NISTIR 5531

Extreme Wind Estimates by the Conditional Mean Exceedance Procedure

J.L. Gross, N.A. Heckert, J.A. Lechner and E. Simiu

April 1995
Building and Fire Research Laboratory
National Institute of Standards and Technology
Gaithersburg, MD 20899

U.S. Department of Commerce
Ronald H. Brown, *Secretary*
Technology Administration
Mary L. Good, *Under Secretary for Technology*
National Institute of Standards and Technology
Arati Prabhakar, *Director*

ABSTRACT

We describe work aimed at improving procedures for the estimation of non-tornadic extreme wind speeds, regardless of their direction, in regions not subjected to hurricanes. Using the Generalized Pareto Distribution (GPD) approach and the Conditional Mean Exceedance (CME) estimation method, we analyze 115 17-year to 52-year sets of largest annual speeds and sets drawn from 48 15-year to 26-year records of maximum daily wind speeds. Based on this analysis we attempt an assessment of the widely held belief that the Gumbel distribution with site-dependent location and scale parameters is a universal model of extreme wind speeds. Some of our results suggest that the reverse Weibull distribution is a more appropriate model. This would result in more reasonable estimates of wind-induced failure probabilities and wind load factors than the corresponding estimates based on the Gumbel distribution. However, our assessment is so far only tentative owing to uncertainties inherent in our results. Future work based on lower thresholds (larger data samples) and alternative estimation methods is planned.

1. INTRODUCTION

Until recently methods for the estimation of extreme wind speeds were based solely on classical extreme value theory (Gumbel, 1958). Although such methods can be used to obtain credible estimates of wind speeds with relatively short mean return periods (50 years, say), questions remain as to their capability to estimate distribution tails reliably.

In the last two decades a novel theory known as the "peaks over threshold" approach was developed that offers the potential for more realistic estimates of the tails. This would allow the estimation by statistical methods of wind load factors, which have to date been specified in building standards on the basis of engineering guesses passed from one generation of standards to the next. This paper is part of a long-term project aimed at improving estimates of wind speed distribution tails and wind load factors. The "peaks over threshold" approach rests on the application of the Generalized Pareto Distribution (GPD) to the excess of the extreme variates over a fixed threshold. For terminology and notations, see Gross et al. (1994).

Unlike classical methods, the "peaks over threshold" approach is applicable to the analysis of the set consisting of all data exceeding a sufficiently high threshold. In addition, it is applicable to data taken from sets of epochal extremes (i.e., maxima over samples of fixed size, such as largest annual wind speeds). According to classical theory, in the asymptotic limit a set of epochal extremes must fit the tail of one of the three extreme value distributions. The epochal extremes that exceed a sufficiently high threshold must therefore fit the GPD with $c>0$, $c \to 0$, or $c<0$.

We review briefly in Section 2 the expression for the GPD, the GPD-based estimator used in this work, and the estimation within the framework of the "peaks over threshold" approach of variates with specified mean return periods. In Section 3 we analyze 115 sets of observed largest annual wind speeds taken from 17- to 52-year records. Section 4 is devoted to analyses of sets taken from 48 15- to 26-year records of largest daily speeds. Section 5 presents our conclusions and outlines future work.

2. GENERALIZED PARETO DISTRIBUTION, AND DESCRIPTION OF ESTIMATORS

We review here the expression for the Generalized Pareto Distribution and the Conditional Mean Exceedance (CME) method for estimating distribution parameters.

Generalized Pareto Distribution (GPD)

The expression for the GPD is

$$G(y) = \mathrm{Prob}[Y \leq y] = 1-\{[1+(cy/a)]^{-1/c}\} \quad a>0,\ (1+(cy/a))>0 \qquad (1)$$

Equation 1 can be used to represent the conditional cumulative distribution of the excess $Y=V-u$ of the variate V over the threshold u, given $V>u$ for u sufficiently large; c and a are distribution parameters. The cases $c>0$, $c=0$ and $c<0$ correspond, respectively, to Fréchet, Gumbel, and reverse Weibull (right tail-limited) domains of attraction. For $c=0$ the expression between braces is understood in a limiting sense as the exponential $\exp(-y/a)$ (Castillo, 1988, p. 215). For $c<0$ the shape parameter of the corresponding distribution is $\gamma=-1/c$ (Smith, 1989).

Cumulative Mean Exceedance (CME) Method

The CME is the expectation of the amount by which a value exceeds a threshold u, conditional on that threshold being attained. If the exceedance data are fitted by the GPD model and $c < 1$, $u > 0$, and $a + uc > 0$, then the CME vs. u plot should follow a line with intercept $a/(1-c)$ and slope $c/(1-c)$ (Davison et al., 1990). The linearity of the plot is an indicator of the appropriateness of the GPD model. Estimates of c and a can be obtained from the slope and intercept of a straight line fit to the CME vs. u plot.

Estimation of variates with specified mean return periods

The mean return period R, in years, of a given wind speed is defined as the inverse of the probability that that wind speed will be exceeded in any one year. In this section we give expressions that allow the estimation from the GPD of the value of the variate corresponding to probability $1 - 1/(\lambda R)$, where λ is the mean crossing rate of the threshold u per year (i.e., the average number of data points above the threshold u per year), and R is the mean recurrence interval in years. We have

$$\mathrm{Prob}(Y < y_R) = 1 - 1/(\lambda R) \tag{2}$$

$$1 - [1 + c y_R/a]^{-1/c} = 1 - 1/(\lambda R) \tag{3}$$

$$y_R = -a[1 - (\lambda R)^c]/c \tag{4}$$

$$V_R = y_R + u \tag{5}$$

where V_R is the R-year wind speed (e.g., V_{50}=50-year speed) and u is the threshold used to estimate c and a. For epochal sets consisting of the largest annual wind speeds, $\lambda = 1$. Note that, given u, λ, c, R and V_R, Eqs. 4 and 5 yield the parameter a inherent in the estimation of V_R.

3. ANALYSIS OF LARGEST ANNUAL WIND SPEEDS

Table A1 given in Appendix A shows estimated values of the tail length parameter, $\hat{c}$. The estimates were obtained by applying the CME method to data samples taken from 115 N-year records of observed largest annual wind speeds adjusted to a 10 m elevation above ground ($17 < N < 52$). Stations where strong winds are predominantly due to hurricanes were not included in Table A1. All wind speeds are given by the Weather Service in terms of fastest miles. For this report, wind speeds have been converted to SI units (1 mph=0.44704 m/s). In order to include only the strongest winds in each set -- the winds most likely to approach the asymptotic condition inherent in the GPD approach -- we used the CME estimator based for each record on a relatively high threshold. We chose this threshold to be equal to the record's median wind speed, V_{med}. All the CME-based results of Table A1 are based on this threshold. In our calculations for observed data a wind speed, V, was defined as exceeding the threshold if $V \geq V_{med}$, that is, the actual threshold is actually smaller (by an infinitesimal amount) than the nominal threshold.

For the threshold V_{med}, the sample average number of exceedances was $E(n_{ex})=21$, and $E(V_{med})=50$, $SD(V_{med})=6.5$ (E and SD denote sample mean and standard deviation). The mean and standard deviation of the estimated values of c listed in Table A1 are:

$E(\hat{c})=-0.26$, $SD(\hat{c})=0.38$.

We denote by V_{med-} and V_{med+} the speed preceding V_{med} and the speed following V_{med}, respectively, in the set of ordered speeds of which V_{med} is the median. Using a threshold V_{med-}, $E(n_{ex})=24$ and a threshold V_{med+}, $E(n_{ex})=17$, results not listed in Table A1 yielded

$E(\hat{c})=-0.24$, $SD(\hat{c})=0.34$

$E(\hat{c})=-0.27$, $SD(\hat{c})=0.48$,

respectively. For lower thresholds $E(\hat{c})$ was found to increase.

The following results were obtained from Monte Carlo simulations. For 500 25-year samples with mean exceedance rate $\lambda=1$ and $\hat{c}$ estimated by the CME method,

$E(\hat{c})=-0.09$, $SD(\hat{c})=0.27$ (population with Gumbel distribution)

$E(\hat{c})=-0.33$, $SD(\hat{c})=0.24$ (population with reverse Weibull distr., $\gamma=-1/c=1/0.275$).

A comparison between the results based on the observed data on the one hand and on the simulated data on the other would suggest that a reverse Weibull distribution with shape parameter $\gamma\approx5$ ($c=-0.2$) is a more appropriate model than the Gumbel distribution. (The Gumbel distribution can be interpreted as the limit of a family of three-parameter extreme value distributions as the shape parameter approaches infinity -- see proof in Simiu et al., 1986)

Let us now hypothesize, nevertheless, that the Gumbel distribution is an appropriate universal model of extreme wind speeds, that is, that for every station the true tail length parameter is $c=0$. The results of the Monte Carlo simulations just shown indicate a bias of about -0.1 in the estimation of c, so let us allow for a bias as large as -0.1 in estimating c. Using a binomial distribution model (with mean $n/2=57.5$ and standard deviation $(n)^{1/2}/2=5.36$), one would expect that about half of the 115 estimated values of c would be below -0.1. Actually, 77 estimated values (significantly more than half) are below -0.1; this number is almost four standard deviations higher than the mean, and would lead to a rejection of the hypothesis that the Gumbel distribution is a universal model for the extreme speeds. However, this tentative conclusion may not be warranted. Indeed, each station may have a different true c, and the sample sizes for the various stations differ. Instead of the average $E(\hat{c})$ for the observed data, it would therefore be appropriate to consider a weighted average of c, where each weight is equal to the inverse of the variance of the estimate of c. Standard deviations of these estimates are listed in Table A1 and were obtained by the expression

$$SD(\hat{c}) = \frac{[\Sigma(n-i)]^{1/2}[\Sigma(n-i)(y_i-\text{intercept}-x_i\cdot(\text{slope}))^2]^{1/2}}{(n-3)^{1/2}(1+\text{slope})^2\cdot\{[\Sigma(n-i)][\Sigma(n-i)x_i^2]-[\Sigma(n-i)x_i]^2\}^{1/2}} \tag{6}$$

where n is the number of data in the set, x_i are the speeds ($i=1,2,..,n-1$), and y_i are CME values (see Appendix B). The weighted mean of the c estimates, based on these standard deviations, is close to -0.1, and its standard deviation is about 0.32. Note that the simplifying assumption implicit in Eq. 6 that the errors in the estimation of y_i for various i's are independent is not correct, and the standard deviations of the c estimates are actually larger than those given by Eq. 6 by factors that pilot Monte Carlo simulations suggested can be as high as two or even more. We conclude that owing to the small sample sizes we used, we do not get a sufficiently good estimate of the weighted average of $\hat{c}$, and the inference made earlier on the basis of the binomial distribution cannot be relied upon with confidence.

A comparison between the tabulated values of CME-based estimated V_N's and values of the maximum speeds on record, V_{max}, shows that the performance of the CME estimator of V_N is very good. We note, however, that a worse set of V_N estimates was obtained, where the CME method was applied to data samples in which identical speeds were made distinct by addition of multiples of 0.001. Though the estimates of c were not much affected by this change, this sensitivity of the CME method appears to cast some shadow upon its dependability.

The CME estimates of V_{100000} appear to be worse than those of V_N: in some cases they differ minimally from the estimates of V_N; in others they can be ridiculously large. We also show in Table A1 wind speed estimates based on the Gumbel model. These were obtained by the probability plot correlation coefficient method (PPCC). It is seen that estimates of V_N based on the Gumbel model are comparable to those based on the CME method.

Table A1 also lists CME-based estimated speeds with mean return period 100000 years, $\hat{V}_{100000}$, where the parameter a is based on the CME-based estimated value of V_N, as indicated in the remark following Eqs. 4 and 5, and on a specified c=-0.2.

Load Factors

Let R_u denote the mean return period of the ultimate load. If the wind load predominates (i.e., no load combination need be considered), the wind load factor is

$$\phi = (V_{Ru}/V_{50})^2 \tag{7}$$

Table A1 lists estimated values of ϕ based on Eq. 7, where V_{50} was based on the CME estimates of V_N, and V_{Ru}, corresponding to asymptotically large R_u, was based on a parameter a estimated from V_N by using Eqs. 4 and 5, and the specified parameter c=-0.2 Depending upon the site, the estimates of ϕ vary between 1.24 and 1.68. Their average is $\phi=1.42$, as compared to $\phi=1.3$ specified in the ASCE Standard 7-93 and earlier versions thereof.

Structural Reliability Implications

Consider, for example, the Fresno, CA data set. Under the assumption that the Gumbel distribution best fits the extremes, for $R_u=10^3$ years, 10^5 and 10^6 years, the estimated wind speeds are 26, 34 and 38 m/s (59, 77 and 86 mph), respectively (Simiu et al., 1979). Under the assumption that the reverse Weibull with $\gamma=-1/c=1/0.20$ holds, they are 24, 26 and 27 m/s (54, 59, and 60 mph), respectively. It is seen that the tail is considerably shorter for the reverse Weibull than for the Gumbel.

Failure probabilities for wind-sensitive structures designed in accordance with U.S. building code requirements (or safety indices reflecting those probabilities) have been estimated on the basis of the Gumbel model. Ellingwood et al. (1980) found such estimates to be substantially higher than for other types of structures. Experience shows that the number of structural failures caused by non-tornadic and non-hurricane winds is vastly smaller than those estimates would indicate. One possible flaw of those probability estimates is in our opinion the fact that they are based on the Gumbel distribution which, as suggested by our results, overestimates extreme winds corresponding to long mean return periods.

The result that the upper tail of the extreme wind speed distribution is finite would invalidate the notion that probabilities of failure of a structure subjected only to wind loading, conditional on the structural strength being sufficiently large, are always larger than zero: if the structural strength corresponded to a wind speed larger than the length of the finite distribution tail, then the conditional failure probability would be zero.

4. ANALYSES OF DATA BASED ON SETS OF LARGEST DAILY WIND SPEEDS

In this section we first analyze data sets that reflect not only extreme winds occurring at various sites, but also ordinary winds. The analyses are intended to verify whether such sets can provide information on the parent population of the extremes. Next, we use a GPD-based approach to analyze sets of data that exceed relatively high thresholds.

Data Selection

From sets of largest daily wind speeds we obtained data samples that: (1) are relatively large so that sampling errors are acceptably small, and (2) have reduced mutual dependence among the data. The procedure for obtaining the data is as follows: Partition the set of daily maxima into small periods of size equal to or larger than the duration of typical storms in days. (A reasonable choice of the length of the period is eight days, but we also use sets based on four-day periods, and compare results of analyses based on the two choices.) Pick the largest value in each period. If the maxima of two adjacent periods are less than half a period apart, replace the smaller of the two maxima by the next smaller value in the respective period which is at least half a period apart from the larger maximum. A data set is thus obtained in which adjacent data are one period apart on the average and never less than half a period apart. We show below the daily maxima at Boise, Idaho in the first six eight-day periods of the year 1965. The periods are separated by vertical bars. The data selected by the procedure just described are in bold type. In the sixth period we underlined the period maximum (26), discarded and replaced by the next largest value (18) because of the proximity to the larger maximum (31) of the adjacent period.

23,32,**35**,20,26,24,24,14 | 13,**16**, 5,11, 5,12,12, 7 | 6, 6, 9, 9,11,12,25,**26** |

15,12,12, 7,15,12,**29**,10 | 7,10,15,20,20,17,24,**31** | 26,9,16,14,**18**,16,14,12|

Our investigation attempts to ascertain whether sets of data selected by this procedure from a set of daily maxima could possibly constitute samples from the parent populations of the extremes. Even though small correlations among data might subsist, we refer to a set obtained by the selection procedure just described as an uncorrelated data set based on eight-day (four-day) intervals or, for short, an eight-day (four-day) interval set.

Analysis of Uncorrelated Data Sets

We considered 48 uncorrelated data sets based on eight-day intervals, with length N ranging from 15 to 26 years. First we analyzed separately the sets of spring, summer, fall and winter data (seasonal data analyses). Next, we analyzed the data sets unsegregated by seasons. In both cases we estimated the best-fitting distributions (i.e., distributions with the largest PPCC) from among a set of seven distributions or families of distributions (normal, double exponential, lognormal, Gumbel, Fréchet, Weibull, and reverse Weibull).

Seasonal data sets

Our goal in performing the seasonal analyses was to attempt to fit to the spring, summer, fall and winter data, respectively, cumulative distributions $P_{sp}(v)$, $P_s(v)$, $P_f(v)$ and $P_w(v)$. Given these distributions, the distribution for all the uncorrelated data is

$$P(v) = P_{sp}(v)P_s(v)P_f(v)P_w(v). \tag{8}$$

We analyzed, for each season, 48 sets based on eight-day intervals. According to our results, for the spring, fall and winter records the best fitting distribution was predominantly reverse Weibull with shape $4 \leq \gamma \leq 30$. However, 29 summer records were better fitted by Gumbel distributions than by the reverse Weibull; the reverse Weibull (for the stations where it fitted the data better than the Gumbel distribution), and the Gumbel distribution (for the other stations), yielded estimated speeds with mean return period N years, V_N, that in most cases underpredicted the maximum speed recorded during N years, $V_{max,N}$. For summer records underpredictions were 15 percent or more for 16 sets, and 8 to 15 percent for 9 sets; there were only two overpredictions, both less than 8 percent. For spring records there were 12 underpredictions by 8 to 18 percent, and only three overpredictions, all less than 5 percent; comparable results were obtained for fall and winter. The results did not depend significantly on whether eight-day interval sets or four-day interval sets were used. From these and additional analyses we concluded that: (1) inferences from seasonal data sets (obtained as was described earlier from samples of largest daily data) do not provide a dependable basis for estimating extremes, but are likely to underestimate the extreme speeds. In other words, those sets are not drawn from populations underlying the extreme winds, but from mixed populations; (2) a similar conclusion applies to the sets consisting of all largest daily data for each season; (3) for these reasons the approach embodied in Eq. 8 appears to be inapplicable if all the data of the 8-day interval sets are considered.

Some researchers have indicated that the Weibull (as opposed to reverse Weibull) distribution best fits the sets of largest daily data. However, our analysis showed that the Weibull distribution fitted the seasonal data best only for less than ten percent of the sets.

Data sets unsegregated by seasons

The analysis of 48 sets based on eight-day intervals showed that the reverse Weibull (with $4 \leq \gamma \leq 22$) was the best fitting distribution for 27 sets, and fitted the data better than the Gumbel distribution for 41 sets. For 25 sets out of these 41 sets, including 12 sets for which it was optimal, the reverse Weibull underpredicted $V_{max,N}$ by 8 to 25 percent. For the 48 sets there were only 4 overpredictions, all smaller than 5 percent. In addition, the availability of largest annual data for periods N_1 ranging from 30 to 49 years allowed us to check the predictive capability of models inferred from sets based on eight-day intervals by comparing the estimated speed with mean return period N_1, V_{N1}, to the maximum speed recorded during an N_1-year period, $V_{max,N1}$, where $30 \leq N_1 \leq 49$. The underpredictions of the N_1-year speeds were more frequent and drastic than those of the N-year speeds. We concluded that estimated distributions of data sets unsegregated by seasons are too affected by the bulk of the non-extreme data to yield satisfactory estimates of extremes. Each of our conclusions for data segregated by seasons were found to be valid for data unsegregated by seasons as well.

Numerical Experiments

The analyses reported in the preceding paragraph showed that even where other distributions best fitted the data, the reverse Weibull was in most cases very close to being the best fitting distribution, i.e., its PPCC differed only in the fourth or even fifth significant figure from the PPCC of the best fitting distribution. We therefore reanalyzed the data based on eight-day intervals by assuming that the populations for all stations have a single reverse Weibull distribution with site-dependent location and scale parameters. This was done by calculating, for each station, the PPCC's based on the assumption that the shape parameter γ is 1,2,3,...50. For samples of data based on eight-day intervals and unsegregated by seasons the mean value of the PPCC's, taken over all the stations, was largest for $\gamma=11$, and the median PPCC was largest for $\gamma=13$. This is an indication that a reverse Weibull population with $\gamma=12$ would explain the results of the analyses. To see whether this is in fact the case, 48 samples of 730 data points each (corresponding to an 18-year record length based on 8-day intervals) were generated from reverse Weibull populations with (1) $\gamma=8$, (2) $\gamma=12$, and (3) $\gamma=16$. The number of simulated sets for which the best fitting reverse Weibull distribution had shape parameters with $\gamma\leq 12$, $13\leq\gamma\leq 20$, and $\gamma\geq 21$ are shown in Table 1. Also shown in Table 1 are the numbers of observed sets (average sample size 18 years) with $\gamma\leq 12$, $13\leq\gamma\leq 20$, and $\gamma\geq 21$. The results of Table 1 suggest that a reverse Weibull distribution with $\gamma\approx 12$ is an appropriate model for the population of extreme winds representing data based on 8-day intervals unsegregated by seasons, except for the larger number of samples with $\gamma\geq 21$ among the observed samples than among the simulated samples. We interpret this larger number as reflecting the relatively frequent presence of outliers among the observed samples. In our opinion this interpretation reinforces the point made earlier that, because wind speed populations which include ordinary speeds in addition to extremes are mixed, samples taken from such populations are not a sound basis for inferences on extremes. It is therefore necessary to "let the tails speak for themselves." This is done by applying to the data the GPD-based "peaks over threshold" approach.

Table 1. Numbers of Sets Best Fitted by Distributions with Various Values of γ

	$\gamma\leq 12$	$13\leq\gamma\leq 20$	$\gamma\geq 21$
Simulated sets, $\gamma=8$	48	0	0
Simulated sets, $\gamma=12$	27	17	4
Simulated sets, $\gamma=16$	8	24	16
Observed sets	26	12	10

"Peaks over Threshold" Analyses

In carrying out "peaks over threshold" analyses it is tempting to use a relatively low threshold in order to increase the number of data and thus reduce sampling errors. However, this introduces in the samples data that are not representative of the extremes and tend to bias the results. So that this does not happen the threshold being selected should be as high as possible, without reducing the size of the sample being analyzed to the point where the sampling errors become too large.

We selected the largest possible threshold subject to the restriction that the resulting sample size of the exceedances not be smaller than 15. Based on this selection, the average number of exceedances for our 48 sets based on 8-day intervals was $E(n_{ex})=16$, and the average threshold was $E(V_T)=45$, that is, less than the average median, $E(M)=50$, for the largest yearly speed samples analyzed in Section 3. For these thresholds we obtained $E(\hat{c})=-0.22$ and $SD(\hat{c})=0.44$. The results were virtually the same for the 48 sets based on 4-day intervals. These results would appear to lend support to the tentative conclusion of Section 3 that the extreme winds are described by a reverse Weibull distribution with shape parameter $\gamma \approx 5$, or perhaps somewhat larger, rather than by a Gumbel distribution. However, the weighted mean of the estimated c's, obtained as was shown for the results of Table A1, was close to zero. In addition, there were about as many estimated c's larger than -0.1 as there were smaller than -0.1. These results would suggest that the Gumbel distribution is appropriate. However, given the very wide confidence bands for our results, we conclude that no statement on whether the Gumbel or the reverse Weibull distribution is more appropriate can be made on the basis of this analysis.

In principle, the approach inherent in Eq. 8 may be based on "peaks over threshold" analyses. However, given that the records at our disposal are relatively short and the number of data exceeding a sufficiently high threshold for each of the seasons was judged to be too small, no attempt to perform "peaks over threshold" seasonal analyses was made in this work.

5. CONCLUSIONS

It is currently assumed in engineering loading models that non-hurricane and non-tornadic extreme wind speeds, regardless of their direction, are described by the Gumbel distribution (which corresponds to a shape parameter $\gamma=-1/c$ approaching infinity). The Gumbel distribution has infinite upper tail. The objective of this paper was to gain insights into the question of whether extreme wind speeds can be described by an extreme value distribution with limited upper tail, that is, by the reverse Weibull distribution.

We used in our analyses observed data, consisting of (a) sets of largest annual wind speeds, and (b) sets of largest daily wind speeds from which we extracted subsets suitable for extreme value analysis; and simulated data. Our results appear to suggest that extreme winds are better described by the reverse Weibull distribution than by the Gumbel distribution. However, given the small sample sizes used in our analyses, the superiority of one of the distributions over the other cannot be affirmed with confidence.

The tentative assumption that the extreme wind distributions are reverse Weibull, with shape parameter $\gamma \approx 5$ (GPD tail length parameter $c=-0.2$) and site-dependent location and scale parameters, yields wind load factors with an average value $\phi \approx 1.4$. This assumption, if confirmed, would invalidate earlier approaches to the estimation of the reliability of wind-sensitive structures, which depend on an infinite-tailed model of extreme wind speeds and therefore yield unrealistically high failure probabilities.

The "peaks over threshold" analyses were based in this paper on the Cumulative Mean Exceedance (CME) approach, which appears to be extremely sensitive to whether identical values of the variate in a set are left identical or modified by the addition to each of a different number much smaller than unity. Future work aimed at verifying the tentative conclusions of this paper will therefore include analyses based on different estimation procedures, including the de Haan procedure (Dekkers et al., 1989). In addition, we plan to perform analyses based on larger data sets, and more elaborate Monte Carlo simulations, in which the sets of samples generated by simulation will have the same sizes as the observed data sets being analyzed, rather than having a constant size. Finally, investigations are envisaged into the possibility that the shape parameter of the extreme wind speed distributions is site-dependent. This would be a departure from current practice, in which it is assumed that extreme winds are described by an extreme value distribution with universal shape parameter (that is, by the Gumbel distribution, which corresponds to a GPD tail length parameter $c \approx 0$), and site-dependent location and scale parameters.

6. ACKNOWLEDGMENTS

We thank S. Coles, D. Walshaw and P. Bortot for valuable suggestions and comments.

7. REFERENCES

ASCE Standard A7-93 (1994), Minimum Design Loads for Buildings and Other Structures, American Society of Civil Engineers, New York.

Castillo, E. (1988), Extreme Value Theory in Engineering, Academic Press, New York.

Davison, A.C., and Smith, R.L. (1990), "Models of Exceedances Over High Thresholds," Journal of the Royal Statistical Society, Series B, Vol. 52, pp. 339-442.

Dekkers, A.L.M., Einmahl, J.H.J., and de Haan, L. (1989b), "A Moment Estimator for the Index of an Extreme-value Distribution," Annals of Statistics, Vol. 17, pp. 1833-1855.

Draper, N. R. and Smith, H. (1966), Applied Regression Analysis, John Wiley & Sons, New York, pp. 77-80.

Ellingwood, B. R. et al., (1980), Development of a probability-based load criterion for Amer. National Standard A58, NBS Special Publ. 577, Natl. Bureau of Standards, Washington, D.C.

Gross, J. L., Heckert, A.N., Lechner, J.A., and Simiu, E. (1994), "Novel Extreme Value Estimation Procedures: Application to Extreme Wind Data," Proc., International Conference on Extreme Value Theory and Its Applications, J. Galambos et al. (eds.), Kluwer , Boston.

Gumbel, E. (1958), Statistics of Extremes, Columbia University Press, New York.

Pickands, J.(1975), "Statistical Inference Using Order Statistics," Annals Stat., Vol.3, 119-131.

Simiu, E., Changery, M. J. and Filliben, J.J.(1979), Extreme Wind Speeds at 129 Stations in the Contiguous United States, NBS BSS 118, Natl. Bureau of Standards, Washington, D.C.

Simiu, E., & Scanlan, R.H.(1986), Wind Effects on Structures, Wiley-Interscience, New York.

Smith, R. L. (1989), Extreme Value Theory, in Handbook of Applicable Mathematics, Supplement (W. Ledermann et al., editors), pp. 437-472, John Wiley and Sons, New York.

APPENDIX A

Table A1. Results of Analyses of Sets of Largest Annual Data

Station	N	V_{max}	V_{med}	n_{ex}	$\hat{c}$	SD($\hat{c}$)	V_N Gum.	V_N CME	V_{50} CME	V_{100000} CME	V_{100000} Gum.	V_{100000} c=-.20	ϕ
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
1. BIRMINGHAM, AL	34	62	44	22	-0.495	0.084	64	62	65	69	104	81	1.47
2. MONTGOMERY, AL	34	77	45	17	0.442	0.022	66	65	72	1071	117	95	1.56
3. TUCSON, AZ	39	78	50	21	0.187	0.086	76	71	79	273	121	103	1.54
4. YUMA, AZ	39	65	46	20	-0.701	0.056	65	66	66	68	116	83	1.44
5. FORT SMITH, AR	31	64	45	17	-0.798	0.158	63	64	64	65	113	80	1.43
6. LITTLE ROCK, AR	39	72	44	21	-0.395	0.070	72	71	73	83	136	97	1.59
7. FRESNO, CA	37	47	34	21	-0.204	0.036	47	46	48	59	77	59	1.42
8. RED BLUFF, CA	42	67	49	25	-0.751	0.115	67	69	67	68	114	81	1.38
9. SACRAMENTO, CA	39	63	43	20	-0.668	0.106	62	64	62	65	114	78	1.45
10. SAN DIEGO, CA	48	61	35	27	0.384	0.022	56	50	57	584	85	75	1.57
11. DENVER, CO	33	61	48	17	-0.265	0.112	60	61	61	69	93	72	1.30
12. GRAND JUNCTION, CO	33	70	52	17	0.104	0.065	67	66	70	146	103	86	1.37
13. PUEBLO, CO	43	79	61	25	-0.477	0.060	80	81	80	85	126	95	1.34
14. HARTFORD, CT	44	67	43	28	-0.238	0.028	67	64	68	84	109	87	1.54
15. WASHINGTON, DC	39	66	47	24	-0.325	0.055	67	65	68	77	106	84	1.44
16. ATLANTA, GA	42	76	46	22	-0.034	0.030	74	72	76	134	131	101	1.59
17. MACON, GA	33	64	45	17	-0.271	0.073	64	65	66	79	118	84	1.46
18. BOISE, ID	48	62	47	24	-0.099	0.068	62	61	62	84	92	75	1.35
19. POCATELLO, ID	48	72	53	25	0.019	0.157	75	71	75	132	113	94	1.43
20. CHICAGO MIDWAY, IL	37	63	46	22	-0.128	0.028	63	61	64	86	97	79	1.41
21. MOLINE, IL	44	72	52	25	-0.665	0.082	72	75	73	75	124	90	1.41
22. PEORIA, IL	42	72	50	22	-0.302	0.075	70	72	71	82	122	89	1.43
23. SPRINGFIELD, IL	32	71	54	16	-0.199	0.118	68	69	69	83	110	83	1.32
24. EVANSVILLE, IN	44	61	47	23	-0.130	0.106	63	63	63	82	103	77	1.36
25. FORT WAYNE, IN	46	69	52	23	-0.163	0.063	69	70	70	88	110	85	1.37
26. INDIANAPOLIS, IN	36	93	53	18	0.023	0.064	81	81	85	171	151	113	1.56
27. BURLINGTON, IA	23	72	55	13	-1.751	0.515	69	76	69	69	140	80	1.28
28. DES MOINES, IA	37	80	56	19	0.042	0.034	79	77	81	155	133	102	1.45
29. SIOUX CITY, IA	46	88	57	24	0.063	0.065	85	80	85	180	132	110	1.49
30. CONCORDIA, KS	20	74	56	11	-0.933	0.188	73	76	74	75	145	89	1.35
31. DODGE CITY, KS	41	72	59	22	-0.709	0.048	72	75	72	73	114	83	1.25
32. TOPEKA, KS	34	79	54	17	-0.180	0.160	71	74	73	91	128	90	1.38
33. WICHITA, KS	41	89	57	22	0.218	0.056	83	80	86	325	134	110	1.49
34. LOUISVILLE, KY	39	66	49	21	-0.085	0.055	63	65	64	87	107	77	1.34
35. PORTLAND, ME	45	73	46	23	-0.283	0.088	73	70	73	90	122	97	1.55
36. BALTIMORE, MD	39	71	54	20	-0.274	0.078	70	71	71	81	116	85	1.34
37. BOSTON, MA	50	85	54	26	-0.126	0.039	84	81	84	119	139	109	1.52
38. NANTUCKET, MA	23	71	55	14	-1.402	0.217	71	73	72	72	128	85	1.33
39. DETROIT, MI	46	68	49	25	-0.207	0.091	67	68	67	81	111	82	1.38
40. GRAND RAPIDS, MI	29	67	47	15	-0.930	0.128	69	71	70	71	136	89	1.47
41. LANSING, MI	38	67	51	21	-0.646	0.063	68	69	68	71	110	83	1.36
42. SAULT STE MARIE, MI	47	67	46	24	-0.351	0.059	65	67	65	74	113	82	1.43
43. DULUTH, MN	36	70	49	19	-0.310	0.033	69	70	70	81	122	88	1.44
44. MINNEAPOLIS, MN	42	82	46	27	-0.023	0.048	79	72	81	149	127	108	1.65
45. JACKSON, MS	29	64	44	16	-0.353	0.073	64	62	66	75	109	84	1.48
46. COLUMBIA, MO	35	65	51	20	-0.103	0.075	66	66	67	89	109	81	1.35
47. KANSAS CITY, MO	51	75	49	31	-0.116	0.038	74	72	74	104	120	94	1.49
48. ST.LOUIS, MO	21	66	46	14	0.035	0.075	64	62	70	135	111	89	1.51

Station	N	V_{max}	V_{med}	n_{ex}	$\hat{c}$	SD($\hat{c}$)	V_N Gum.	V_N CME	V_{50} CME	V_{100000} CME	V_{100000} Gum.	V_{100000} c=-.20	ϕ
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
49. SPRINGFIELD, MO	44	71	49	22	-0.127	0.068	68	67	69	93	111	86	1.42
50. BILLINGS, MT	49	84	58	26	-0.051	0.031	84	81	84	130	132	106	1.45
51. GREAT FALLS, MT	44	75	59	24	-0.400	0.078	74	78	74	80	125	87	1.29
52. HAVRE, MT	27	78	57	14	-0.229	0.227	77	76	80	97	135	99	1.41
53. HELENA, MT	48	71	55	25	-0.241	0.114	70	71	70	81	111	83	1.31
54. MISSOULA, MT	43	71	47	22	-0.157	0.050	71	65	72	97	105	92	1.50
55. NORTH PLATTE, NE	31	74	61	16	-0.709	0.044	74	77	75	76	123	87	1.26
56. OMAHA, NE	51	104	50	29	0.294	0.054	92	81	91	640	145	125	1.68
57. VALENTINE, NE	27	74	61	15	-0.574	0.120	74	78	76	78	130	88	1.27
58. ELY, NV	49	70	51	28	-0.191	0.023	70	68	70	87	107	86	1.39
59. LAS VEGAS, NV	20	70	55	12	-0.938	0.362	68	70	69	69	121	80	1.28
60. RENO, NV	45	77	55	24	-0.463	0.053	77	78	77	83	129	96	1.41
61. WINNEMUCCA, NV	38	63	47	21	-1.102	0.071	63	67	64	64	114	77	1.37
62. CONCORD, NH	46	68	41	23	-0.040	0.038	66	63	66	116	111	89	1.58
63. ALBUQUERQUE, NM	52	85	56	26	0.139	0.023	79	78	78	195	124	98	1.41
64. ROSWELL, NM	36	82	57	18	0.088	0.119	81	81	84	190	145	108	1.48
65. ALBANY, NY	46	68	46	29	-0.086	0.081	66	64	67	95	103	83	1.45
66. BINGHAMPTON, NY	35	65	48	18	-0.370	0.081	64	66	65	72	115	80	1.38
67. BUFFALO, NY	44	79	52	22	0.463	0.050	73	69	74	1049	109	94	1.45
68. LA GUARDIA, NY	33	73	57	17	0.314	0.032	71	71	74	344	113	89	1.33
69. ROCHESTER, NY	45	66	52	23	-0.717	0.094	66	69	66	67	106	78	1.30
70. SYRACUSE, NY	45	67	51	23	0.014	0.089	66	65	66	105	100	79	1.33
71. CHARLOTTE, NC	29	65	42	17	-0.481	0.244	65	62	67	73	112	87	1.55
72. GREENSBORO, NC	50	67	41	25	-0.089	0.073	62	62	62	94	109	81	1.51
73. BISMARCK, ND	40	69	58	20	-0.582	0.145	69	72	69	71	107	79	1.23
74. FARGO, ND	45	100	57	25	0.252	0.032	93	86	95	484	151	127	1.60
75. WILLISTON, ND	18	69	56	9	-0.469	0.127	68	68	71	75	116	84	1.31
76. CLEVELAND, OH	35	69	53	19	-0.224	0.080	68	69	69	81	111	82	1.33
77. COLUMBUS, OH	30	61	49	15	-1.236	0.106	60	64	60	60	106	70	1.26
78. DAYTON, OH	41	72	52	24	-0.170	0.072	74	72	75	97	121	94	1.45
79. TOLEDO, OH	45	82	48	24	0.207	0.050	76	72	78	308	125	103	1.57
80. OKLAHOMA CITY, OK	30	69	53	15	-0.135	0.081	67	67	69	89	105	84	1.35
81. TULSA, OK	35	68	49	18	0.019	0.107	63	65	65	106	109	79	1.35
82. PORTLAND, OR	38	88	49	19	0.268	0.077	80	75	85	503	138	116	1.64
83. HARRISBURG, PA	38	64	45	19	-0.599	0.087	63	65	64	67	114	80	1.42
84. PHILADELPHIA, PA	33	62	47	21	-0.507	0.094	63	64	64	67	105	77	1.38
85. PITTSBURGH, PA	18	60	47	11	-0.591	0.132	61	60	63	65	102	75	1.36
86. SCRANTON, PA	33	57	44	17	-0.387	0.092	56	57	57	62	89	68	1.32
87. BLOCK ISLAND, RI	31	86	60	16	-0.182	0.120	82	82	84	107	138	105	1.42
88. GREENVILLE, SC	43	72	46	22	-0.483	0.069	69	75	70	76	143	90	1.50
89. HURON, SD	49	79	59	26	-0.447	0.059	80	82	80	86	132	98	1.38
90. RAPID CITY, SD	43	70	62	22	-0.341	0.096	71	73	72	76	102	80	1.18
91. CHATTANOOGA, TN	35	76	46	18	-0.336	0.066	76	73	78	93	141	105	1.61
92. KNOXVILLE, TN	33	66	50	18	0.007	0.042	66	65	68	111	109	83	1.38
93. MEMPHIS, TN	21	61	45	11	-0.370	0.369	57	58	59	65	103	71	1.34
94. NASHVILLE, TN	34	70	45	17	-0.178	0.092	67	66	70	94	119	91	1.53
95. ABILENE, TX	36	100	54	19	0.550	0.047	79	78	85	2491	141	111	1.54
96. AMARILLO, TX	34	81	62	17	0.201	0.082	77	78	79	214	124	95	1.32
97. AUSTIN, TX	37	58	45	19	-0.189	0.057	57	58	58	69	91	68	1.31
98. DALLAS, TX	32	67	48	17	-0.233	0.043	65	64	67	81	107	83	1.41
99. EL PASO, TX	32	67	55	17	-0.187	0.103	68	67	69	82	99	81	1.28

							V_N		V_{50}	V_{100000}			
Station	N	V_{max}	V_{med}	n_{ex}	$\hat{c}$	SD($\hat{c}$)	Gum.	CME	CME	CME	Gum.	c=-.20	ϕ
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
100. SAN ANTONIO, TX	36	80	46	19	0.236	0.041	70	68	73	328	123	96	1.55
101. SALT LAKE CITY, UT	46	69	49	28	-0.333	0.038	70	69	70	80	113	88	1.44
102. BURLINGTON, VT	40	66	44	23	-0.232	0.093	66	63	67	84	107	86	1.51
103. LYNCHBURG, VA	44	53	39	22	-0.750	0.154	52	56	52	53	94	63	1.37
104. RICHMOND, VA	33	61	42	20	0.192	0.058	59	56	63	204	95	79	1.48
105. NORTH HEAD, WA	41	104	67	28	-0.250	0.055	105	96	106	130	158	136	1.55
106. QUILLAYUTE, WA	21	45	35	12	-0.184	0.018	45	44	47	58	72	57	1.37
107. SEATTLE, WA	20	59	43	11	0.027	0.069	56	56	62	112	97	77	1.44
108. SPOKANE, WA	47	65	48	24	0.028	0.040	64	64	65	111	101	79	1.37
109. TATOOSH ISLAND, WA	54	86	66	27	-0.290	0.045	86	85	86	98	128	104	1.34
110. GREEN BAY, WI	36	103	54	18	0.431	0.089	85	82	92	1408	153	126	1.63
111. MADISON, WI	41	75	48	21	-0.516	0.058	75	75	76	82	134	100	1.54
112. MILWAUKEE, WI	42	68	54	21	-0.423	0.059	67	70	67	72	109	79	1.29
113. CHEYENNE, WY	46	73	61	24	-0.476	0.051	74	76	74	77	113	85	1.24
114. LANDER, WY	42	80	58	21	-0.621	0.067	77	83	77	80	142	94	1.35
115. SHERIDAN, WY	44	82	61	24	0.125	0.073	82	80	83	184	128	101	1.38
MEAN	38.1	71.9	50.4	20.5	-0.257 (unweighted)								1.42
SD	8.2	10.5	6.3	4.5	0.384 (unweighted)								0.10

Key:

Col.	Notation	Description
(1)	Station	Name of NWS Station
(2)	N	Sample size
(3)	V_{max}	Maximum observed wind speed (mph)
(4)	V_{med}	Median observed wind speed (mph)
(5)	n_{ex}	Number of exceedances
(6)	$\hat{c}$	Estimated c
(7)	SD($\hat{c}$)	Standard deviation of $\hat{c}$
(8)	V_N Gum	Estimated N-yr wind based on Gumbel model ($\hat{c}$ from Col. 6)
(9)	V_N CME	Estimated N-yr wind based on CME method ($\hat{c}$ from Col. 6)
(10)	V_{50} CME	Estimated 50-yr wind based on CME method ($\hat{c}$ from Col. 6)
(11)	V_{100000} CME	Estimated 100,000-yr wind based on CME method ($\hat{c}$ from Col. 6)
(12)	V_{100000} Gum	Estimated 100,000-yr wind based on Gumbel model ($\hat{c}$ from Col. 6)
(13)	V_{100000} c=-.2	Estimated 100,000-yr wind based on CME method (c=-0.20)
(14)	ϕ	Load factor based on c=-0.20

Note: 1 mph = 0.44704 m/s

APPENDIX B

Estimation of SD of $\hat{c}$

We consider the problem of fitting CME_i = intercept + slope $\cdot X_i + \epsilon_i$ where the variance of ϵ_i is proportional to $1/(n-i)$ for $i=1,2,\ldots,n-1$.

GIVEN:

1) Speeds x_i, $i=1,2,\ldots,n$
2) CME values y_i, $i=1,2,\ldots,(n-1)$

Define "relative variances" $v_i=(n-i)^{-1}$, $i=1,2,\ldots,(n-1)$ and $\underline{V} = \mathrm{diag}(\underline{v})$.

Let $\underline{B}^T$ = [intercept slope] $\underline{x}^T = [x_1, x_2, \ldots, x_{n-1}]$ and $\underline{X} = [\underline{1}\ \underline{x}]$ (T denotes transpose).

Also, let $\underline{M} = [\underline{X}^T V^{-1} \underline{X}]^{-1}$

$$= \begin{bmatrix} \Sigma(n-i) & \Sigma(n-i)\,x_i \\ \Sigma(n-i)\,x_i & \Sigma(n-i)\,x_i^2 \end{bmatrix}^{-1} = (SSX)^{-1} \begin{bmatrix} \Sigma(n-i)\,x_i^2 & -\Sigma(n-i)\,x_i \\ -\Sigma(n-i)\,x_i & \Sigma(n-i) \end{bmatrix}$$

(All summations are for $i=1,2,\ldots,(n-i)$.)

Here and below, $SSX = [\Sigma(n-i)]\,[\Sigma(n-i)x_i^2] - [\Sigma(n-i)x_i]^2$.

Then the parameter estimates $\underline{b}$ are given by:

$$\underline{b} = \underline{M} \begin{bmatrix} \Sigma(n-i)\,y_i \\ \Sigma(n-i)\,x_i\,y_i \end{bmatrix};$$

with covariance matrix $\sigma^2 \underline{M}$;

and $\hat{\sigma}^2$ is $(1/(n-i))\,(\underline{y}-\underline{x}^T\underline{b})^T\,\underline{V}^{-1}(\underline{y}-\underline{x}^T\underline{b})$

$$= (1/(n-1))\ \Sigma(n-i)\,(y_i - \text{intercept} - \text{slope} \cdot x_i)^2.$$

Our estimate of $\hat{c}$ is slope/(1+slope); and the estimated standard deviation of the c-estimate is $(1+\text{slope})^{-2}$ times the standard deviation of the slope, or to within a constant (see Draper and Smith, 1966)

$$SD(\hat{c}) = \frac{[\Sigma(n-i)]^{1/2}\,[\Sigma(n-i)\,(y_i - \text{intercept} - \text{slope} \cdot x_i)^2]^{1/2}}{(n-3)^{1/2}(1+\text{slope})^2\,(SSX)^{1/2}}$$