25th International Congress on Modelling and Simulation, Darwin, NT, Australia, 9 to 14 July 2023
mssanz.org.au/modsim2023

On reducing bias at high return levels for extreme value analysis

C.-H. Wang

CSIRO Energy, Melbourne, Australia
Email: chi-hsiang.wang@csiro.au

Abstract: Estimates of extremal return levels at high average recurrence intervals (ARIs) are strongly dependent on the shape parameter of the statistical model. Because of scarcity in occurrences, however, many existing extremal data span only a few decades, often resulting in large bias and uncertainty in the estimated shape parameter of the extreme hazard model. This in turn leads to unreliable predicted extreme values at high ARIs. A common approach to ameliorate this shortcoming is the 'super-station' (or station-year) approach which extends the length of record and should reduce the uncertainty in high ARIs. However, the problem of predicted bias remains for return levels beyond the record length of the super-station.

This paper illustrates a statistical method that provides a mechanism to obtain a hazard model that produces return levels at high ARIs with reduced bias. For an ensemble of independently collected records from a number of observational sites, the method makes use of the maximum recorded value of each of the extremal data of the ensemble, as shown in Figure 1 as the starred points (green points represent synoptic and red points non-synoptic wind gusts). The logarithmically transformed probability of the maximum recorded value at a site is shown to follow the Gumbel (Type I extreme-value) distribution, therefore multiple, say m, sites provide a sample of size m transformed probabilities of extreme values, each from a distinct site. The sample can be treated as being drawn from a Gumbel distribution, irrespective of the underlying hazard-generating mechanisms or the statistical hazard models. The method is demonstrated by an analysis of the extreme wind gust data in South Australia. The results are compared to the specifications in the Australian standard AS/NZS 1170.2:2021 and indicates that the standard may have overestimated the wind gust hazard, hence the specified design wind speeds may fall on the conservative side for South Australia.

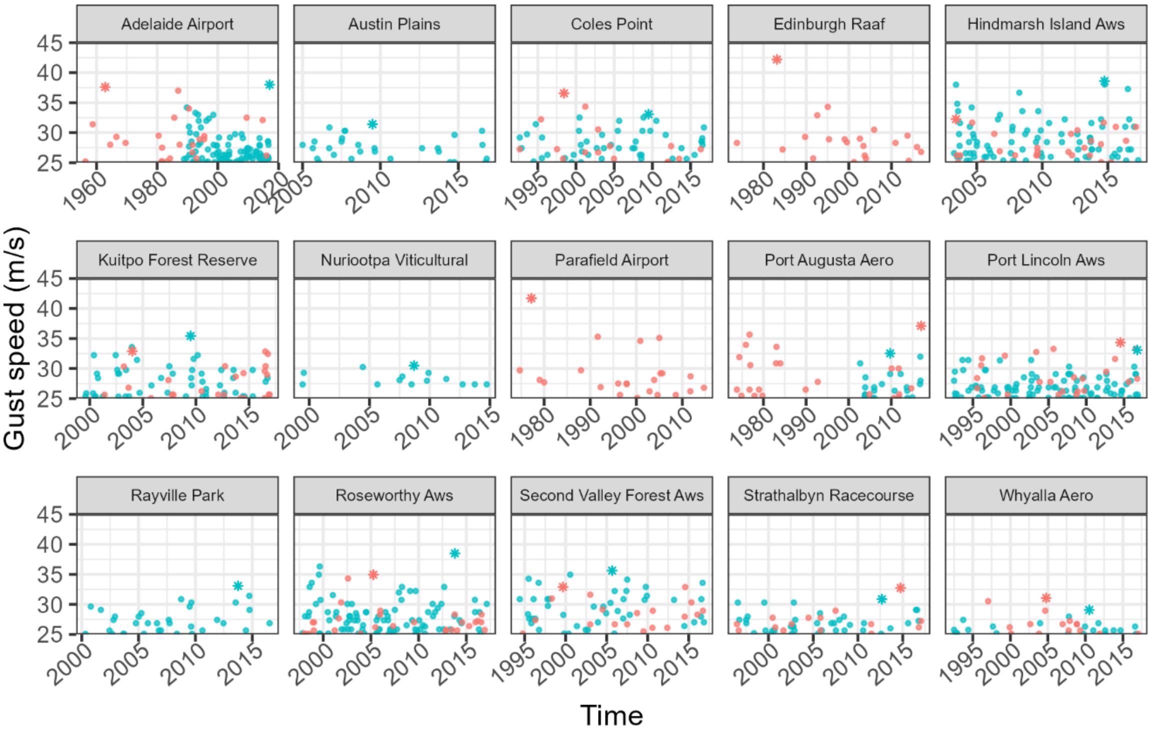

Figure 1. Three-second wind gust speeds exceeding 25 m/s (blue: synoptic; red: non-synoptic)

Keywords: *Extreme values, parameter estimation, tail index, bias reduction, wind hazard*

1. INTRODUCTION

Estimates of extremal return levels at high average recurrence intervals (ARIs) are strongly dependent on the shape parameter of the statistical model. Especially with only a few tens of years of observed data, high bias and uncertainty invariably exist for parameter estimation based on data from an individual observational stations. For instance, with a short record of extreme wind gust data, say around 20 years, the unreliably estimated shape parameter could lead in some cases to the prediction errors of 1000-year wind speeds up to a few hundred percent (Simiu and Filliben, 1975). A common approach to ameliorate this shortcoming is the 'super-station' (or station-year) approach (Wang et al., 2013) which takes advantage of having multiple weather stations with valid records in a climatologically uniform region. A 'super-station' extends the length of record and should reduce the uncertainty in high ARIs; however, the problem of potentially high bias remains for return levels beyond the accumulated record length of the super-station.

With regards to natural hazards, for instance, if multiple regions of different climatic conditions are considered, the hazard in each region is analysed based on the data collected in the region, hence each region has its own hazard model parameters, and in such cases the model estimation processes would most likely give different values of the shape parameters for different regions. In practices, it may be decided to use a specific shape parameter value across all the regions. For instance, the Australian design standard for wind actions (Australian / New Zealand Standard, 2021) uses a shape parameter of 0.1 for all of the four wind regions specified. However, the specific values were chosen based either on the consensus of judgment or heuristic averaging of the set of shape parameters derived from analysing the wind gust data (Holmes, 2002). Either way, there is a lack of objective criteria or theoretical basis to derive the best shape parameter value. In addition, insufficient record length of observational data poses a serious challenge to derive a shape parameter value with high confidence for extrapolation to hundreds of years beyond the record length.

van den Brink and Können (2011) introduced a concept in terms of return period for modelling the probabilities of occurrences of the maximum values from each record of an ensemble of independently collected records and found that the logarithmically transformed return period of a maximum value is approximately a standard Gumbel variate. Instead of return period, this paper derives the concept by means of the relationship between the exceedance probability and ARI, and shows that the log-transformed ARI follows exactly the standard Gumbel distribution. In addition, the use of ARI along with exceedance probability enables more flexible choice between the two most commonly used distributions: the generalised extreme-value distribution (GEV) and the generalised Pareto distribution (GPD) (Wang and Holmes, 2020).

For large-scale synoptic wind gusts, extracting annual extremes is straightforward, whereas for non-synoptic winds such as thunderstorms, downbursts, and tropical cyclones, which may not occur every year, the wind gusts over a sufficiently high threshold from independent events may be taken for analysis. In this context, the GEV is conceptually a block-maxima (BM) method, whereas the GPD conforms to the peaks-overt-threshold (POT) method. Despite the conceptual distinction in data extraction, the GEV and GPD possess a duality relationship that admits the parameters of one distribution to be converted to that of another (Wang and Holmes 2020). Since the GEV requires one less (i.e. the rate of threshold exceedance) model parameters than the GPD, it is employed for gust hazard analysis.

2. METHOD

The method described herein is applicable to an ensemble of extremal data records that may be produced by different mechanisms from different geographical regions as it involves only the maximum value of each record. For derivation, we will concentrate on the POT method. If the occurrence of extremes exceeding a high threshold follows a Poisson process, the relationship between return period *R* and average recurrence interval *A* can be shown to be $1/R = 1 - \exp(-1/A)$, where *R* is conventionally defined as the inverse of the probability of exceedance (Wang and Holmes 2020). Assuming that the probability distribution of annual extreme *Y* is $F_Y(y) = P\{Y \le y\}$, then the probability of *Y* less than y_a, the *y* value at *a*-year ARI, is

$$F_Y(y) = P\{Y \le y_a\} = e^{-1/a} \tag{1}$$

Let Y_n be the extreme value of *Y* in an *n*-year time interval, then $P\{Y_n \le y_a\} = \left[P\{Y \le y_a\}\right]^n = e^{-n/a}$. If *Y* is continuous, for every *y* in *Y*, there is a unique *a* in *A*. If $F_A(a)$ is the distribution function of *A*, then $F_A(a) = P\{A \le a\} = e^{-1/a}$. Let A_n be the corresponding ARI (in a reference time interval of *n* years) of Y_n,

$$P\{A_n \le a\} = \left[P\{A \le a\}\right]^n = e^{-n/a} = \exp\left[-\exp\left(-\left(\ln a - \ln n\right)\right)\right] \tag{2}$$

If we further define $\Delta L_A = \ln A - \ln n$, then ΔL_A is also a random variable with the probability distribution,

$$P\left\{\Delta L_A \le \Delta \hat{L}_A\right\} = \exp\left[-\exp\left(-\Delta \hat{L}_A\right)\right] \tag{3}$$

which is the standard Gumbel distribution. If $F_Y(y)$ is perfectly known, any given y_a and the corresponding *a* will be known. In this case, if we choose, e.g. $a = n$, $\Delta \hat{L}_A$ will be zero. In practical situations, however, $F_Y(y)$ is not known, an observed quantity y_a would correspond to an unknown *a*, which effectively constitutes a random sampling problem with interest of obtaining an estimated *a* or equivalently $\Delta \hat{L}_A$.

Suppose that there is a sample $\Delta \hat{L}_{A_i}$ of size *m*, $i = 1,\ldots,m$, from *m* stations, where $\Delta \hat{L}_{A_i}$ is the maximum among the N_i values recorded at station *i*, then provided that the $\Delta \hat{L}_{A_i}$'s are independent, the extreme-value distribution functions $F_{Y_{n_i}}\left(y_{a_i}\right)$ used to model each of the *m* records are allowed to be different. Consequently, records from different climatological regions may be combined to gauge the conformance of $\Delta \hat{L}_{A_i}$'s to the standard Gumbel distribution. In other words, this property allows different records of extremes to 'learn' from the experience of others by pooling the $\Delta \hat{L}_{A_i}$'s as if they were a sample drawn from a standard Gumbel variate.

Even though ΔL_A is independent of the underlying $F_{Y_i}(y_i)$'s, in practical situations, $F_{Y_i}(y_i)$'s are typically unavailable and still needs to be substituted by an empirically determined distribution $\tilde{F}_{Y_i}(y_i)$ with the distribution parameters being estimated from observational data. That is, if the N_i extreme values of Y_i are arranged in ascending order, $y_{(1_i)} \le y_{(2_i)} \le \cdots \le y_{(N_i)}$, then $\Delta \hat{L}_{A_i}$ based on $\tilde{F}_{Y_i}\left(y_{(N_i)}\right)$ is obtained. As $\Delta \hat{L}_{A_i}$ is determined by the largest value from station *i*, special attention should be paid to ensure that all the *m* largest values are contributed by independent extreme events. If one event contributes to multiple $\Delta \hat{L}_{A_i}$'s, the $\Delta \hat{L}_{A_i}$ value representing the highest ARI among them is kept in the analysis but all others triggered by the same event should be discarded.

The duality of the GEV and GPD ensures that the outcomes and conclusion drawn for the GEV should be equally applicable to the GPD (Wang and Holmes, 2020). The GEV may be expressed as

$$P\{Y \le y_a\} = \exp\left[-\left(1 - k\left(y_a - \eta\right)/\sigma\right)^{1/k}\right] \tag{4}$$

where *η*, *σ*, and *k* are the location, scale, and shape parameters, respectively, of the distribution. y_a can be related to its corresponding *a* as follows,

$$y_a = \begin{cases} \eta + \sigma\left(1 - a^{-k}\right)/k, & \text{if } k \ne 0; \\ \eta + \sigma \ln a, & \text{otherwise.} \end{cases} \tag{5}$$

For a given wind type (e.g. non-synoptic) at a station, suppose there are *N* extreme gust speeds exceeding a specified threshold in *n* years and the occurrence of exceedance obeys a Poisson process, then an unbiased estimate of the rate of exceedance, $\lambda_j, j = 1,\ldots,N$, with respect to the *j*-th smallest gust speed may be estimated by $\lambda_j = \left(N - j + 1\right)/n$ (Ang and Tang 2007). For simplicity and without loss of generality, the least-squares linear (for cases with fixed shape parameter) and nonlinear (for cases with free shape parameter) regression techniques for the wind gust speed on ARI were used for model parameter estimation.

3. DATA

The first Dines anemometer in South Australia became operational around 1956. Three datasets of 3-second wind gust speeds, recorded at 10 meters high, were acquired: half-hourly data (up to January 2015) from 64 stations, daily data (up to May 2017) from 76 stations, and one-minute data (up to May 2017) from 69 stations. After data screening, some of the stations were eliminated because of a high percentage of missing data, suspect recordings, or complicated topographical surroundings that make highly doubtful a gust speed could be corrected to terrain category 2 (i.e. open terrain) exposure as specified by Australian / New Zealand Standard (2021). This gives the longest record length of around 30 years among all the stations. Because of the short record lengths of the datasets, the convective windstorms such as downbursts, thunderstorms, and tornadoes were grouped as non-synoptic wind events. Other non-convective, large-scale events were grouped as synoptic wind events. Synoptic and non-synoptic winds were considered separately and only the records with data length ≥ 10 years were kept for analysis. This leaves 13 stations for synoptic and 12 stations for non-synoptic winds. Ten of the stations had both synoptic and non-synoptic wind records.

Extreme wind gust analysis requires different recorded gust speeds be generated by different storm events. To reduce the inadvertent inclusion of multiple peak gust speeds from the same wind event, minimum separation intervals of 4 days for synoptic and of 12 hours for non-synoptic wind gusts (Lombardo et al. 2009) were specified. The recorded 3-second gust speeds were corrected for the effects of terrain, topography and of shielding by nearby plantation and construction in the cardinal and inter-cardinal directions around each station in accordance with the AS/NZS 1170.2:2021. The 3-second gust speeds were then converted to 0.2-second gust speeds (Holmes and Ginger 2012). For the wind gust hazard modelling of both wind types in the next section, only those exceeding a threshold of 25 m/s were retained for analysis, as shown in Figure 1.

4. RESULTS

4.1. Shape parameters for component wind hazards

Different shape parameter values of a GEV distribution fitted to a wind gust record led to different estimates $\Delta\hat{L}_A$ of ΔL_A. This section illustrates computation of $\Delta\hat{L}_{A_i}$'s given m data series of a wind hazard type (synoptic, non-synoptic, or combined wind hazard) and determination of a shape parameter. The resulting shape parameter gives the best fit of $\Delta\hat{L}_A$ to the standard Gumbel distribution.

The m independent $\Delta\hat{L}_{A_i}$'s are plotted against the theoretical quantiles of standard Gumbel variate determined by a plotting position formula (e.g. Cunnane, 1978). If the plotted data points fall closely along the diagonal line, the chosen hazard model and its assumed parameters are consistent with that implied in (3). However, the m data points falling below (above) the diagonal line means the model underestimate (overestimate) the ARI value. If the data points form a linear trend that crosses the diagonal line with slope < 1, then it means the hazard model may have too many parameters and hence may be inappropriate for predicting the extreme values of ARIs beyond the record length (van den Brink and Können, 2008).

A range of shape parameter k values was used to fit the 13 synoptic and 12 non-synoptic wind records. The root mean squared errors (RMSE's) between $\Delta\hat{L}_A$ and its idealised counterpart from the standard Gumbel variate were computed. The best k value was chosen based on the minimisation of RMSE. Figure 2 shows the RMSE versus k, in which $k = 0.2$ for synoptic and $k = 0.25$ for non-synoptic (shown as star-shaped points) were revealed to be optimal.

The Gumbel quantile-quantile (Q-Q) plots in Figure 3(a) shows that the GEV models with fixed $k = 0.2$ for synoptic and $k = 0.25$ for non-synoptic approximately follow the diagonal line, hence agrees with the theory implied in (3). The combined winds (red connected points, Figure 3(a)) follow also the standard Gumbel distribution, as asserted in Section 2.

Instead of fixing the k values, if all three GEV distribution parameters were determined for each of the m data records, Figure 3(b) shows that the lines connecting $\Delta\hat{L}_A$'s cross the diagonal line, meaning that they are overestimated (above the diagonal line) in the lower-value range but underestimated (below the diagonal line) in the higher-value range, and hence do not follow the standard Gumbel. This implies that the fitted models may be biased and hence inappropriate for extrapolation to high ARI levels. The conundrum may be of a consequence that, with free shape parameter, the GEV has too many parameters such that the fitted models exhibit unacceptably high extrapolation bias to

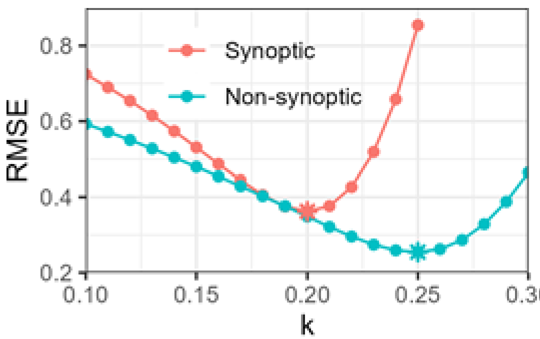

Figure 2. RMSE's between $\Delta\hat{L}_A$ and standard Gumbel variate versus shape parameters for synoptic and non-synoptic wind hazards

high return levels. Fixing the k value would avoid such unfavourable bias, and hence a sensible decision for more reliably determining the design wind speeds at ARIs beyond the available data lengths.

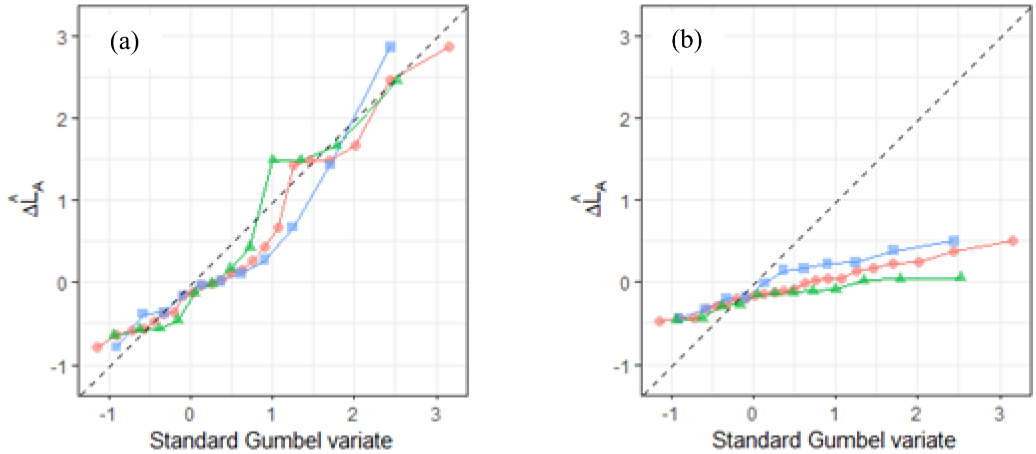

Figure 3. $\Delta\hat{L}_A$ for wind gust models with (a) fixed and (b) free shape parameters on Gumbel Q-Q plots (red: commingled; blue: synoptic; green: non-synoptic)

4.2. Shape parameters for combined wind hazard

The synoptic and non-synoptic wind events at a location pose different extent of threats as they are induced by different climatic mechanisms. Probability theory can be employed to combine hazards of different, independent mechanisms (Wang and Holmes 2020). However, no closed-form probability distribution is available for the combined hazard, and no simple expression exists for the shape parameter of the combined model. As a result, Monte-Carlo simulation was conducted to generate annual gust speeds for 1000 years from the best models (i.e. $k = 0.2$ for synoptic and $k = 0.25$ for non-synoptic) for each of the 10 locations where the records of the two wind types were available. For a given year at a specific site, the maximum of the two generated gust speeds was taken as the extreme speed of the year.

As an approximation and for comparison with the Australian Standard, the generated combined gust speeds were fitted to a GEV distribution. Similar to the model parameter estimation described in Section 4.1, we computed the RMSE between estimated $\Delta\hat{L}_A$ and the standard Gumbel quantiles, and plotted it over a range of k values (Figure 4). It shows that the optimal k is about 0.16. Incidentally, this k value is close to $k = 0.161$ by Holmes and Moriarty (1999) for thunderstorm downbursts at Moree, New South Wales, Australia, which is located in wind hazard region A, the same region as the locations studied herein.

For the combined hazards, Figure 5 shows the Gumbel Q-Q plot of $\Delta\hat{L}_A$ by the GEV models with $k = 0.16$, which indicates that the simulated wind hazards agree with the theory, whereas with k =0.10 the hazard models overestimate the hazard (i.e. underestimate the ARI). The simulated maximum wind speed of 1,000 years among the 10 stations is $V_{\max}$ = 46.2 m/s at Port Augusta Aero. For $k = 0.16$ with $\Delta\hat{L}_A = 2.33$, $V_{\max}$ is predicted to have an ARI of 10,229 years (close to 10,000 years inferred by the 10 stations simulated independently for 1000 years), whereas for $k = 0.10$ with $\Delta\hat{L}_A = 0.22$, it is predicted to have an ARI of 1,250 years. Incidentally, in AS/NZS 1170.2:2021, k = 0.10 is used for all the four wind regions, the regional wind speed of ARI =1000 years for Region A (where the studied area is located) is 46 m/s. This comparison shows that the computed results agree well with the Australian standard and implies that the standard may have overestimated the wind gust hazard for South Australia. Incidentally, a recent study (El Rafei et al. 2023) on the wind gust hazard in New South Wales, Australia, using high-resolution

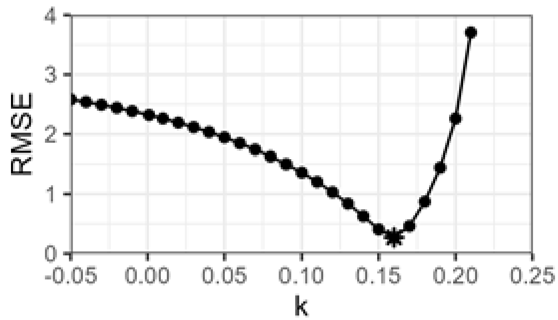

Figure 4. RMSE between $\Delta\hat{L}_A$ and standard Gumbel variate versus shape parameter for the combined wind hazard

Australian regional reanalysis found that using k = 0.10 overestimates the 500-year ARI gust speeds, when compared to that using variable k values, by approximately 4% for non-synoptic and 2.5% for synoptic events.

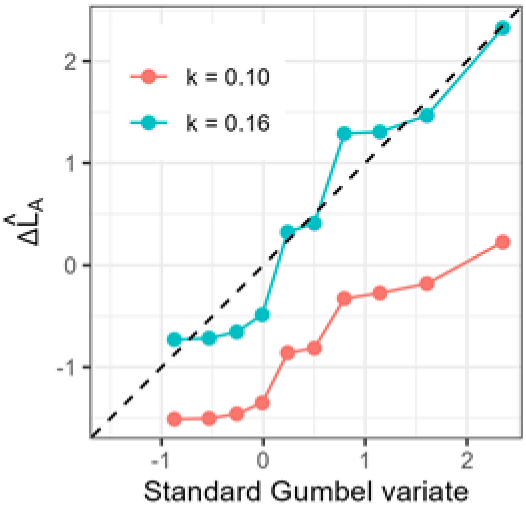

Figure 5. $\Delta\hat{L}_A$ for combined wind gust models on Gumbel Q-Q plot

Figure 6 illustrates the fitted combined hazard models with k = 0.10 and k = 0.16 along with the simulated annual extreme gust speeds for the 10 locations. Compared with k = 0.10, the curves with k = 0.16 provide closer fit to the data points in most locations and result in lower gust speeds at high ARI years. On average, the models with k = 0.10 give about 2.9% and 3.5% higher gust speed estimates than that with k = 0.16 for ARIs of 500 and 1000 years, respectively. That is, the Australian standard-specified design wind speeds for South Australia generally fall on the conservative side with respect to the design for wind actions of structures specified as of importance levels 2 (domestic housing and structures under normal operations) and 3 (construction designed to contain a large number of people) in the 2019 National Construction Code of Australia (Australian Building Codes Board 2019). Nevertheless, dependent upon the balance of the benefits gained versus the costs incurred due to the more conservative design wind speed, the resulting higher cost but more conservative construction may be justified if the additional benefits gained are deemed to outweigh the extra costs incurred.

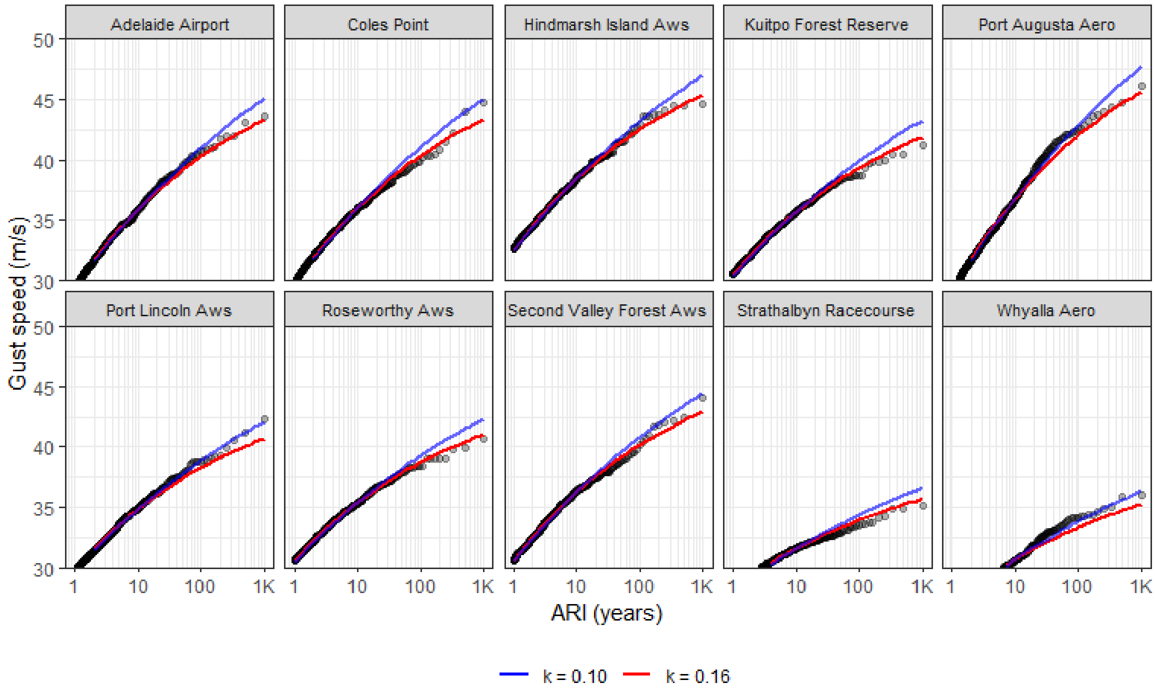

Figure 6. Simulated versus fitted gust speeds for combined gust hazards.

5. CONCLUSION

Because many currently available extreme data of climatic events such as observational extreme wind gusts span only a few tens of years, resulting in high bias and uncertainty of distribution parameter values estimated based on the data, and hence unreliable predicted extreme values when extrapolated to high ARIs beyond the range of data length. For the quality of model fitting, the typical goodness-of-fit tests allows assessment within the record length, but unable to test the fitness for higher ARIs. The approach used in this study serves to gauge whether the fitted model is appropriate and unbiased for extrapolation, providing a mechanism to safeguard the accuracy of fitted values longer than the record length or at high ARIs, which is what's typically needed for engineering design and reliability assessment.

Instead of return period, this paper demonstrates the advantage of using ARI for modelling the occurrence of extremes as the ARI is proved to follow an inverted exponential distribution and the log-transformed ARI (i.e. $\Delta\hat{L}_A$) becomes Gumbel distributed. Moreover, the ability of the method in pooling the ARIs of maximum recorded data points from all observational stations, even when they are from regions of different hazard-generating mechanisms, may be regarded as a generalisation of the 'super-station' approach that has been used to commingle all the extreme wind gust or rainfall data from a climatologically uniform region. Therefore, the method is useful for cases such as the wind gust speed specification in the Australian standard in which a shape parameter value is applied across the four wind regions.

Although it often occurs in Australia that non-synoptic wind gusts dominate the extreme wind climate, particularly at larger ARIs, consideration of both synoptic and non-synoptic is necessary as the combined wind hazard tends to have a smaller shape parameter value that, compared to a larger one, leads to higher wind gust values at high ARIs. In addition, synoptic wind gusts at some locations dominate at smaller ARIs, which is important for construction of temporary and secondary structures such as formwork, circus tents, and farm shelters that are intended to be in services for only a short period of time. The analysis of wind gust data from South Australia indicates that the shape parameter value of 0.1 used in the Australian standard, AS/NZS 1170.2:2021, may be too low that it appears to lead to overestimate the wind hazard, and hence fall on the conservative side, in South Australia.

As with typical experimental and observational studies, the accuracy of estimation by the method used in this paper clearly depends on the accuracy of measurement, the quality of the data collection and processing, and the classification of the right hazard-generating mechanism when heterogeneous mechanisms are concerned. In addition, the method assists only in selecting the hazard model and its fitted parameters for bias reduction of prediction at high ARIs but does not provide uncertainty estimates of the prediction. If preferred, uncertainty quantification may be conducted by the bootstrap or Bayesian analysis. On the other hand, the method can be used in conjunction with uncertainty minimization techniques as well as within-data goodness-of-fit tests to help obtain a model of minimised extrapolation bias and variance.

REFERENCES

Ang, A.H-S., Tang, W.H., 2007. Probability concepts in engineering: emphasis on applications in civil & environmental engineering. 2nd ed. John Wiley, New York.

Australian / New Zealand Standard., 2021. Structural design actions Part 2: Wind actions. AS/NZS 1170.2:2021.

Australian Building Codes Board. 2019. National Construction Code Volume 1.

Cunnane, C., 1978. Unbiased plotting positions — A review. Journal of Hydrology 37 (3-4), 205–22.

El Rafei, M., Sherwood, S., Evans, J.P., Ji, F., 2023. Analysis of extreme wind gusts using a high-resolution Australian Regional Reanalysis. Weather and Climate Extremes 39, 100537.

Holmes, J. D., 2002. A Re-analysis of Recorded Extreme Wind Speeds in Region A. Australian Journal of Structural Engineering 4 (1), 29–40.

Holmes, J.D., Ginger, J.D., 2012. The gust wind speed duration in AS/NZS 1170.2. Australian Journal of Structural Engineering 13 (3), 207–16.

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 83 (1-3), 1–10.

Lombardo, F.T., Main, J.A., Simiu, E., 2009. Automated extraction and classification of thunderstorm and non-thunderstorm wind data for extreme-value analysis. Journal of Wind Engineering and Industrial Aerodynamics 97 (3-4), 120–31.

Simiu, E., Filliben, J., 1975. Statistical Analysis of Extreme Winds. National Bureau of Standards.

van den Brink, H.W., Können, G.P., 2008. The statistical distribution of meteorological outliers. Geophysical Research Letters 35 (23), 1–5.

van den Brink, H.W., Können G.P., 2011. Estimating 10000-year return values from short time series. International Journal of Climatology 31 (1), 115–26.

Wang, C.-H., Holmes, J.D., 2020. Exceedance rate, exceedance probability, and the duality of GEV and GPD for extreme hazard analysis. Natural Hazards 102 (3), 1305–21.

Wang, C.-H., Wang, X., Khoo, Y.B., 2013. Extreme wind gust hazard in Australia and its sensitivity to climate change. Natural Hazards 67 (2).