On estimating long period wind speed return levels from annual maxima

M.A. Ben Alaya, Francis W. Zwiers & Xinlong Zhang

2021

Original citation:

Ben Alaya, M. A., Zwiers, F. W., & Zhang, X. (2021). On estimating long period wind speed return levels from annual maxima. *Weather and Climate Extremes, 34*, 100388. https://doi.org/10.1016/j.wace.2021.100388

Weather and Climate Extremes 34 (2021) 100388

Contents lists available at ScienceDirect

Weather and Climate Extremes

journal homepage: www.elsevier.com/locate/wace

On estimating long period wind speed return levels from annual maxima

M.A. Ben Alaya [a,*], F.W. Zwiers [a,b], X. Zhang [c]

[a] Pacific Climate Impacts Consortium, University of Victoria, PO Box 1700 Stn CSC, Victoria, BC, V8W2Y2, Canada
[b] Nanjing University of Information Science and Technology, Nanjing, China
[c] Climate Research Division, Environment and Climate Change Canada, Toronto, ON, M3H 5T4, Canada

ARTICLE INFO

Keywords:
Long period wind speed return levels
Generalized extreme value distribution
Regional climate model

ABSTRACT

The uniform risk engineering practices that are increasingly being adopted for structural design require estimates of the extreme wind loads with very low annual probabilities of exceedance, corresponding to return periods of up to 3000-years in some cases. These estimates are necessarily based on observational wind data that typically spans only a few decades. The estimates are therefore affected by both large sampling uncertainty and, potentially, non-negligible biases. Design practices that aim to meet mandated structural reliability criteria take the sampling uncertainty of long period wind speed or wind pressure estimates into account, but reliability could be compromised if estimates are also biased. In many circumstances, estimates are obtained by fitting an extreme value distribution to annual maximum wind speed observed over a few decades. A key assumption implicit in doing so is that wind speed annual maxima are max-stable. Departures from max-stability can exacerbate the uncertainty of long-period return level estimates by inducing systematic estimation bias as well. Observational records, however, are generally too short to assess max-stability. We therefore use wind speed data from a large (50-member) ensemble of CanRCM4 historical simulations over North America to assess whether wind speed annual maxima are max-stable. While results are generally reassuring at the continental scale, disquieting evidence of a lack of max-stability is often found in the central and southern parts of the continent. Results show that when annual maximum wind speeds are not max-stable, long period return level extreme wind speeds tend to be underestimated, which would compromise reliability if used to design infrastructure such as tall buildings and towers.

1. Introduction

Accurate estimation of the occurrence of extreme wind speeds is needed in many design projects such as buildings, bridges, wind turbines and radio masts to are needed ensure the safety and reliability of these structures. Ideally, these estimates should be both unbiased and possess low uncertainty so as to avoid compromised safety and reliability as a consequence of under-design, and excessive construction costs due to over-design. To this end, the occurrence of extreme wind speeds is often expressed in terms of a return level (RL) X_T, which is the maximum wind speed that is likely to be exceeded, on average once every T-years. These RLs are generally estimated using a theoretical statistical distribution that is fitted to a sample of observed wind speed data.

Traditionally, wind speed design values, expressed as RLs, and corresponding wind pressure design values, have been required by building and construction codes for return periods ranging from 10- to 100-years (CSA 2019). Recently, however, changes in building codes, such as the American Society for Civil Engineers standard (ASCE 7–16 2017), are leading to requirements for wind load RLs corresponding to much longer return periods. The National Research Council of Canada, which is responsible for the National Building Code of Canada, is also considering this possibility. In the case of ASCE 7, return periods have been set at 300, 700, 1700 and 3000-years respectively for structures in ASCE Risk Categories I-IV respectively, where Risk Category III refers primarily to public assembly facilities and Risk Category IV refers to essential facilities such as hospitals (McAllister et al., 2018). These evolving requirements are in contrast with the stark reality that in most locations, only 10–50 years of wind observations are available (Cook 1986; Holmes 2018) for the estimation of design wind speeds and pressures. The consequent need for extrapolation beyond the information that is contained in the observations usually leads to the use of the extreme value theory (Reiss et al., 2007), which often involves the application of the block maximum approach (BM) to samples of annual maxima and therefore the use of the generalized extreme value distribution (Cheng

---

* Corresponding author.
E-mail address: mohamedalibenalaya@uvic.ca (M.A. Ben Alaya).

https://doi.org/10.1016/j.wace.2021.100388
Received 24 December 2020; Received in revised form 29 July 2021; Accepted 6 October 2021

and Yeung 2002; Palutikof et al., 1999; Perrin et al., 2006).

The application of the BM approach involves the assumption that the sample of block maxima exhibits the property of max-stability. Max-stability reflects a regular behavior of extreme values according to which the distribution of higher length block maxima should be equal to the initial one, except for a linear transformation (Gumbel 1958). A distribution $F(x)$ is max-stable if and only if for any $n \in \mathbb{N}$, there exist real numbers $a_n > 0$ and b_n such that

$$F^n(b_n + a_n x) = F(x), \quad x \in \mathbb{R}. \tag{1}$$

The max-stability property is a necessary condition for the existence of a limiting distribution for maxima (Fisher and Tippett 1928), and this property is only satisfied by the GEV distribution (Gnedenko 1943). The data, to which the distribution is fitted, however, may not be max-stable, therefore begging questions about the practical use of the fitted distribution to extrapolate beyond available samples of extremes. Indeed, typical 10 to 50-year data records might not be large enough to reveal important aspects of irregularity that might be present in annual maxima.

Analysis of a large ensemble of climate model simulations can be helpful to study questions that cannot be studied with available observations (Ben Alaya et al., 2020; Huang et al. 2016, 2020; Li et al., 2019). For instance, in the case of extreme precipitation, Ben Alaya et al. (2020) used a large ensemble of regional climate simulations over North America and found that annual maxima of precipitation tends not to be max-stable in the simulated climate. Their study demonstrates how the lack of max-stability might have serious implications for estimation of very long period precipitation return levels (such as 1000-year return RLs), suggesting that we should also think more deeply about this issue for other climate variables when ambitious extrapolation to the deep upper tail is required.

The need for wind speed return levels corresponding to return periods as long as 3000-years (Holmes 2018; Jain et al., 2001) implies a need to assess whether a possible lack of max-stability should be of concern. In this paper, we assess whether extreme wind speed simulated from the Canadian Regional Climate Model (CanRCM4) can be well described by a max-stable distribution using a large ensemble (50-member) of historical simulations. We therefore explore the implications of lack of max-stability on the estimation of very long period wind speed return levels for return periods such as 1000-years. The remainder of this paper is structured as follows: The data and the methods are described in Section 2, results are presented in Section 3 and conclusions and some further discussion is provided in Section 4.

2. Data and methods

2.1. Data

We use daily maximum of "instantaneous" (described further below) near surface (10 m anemometer height) wind speed from a 50-member large ensemble simulation of the Canadian Regional Climate Model (CanRCM4) covering North America at 0.44° spatial horizontal resolution (~50 km) over the historical period 1951–2000. The simulations were driven by a corresponding 50-member large ensemble simulation produced with the second generation of Canadian earth system model (CanESM2) using historical greenhouse gas, aerosols, land use, solar and volcanic forcing. CanRCM4 is developed by the Canadian Center for Climate Modelling and Analysis and is a participant in the Coordinated Regional climate Downscaling Experiment (CORDEX) framework (Giorgi et al., 2009). Further details about CanRCM4 can be found in Scinocca et al. (2016) and von Salzen et al. (2013).

The 50-ensemble members combined provide 2500 annual maxima for the 1951–2000 period at each grid point that are representative of the various physical processes that produce extreme wind speed in CanESM2/CanRCM4. Since historical forcings were used in those simulations, we tested whether they contain discernible trends in annual maximum wind speed over that period using the Mann Kendal test, considering each member and each grid box separately. The no-trend null hypothesis was rejected in fewer than 5% of tests when conducting the test at 5% significance level, indicating that there is no evidence of a discernible field significant response to the historical external forcing in annual maximum wind speeds. We therefore assumed quasi-stationarity over the 1951–2000 period.

There is only a very limited body of research that has assessed the ability of regional climate models to simulated observed near surface winds, and very little climate research that is specifically focused on CanRCM4 or on wind speed extremes. A challenge in this regard is the large discrepancy between the spatial scales represented by in-situ anemometer observations and the scales that are represented by climate models. Extreme damaging winds often reflect intense short duration small scale flows with spatial scales of a few km and time scales of minutes, such as downdrafts in convective cells or tornadic behaviour (Klemp 1987). They can also be strongly affected by the details of the local topography and other local land surface characteristics in the area surrounding the anemometer (Petersen et al., 1998). In contrast, climate models such as CanRCM4 simulate grid box mean wind speeds that varies according the equations governing the atmospheric circulation at intervals of the model time-step, which is 20 min in the case of CanRCM4. Such extreme model simulated near surface "instantaneous" wind speeds are not directly comparable with the short interval mean wind speeds observed at meteorological stations. Nevertheless, despite these scale differences, Jeong et al. (2020) report that the model reasonably reproduces the observed annual mean wind climatology. It also well reproduces observed driving rain wind pressures (DRWP), which are also used for engineering design. Specifically, in the National Building Code of Canada (NRCA 2015), DRWP is defined as the 5-year return level for wind pressure (which is proportional to wind speed squared) when it is raining heavily at a rate of at least 1.8 mm/h. Thus, this is a quantile that reflects relatively frequent, moderately large, wind speed extremes in most locations. Further, we have compared climatological annual maximum hourly wind speeds for 1986–2016 from a network of 130 weather station located in Canada with the corresponding CanRCM4 climatology (Fig. 1), finding reasonable correspondence between models and observations in terms of both pattern and magnitude, consistent with the assessment reported in Cannon et al. (2020). Given the differences in scale discussed above and the fact that the model winds will be very dependent on the parameterization of the model's land surface and its interaction with the atmospheric boundary layer, the correspondence between models and observations shown in Fig. 1b is perhaps as good as can be expected.

With these limitations in mind, it is worth briefly commenting on the magnitude of the long period return levels, which based on empirical estimates of the 1000-year return wind speeds from the 2500 available annual maxima, vary between 13.4 ms^{-1} (48.2 km/h) and 48.5 ms^{-1} (175 km/h) depending on location. These speeds, particularly at the lower end of the range, may appear to be low relative to the observed intensity of local extreme short duration winds, but it should be borne in mind these are grid box mean 20-min average wind speeds for grid boxes that cover areas of about 2500 km^2, and thus their magnitudes should be understood in that context.

2.2. Methods

Our analysis seeks to assess estimates of long period return levels (1000- and 2000-year events) obtained from GEV distributions fitted to wind speed annual maxima via the method of maximum likelihood. The distribution function for the GEV distribution is given by

$$G\left(z\right) = \begin{cases} \exp\left\{ -\left[1 + \xi(z-\mu)/\sigma\right]_+^{-1/\xi} \right\}, \ \ 1 + \xi z > 0, \quad if \ \ \xi \neq 0, \\ \exp\{ -\exp[-(z-\mu)/\sigma]\}, \ \ z \in \mathbb{R} \end{cases} , if \ \ \xi = 0, \tag{2}$$

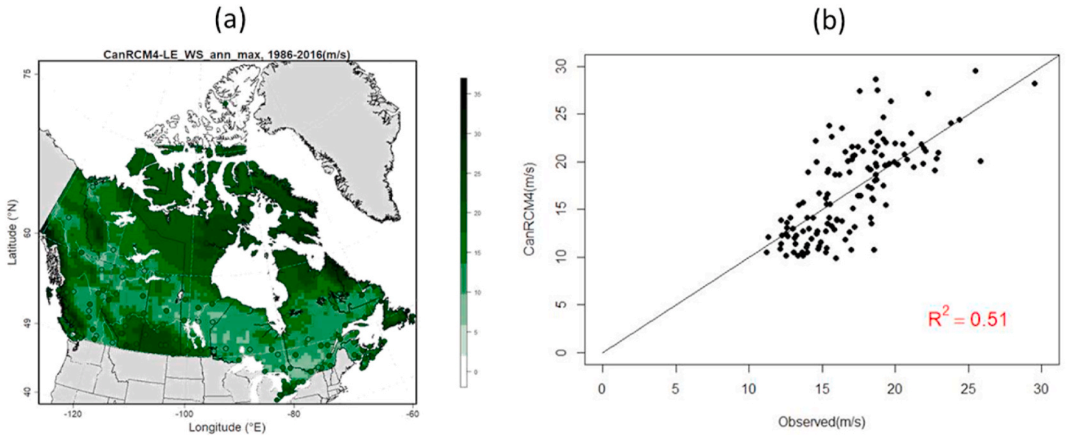


Fig. 1. Spatial (a) and scatter plots (b) of CanRCM4 ensemble simulation and 130 weather station observations for the average over the period 1986–2016 of annual maxima of hourly mean wind speed over the period 1986–2016.

where μ, σ and ξ are the location, scale and shape parameters respectively, with $\xi > 0$ corresponding to the heavy-tailed (Fréchet) version of the distribution. The shape parameter governs the tail behaviour of the distribution and quantify the rate at which the scale change linearly with the increasing location as the block length increases.

The evaluation of return level estimation models used in climatology and hydrology has typically involved the use of classical goodness-of-fit (GOF) tests and empirical quantile estimates as references. Classical goodness-of-fit tests are useful for checking at least the central part of the distribution, but are less informative about behaviour in the far upper tail. Criteria based on empirical quantiles are useful when sufficient data is available for their estimation because the approach focuses directly on the quantity of interest. We use both approaches in this paper, and in addition, we also focus on how the largest values get larger. To this end we use the (Fisher and Tippett 1928) approach to approximate series of distributions of block maxima by limiting forms satisfying the max-stability property as the block length increases. Other approaches may be possible, but have not been explored in this paper because of the availability of the large sample of annual maxima provided by the CanRCM4 large ensemble that we use. The evaluation and approximation of the shape of the upper right-hand tail of the extreme value distribution is an active research topic in the statistical literature of extreme values (Anderson 1971; Galambos 1978; Gomes 1984; Smith 1987). In the context of block maxima, these approximations could help avoid increasing block length and therefore lead to more efficient use of data. The approaches that have been studied include the use of a power series of scaling sequences (Zarfaty et al., 2021) and the use of additional criteria reflecting a second order regularization to characterize the rate of convergence towards a max-stable limit (Gomes et al., 2007; Gomesa and Martins 2002) as well as others that are discussed in many review papers (Beirlant et al., 2012; Gomes 1994; 2020; Gomes and Guillou 2015). Their application has, however, not yet been studied extensively in climatology and hydrology, where the parent process that produces daily values of the variable in question can be the result of a complex mixture of physical processes.

Max-stability, ie., regular tail behaviour for which the value of the shape parameter remains stable as block length is increased, is assumed when the GEV distribution to estimate long period return levels that correspond to points in the far upper right hand tail of the distribution of block maxima. We use a similar analysis strategy as was used to study the max stability of annual maximum 1-day precipitation in Ben Alaya et al. (2020). We fit GEV distributions at grid boxes to samples of block maxima for blocks of different lengths, ranging from 1-year to 20-years. The GEV distributions are fitted via the method of maximum likelihood, but results are insensitive to the use of another fitting method (not shown), such as the commonly used method of probability weighted moments (Hosking et al., 1985). The regularity (or lack of regularity) of tail behaviour is evaluated by studying changes in the GEV shape parameter as block length is varied and by comparing changes in estimated long-period return levels, such as the 1000-year level, that occur as block length is varied. Since the available sample of annual maximum wind speeds is very large, it is also possible to compare return level estimates derived from the fitted GEV distributions with empirical return level estimates obtained directly from the 2500-year samples. In addition, we assess the goodness of fit of the GEV distributions fitted to the very large samples of annual maxima, which provides another line of evidence that can be used to assess the GEV approximation the main body of the empirical distribution of annual maxima. We use the Anderson Darling (AD) goodness of fit test with unknown GEV parameters, for this purpose.

The bias in estimating very long return periods from the GEV fitted to annual maxima depends on estimation bias affecting the GEV shape parameter when the shape estimate differs from the true shape value representing the deep upper tail of the parent distribution for which tail stability is reached. This bias depends on two main aspects: i) the use of relatively short data records (Martins and Stedinger 2000) and ii) the possibility that the block length may not be high enough to ensure that convergence to a max-stable limit has effectively been achieved (Dombry 2015). In our approach, estimation bias affecting the shape parameter due to the first aspect should not be of great concern when using 1-year blocks because of the very large sample that is available for parameter estimation. This aspect of bias becomes a somewhat greater concern as block length is increased and thus we limit block length to no more than 20 years, which nevertheless allows a sample of 125 block maxima. Limiting the bias due to the first aspect by avoiding small samples allows us to focus on the variation of the shape parameter as block length increases to assess biases in long period return level estimations due to the second aspect.

Finally, we note a possible concern due to the fact the samples of block maxima that we use are dependent since they are nested (e.g., the sample of 10-year block maxima is a subsample of the sample of annual maxima). Nevertheless, the information extracted from fits of the GEV distribution to samples of block maxima is not strongly dependent for blocks of different length because the fit of the GEV to each sample is most strongly influenced by the central part of each sample; GEV-10 in effect gives relatively greater weight to the upper tail by considering only very large annual maxima.

3. Results

Fig. 2 shows maps of the estimated shape parameter of the GEV distribution fitted to block maxima of CanRCM4 simulated wind speed

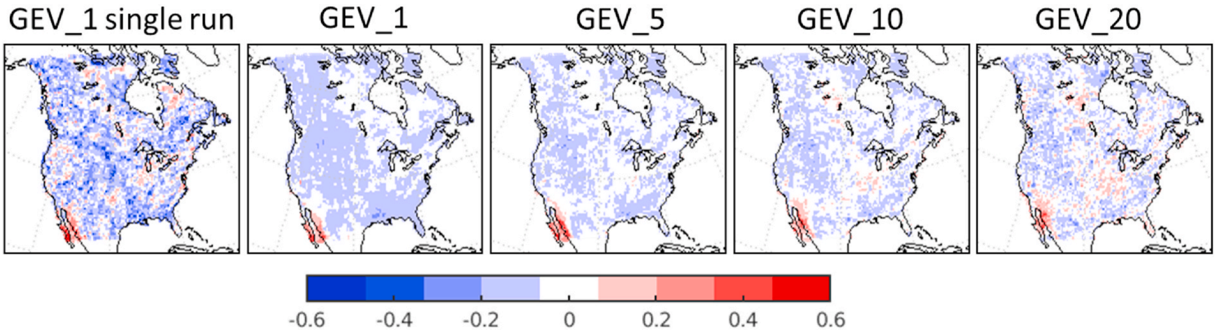


Fig. 2. Estimated shape parameter of the GEV distribution for CanRCM4 simulated annual maximum of instantaneous wind speeds for the historical period 1951–2000 over North America. The first column shows the shape parameter estimates when the GEV distribution is fitted to a sample of 50 annual maximum values from only the first of the 50 CanRCM4 simulation. The remaining columns shows shape parameter estimates using block maxima from the ensemble of 50 CanRCM4 simulations for 1-year (GEV-1; 2500 blocks), 5-year (GEV-5; 500 blocks), 10-year (GEV-10; 250 blocks) and 20-year blocks (GEV_20; 125 blocks).

for the historical period 1951–2000 over North America. The map obtained using a single 50-year CanRCM4 simulation is obviously noisy, which reflects the high uncertainty in shape parameter estimates based on 50-year samples. In contrast, the map obtained using the 2500 annual maxima pooled from the 50 ensemble members is smoother due to the substantial reduction of sampling uncertainty. Increasing the length of the blocks to 5, 10 and 20-years again increases the noise in the shape parameter map due to the declining number of available blocks. At first sight Fig. 2 does not seem to reveal strong evidence of substantial variation of the shape parameter with increasing block length. Nevertheless, small increases in the shape parameter from negative to near zero values that can be seen in grid boxes scattered over the domain, but mostly located in the southern and the central parts of the continent, which raises some doubt about the overall regularity of wind speed annual maxima.

Assessing the appropriateness of the GEV distribution for a given block length with a statistical goodness-of-fit (GOF) test may help to determine whether the series of block maxima exhibits regular behavior by evaluating the ability of the fitted model to approximate the main body of the distribution. We therefore use the Anderson Darling (AD) test to check if the samples of the 50 maxima (from the first single member) and 2500 (from the 50-member) maxima follow the fitted GEV distributions. All GOF tests are conducted at the 5% significance level at each grid box. Results, which are shown in Fig. 3, indicate considerable evidence of lack of fit when using the full 2500-year sample, suggesting that the annual maxima of wind speed simulated by CanRCM4 may indeed not be max-stable. Test results based on the fits of the samples of 50 annual maxima from the first CanRCM4 simulation to the GEV distribution are not informative (Fig. 3a). Rejection occurs at 1.4% of grid points, which is less than the nominal 5% significance level, perhaps partly because a sample of only 50 annual maxima does not provide enough information about the far upper tail to assess lack of fit. In contrast, we should expect a GOF test based on a large sample of 2500 annual maxima to have much higher power to detect lack of fit. In this case, the null hypothesis that these samples follow the GEV distribution (Fig. 3b) was rejected at approximately 20.5% of grid boxes, indicating considerable evidence of lack of fit and thus raising concerns that max-stability may not be valid. While the pattern of occurrence of rejection is quite noisy, some spatial organization is apparent, with the highest densities of occurrence being located in the central and southern parts of the continent, consistent with the change in shape parameter that occurs when increasing the block length from 1 year to 10 years (Fig. 3c). A way to measure the correspondence between these patterns is to calculate the mean absolute change in shape parameter at locations where rejection of the GOF occurs, and to compare that change with the mean absolute change in shape parameter that occurs at all other locations. We found

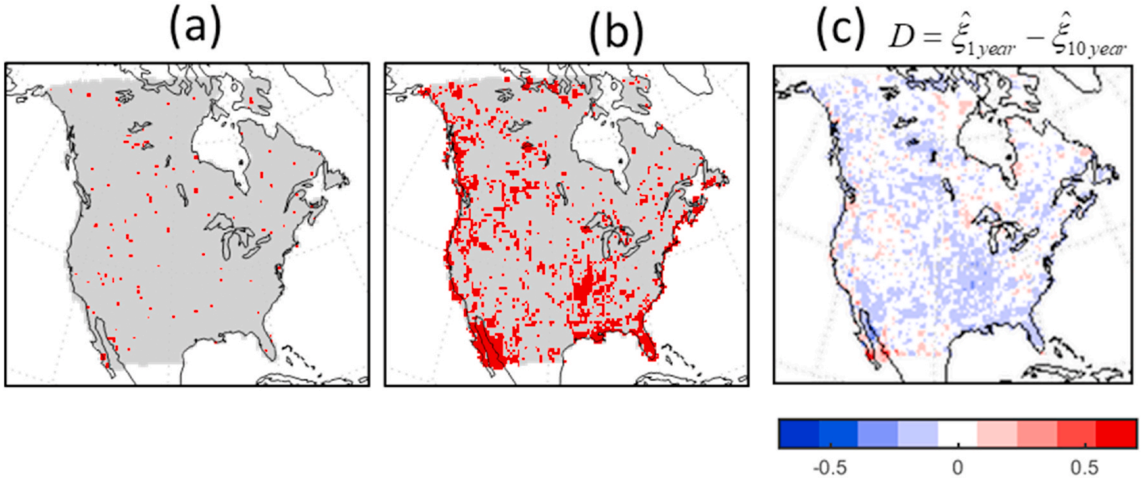


Fig. 3. Goodness of fit test results for each grid box over North America. Results of Anderson Darling tests of the null hypothesis that the samples of 50 annual maxima follow the fitted GEV distribution are shown in (a) and of the null hypothesis that the samples of 2500 annual maxima follow the GEV distributions fitted to those larger samples are shown in (b). Red points show grid boxes where the null hypothesis is rejected at 5% significance level. (c) Estimated difference between the shape parameters of GEV distributions fitted to annual wind speed maxima and those of GEV distributions fitted to 10-year wind speed maxima for the 50 CanRCM4 historical simulations of the period 1951–2000. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

the mean absolute change in shape parameter is 0.067 with a standard deviation of 0.09 at locations where rejection of the GOF test occurs. At all the other locations, the mean absolute change in shape parameter is 0.02 with a standard deviation 0.05. Change in the shape parameter when increasing block length is clearly larger where the fit is found to be poor. We can conclude therefore, that annual maximum wind speed is not max-stable over substantial parts of North America in the climate simulated by CanRCM4.

We next explore how the lack of max-stability affects the estimation of very long period wind speed return levels. We note that biases may result in two ways: (i) the well-known difficulty of estimating GEV parameters from small samples, and (ii) the possibility that the GEV approximation may not be appropriate because block maxima do not exhibit regular behaviour.

Fig. 4 shows plots of estimated return levels as a function of return periods at five different locations A, B, C, D and E highlighting different situations regarding the variation of the shape parameter with increasing block length. The geographical positions of the five points are presented in Fig. 4a. To show that the validity of max-stability assumption is very important for reliable high return level estimation, we need to substantially reduce return level estimation errors that are due to sampling uncertainty that affects the fitting of the GEV distribution. The 2500 annual maxima are used for this purpose. Thus, our interpretation of biases that are due to max-stability will be based on the red curves shown in Fig. 4, which correspond to results from the GEV fitted to the 2500 annual maxima. On the other hand, adding results from fitting the GEV to only 50 annual maxima (blue curves in Fig. 4) is helpful to see the real gain of using 2500 annual maxima in both parameter and RL estimates. The shading in Fig. 4 indicates 80% confidence intervals obtained by bootstrapping.

For location A, Fig. 4g illustrates a situation where the shape parameter decreases from a small positive value at annual maxima and attains negative values when using blocks of approximately 5 years or longer. This behaviour reflects the complex fluctuations in the upper tail of the distribution of extreme wind speeds at the location A, as can be seen from the empirical distribution function that is indicated by the dots shown in Fig. 4b. The transition from positive to negative shape parameters as the block length increases suggests that max-stability may not be satisfied. As can be seen from the blue curve in Fig. 4b, the GEV fitted to a sample of 50 annual maxima seems to overestimate high return levels. Over-estimation also seems to persist, but to a lesser extent, when increasing the sample size to 2500 annual maxima, even though the much larger sample substantially reduces the uncertainty of the GEV parameter estimates. This is consistent with the decrease of the shape parameter beyond annual block maxima in the wind speed CanRCM4 climate. Indeed, the distribution that is fitted for block lengths longer than about 5-years is bounded above, while the distribution fitted to annual maxima is mildly heavy tailed.

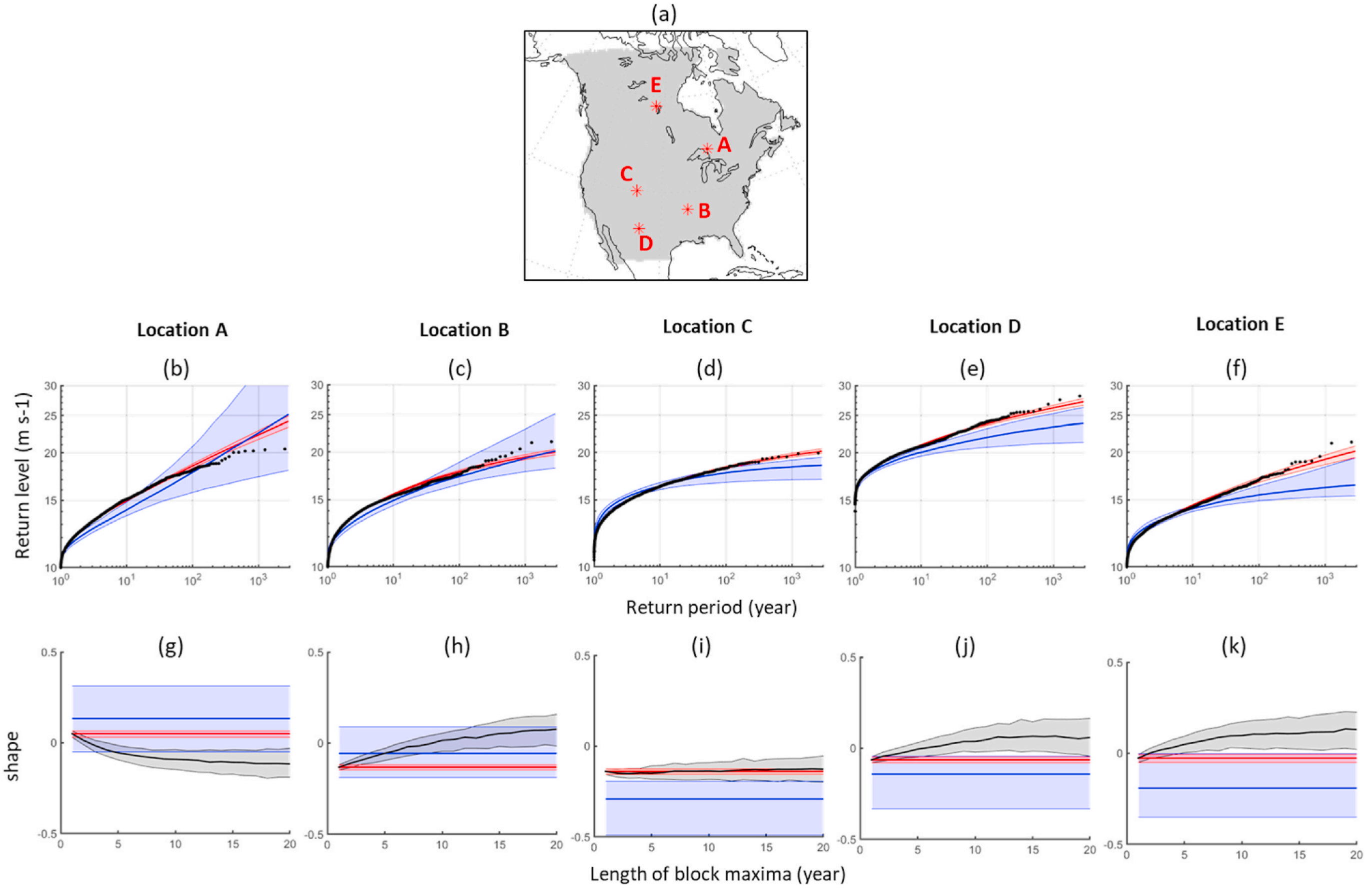


Fig. 4. Return level estimates based on fitting the GEV distribution to annual maxima of instantaneous wind speed (using the ML method) at five different locations A (in (b)), B (in (c)), C (in (d)), D (in (e)) and E (in (f)) using one CanRCM4 simulation of 1951–2000 (50 annual maxima, in blue) and the 50 simulations (2500 annual maxima, in red). Geographical positions of the five locations are shown in (a). Black dots in panels (b), (c), (d), (e) and (f) show empirical quantile estimates obtained using the 2500 annual maxima. Estimates of the shape parameter versus block length based on 2500 years of CanRCM4 simulations are shown by the black line for the five locations A (in (g)), B (in (h)), C (in (i)), D (in (j)) and E (in (k)). These panels also show estimated shape parameters based on annual maxima from a single CanRCM4 simulation (in blue) and the 50 ensemble members (in red), with the extension to longer blocks reflecting the max-stability assumption. Shading indicates 80% confidence intervals obtained by bootstrapping. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

In contrast to location A, the apparent underestimation of long period return levels when the GEV is fitted to the 2500 annual maxima at location B (Fig. 4c) can be explained by the increase in the shape parameter that occurs as the block length increases (Fig. 4h). In this case, a distribution with a finite upper bound is obtained when fitting the GEV to the annual maxima, while distributions that are heavy tailed are obtained when fitting to block maxima longer than about 5-years. In both cases, this behaviour is linked to the fluctuation of the high upper tail of the empirical distribution, which receives more weight in fitting the GEV when longer block lengths are used. These fluctuations are missed when only 50-years of data is used, and they do not receive sufficient weight to represent far upper tail behaviour well when annual maxima are used. Comparable behavior and results to those at location B, regarding the behaviour of the shape parameters are also seen at locations D and E.

Location C represents an example where there is little change in the shape parameter with increasing block length (Fig. 4i), and where the GEV distribution fitted to annual maxima corresponds well to the empirical distribution (Fig. 4d). Thus, max-stability for annual maxima is not contradicted. On the other hand, extreme quantiles are underestimated when using the GEV fitted to a sample of only 50 annual maxima (the blue curve in Fig. 4d), which simply reflects the large uncertainty in estimating GEV parameters from such a small sample. Repeating the exercise with a different 50-year sample could as easily have resulted in extreme quantile estimates that are overestimated. Evidently, bootstrap sampling of the small 50-year sample is inadequate to fully represent the uncertainty of the shape parameter estimate.

To further assess our previous premise about biases at the five locations, a bootstrap sampling approach is used to estimate the uncertainty in the empirical and GEV based RL estimates. Fig. 5 shows box-and-whisker plots of 1000 1000-year RL estimates of wind speed based on the GEV distributions fitted to each of 1000 bootstrap samples of the available 2500 annual maxima. Results are displayed for different block lengths at locations A-E. The median and the interquartile range of empirical RL estimates obtained from the 1000 bootstrap samples are plotted using black and dashed horizontal lines respectively for comparison. The results confirm the biases at locations A, B, D and E where the variations in estimated shape parameter with block length indicate that annual maxima lack of max-stability. For higher block length, however, results show that GEV estimates are more consistent with empirical estimates.

We now return to Fig. 3c, which shows a map of the differences between the shape parameter ξ_{1year}estimated by fitting GEV distributions to the sample of 2500 annual maxima, and the shape parameter ξ_{10year} estimated by fitting a GEV distribution to 10-year block maxima from the same sample. The absolute value of D gives an indication of the complexity of the upper tail behaviour of extreme wind speed. Negative values (shown in blue) suggest underestimation of very long period return levels when using the GEV fitted to annual maxima, while positive values suggest possible overestimation of high RLs. Fig. 3c shows that D is typically near to zero but with a tendency for D to be negative more often than positive, with larger negative values occurring mainly in the southern and central parts of the continent. These negative values raise concerns about possible underestimation of long period wind speed RLs. Furthermore, a puzzling local region with positive and negative values are seen over Baja California and coastal northwestern Mexico, suggesting similar concerns to location A and B respectively that has been presented in Fig. 4. One possible explanation could be a discrepancy between the land mask of the driving global model, CanESM2, which considers this region ocean, and that of the higher resolution CanRCM4 model, which treats the region as land.

Fig. 6 shows the relative differences in 100-, 1000- and 2000-year RL

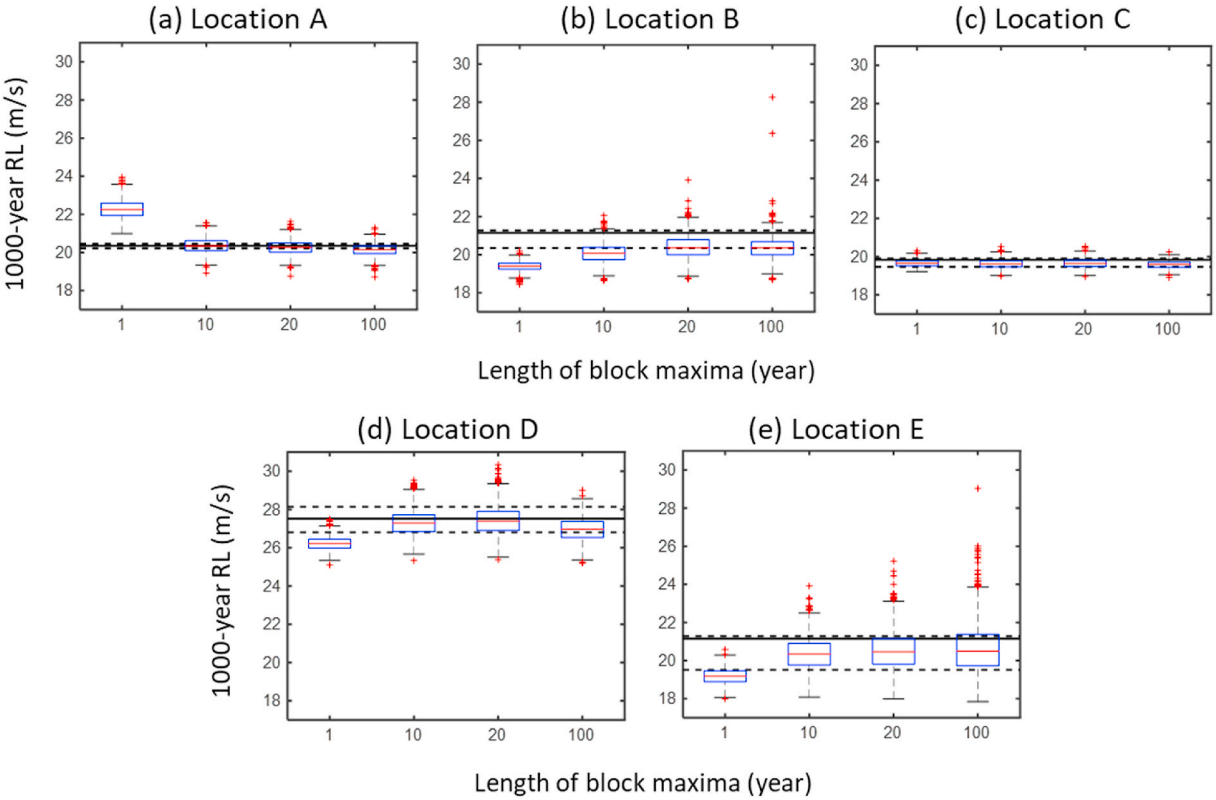


Fig. 5. Box-and-whisker plots of 1000 1000-year RL estimates of wind speed based on the GEV distributions fitted to each of 1000 bootstrap samples of 2500 annual maxima of CanRCM4 simulated wind speed. Results are displayed for different block lengths at the five different locations A, B, C, D and E (presented in Fig. 3a). GEV parameters are estimated using the ML method. Black and dashed horizontal lines show the median and the interquartile range of empirical RL estimates obtained from the 1000 bootstrap samples.

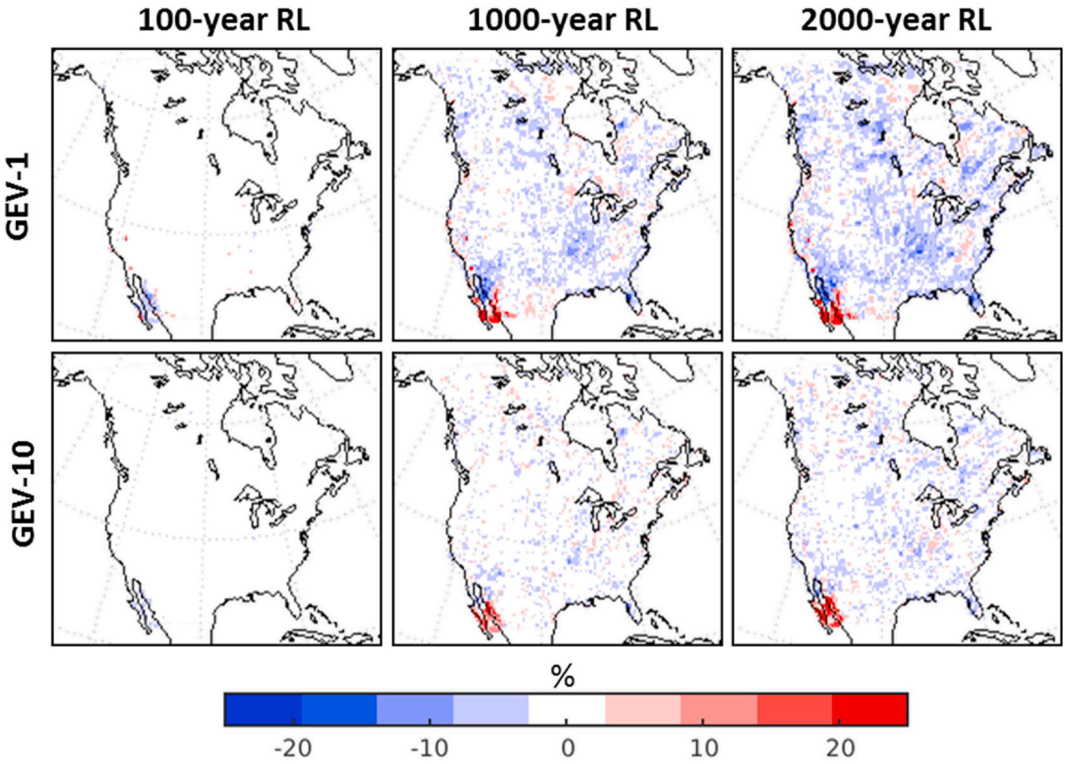


Fig. 6. Maps of the differences in % with empirical return levels estimates for GEV-1 distributions (GEV fitted to annual maximum wind speed) and GEV-10 distributions (GEV fitted to 10-year maxima of wind speed) using the 50 CanRCM4 historical simulations of the period 1951–2000. Bias is expressed in % relative to the corresponding empirical estimates.

estimates between estimates obtained directly from the empirical distribution and estimates from GEV distributions fitted to samples of 2500 annual maxima. In additions, the same differences when using the GEV fitted to samples of 250 10-year block maxima are also shown in Fig. 6. As can be seen, while the GEV distribution fitted to annual maxima provides relatively reliable estimates for the 100-year RL, the apparent differences in estimating the 1000- and 2000-year RLs cannot be neglected. As it is expected, the spatial pattern of D values seen in Fig. 3c is very similar to that of the bias in 1000- and 2000-year RL estimates. To illustrate this quantitatively, Fig. 7 displays a scatter plot that contrasts values at grid boxes with the percent bias in the estimated 1000-year RL. We see in this figure that an increase in the shape parameter when lengthening the block to 10-years tends to correspond to 1000-year return levels estimated from a GEV distribution fitted to annual maxima that are lower than empirical estimates. Conversely, a decrease in the shape parameter tends to correspond to values that are higher than empirical estimates. Careful application of the GEV by increasing the block length, however, effectively helps to substantially reduce the differences with empirical estimates as shown in Figs. 6 and 7. More generally, the deeper the extrapolation into the upper tail of the distribution, the larger the role of max-stability in providing reliable return level estimates.

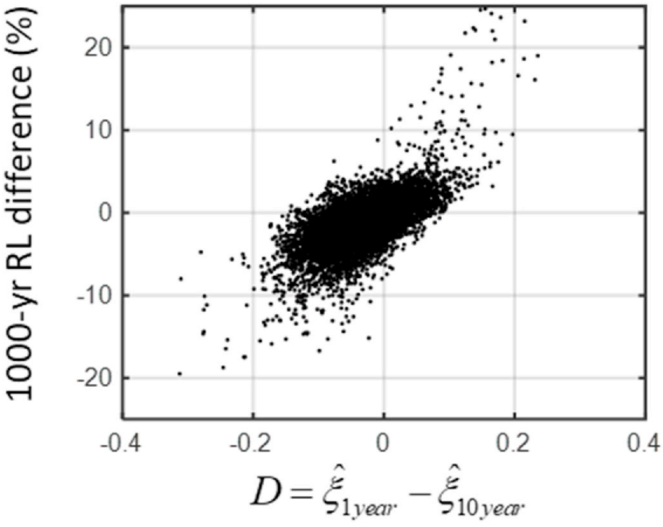


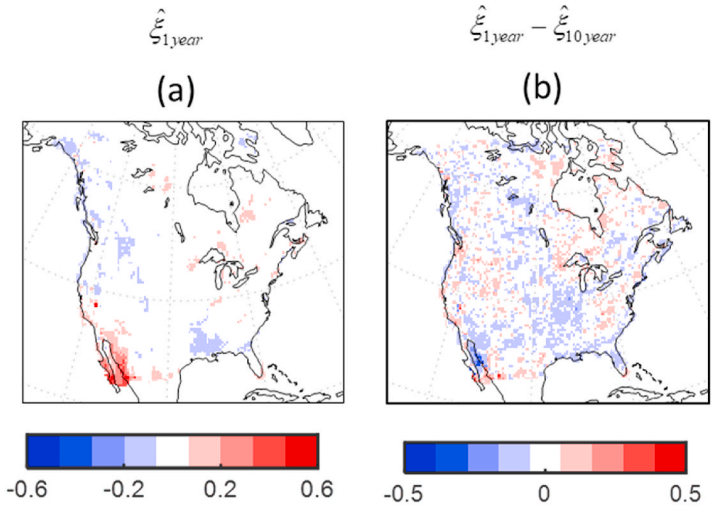
Fig. 7. Differences in GEV-based 1000-yr return level estimates relative to empirical return level estimates based on samples of 2500 annual maxima of CanRCM4 simulated wind speed shown as a function of the change D in the estimated GEV shape parameter when the distribution is fitted to 1- and 10-yr block maxima.

As mentioned in Section 2, wind loads on buildings are often given in pressure units that are proportional to wind speed squared, and thus a small bias in wind speed becomes more of an issue when wind pressure is of interest. Previous studies have recommended (Palutikof et al., 1999) that estimates of extreme wind loads should be obtained by analysing the extremes of squared wind speeds rather than by squaring the high quantiles of wind speed. For wind speed data the parent distribution is generally assumed to be Weibull (Justus et al., 1978; Lun and Lam 2000; Seguro and Lambert 2000), which belongs to the extremal Gumbel domain. That is, the distribution of block maxima of samples from a Weibull distribution converges to the Gumbel distribution (a GEV distribution with shape parameter $\xi = 0$) as the block length increases without bound. Empirical evidence suggests, however, that a GEV distribution with $\xi > 0$ may be more suitable for extreme wind speed than the Gumbel form (Perrin et al., 2006; Simiu and Heckert 1996), as we also find for the majority of locations, even when using 20-year blocks (Fig. 2). In this context, a number of authors have recommended (Cook 1986; Harris 1996, 2004) analysing values of wind pressure (i.e., wind speed squared) instead of directly analysing wind speed to enhance the chances of achieving rapid convergence toward the Gumbel. In reality the tail of the parent distribution is not known, and therefore the necessary conditions to ensure the convergence to a stable decay cannot be checked. Furthermore, the existence of a stable limit is not guaranteed by the extremal types theorem (Fisher-Tippett-Gnedenko Theorem (Fisher and Tippett 1928; Gnedenko 1943),). As can be seen in Fig. 8a, annual maxima that were previously better approximated by a GEV distribution with $\xi < 0$when using wind speeds, are more suitably approximated with a Gumbel distribution when using squared wind speed. While this slightly reduces the negative biases in 1000-year RLs in some locations, the approach also leads to worst results in many other locations by inducing substantial positive biases (see Fig. 8c, d and 8e) that are likely due to a change of the shape parameter beyond annual maxima (see Fig. 8b).

4. Discussion and conclusions

We have used a large ensemble simulation from the CanRCM4 regional climate model to obtain samples of 2500 wind speed annual maxima that are orders of magnitude larger than observed samples, and have used those samples to study whether the model simulated wind speed annual maxima are max-stable. Max-stability is the key assumption that allows extrapolation of extreme value distributions that are fitted to observed annual maxima beyond the range of those observations. This is of critical importance because the transition to so-called uniform risk engineering design that is now occurring requires return level estimates for extreme wind speeds for return periods ranging from 700 to 3000-years (ASCE/SEI 7–16, 2017; McAllister et al., 2018), while available observational records typically range between 10- and 50-years in length. Our findings are subject to the caveat that results are specific to the climate simulated by CanRCM4, which is certainly not identically the same as the observed climate. Nevertheless, if ‘surprises’ are found in the far upper tail of model simulated wind speed distribution, then we should expect that similar surprises may also exist in the upper tail of observed wind speed.

We found that CanRCM4 simulated wind speed annual maxima do appear to be approximately max-stable at the majority of model grid boxes over North America. Nevertheless, a substantial minority of model grid boxes have wind speed annual maxima that appear not to be max-stable, particularly in the southern and central parts of the continent. Generally, sampling deeper in the tail, for example, by using multi-year



Percent Biases (PB) in 1000-year RL squared wind speed

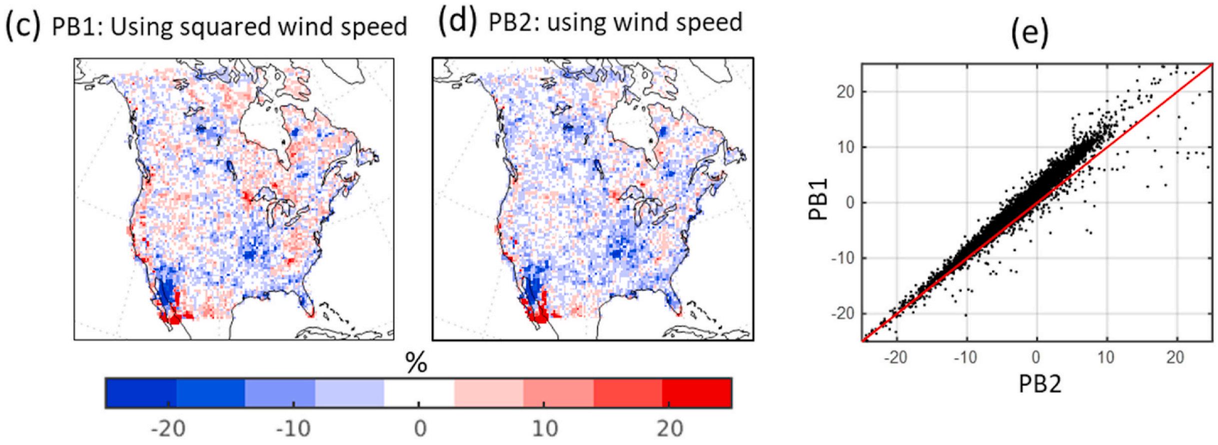


Fig. 8. Main results obtained by taking the square of annual maxima hourly mean wind speeds using the 50 CanRCM4 historical simulations of the period 1951–2000. (a) Maps of the shape parameter of the GEV fitted to squared wind speed annual maxima. (b) Estimated difference between the shape parameters of GEV distributions fitted to annual maximum squared wind speed and those of GEV distributions fitted to 10-year maxima of squared wind speed. (c) Maps of the percent bias, PB1, in 1000-year squared wind speed RLs estimated using GEV distributions fitted to annual maxima of squared wind speed. (d) Maps of the percent bias, PB2, in 1000-year squared wind speed RLs obtained by taking the square of 1000-year wind speed RLs estimated using the GEV directly fitted to wind speed annual maxima. The percent bias is calculated relative to the corresponding empirical estimates from the available 2500 annual maxima. (e) Percent bias PB1 as function of percent bias PB2.

blocks, led to GEV distribution fits that better represent the upper tail at locations lacking max-stability and provide long-period return level estimates with reduced bias compared to direct empirical estimates from the climate model output. In the climate of CanRCM4, most locations that are found to lack max-stability have wind speed annual maxima with GEV shape parameters that increase as block length increases beyond one year, shifting from fitted GEV distributions for annual maxima that are mostly light-tailed (negative shape parameter) to fitted distributions apparently converging to the Gumbel distribution (close to zero shape parameter) as block length increases. As a result, the GEV fitted to annual maxima might often underestimate very long period (e.g., 1000-year) return levels. It does, however, produce quite reasonable estimates of return levels for shorter (e.g., 100-year) return periods.

Unfortunately, the instrumental observational record is insufficient to similarly assess whether max-stability holds in the observed climate. Nevertheless, the potential for a lack of max-stability in the real world is almost certainly greater than in the climate simulated by CanRCM4 given that it does not simulate the full range of physical phenomena, such as tornadoes, that produce extreme winds. This potential has been well recognized by practitioners, who have developed and apply a number of techniques that attempt to account for the different types of mechanisms that may produce extreme wind loads (Gomes and Vickery 1978; Harris 2017; Zhang et al., 2018). Such techniques often involve the classification of extreme wind speed events into different types, resulting in subsetting of the limited observational resource. Indeed, Twisdale and Vickery (1992) warn against methods for estimating design wind speeds that use mixed observations coming from separate statistical distributions corresponding to different meteorological phenomena. This occurs not only in areas affected by tropical cyclones, but also in other areas that experience *"mixed wind climates"*, such as regions where strong winds might originate from either large-scale extratropical cyclones or small but intense thunderstorms. Brabson and Palutikof (2000) showed using properties of the generalized Pareto distribution (GPD) that variation of the shape parameter as a function of the threshold may be indicative of a mixed climate. They suggest that the GPD shape parameter should be constant with the threshold for a homogeneous data set generated by a single mechanism. This is analogous to our interpretation that is based on the BM approach shape parameter in which the location parameter plays the same role as the threshold in the POT method. With either approach, reliable estimation of quantiles deep in the upper tail of wind speed distribution requires a sampling strategy that samples extreme winds from the mechanism that produces the strongest winds experienced at a given location sufficiently often so that the upper tail can be well represented by the fitted model. Sampling more deeply in the tail by increasing block length appears to approach a point in the tail where samples are approximately max-stable, and thus achieves that goal, provided large enough samples are available (e.g., from a climate or weather model that is known to be reliable). It remains to be determined whether other approaches that distinguish between different mechanisms can achieve the same goal, and whether that can be done with available observational records.

By quantifying the impact of lack of max-stability on quantiles, we reiterate via our study that this should be given careful consideration particularly when ambitious extrapolation deep into the upper tail is required. While increasing the block length to 10-years or more should help to reduce biases due to lack of max-stability, this is not a solution that practitioners would be able to apply, for example, in a context where observational records are only a few decades in length. Thus, for the moment, this may mean continuing to rely on solutions that involve the use of a mixture of distributions or involve the incorporation of covariates.

CRediT authorship contribution statement

M.A. Ben Alaya: Conceptualization, Methodology, Formal analysis, Writing – original draft, Writing – review & editing. **F.W. Zwiers:** Methodology, Formal analysis, Writing – review & editing. **X. Zhang:** Formal analysis, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

M.A. Ben Alaya was supported by the Climate Related Precipitation Extremes project of the Global Water Futures program. We also very much appreciate the assistance provided by Guilong Li in providing large ensemble CanRCM4 wind speed data. We thank two anonymous reviewers for constructive comments that helped to improve this paper.

References

Anderson, C., 1971. Contributions to the Asymptotic Theory of Extreme Values.

ASCE 7 -16, A. S., 2017. Minimum Design Loads and Associated Criteria for Buildings and Other Structures. American society of civil engineers.

Gomes, M.I., 1994. Penultimate Behaviour of the Extremes. *Extreme Value Theory and Applications*. Springer, pp. 403–418.

Harris, R., 2004. Extreme value analysis of epoch maxima—convergence, and choice of asymptote. J. Wind Eng. Ind. Aerod. 92, 897–918.

Gomes, M.I., 2020. Revisiting Rates of Convergence and Penultimate Approximations for Extremes. Annales Univ. Sci. Budapest., Sect. Comp, pp. 135–149.

Beirlant, J., Caeiro, F., Gomes, M.I., 2012. An overview and open research topics in statistics of univariate extremes. Rev. Stat. 10, 1–31.

Ben Alaya, M.A., Zwiers, F.W., Zhang, X., 2020. An evaluation of block-maximum-based estimation of very long return period precipitation extremes with a large ensemble climate simulation. J. Clim. 33, 6957–6970.

Brabson, B., Palutikof, J., 2000. Tests of the generalized Pareto distribution for predicting extreme wind speeds. J. Appl. Meteorol. 39, 1627–1640.

Cannon, A.J., Jeong, D.I., Zhang, X., Zwiers, F.W., 2020. Climate-Resilient Buildings and Core Public Infrastructure: an Assessment of the Impact of Climate Change on Climatic Design Data in Canada978-0-660-36478-0, p. 105.

Cheng, E., Yeung, C., 2002. Generalized extreme gust wind speeds distributions. J. Wind Eng. Ind. Aerod. 90, 1657–1669.

Cook, N.J., 1986. Designers Guide to Wind Loading of Building Structures.

CSA, 2019. CAN/CSA-S6-19, Canadian Highway Bridge Design Code. CSA, Mississauga, Ont.

Dombry, C., 2015. Existence and consistency of the maximum likelihood estimators for the extreme value index within the block maxima framework. Bernoulli 21, 420–436.

Fisher, R.A., Tippett, L.H.C., 1928. Limiting forms of the frequency distribution of the largest or smallest member of a sample. Mathematical Proceedings of the Cambridge Philosophical Society. Cambridge University Press, pp. 180–190.

Galambos, J., 1978. The Asymptotic Theory of Extreme Order Statistics.

Giorgi, F., Jones, C., Asrar, G.R., 2009. Addressing climate information needs at the regional level: the CORDEX framework. World Meteorol. Organ. Bull. 58, 175.

Gnedenko, B.V., 1943. Sur la distribution limite du terme maximum of d'unesérie Aléatorie. Ann. Math. 44, 423–453.

Gomes, M.I., 1984. Penultimate limiting forms in extreme value theory. Ann. Inst. Stat. Math. 36, 71–85.

Gomes, M.I., Guillou, A., 2015. Extreme value theory and statistics of univariate extremes: a review. Int. Stat. Rev. 83, 263–292.

Gomes, L., Vickery, B., 1978. Extreme wind speeds in mixed wind climates. J. Wind Eng. Ind. Aerod. 2, 331–344.

Gomes, M.I., Martins, M.J., Neves, M., 2007. Improving second order reduced bias extreme value index estimation. Rev. Stat. 5, 177–207.

Gomesa, M.I., Martins, M.J., 2002. "Asymptotically unbiased" estimators of the tail index based on external estimation of the second order parameter. Extremes 5, 5–31.

Gumbel, E.J., 1958. Statistics of Extremes. Columbia University Press, p. 375.

Harris, R., 1996. Gumbel re-visited-a new look at extreme value statistics applied to wind speeds. J. Wind Eng. Ind. Aerod. 59, 1–22.

Harris, R.I., 2017. The Level Crossing Method applied to mean wind speeds from "mixed" climates. Struct. Saf. 67, 54–61.

Holmes, J.D., 2018. Wind Loading of Structures. CRC press.

Hosking, J.R.M., Wallis, J.R., Wood, E.F., 1985. Estimation of the generalized extreme-value distribution by the method of probability-weighted moments. Technometrics 27, 251–261.

Huang, W.K., Stein, M.L., McInerney, D.J., Sun, S., Moyer, E.J., 2016. Estimating changes in temperature extremes from millennial-scale climate simulations using generalized extreme value (GEV) distributions. Adv. Stat. Clim. Meteorol. Oceanogr. 2, 79–103.

Huang, W.K., Monahan, A.H., Zwiers, F.W., 2020. Estimating Concurrent Climate Extremes: A Conditional Approach. Weather and Climate Extremes.

Jain, A., Srinivasan, M., Hart, G.C., 2001. Performance based design extreme wind loads on a tall building. Struct. Des. Tall Build. 10, 9–26.

Jeong, D.I., Cannon, A.J., Morris, R.J., 2020. Projected changes to wind loads coinciding with rainfall for building design in Canada based on an ensemble of Canadian regional climate model simulations. Climatic Change 1–15.

Justus, C., Hargraves, W., Mikhail, A., Graber, D., 1978. Methods for estimating wind speed frequency distributions. J. Appl. Meteorol. 17, 350–353.

Klemp, J.B., 1987. Dynamics of tornadic thunderstorms. Annu. Rev. Fluid Mech. 19, 369–402.

Li, C., Zwiers, F., Zhang, X., Li, G., 2019. How much information is required to well constrain local estimates of future precipitation extremes? Earth's Future 7, 11–24.

Lun, I.Y., Lam, J.C., 2000. A study of Weibull parameters using long-term wind observations. Renew. Energy 20, 145–153.

Martins, E.S., Stedinger, J.R., 2000. Generalized maximum-likelihood generalized extreme-value quantile estimators for hydrologic data. Water Resour. Res. 36, 737–744.

McAllister, T.P., Wang, N., Ellingwood, B.R., 2018. Risk-Informed Mean Recurrence Intervals for Updated Wind Maps in ASCE 7-16. Journal of structural engineering, New York, NY), p. 144.

NRCA, 2015, 2015. National Building Code of Canada. National Research Council Canada.

Palutikof, J., Brabson, B., Lister, D., Adcock, S., 1999. A review of methods to calculate extreme wind speeds. Meteorol. Appl. 6, 119–132.

Perrin, O., Rootzén, H., Taesler, R., 2006. A discussion of statistical methods used to estimate extreme wind speeds. Theor. Appl. Climatol. 85, 203–215.

Petersen, E.L., Mortensen, N.G., Landberg, L., Højstrup, J., Frank, H.P., 1998. Wind power meteorology. Part II: siting and models. Wind Energy: An International Journal for Progress and Applications in Wind Power Conversion Technology 1, 55–72.

Reiss, R.-D., Thomas, M., Reiss, R., 2007. Statistical Analysis of Extreme Values, vol. 2. Springer.

Scinocca, J., et al., 2016. Coordinated global and regional climate modeling. J. Clim. 29, 17–35.

Seguro, J., Lambert, T., 2000. Modern estimation of the parameters of the Weibull wind speed distribution for wind energy analysis. J. Wind Eng. Ind. Aerod. 85, 75–84.

Simiu, E., Heckert, N., 1996. Extreme wind distribution tails: a "peaks over threshold" approach. J. Struct. Eng. 122, 539–547.

Smith, R.L., 1987. Approximations in Extreme Value Theory.

Twisdale, L.A., Vickery, P.J., 1992. Research on thunderstorm wind design parameters. J. Wind Eng. Ind. Aerod. 41, 545–556.

von Salzen, K., Coauthors, 2013. The Canadian fourth generation atmospheric global climate model (CanAM4). Part I: representation of physical processes. Atmos.-Ocean 51, 104–125.

Zarfaty, L., Barkai, E., Kessler, D.A., 2021. Accurately approximating extreme value statistics. J. Phys. Math. Theor. 54.

Zhang, S., Solari, G., Yang, Q., Repetto, M.P., 2018. Extreme wind speed distribution in a mixed wind climate. J. Wind Eng. Ind. Aerod. 176, 239–253.