Structures Congress XII, Vol. 2, 1994

Extreme Winds Estimation by 'Peaks Over Threshold' and Epochal Methods

J. L. Gross[1], N. A. Heckert,[2] J. A. Lechner[2] and E. Simiu[1]

ABSTRACT

With a view to applying the "peaks over threshold" method to the estimation of extreme wind speed data, we perform Monte Carlo simulations for which the parameters of the population distributions were estimated from sets of actual extreme wind speed data. We summarize results concerning (1) the relative efficiency of several estimation procedures used in such methods, (2) the optimal threshold for any given set of data, and (3) estimates based on the "peaks over threshold" method as compared to estimates based on the epochal approach. This work is part of a long-term effort conducted by NIST aimed at assessing and utilizing "peaks over threshold" methods for the estimation of extreme wind loads.

INTRODUCTION

Classical extreme value theory is based on the analysis of data consisting of the largest value in each of a number of basic comparable sets called epochs. In wind engineering, it has been customary to define epochs by calendar years. For

---

[1]Building and Fire Research Laboratory, National Institute of Standards and Technology, Gaithersburg, MD 20899.

[2]Computing and Applied Mathematics Laboratory, National Institute of Standards and Technology, Gaithersburg, MD 20899.

independent, identically-distributed variates, the distribution of the largest of a set of values converges to one of three types of asymptotic extreme value distributions, known, in order of decreasing tail length, as the Frechet (or Fisher-Tippett Type II), Gumbel (Type I), and Weibull (Type III) distributions.

In contrast to classical extreme value theory, recent developments in statistical estimation procedures applicable to extreme data allow the analysis of all data exceeding a specified threshold, regardless of whether they are the largest in the respective sets or not.

An asymptotic distribution -- the Generalized Pareto Distribution (GPD) -- has been developed using the fact that exceedances of a sufficiently high threshold are rare events to which the Poisson distribution applies. The expression for the GPD contains a tail length parameter, c, such that $c>0$, $c=0$ and $c<0$ correspond respectively to Frechet, Gumbel, and Weibull (right tail-limited) domains of attraction. The Generalized Pareto Distribution is given by:

$$G(y) = Prob[Y<y] = 1 - \left[1 + \left(\frac{cy}{a}\right)\right]^{-1/c} \qquad a>0,\ 1+\left(\frac{cy}{a}\right)>0 \tag{1}$$

For wind engineering purposes, estimates of the wind speeds corresponding to various mean recurrence intervals are of interest. The value of the random variable which corresponds to a given mean recurrence interval, R, (in years) is found by setting

$$p = 1 - \frac{1}{\lambda R} \tag{2}$$

where λ is the mean crossing rate of the threshold, u, per year, since one expects λR data values in R years. From Eqs. 1 and 2

$$y = \frac{-a[1-(\lambda R)^c]}{c} \tag{3}$$

The maximum wind speed corresponding to a mean recurrence interval, R, is

$$x_R = y + u \tag{4}$$

where u is the threshold used in the estimation of c and a.

Our objective was to obtain, by Monte Carlo simulations, results allowing comparisons to be made (1) between the performance of the various estimation procedures, (2) between the quality of the estimates by the 'peaks over threshold' approach for various mean exceedance rates (or, equivalently, for various threshold levels), and (3) between the optimal (or almost optimal) estimates obtained by the 'peaks over threshold' approach, on the one hand, and the epochal approach on the other.

MONTE CARLO SIMULATIONS

Maximum daily wind speed data from several sites were studied to determine the mean and standard deviation of the data exceeding a specified threshold. Next, samples with this mean and standard deviation were generated using the Monte Carlo Method for Gumbel, Weibull and Normal distributions. The Conditional Mean Exceedance (CME), Pickands and Dekkers-Einmahl-de Haan methods (see Lechner et al., 1993) were then applied to subsets of the populations representing various thresholds (or mean crossing rates) resulting in estimates of the GPD parameters, c and a. Based on these parameters, estimates of extreme winds for several recurrence intervals were computed.

Distribution Parameters

Daily maximum wind speed data from two sites, Toledo, Ohio and Boise, Idaho were selected for this study. It was found that in each set consisting of roughly 25 years of data, there were approximately 750 data exceeding a threshold of 8 m/s (18 mph). Thus, the mean exceedance rate was approximately 30/yr. The mean and standard deviation of the wind speeds exceeding 8 m/s (18 mph) was found to be 12.96 m/s (29 mph) and 2.91 m/s (6.5 mph), respectively. The estimate of the tail length parameter, c, (using the CME method) was approximately -0.275.

Population Percentage Point Functions

Percentage point functions for the Gumbel and Weibull distributions are found from the expressions for the respective cumulative distribution functions, $F_G(x)$ and $F_W(x)$ which are set equal to $1-1/(\lambda R)$.

Simulations

Monte Carlo simulations were run to generate 500 samples from (1) Gumbel distribution (c = 0), (2) Weibull distribution with tail length parameter $\gamma = -1/c = -1/(-0.275) = 3.64$, and (3) Normal distribution, each with mean, $E(x)$ = 12.96 m/s (29 mph) and standard deviation, $s(x)$ = 2.91 m/s (6.5 mph). Samples of size 750 (25 years of data) and 1200 (40 years of data) were generated. For various crossing rates (λ = 20, 15, 10, 5, 2 and 1), the CME, Pickands and de Haan estimation methods were applied to obtain estimates of the GPD parameters c and a and, based on these parameters, estimates of extreme wind speeds for mean recurrence intervals of 50, 500 and 5000 years were computed.

RESULTS

Comparison of Estimation Methods

First we compare the relative efficiency (based on the root-mean-square-error) of estimation methods considering the CME, Pickands and de Haan methods. From results presented by Gross et al. (1993) it follows that the CME and de Haan methods tend to be competitive. Further, both methods are superior to the Pickands method. Results for the Gumbel distribution (25 years of data) are given in Table 1 for crossing rates that result in sets ranging from 25 to 500 data. Results for sets of 40 years were comparable. Results from this study and a previous study (Lechner et al., 1993) suggest that the de Haan method gives better estimates than the CME method for extremes with a large mean recurrence interval.

Optimal Crossing Rate

A high threshold reduces the bias since it conforms best with the asymptotic assumption on which the GPD distribution is based; however, because it results in a small number of data, it increases the sampling error. Estimates presented in Gross et al. (1993) suggest that, with no significant error, an approximately optimal threshold may be assumed to correspond to a mean exceedance rate of 5/yr to 15/yr.

Comparison of Threshold and Epochal Procedures

Next we compare the epochal procedure, traditionally used in the estimation of extreme wind speeds, with the threshold procedure. Epochal extreme wind data were generated by taking the largest variate in each successive set of 30 (maximum crossing rate) data from the parent populations. This resulted in 500 sets of 25 and 40 yearly maxima. The CME, Pickands, de Haan and Probability Plot Correlation Coefficient (PPCC) estimation methods were applied to the data. Results for the Gumbel distribution are given in Table 2 for sets of 25 years. Results for sets of 40 years were similar.

First we note that, for the epochal procedure, the CME and the de Haan methods are competitive with the de Haan method performing better for extremes with large mean recurrence intervals.

Results for the epochal procedure using the optimal estimation method were compared with results for the threshold procedure using a crossing rate of 10/yr. It was found that for the Gumbel and Weibull distributions, the threshold procedure produced better estimates (based on the root-mean-square-error). However, for the Normal distribution, the epochal method produced better extreme wind speed estimates.

CONCLUSIONS

The following conclusions may be drawn from this study. The reader should note that the results may depend on the choice of population distributions and are therefore tentative.

- For both the 'peaks over threshold' and the epochal procedures, the CME and de Haan methods are competitive with the de Haan method performing better than the CME method for extremes with large mean recurrence intervals. Both methods are significantly better than the Pickands method.

- Thresholds corresponding to mean exceedance rates of between 5/yr and 15/yr appear to be optimal.

- For a mean exceedance of 10/yr (near optimal), the threshold procedure produces slightly better estimates than does the epochal procedure for both the Gumbel and Weibull distributions. However, for the normal distribution, the epochal procedure produces better estimates than does the threshold procedure.

ACKNOWLEDGMENT

The authors acknowledge with thanks partial support by the National Science Foundation (Grant MSM-9013116 to Department of Civil Engineering, The Johns Hopkins University) and the helpful interaction on this project with R.B. Corotis of the Johns Hopkins University.

REFERENCES

Gross, J. L., Heckert, N. A., Lechner, J. A. and Simiu, E., (1993) "Modeling of Extreme Loading by 'Peaks Over Threshold' Methods," *Dynamic Response and Progressive Failure of Special Structures*, ASCE, in press.

Lechner, J.A., Simiu, E. and Heckert, N.A. (1993), "Assessment of 'peaks over threshold' methods for estimating extreme value distribution tails," *Structural Safety*, Elsevier Science Publishers, in press.

Data	Method	R = 50 yrs			R = 500 yrs			R = 5000 yrs		
		M (m/s)	SD (m/s)	RMSE (m/s)	M (m/s)	SD (m/s)	RMSE (m/s)	M (m/s)	SD (m/s)	RMSE (m/s)
500	CME	27.34	2.06	2.24	31.31	3.94	4.48	34.92	6.49	7.50
	DEH	24.46	1.68	4.12	26.44	2.47	7.42	27.84	3.21	11.28
	PIC	23.49	1.58	4.99	24.89	2.30	8.86	25.78	2.93	13.20
375	CME	27.71	2.21	2.28	32.20	4.54	4.71	36.58	8.08	8.35
	DEH	26.29	1.99	2.77	29.50	3.27	5.12	32.21	4.69	7.98
	PIC	25.21	2.32	3.80	27.65	3.79	6.92	29.58	5.40	10.56
250	CME	27.92	2.33	2.35	32.88	5.19	5.22	38.04	10.09	10.11
	DEH	27.38	2.30	2.45	31.61	4.11	4.50	35.59	6.40	7.09
	PIC	26.98	3.31	3.53	31.06	6.55	6.97	35.13	11.37	11.90
125	CME	28.00	2.45	2.47	33.40	6.22	6.22	39.68	14.00	14.04
	DEH	27.96	2.71	2.72	33.00	5.41	5.43	38.26	9.36	9.37
	PIC	28.31	4.68	4.68	34.67	12.39	12.45	43.41	29.91	30.28
50	CME	27.91	2.49	2.51	33.74	7.35	7.35	41.93	19.85	20.12
	DEH	27.92	2.82	2.83	33.17	6.18	6.19	39.15	11.73	11.74
	PIC	28.71	5.78	5.80	38.36	26.40	26.86	50.65	47.68	49.17
25	CME	27.88	2.61	2.63	33.82	8.31	8.31	43.63	26.09	26.56
	DEH	28.08	3.04	3.04	33.69	7.41	7.41	40.84	15.71	15.86
	PIC	29.60	6.69	6.83	46.55	47.02	48.82	61.20	65.07	68.86
Pop.		28.22			33.44			38.66		

Table 1 - Threshold Procedure / Gumbel Distribution / 25 Years of Data

Data	Method	R = 50			R = 500			R = 5000		
		M (m/s)	SD (m/s)	RMSE (m/s)	M (m/s)	SD (m/s)	RMSE (m/s)	M (m/s)	SD (m/s)	RMSE (m/s)
25	CME	27.54	2.35	2.45	31.38	6.28	6.61	35.79	16.02	16.27
	DEH	24.25	2.52	4.70	24.97	3.62	9.21	25.37	4.61	14.06
	PIC	25.44	2.48	3.73	27.13	5.51	8.38	28.58	13.87	17.15
	PPCC	29.12	3.79	3.89	39.99	29.54	30.25	77.74	285.83	288.48
Pop.		28.22			33.44			38.66		

Table 2 - Epochal Procedure / Gumbel Distribution / 25 Years of Data