Q95, and Sw3q90 EWI. The climatologies using EWI based on local wind climatologies (Sfq95 and Sfq95q99) are quite similar in magnitude because of the normalizing effect inherit in these indices. Differences in the mean circulation are reflected in the overall differences between the climatologies (see Fig. 3), with SYS3 having higher intensity values than ERA-40 in all intensity-based EWI. SYS2 and EC-DEM are lower in wind intensity than ERA-40 but are similar to each other, which is expected since these models have similar physics and resolution (Figs. 6a–f). The widths of the CIs for each of the s2d models are much smaller than those of ERA-40, and there is clearly a less negative shape parameter in the climatologies of the s2d models using intensity-based EWI relative to ERA-40 (Table 3). As stated above, the LM GPD fits have higher GOF. The GOF test is more sensitive to lack of fit in the extreme tail of the GPD than it is to the lower tail (Choulakian and Stephens 2001). The high GOF with LM is somewhat contrary to the observed quality of the fits in the extreme tails. Comparing Figs. 6b,d,f with Figs. 6a,c,d, the LM-fitted GPD does not seem to fit the data as well as it does using the ML method in the range of approximately 20–300 yr. The ML fits seems to take more of a “middle path” through the observations, whereas the LM fits seem to be conditioned more by the lower RP EWI observations. It could be that the sensitivity of the

GOF test diminishes with a high number of samples, such as used here.

Other differences (Fig. 6) lie in the position along the x axis of the first DPOT (windstorms, colored dots). The varying position indicates 1) that in some cases a different threshold was chosen (based on the GOF test; see Table 2) or 2) that the mean number of storm occurrences per season λ is different for each dataset (Table 2). Histograms in Fig. 7 show the empirical and theoretical distribution of the number of windstorms identified in each season using Sw3q90 choosing a 97.5% quantile threshold for all datasets. All distributions show some underdispersion. This has implications for the clustering (in time) of storm events (Mailier et al. 2006).

The reasons for the systematic differences in both the atmospheric circulation at 850 hPa and extreme wind climatologies are not fully known, but certainly some of the differences are due to model resolution (Jung et al. 2006). SYS2 and EC-DEM have a horizontal resolution of T95, whereas SYS3 and ERA-40 have a resolution of T159; this may explain why both ERA-40 and SYS3 have higher wind speeds (cf. Fig. 6). Model physics are also likely to play a role in the differences shown here; SYS2, EC-DEM, and ERA-40 all have similar physics (ECMWF 2003), whereas SYS3 has substantially revised physics (ECMWF 2007). Overall, these results indicate that a direct comparison of windstorm climatologies is not possible and that a calibration technique must be used.

b. Estimates of windstorm return periods

1) CALIBRATION OF S2D AND ERA-40 WINDSTORM CLIMATOLOGIES

The three different calibration techniques, PERC, 95PERC, and MEAN, were applied to calibrate each s2d EWI series to the corresponding ERA-40 climatology, obtaining $C_{s2d_to_ERA}$. Figure 8 shows the calibration curves using the PERC method and the Sw3q90 series. The PERC curve for SYS3 Sw3q90 suggests increasingly negative adjustments to SYS3 with increasing Sw3q90 magnitude to be compatible with ERA-40 (0–80 nondimensional units). Both SYS2 and EC-DEM display mostly positive adjustments (2–5 nondimensional units) over most of the distribution; however, at higher percentiles (above the 95% quantile) the adjustments change quickly from being positive to negative. This occurs since the EWI climatologies of both SYS2 and EC-DEM cross ERA-40 at around 145 nondimensional units (Fig. 6e). Note in Figs. 8c,f there are minor discontinuities at the point where the calibration technique changes from the empirical method to the GPD method. While this is not desirable, it does not affect the subsequent

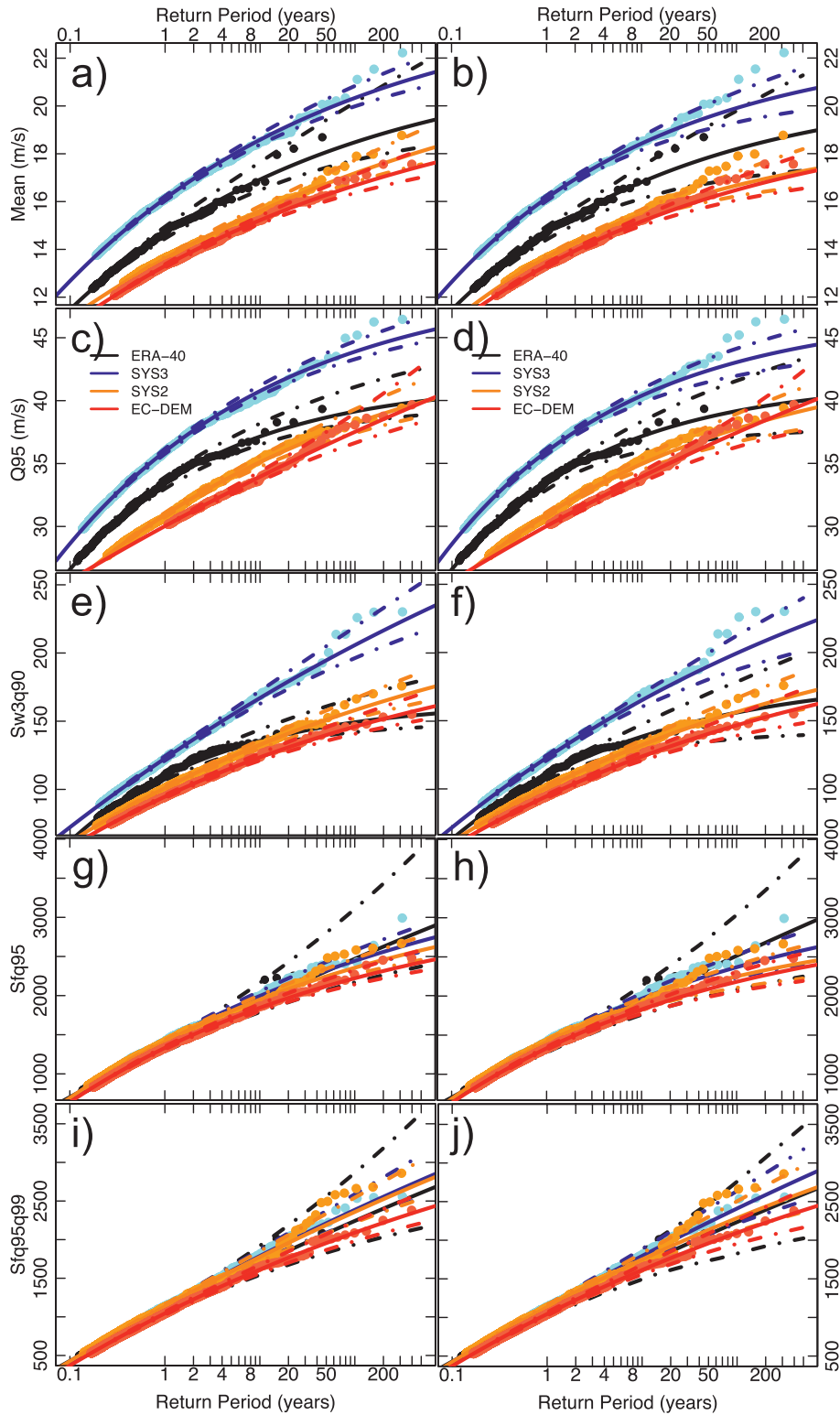


FIG. 6. The RP and RL of the GPD fit [solid colored lines; Eq. (2)] to each DPOT EWI series. GPD are fitted (a),(c),(e),(g),(i) using ML and (b),(d),(f),(h),(j) using LM. The name of each index is on the left-hand side of each row. Datasets are colored according to the legend in (c) and (d). The solid filled colored dots represent the DPOT series (see example in Fig. 5a). Dash-dotted colored lines show the upper and lower bounds of the CI of both RL and RP calculated using profile log-likelihood (parametric resampling) method for ML (LM), respectively.

TABLE 3. A comparison of the GPD parameters to noncalibrated DPOT Sw3q90 EWI for each dataset. The columns titled lower and upper represent the lower and upper bounds of the CI calculated from the likelihood profile. The parameters λ , σ , and ξ are from Eqs. (1) and (2).

Dataset	Lower	λ	Upper
ERA-40	4.5	5.2	5.8
SYS3	4.6	4.8	5.0
SYS2	5.0	5.2	5.5
EC-DEM	3.5	3.7	3.9

	Lower	σ	Upper
ERA-40	20	23	28
SYS3	21	22	24
SYS2	17	18	19
EC-DEM	15	16	17

	Lower	ξ	Upper
ERA-40	-0.38	-0.27	-0.13
SYS3	-0.10	-0.06	-0.01
SYS2	-0.14	-0.10	-0.05
EC-DEM	-0.15	-0.11	-0.06

analysis since the storm climatologies are based on the extreme tail of the distribution.

Applying the calibration curves and then going through the storm selection and GPD fitting procedure again results in calibrated climatologies. Figure 9 contrasts the various EVA climatologies of Sw3q90. Figures 9a,c,e compare SYS3 and SYS2 with 12-hourly aliased ERA-40 for each of the calibration techniques: PERC, 95PERC, and MEAN, respectively. Figures 9b,d,f compare EC-DEM with 24-hourly aliased ERA-40 for each of the calibration techniques, as for Figs. 9a,c,e, respectively. The PERC technique results in SYS2 and SYS3 and EC-DEM windstorm climatologies being almost exactly the same (Figs. 9a,b) as the aliased ERA-40 climate (see Table 4). The 24-hourly aliased ERA-40 climate (Fig. 9b) is very different from the 12-hourly aliased ERA-40 climate (Fig. 9a). The 24-hourly aliasing of ERA-40 has a high impact on the shape of the fit, resulting in a much shorter tail (Fig. 9b, black line, compared to Fig. 9a, black line). The aliasing of ERA-40 to 24-hourly resolution has resulted in similar windstorms being selected, that is, major storms such as Daria are identified; however, the maximum intensity of each storm in the aliased series is sometimes lower. The 95PERC calibration allows greater differences in the windstorm climatologies to remain. Both SYS3 and SYS2 (Fig. 9c) indicate that the RP of storms in ERA-40 are underestimated. The best-fit SYS2 curves lies within the CI of ERA-40 12-hourly aliased series, whereas the SYS3 estimate lies outside of the CI. Again we see the effects of

the aliasing and calibration technique in Fig. 9d. Given these results we will not consider climatologies based on EC-DEM any further in this study. Calibrating using the MEAN technique results in an even more extreme windstorm climate based on SYS3 and SYS2 (Figs. 9e,f) than for the 95PERC calibration compared to ERA-40, and it is likely that this type of calibration does not result in a meaningful comparison.

2) SAMPLING AND FIT UNCERTAINTY ESTIMATES

We chose to investigate SYS3 PERC-calibrated results in more detail since the model resolution is the same as ERA-40 [shown in Jung et al. (2006) to have a high impact on extratropical cyclone numbers and in Weisheimer et al. (2003) to have an influence on large-scale circulation modes]. However, we could have equally decided to use SYS2 results since SYS2 and ERA-40 have similar physics and SYS3 physics display different biases to SYS2 and EC-DEM (Fig. 3).

In Fig. 9a the windstorm climatology based on SYS3 is almost exactly the same as ERA-40. Effectively the PERC method has removed any differences between the SYS3 EWI climatologies and the ERA-40 climatology. Therefore, at this stage we are no closer to assessing whether the ERA-40 climate is close to the true climate; that is, we cannot answer the question whether the ERA-40 is a representative, unbiased sample of the true winter windstorm climate. One way to help answer this question is to utilize the longer s2d data by performing sampling experiments. Figures 10a,b show the bias and rate of convergence of the scale and shape parameter of the ML-fitted Sw3q90 SYS3 GPD as a function of the number of seasons (of DPOT data) used in the fit. Neither the scale nor the shape parameter has converged for a 45-season climatology, and they are different from the long-term parameter estimates. The magnitude of the biases represent approximately +16% and -22% of the IQR of samples taken using 45 season lengths for the scale and shape parameters, respectively. Approximately 65% of the shape and scale parameters lie below and above the long-term (316 seasons) SYS3 climatology, respectively (similar results are found for SYS2). A more negative shape parameter combined with a higher scale parameter than the converged values implies a shorter-tailed climatology when only 45 seasons are used. Figures 10c-f reveal that there is also a lack of convergence of the RPs [Eqs. (1) and (2)] for 45-season samples. For example, Figs. 10c-e all show that ERA-40 length datasets are likely to overestimate the RP of 30-, 45-, and 100-yr RP events by approximately 7, 11, and 21 yr, respectively. For RPs greater than approximately 300 yr ERA-40 length datasets tend to underestimate RP (Fig. 10f). Note, however, that the spread of possible

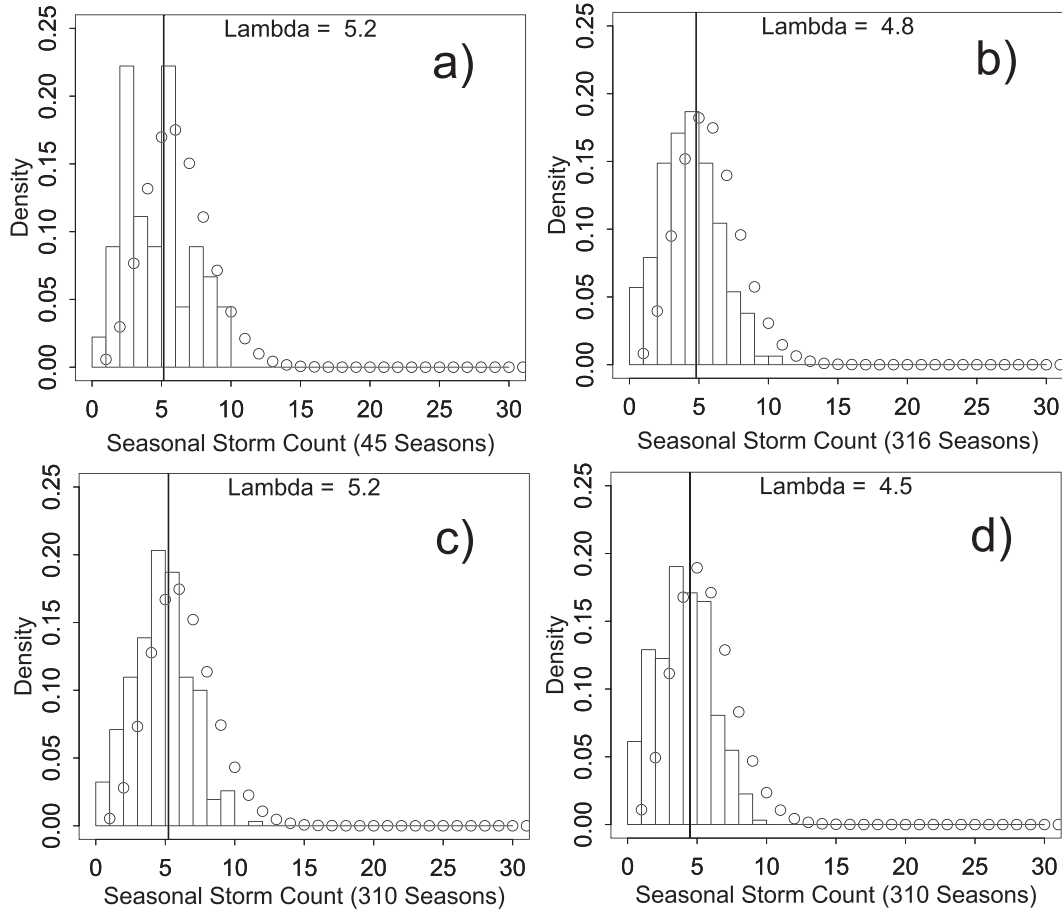


FIG. 7. Histograms of the number of Sw3q90-based windstorm occurrences per extended winter together with the ML-fitted Poisson probability density function (open circles) for (a) ERA-40, (b) SYS3, (c) SYS2, and (d) EC-DEM. At the bottom of each panel is the number of seasons in each dataset. The solid black vertical lines denote λ [in Eq. (2)]. See Table 1 for the number of windstorms identified in each dataset and Table 3 for error estimates of the λ parameter.

RP estimates at 45-season lengths is large (IQR approximately 30–120 yr) compared to the apparent bias of approximately 11 yr. Therefore the issue of bias for 45-season length storm climatologies is secondary in importance to sampling variability. Figure 11 shows the same analyses as in Fig. 10 but using LM method of fitting the GPD. The LM-fitted scale and shape parameters are less obviously biased (Figs. 11a,b) than their ML counterparts. The small biases in these parameters result in small negative biases in the 30- and 45-yr RP around 1.1 and 2.1 yr, respectively. At higher RPs the negative biases become proportionally larger in magnitude than for the lower RPs, with the biases at 100- and 500-yr RP being 14 and 220 yr, respectively. Investigation of these biases using random storms simulated from the fitted ML and LM GPDs are shown in Figs. 12 and 13. These figures are very similar to Figs. 10 and 11, respectively. Almost all of the biases we see in the scale,

shape, and RP estimates are due entirely to biases inherent to the ML and LM methods of fitting the GPD. Similar calculations reveal that these biases are common to all other EWI, although the magnitude of the RP biases using ML with the index Sf_{q95q99} are much lower than the RP biases in the magnitude-based EWI. This seems to be a function of the magnitude of the shape parameter: the more negative the shape parameter is, the larger is the ML bias (Zhang and Stephens 2009).

Assuming that the bias can be explained by the GPD fitting method we have applied a bias correction to the ERA-40 estimated ξ and σ of the form $(\tilde{\xi}, \tilde{\sigma})' = (\xi, \sigma)' - B'$, where $B = \text{Bias}(\xi, \sigma)$ and is measured from the difference between the converged SYS3 EWI GPD parameter value and the median SYS3 EWI GPD parameter value at 45 seasons (e.g., shown as text in Figs. 10a,b and Figs. 11a,b). Applying the bias correction we obtain new estimates of the European windstorm

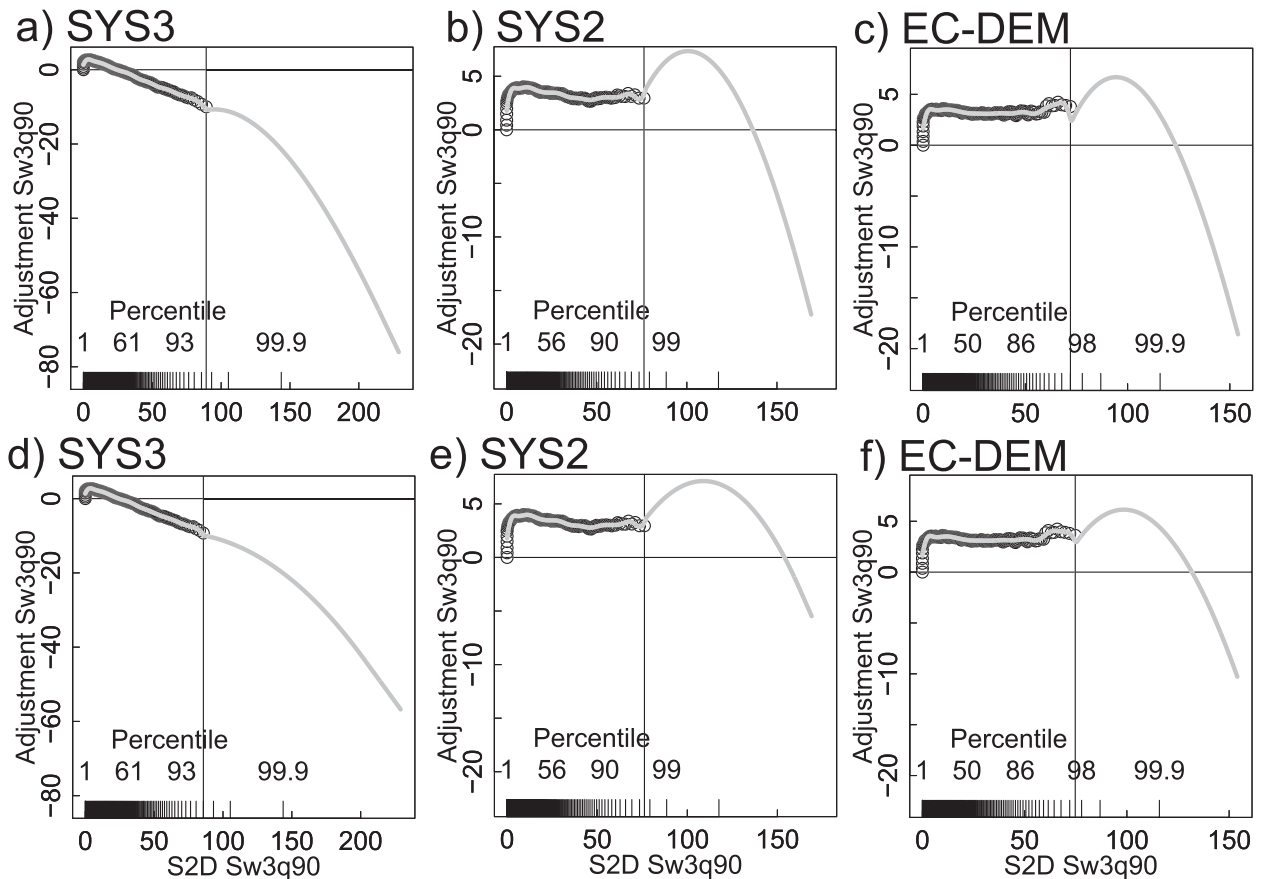


FIG. 8. The calibration curves (gray lines) applied to the Sw3q90 time series of each s2d model (a)–(c) using ML and (d)–(e) using LM. These gray lines represent the cubic spline fitted values to the differences in the empirical cumulative distribution function [black open circles; Eq. (3)]. The x axes show the Sw3q90 and the Sw3q90 empirical percentiles (vertical black lines).

climate as defined by each EWI and each fitting method presented in Table 5. All ML-fitted EWI windstorms have lower RPs after bias correction (Table 5, column 6). The magnitude of the bias in RP is greatest for the Mean EWI with a 50-yr event bias corrected becoming a 37.5-yr event, whereas the magnitude of the bias in RP is lowest for the Sf_{q95q99} index showing only a small reduction in a 50-yr event to a 47.6-yr event. According to the results of Zhang and Stephens (2009) the magnitude of the bias correction to the shape parameter in their experiments increases for decreasing values of ξ . This may explain the results since the shape parameter of the Sf_{q95q99} is less negative (-0.03) than, say, the shape parameter of Sw3q90 (-0.28). All LM-fitted EWI windstorms have a higher RP after bias correction (Table 5, column 12) except the Sw3q90 index, which results in slightly negative corrections. As for the ML method the Mean index displays the largest change in RP: a 50-yr event becomes a 55.5-yr event after LM bias correction. All other LM RP differences are small. This indicates that the use of

LM results in less biased RP estimates of the European windstorm climate. Possible reasons for Sw3q90 displaying slightly negative RP adjustments using the LM method (Table 5), when results presented in Figs. 11c–f indicate that positive RP adjustments may be needed, especially for higher RPs, is believed to be a function of the heteroscedasticity of the relationship between the dynamically resampled scale and shape parameters, which increases for more negative shape parameters. This has a nonlinear effect on the RP calculation and leads to the observed bias in Figs. 11c–f. This information is lost when only the median bias in the shape and scale parameters is used to correct the RPs.

Summarizing the results: if SYS3 windstorms are representative of the parent distribution (i.e., when calibrated) then dynamical–statistical sampling implies that ERA-40, by virtue of its length only and the ML method of fitting the GPD, is probably underestimating the severity (by overestimating the RPs of storms) of the current windstorm climate for windstorms with RPs

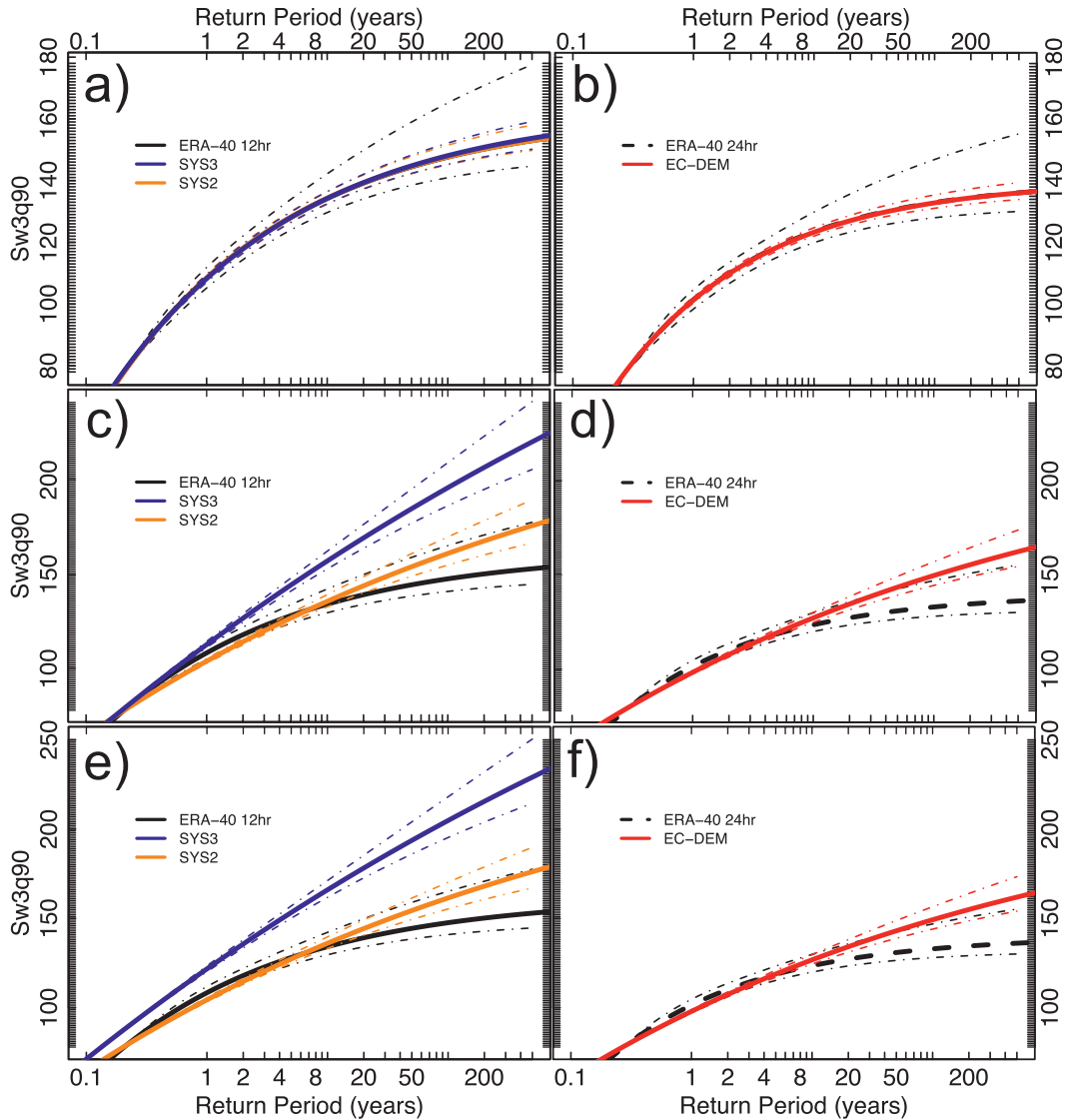


FIG. 9. An intercomparison of the effect of three different calibration techniques and aliasing on the RP and RL of the ML-fitted GPD to the Sw3q90 DPOT series for each s2d dataset and ERA-40. Using the (a),(b) PERC method [Eq. (3)]; (c),(d) 95PERC technique; and (e),(f) MEAN method. Each dataset is colored according to the legend in the top left of each panel. Panels (a),(c),(e) show SYS3 and SYS2 with the aliased 12-hourly ERA-40 Sw3q90, and panels (b),(d),(f) show the EC-DEM model with the aliased 24-hourly ERA-40 Sw3q90. Dot-dashed colored lines indicate the CIs of the fitted GPD.

between 10 and 300 yr and overestimating the severity of the windstorm climate (underestimating the RPs of storms) for storms with RPs longer than 300 yr. Whereas, if the ERA-40 storm climate is fitted using LM, the climatology is relatively well represented except at higher RPs where the method tends to overestimate the severity of the windstorm climate.

Figure 14 presents the sampling uncertainty and fit uncertainty (described in section 3d) of ERA-40 based on calibrated SYS3 for two EWIs, Sw3q90 and Sf95q99, contrasting the two parameter estimation methods.

Focusing first on the gray shaded regions of Figs. 14a,c, representing the dynamical sampling uncertainty, we see that the ERA-40 and SYS3 windstorm climatologies overlap each other and lie within, but always in the upper part of, the 45%–55% percentile range of randomly sampled 45-season SYS3 GPDs (Figs. 14a,c; also shown in Fig. 10d). This result confirms that ML estimates of the windstorm climatology are on average biased. Interestingly, approximately 20% of the randomly sampled 45-season SYS3 GPDs (Figs. 14a,c, medium gray shading) are lower than the ERA-40 GPD fit lower CI

TABLE 4. A comparison of the GPD parameters fitted to the PERC [Eq. (3)] DPOT Sw3q90 EWI for each s2d dataset. The columns titled lower and upper represent the lower and upper bounds of the CI calculated from the likelihood profile. The parameters λ , σ , and ξ are from Eqs. (1) and (2). ERA-40 12hr and ERA-40 24hr refer to the aliased DPOT Sw3q90.

Dataset	Lower	λ	Upper
ERA-40 12hr	4.0	4.6	5.3
SYS3	4.6	4.8	5.0
SYS2	4.3	4.5	4.8
ERA-40 24hr	3.1	3.6	4.2
EC-DEM	3.5	3.7	3.9
	Lower	σ	Upper
ERA-40 12hr	20	24	28
SYS3	22	23	25
SYS2	23	25	26
ERA-40 24hr	18	23	28
EC-DEM	22	23	25
	Lower	ξ	Upper
ERA-40 12hr	-0.39	-0.28	-0.14
SYS3	-0.31	-0.28	-0.24
SYS2	-0.33	-0.30	-0.26
ERA-40 24hr	-0.51	-0.37	-0.19
EC-DEM	-0.42	-0.38	-0.34

(black dot-dashed line). This indicates that the ML fit uncertainty (profile log-likelihood method) of ERA-40 may be underestimating the possibility for less extreme wind climatology than what is possible by drawing a random 45-season sample from the SYS3 climatology. Similarly, the upper boundary of the gray shading is much lower than the ERA-40 GPD fit upper CI (black dot-dashed line). This indicates that the ERA-40 GPD fit uncertainty overestimates the width of the upper windstorm intensity uncertainty. In other words, ERA-40 ML fit uncertainties seem to underestimate the possibility for a less intense windstorm climate and seem to overestimate the possibility for a more intense windstorm climate (these results are also true based on SYS2 PERC-calibrated data, not shown). Another conclusion of these experiments is that the median SYS3 fit uncertainty, denoted by the yellow lines in Figs. 14a,c, indicates that the ERA-40 GPD fit uncertainty (black dot-dashed lines) is a good estimate of the typical CI widths associated with the GPDs fitted to the randomly sampled 45-season SYS3 climatologies; that is, the yellow lines more or less match the edge of the medium gray shaded regions. This signifies that, given the sample of the observed climate (i.e., ERA-40), the fit uncertainty estimates are not biased once the overall bias in the GPD fit is removed. All results described above remain valid for the other ML-fitted EWI for the range of thresholds higher than

those presented in Table 2. In contrast to the ML-fitted climatologies, the 45-season LM-fitted climatologies of Sw3q90 and Sfq95q99 are not biased (Figs. 14b,d); that is, the ERA-40 and calibrated SYS3 climatologies fall in the middle of the middle 10% of 45-season sampled climatologies (light gray shaded region). However, the width of the 45-season sampled climatologies (gray shaded areas) cover a wider range of RP estimates than the ML estimates. The wider range of LM climatologies shows that the LM method is less accurate in determining the long-term windstorm climate given a 45-season sample than the ML bases climatologies, which have a lower range (Figs. 14a,c gray shaded regions compared with Figs. 14b,d shaded regions). The larger spread of LM-based climatologies is being driven by the wider range of fitted scale and shape parameters compared with ML (Figs. 10a,b compared with Figs. 11a,b). Therefore the use of ML results in more representative estimates of the long-term storm climate given 45 seasons than the LM-based climates after the ML bias has been removed. On the other hand, the LM-based climates are not biased but the spread of possible sampled climates is wider given 45 seasons of data. Similar conclusions can be drawn from the SYS3 climatology based on the 95PERC technique, yet the proposed windstorm climate is far more extreme than ERA-40, implying that the 95PERC has not adequately accounted for model differences. Therefore these 95PERC-based results (not shown) should only be seen as speculative.

The sampling experiment results give us details about how reliable the LM and ML methods are. Alternatively, the s2d climatologies can be used to simply quantify the reduction in GPD fit uncertainty gained from having a larger sample. The results are demonstrated in Fig. 15. Here the width of the ML return level (nondimensional units) and RP (yr) uncertainties (CIs) are plotted as functions of the RP estimated from ERA-40. S2d RL uncertainties are approximately one-third of the size of ERA-40 RL uncertainties (Fig. 15a). The upper bound of the return level CI is larger in magnitude than the lower bound of the return level CI, showing the asymmetry in the profile log-likelihood uncertainty estimates. Figure 15b shows the improvements in the width of the RP CI by using s2d data. For a given RP, say 20 yr, the s2d upper (lower) bound is approximately 16% (40%) of ERA-40 upper (lower) bound. Clearly, the greater length of the s2d climatologies results in markedly reduced CI widths.

c. Estimates of windstorm-related loss

Windstorm fields identified in ERA-40 and s2d datasets are used as input to catXos. The first step in using a windstorm field is to calibrate it with the Swiss Re

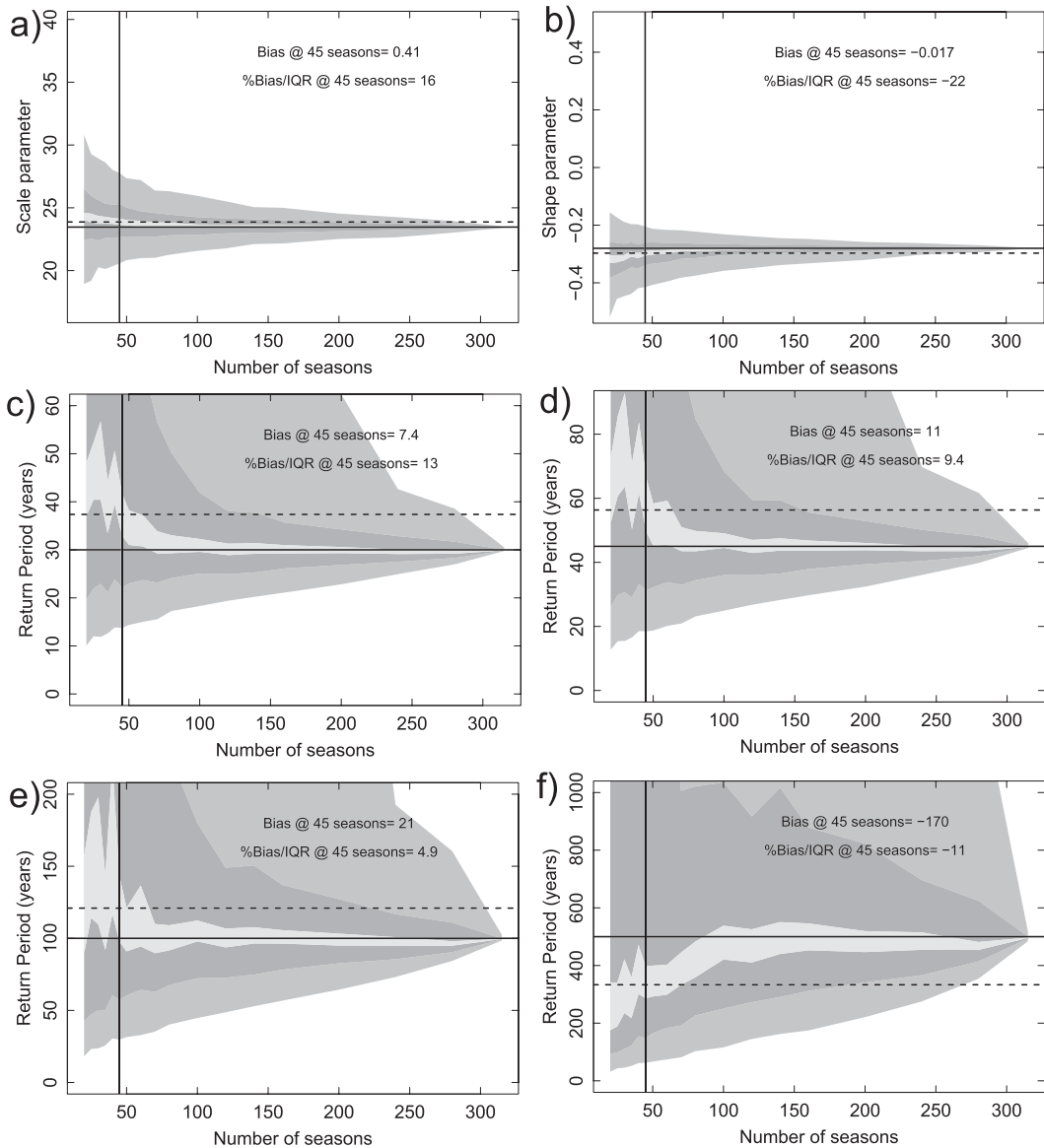


FIG. 10. The convergence of (a) σ and (b) ξ and the (c) 30-, (d) 45-, (e) 100-, and (f) 500-yr RP of the calibrated SYS3 Sw3q90 ML-fitted GPD as a function of the number of seasons taken from the 315-season integration. The light gray shading shows the 45th–55th percentile range of the resampled SYS3 dataset, whereas the dark gray and medium gray shading show the interquartile (50% confidence interval) and CI, respectively. The horizontal black lines show the converged, σ , ξ , and RP estimates using all available data. The horizontal dashed black line is the median of the resampled estimates at 45-season length. The vertical black line denotes 45 seasons. In the top right of each figure is magnitude of the bias at 45 seasons and percentage ratio of the bias to the interquartile range.

windstorm dataset (SR Hist). The calibration is described in section 3c as $C_{s2d_to_SR}$. The $C_{ERA_to_SR}$ PERC adjustment curve (Fig. 16a, thick gray line) shows that ERA-40 GW needs to be adjusted negatively, which is expected since the GW is a measure of the free atmosphere wind. Generally the negative adjustment increases with GW and displays a rather unusual behavior between ERA-40 GW values of 10–30 $m s^{-1}$. The cause of this “bump” is related to the way

in which the Swiss Re windstorm fields are created. Generally, only a limited number of pressure and wind gust speeds observations were available in the wind field reconstruction, usually concentrated in regions where the windstorm was known to have caused damage. These wind gust speeds are generally above 20 $m s^{-1}$. This has had the effect that the Swiss Re windstorm fields may be unreliable for wind gust speeds less than 20 $m s^{-1}$ in regions that were not affected by the windstorm. While this

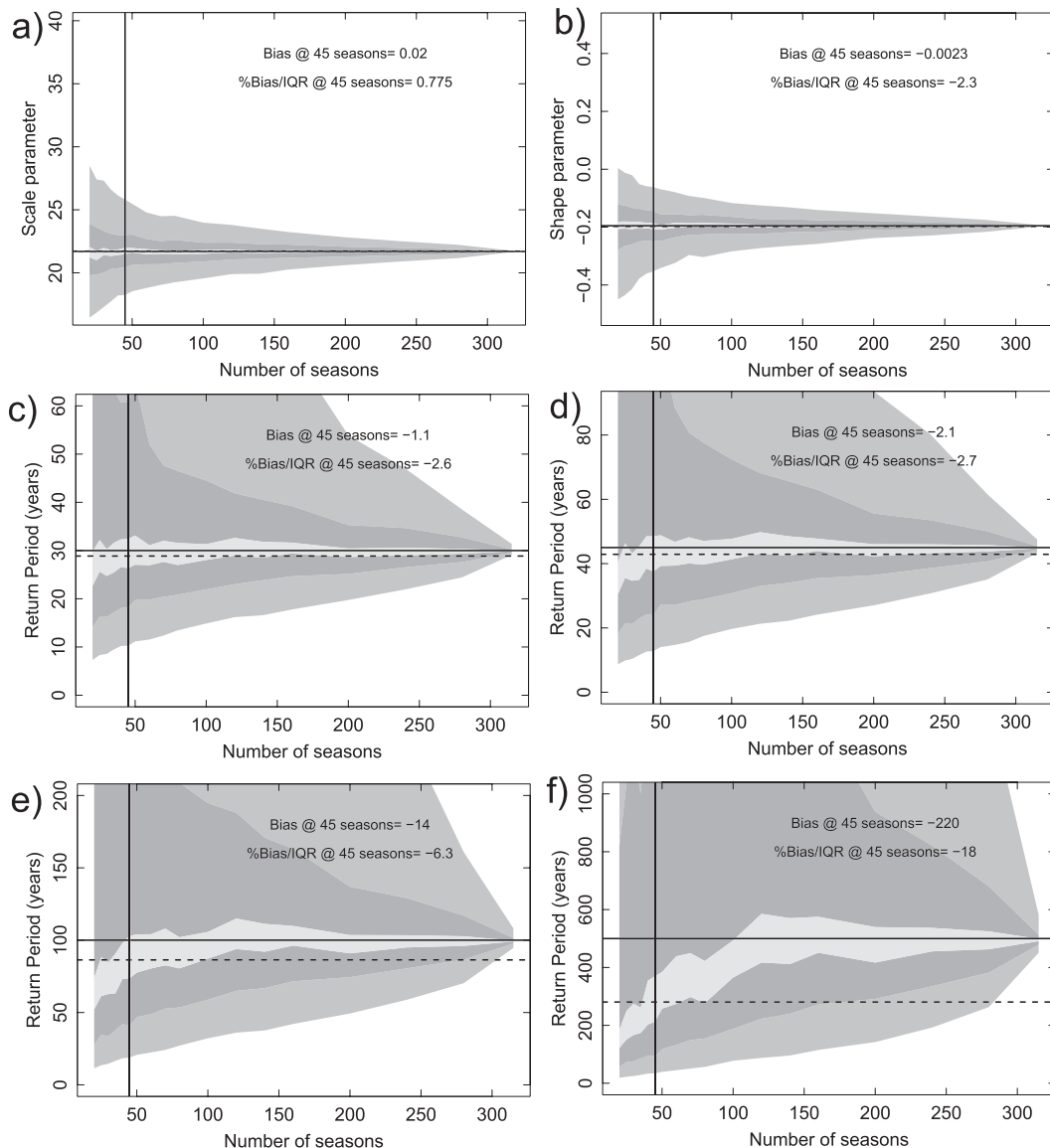


FIG. 11. As in Fig. 10, but using LM method.

has no effect on the actual loss values (see Fig. 2a), it exposes differences between the datasets. The $C_ERA_to_SR$ adjustment uncertainty (Fig. 16a, thin vertical black lines) calculated using resampled kernel density estimates, shows some nonsymmetric behavior in the adjustment amounts especially at higher percentiles. Figures 16b–d show the $C_s2d_to_ERA$ PERC adjustment curves for each of the s2d datasets. The calibration of the windstorm fields show similar adjustment curves to the calibration of the Sw3q90 (Fig. 8); however, all curves show a monotonic increasing (SYS3) or decreasing (SYS2) adjustment amount with increasing GW unlike their EWI equivalents in Figs. 8b,c,e,f for SYS2 and EC-DEM.

Figure 17 presents a comparison between loss estimates derived from s2d, ERA-40, and Swiss Re windstorm datasets. In presenting these results we wish to advise readers that the loss figures derived from ERA-40 and s2d are modeled loss estimates, which are critically dependent on methods chosen to compare the model climates and are not calibrated against loss data. In contrast, the Swiss Re loss estimates are based on and tested against independent loss data. Figure 17a compares the SR Hist and SR Stoch loss-frequency curve(s) (LFC) with those derived from ERA-40 windstorms. First, the $C_ERA_to_SR$ method brings the raw LFC of ERA-40 (black solid and dashed lines in the top left of Fig. 17a) to lie between the Swiss Re historic and

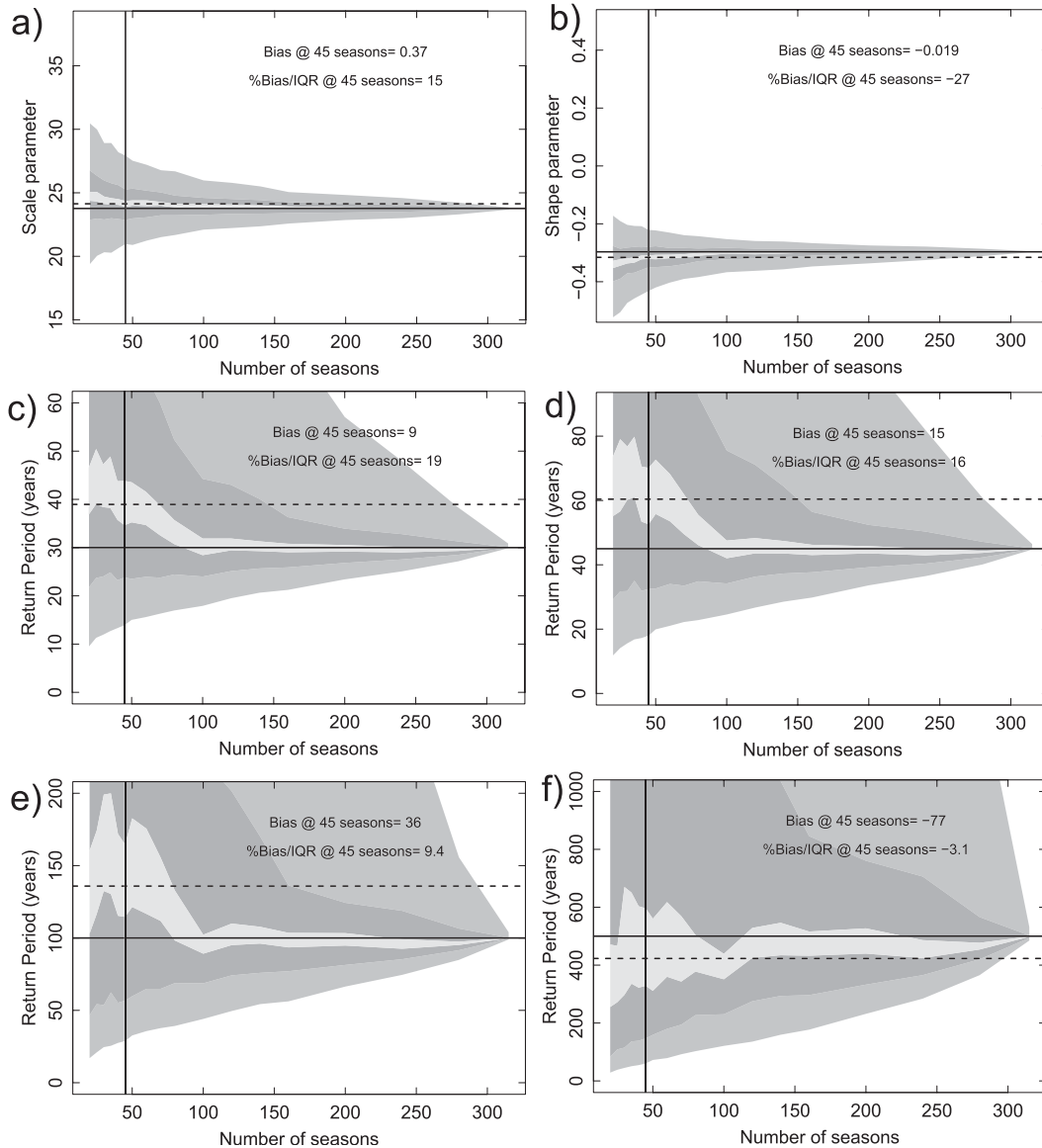


FIG. 12. As in Fig. 10, but using random samples from the GPD fitted to SYS3 using ML.

stochastic LFCs (green lines). The ERA-40 LFC based on 121 windstorms (labeled SR ERA-40 C_ERA_to_SR, the subset of storms in ERA-40 that match in time with the Swiss Re storms; black dashed line) is generally below the MeteoSwiss (MS) ERA-40 C_ERA_to_SR and could be due to some storms missing from SR Hist, in terms of their potential loss. It is noticeable that each of the ERA-40 LFCs and the Swiss Re historical LFC reach a point where the LFC is a horizontal straight line. LFCs produced by catXos have not been modeled and represent empirical loss-frequency estimates. The LFC curves remain horizontal at a RP equal to the length of the dataset. This is also the reason why the LFCs tend to display “jumps” at higher RPs because

of sampling. Figure 17b is similar to Fig. 17a but shows the LFCs based on each of the s2d C_s2d_to_SR datasets. The combination of C_ERA_to_SR and C_s2d_to_ERA (C_S2d_to_SR) shows that the s2d LFCs are close to the ERA-40 LFC (black line) for RPs between 1 and 45 yr. Improved loss estimates are gained from the s2d datasets in the range of 45–300 yr and are comparable but lower than the SR Stoch results.

Figures 17c,d quantify the uncertainties; first in Fig. 17c by exploring the sampling errors in loss estimates due to a limited length of ERA-40 data. We use a similar resampling technique as described in section 3d (see Fig. 14). Taking 20 (a limited number because of computing and data storage requirements) random 45-season samples

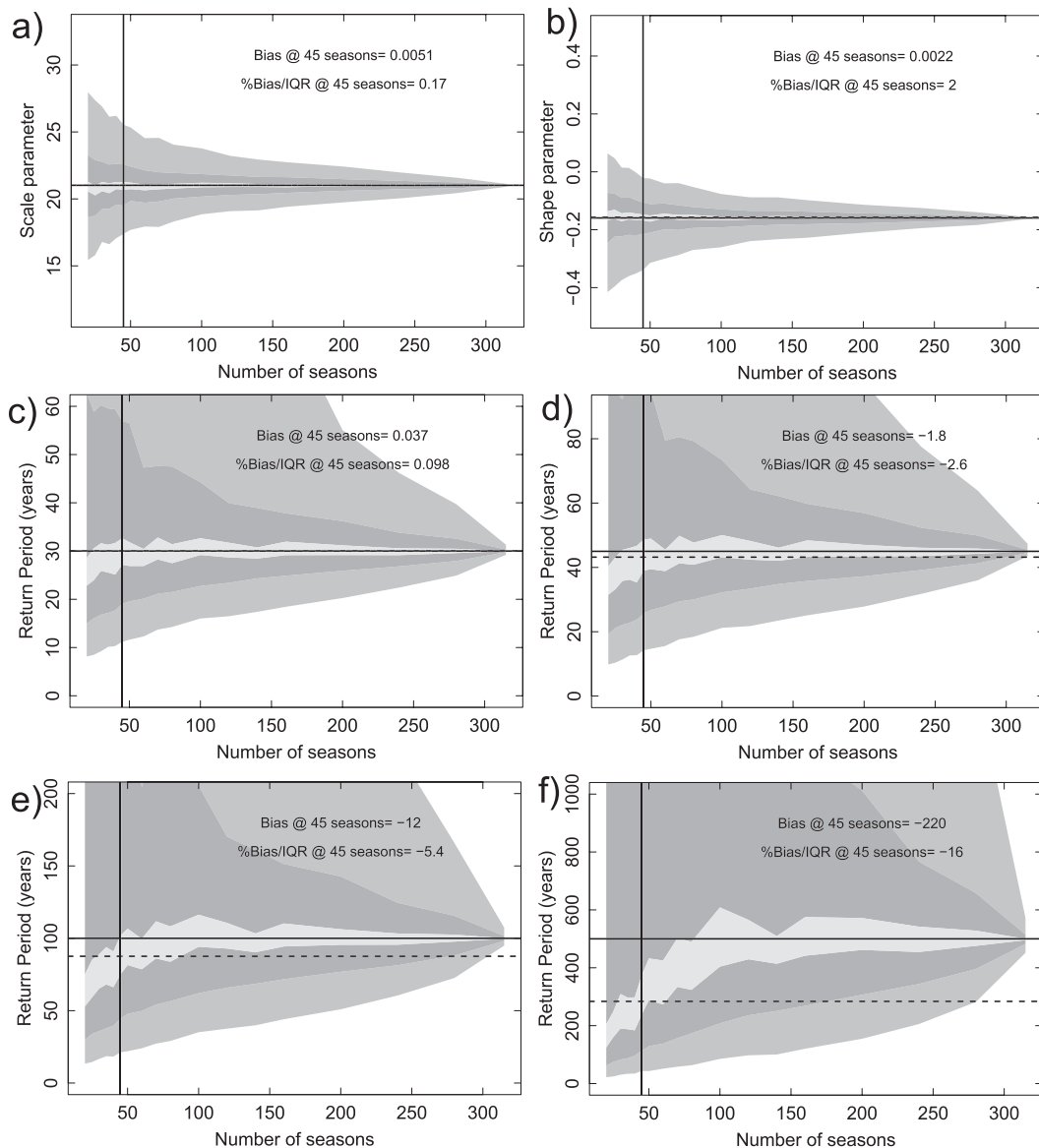


FIG. 13. As in Fig. 12, but using LM method.

of windstorms from SYS3 to calculate their expected loss. The approximate spread of the LFCs is shown by gray shading in Fig. 17c and quantifies the sampling uncertainty in loss estimates due to using a dataset that is only 45 seasons in length. The width of the approximate confidence interval encompasses both the SR Hist and SR Stoch LFCs (green lines). This figure demonstrates the need to use alternative methods to estimate the RP of very extreme wind-related loss events, such as those currently used by Swiss Re (i.e., stochastic datasets) or the alternative that we propose using s2d models.

Figure 17d compares the spread in the LFCs by re-sampling the $C_ERA_to_SR$ and $C_s2d_to_ERA$ adjustment curves (Fig. 16) within their uncertainty estimates.

Taking 20 random samples from the 200 resampled kernel density estimates of the SYS3 adjustment curve (described in section 3a). This results in the LFCs (gray shaded area, Fig. 17d), which also encompass the two Swiss Re LFCs. This figure demonstrates one of the caveats for using s2d data, namely, that there are high uncertainties associated with calibration procedure which should not be ignored. The approximate interquartile range of the LFCs, shown by the dark gray shading, is lower than the best-estimate SYS3 LFC (blue line). This is caused by the distribution of the calibration curve error (Fig. 16b) having a skewed distribution with higher densities at lower adjustment values. It is not clear whether the kernel estimators are biased, especially at such high percentiles.

TABLE 5. New bias-corrected estimates of the RP (yr) of windstorms using the SYS3 windstorm climatology of each EWI calibrated using the PERC method. Columns 2–4 show the RP ERA-40 storms together with the lower and upper bound of the CI (columns 2 and 4) estimated using ML. Columns 5–7 show new estimates of RP in column 6 after correcting the bias caused by the ML method, together with the lower and upper bounds of the SYS3 EWI CI (columns 5 and 7). Columns 8–10 and 11–13 are as in columns 2–4 and 5–7, respectively, but using the LM method.

EWI	ERA-40 ML RP Biased			SYS3 ML RP Bias corrected			ERA-40 LM RP Biased			SYS3 LM RP Bias corrected		
	Lower	RP	Upper	Lower	RP	Upper	Lower	RP	Upper	Lower	RP	Upper
Mean	0.8	1	1.3	0.9	1	1.1	0.8	1	1.3	0.9	1	1.1
	3.2	5	8.2	4.1	4.7	5.8	3.1	5	10.6	4.2	5.2	6.6
	5.2	10	20.2	7.8	8.9	12.5	5.2	10	40	8.1	10.5	17.2
	10.6	30	109	20.7	24	44.3	11.1	30	1354	22	32.6	112
	14.2	50	281	32	37.5	82.6	15.1	50	8751	34.4	55.5	330
	17.1	70	574	42.4	50.1	126	18.4	70	19 415	46.3	78.9	749
	20.5	100	1215	56.8	67.7	201	22.1	100	36 864	62.7	115	1633
	28.7	200	3837	99.2	120	521	30.8	200	90 692	110	239	5063
Q95	0.8	1	1.2	0.9	1	1.1	0.8	1	1.3	0.9	1	1.1
	3.2	5	7.9	4.1	4.9	5.7	3	5	11.7	4.5	5.1	7.4
	5.3	10	19.2	7.7	9.4	11.9	4.9	10	48.7	8.5	10.2	20.9
	10.5	30	96.1	19.9	26	40.1	9.4	30	4287	22.5	31.1	190
	14	50	238	30.2	41.1	72.3	12.3	50	16 311	34.9	52.2	922
	16.6	70	476	39.4	55.4	108	14.5	70	29 604	46.4	73.5	2389
	19.8	100	1015	52	75.5	167	17.1	100	48 369	62.5	106	5146
	26.8	200	3174	86.9	136	409	22.8	200	97 205	108	215	14 319
Sw3q90	0.8	1	1.3	0.9	1	1.1	0.8	1	1.3	0.9	1	1.1
	3.2	5	8.1	4.2	4.9	6	3	5	11	4.1	5	6.2
	5.1	10	20.4	7.8	9.4	12.7	5.3	10	56.1	7.4	9.8	14.1
	9.5	30	128	19	25.9	43.8	10.9	30	2250	16.8	29.1	57.7
	12.2	50	438	27.9	40.8	80.7	14.2	50	4623	23.5	48	122
	14.1	70	1138	35.4	54.7	123	16.9	70	6247	28.9	66.7	210
	16.2	100	2612	45	74.2	194	20	100	7886	35.8	94.5	390
	20.7	200	7811	69.8	132	507	26.4	200	10 574	53.2	186	1307
Sfq95	0.8	1	1.3	0.9	1	1.1	0.8	1	1.3	0.9	1	1.1
	3.1	5	8.4	4.2	4.8	6	3.2	5	10.5	4.3	5.1	6.5
	5.2	10	21.2	8	9.4	13.3	5.4	10	35.5	8.2	10.2	15.8
	10.8	30	116	21.4	26.9	49.4	11.2	30	479	22.1	31.1	72.2
	14.9	50	292	33.5	43.5	94.1	15.1	50	1979	34.9	52.3	156
	18.2	70	566	44.8	59.6	146	18.4	70	4143	46.8	73.7	261
	22.3	100	1130	60.8	83.2	235	22.7	100	7835	63.2	106	464
	32.7	200	3463	109	158	613	33.3	200	20 250	114	216	1389
Sfq95q99	0.8	1	1.2	0.9	1	1.1	0.8	1	1.3	0.9	1	1.1
	3.1	5	8.4	4.2	4.9	6	3.2	5	11	4.3	5	6.7
	5.2	10	21.1	7.9	9.8	13.5	5.5	10	37.2	8.5	10.1	16.9
	11	30	112	21.3	28.9	52.2	12.2	30	519	24.5	30.7	92.8
	15.2	50	272	33.5	47.6	102	17.2	50	1996	39.4	51.7	227
	18.7	70	512	45	66.2	161	21.3	70	4096	54.2	72.9	438
	23.2	100	1004	61.4	93.7	266	26.6	100	7670	76.2	105	890
	34.4	200	3068	111	184	728	40.4	200	19 722	147	214	2862

5. Discussion

The choice of threshold used for the GPD can have a large influence on the windstorm climatology, especially if the data are not sampled from the extreme tail of the distribution (Coles 2001). We performed extensive experiments on the sensitivity of our results to the

choice of threshold combined with qualitative and quantitative measures of the GOF the GPD has to the EWI data. We found that all conclusions remain valid for thresholds above those shown in Table 2 and up to the 99.5% quantile threshold. The magnitude of biases in ML- and LM-based climatologies varies in magnitude (nonmonotonically with increasing threshold); however,

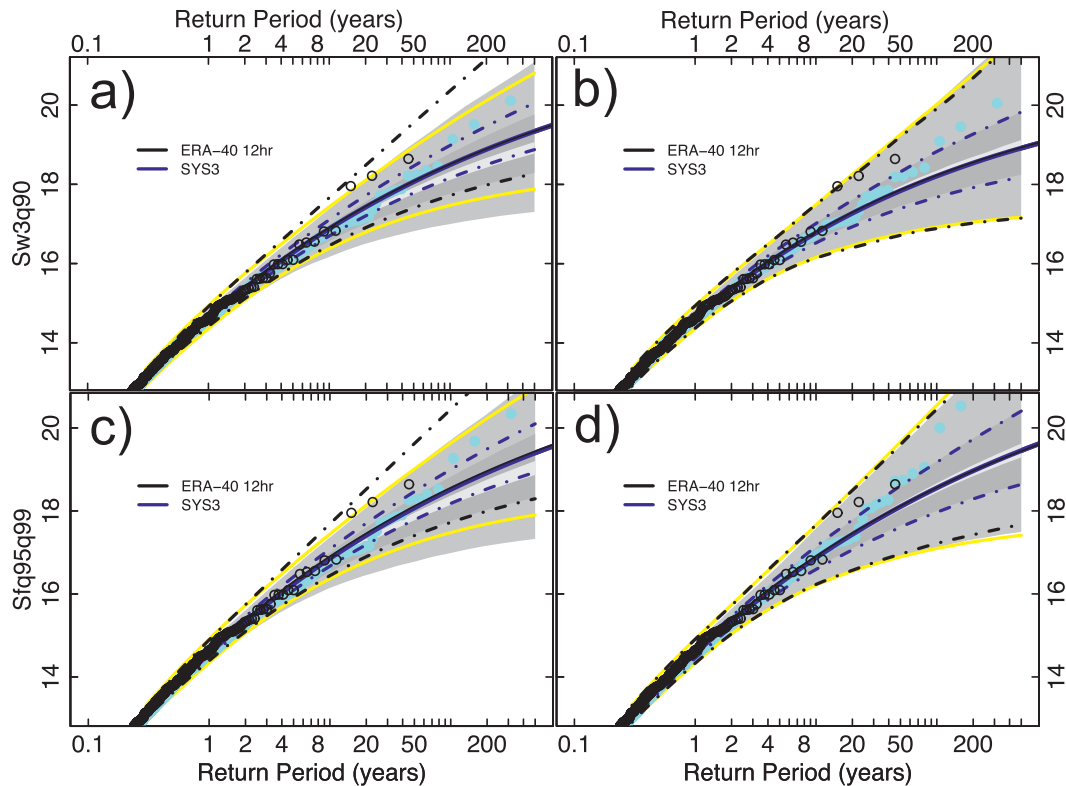


FIG. 14. The sampling and fit uncertainty in the RP and RL of ERA-40 using SYS3 DPOT series calibrated using the percentile method for (a) Sw3q90 and (b) Sf95q99. The light gray shading is 45th–55th percentile range of the resampled SYS3 dataset, whereas the dark gray and medium gray shading show the interquartile (50% confidence interval) and CI, respectively. The yellow lines denote the median width of the upper and lower SYS3 fit uncertainties which have been added to the ERA-40 best-fit GPD (black line). Each dataset is colored according to the legend in the top left of each panel. Dot-dashed colored lines indicate the CIs of the fitted GPD. See section 3d for more information.

the sign of the biases remains the same as those shown in Table 5 and in Figs. 10–13.

To use s2d climatologies to estimate wind-induced loss we applied two stages of calibration. The first to make ERA-40 comparable with SR Hist ($C_{ERA_to_SR}$) and the second to make s2d comparable with ERA-40 ($C_{s2d_to_ERA}$). Conceivably we could remove the need for $C_{ERA_to_SR}$ by using ERA-40 to define the vulnerability function (cf. Fig. 2a). However, some ERA-40 parameters may not have the accuracy or the resolution for the purposes of modeling wind-related loss (e.g., ERA-40 wind gust; Della-Marta et al. 2009). There remains the need to apply $C_{s2d_to_ERA}$ if meaningful loss estimates are to be produced.

Previous work on the impact of GCM resolution on cyclone frequencies (Jung et al. 2006) and interannual–decadal variability of large-scale circulation (Weisheimer et al. 2003) suggests that the statistics of extreme winds are likely to be affected; therefore we place more emphasis on SYS3 results and conclusions. There is strong

evidence in Jung et al. (2006) that horizontal model resolution has a large impact on the number and intensity of cyclones simulated in the ECMWF atmospheric model. They note that the dynamical effect of changing horizontal resolution dominates over the truncation effect for intense cyclones, whereas the truncation effect dominates for shallow cyclones. They also suggest that the ECMWF atmospheric model, tuned to be particularly skillful in medium-range weather forecasting, may have defects when it comes to extended-range integrations at relatively low horizontal resolution. They base this argument on Bengtsson et al. (2006), who found that their GCM produced a realistic representation of extratropical cyclone characteristics, despite the relatively coarse resolution used (T63). All climate models—whether they be oceanic, atmospheric, or coupled models—display inconsistencies with observations. Sometimes these inconsistencies are in the climate of the model and/or its variability. Key components of the model may be parameterized, leading to unresolved scales of interaction

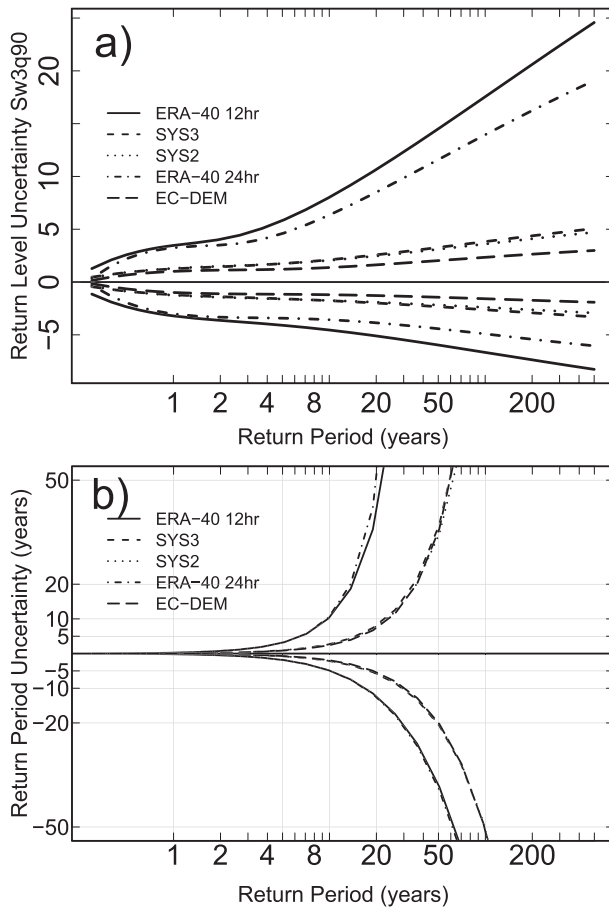


FIG. 15. A comparison of Sw3q90 RL and RP GPD fit uncertainties as a function of ERA-40 windstorm RP. (a) The width of the GPD RL CI for ERA-40 and s2d datasets. (b) The width of the GPD RP CI for ERA-40 and s2d datasets. The datasets are depicted according to the legend shown in the top left of the panels. Note that the x axis of each panel is logarithmic.

within the model, leading to biases or misrepresentations. Unfortunately, given the complexity of dynamical models, it is often difficult to pinpoint the exact reason for a large-scale bias, for example, in circulation, as seen in Fig. 3 of this study. Readers are encouraged to reference Anderson et al. (2003, 2007) for a more complete analysis of ECMWF season forecast model biases and ECMWF (2003, 2007) for a full specification of the model physics.

The 850-hPa GW is suitable for the intercomparison of different s2d data since it is independent of surface boundary layer parameterization. It is available for basically any GCM run. However, the GW at 850 hPa is not fully representative of the wind gusts important for modeling of wind-related loss. GW is an analyzed variable, leading to aliasing of wind intensity (Della-Marta et al. 2009). Thus, we must consider the loss estimation figures with caution.

The aliasing technique was necessary because of the sensitivity of the storm selection method to the temporal resolution of the EW. Using another storm selection technique (e.g., Lagrangian-based storm tracking) probably would not have helped since these methods are also sensitive to spatial and temporal resolution (Bhend 2005; Raible et al. 2008). In the absence of a reliable integrated model parameter (e.g., maximum wind gust) the spatial and temporal sampling must be adequate to analyze the variability of interest.

Do the s2d hindcasts represent the current climate? Each of the extended winter seasons are constructed from seasonal forecasts initialized at different times (see Table 1, e.g., September and November forecasts combined), possibly conditioning the dynamics by the observed initial states during the hindcast period. The longer the hindcast the more the dynamics are able to sample a wider range of initial states leading to a fuller range of possible simulated weather. Likewise, an unequal number of ensembles within hindcasts could lead to dynamics biased to the initial states. The limited resolution of s2d models indicates that the statistics of extremes should only be inferred for synoptic-scale features.

A related point to that above is the assumption that our seasonal resampling (cf. Figs. 10–14, which is white in spectrum) represents likely interannual climate variability. For instance, it would be useful to test the influence of multiannual and decadal climate variability on the convergence of the parameters shown in Figs. 10–14. Resampling with a reddened spectrum would probably result in a slower convergence of the parameters to their long-term values.

In Fig. 7 we showed a certain amount of underdispersion in the Poisson fits to the extended winter season storm frequency. On the contrary, Mailier et al. (2006) show that storms tend to cluster in time (i.e., the Poisson distribution is overdispersed) in this region. Further investigation and comparison with Della-Marta et al. (2009, see their Fig. 9a) show that it is related to the choice of domain and the continental perspective of the EW used in this study. Extreme winds events in the Mediterranean region exhibit a clear tendency not to cluster in time as much as in the northwestern European and adjacent Atlantic Ocean regions.

6. Conclusions

This study has investigated the use of seasonal-to-decadal AOGCM forecast data to estimate the current climate of extreme winds associated with synoptic-scale cyclones over the eastern North Atlantic and western European domain. Differences between the physics, spatial and temporal resolution, and the hindcast length

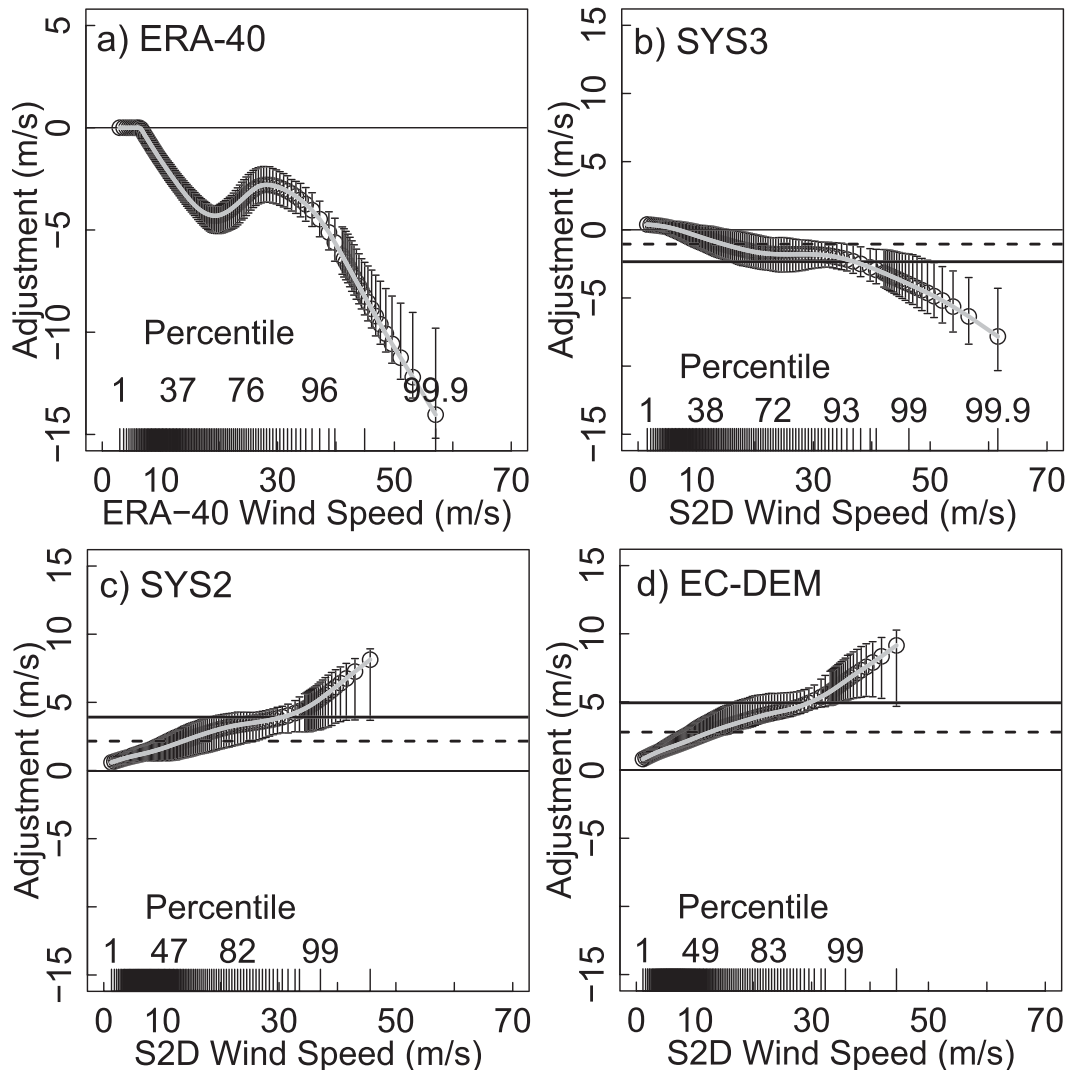


FIG. 16. As in Fig. 8, but for $C_ERA_to_SR$ and $C_s2d_to_SR$ applied to GW fields. (a) The PERC calibration curve (gray line) applied to ERA-40 windstorm fields ($C_ERA_to_SR$, see section 3c). The calibration curves $C_ERA_to_SR$ (gray lines, see section 3c) applied to each windstorm field in each of the s2d datasets (b) SYS3 and (c) SYS2. The error bars represent the CI of the fitted cubic spline (gray line).

result in different windstorm climatologies to ERA-40 (cf. Figs. 3 and 6). After the application of a percentile calibration technique the s2d EWI-based windstorm climates were made very similar to ERA-40 windstorm climate (cf. Figs. 9a,b). This effectively removed the possibility for a different estimate of the current windstorm climate as defined by ERA-40. However, dynamical sampling experiments, utilizing the long s2d datasets, revealed that both the ML and LM methods of estimating the parameters of the GPD [Eq. (1)] contain biases at equivalent sample sizes (around 300 storms; cf. Table 2) from 45 seasons of reanalysis data.

New estimates of the European windstorm climate that have been calibrated to ERA-40 and then bias corrected

suggest that the windstorm climate estimated from ERA-40 using ML method underestimates the severity of the windstorm climate, whereas using the LM method overestimates the severity of the windstorm climate at high RP. This suggests that, for example, a windstorm event that is estimated by ERA-40 to have an RP of 50 yr is more likely to be approximately a 40-yr event (cf. Table 5) when the ML method is used. Evidence to support this hypothesis is seen in the lack of convergence of the GPD parameters using datasets, which are 45 seasons in length (cf. Fig. 10). Experiments using random storms (cf. Fig. 12) indicate that the likely cause of this bias is inherent in the ML GPD parameter estimation method. Thus, by virtue of its length and the ML method, ERA-40-based

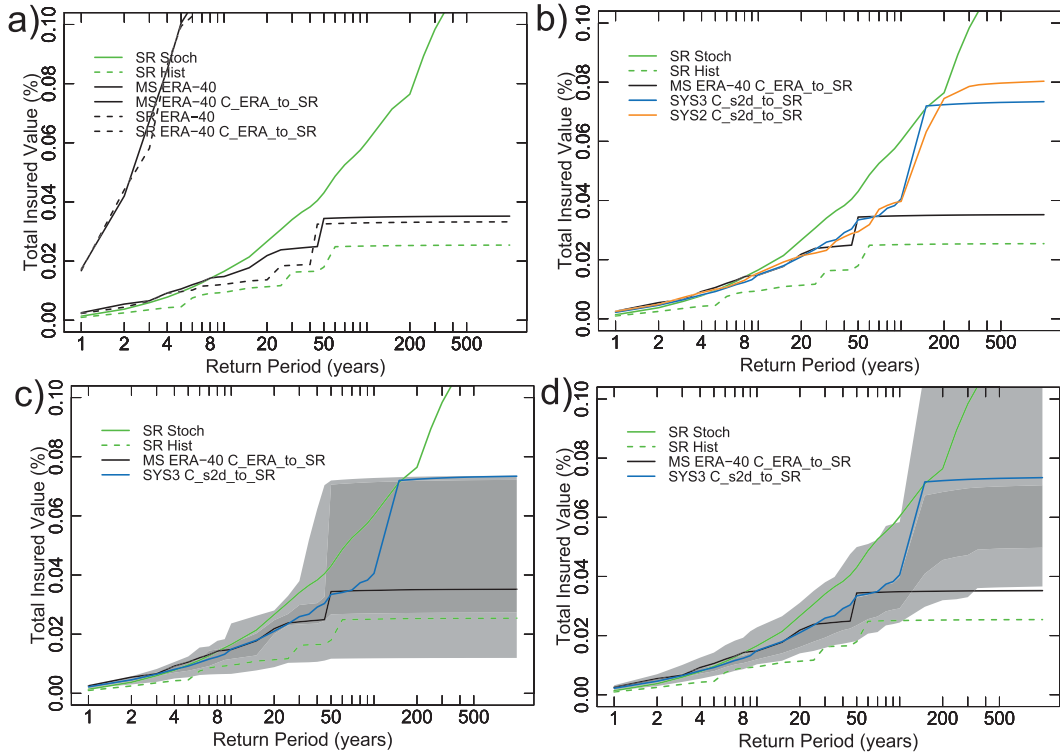


FIG. 17. Loss-frequency curve comparison for various datasets and calibration configurations. (a) A comparison of SR Hist and SR Stoch losses (from the educational catXos model) and the uncalibrated and calibrated loss curves derived from matched storms (SR ERA-40 and SR ERA-40 C_ERA_to_SR, respectively) as well as the MeteoSwiss-identified windstorms both uncalibrated and calibrated (MS ERA-40 and MS ERA-40 C_ERA_to_SR, respectively) using the PERC method) according to the legend in the top left of the panel. (b) SR Hist and SR Stoch loss frequencies are the same as in (a) and are compared with the loss frequencies of C_s2d_to_ERA SR s2d data according to the legend in the top left of the panel. (c) As in (b), but only for SYS3 C_s2d_to_ERA SR loss-frequency data (blue line) and ERA-40 C_ERA_to_SR loss frequencies. The shaded regions show the approximate maximum and minimum (light gray) range and the interquartile range of (dark gray) of loss frequency of 20 random subsamples of 45 seasons in length from the full 315-season dataset. (d) As in (c), but the shaded regions show the approximate maximum and minimum (light gray) range and the approximate interquartile range of (dark gray) of loss frequency based on 20 resampled SYS3 calibration curves (C_s2d_to_SR).

windstorm climatologies (e.g., Della-Marta et al. 2009) are probably underestimating the severity of the current windstorm climate for windstorms with RPs between 10 and 300 yr and overestimating RPs of storms greater than 300 yr. However, these biases are relatively small compared to the role of dynamical sampling variability, which indicates, given 45 seasons to estimate the windstorm climate, the biases represent only around 20% (5% for L-moments) of the interquartile range of possible climates.

Using s2d allows various sampling experiments to quantify uncertainties associated with having ERA-40 length datasets. The ML ERA-40 climatology lies within the middle 10% of 45-season subsampled SYS3 windstorm climatologies (cf. Figs. 14a,c). This implies that the GPD fit to ERA-40 is accurate given that it is only 45 seasons in length, yet the windstorm climate of ERA-40 is not likely (~65% chance) to have converged to the parent

distribution (cf. Figs. 14a,c, blue line). The second conclusion from this analysis is that the profile log-likelihood ERA-40 GPD fit uncertainty overestimates the width of the lower bound of the RP of windstorms. Conversely, we also find that the upper bound of the ERA-40 RP fit uncertainty is underestimated. For the LM method the width of the 45-season sampled climatologies are not biased but cover a wider range of RP estimates than the ML estimates. The wider range of LM climatologies implies (cf. Figs. 14b,d) that the LM method is less accurate in determining the long-term windstorm climate given a 45-season sample than the ML-based climatologies.

We have shown that there is possibly useful skill in the prediction of the intensity of wind over the western European domain from the ECMWF s2d models during the first month of the November initialized forecasts (cf. Fig. 4) with little skill thereafter. Therefore, there is

potential value in monthly forecast systems predicting four weeks in advance (Weigel et al. 2008a).

Our analysis has demonstrated that it is possible to use s2d data to gain useful estimates of the current windstorm climate and related loss. These data give opportunities to compliment the Swiss Re stochastically generated windstorm events with a completely independent set of physically consistent simulations of extreme wind-related weather in the current climate. New loss uncertainty estimates associated with sampling 45 seasons from SYS3 advocate the use of either a stochastically generated windstorm dataset as used by Swiss Re, or a dynamical ensemble-based dataset to improve the estimates of loss associated with high RP events. While there remains considerable uncertainty in long RP loss estimates because of the need to propagate calibration errors through the loss model, we demonstrate that s2d models are a powerful tool for exploring the climate of extremes and offer another perspective on the windstorm climate over the eastern North Atlantic and western Europe.

Acknowledgments. We acknowledge Swiss Reinsurance Company who substantially sponsored this research through the Swiss National Science Foundation and the National Centre for Competence in Research Climate (NCCR-Climat). We thank Christoph Frei for discussions on the ML and LM methods as well as providing extreme value analysis software, and Stefan Wunderlich, Malcolm Haylock, and Marc Wüest for useful discussions and input to this paper. We thank three reviewers who provided detailed suggestions and comments that greatly improved the content of the paper. Paul Della-Marta was partially sponsored by the 6th EU framework program ENSEMBLES contract number GOCE-CT-2003-505539 and by the Swiss National Science Foundation and the National Centre for Competence in Research Climate (NCCR-Climat).

APPENDIX A

List of Acronyms

95PERC	The 95th percentile method of calibration
AOGCMs	Atmosphere–ocean general circulation models
C_ERA_to_SR	The calibration of ERA-40 wind climatology to the SR Stoch wind climatology
C_s2d_to_ERA	The calibration of s2d wind climatology to the ERA-40 wind climatology
C_s2d_to_SR	The calibration of s2d wind climatology to the SR Stoch wind climatology

catXos	Educational version of the Swiss Re windstorm loss model
CI	Confidence interval (in most cases the 95% confidence interval)
DPOT	Declustered peaks over threshold series
EC-DEM	ECMWF s2d model used in the DEMETER project
ECMWF	European Centre for Medium-Range Weather Forecasts
ERA-40	40-yr ECMWF Re-Analysis
EVA	Extreme value analysis
EWI	Extreme wind index
GCM	General circulation model
GOF	Goodness of fit (assessed using the Anderson–Darling test statistic)
GPD	Generalized Pareto distribution
GPH	Geopotential height
GW	Geostrophic wind speed
IQR	Interquartile range
LFC	Loss-frequency curve
LM	The L-moments method
Mean	An EWI defined in appendix B
MEAN	The mean method of calibration
ML	The maximum likelihood method
PERC	The percentile method of calibration
Q95	An EWI defined in appendix B
RL	Return level
RP	Return period
RPSSd	The debiased ranked probability skill score
s2d	Seasonal-to-decadal climate forecasts
Sfq95	An EWI defined in appendix B
Sfq95q99	An EWI defined in appendix B
SR Hist	Swiss Re historical wind field dataset
SR Stoch	Swiss Re stochastic wind field dataset
Sw3q90	An EWI defined in appendix B
Swiss Re	Swiss Reinsurance Company
SYS2	ECMWF system2 s2d model
SYS3	ECMWF system 3 s2d model

APPENDIX B

Mathematical Notation of Extreme Wind Indices

This appendix describes each of the extreme wind indices in mathematical notation for the benefit of readers who may wish to implement such indices. The content of this appendix has been taken from Della-Marta et al. (2009). The indices are denoted in terms of a generic wind variable W . Where possible we tried to take into account the unequal areas of each grid box by weighting sums and multipliers by the cosine of the latitude of each grid point.

a. *Mean: Mean wind ($m s^{-1}$)*

$$\bar{X}(t) = \frac{1}{N_{\kappa\delta}} \sum_{x,y \in \delta} \kappa(x, y) w(w, y, t), \quad (B1)$$

where κ are the individual gridpoint weights, which only depend on y , $\kappa(x, y) = \cos[\text{latitude}(y)]$, $N_{\kappa\delta} = \sum_{x,y \in \delta} \kappa(x, y)$, and δ denotes the domain.

b. *Q95: The spatial 95% quantile wind ($m s^{-1}$)*

$$Q95(t) = F_*^{-1}(p) = \min\{w : p \leq F_*(W)\}, \quad (B2)$$

where $p = 0.95$ and F_* is the latitude-weighted empirical cumulative distribution function of $\{w(x, y, t) : (x, y) \in \delta\}$.

$$F_*(W) = \frac{1}{N_{\kappa\delta}} \sum_{x,y \in \delta} \kappa(x, y) \mathbb{1}[(x, y, t) \leq W], \quad (B3)$$

where $\mathbb{1} = \begin{cases} 1 : w(x, y, t) \leq W \\ 0 : \text{otherwise} \end{cases}$.

c. *Sw3q90: Cube root of the sum of wind cubed above the domain climatological 90% quantile (nondimensional)*

$$Sw3q90(t) = \left(\sum_{x,y \in \delta} \left\{ \mathbb{1}_{\{>\bar{q}90\}} [w(x, y, t) > \bar{q}90] \kappa(x, y) \times [w(x, y, t) - \bar{q}90]^3 \right\} \right)^{1/3}, \quad (B4)$$

where $\mathbb{1}_{\{>\bar{q}90\}} = \begin{cases} 1 : w(x, y, t) > \bar{q}90 \\ 0 : \text{otherwise} \end{cases}$. The domain mean quantile function $\bar{q}90$ is given by

$$\frac{1}{N_{\delta}} \sum_{x,y \in \delta} q90(x, y), \quad (B5)$$

where $q90(x, y) = F^{-1}(p) = \min\{w : p \leq F(W)\}$, $p = 0.90$; F is the empirical cumulative distribution function of $\{w(x, y, t) : t \in \text{ONDJFMA}\}$.

d. *Sfq95: Sum of the fraction of wind divided by the gridpoint climatological 95% quantile (nondimensional)*

$$Sfq95(t) = \sum_{x,y \in \delta} \mathbb{1}_{\{>1\}} \left[\frac{w(x, y, t)}{q95(x, y)} \right] \kappa(x, y) \frac{w(x, y, t)}{q95(x, y)}, \quad (B6)$$

where $\mathbb{1}_{\{>1\}} = \begin{cases} 1 : \left[\frac{w(x, y, t)}{q95(x, y)} \right] > 1 \\ 0 : \text{otherwise} \end{cases}$. The gridpoint quantile function $q95$ is given by

$$q95(x, y) = F^{-1}(p) = \min\{w : p \leq F(W)\}, \quad (B7)$$

where $p = 0.95$.

e. *Sfq95q99: Sum of the fraction of extreme wind divided by the length of the distribution tail (nondimensional)*

$$Sfq95q99(t) = \sum_{x,y \in \delta} \mathbb{1}_{\{>1\}} \left[\frac{w(x, y, t) - q95(x, y)}{q99(x, y) - q95(x, y)} \right] \times \kappa(x, y) \frac{w(x, y, t) - q95(x, y)}{q99(x, y) - q95(x, y)}, \quad (B8)$$

where $\mathbb{1}_{\{>0\}} = \begin{cases} 1 : \left[\frac{w(x, y, t) - q95(x, y)}{q99(x, y) - q95(x, y)} \right] > 0 \\ 0 : \text{otherwise} \end{cases}$. The gridpoint quantile functions $q95$ and $q99$ are given above.

REFERENCES

Anderson, D. L. T., 2008: Overview of seasonal forecasting. *Seasonal Climate: Forecasting and Managing Risk*, A. Troccoli et al., Eds., NATO Science Series IV: Earth and Environmental Sciences, Springer, 45–65.

—, and Coauthors, 2003: Comparison of the ECMWF seasonal forecast systems 1 and 2, including the relative performance for the 1997/8 El Niño. European Centre for Medium-Range Weather Forecasts Tech. Rep. 404, 95 pp.

—, and Coauthors, 2007: Development of the ECMWF seasonal forecast system 3. European Centre for Medium-Range Weather Forecasts Tech. Rep. 503, 58 pp.

Balakrishnan, V., C. Nicolis, and G. Nicolis, 1995: Extreme value distributions in chaotic dynamics. *J. Stat. Phys.*, **80**, 307–336.

Bengtsson, L., K. I. Hodges, and E. Roeckner, 2006: Storm tracks and climate change. *J. Climate*, **19**, 3518–3543.

Bhend, J., 2005: North Atlantic and European cyclones: Their variability and change from 1881 to 2003. M.S. thesis, Institute of Geography, University of Berne, 102 pp.

Choulakian, V., and M. A. Stephens, 2001: Goodness-of-fit tests for the generalized Pareto distribution. *Technometrics*, **43**, 478–484.

Coles, S., 2001: *An Introduction to Statistical Modeling of Extreme Values*. Springer, 224 pp.

Cox, D. R., and E. J. Snell, 1968: A general definition of residuals. *J. Roy. Stat. Soc.*, **30**, 248–275.

Davison, A. C., and R. L. Smith, 1990: Models for exceedances over high thresholds. *J. Roy. Stat. Soc.*, **52**, 393–442.

Della-Marta, P. M., and H. Wanner, 2006: A method of homogenizing the extremes and mean of daily temperature measurements. *J. Climate*, **19**, 4179–4197.

—, H. Mathis, C. Frei, M. A. Liniger, J. Kleinn, and C. Appenzeller, 2009: The return period of wind storms over Europe. *Int. J. Climatol.*, **29**, 437–459.

Dukes, M. D. G., and J. P. Palutikof, 1995: Estimation of extreme wind speeds with very long return periods. *J. Appl. Meteor.*, **34**, 1950–1961.

- ECMWF, 2003: IFS documentation cycle CY23r4, part IV: Physical processes. European Centre for Medium-Range Weather Forecasts Tech. Rep., 166 pp.
- , 2007: IFS documentation cycle CY31r1, part IV: Physical processes. European Centre for Medium-Range Weather Forecasts Tech. Rep., 155 pp.
- Efron, B., and G. Gong, 1983: A leisurely look at the bootstrap, the jackknife, and cross-validation. *Amer. Stat.*, **37**, 36–48.
- Frei, C., R. Schöll, S. Fukutome, J. Schmidli, and P. L. Vidale, 2006: Future change of precipitation extremes in Europe: Inter-comparison of scenarios from regional climate models. *J. Geophys. Res.*, **111**, D06105, doi:10.1029/2005JD005965.
- Giles, D. E., and H. Feng, 2009: Bias-corrected maximum likelihood estimation of the parameters of the generalized Pareto distribution. University of Victoria Econometric Working Paper EWP0902, 16 pp.
- Gumbel, E. J., 1958: *Statistics of Extremes*. Columbia University Press, 375 pp.
- Heneka, P., T. Hofherr, B. Ruck, and C. Kottmeier, 2006: 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.
- Holton, J. R., 2004: *An Introduction to Dynamic Meteorology*. 4th ed. Academic Press, 535 pp.
- Hosking, J. R. M., 1990: L-moments: Analysis and estimation of distributions using linear combinations of order statistics. *J. Roy. Stat. Soc.*, **52**, 105–124.
- Jung, T., S. K. Gulev, I. Rudeva, and V. Soloviev, 2006: Sensitivity of extratropical cyclone characteristics to horizontal resolution in the ECMWF model. *Quart. J. Roy. Meteor. Soc.*, **132**, 1839–1857.
- Kharin, V. V., and F. W. Zwiers, 2000: Changes in the extremes in an ensemble of transient climate simulations with a coupled atmosphere–ocean GCM. *J. Climate*, **13**, 3760–3788.
- Klawa, M., and U. Ulbrich, 2003: 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.
- Lamb, H. H., and K. Fyfe, 1991: Grading of storms. *Historic Storms of the North Sea, British Isles and Northwest Europe*, Cambridge University Press, 7–32.
- Leckebusch, G. C., U. Ulbrich, L. Froehlich, and J. G. Pinto, 2007: Property loss potentials for European midlatitude storms in a changing climate. *Geophys. Res. Lett.*, **34**, L05703, doi:10.1029/2006GL027663.
- Liniger, M. A., H. Mathis, C. Appenzeller, and F. J. Doblas-Reyes, 2007: Realistic greenhouse gas forcing and seasonal forecasts. *Geophys. Res. Lett.*, **34**, L04705, doi:10.1029/2006GL028335.
- Mailier, P. J., D. B. Stephenson, C. A. T. Ferro, and K. I. Hodges, 2006: Serial clustering of extratropical cyclones. *Mon. Wea. Rev.*, **134**, 2224–2240.
- Palmer, T. N., and Coauthors, 2004: Development of a European Multimodel Ensemble System for Seasonal-to-Interannual Prediction (DEMETER). *Bull. Amer. Meteor. Soc.*, **85**, 853–872.
- Palutikof, J. P., B. B. Brabson, D. H. Lister, and S. T. Adcock, 1999: A review of methods to calculate extreme wind speeds. *Meteor. Appl.*, **6**, 119–132.
- Pandey, M. D., P. H. A. J. M. Van Gelder, and J. K. Vrijling, 2003: Bootstrap simulations for evaluating the uncertainty associated with peaks-over-threshold estimates of extreme wind velocity. *Environmetrics*, **14**, 27–43.
- Pinto, J. G., E. L. Froehlich, G. C. Leckebusch, and U. Ulbrich, 2007: 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.
- Raible, C. C., P. M. Della-Marta, C. Schwierz, H. Wernli, and R. Blender, 2008: Northern Hemisphere midlatitude cyclones: A comparison of detection and tracking methods and different reanalyses. *Mon. Wea. Rev.*, **136**, 880–897.
- R Development Core Team, cited 2005: R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, 1706 pp. [Available online at <http://cran.r-project.org/doc/manuals/refman.pdf>.]
- Schwierz, C., C. Appenzeller, H. C. Davies, M. A. Liniger, W. Muller, T. F. Stocker, and M. Yoshimori, 2006: Challenges posed by and approaches to the study of seasonal-to-decadal climate variability. *Climatic Change*, **79**, 31–63.
- , P. Köllner-Heck, E. Zenklusen Mutter, D. N. Bresch, P.-L. Vidale, M. Wild, and C. Schär, 2009: Modelling European winter wind storm losses in current and future climate. *Climatic Change*, **101**, 485–514.
- Stephenson, A., and E. Gilleland, 2006: Software for the analysis of the extreme events: The current state and future directions. *Extremes*, **8**, 87–109.
- Swiss Re, 2000: Storm over Europe—An underestimated risk. Swiss Reinsurance Company, 29 pp.
- Uppala, S. M., and Coauthors, 2005: The ERA-40 Re-Analysis. *Quart. J. Roy. Meteor. Soc.*, **131**, 2961–3012.
- van den Brink, H. W., G. P. Konnen, and J. D. Opsteegh, 2004a: Statistics of extreme synoptic-scale wind speeds in ensemble simulations of current and future climate. *J. Climate*, **17**, 4564–4574.
- , —, —, G. J. van Oldenborgh, and G. Burgers, 2004b: Improving 10⁴-year surge level estimates using data of the ECMWF seasonal prediction system. *Geophys. Res. Lett.*, **31**, L17210, doi:10.1029/2004GL020610.
- , —, —, —, and —, 2005: Estimating return periods of extreme events from ECMWF seasonal forecast ensembles. *Int. J. Climatol.*, **25**, 1345–1354.
- Vannitsem, S., 2007: Statistical properties of the temperature maxima in an intermediate order quasi-geostrophic model. *Tellus*, **59A**, 80–95.
- Weigel, A. P., M. A. Liniger, and C. Appenzeller, 2007: The discrete Brier and ranked probability skill scores. *Mon. Wea. Rev.*, **135**, 118–124.
- , D. Baggenstos, M. A. Liniger, F. Vitart, and C. Appenzeller, 2008a: Probabilistic verification of monthly temperature forecasts. *Mon. Wea. Rev.*, **136**, 5162–5182.
- , M. A. Liniger, and C. Appenzeller, 2008b: Can multi-model combination really enhance the prediction skill of probabilistic ensemble forecasts? *Quart. J. Roy. Meteor. Soc.*, **134**, 241–260.
- Weisheimer, A., M. V. Kurgansky, K. Dethloff, and D. Handorf, 2003: Extratropical low-frequency variability in a three-level quasi-geostrophic atmospheric model with different spectral resolution. *J. Geophys. Res.*, **108**, 4171, doi:10.1029/2001JD001282.
- Zhang, J., and M. A. Stephens, 2009: A new and efficient estimation method for the generalized Pareto distribution. *Technometrics*, **51**, 316–325.
- Zimmerli, P., 2003: Natural catastrophes and reinsurance. Swiss Reinsurance Company Tech. Rep., 48 pp. [Available online at http://media.swissre.com/documents/Nat_Cat_reins_en.pdf.]