Nineteenth Australasian Wind Engineering Society Workshop, April 4-6, 2018, Torquay, Victoria

The Equivalence of Generalized Extreme-Value Distribution and Generalized Pareto Distribution for Wind Gust Analysis

Chi-Hsiang Wang[1], John D. Holmes[2]

[1]*Land and Water, CSIRO, Private Bag 10, Clayton South, VIC 3169, Australia; chi-hsiang.wang@csiro.au.*

[1]*JDH Consulting, 6 Charman Road, Mentone, VIC 3194, Australia; john.holmes@jdhconsult.com.*

ABSTRACT

Conventionally the generalized extreme-value distribution is applied to peak wind gust datasets collected as block maxima, whereas the generalized Pareto distribution is applied to the datasets that exceed specified high thresholds. We show that this distinction in applications of the two statistical models is unnecessary as the parameters of one model can be computed from the parameters of the other. The equivalence is derived from the fact that the gust occurrences exceeding a high threshold are approximated by a Poisson process and the relationship between the mean inter-arrival time of the Poisson process and the commonly used return period. We demonstrate the equivalence of the two models by analysis of wind gust data collected at Woomera, South Australia.

1. Introduction

For statistical analysis of extreme wind gust speeds, the generalized extreme value distribution (GEV) and generalized Pareto distribution (GPD) are among the most widely used statistical models. Because of their statistical characteristics, the GEV is regarded as a block-maximum method that applies to a sample of maximal gust speeds collected from fixed time intervals (e.g. annual maxima), whereas the GPD is a peaks-over-threshold method that applies to a sample of gust speeds exceeding a sufficiently high threshold (Wang *et al.* 2013; Holmes 2015).

This paper shows that, in applications, the GEV and GPD are equivalent statistical models. That is, given a set of model parameter values of a distribution, the model parameter values of the other can be determined accordingly. For peaks-over-threshold methods, it is assumed that the occurrence of wind gusts exceeding a high threshold follows a Poisson process, hence the inter-arrival time of a gust speed follows an exponential distribution. For block-maxima methods, on the other hand, the maximum in a defined time interval (typically one year) is taken, hence the occurrences of annual maxima constitute Bernoulli sequences with geometrically distributed recurrence time; its mean recurrence time is popularly known in engineering as the return period. Because the mean inter-arrival time can be related to return period, a sample of gust speeds, whether it be produced by way of block maxima or peaks-over-threshold, may be analysed either by GEV or GPD.

We present the equivalence between the model parameters of the GEV and GPD, and demonstrate model parameter estimation using wind gust data recorded by the Australian Bureau of Meteorology at automatic weather station Woomera, South Australia. The findings allow researchers to use either the GEV or GPD for wind gust hazard analysis, regardless of whether the wind gust data are collected as block maxima or via peaks over a threshold.

2. Exceedance Rate, Mean Inter-Arrival Time, and Return Period

Assume that the occurrence of a wind gust exceeding a high threshold (e.g. 25 m/s) follows a Poisson process. Then the inter-arrival time of gusts follows an exponential distribution with mean inter-arrival time, T , as the scale parameter, and $1/T$ as the rate of exceedance. The probability of T greater than a value t is

$$P(T > t) = 1 - e^{-1/t} \tag{1}$$

By conventional definition in engineering applications, the inverse of exceedance probability in a given time interval is termed return period, R. Therefore,

$$\frac{1}{R} = 1 - e^{-1/t} \tag{2}$$

Figure 1 shows that, for gust hazard estimation, return period and mean inter-arrival time become virtually the same for return periods greater than 10.

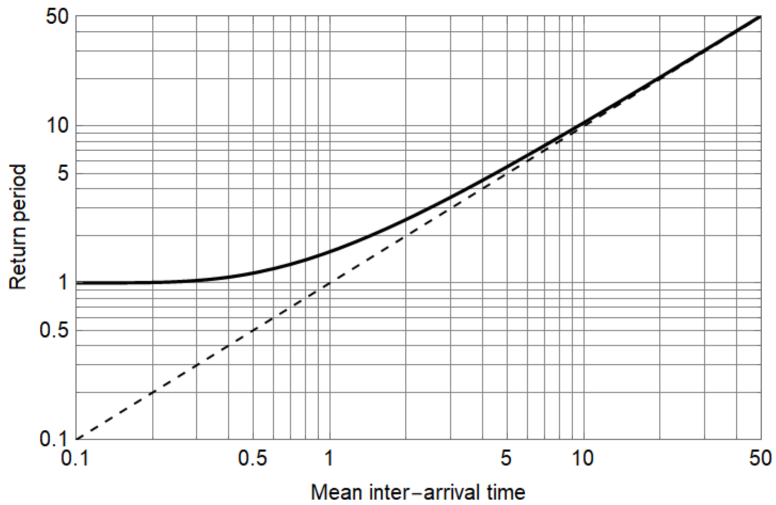

Figure 1. Relationship between return period and mean inter-arrival time.

For a sample of size n observed gust speeds in N years, the rate of exceedance λ_i of the i -th smallest gust speed may be estimated by

$$\lambda_i = \frac{n - i + 1}{N} \tag{3}$$

The empirical mean inter-arrival time t_i is the inverse of λ_i. Hence the corresponding return period R_i of the i -th smallest gust speed is

$$\frac{1}{R_i} = 1 - \exp\left(-\frac{n - i + 1}{N}\right) \tag{4}$$

Incidentally, the term mean inter-arrival time, commonly used in stochastic processes to signify the mean recurrence time of events, is coined in a guide for Australian rainfall and runoff (Institution of Engineers Australia, 1987) as "average recurrence interval", which is adopted also by the Australian Bureau of Meteorology (http://www.bom.gov.au/water/designRainfalls/ifd/glossary.shtml, accessed 10 March 2018).

3. Relationship between GPD and GEV Distributions

The generalized Pareto distribution is expressed by

$$F_Y(y) = 1 - \left(1 - k_p \frac{y}{\sigma_p}\right)^{1/k_p} \tag{5}$$

where y (m/s) is the excess over a threshold; σ_p (m/s) and k_p are the scale and the shape parameters, respectively, of the distribution.

For determination of the wind speed, v_t (m/s), at a mean inter-arrival time, t (years), the rate of exceedance, λ_p (times/year), of the threshold, v_p, is required. If we equate the mean crossing rate of the level v_t to $1/t$; i.e. $\lambda_p\left[1 - k_p\left(v_t - v_p\right)/\sigma_p\right]^{1/k_p} = 1/t$, then after rearranging we have

$$v_t = \begin{cases} v_p + \dfrac{\sigma_p}{k_p}\left[1 - \left(\lambda_p t\right)^{-k_p}\right], & \text{when } k_p \neq 0; \\ v_p + \sigma_p \ln\left(\lambda_p t\right), & \text{when } k_p = 0. \end{cases} \tag{6}$$

When $k_p > 0$ and $R \to \infty$, $v_t \to v_p + \sigma_p / k_p$, meaning that the predicted extreme wind speeds have a limiting value at high return periods; when $k_p \leq 0$, v_t is unbounded from above.

The generalized extreme-value distribution may be expressed by

$$F_V(v_t) = e^{-\left[1 - k_g\left(\frac{v_t - v_g}{\sigma_g}\right)\right]^{1/k_g}} \tag{7}$$

where v_g (m/s), σ_g (m/s), and k_g, respectively, are the location, scale, and shape parameters of the distribution, and v is a gust speed (m/s).

Taking logarithm of Eq. (7) and using Eq. (2), we obtain

$$t = \left(1 - k_g \frac{v_t - v_g}{\sigma_g}\right)^{-1/k_g} \tag{8}$$

Therefore,

$$v_t = \begin{cases} v_g + \dfrac{\sigma_g}{k_g}\left(1 - t^{-k_g}\right), & \text{for } k_g \neq 0; \\ v_g + \sigma_g \ln t, & \text{for } k_g = 0. \end{cases} \tag{9}$$

To be equivalent, the shape parameters of GPD and GEV must have the same value; i.e. $k_p = k_g = k$. Comparing Eqs. (6) and (9), the location and scale parameters of GEV can be expressed in terms of the GPD parameters as follows,

$$\sigma_g = \sigma_p \lambda_p^{-k}$$

$$v_g = \begin{cases} v_p + \dfrac{\sigma_p}{k}\left(1 - \lambda_p^{-k}\right) & \text{when } k \neq 0; \\ v_p + \sigma_p \ln \lambda_p & \text{when } k = 0. \end{cases} \tag{10}$$

Conversely, for a given λ_p, the location and scale parameters of GPD can be expressed in terms of the GEV parameters as follows,

$$\sigma_p = \sigma_g \lambda_p^{k}$$

$$v_p = \begin{cases} v_g + \dfrac{\sigma_g}{k}\left(1 - \lambda_p^{k}\right) & \text{when } k \neq 0; \\ v_g - \sigma_g \ln \lambda_p & \text{when } k = 0. \end{cases} \tag{11}$$

If v_p is given instead of λ_p, then λ_p and σ_p are determined as follows,

$$\sigma_p = \sigma_g + k\left(v_g - v_p\right)$$

$$\lambda_p = \begin{cases} \left(1 + \dfrac{v_g - v_p}{\sigma_g} k\right)^{1/k} & \text{when } k \neq 0; \\ \exp\left(\dfrac{v_g - v_p}{\sigma_g}\right) & \text{when } k = 0. \end{cases} \tag{12}$$

4. Example: Analysis of Woomera AWS Station Data (1991—2017)

The Bureau of Meteorology weather station at Woomera, SA, was changed to automatic weather station since 1991, hence a total of 26 years of wind gust data up were recorded up to 2017. We considered the peak gust wind speeds exceeding 25 m/s from independent non-synoptic (i.e. thunderstorms and frontals) events, resulting in a sample of 23 gust speeds exceeding the threshold.

For the dataset considered, the rate of threshold exceedance λ_p is estimated as $23/26 = 0.8846$ with a threshold v_p of 25 m/s. If $\lambda_p = 0.8846$ is used, all the other GPD model parameters (i.e. v_p, σ_p, and k) are estimated by least-squares regression analysis, as shown in Table 1, along with the GEV model parameters determined by Eq. (10). The fitted models are shown in Figure 2 for wind speed vs return period and wind speed vs mean inter-arrival time. We see that the GPD and GEV give identical fitted results.

Table 1. Estimated GPD and GEV model parameters assuming $\lambda_p = 0.8846$.

Model	λ_p	v_p or v_g	σ_p or σ_g	k
GPD	0.8846	25.50	5.677	0.319
GEV	—	24.79	5.903	0.319

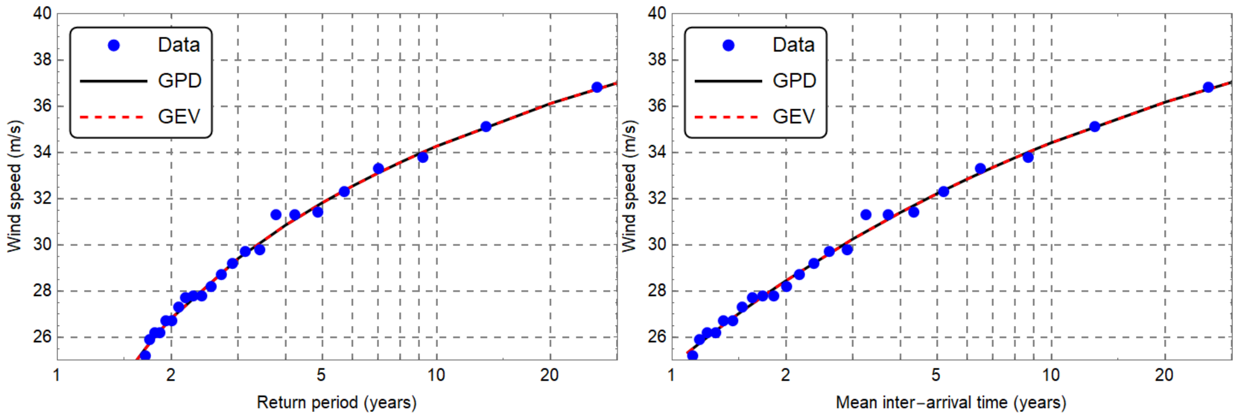

Figure 2. Fitted GPD and GEV models assuming $\lambda_p = 0.8846$; (left panel) wind speed vs return period; (right panel) wind speed vs mean inter-arrival time.

Due to sampling error of finite datasets, the k value determined above may be deemed unrealistically large. If, to be consistent with AS/NZS 1170.2:2011 (Standards Australia 2011), the shape parameter $k = 0.1$ is pre-specified in addition to the exceedance rate. The estimated model parameters are shown in Table 2 and the fitted results are shown in Figure 3. The fitting in Figure 3 in terms of least-squares criterion is not as good as in Figure 2 because more constraints are imposed by specifying $k = 0.1$, but it gives higher estimated gust speed (i.e. more conservative for design) than $k = 0.319$.

Table 2. Estimated GPD and GEV model parameters assuming $\lambda_p = 0.8846$ and $k = 0.1$.

Model	λ_p	v_p or v_g	σ_p or σ_g	k
GPD	0.8846	25.91	4.342	0.100
GEV	—	25.38	4.396	0.100

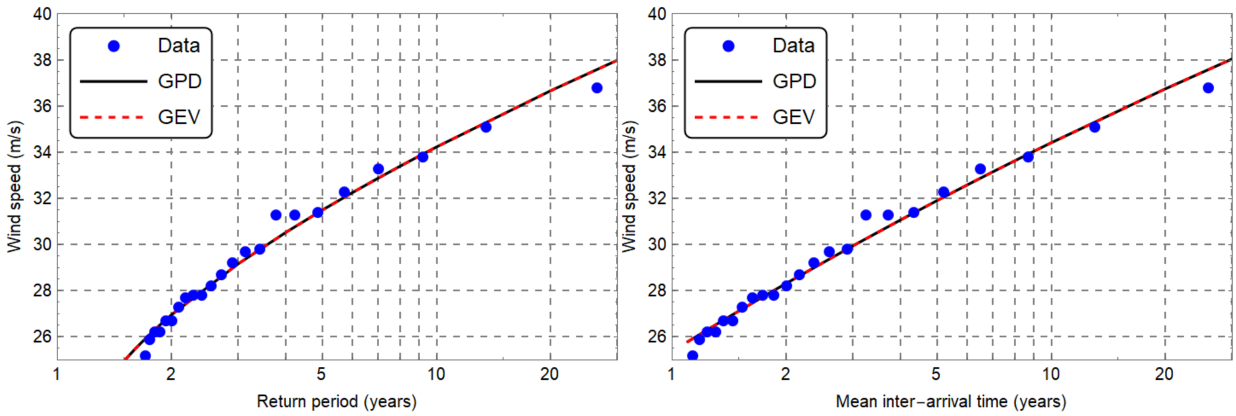

Figure 3. Fitted GPD and GEV models assuming $\lambda_p = 0.8846$ and $k = 0.1$; (left panel) wind speed vs return period; (right panel) wind speed vs mean inter-arrival time.

In practice, the empirical exceedance rate and gust speed threshold are often specified to be equal to the observed values (i.e. $\lambda_p = 0.8846$, $v_p = 25\,\text{m/s}$). If $k = 0.1$ is specified as well, then only σ_p in the GPD model is to be determined. A least-squares regression gives $\sigma_p = 4.982$. Figure 4 compares the GPD model derived herein with that shown in Figure 3. We see that the GPD fitting becomes deviant from the data points. Therefore, pre-specifying three of the four GPD model parameters may impose unacceptable constraints that result in sub-optimal model fitting.

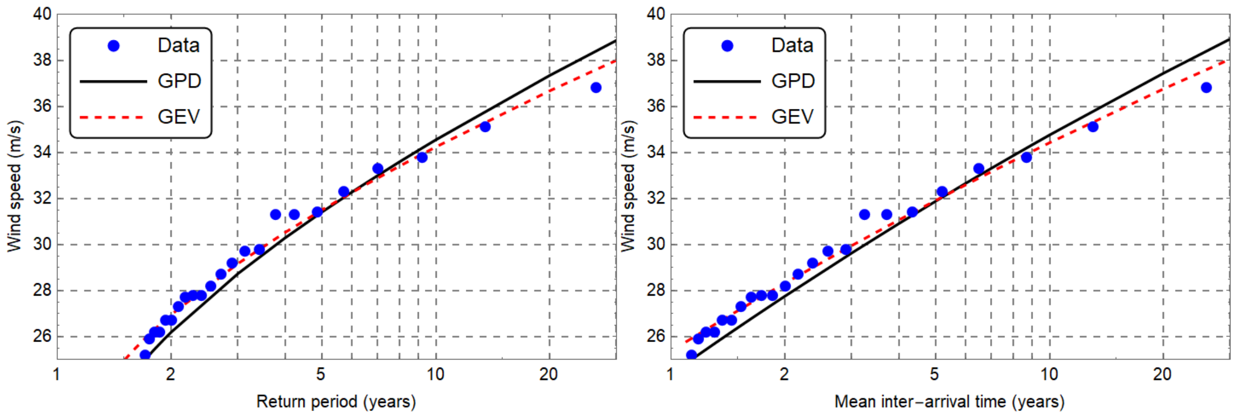

Figure 4. Fitted GPD and GEV models assuming $\lambda_p = 0.8846$, $v_p = 25\,\text{m/s}$, and $k = 0.1$; (left panel) wind speed vs return period; (right panel) wind speed vs mean inter-arrival time.

5. Concluding Remarks

Because of their statistical characteristics, conventionally the generalized extreme-value distribution is applied to datasets collected as block maxima, whereas the generalized Pareto distribution is applied to that exceeding a specified high threshold. We demonstrate that the model parameters of the two can be computed from that of the other. This allows the analyst to choose a model of his/her preference for analysis, whether the dataset is collected via block maxima or peaks-over-threshold, as long as the empirical return periods (or mean inter-arrival times) of the data points are properly estimated. In addition, a couple observations are:

- For data obtained by threshold exceedance, the empirical rates of exceedance and mean inter-arrival times are estimated for regression analysis. For data obtained by block maxima, the empirical exceedance probability and return periods are first estimated and then converted to mean inter-arrival times for analysis.
- In the GPD, the exceedance rate and threshold are dependent, therefore we suggest that, for regression analysis, only one of them, but not both, is specified from the observational data. Specifying both of them may impose unacceptable constraints on regression, which will result in a sub-optimal fitted model.

References

Holmes, J.D., (2015), "Wind loading of structures", 3rd Edition, CRC Press Boca Raton, Florida, USA

Standards Australia, (2011), "Structural design actions. Part 2 Wind actions", Australian/New Zealand Standard, AS/NZS 1170.2:2011.

Wang, C-H., Wang, X., Khoo, YB. (2013), Extreme wind gust hazard in Australia and its sensitivity to climate change. *Natural Hazards*. 67:549-567.