Biases in estimating long-term recurrence intervals of extreme events due to regionalised sampling

Moutassem El Rafei[1], Steven Sherwood[2], Jason Peter Evans[2], Andrew J. Dowdy[3], and Fei Ji[4]

[1]Climate Change Research Centre
[2]University of New South Wales
[3]Bureau of Meteorology (Australia)
[4]NSW Department of Planning and Environment

March 9, 2023

Abstract

Preparing for environmental risks requires estimating the frequencies of extreme events, often from data records that are too short to confirm them directly. This requires fitting a statistical distribution to the data. To improve precision, investigators often pool data from neighboring sites into single samples, referred to as “superstations,” before fitting. We demonstrate that this technique can introduce unexpected biases in typical situations, using wind and rainfall extremes as case studies. When the combined locations have even small differences in the underlying statistics, the regionalization approach gives a fit that may tend toward the highest levels suggested by any of the individual sites. This bias may be large or small compared to the sampling error, for realistic record lengths, depending on the distribution of the quantity analysed. The results of this analysis indicate that previous analyses could potentially have overestimated the likelihood of extreme events arising from natural weather variability.

Biases in estimating long-term recurrence intervals of extreme events due to regionalised sampling

Moutassem El Rafei [1], Steven Sherwood [1,2], Jason Evans [1,2], Andrew Dowdy [3] and Fei Ji [4]

[1]Climate Change Research Centre, University of New South, Wales, Sydney, Australia
[2]ARC Centre of Excellence for Climate Extremes, University of New South Wales, Sydney, Australia
[3]Australian Bureau of Meteorology
[4]NSW Department of Planning and Environment, Sydney, Australia

Key Points:

- Grouping data of nearby locations into one larger sample or "superstation" can induce biases at long recurrence intervals
- The superstation fit tends to the highest levels suggested by any of the pooled locations
- The bias may be large or small compared to random uncertainty, depending on the distribution of the extreme event analysed

Corresponding author: Moutassem El Rafei, m.el_rafei@unsw.edu.au

Abstract

Preparing for environmental risks requires estimating the frequencies of extreme events, often from data records that are too short to confirm them directly. This requires fitting a statistical distribution to the data. To improve precision, investigators often pool data from neighboring sites into single samples, referred to as "superstations," before fitting. We demonstrate that this technique can introduce unexpected biases in typical situations, using wind and rainfall extremes as case studies. When the combined locations have even small differences in the underlying statistics, the regionalization approach gives a fit that may tend toward the highest levels suggested by any of the individual sites. This bias may be large or small compared to the sampling error, for realistic record lengths, depending on the distribution of the quantity analysed. The results of this analysis indicate that previous analyses could potentially have overestimated the likelihood of extreme events arising from natural weather variability.

Plain Language Summary

We report a previously unknown bias in a common method for estimating how often extremely rare events such as extreme wind bursts or rain events will occur, when return periods are longer than the available data record. The method analysed is one where an investigator combines data from nearby locations to reduce sampling error. We find by looking at new, high-resolution data that variations in behavior across sites can sometimes produce biases much larger than the sampling error. The implication is that some observed extreme events are even less likely to have occurred than previously thought, assuming the underlying distribution hasn't changed over the period of observation.

1 Introduction

The statistical analysis of extreme-event frequencies and intensities is important to many risk management problems. For example, estimating the appropriate design wind speed requires the statistical analysis of historical wind data to estimate the strongest wind that might occur over a long time interval (AS/NZS1170.2:2021, 2021; Holmes, 2002; El Rafei et al., 2022). The design of offshore and coastal marine structures is governed by statistical estimates of extreme waves (Gulev & Grigorieva, 2004; Meucci et al., 2020). Similarly, statistical estimates of extreme rainfall values are essential for calculating flood risk and designing stormwater infrastructure (Green et al., 2012; Johnson & Green, 2018). To meet this need, extreme value theory has been widely used to estimate the probability of events larger than any on record so far (Brabson & Palutikof, 2000; Coles, 2001; Church et al., 2006; Wang et al., 2013).

The two usual approaches of extreme value theory are the generalized extreme value distribution (GEV) and the generalized Pareto distribution (GPD). The GEV approach, which combines three different statistical families (Weibull, Gumbel and Frechet), uses block maxima in which the dataset is divided into blocks and the maximum over each block is modelled (Gumbel, 1958; Palutikof et al., 1999; Coles, 2001). The GPD approach is instead based on the peaks-over-threshold method for which a threshold value is specified and all the values above this chosen threshold are used to fit the model (Pickands, 1975; Coles, 2001; Holmes & Moriarty, 1999). No matter which approach is used, the extrapolation to very rare events is subject to significant sampling errors when using short data ranges and uncertainties are unavoidable as accurate observational records are commonly short and/or geographically sparse.

To reduce statistical uncertainty, regionalization techniques have been used, whereby a larger sample is created by combining independent records of neighboring stations (J. Peterka, 1992; J. A. Peterka & Shahid, 1998; Holmes, 2002; Wang et al., 2013; Holmes, 2019) into what is sometimes called a "superstation". For example, regionalization has been used for extreme wind assessment in the United States (J. A. Peterka & Shahid, 1998; ANSI/AS CE 7-98, 1998; ASCE/SE I 7-16, 2016) and Australia (Holmes, 2002; AS/NZS1170.2:2021, 2021) to specify single design wind speed by compositing data from multiple stations; for regional flood frequency estimates (Haddad & Rahman, 2012); and for Intensity-Frequency-Duration (IFD) rainfall relationships (Wallis et al., 2007; Norbiato et al., 2007; Green et al., 2012; Johnson & Green, 2018). For IFD applications, nearby stations are pooled together assuming they share a common distribution of rainfall and are independent. Another place where this approach has been used is for so-called "regional frequency analysis" of extreme wave heights, where data from sites with similar wave statistics are used to estimate the distribution for a presumed homogeneous region (Van Gelder et al., 2001; Bernardara et al., 2011; Lucas et al., 2017).

While regionalization allows the estimation of distribution parameters using a larger dataset, the biases of this strategy are not explicitly quantified in the literature. Here we report unexpected biases in estimating long-term recurrence intervals of extreme events via regionalization, considering wind and rainfall extremes as case studies.

2 Data and Methods

2.1 Wind Data and Distribution Parameters

Wind data are obtained from the 1.5 km Bureau of Meteorology Atmospheric high-resolution Regional Reanalysis for Australia Sydney region (BARRA-SY) and cover the period from 1996 to 2019. This fine-scale reanalysis may provide better representation of storm systems and mesoscale phenomena, which can allow better modelling and higher accuracy of the wind field compared to lower resolution models. The reader is referred to (Jakob et al., 2017; Su et al., 2019; Su et al., 2021) for more details.

The maximum hourly wind gust speed data was used to calculate the daily maxima, which are then used to fit the GPD model. For computational tractability, we subsampled the BARRA-SY data to 5324 grid points covering New South Wales (NSW), with a spacing of approximately 10 km. This is judged to be justified because gusts at locations that are very close (i.e., 1.5 km apart) would capture the same gust events multiple times hence not be independent. The domain is then decomposed into groups (superstations) of approximately 25 grid points (5×5 grid points neighborhood) such that approximately 575 station years are obtained for each superstation.

The fully heterogeneous numerical experiment (Fig. 2B) considers threshold values between 15 and 23 m/s, shape factors between -0.2 to -0.1 and scale parameters ranging from 2 to 3. These ranges are representative of wind gust GPD distributions obtained using the BARRA-SY dataset. The partially heterogeneous case (Fig. 2B) considers a pre-fixed shape factor of -0.1 (similarly to what is used in the Australian standards) and a threshold value of 20 m/s. The range of scale factor values (between 3 and 4) was arbitrarily selected to represent a gradient of this parameter between the pseudo-stations.

2.2 Rainfall Parameters

To provide parameter values for rainfall tests, 30 years (1992-2021) of daily rainfall data were obtained from the Australian Bureau of Meteorology for five locations in the Sydney area (Sydney Airport, Sans Suci, Randwick, Rose Bay and Peakhurst) providing good data quality, using only low elevation stations ($\leq$ 100 m). Incorrect values (flagged as wrong via the quality checks) were removed, and the annual maxima were used to fit a GEV model at each station. The GEV analysis suggests threshold values ranging from 66.2 to 77.3, shape factors ranging from 0.08 to 0.27 and scale factors ranging from 22.7 to 25.3. These ranges were then used to generate synthetic random samples for the fully heterogeneous scenario (Fig. 4B). The partially heterogeneous scenario considers pre-fixed threshold and shape factor ($u_0 = 77.3$ and $\xi = 0.19$), which are selected from one of the stations distributions, and a range of scale factors from 17 to 25. The range of scale factors is randomly selected for this scenario to represent a gradient of this parameter.

2.3 Extreme Value Theory

The GEV approach uses the block maxima technique where the maximum yearly value is considered for the fit. The GEV cumulative distribution function can be expressed as follows:

$$G(u_g) = \begin{cases} e^{-\left[1+\xi_g\left(\frac{u_g-u_{0g}}{\sigma_g}\right)\right]^{-1/\xi_g}}, & \text{for } \xi_g \neq 0 \\ e^{-e^{(u_g-u_{0g})/\sigma_g}}, & \text{for } \xi_g = 0 \end{cases}$$

where u_{0g}, ξ_g and σ_g are the location, shape and scale parameters respectively. The return levels are given in terms of ARI as:

$$U_{Rg} = u_{0g} - \frac{\sigma_g}{\xi_g}[1 - (R)^{\xi_g}] \tag{1}$$

where R is the ARI.

The GPD model is based on the peaks-over-threshold approach (Palutikof et al., 1999; Coles, 2001) where all the data above a specified threshold are modelled. The GPD distribution can be expressed as:

$$H(u_p) = \begin{cases} 1 - \left(1 + \xi_p \frac{u_p - u_{0p}}{\sigma_p}\right)^{-1/\xi_p}, & \text{for } \xi_p \neq 0 \\ 1 - e^{\frac{u_p - u_{0p}}{\sigma_p}}, & \text{for } \xi_p = 0 \end{cases}$$

where u_{0p}, ξ_p and σ_p are the threshold, shape and scale parameters respectively. This expression is defined on $\{u_p - u_{0p} > 0$ and $(1 + \xi_p(u_p - u_{0p})/\tilde{\sigma_p}) > 0\}$ where $\tilde{\sigma_p} = \sigma_p + \xi_p(u_p - u_{0p})$.

The return levels are given in terms of the average recurrence interval using:

$$U_{Rp} = u_{0p} - \frac{\sigma_p}{\xi_p}[1 - (\lambda R)^{\xi_p}] \tag{2}$$

where λ is the number of crossings of the threshold per year.

The rainfall estimates calculated using the GPD approach use the 99$^{\text{th}}$ percentile as a threshold value (Lazoglou & Anagnostopoulou, 2017). For the wind analysis, we used a threshold selection algorithm that selects an optimal threshold value at each location. The selection algorithm first classifies the gusts into convective (i.e., thunderstorms) and synoptic (e.g., east coast lows, frontal systems) events. This is because gusts produced by different mechanisms can have different statistical properties and distributions. A range of thresholds is then tested to select the best fit. The reader is referred to (El Rafei et al., 2022) for more details on the threshold selection algorithm and the storm classification technique. The current study only presents results of convective events, as the same bias occur with either type using the superstation technique.

3 Results

3.1 Extreme Wind Gust Example

3.1.1 Reanalysis Data Study

We begin by examining a high-resolution (1.5-km horizontal grid spacing) regional Australian reanalysis dataset (BARRA-SY) of 23 years length from Year 1996 to Year

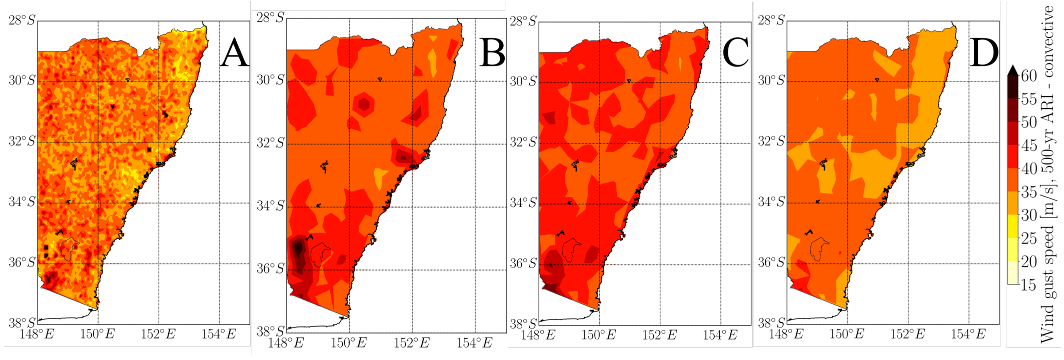

Figure 1: 500-year return wind gust speeds from convective events estimated several ways. Results are calculated using: A) the GPD algorithm applied at each grid point. B) the "superstation" approach to aggregate data from neighboring grid points. C) the nearby 90$^{\text{th}}$ percentile and D) mean of values shown in panel A, computed in the same neighborhood as the "superstations".

2019, using the GPD approach. The availability of this new high-resolution gridded dataset has motivated this study, since wind data are available at sufficient density to directly test the performance of the regionalization approach. The GPD method and reanalysis data are detailed in the Data and Methods section. Maps of 500-year convective (i.e., thunderstorm) wind gust speeds calculated in several ways are shown in Fig. 1. We show the 500-year ARI as this is used for the design criteria of many buildings in Australia (Wang et al., 2013). Results of higher recurrence intervals (e.g. 1000- and 2000-year ARIs, not shown) present similar patterns.

The results obtained by GPD fitting and calculating return levels independently at each grid cell (Fig. 1A) show speckling due to small differences in the estimated GPD parameters. We expect this is mainly from sampling error due to the short record length, as it looks random rather than geophysical. The regionalization approach (Fig. 1B), in which data from 25 adjacent grid cells are combined before fitting, yields smoother results but they show higher wind gust levels everywhere by about 10% compared to Fig. 1A. They are also higher than when the grid-point results are instead spatially smoothed by taking neighborhood means (2-D boxcar mean smoother of 5$\times$5 grid points), as shown in Fig. 1D. Indeed they are very close to the neighborhood 90$^{\text{th}}$ percentile results in Fig. 1C (calculated as the 90$^{\text{th}}$ percentile wind gust speed of 25 adjacent grid points). This is true even for regional maxima in neighborhood spread such as in the southeast corner of the state.

This analysis suggests that the superstation technique gives a fit that is close to the highest levels suggested by any of the neighboring, noise-influenced, sites. We now examine two hypotheses for why this is happening: first, that we bias the return levels by using short data ranges; and second, that it is an effect of combining locations into a superstation.

3.1.2 Simulation Experiments

A set of numerical experiments have been carried out to test these hypotheses, by generating synthetic records from pseudo-stations. Five synthetic "neighbor" data records are analytically generated from assumed GPD distributions with parameters based on wind gust distributions estimated using the BARRA-SY reanalysis, in two test scenarios. For the first, partially heterogeneous scenario (Fig. 2A,C), the five records are generated using the same threshold ($u_0 = 20$ m/s) and shape factor ($\xi = -0.1$) but a range of scale factors from 3-4. Pre-fixing the threshold and shape factor is similar to to approaches sometimes used in structural design standards (AS/NZS1170.2:2021, 2021). For the second, fully heterogeneous scenario (Fig. 2B,D), all the distribution parameters vary among the records.

We first examine outcomes with long records (Figs. 2A and 2B), generating 1000 years of data, with a total of 5000 data points used by the GPD model, for each record (i.e., five threshold exceedence events per year on average). Figs. 2A and B show that, even when differences in the underlying distributions between the locations are small, the superstation fit is higher than the mean and the median of the individual fits and tends to the highest levels suggested by the individual stations, which is consistent with the wind data results in Fig. 1. Furthermore, in both simulations the bias increases at longer recurrence intervals. The superstation bias in the fully heterogeneous scenario is more significant, even at short recurrence intervals, because some locations are contributing a lot more events than others to the superstation. This is not seen in the partially homogeneous scenario where all station distributions have the same threshold and shape factor and hence contribute similarly.

We next consider the effect of sample size by considering dataset lengths that range from 30 to 100000 years. For each length we repeated the test 1000 times to yield a PDF, the mean and 90$^{\text{th}}$ percentile of which is shown in Figs. 2C,D. The sampling uncertainty monotonically reduces for longer datasets and gradually converges to the true superstation bias. At 30 and 50 year record lengths (typical of real-world datasets), the bias may be exceeded in magnitude by the random error as depicted by the large spread of the PDF. Importantly however, all PDFs are centred on the true bias, showing that short record lengths do not cause biases, only random sampling errors. The PDFs of error (Fig. 2C,D) are narrow in both simulations when datasets are longer than 1000 years, implying a consistent bias site to site (or realisation to realisation), as implied by the geographic uniformity of the difference between Fig. 1B vs. D. Moreover, as seen before, the simulated superstation result is higher than the mean return level of the stations in the neighborhood. Hence we conclude from these tests that while the noise seen in the BARRA-SY return-period map is from sampling errors due to the short record, the ubiquitous bias toward high values is caused by regional pooling of data from nearby locations.

To understand what gives rise to this systematic bias, we compared the PDFs of gust speed (Figs. 3A and B) corresponding to the exceedence curves shown in Fig. 2A, B. In both scenarios, the slope of the superstation gust PDF at high gust speed tends to be dictated by the stations that have the heaviest tail (i.e., where the most extremes are recorded hence contributing most heavily to the superstation sample). Moreover, this phenomenon increases as one goes farther out on the tail of the PDF; if for example a very high threshold is used, nearly all data meeting the threshold come from one station (open circles in Fig. 3B). This level-dependent bias imparts a shallower slope to the PDF tail, which means that when calculating very long recurrence intervals, the extrapolation would tend to levels suggested by the locations that experience the most extreme events, or possibly even higher.

3.2 Rainfall Example

We now repeat the above analysis for another type of natural hazard, extreme rainfall, to explore the generality of the result. In this case rather than GPD we use GEV, since this is how published rainfall estimates are typically calculated, but this choice does not substantially affect results (Data and Methods section). The synthetic data records are generated as before, considering the same two scenarios except for the parameter values. The range of rainfall scale factors varies from 17 to 25, based on distributions estimated from observed daily accumulated rainfall from weather stations in Sydney area.

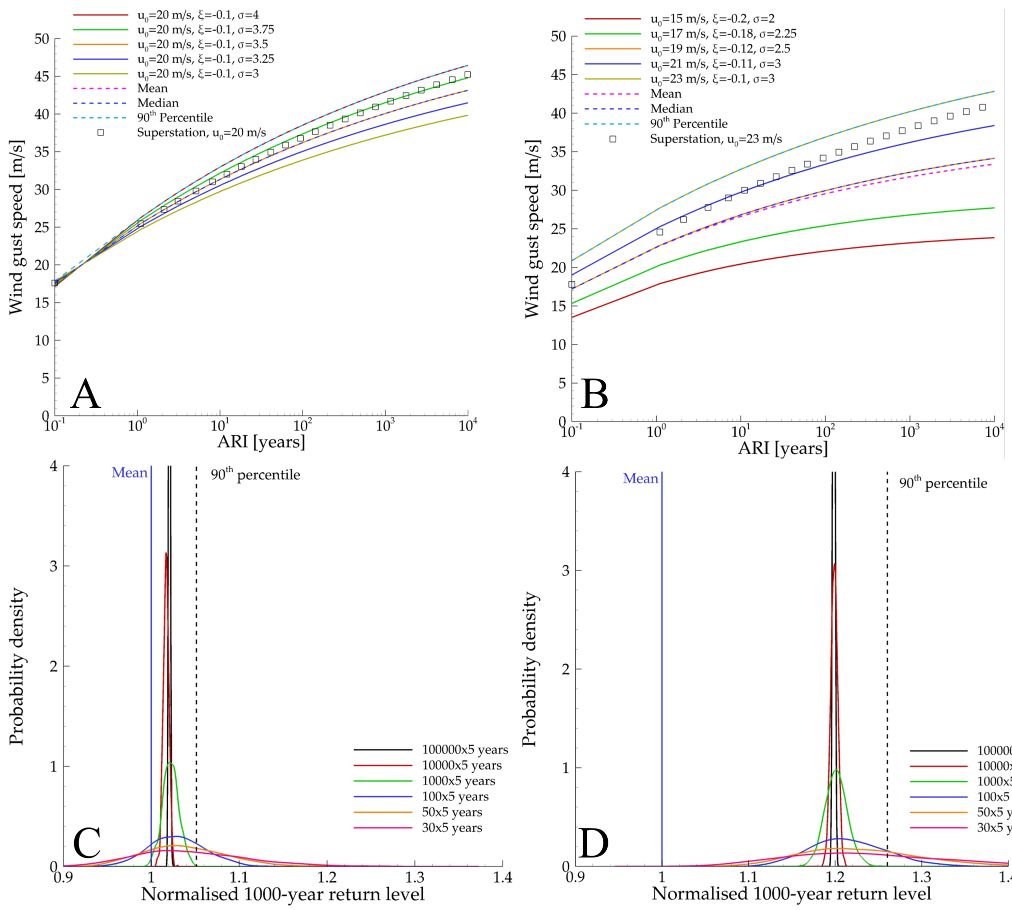

Figure 2: Simulated results for two different test cases: five pooled locations with (A,C) common threshold and shape factor values but varying scale factors, (B,D) all GPD distribution parameters varying among locations. Panels (A,B) show value vs. return period given 1000 years of data, and (C,D) show the PDF of the 1000-year return-period value given different data record lengths indicated in the legend (and assuming five threshold exceedences per year).

The return behavior of rainfall (Fig. 4A,B) is quite different from that of wind. Rainfall is highly intermittent with a long tail on the PDF, and the fits are unbounded with upward curvature to very high rain rates at extremely long return periods, due to the positive shape factor of the distribution, unlike the case for wind which has a negative shape factor and appears to asymptote toward a maximum possible value. Nonetheless the rainfall superstation fit is again higher than the true mean and median, and tends toward the 90$^{\text{th}}$ percentile of the neighborhood, consistent with the results based on wind distributions. For dataset lengths less than 1000 years, however (Fig. 4C,D), this bias is significantly outweighed by the sampling error such that the observed error in a single realization can be of either sign. This was not the case for gust distributions, where the superstation bias stands out even with short records (Figs. 2C,D).

The large sampling errors observed in the rainfall case are independent of the distribution model, as shown in Fig. 5, where the results from Fig. 4D are compared between the GPD and GEV approaches. Both show a similar level of bias for all record lengths, although the biases are slightly smaller if GPD is used instead of the usual (for rainfall) GEV. Sampling uncertainties exceed the superstation bias regardless of the model,

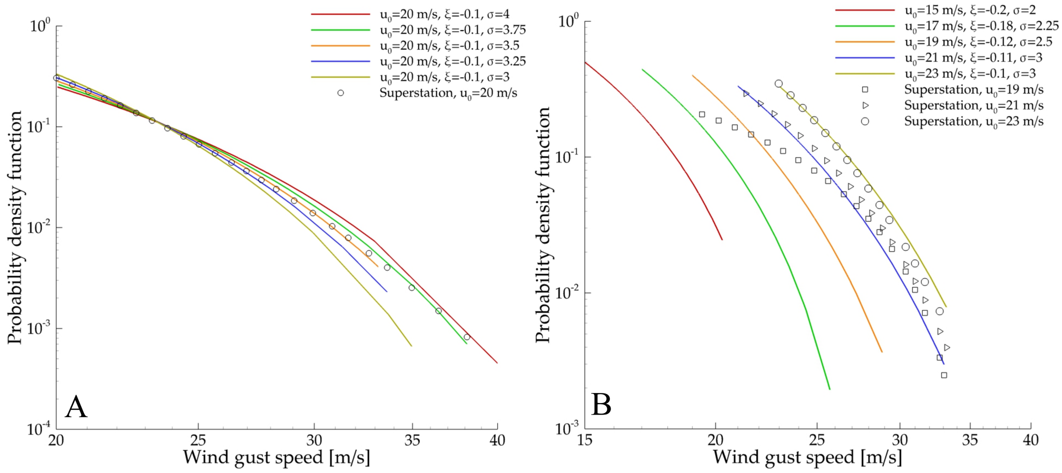

Figure 3: Probability density functions of the pseudo-station and superstation synthetic data shown in Figs. 2A, 2B. Open symbols in panel B show PDFs for three different threshold values.

for realistic dataset lengths. Thus, rainfall estimates are harder to constrain and the bias identified here is much less important compared to sampling error.

4 Conclusion

We demonstrate a previously unreported bias in estimating long-term recurrence intervals of extreme events that results from the common practice of regionalization or grouping data of nearby locations into one larger sample or ”superstation”. Wind gust and rainfall extremes have been considered for this analysis, but the results are also likely applicable to other types of weather extremes. Regionalisation assumes that all locations grouped have the same underlying distribution. According to newly available, high-resolution simulations of wind events in eastern Australia, differences in the underlying distribution can be large enough to induce biases at long recurrence intervals that dominate sampling uncertainty. The superstation fit tends to the highest levels suggested by any of the pooled locations and this bias increases with longer recurrence intervals. The tail of the superstation distribution tends to get its slope from the locations that experienced the most extremes. Moreover, the superstation PDF slope in our calculations is shallower than the those of any of the contributing stations, such that extrapolation will result in increasingly biased estimates at longer recurrence intervals. Our analysis suggests that for highly intermittent processes with unbounded behavior at the extreme tail such as rainfall with positive skewness for large values, the bias may be outweighed by random uncertainty and so may not matter in some cases, but will become important for bounded distributions. Since the importance of this bias depends on the distribution of the variable examined, we suggest that researchers should test for this bias before applying any regionalization method.

Conflict of Interest

The authors declare no conflicts of interest relevant to this study.

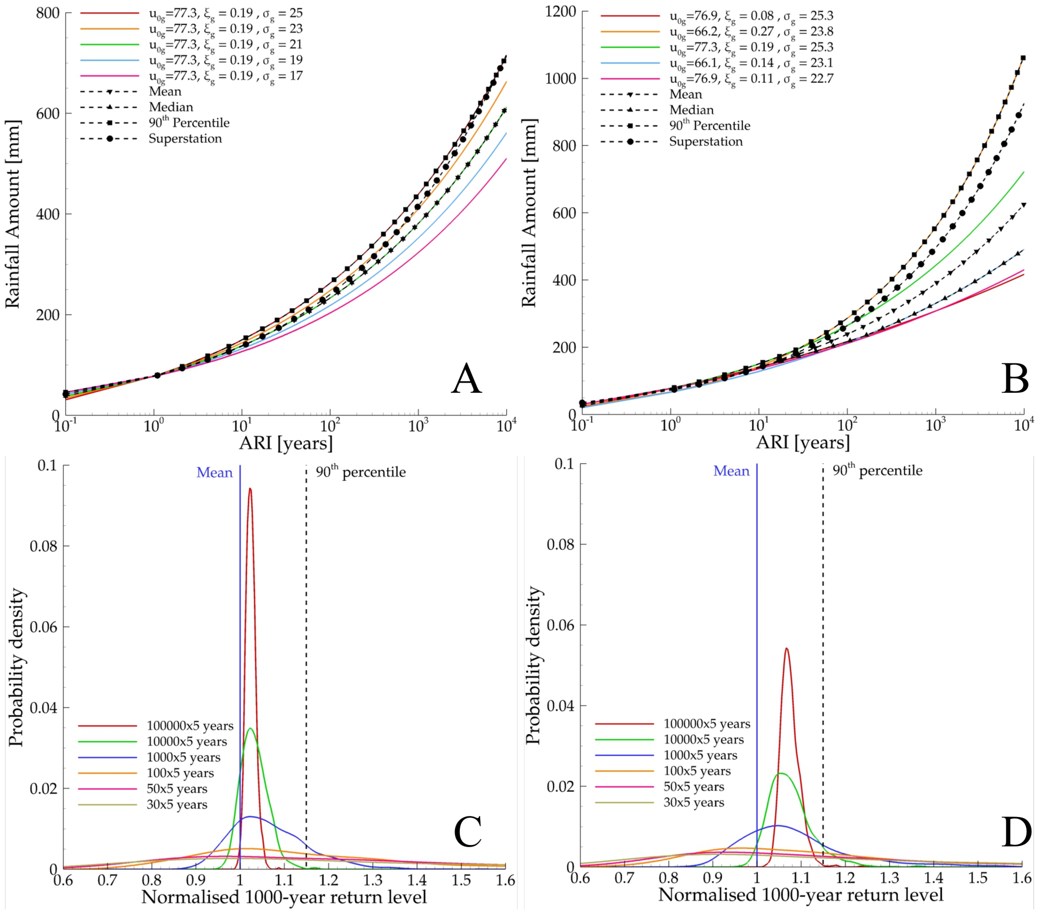

Figure 4: Same as Fig. 2 but for rainfall estimates.

Data Availability Statement

BARRA-SY data are available in an-open repository on the NCI (`http://dx.doi.org/10.4225/41/5993927b50f53`).

Acknowledgments

This research was supported by funding from the Australian Research council (Grant Number: LP200100138).

References

Bernardara, P., Andreewsky, M., & Benoit, M. (2011). Application of regional frequency analysis to the estimation of extreme storm surges. *Journal of Geophysical Research: Oceans*, *116*(C2). Retrieved from `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2010JC006229` doi: https://doi.org/10.1029/2010JC006229

Brabson, B. B., & Palutikof, J. P. (2000). Tests of the generalized Pareto distribution for predicting extreme wind speeds. *Journal of Applied Meteorology*, *39*, 1627–1640.

Church, J. A., Hunter, J. R., McInnes, K. L., & et. al. (2006). Sea-level rise around the Australian coastline and the changing frequency of extreme sea-level

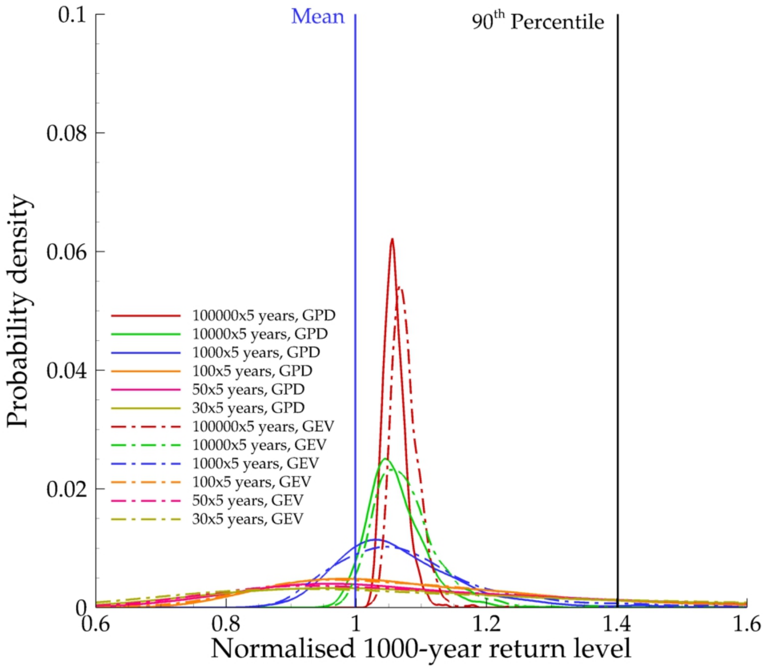

Figure 5: Same as Fig. 4D except comparing GPD and GEV models.

events. *Australian Meteorological Magazine*, *55*, 253–260.

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

El Rafei, M., Sherwood, S., Evans, J., & Dowdy, A. (2022). Analysis and characterisation of extreme wind gust hazards in New South Wales, Australia. *PREPRINT (Version 1) available at Research Square..* doi: https://doi.org/10.21203/rs.3.rs-1563734/v1

Green, J., Xuereb, K., Johnson, F., Moore, G., & The, C. (2012). The revised intensity-frequency-duration (ifd) design rainfall estimates for australia - an overview. In *Hydrology and water resources symposium 2012, 1, 1, 2012, 808-815.* Barton, ACT.

Gulev, S. K., & Grigorieva, V. (2004). Last century changes in ocean wind wave height from global visual wave data. *Geophysical Research Letters*, *31*(24). Retrieved from `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2004GL021040` doi: https://doi.org/10.1029/2004GL021040

Gumbel, E. J. (1958). Statistics of extremes. New York: Columbia University Press.

Haddad, K., & Rahman, A. (2012). Regional flood frequency analysis in eastern australia: Bayesian gls regression-based methods within fixed region and roi framework – quantile regression vs. parameter regression technique. *Journal of Hydrology*, *430-431*, 142-161. Retrieved from `https://www.sciencedirect.com/science/article/pii/S0022169412001102` doi: https://doi.org/10.1016/j.jhydrol.2012.02.012

Holmes, J. D. (2002). A re-analysis of recorded extreme wind speeds in region a. *Australian Journal of Structural Engineering*, *4*, 29–40.

Holmes, J. D. (2019). Extreme wind prediction – the Australian experience. In *Proceedings of the xv conference of the italian association for wind engineering* (pp. 365–375).

Holmes, J. D., & Moriarty, W. W. (1999). Application of the generalized pareto distribution to extreme value analysis in wind engineering. *Journal of Wind Engineering and Industrial Aerodynamics*, *93*, 1–10.

Jakob, D., Su, C.-H., Eizenberg, N., Steinle, P., , Fox-Hughes, P., & Bettio, L. (2017). An atmospheric high-resolution regional reanalysis for australia. *Bulletin of The Australian Meteorological and Oceanographic Society*, *30*, 16–23.

Johnson, F., & Green, J. (2018). A comprehensive continent-wide regionalisation investigation for daily design rainfall. *Journal of Hydrology: Regional Studies*, *16*, 67-79. Retrieved from `https://www.sciencedirect.com/science/article/pii/S2214581817302860` doi: https://doi.org/10.1016/j.ejrh.2018.03.001

Lazoglou, G., & Anagnostopoulou, C. (2017). An overview of statistical methods for studying the extreme rainfalls in mediterranean. *Proceedings*, *1*(5). Retrieved from `https://www.mdpi.com/2504-3900/1/5/681` doi: 10.3390/ecas2017-04132

Lucas, C., Muraleedharan, G., & Guedes Soares, C. (2017). Regional frequency analysis of extreme waves in a coastal area. *Coastal Engineering*, *126*, 81-95. Retrieved from `https://www.sciencedirect.com/science/article/pii/S0378383917302910` doi: https://doi.org/10.1016/j.coastaleng.2017.06.002

Meucci, A., Young, I. R., Hemer, M., Kirezci, E., & Ranasinghe, R. (2020). Projected 21st century changes in extreme wind-wave events. *Science Advances*, *6*(24), eaaz7295. Retrieved from `https://www.science.org/doi/abs/10.1126/sciadv.aaz7295` doi: 10.1126/sciadv.aaz7295

Minimum design loads for buildings and other structures (Standard). (1998). New York: American Society of Civil Engineers.

Minimum design loads for buildings and other structures (Standard). (2016). New York: American Society of Civil Engineers.

Norbiato, D., Borga, M., Sangati, M., & Zanon, F. (2007). Regional frequency analysis of extreme precipitation in the eastern italian alps and the august 29, 2003 flash flood. *Journal of Hydrology*, *345*(3), 149-166. Retrieved from `https://www.sciencedirect.com/science/article/pii/S0022169407004313` doi: https://doi.org/10.1016/j.jhydrol.2007.07.009

Palutikof, J. P., Brabson, B. B., Lister, D. H., & Adcock, S. T. (1999). A review of methods to calculate extreme wind speeds. *Meteorological Applications*, *6*, 119–132.

Peterka, J. (1992). Improved extreme wind prediction for the united states. *Journal of Wind Engineering and Industrial Aerodynamics*, *41*(1), 533-541. Retrieved from `https://www.sciencedirect.com/science/article/pii/016761059290459N` doi: https://doi.org/10.1016/0167-6105(92)90459-N

Peterka, J. A., & Shahid, S. (1998). Design gust wind speeds in the united states. *Journal of Structural Engineering*, *124*(2), 207-214. Retrieved from `https://ascelibrary.org/doi/abs/10.1061/%28ASCE%290733-9445%281998%29124%3A2%28207%29` doi: 10.1061/(ASCE)0733-9445(1998)124:2(207)

Pickands, J. (1975). Statistical inference using order statistics. *Annals of Statistics*, *3*, 119–131.

Structural design actions. Part 2: Wind actions (Standard). (2021). Sydney: Standards Australia.

Su, C.-H., Eizenberg, N., Jakob, D., Fox-Hughes, P., Steinle, P., White, C. J., & Franklin, C. (2021, July). BARRA v1.0: kilometre-scale downscaling of an Australian regional atmospheric reanalysis over four midlatitude domains. *Geoscientific Model Development*, *14*(7), 4357-4378. doi: 10.5194/gmd-14-4357-2021

Su, C.-H., Eizenberg, N., Steinle, P., Jakob, D., Fox-Hughes, P., White, C. J., ... Zhu, H. (2019). Barra v1.0: the bureau of meteorology atmospheric high-resolution regional reanalysis for australia. *Geoscientific Model Development*, *12*(5), 2049–2068. Retrieved from `https://gmd.copernicus.org/articles/12/2049/2019/` doi: 10.5194/gmd-12-2049-2019

Van Gelder, P., De Ronde, J., Neykov, N. M., & Neytchev, P. (2001). Regional frequency analysis of extreme wave heights: Trading space for time. In *Coastal engineering conference* (Vol. 2, pp. 1099–1112).

Wallis, J. R., Schaefer, M. G., Barker, B. L., & Taylor, G. H. (2007, January). Regional precipitation-frequency analysis and spatial mapping for 24-hour and 2-hour durations for Washington State. *Hydrology and Earth System Sciences*, *11*(1), 415-442. doi: 10.5194/hess-11-415-2007

Wang, C., Wang, X., & Khoo, Y. B. (2013). Extreme wind gust hazard in Australia and its sensitivity to climate change. *Natural Hazards*, *67*, 549–567.

Biases in estimating long-term recurrence intervals of extreme events due to regionalised sampling

Moutassem El Rafei [1], Steven Sherwood [1,2], Jason Evans [1,2], Andrew Dowdy [3] and Fei Ji [4]

[1] Climate Change Research Centre, University of New South, Wales, Sydney, Australia
[2] ARC Centre of Excellence for Climate Extremes, University of New South Wales, Sydney, Australia
[3] Australian Bureau of Meteorology
[4] NSW Department of Planning and Environment, Sydney, Australia

Key Points:

- Grouping data of nearby locations into one larger sample or "superstation" can induce biases at long recurrence intervals
- The superstation fit tends to the highest levels suggested by any of the pooled locations
- The bias may be large or small compared to random uncertainty, depending on the distribution of the extreme event analysed

Corresponding author: Moutassem El Rafei, m.el_rafei@unsw.edu.au

Abstract

Preparing for environmental risks requires estimating the frequencies of extreme events, often from data records that are too short to confirm them directly. This requires fitting a statistical distribution to the data. To improve precision, investigators often pool data from neighboring sites into single samples, referred to as "superstations," before fitting. We demonstrate that this technique can introduce unexpected biases in typical situations, using wind and rainfall extremes as case studies. When the combined locations have even small differences in the underlying statistics, the regionalization approach gives a fit that may tend toward the highest levels suggested by any of the individual sites. This bias may be large or small compared to the sampling error, for realistic record lengths, depending on the distribution of the quantity analysed. The results of this analysis indicate that previous analyses could potentially have overestimated the likelihood of extreme events arising from natural weather variability.

Plain Language Summary

We report a previously unknown bias in a common method for estimating how often extremely rare events such as extreme wind bursts or rain events will occur, when return periods are longer than the available data record. The method analysed is one where an investigator combines data from nearby locations to reduce sampling error. We find by looking at new, high-resolution data that variations in behavior across sites can sometimes produce biases much larger than the sampling error. The implication is that some observed extreme events are even less likely to have occurred than previously thought, assuming the underlying distribution hasn't changed over the period of observation.

1 Introduction

The statistical analysis of extreme-event frequencies and intensities is important to many risk management problems. For example, estimating the appropriate design wind speed requires the statistical analysis of historical wind data to estimate the strongest wind that might occur over a long time interval (AS/NZS1170.2:2021, 2021; Holmes, 2002; El Rafei et al., 2022). The design of offshore and coastal marine structures is governed by statistical estimates of extreme waves (Gulev & Grigorieva, 2004; Meucci et al., 2020). Similarly, statistical estimates of extreme rainfall values are essential for calculating flood risk and designing stormwater infrastructure (Green et al., 2012; Johnson & Green, 2018). To meet this need, extreme value theory has been widely used to estimate the probability of events larger than any on record so far (Brabson & Palutikof, 2000; Coles, 2001; Church et al., 2006; Wang et al., 2013).

The two usual approaches of extreme value theory are the generalized extreme value distribution (GEV) and the generalized Pareto distribution (GPD). The GEV approach, which combines three different statistical families (Weibull, Gumbel and Frechet), uses block maxima in which the dataset is divided into blocks and the maximum over each block is modelled (Gumbel, 1958; Palutikof et al., 1999; Coles, 2001). The GPD approach is instead based on the peaks-over-threshold method for which a threshold value is specified and all the values above this chosen threshold are used to fit the model (Pickands, 1975; Coles, 2001; Holmes & Moriarty, 1999). No matter which approach is used, the extrapolation to very rare events is subject to significant sampling errors when using short data ranges and uncertainties are unavoidable as accurate observational records are commonly short and/or geographically sparse.

To reduce statistical uncertainty, regionalization techniques have been used, whereby a larger sample is created by combining independent records of neighboring stations (J. Peterka, 1992; J. A. Peterka & Shahid, 1998; Holmes, 2002; Wang et al., 2013; Holmes, 2019) into what is sometimes called a "superstation". For example, regionalization has been used for extreme wind assessment in the United States (J. A. Peterka & Shahid, 1998; ANSI/AS CE 7-98, 1998; ASCE/SE I 7-16, 2016) and Australia (Holmes, 2002; AS/NZS1170.2:2021, 2021) to specify single design wind speed by compositing data from multiple stations; for regional flood frequency estimates (Haddad & Rahman, 2012); and for Intensity-Frequency-Duration (IFD) rainfall relationships (Wallis et al., 2007; Norbiato et al., 2007; Green et al., 2012; Johnson & Green, 2018). For IFD applications, nearby stations are pooled together assuming they share a common distribution of rainfall and are independent. Another place where this approach has been used is for so-called "regional frequency analysis" of extreme wave heights, where data from sites with similar wave statistics are used to estimate the distribution for a presumed homogeneous region (Van Gelder et al., 2001; Bernardara et al., 2011; Lucas et al., 2017).

While regionalization allows the estimation of distribution parameters using a larger dataset, the biases of this strategy are not explicitly quantified in the literature. Here we report unexpected biases in estimating long-term recurrence intervals of extreme events via regionalization, considering wind and rainfall extremes as case studies.

2 Data and Methods

2.1 Wind Data and Distribution Parameters

Wind data are obtained from the 1.5 km Bureau of Meteorology Atmospheric high-resolution Regional Reanalysis for Australia Sydney region (BARRA-SY) and cover the period from 1996 to 2019. This fine-scale reanalysis may provide better representation of storm systems and mesoscale phenomena, which can allow better modelling and higher accuracy of the wind field compared to lower resolution models. The reader is referred to (Jakob et al., 2017; Su et al., 2019; Su et al., 2021) for more details.

The maximum hourly wind gust speed data was used to calculate the daily maxima, which are then used to fit the GPD model. For computational tractability, we sub-sampled the BARRA-SY data to 5324 grid points covering New South Wales (NSW), with a spacing of approximately 10 km. This is judged to be justified because gusts at locations that are very close (i.e., 1.5 km apart) would capture the same gust events multiple times hence not be independent. The domain is then decomposed into groups (superstations) of approximately 25 grid points (5×5 grid points neighborhood) such that approximately 575 station years are obtained for each superstation.

The fully heterogeneous numerical experiment (Fig. 2B) considers threshold values between 15 and 23 m/s, shape factors between -0.2 to -0.1 and scale parameters ranging from 2 to 3. These ranges are representative of wind gust GPD distributions obtained using the BARRA-SY dataset. The partially heterogeneous case (Fig. 2B) considers a pre-fixed shape factor of -0.1 (similarly to what is used in the Australian standards) and a threshold value of 20 m/s. The range of scale factor values (between 3 and 4) was arbitrarily selected to represent a gradient of this parameter between the pseudo-stations.

2.2 Rainfall Parameters

To provide parameter values for rainfall tests, 30 years (1992-2021) of daily rainfall data were obtained from the Australian Bureau of Meteorology for five locations in the Sydney area (Sydney Airport, Sans Suci, Randwick, Rose Bay and Peakhurst) providing good data quality, using only low elevation stations ($\leq$ 100 m). Incorrect values (flagged as wrong via the quality checks) were removed, and the annual maxima were used to fit a GEV model at each station. The GEV analysis suggests threshold values ranging from 66.2 to 77.3, shape factors ranging from 0.08 to 0.27 and scale factors ranging from 22.7 to 25.3. These ranges were then used to generate synthetic random samples for the fully heterogeneous scenario (Fig. 4B). The partially heterogeneous scenario considers pre-fixed threshold and shape factor ($u_0 = 77.3$ and $\xi = 0.19$), which are selected from one of the stations distributions, and a range of scale factors from 17 to 25. The range of scale factors is randomly selected for this scenario to represent a gradient of this parameter.

2.3 Extreme Value Theory

The GEV approach uses the block maxima technique where the maximum yearly value is considered for the fit. The GEV cumulative distribution function can be expressed as follows:

$$G(u_g) = \begin{cases} e^{-\left[1+\xi_g\left(\frac{u_g-u_{0g}}{\sigma_g}\right)\right]^{-1/\xi_g}}, & \text{for } \xi_g \neq 0 \\ e^{-e^{(u_g-u_{0g})/\sigma_g}}, & \text{for } \xi_g = 0 \end{cases}$$

where u_{0g}, ξ_g and σ_g are the location, shape and scale parameters respectively. The return levels are given in terms of ARI as:

$$U_{Rg} = u_{0g} - \frac{\sigma_g}{\xi_g}[1 - (R)^{\xi_g}] \tag{1}$$

where R is the ARI.

The GPD model is based on the peaks-over-threshold approach (Palutikof et al., 1999; Coles, 2001) where all the data above a specified threshold are modelled. The GPD distribution can be expressed as:

$$H(u_p) = \begin{cases} 1 - \left(1 + \xi_p \frac{u_p - u_{0p}}{\sigma_p}\right)^{-1/\xi_p}, & \text{for } \xi_p \neq 0 \\ 1 - e^{\frac{u_p - u_{0p}}{\sigma_p}}, & \text{for } \xi_p = 0 \end{cases}$$

where u_{0p}, ξ_p and σ_p are the threshold, shape and scale parameters respectively. This expression is defined on $\{u_p - u_{0p} > 0$ and $(1 + \xi_p(u_p - u_{0p})/\tilde{\sigma_p}) > 0\}$ where $\tilde{\sigma_p} = \sigma_p + \xi_p(u_p - u_{0p})$.

The return levels are given in terms of the average recurrence interval using:

$$U_{Rp} = u_{0p} - \frac{\sigma_p}{\xi_p}[1 - (\lambda R)^{\xi_p}] \tag{2}$$

where λ is the number of crossings of the threshold per year.

The rainfall estimates calculated using the GPD approach use the 99^{th} percentile as a threshold value (Lazoglou & Anagnostopoulou, 2017). For the wind analysis, we used a threshold selection algorithm that selects an optimal threshold value at each location. The selection algorithm first classifies the gusts into convective (i.e., thunderstorms) and synoptic (e.g., east coast lows, frontal systems) events. This is because gusts produced by different mechanisms can have different statistical properties and distributions. A range of thresholds is then tested to select the best fit. The reader is referred to (El Rafei et al., 2022) for more details on the threshold selection algorithm and the storm classification technique. The current study only presents results of convective events, as the same bias occur with either type using the superstation technique.

3 Results

3.1 Extreme Wind Gust Example

3.1.1 Reanalysis Data Study

We begin by examining a high-resolution (1.5-km horizontal grid spacing) regional Australian reanalysis dataset (BARRA-SY) of 23 years length from Year 1996 to Year

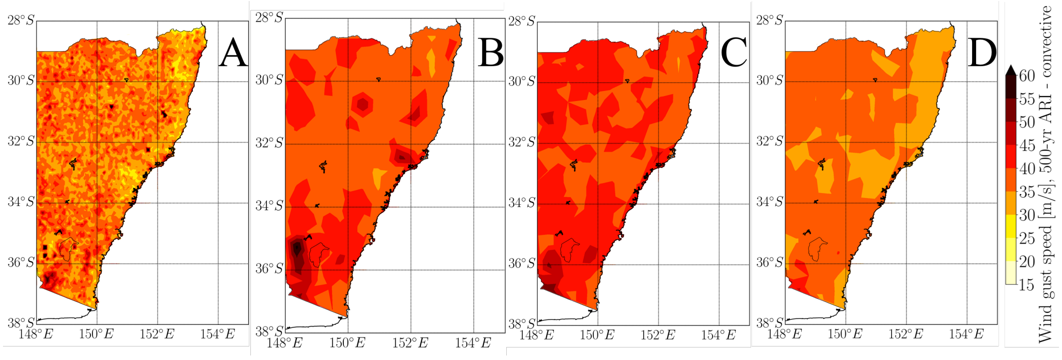

Figure 1: 500-year return wind gust speeds from convective events estimated several ways. Results are calculated using: A) the GPD algorithm applied at each grid point. B) the "superstation" approach to aggregate data from neighboring grid points. C) the nearby 90$^{\text{th}}$ percentile and D) mean of values shown in panel A, computed in the same neighborhood as the "superstations".

2019, using the GPD approach. The availability of this new high-resolution gridded dataset has motivated this study, since wind data are available at sufficient density to directly test the performance of the regionalization approach. The GPD method and reanalysis data are detailed in the Data and Methods section. Maps of 500-year convective (i.e., thunderstorm) wind gust speeds calculated in several ways are shown in Fig. 1. We show the 500-year ARI as this is used for the design criteria of many buildings in Australia (Wang et al., 2013). Results of higher recurrence intervals (e.g. 1000- and 2000-year ARIs, not shown) present similar patterns.

The results obtained by GPD fitting and calculating return levels independently at each grid cell (Fig. 1A) show speckling due to small differences in the estimated GPD parameters. We expect this is mainly from sampling error due to the short record length, as it looks random rather than geophysical. The regionalization approach (Fig. 1B), in which data from 25 adjacent grid cells are combined before fitting, yields smoother results but they show higher wind gust levels everywhere by about 10% compared to Fig. 1A. They are also higher than when the grid-point results are instead spatially smoothed by taking neighborhood means (2-D boxcar mean smoother of 5×5 grid points), as shown in Fig. 1D. Indeed they are very close to the neighborhood 90$^{\text{th}}$ percentile results in Fig. 1C (calculated as the 90$^{\text{th}}$ percentile wind gust speed of 25 adjacent grid points). This is true even for regional maxima in neighborhood spread such as in the southeast corner of the state.

This analysis suggests that the superstation technique gives a fit that is close to the highest levels suggested by any of the neighboring, noise-influenced, sites. We now examine two hypotheses for why this is happening: first, that we bias the return levels by using short data ranges; and second, that it is an effect of combining locations into a superstation.

3.1.2 Simulation Experiments

A set of numerical experiments have been carried out to test these hypotheses, by generating synthetic records from pseudo-stations. Five synthetic "neighbor" data records are analytically generated from assumed GPD distributions with parameters based on wind gust distributions estimated using the BARRA-SY reanalysis, in two test scenarios. For the first, partially heterogeneous scenario (Fig. 2A,C), the five records are generated using the same threshold ($u_0 = 20$ m/s) and shape factor ($\xi = -0.1$) but a range of scale factors from 3-4. Pre-fixing the threshold and shape factor is similar to to approaches sometimes used in structural design standards (AS/NZS1170.2:2021, 2021). For the second, fully heterogeneous scenario (Fig. 2B,D), all the distribution parameters vary among the records.

We first examine outcomes with long records (Figs. 2A and 2B), generating 1000 years of data, with a total of 5000 data points used by the GPD model, for each record (i.e., five threshold exceedence events per year on average). Figs. 2A and B show that, even when differences in the underlying distributions between the locations are small, the superstation fit is higher than the mean and the median of the individual fits and tends to the highest levels suggested by the individual stations, which is consistent with the wind data results in Fig. 1. Furthermore, in both simulations the bias increases at longer recurrence intervals. The superstation bias in the fully heterogeneous scenario is more significant, even at short recurrence intervals, because some locations are contributing a lot more events than others to the superstation. This is not seen in the partially homogeneous scenario where all station distributions have the same threshold and shape factor and hence contribute similarly.

We next consider the effect of sample size by considering dataset lengths that range from 30 to 100000 years. For each length we repeated the test 1000 times to yield a PDF, the mean and 90^{th} percentile of which is shown in Figs. 2C,D. The sampling uncertainty monotonically reduces for longer datasets and gradually converges to the true superstation bias. At 30 and 50 year record lengths (typical of real-world datasets), the bias may be exceeded in magnitude by the random error as depicted by the large spread of the PDF. Importantly however, all PDFs are centred on the true bias, showing that short record lengths do not cause biases, only random sampling errors. The PDFs of error (Fig. 2C,D) are narrow in both simulations when datasets are longer than 1000 years, implying a consistent bias site to site (or realisation to realisation), as implied by the geographic uniformity of the difference between Fig. 1B vs. D. Moreover, as seen before, the simulated superstation result is higher than the mean return level of the stations in the neighborhood. Hence we conclude from these tests that while the noise seen in the BARRA-SY return-period map is from sampling errors due to the short record, the ubiquitous bias toward high values is caused by regional pooling of data from nearby locations.

To understand what gives rise to this systematic bias, we compared the PDFs of gust speed (Figs. 3A and B) corresponding to the exceedence curves shown in Fig. 2A, B. In both scenarios, the slope of the superstation gust PDF at high gust speed tends to be dictated by the stations that have the heaviest tail (i.e., where the most extremes are recorded hence contributing most heavily to the superstation sample). Moreover, this phenomenon increases as one goes farther out on the tail of the PDF; if for example a very high threshold is used, nearly all data meeting the threshold come from one station (open circles in Fig. 3B). This level-dependent bias imparts a shallower slope to the PDF tail, which means that when calculating very long recurrence intervals, the extrapolation would tend to levels suggested by the locations that experience the most extreme events, or possibly even higher.

3.2 Rainfall Example

We now repeat the above analysis for another type of natural hazard, extreme rainfall, to explore the generality of the result. In this case rather than GPD we use GEV, since this is how published rainfall estimates are typically calculated, but this choice does not substantially affect results (Data and Methods section). The synthetic data records are generated as before, considering the same two scenarios except for the parameter values. The range of rainfall scale factors varies from 17 to 25, based on distributions estimated from observed daily accumulated rainfall from weather stations in Sydney area.

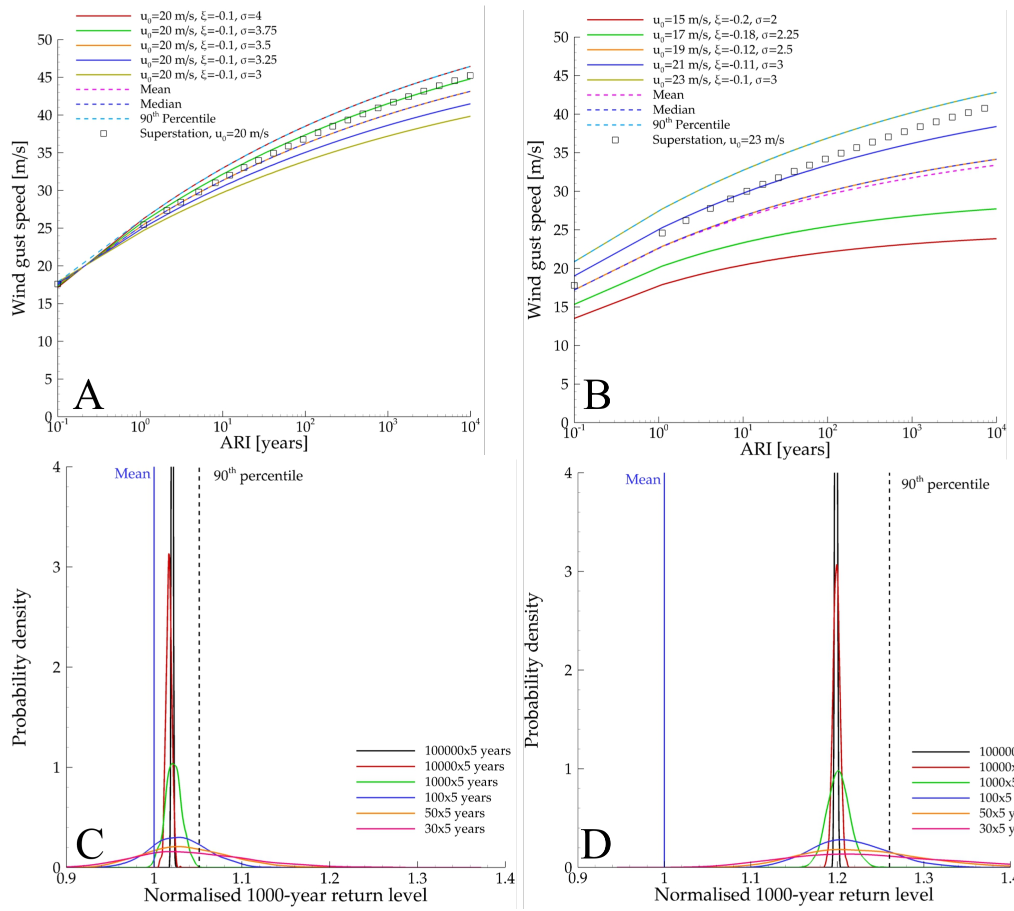

Figure 2: Simulated results for two different test cases: five pooled locations with (A,C) common threshold and shape factor values but varying scale factors, (B,D) all GPD distribution parameters varying among locations. Panels (A,B) show value vs. return period given 1000 years of data, and (C,D) show the PDF of the 1000-year return-period value given different data record lengths indicated in the legend (and assuming five threshold exceedences per year).

The return behavior of rainfall (Fig. 4A,B) is quite different from that of wind. Rainfall is highly intermittent with a long tail on the PDF, and the fits are unbounded with upward curvature to very high rain rates at extremely long return periods, due to the positive shape factor of the distribution, unlike the case for wind which has a negative shape factor and appears to asymptote toward a maximum possible value. Nonetheless the rainfall superstation fit is again higher than the true mean and median, and tends toward the 90$^{\text{th}}$ percentile of the neighborhood, consistent with the results based on wind distributions. For dataset lengths less than 1000 years, however (Fig. 4C,D), this bias is significantly outweighed by the sampling error such that the observed error in a single realization can be of either sign. This was not the case for gust distributions, where the superstation bias stands out even with short records (Figs. 2C,D).

The large sampling errors observed in the rainfall case are independent of the distribution model, as shown in Fig. 5, where the results from Fig. 4D are compared between the GPD and GEV approaches. Both show a similar level of bias for all record lengths, although the biases are slightly smaller if GPD is used instead of the usual (for rainfall) GEV. Sampling uncertainties exceed the superstation bias regardless of the model,

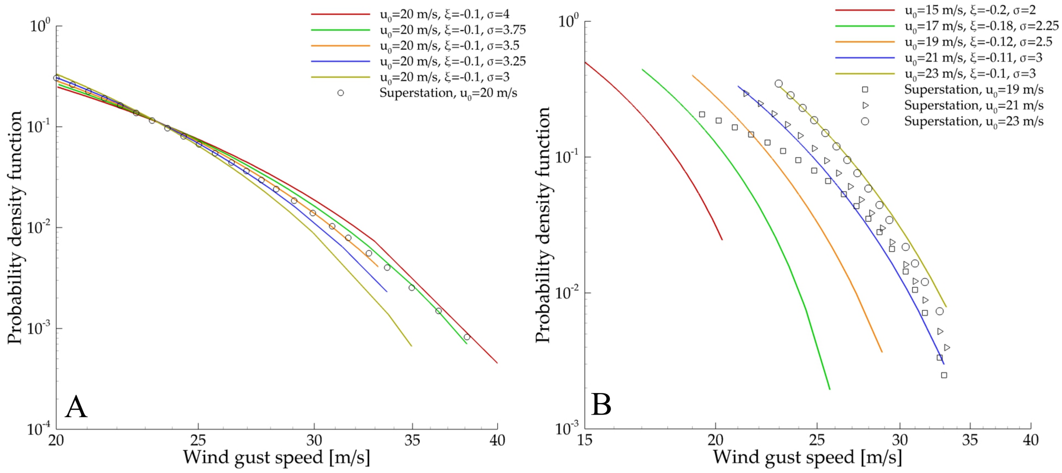

Figure 3: Probability density functions of the pseudo-station and superstation synthetic data shown in Figs. 2A, 2B. Open symbols in panel B show PDFs for three different threshold values.

for realistic dataset lengths. Thus, rainfall estimates are harder to constrain and the bias identified here is much less important compared to sampling error.

4 Conclusion

We demonstrate a previously unreported bias in estimating long-term recurrence intervals of extreme events that results from the common practice of regionalization or grouping data of nearby locations into one larger sample or ”superstation”. Wind gust and rainfall extremes have been considered for this analysis, but the results are also likely applicable to other types of weather extremes. Regionalisation assumes that all locations grouped have the same underlying distribution. According to newly available, high-resolution simulations of wind events in eastern Australia, differences in the underlying distribution can be large enough to induce biases at long recurrence intervals that dominate sampling uncertainty. The superstation fit tends to the highest levels suggested by any of the pooled locations and this bias increases with longer recurrence intervals. The tail of the superstation distribution tends to get its slope from the locations that experienced the most extremes. Moreover, the superstation PDF slope in our calculations is shallower than the those of any of the contributing stations, such that extrapolation will result in increasingly biased estimates at longer recurrence intervals. Our analysis suggests that for highly intermittent processes with unbounded behavior at the extreme tail such as rainfall with positive skewness for large values, the bias may be outweighed by random uncertainty and so may not matter in some cases, but will become important for bounded distributions. Since the importance of this bias depends on the distribution of the variable examined, we suggest that researchers should test for this bias before applying any regionalization method.

Conflict of Interest

The authors declare no conflicts of interest relevant to this study.

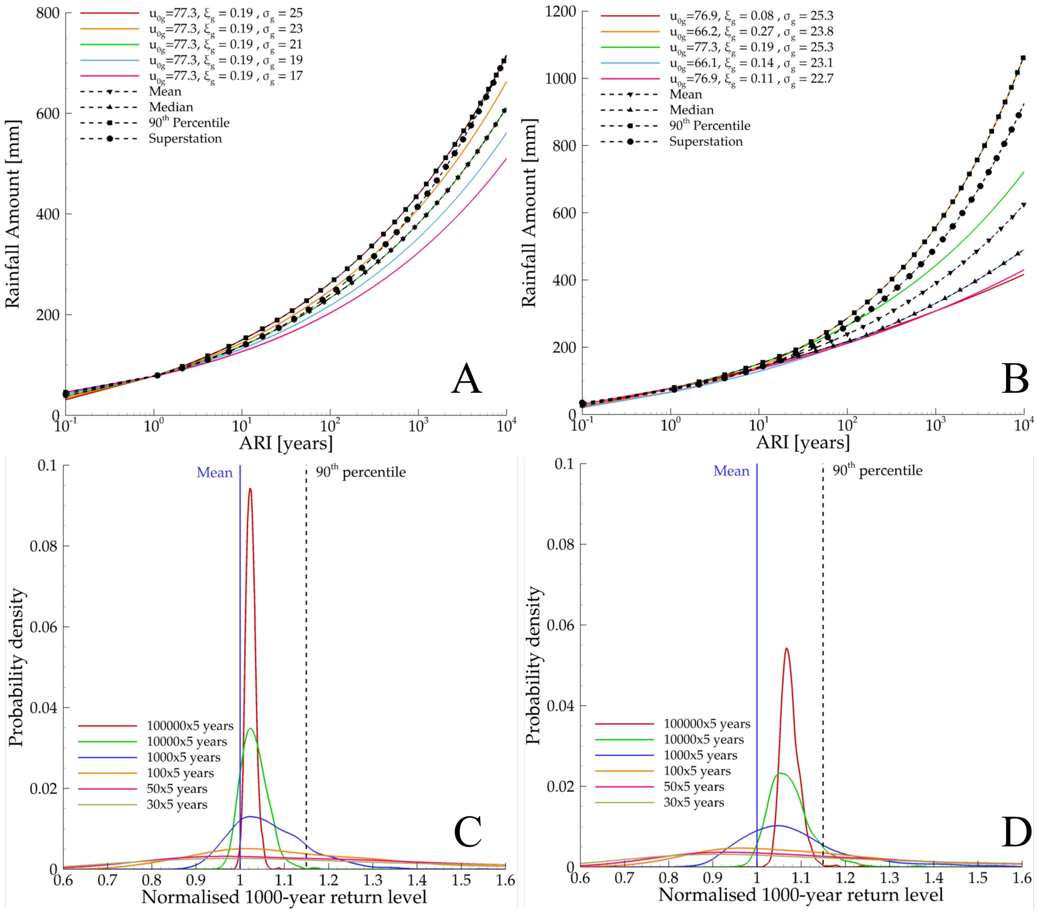

Figure 4: Same as Fig. 2 but for rainfall estimates.

Data Availability Statement

BARRA-SY data are available in an-open repository on the NCI (`http://dx.doi.org/10.4225/41/5993927b50f53`).

Acknowledgments

This research was supported by funding from the Australian Research council (Grant Number: LP200100138).

References

Bernardara, P., Andreewsky, M., & Benoit, M. (2011). Application of regional frequency analysis to the estimation of extreme storm surges. *Journal of Geophysical Research: Oceans*, *116*(C2). Retrieved from `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2010JC006229` doi: https://doi.org/10.1029/2010JC006229

Brabson, B. B., & Palutikof, J. P. (2000). Tests of the generalized Pareto distribution for predicting extreme wind speeds. *Journal of Applied Meteorology*, *39*, 1627–1640.

Church, J. A., Hunter, J. R., McInnes, K. L., & et. al. (2006). Sea-level rise around the Australian coastline and the changing frequency of extreme sea-level

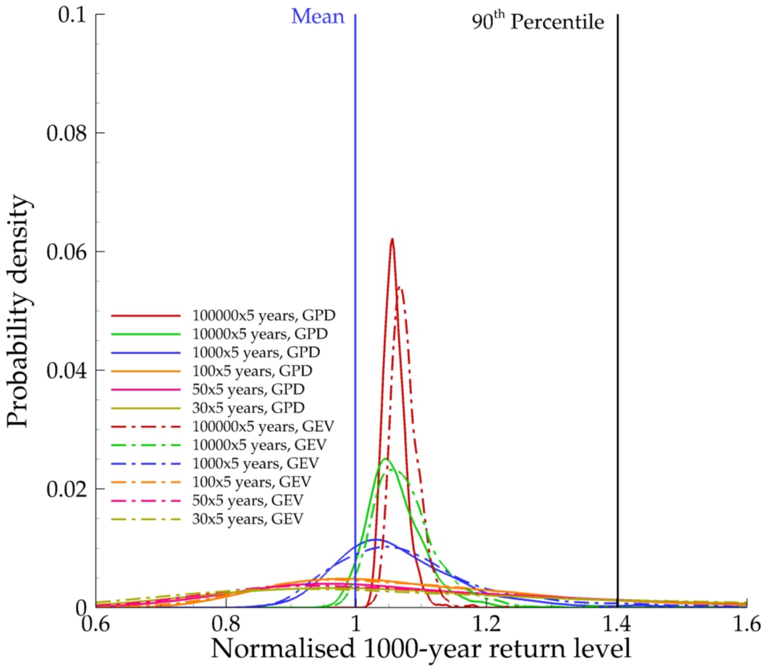

Figure 5: Same as Fig. 4D except comparing GPD and GEV models.

events. *Australian Meteorological Magazine*, *55*, 253–260.

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

El Rafei, M., Sherwood, S., Evans, J., & Dowdy, A. (2022). Analysis and characterisation of extreme wind gust hazards in New South Wales, Australia. *PREPRINT (Version 1) available at Research Square.*. doi: https://doi.org/10.21203/rs.3.rs-1563734/v1

Green, J., Xuereb, K., Johnson, F., Moore, G., & The, C. (2012). The revised intensity-frequency-duration (ifd) design rainfall estimates for australia - an overview. In *Hydrology and water resources symposium 2012, 1, 1, 2012, 808-815.* Barton, ACT.

Gulev, S. K., & Grigorieva, V. (2004). Last century changes in ocean wind wave height from global visual wave data. *Geophysical Research Letters*, *31*(24). Retrieved from `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2004GL021040` doi: https://doi.org/10.1029/2004GL021040

Gumbel, E. J. (1958). Statistics of extremes. New York: Columbia University Press.

Haddad, K., & Rahman, A. (2012). Regional flood frequency analysis in eastern australia: Bayesian gls regression-based methods within fixed region and roi framework – quantile regression vs. parameter regression technique. *Journal of Hydrology*, *430-431*, 142-161. Retrieved from `https://www.sciencedirect.com/science/article/pii/S0022169412001102` doi: https://doi.org/10.1016/j.jhydrol.2012.02.012

Holmes, J. D. (2002). A re-analysis of recorded extreme wind speeds in region a. *Australian Journal of Structural Engineering*, *4*, 29–40.

Holmes, J. D. (2019). Extreme wind prediction – the Australian experience. In *Proceedings of the xv conference of the italian association for wind engineering*

(pp. 365–375).
Holmes, J. D., & Moriarty, W. W. (1999). Application of the generalized pareto distribution to extreme value analysis in wind engineering. *Journal of Wind Engineering and Industrial Aerodynamics*, *93*, 1–10.
Jakob, D., Su, C.-H., Eizenberg, N., Steinle, P., , Fox-Hughes, P., & Bettio, L. (2017). An atmospheric high-resolution regional reanalysis for australia. *Bulletin of The Australian Meteorological and Oceanographic Society*, *30*, 16–23.
Johnson, F., & Green, J. (2018). A comprehensive continent-wide regionalisation investigation for daily design rainfall. *Journal of Hydrology: Regional Studies*, *16*, 67-79. Retrieved from `https://www.sciencedirect.com/science/article/pii/S2214581817302860` doi: https://doi.org/10.1016/j.ejrh.2018.03.001
Lazoglou, G., & Anagnostopoulou, C. (2017). An overview of statistical methods for studying the extreme rainfalls in mediterranean. *Proceedings*, *1*(5). Retrieved from `https://www.mdpi.com/2504-3900/1/5/681` doi: 10.3390/ecas2017-04132
Lucas, C., Muraleedharan, G., & Guedes Soares, C. (2017). Regional frequency analysis of extreme waves in a coastal area. *Coastal Engineering*, *126*, 81-95. Retrieved from `https://www.sciencedirect.com/science/article/pii/S0378383917302910` doi: https://doi.org/10.1016/j.coastaleng.2017.06.002
Meucci, A., Young, I. R., Hemer, M., Kirezci, E., & Ranasinghe, R. (2020). Projected 21st century changes in extreme wind-wave events. *Science Advances*, *6*(24), eaaz7295. Retrieved from `https://www.science.org/doi/abs/10.1126/sciadv.aaz7295` doi: 10.1126/sciadv.aaz7295
Minimum design loads for buildings and other structures (Standard). (1998). New York: American Society of Civil Engineers.
Minimum design loads for buildings and other structures (Standard). (2016). New York: American Society of Civil Engineers.
Norbiato, D., Borga, M., Sangati, M., & Zanon, F. (2007). Regional frequency analysis of extreme precipitation in the eastern italian alps and the august 29, 2003 flash flood. *Journal of Hydrology*, *345*(3), 149-166. Retrieved from `https://www.sciencedirect.com/science/article/pii/S0022169407004313` doi: https://doi.org/10.1016/j.jhydrol.2007.07.009
Palutikof, J. P., Brabson, B. B., Lister, D. H., & Adcock, S. T. (1999). A review of methods to calculate extreme wind speeds. *Meteorological Applications*, *6*, 119–132.
Peterka, J. (1992). Improved extreme wind prediction for the united states. *Journal of Wind Engineering and Industrial Aerodynamics*, *41*(1), 533-541. Retrieved from `https://www.sciencedirect.com/science/article/pii/016761059290459N` doi: https://doi.org/10.1016/0167-6105(92)90459-N
Peterka, J. A., & Shahid, S. (1998). Design gust wind speeds in the united states. *Journal of Structural Engineering*, *124*(2), 207-214. Retrieved from `https://ascelibrary.org/doi/abs/10.1061/%28ASCE%290733-9445%281998%29124%3A2%28207%29` doi: 10.1061/(ASCE)0733-9445(1998)124:2(207)
Pickands, J. (1975). Statistical inference using order statistics. *Annals of Statistics*, *3*, 119–131.
Structural design actions. Part 2: Wind actions (Standard). (2021). Sydney: Standards Australia.
Su, C.-H., Eizenberg, N., Jakob, D., Fox-Hughes, P., Steinle, P., White, C. J., & Franklin, C. (2021, July). BARRA v1.0: kilometre-scale downscaling of an Australian regional atmospheric reanalysis over four midlatitude domains. *Geoscientific Model Development*, *14*(7), 4357-4378. doi: 10.5194/gmd-14-4357-2021

Su, C.-H., Eizenberg, N., Steinle, P., Jakob, D., Fox-Hughes, P., White, C. J., ... Zhu, H. (2019). Barra v1.0: the bureau of meteorology atmospheric high-resolution regional reanalysis for australia. *Geoscientific Model Development*, *12*(5), 2049–2068. Retrieved from `https://gmd.copernicus.org/articles/12/2049/2019/` doi: 10.5194/gmd-12-2049-2019

Van Gelder, P., De Ronde, J., Neykov, N. M., & Neytchev, P. (2001). Regional frequency analysis of extreme wave heights: Trading space for time. In *Coastal engineering conference* (Vol. 2, pp. 1099–1112).

Wallis, J. R., Schaefer, M. G., Barker, B. L., & Taylor, G. H. (2007, January). Regional precipitation-frequency analysis and spatial mapping for 24-hour and 2-hour durations for Washington State. *Hydrology and Earth System Sciences*, *11*(1), 415-442. doi: 10.5194/hess-11-415-2007

Wang, C., Wang, X., & Khoo, Y. B. (2013). Extreme wind gust hazard in Australia and its sensitivity to climate change. *Natural Hazards*, *67*, 549–567.