Extreme Wind Speed Analysis for the Natural Shoreline Infrastructure Project Area

Prepared by: Jeffrey K. Anderson

Affiliation: Northern Hydrology & Engineering

Date: 29 April 2022

Introduction

To support wind-wave modeling an extreme value analysis (EVA) of maximum 2-min average daily wind speeds recorded at the Arcata/Eureka Airport was conducted. Two distributions were used for the EVA: (1) the Generalized Extreme Value distribution using annual maxima, and (2) the peaks-over-threshold and Generalized Pareto Distribution using daily maximum values. Although wind data is available at other locations in closer proximity to Humboldt Bay, it was determined that the Arcata/Eureka Airport data provided by the National Centers for Environmental Information was the longest and most reliable maximum wind speed data for EVA. All extreme value and data analysis was conducted in R (R Core Team 2021). The analysis was conducted in SI units (wind speed as meters per second (mps)), and only final tabulated results will be presented in English units.

Methods

The following sections describe the methods used for processing the wind data and conducting the extreme value analysis.

Wind Data

The Arcata/Eureka Airport (WBAN: 24283; USAF: 725945; Call: KACV) wind data were downloaded from NOAA using the R package rnoaa (Chamberlain 2021). The data used was the Integrated Surface Data (ISD) database that contains global hourly and synoptic surface observations from several data sources. For this analysis the hourly (or more frequent) and daily summary wind speed and direction observations were extracted from the ISD and further processed as described below. This is the same general wind speed data set used by the National Institute of Standards and Technology (NIST) for developing design wind speeds reported in the *Minimum Design Loads for Buildings and Other Structures* (Pinter et al. 2015, ASCE 2017).

The primary source of wind data used in the EVA was from the automated surface observing system (ASOS) which automated and standardized surface weather data collection in the United States (Powell 1993). The ISD contains pre-ASOS wind speed data spanning a much longer interval than used in the EVA. However, much of the reported pre-ASOS data was mandatory observations (e.g. hourly 1-min average wind speed) that did not represent the maximum or fastest daily wind speed, and contained a range of reported averaging periods. Two wind data types were accessed and used in the analysis: (1) the mandatory wind observations which reports approximately hourly, and (2) the daily summary wind observation which reports the fastest 2-min average wind speed and direction for the day. The hourly wind data were used for a preliminary assessment of maximum wind speed by direction due to the length and number of wind observations. The fastest 2-min average wind speed from the daily summary were the wind data used in the EVA.

Following wind data download from the ISD, several quality control (QC) checks and data processing steps were conducted on the raw data to develop a maximum daily 2-min wind speed series. To better support the EVA analysis missing daily maximum wind speed gaps were filled in with estimated values. The following describes the basic QC and data processing procedures for the hourly wind data:

1. Remove any ISD reported suspect, erroneous, or missing data (NOAA 2018).
2. Conduct first differences analysis (IOOS 2014) and remove any hourly wind speed data that is equal to or greater than 10 times the standard deviation of the adjacent lead and lag wind speeds. This represents extreme outlier removal.
3. Replace missing wind direction values in following order:
    a. Set wind direction to lag value,
    b. Set wind direction to lead value,
    c. Set wind direction to the mean daily value for the period of record.
4. Conduct first differences analysis and identify, but do not delete, any wind speed that is equal to or greater than 3 times the standard deviation of the adjacent lead and lag wind speeds.
5. Adjust wind speed to 10 m height using the following equations (CEM 2015; Pinter et al. 2015):

$$U^* = \frac{U_z(k)}{ln\left(\frac{z}{z_0}\right)} \quad (1)$$

$$U_{10} = \frac{U^*}{k} ln\left(\frac{10}{z_0}\right) \quad (2)$$

   where U^* is the friction velocity, U_z is the wind speed at height z above the surface (e.g. U_{10} is wind speed at 10 m), k is the von Karman constant (~0.41), and z_o is the roughness height at the surface (0.03 m in this analysis).

6. Adjust wind speed to 2-min average wind speed (wind speed typically reported as 1-min average or 2-min average) using the following equations (CEM 2015):

$$U_t/U_{3600} = [1.277 + 0.296 \tanh(0.9 \log(45/t))] \quad \text{for } t < 3{,}600 \text{ s} \quad (3)$$

$$U_t/U_{3600} = 1.5334 - 0.15\, log(t) \quad \text{for } 3{,}600 < t < 36{,}000 \text{ s} \quad (4)$$

   with wind speed t, U_t the wind speed at specified duration, and U_{3600} the 1-hour (3,600 s) wind speed.

7. Determine maximum daily windspeed from the available hourly data.
8. Fill in missing daily maximum wind speed and direction data by averaging the lead and lag values.

The general QC and data processing procedures for the daily summary data (this was the ASOS data that reported maximum 2-min average daily wind speed) is described below:

1. Remove any ISD reported erroneous or missing data (NOAA 2018).
2. Adjust wind speed to 10 m using equations (1) and (2) above; and convert 1-min to 2-min average wind speed using equations (3) and (4). Only six values prior to 8 Feb 2001 had to be converted.

3. Fill in missing daily maximum wind speed and direction data by using the maximum daily values from the above processed hourly data.
4. Identify any ISD (NOAA 2018) reported suspect data, but do not delete the data.
5. Filter suspect data from step 4 using the following approach:
    a. Find the maximum daily 3-sec peak wind gust data for Arcata/Eureka Airport from the Pinter et al. 2015 data set, which, by definition, should be greater than any 2-min maximum daily wind speed value. Convert the 3-sec peak gust wind speed to a 2-min threshold wind speed (upper threshold limit) using equations (3) and (4).
    b. Set any suspect data greater than the upper threshold limit to the maximum daily value from the processed hourly data.
    c. Set any suspect wind data equal to or less than the upper threshold limit, and less than the maximum daily value from the processed hourly data to the hourly data value.
    d. Keep any suspect wind data equal to or less than the upper threshold limit, and equal to or greater than the maximum daily value from the processed hourly data.

Figure 1 shows the maximum daily 2-min wind speeds from the ASOS daily summary data after steps 1 to 4 above, but prior to removal of the data (step 5) above the upper threshold limit of 23.41 mps. The upper limit threshold was determined as the maximum 3-sec peak wind speed (67.6 mph) from the Pinter et al. 2015 data set adjusted to 2-min wind speed.

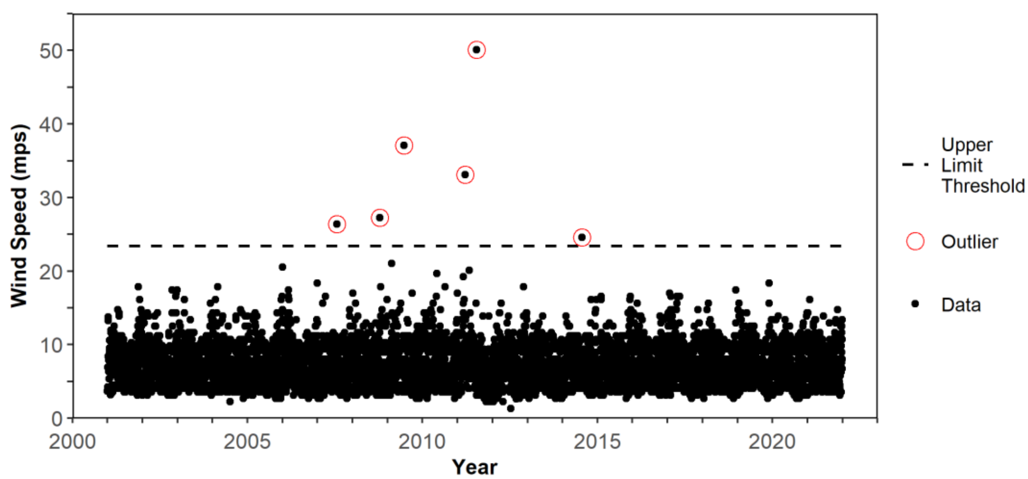

Figure 1 Maximum daily 2-min wind speeds for the ASOS daily summary data adjusted to 10 m elevation prior to removing outliers above the upper limit threshold of 23.41 mps.

The above procedures create two distinct maximum daily 2-min wind speed and direction data sets based on the ISD data used (Table 1). At first glance the longer period of record for the hourly wind data (Temp Max Day Wind Speed) would appear to be the best data for conducting the EVA. However, Figure 2 clearly shows that the shorter period data set (Final Max Day 2-min Wind Speed), based on the daily summary data, has higher maximum wind speeds and is the best data set to use for the EVA.

Table 1 Summary of processed Arcata/Eureka Airport wind speed and direction data from the ISD database used in the extreme value analysis.

Data Set	Period of Record	Number of Processed Maximum Annual Values	Number of Processed Maximum Daily Values	Data Type and Source
Temp Max Day Wind Speed	Dec. 1949 to 2021	72	26,329	Hourly 1-min and 2-min wind speed; multiple data sources
Final Max Day 2-min Wind Speed	2001 to 2021	21	7,669	Maximum daily 2-min wind speed; ASOS

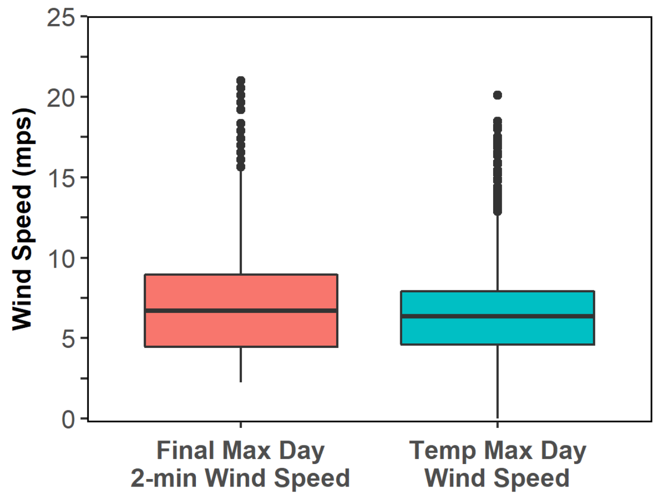

Figure 2 Box plot of maximum daily 2-min wind speeds for the Final Max Day 2-min Wind Speed data set (n = 7,669, generated from daily summary data) and the Temp Max Day Wind Speed (n = 26,329, generated from hourly data).

Extreme Value Analysis

Two different distribution functions were used in the EVA: (1) the generalized extreme value distribution using block maxima (e.g. annual maximum), and (2) the peaks-over-threshold and Generalized Pareto Distribution using daily maximum values. A theoretical definition, more detailed information, and an explanation of the parameter estimation process for both distributions can be found in Coles (2001) and Reiss and Thomas (2007). The EVA and parameter estimation were conducted with the R package extRemes (Gilleland and Katz 2016). All model distribution parameters were estimated with the maximum likelihood estimation (MLE) approach (Coles 2001).

Generalized Extreme Value Distribution

The generalized extreme value (GEV) distribution describes the distribution of block maxima, typically annual maximums. The GEV distribution cumulative function has the form

$$G(z) = exp\left\{-\left[1 + \xi\left(\frac{z-\mu}{\sigma}\right)\right]^{-1/\xi}\right\} \quad (5)$$

where z are independent random variables, and μ, σ, and ξ describe the location, scale and shape parameters, respectively.

The GEV combines the Gumbel (Type I), Frechet (Type II) and Weibull (Type III) families into one general distribution. The shape parameter defines the type of distribution, with Gumbel (Type I), Frechet (Type II) and Weibull (Type III) classes corresponding to cases with ξ = 0, ξ > 1, and ξ < 1, respectively. The shape parameter also determines the shape of the function $G(z)$. Negative shape factors (ξ < 1) have convex return curves with an upper bound return height defined by $\mu - \sigma/\xi$; positive shape factors (ξ > 1) have concave curves with unbounded return heights; and a shape parameter of zero (ξ = 0) has a linear curve with linear return heights (Coles 2001).

Once the parameters are estimated, the exceedance probability z_p is determined by inverting equation 5:

$$z_p = u - \frac{\sigma}{\xi}\left[1 - \{-log(1-p)\}^{-\xi}\right] \qquad \text{for } \xi \neq 0\text{, and} \qquad (6a)$$

$$z_p = u - \sigma \log\{-log(1-p)\} \qquad \text{for } \xi = 0 \qquad (6b)$$

where z_p is the return level with return period $1/p$.

A requirement of EVA is that the data be independent random variables (i.e. independent and identically distributed), which requires that only one wind speed from each storm system be used (Palutikof et al. 1999, Coles 2001, Lombardo et al. 2009). For wind data the use of annual maximums for the GEV analysis typically produce independent extremes (Palutikof et al. 1999). The annual maximums for the Final Max Day 2-min Wind Speed data (Table 1) are shown in Figure 3. Although there appears to be a small negative trend in the annual data, the trend slope is not significant (p > 0.1) indicating stationary data.

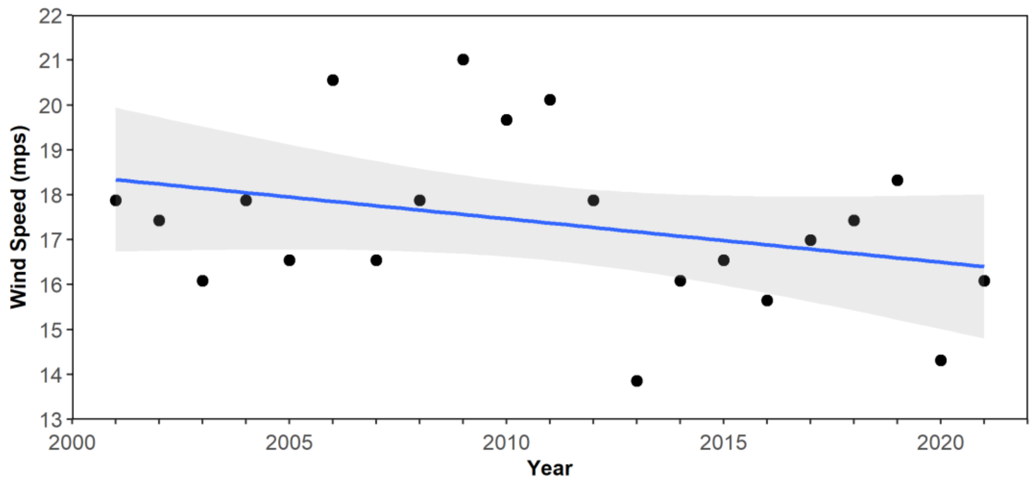

Figure 3 Annual maximum wind daily 2-min wind speeds for the Final Max Day 2-min Wind Speed data set and the linear trend slope and 95% confidence intervals.

Peaks-Over-Threshold and Generalized Pareto Distribution

A disadvantage of the GEV is the potential loss of data associated with block maxima modeling, such as the annual maximum (Coles 2001). An alternative approach which avoids this disadvantage is the peaks-over-threshold (POT) method that uses all extreme events over a high threshold and fits these exceedances to a generalized pareto distribution (GPD), henceforth known as the POT/GPD method. The GPD distribution cumulative function has the form

$$H(y) = 1 - \left(1 + \frac{\xi y}{\sigma_u}\right)^{-1/\xi} \tag{7}$$

where $y = x - u$ is the threshold excess conditional on $x > u$, u is the threshold value, x are independent random variables, and $\sigma_u = \sigma + \xi(u - \mu)$, with the remaining parameters (μ, σ, and ξ) the same as equation (5).

Like the GEV, the shape parameter defines the curve shape of the function $H(y)$. For ξ > 1 (Type II) the distribution of excesses has no upper limit; for ξ = 0 (Type I) the distribution is linear and unbounded; and for ξ < 1 (Type III) the distribution has an upper bound of $\mu - \sigma_u/\xi$ (Coles 2001).

After parameter estimation, the level x_m that is exceeded on average once every m observations is determined by

$$x_m = u + \frac{\sigma_u}{\xi}\left[(m\zeta_u)^{-\xi} - 1\right] \qquad \text{for } \xi \neq 0\text{, and} \tag{8a}$$

$$x_m = u + \sigma \log(m\zeta_u) \qquad \text{for } \xi = 0 \tag{8b}$$

with x_m the m observation return level with return period $1/m$, and ζ_u the probability of an individual observation exceeding the threshold (u).

For N-year return levels, let n_y be the number of observations per year so that $m = Nn_y$. Therefore, equations 8a and 8b become

$$x_N = u + \frac{\sigma_u}{\xi}\left[\left(Nn_y\zeta_u\right)^{-\xi} - 1\right] \qquad \text{for } \xi \neq 0\text{, and} \tag{9a}$$

$$x_N = u + \sigma \log\left(Nn_y\zeta_u\right) \qquad \text{for } \xi = 0 \tag{9b}$$

Coles (2001) defines a natural estimator of the exceedance probability ζ_u as

$$\zeta_u = \frac{k}{n} \tag{10}$$

with k the number of exceedances, and n the total number of measurements.

POT/GPD Threshold Selection and Declustering

An initial threshold (u) estimate of 13.9 mps was determined as the minimum annual maximum value from the Final Max Day 2-min Wind Speed data (Table 1). This threshold value represents the approximate 98.5th percentile of the 7,669 daily wind speed observations. Review of the mean residual life plot and threshold plots (not shown) indicate a threshold value of 13.9 mps appears reasonable (refer to Coles (2001) for more details on mean residual and threshold plots).

As previously mentioned, a requirement for EVA is independent and identically distributed data, which requires that only one wind speed from each storm system be used. Since windstorms, and other environmental variables, tend to form clusters of wind speeds, wind data can have strong serial correlation (Palutikof et al. 1999). To satisfy the data independence requirement it becomes necessary to decluster the data, that is extract the maximum wind speed (above the threshold (u)) from the clustered data and provide a minimum time length that observations need to be below the threshold before the next event can be considered independent from the previous. For this analysis a declustering time of 4 days is used, which has been used in other extreme wind speed analysis (Lombardo et al. 2009, Pinter et al. 2015).

Using a threshold (u) value of 13.9 mps and a declustering time of 4 days results in 87 exceedances over 21 years for the Final Max Day 2-min Wind Speed data (Table 1). This provides an average number of exceedances per year of 4.14 (87/21); and an exceedance probability ζ_u of 0.01134 (87/(21x365.25)).

Figure 4 shows the declustered wind speed exceedances above the threshold for the Final Max Day 2-min Wind Speed data (Table 1). The data also appear stationary over time due to the non-significant slope in the trend (p > 0.1).

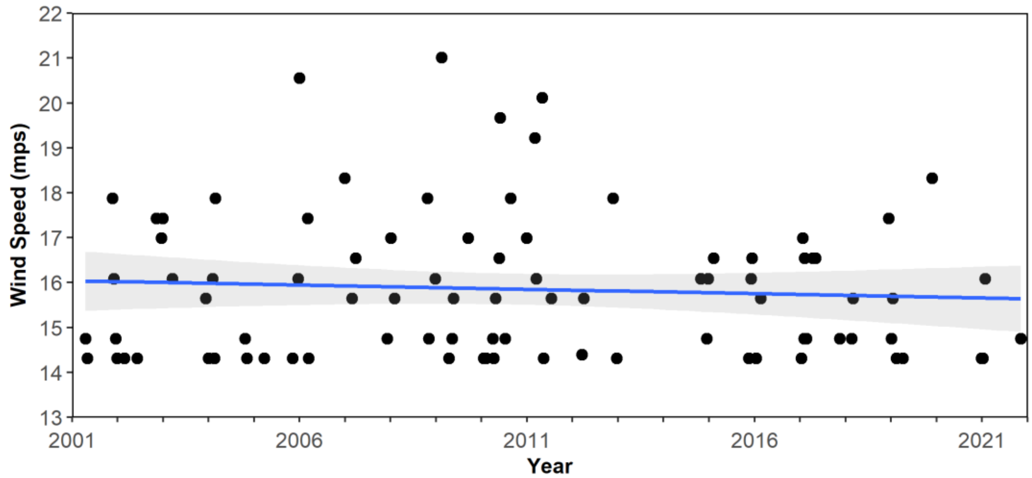

Figure 4 Declustered 2-min maximum wind speed exceedances above the threshold for the Final Max Day 2-min Wind Speed data set and the linear trend slope and 95% confidence intervals.

Results and Discussion

Results of the conducted extreme value analysis for the Arcata/Eureka Airport wind speed data are provided in the following sections.

Peak Wind Speed by Wind Direction

Given the relatively small size of the Final Max Day 2-min Wind Speed data sets, it was not possible to conduct an EVA of wind speed by wind direction, due to the lack of daily maximum extreme wind speeds from some wind directions (e.g. wind from the east). However, to provide some insight into peak wind speeds by direction, a Gumbel (Type I) distribution was fit to the maximum daily wind speeds by direction (22.5° increments) for the larger Temp Max Day Wind Speed data set (Figure 5). Peak wind speeds for three return levels (2-yr, 10-yr and 100-yr) were estimated. The use of the Gumbel provides a consistent distribution for each direction with curve tails not affected by the shape parameter.

Results demonstrate that the fastest peak wind speeds tend to be from the southeast (112.5°) to north (360°) directions, with peak winds from easterly directions being much lower. The maximum peak winds appear to come from two dominant and opposing directions, southeast (135°) and northwest (315°). Southeast winds are associated with winter storms from the south, and spring and summer high pressure winds from the north.

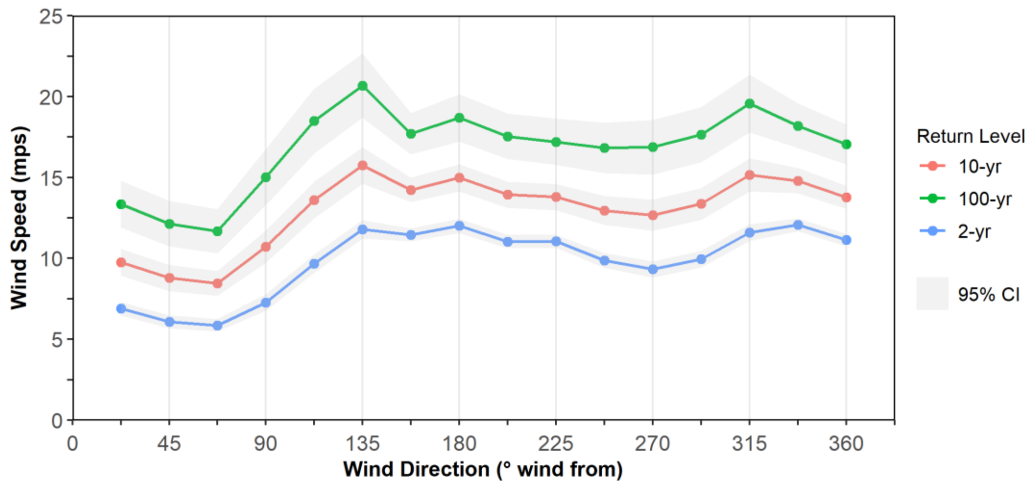

Figure 5 Peak wind speed estimates and 95% confidence intervals by wind direction from a Gumbel distribution using the Temp Max Day Wind Speed data set for the 2-yr, 10-yr and 100-yr return levels.

GEV and POT/GPD Analysis

The GEV and POT/GPD analysis were based on the Final Max Day 2-min Wind Speed data. The GEV distribution was fit to the annual block maxima, and the GPD fitted to the daily maximums using a threshold (u) value of 13.9 mps and a declustering time of 4 days as previously described. The estimated parameters for both distributions are listed in Table 2, and summary diagnostic plots (from R package extRemes) are provided in Figure 6 and Figure 7 for the GEV and GPD distributions, respectively.

Table 2 Summary of the GEV and POT/GPD distribution fits to the Arcata/Eureka Airport Final Max Day 2-min Wind Speed data set.

Distribution Fitted Variables		Units	Extreme Value Distribution Types	
			GEV	**POT/GPD**
Number of Samples		#	21	87
Parameters (Standard Error)	Location (μ)	mps	16.692 (0.430)	--
	Scale (σ)	mps	1.753 (0.304)	2.528 (0.353)
	Shape (ξ)	--	-0.236 (0.167)	-0.267 (0.094)
	Threshold (u)	mps	--	13.86
[95% Confidence Intervals]	Location (μ)	mps	[15.849, 17.534]	--
	Scale (σ)	mps	[1.156, 2.349]	[1.836, 3.219]
	Shape (ξ)	--	[-0.564, 0.091]	[-0.451, -0.083]
Return Level Estimates [95% Confidence Interval]	2-yr	mps	17.31 [16.43, 18.18]	17.94 [16.82, 19.07]
	10-yr	mps	19.75 [18.67, 20.83]	19.82 [18.11, 21.53]
	100-yr	mps	21.61 [19.28, 23.94]	21.43 [18.25, 24.60]

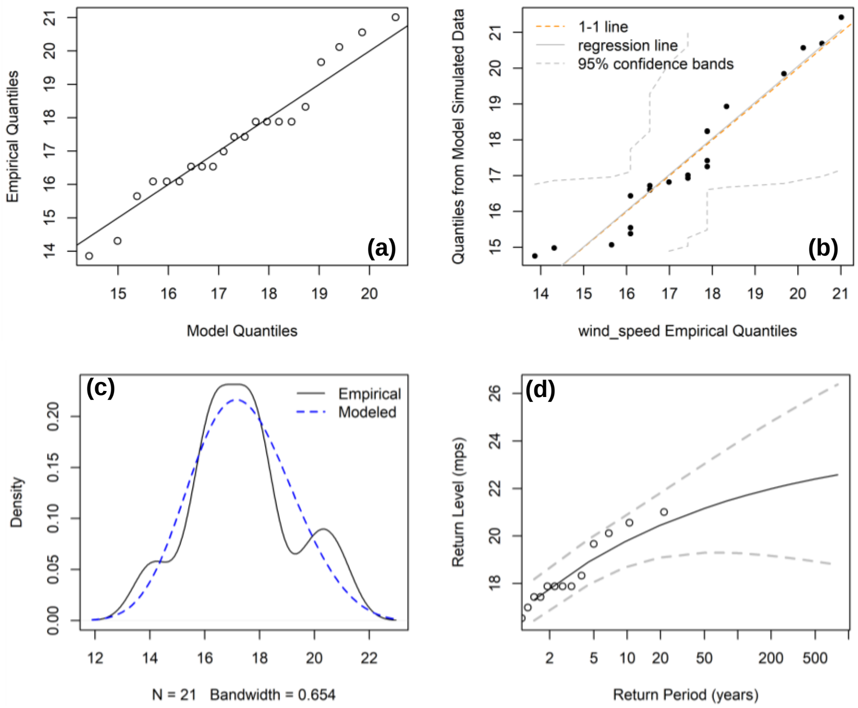

Figure 6 Diagnostic plots from R package extRemes (Gilleland and Katz 2016) for the annual maximum GEV distribution for the Final Max Day 2-min Wind Speed data set. Quantile-quantile plot (a); quantiles from distribution simulated data and empirical quantiles (b); density plots of empirical data and fitted distribution (c); and return level plot (d).

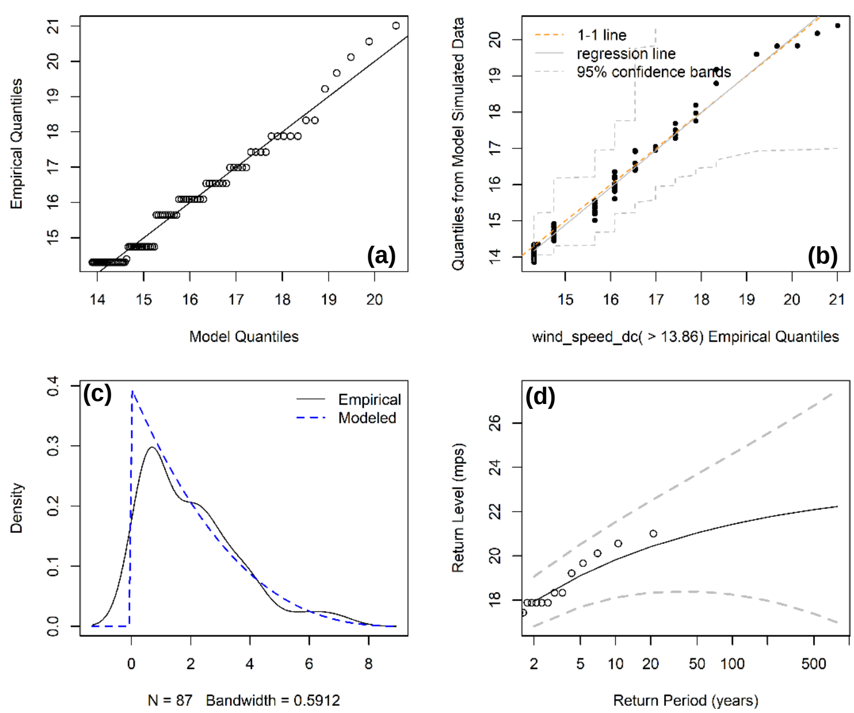

Figure 7 Diagnostic plots from R package extRemes (Gilleland and Katz 2016) for the daily maximum GPD distribution for the Final Max Day 2-min Wind Speed data set. Quantile-quantile plot (a); quantiles from distribution simulated data and empirical quantiles (b); density plots of empirical data and fitted distribution (c); and return level plot (d).

The diagnostic plots (Figure 6 and Figure 7) for both the GEV and GPD fitted models look reasonable. All quantile plots (top row plots) appear linear, and all samples from the fitted distributions fall within the 95% confidence bands (top right plots). Density function plots for both distributions are consistent with the empirical data (bottom left plots), although the GPD distribution appears to best fit the empirical data. Due to the negative shape factors (ξ) both return level curves are convex with an asymptotic upper limit (bottom right plots). Both curves appear to represent the empirical data, and all data are within the 95% confidence bands.

Return level estimates (Table 2) for both the GEV and GPD distributions are very similar, providing confidence in both fitted distributions. The 95% confidence interval for the GEV shape factor does contain zero indicating the possibility of shape factors being positive (Frechet), negative (Weibull), or zero (Gumbel). However, the 95% shape factor interval contains almost all negative values providing evidence that the fitted negative shape factor for the GEV is reasonable. The 95% confidence interval for the GPD fitted shape factor contains all negative values providing strong evidence to support the negative GPD shape factor.

Based on the longer sample record of the POT/GPD (87 samples) compared to the shorter GEV record (21 samples), and the greater confidence in the negative GPD shape factor, the POT/GPD analysis was considered the best approach for determining extreme wind speeds for the Arcata/Eureka Airport data. The final extreme wind speed estimates are only provided for the POT/GPD analysis.

Final Extreme Wind Speed Estimates

The final extreme 2-min wind speed estimates from the POT/GPD analysis for the Arcata/Eureka Airport data are provided in Table 3. The final wind speeds are provided in units of meters per second (mps) and miles per hour (mph).

Table 3 Final extreme 2-min wind speed estimates from the POT/GPD analysis of the Arcata/Eureka Airport wind speed data. All wind speeds have been adjusted to 2-min average duration and 10 m height.

Return Level (yr)	Annual Exceedance (%)	Extreme 2-min Wind Speed (mps)		Extreme 2-min Wind Speed (mph)	
		Estimate	95% CI	Estimate	95% CI
~1	~100	16.85	[15.90, 17.79]	37.7	[35.6, 39.8]
1.053	95	16.94	[15.98, 17.90]	37.9	[35.7, 40.0]
1.25	80	17.22	[16.22, 18.23]	38.5	[36.3, 40.8]
1.5	66.67	17.51	[16.46, 18.57]	39.2	[36.8, 41.5]
2	50	17.94	[16.82, 19.07]	40.1	[37.6, 42.7]
5	20	19.11	[17.69, 20.53]	42.7	[39.6, 45.9]
10	10	19.82	[18.11, 21.53]	44.3	[40.5, 48.2]
25	4	20.58	[18.38, 22.78]	46.0	[41.1, 51.0]
50	2	21.04	[18.39, 23.70]	47.1	[41.1, 53.0]
100	1	21.43	[18.25, 24.60]	47.9	[40.8, 55.0]
200	0.5	21.75	[17.97, 25.52]	48.6	[40.2, 57.1]
500	0.2	22.09	[17.38, 26.79]	49.4	[38.9, 59.9]

References

American Society of Civil Engineers (ASCE). 2017. Minimum Design Loads and Associated Criteria for Buildings and Other Structures. American Society of Civil Engineers (ASCE Standard - ASCE/SEI 7-16).

Chamberlain, S. 2021. rnoaa: 'NOAA' Weather Data from R. R package version 1.3.4. https://CRAN.R-project.org/package=rnoaa.

Coles, S. 2001. An introduction to statistical modeling of extreme values. Springer-Verlag, London.

Coastal Engineering Manual (CEM). 2015. Chapter 2, Meteorology and Wave Climate. U.S. Army Corps of Engineers. Washington D.C. EM 1110-2-1100 (Part II).

Gilleland, E. and R.W. Katz. 2016. extRemes 2.0: An Extreme Value Analysis Package in R. Journal of Statistical Software, 72(8), 1-39. doi:10.18637/jss.v072.i08.

Lombardo, F.T, Main, J.A. and E. Simiu. 2009. Automated extraction and classification of thunderstorm and non-thunderstorm wind data for extreme-value analysis. Journal of wind Engineering and Industrial Aerodynamics, 97:120–131.

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

Pintar, A.L., Simiu, E., Lombardo, F.T. and M. Levitan. 2015. Maps of Non-Hurricane Non-tornadic Wind Speeds With Specified Mean Recurrence Intervals for the Contiguous United States Using a Two-Dimensional Poisson Process Extreme Value Model and Local Regression. National Institute of Standards and Technology (NIST) Special Publication 500-301.

Powell, M.D. 1993. Wind measurement and archival under the Automated Surface Observing System (ASOS): User concerns and opportunity for improvement. Bull. Amer. Meteorol. Soc., vol. 74, no. 4, pp. 615-624.

R Core Team. 2021. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Reiss, R.D. and M. Thomas. 2007. Statistical Analysis of Extreme Values: From Insurance, Finance, Hydrology, and Other Fields, Third Edition. Birkhäuser, Verlag.

U.S. Integrated Ocean Observing System (IOOS), 2014. Manual for Real-Time Quality Control of Wind Data: A Guide to Quality Control and Quality Assurance of Coastal and Oceanic Wind Observations. 45pp.