Wind and Structures, Vol. 1, No. 1 (1998) 25-42
DOI: http://dx.doi.org/10.12989/was.1998.1.1.025

Improving the Gumbel analysis by using *M*-th highest extremes

Nicholas J. Cook†

Wind Engineering Consultant, 10 Arretine Close, St Albans AL3 4JL, U.K.

Abstract. Improvements to the Gumbel method of extreme value analysis of wind data made over the last two decades are reviewed and illustrated using sample data for Jersey. A new procedure for extending the Gumbel method to include *M*-th highest annual extremes is shown to be less effective than the standard method, but leads to a method for calibrating peak-over-threshold methods against the standard Gumbel approach. Peak-over-threshold methods that include at least the 3rd highest annual extremes, specifically the modified Jensen and Franck method and the "Method of independent storms" are shown to give the best estimates of extremes from observations.

Key words: extreme value analysis; anemometers; sub-annual extremes; Gumbel analysis; statistical independence; recurrence.

1. Introduction

Gumbel's method of extreme value analysis (Gumbel 1958), which is founded on the extreme-value theory of Fisher and Tippett (1928), has been the basis for determining design wind speeds ever since the first statistical analysis of UK wind speeds by Shellard (1963). The Gumbel analysis has conventionally been performed on annual maxima of wind speed or dynamic pressure, depending on the method of measurement. Annual maxima are used for two principal reasons

1. To contain whole cycles of seasonal trends.
2. To contain a sufficiently large population of independent events for reasonable convergence to the Gumbel (Fisher Tippett type I or FT1) asymptotic model.

Also the length of the data record must be sufficient to meet two requirements

a) To be representative of the wind climate.
b) To give a sufficiently accurate analysis.

The intervening years have seen a number of significant insights to extreme value theory, giving improvements in the methodology. Sample data that are compatible with all the methods are used to illustrate the methodology reviewed. These are "peak over threshold" data, used previously (Cook 1982) and reproduced in Table 1. They comprise the daily maximum gust speed exceeding 29 m/s (56 kts, Beaufort Range 11) at Jersey Airport during the 21-year period 1958 to 1978 inclusive, from which the required data for each method may be extracted. Several features in these data are useful in testing the robustness of the methods

1. There is significant variation in windiness from year to year as expressed by the frequency of exceeding the threshold.

† Professor

2. The 29 m/s threshold was not exceeded in 1971 or 1973, so annual maxima for these years are "missing".
3. The largest value, 48 m/s on 9 October 1964, is now known to be erroneous.

2. Standard Gumbel method

2.1. Methodology

In the standard Gumbel method, the cumulative density function (CDF) of the annual maxima, P, is estimated from the order statistics as follows

1. The population $n = N$ annual maxima are assigned a rank, m, in ascending value, $1 \leq m \leq n$
2. The CDF is estimated from

$$P = \frac{m}{n+1} \tag{1}$$

3. The CDF is fitted to the FT1 asymptotic model

$$P = \exp(-\exp(-y)) \tag{2}$$

where y is the reduced variate, given by

$$y = a(V - U) \tag{3}$$

and the variate, V, may be the speed or dynamic pressure of the wind.

The fit is conventionally performed on a "Gumbel plot," where the ordinate is the reduced variate from the estimated probability in Eqs. (1) and (2), plotted as $y = -\ln(-\ln(P))$ and the abscissa is the variate value V. This "Gumbel plot" transforms the FT1 model equation, Eq.(2), to a straight line, with slope $1/a$ and intercept U. The best fit was originally determined by the method of least squares with y as the independent variable and V as the dependent variable.

Gumbel (1958) demonstrated that the estimate of probability by Eq. (1) is most reliable around the mode and becomes less reliable in the extreme tails of the CDF. He provided confidence limits for the tails and recommended that any value lying outside these limits be rejected from the analysis. This is most likely to occur with the largest value of rank $m = n$, since there is a chance that it represents a mean recurrence interval longer than the record period. For example, the chance of the 50-year return wind speed occurring in any given ten-year record is 1/5.

2.2. Gumbel method applied to the Jersey data

The standard "Gumbel Plot" for the gust dynamic pressure, q, at Jersey is shown as Fig. 1. The reason for using dynamic pressure instead of wind speed is addressed later.

All the annual maxima (shown bold in Table 1) are represented by the □ symbols in Fig. 1. It is not appropriate to reduce the population of annual maxima from n=21 to n=19 to account for the two "missing" values for 1971 and 1973. They are less than 29 m/s, so rank as m=1 and m=2, but simply do not appear or contribute to the fit because the values are unknown. Therefore ranks m=3 to m=21 appear in the plotting positions for a population $n = 21$. The conventional least-mean-squares fit, taking the variate q as the dependent variable, is shown by the dashed line.

What would have happened if the "missing" values had simply been discarded and the

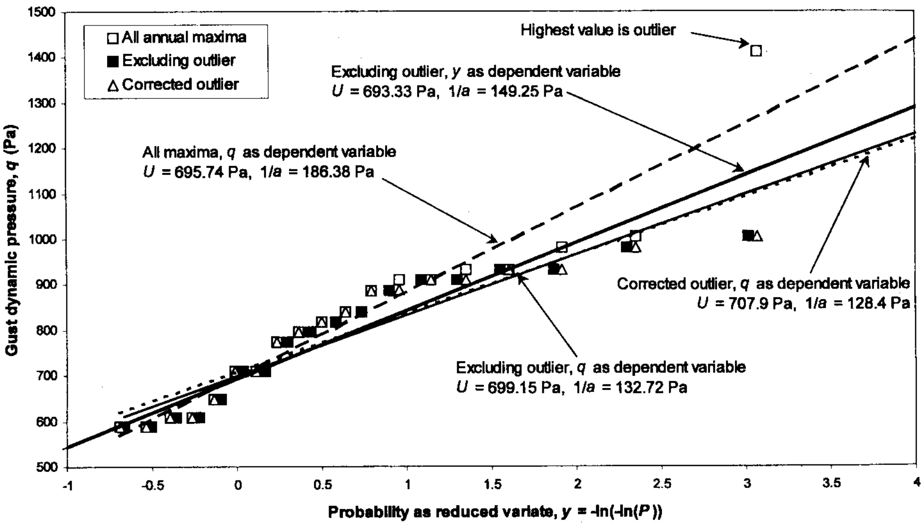

Fig. 1 Standard Gumbel plot for Jersey

population reduced to n=19? There are 210 different combinations of extracting 19 values from a population of 21, but this particular combination uniquely selects the highest possible 19 values. Discarding the "missing" values would have produced an unwanted bias in the plotting positions. However, it is clear that the highest annual maximum, corresponding to 48 m/s on 9 October 1964, lies well outside Gumbel's confidence limits so is an "outlier". In response to this observation, the outlier was discarded, and the data re-plotted using a population of n=20, represented by the ■ symbols, and fitted as before by the thinner continuous line.

There is usually no alternative to discarding the outlier and reducing the population, since the "true" value may lie anywhere in the range, but is most likely to lie near the mode. The occurrence of the outlier can be quite natural – the 1/1000-year extreme will occur sometime, so why not in these twenty years? It can also be due to a different mechanism, such as tornado or lee-waves, so not be a member of the general population – the "mixed population" problem addressed by Gomes and Vickery (1978) (Cook 1985). The reason in this case is more prosaic. Inspection of the original data records has revealed that the observer's entry of 74 kts had later been wrongly transcribed as 94 kts. Fortunately, this corrected gust speed of 38 m/s remains the extreme for 1964, enabling the full population of 21 extremes to be re-plotted, represented by the Δ symbols, and re-fitted by the dotted line. It is clear that the result is closely similar to discarding the outlier, supporting Gumbel's procedure for outliers.

3. Improved methodology

3.1. Gumbel-Lieblein and Harris methodology

Lieblein (1974) made a most significant contribution to the fitting process of the Gumbel method by devising "best linear unbiased estimators" (BLUE) that gave a different weight to

Note: Because it is unlikely in practice that an outlier could be corrected, or would still be the annual extreme after correction, the outlier has been discarded in all subsequent comparisons.

Table 1 Days when gust speeds exceeded 29 m/s at Jersey Airport, 1958-1978

Year	Month	Day	Gust speed (m/s)
1958	Jan	6	32.0
		7	(30.0)
		9	33.5
		11	(30.0)
		21	**38.5**
	Feb	25	31.0
	May	16	35.0
	Dec	11	30.0
1959	Jan	2	29.5
		20	33.0
		22	(29.0)
	Oct	17	34.5
		27	33.0
	Nov	13	34.0
	Dec	26	**37.0**
		27	(36.0)
		28	(34.0)
1960	Nov	1	**34.0**
		2	(32.0)
	Dec	4	31.0
1961	Jan	28	37.0
	Feb	2	**38.0**
		3	(30.0)
		6	(31.0)
		7	31.5
	Jul	13	32.0
	Aug	8	31.5
	Dec	5	33.0
1962	Jan	9	(30.5)
		10	(35.0)
		11	**36.0**
		12	(30.5)
		21	32.0
		22	(32.0)
	Feb	13	29.0
	Apr	4	(29.0)
		5	32.0
		8	31.5
	May	19	30.5
	Nov	18	29.0
	Dec	9	29.5
		15	33.5

Year	Month	Day	Gust speed (m/s)
1963	Jan	16	(32.0)
		18	29.0
	Mar	11	**31.0**
1964	Oct	7	(30.0)
		9	**48.0**
		10	(33.0)
		23	29.0
	Nov	15	(29.5)
		163	31.5
1965	Jan	13	(30.5)
		14	31.0
		17	38.0
		18	(35.0)
	Mar	26	29.5
	June	16	32.0
	Nov	23	31.5
		27	**38.5**
		29	(33.0)
	Dec	2	(30.5)
		3	36.5
		5	(36.0)
		6	(30.5)
		9	37.5
		10	(34.0)
		30	30.0
1966	Jan	1	(32.0)
		2	**40.5**
	Mar	24	30.0
		27	34.5
	Nov	16	31.5
	Dec	1	(32.0)
		2	(33.5)
		12	31.0
		13	(30.0)
1967	Feb	19	(31.4)
		20	35.5
		21	(30.5)
		23	29.0
		28	31.0
	Mar	12	31.5
	Apr	21	30.0
	Oct	4	29.5

Year	Month	Day	Gust speed (m/s)
		15	(29.5)
		17	35.0
		27	33.5
	Oct	28	(29.5)
		31	**39.0**
	Nov	1	(31.5)
		2	(31.0)
		3	(30.0)
		29	33.0
	Dec	6	29.5
		23	(31.5)
		25	32.0
1968	Jan	6	31.0
		9	**39.0**
1969	Jan	17	**34.0**
	Jul	6	29.5
	Nov	9	31.0
1970	Feb	12	**31.0**
1972	Jan	27	**31.5**
	Fed	2	29.0
	Apr	11	29.0
	Nov	12	31.5
		13	(29.0)
		20	29.5
1974	Jan	11	29.5
		16	**36.5**
	Sep	23	29.5
		24	(29.5)
	Dec	11	29.0
1975	Jan	27	**35.5**
1976	Jan	2	29.0
	Oct	14	34.5
	Nov	30	(32.0)
	Dec	1	**40.0**
		2	(31.5)
1977	Jan	13	**30.5**
		14	(30.5)
	Nov	14	32.5
1978	Jan	29	**31.5**
	Feb	2	29.0

Source: Meteorological Office, States of Jersey each extreme value, according to its reliability. He provided BLUE for populations of n=10 to 24 extremes (tabulated in Cook 1985), noting that the estimate of dispersion becomes unreliable for n<10, while n=16 is about optimum, and analysis accuracy is not significantly improved with n> 16. Note that the Lieblein BLUE would not cope with the "missing" ranks 1 and 2 in the Jersey data, since they require to be applied to the full set of values.

Harris (1996) revisited this subject recently and suggested a number of improvements. The first was to improve the plotting positions by transforming the variate, V, to $-\ln(-\ln(V))$ before estimating the CDF. The second was to derive new unbiased estimators for the revised plotting positions in the form of weights to be used with the classic method of least squares. These weights have a practical advantage over the Lieblein BLUE in that they are tolerant of missing data values. If a value is missing (typically the smallest value, because it did not exceed a recording threshold) its weight is set to zero and the weights for all values are re-normalised to unity.

Harris (1996) also presented an eloquent argument for swapping the plotting axes, as suggested by Castillo (1988), treating the measured wind data as the independent variable (ordinate) and the reduced variate from the estimated probability as the dependent variable (abscissa). Accordingly, this complementary form of fit is included in Fig. 1 as the thicker continuous line, which gives a higher dispersion (slope) than before. Harris's argument is based on the way the data are measured in the field – that the measurements of wind speed are exact (independent) and their probability is estimated (dependent). But the reverse applies when the results are applied in design – given a required (independent) design risk, the best estimate of (dependent) wind speed is required – making the conventional fit appear the more appropriate. There is advantage to performing both fits, since the ratio of the two slopes is the regression coefficient R^2 that provides an objective measurement of the accuracy of the fit. With random uncorrelated data, one fitted line will be horizontal, $Y = \overline{X}$, and the other vertical, $X = \overline{Y}$, and $R^2 = 0$. Of course, this can never happen in on Gumbel plot because the ranking process always results in a "staircase" curve, rising from left to right. In a perfect fit, both lines superimpose and $R^2 = 1$. In practice, R^2<1, and the fit with probability as the dependent variable, as argued by Harris, provides conservative and therefore "safer" predictions of extremes.

3.2. Modified Jensen and Franck method

Faced with the requirement to determine design wind speeds for Denmark with no existing data records, Jensen and Franck (1970) devised instrumentation for recording daily maxima. They reported first results with only three years of data, then again with seven years of data. They extracted statistically independent maxima from separate storms by inspection of the record and rejection of extremes occurring on consecutive days. They then formed the CDF of storm maxima and extrapolated into the upper tail to the required design risk. While we may question whether these record lengths are sufficiently representative of the wind climate over the long term, the innovative use of statistically independent storm maxima brought significant improvements in analysis accuracy.

Cook (1982) modified the Jensen and Franck method to be compatible with the Gumbel plot. This method (MJ&F) requires "peak over threshold" data of the form shown in Table 1. Storm maxima separated by at least two days are extracted. The cumulative probability distribution of the storm extremes, P_S, is formed from the order statistics in the conventional manner

$$P_S = \frac{m}{n+1} \tag{4}$$

The cumulative probability of the annual extremes is then estimated from the storm distribution on the assumption of statistical independence

$$P = P_S^r \tag{5}$$

where $r = n/N$ is the observed average annual rate of independent storms exceeding the threshold. The distribution is plotted on the standard Gumbel plot and fitted to the FT1 distribution as before. Neither the Lieblein BLUE nor the Harris weights can be used directly in the fit, because the number of values and their plotting positions differ from the annual maxima. Cook (1982) worked around this problem by interpolating the data to the standard plotting positions. More recently, Harris (1998) solved this problem by devising a procedure to determine the weights for the variable plotting positions produced by Eq. (5).

Cook (1982) also investigated the rate of convergence to the FT1 model, showing that dynamic pressure converges much more rapidly than wind speed, so that the CDF of annual maximum dynamic pressure provides a significantly better fit to the FT1 asymptote than wind speed. This is the reason for the choice of dynamic pressure for the examples in this paper, in Harris (1997, 1998), and also in the UK and European codes of practice for wind loads (BSI 1997a, 1997b).

3.3. "Method of independent storms"

The "Method of independent storms" (MIS) (Cook 1982, Harris 1996), is a similar, but more sophisticated, form of analysis that requires a continuous data record. The record is first searched by a process that identifies individual storms, then the maximum wind speed in each storm is extracted. This method typically identifies 100 to 150 storms per year in the UK. However, because of this increased power of r in Eq. (5), only about 3 storms per year are significant in the analysis. The main advantage of the method is the ability to extract storm maxima by season or to search within storms for maxima within directional sectors. The method has been shown (Cook 1983, Cook and Prior 1987) to give reliable estimates of the FT1 parameters by month and by 30°-wide sectors of direction.

3.4. Modified Jensen and Franck method applied to the Jersey data

The extremes separated by at least two days were extracted for analysis from the peak-over-threshold data for Jersey. Rejected extremes are indicated in Table 1 by parentheses, thus (29.5). This leaves 84 independent extremes in the 21 year period, giving an average annual rate of exactly $r = 4$. These are plotted in Fig. 2, denoted by the □ symbols, with the annual maxima included for comparison and denoted by ■ symbols. The improvement in analysis accuracy is immediately apparent, with the data forming a straighter and less stepped curve, and the two forms of FT1 fit being more closely similar. However, the MJ&F method typically gives higher estimates of mode, U, and lower estimates of dispersion, $1/a$, than the standard Gumbel method. This is addressed later.

Both Cook (1982) and Harris (1996) have commented on the insensitivity of the method to the threshold and hence the number of extremes. This is illustrated by restricting the data to the 1st, 2nd and 3rd highest extremes only, leaving the 46 values extremes denoted by the Δ symbols. The fitted lines on the graph are indistinguishable from those for the full set, the fitted modes and dispersions differing by only a few percent. The reason for this becomes obvious on

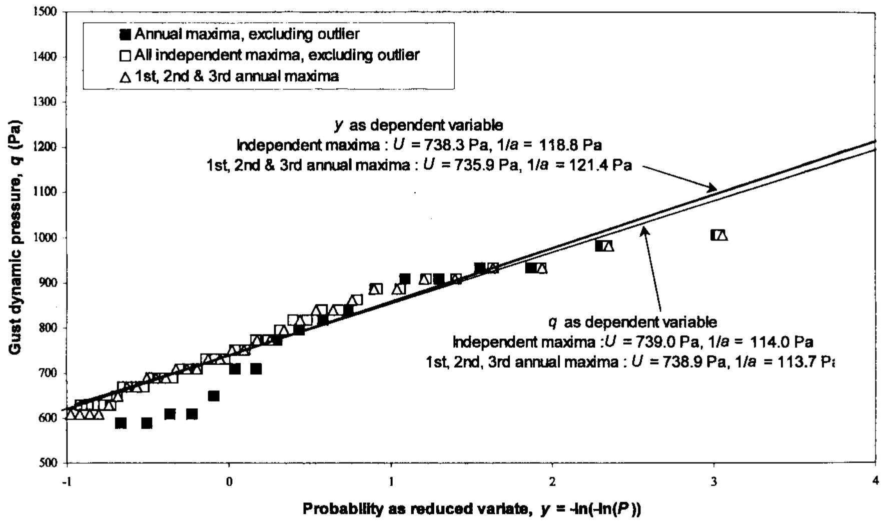

Fig. 2 Modified Jensen and Franck method for Jersey

inspection, since only one of the excluded higher-order extremes, $M \geq 4$, lies to the right of the mode, and the change to r=2.19 gives new plotting positions that compensates for the loss of data.

To check whether the two-day separation was sufficient to ensure independence, the annual rate of independent gusts exceeding 29 m/s, r_i, was determined from CDF of the full set of 123 correlated extremes of Table 1 using Eq. (5). The value of r_i was iterated until the fitted mode matched the previous analysis. The match shown in Fig. 3 was obtained when r_i=5.27, which confirms that the lower rate of r=4.0 obtained by the selection process had indeed ensured statistical independence.

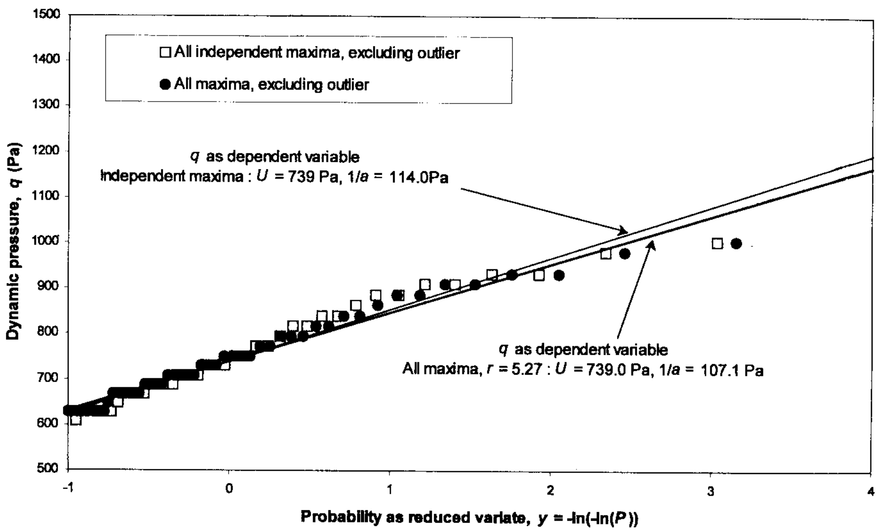

Fig. 3 Determining rate of independent exceedances of threshold for Jersey

4. Gumbel's method incorporating M-th highest extremes

4.1. Methodology

Having prepared this paper so far as an illustrated review, and having noted the efficiency of the MJ&F method using just the 1st, 2nd and 3rd highest annual maxima, the writer reconsidered Gumbel's (1958) treatment of M-th highest extremes. From this, it seemed possible to develop a methodology that accounted for them directly.

Gumbel gives an equation for generating the CDF for any M, P_M, from the CDF of the extreme, P

$$P_M = P^M \sum_{\nu=0}^{M-1} \frac{M^\nu \exp(-\nu y_M)}{\nu!} \tag{6}$$

where y_M is the reduced variate for the M-th highest extreme, given by

$$y_M = y + \ln(M) \tag{7}$$

This gives equations for the CDF of the 2nd and 3rd - highest extremes, P_2 and P_3

$$P_2 = \exp(-2\exp(-y - \ln(2)))[1 + 2\exp(-y - \ln(2))] \tag{8}$$

$$P_3 = \exp(-3\exp(-y - \ln(3))[1 + 3\exp(-y - \ln(3)) + \frac{9}{2}\exp(-2(y + \ln(3)))] \tag{9}$$

Eqs. (8) and (9) are plotted with Eq. (2) in Fig. 4, showing they are curves on the Gumbel plot.

Unlike Eq. (2) for P, Eqs. (8) and (9) cannot be solved directly for y, but do fit excellently to the cubic polynomials shown in Fig. 4, which allow values of y to be determined.

4.2. Applied to the Jersey data

Fig. 5 shows the 1st, 2nd and 3rd highest annual maxima for Jersey, with curves of Eqs. (2),

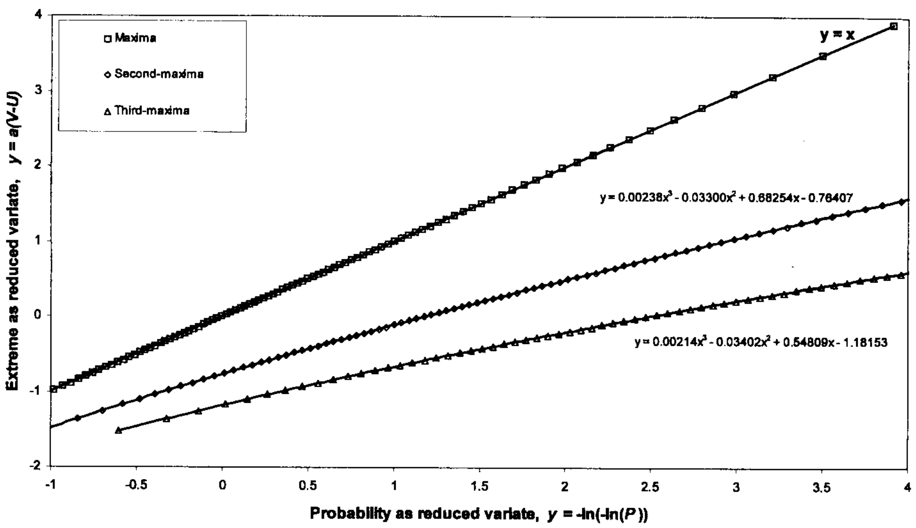

Fig. 4 Gumbel asymptotes for 1st, 2nd and 3rd highest extremes

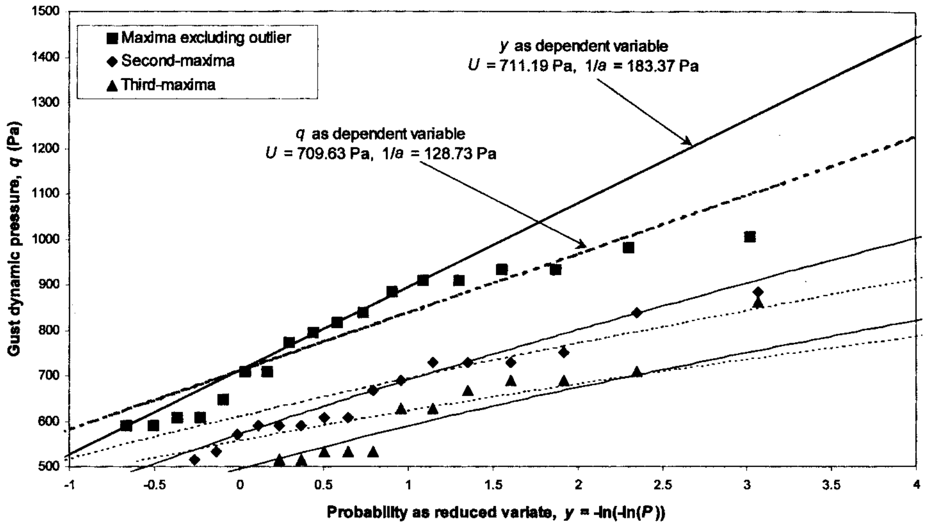

Fig. 5 Gumbel plot with 1st, 2nd and 3rd highest annual maxima for Jersey

(8) and (9) fitted by the method of least squares to all three curves simultaneously. This involved fitting first to the extreme to obtain initial estimates of mode and dispersion, calculating the residuals from each of the curves, then iterating the values of mode and dispersion to give the least mean square of all residuals. This was done for both dynamic pressure and probability as the dependent variable. It is immediately apparent from the larger difference in slope of the two fitted lines to the CDF of the highest extreme that the quality of fit is worse than using the highest extremes alone in Fig. 1.

5. Discussion

5.1. Preamble

A summary of the results for Jersey by each of the illustrated methods is given in Table 2. The reader was promised an objective measure of quality, and this is provided by the regression coefficient R^2. This coefficient gives a direct measure of the spread of the predictions obtained from analysing the data with either the variate or the probability as the dependent variable. On this basis, the "best" result would be obtained using the CDF all the correlated exceedances, but as this requires pre-knowledge of the rate of independent observations, r_i, the method is impractical for analysis. This leaves the MJ&F method, using all independent extremes excluding the outlier, as the best method – followed closely by the same method using 1st, 2nd and 3rd annual maxima. The Gumbel method using 1st, 2nd and 3rd highest extremes appears worse than the standard Gumbel method. To understand this, we need to consider the role of variance errors in the analyses – but first we need to be sure about the validity of the methodology.

5.2. Statistical independence and physical correlation

Some metorologists, most recently Wieringa (1996), doubt the statistical independence of extreme separated by only two days. Wieringa states:

Table 2 Fitted parameters for the Jersey data

Analysis	Figure	Dependent parameter	Mode (Pa)	Dispersion (Pa)	q_{50} (Pa)	R^2
Gumbel, all annual extremes	1	q y	695.7 690.5	186.4 204.1	1423.0 1486.9	0.913
Gumbel, annual extremes excluding outlier	1	q y	699.2 693.3	132.7 149.3	1217.0 1275.7	0.896
Gumbel, annual extremes correcting outlier	1	q y	707.9 700.2	128.4 147.1	1209.0 1274.0	0.879
Modified J&F, excluding outlier	2	q y	739.0 738.3	114.0 118.8	1183.7 1202.0	0.959
Modified J&F, 1st, 2nd & 3rd extremes, excluding outlier	2	q y	738.9 735.9	113.7 121.4	1182.6 1209.4	0.950
All exceedances, excluding outlier, $r_i = 5.25$	3	q y	738.9 738.0	107.1 110.1	1157.0 1167.6	0.973
Gumbel, 1st, 2nd & 3rd excluding outlier	5	q y	709.6 711.2	128.7 183.4	1211.9 1426.7	0.702

"Strong depressions come in pairs, strung like beads on the jetstream, and in Western Europe there is a significant chance of a storm being followed by a similar storm two days later."

Here Wieringa is confusing physical correlation with statistical independence. While a lack of physical correlation is *sufficient* ensure statistical independence, it is not *necessary*. If it were, the very foundations of reliability-based design laid by Davenport, on which many of the world's codes of practice depend, would be seriously undermined. Using spectral methods, Davenport (1967) determined the average annual population of statistically independent wind events in the UK to be $r_i \approx 300$ which, given the average rate of storms, implies 2 to 3 independent events per storm. The same conclusion can be obtained another way

Consider the parent CDF of every hour of wind – this has a Wiebull form and a population $r = 8766$ per year, but is highly correlated from hour to hour. Now consider a process by which values are sampled at intervals sufficiently separated to be statistically independent – this will have exactly the same as the parent CDF, but will have a smaller population and take longer to form because the rate of data is much slower. This parent CFD is related to the CDF of annual extremes by Eq. (5), which is a fundamental statistical relationship. If both the CDF of the parent and of the annual extremes are independently measured, Eq. (5) can be solved for the annual rate of independent events. This, not surprisingly, turns out to be about $r_i \approx 300$.

Raising the parent CDF, which is highly correlated, by the power of the annual rate of independent events to obtain the CDF of annual extremes is a well-established procedure for combining meteorological and wind tunnel data in reliability-based design[1]. What matters is that a two-day separation, used here, ensures statistical independence, even when the meteorological events are physically correlated. Harris (1996) reaches the same conclusion by considering the maximum correlation time, noting

"for correlated data, the original Gumbel form is still valid, but the value of r_i remains to be

[1]This is the "parent-parent" approach favoured by Davenport at the University of Western Ontario, among others. The writer favours the "extreme-extreme" approach – the "Cook-Mayne method" (Cook and Mayne 1979, 1980) – but that is a different point.

determined."

Although physical correlation does not affect the value of statistically independent extremes, it does affect their recurrence. Recurrence of uncorrelated independent events obey the binomial distribution (Cook 1985) so that, given the incidence of the strongest of 100 storms in a year, the probability of the second-strongest occurring consecutively (immediately before or after) is 2/99. However, with the physical correlation described by Wieringa[2] (1996) – "like beads on the jetstream" – this likelihood is higher, and strong winds tend to occur in clumps with intervening calm periods. Significant damage has been observed in consecutive storms in the UK, notably in 1990 and in January 3 to 5 1998, but at different locations due to the differing storm tracks. The correlation at fixed anemograph stations is less than the correlation between storms. Examining the Jersey data in Table 1 reveals that, discounting the extremes in consecutive days which may occur in the same storm, all the rejected extremes separated by two days are less than the value of the mode.

5.3. Role of variance errors in the standard Gumbel analysis

Variance errors in the analysis come from several sources. There may be errors in the data – instrumentation errors, variation is site exposure by direction, etc. There may be departures from

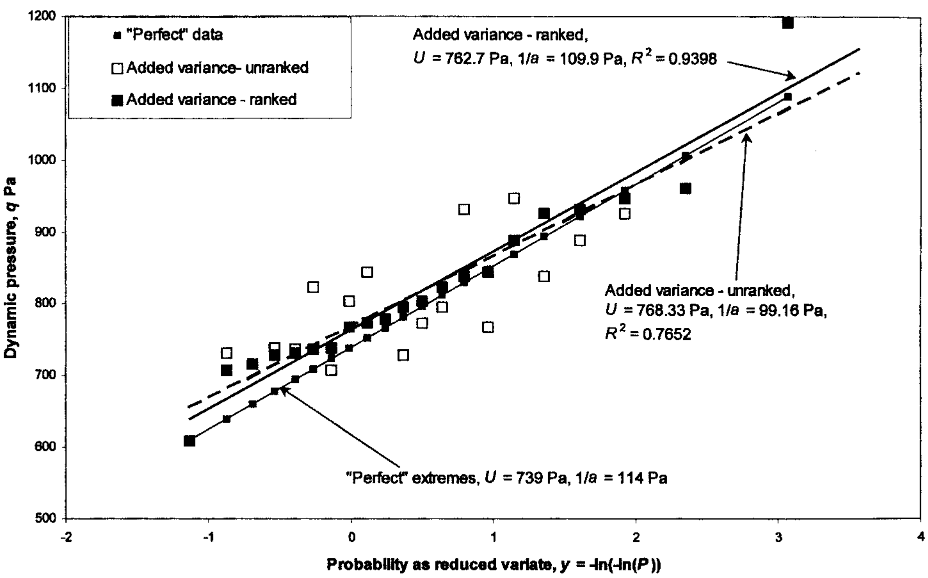

Fig. 6 Effect of data variance on Gumbel plot

[2]Wieringa also cites the tendency for analyses of monthly data (and, by implication, sectorial data) to predict extremes greater that predicated for all year (or all directions). This effect has been noted many times before. It is caused in the standard Gumbel method because there are different populations month-to-month compared with all year (or sector-to-sector compared with all directions). But this effect is automatically corrected by the observed annual rates, r, in Eq. 5. The "Method of independent storms" has been applied to sectorial and monthly analyses of 50 anemograph stations across the UK (Cook and Prior 1987). The monthly predictions of the 50-year return wind speeds for some stations exceeded the all year value by a only few percent in the winter months, while the sectorial predictions never exceeded the all-direction value.

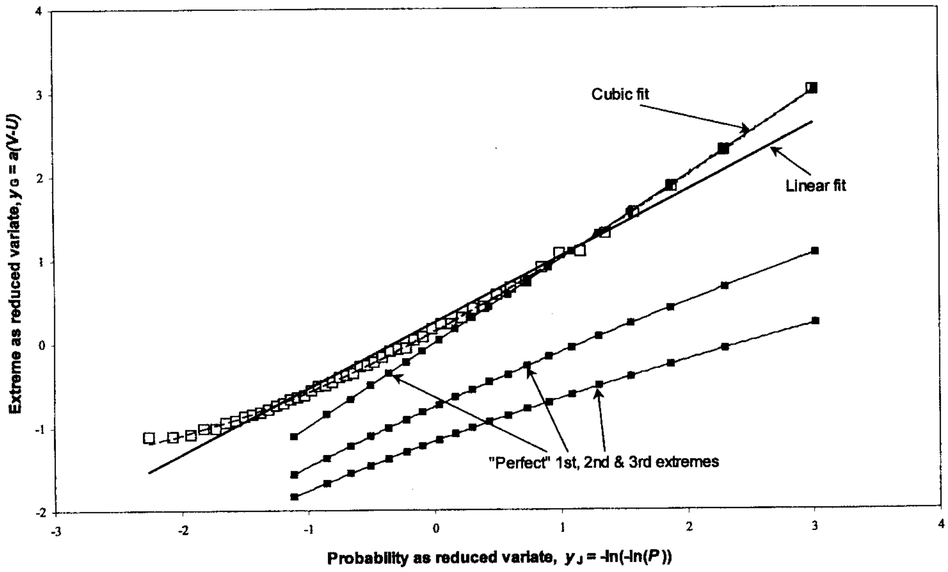

Fig. 7 Calibration of modified Jensen and Franck method for $N = 20$ years, $M = 3$

the statistical ideal – short record lengths, year-to-year variation in windiness, etc. The effect of variance errors in the standard Gumbel analysis may be demonstrated by adding random errors to a "perfect" set of data (i.e., Gumbel's asymptotic values), as follows

1. A set of "perfect" annual maximum dynamic pressures for $N = 20$ years, was generated for $U = 739$ Pa, $1/a = 114$ Pa by Eqs. (2) and (3), and plotted in Fig. 6.
2. Substantial errors were generated randomly and added to the extremes. This added variance appears as random "scatter" of the points indicated by the □ symbols in Fig. 6. If fits are made at this stage, the scatter reduces the regression coefficient to $R^2 = 0.7642$. However, this form of scatter can never appear in a Gumbel plot because of the ranking process that determines the order statistics.
3. When re-ranked, these data give the typical Gumbel "staircase" as shown by the ■ symbols. The ranking re-imposes order on random errors, with "high" values tending to be ranked further to the right and "low" values further to the left. This leads us to expect higher slopes and hence higher values of dispersion. The order re-imposed by the ranking raises the regression coefficient to $R^2 = 0.9398$. It is tempting to view this improvement as spurious, but it is simply the effect of including the random "scatter" as a new component of randomness into the process to be analysed.

Gumbel (1958) shows that the analysis variance is inversely proportional to the population or record length, and that the fitted line is more likely to be steeper than the "perfect" line (63%) than not (37%). Of course, this is not necessarily true in any one individual case, and Fig. 6 is one of the 37% of exceptions. Nevertheless we would expect, on average, higher estimates of dispersion from shorter data records, leading to higher predictions of extremes. In comparing MIS results with standard Gumbel results, Cook and Prior (1987, Fig. 7) demonstrate this to be true more often than expected, indicating an additional systematic bias inherent in the MJ&F and MIS methodology.

5.4. Bias in the modified Jensen and Franck method and the "Method of independent storms"

The close equivalence between the MJ&F method using 1st, 2nd and 3rd highest extremes, the MJ&F method using all independent extremes over threshold and MIS may be exploited to determine the bias in these methodologies with respect to the standard Gumbel method, as follows

1. For any given period of N years ...
2. Generate a set of "perfect" 1st, 2nd and 3rd highest extremes from the Gumbel asymptotes using Eq. (2) and the cubic polynomial fits to Eqs. (8) and (9) in Fig. 4.
3. Set the threshold equal to the smallest annual maximum.
4. Discard the extremes smaller than the threshold.
5. Determine the population, n, and average annual rate, r, of the surviving extremes.
6. Aggregate and rank the extremes, then perform the MJ&F analysis.
7. Compare the MJ&F results with the original "perfect" annual extremes.

This process is illustrated in Fig. 7 for $N = 20$ years in terms of the Gumbel reduced variate, y, where the MJ&F data can be seen to lie on a curve that converges onto Eq. (2) in the upper tail. Fits to linear and cubic polynomials are shown on the figure. If we use the subscripts "G" and "J" to denote Gumbel and equivalent MJ&F parameters, the two fitted equations become

$$y_G = 0.7855y_J + 0.2449 \qquad (N = 20, \textit{linear}) \quad (10)$$

and

$$y_G = -0.0093y_J^3 + 0.0767y_J^2 + 0.808y_J + 0.1378 \qquad (N = 20, \textit{cubic}) \quad (11)$$

Only the linear equation can be used to post-calibrate previously fitted MJ&F parameters to equivalent Gumbel values

$$a_G = 0.7855a_J \qquad (N = 20) \quad (12)$$

and

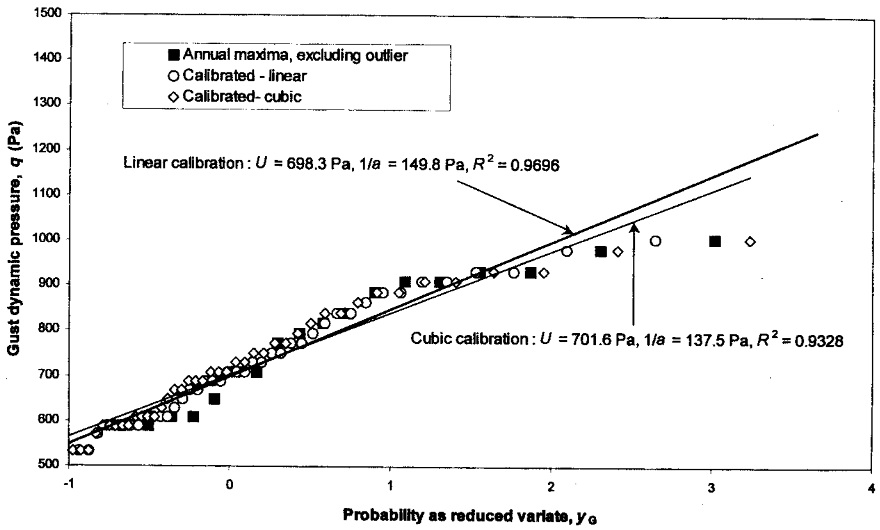

Fig. 8 Calibrated MJ&F method for Jersey

Table 3 Effect of MJ&F calibration on estimates of q_{50} for Jersey

MJ&F calibration	Mode (Pa)	Dispersion (Pa)	q_{50} (Pa)	R^2	Effect on q_{50} compared with MJ&F	Standard Gumbel dependent parameter q	P
Post – linear	711.1	145.1	1277.4	0.950	+7.9%	+5.0%	+0.1%
Inter – linear	698.3	149.8	1282.8	0.970	+8.4%	+5.4%	+0.6%
Inter – cubic	701.6	137.5	1238.2	0.933	+4.6%	+1.7%	– 2.9%

$$U_G = U_J - \frac{0.2449}{a_G} \qquad (N = 20) \quad (13)$$

On the other hand, either Eq. (10) or (11) can be used as an intermediate transformation between the estimate for y_J from Eqs. (4), (5) and (2) and y_G in Eq. (3) for the FT1 parameters to inter-calibrate the MJ&F estimates of probability during the analysis. This is illustrated in Fig. 8, where the ordinate of the Gumbel plot is now y_G from Eq. (12) or (13). Results of post- and inter-calibrating the MJ&F method and the effect of the calibrations on the estimates for q_{50} for Jersey are given in Table 3.

Although this calibration is only valid for $N = 20$ and $M = 3$, it represents a typical calibration of the MJ&F method. A full calibration would require higher extremes, up to $M = 7$ for the threshold used here, and calculations over a wider range of N. Further example calibrations show that the coefficients in Eqs. (10) and (11), and the resulting bias, depend only weakly on N over a wide range $5 \leq N \leq 200$. Insensitivity to the annual rate, r, has already been noted. It follows that the bias for any given period, N, and order of extreme, M, is consistent in terms of the reduced variate, y, and will therefore be inversely proportional to the mode/dispersion ratio, $\Pi = aU$. This means that the MJ&F method has a bias of -9% on q_{50} and -5% on V_{50}, compared with the standard Gumbel method, irrespective of whether the analysis variate is dynamic pressure ($\Pi \approx 5$) or wind speed ($\Pi \approx 10$). Applying the expected bias to the data of Cook and Prior (1987, Fig. 7), restores the expected 63:37 proportion of exceedances.

This bias represents the difference between equally valid methodologies for estimating the same parameters. Each method seeks to derive estimates of the Gumbel (FT1) asymptotic parameters, mode U and dispersion $1/a$. The differences come from the relative efficiency of the order statistics in estimating the probability, P, of each observation. The regression coefficient, R^2, gives an objective measure of goodness of fit to the asymptotic distribution. On this basis, the MJ&F method performs "better" than the standard Gumbel analysis. If the calibration is used to "correct" the analyses, this indicates that Gumbel should be corrected to MJ&F. However, since wind engineering design has been calibrated by historical usage against the Gumbel methodology and always errs on the side of conservatism, it may be more appropriate to correct MJ&F to match Gumbel.

5.5. Modified Jensen and Franck method compared with Gumbel method with M-th highest extremes

The regression coefficient indicates the Gumbel method with M-th highest extremes is worse than the standard Gumbel method, and much worse than all variations of MJ&F – even though these use the same 1st, 2nd and 3rd highest extremes. As the data values are the same, the dif-

Table 4 Sensitivity of methodologies to added variance

Method	R^2	Mode (Pa)	Dispersion (Pa)	q_{50} (Pa)	Change of q_{50}
Gumbel, $M=3$	0.8858	761.5	122.2	1238.3	4.60%
Standard Gumbel	0.9398	762.7	109.9	1191.5	0.65%
MJ&F	0.9865	782.5	86.15	1118.7	−5.50%
Inter-linear MJ&F	0.9865	755.7	109.7	1183.7	−0.01%
Perfect	1	739	114	1183.8	0%

Table 5 Exceedances of 29 m/s threshold expected and observed at Jersey

Year rank	21	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	Total
Expected	6	5	5	4	4	4	3	3	3	3	3	3	2	2	2	2	1	1	1	1	1	59
Observed	6	3	12	2	6	9	6	3	6	4	9	1	2	3	5	2	2	1	2	0	0	84

ference must come from the corresponding estimates of probability, the aggregation of data and the ranking process. Herein lies the key to the reliability and robustness of the MJ&F method and, by implication, also MIS.

The underlying theory of extremes does not actually require abstraction from separate equal periods of time, this is done for convenience and to enclose seasonal trends as noted earlier. Extreme-value theory is based on extracting the extreme from a population of values *selected at random* from the whole population. The aggregation process comes closer to this ideal than the use of fixed year periods, particularly when real data contains year-to-year variations in windiness. Although the original intention of MJ&F and MIS was to increase the population of data, the methods benefit more from the aggregation process, since it improves the estimates of probability as demonstrated in §5.3. With real data it also encloses year-to-year variations in the same way that taking whole-year periods encloses seasonal variations.

On the other hand, segregation of the M-th highest extremes into separate curves has the opposite effect to aggregation. The year of the highest extreme will also have high 2nd and 3rd extremes, while the year of the lowest extreme will also have low 2nd and 3rd extremes, even when the wind climate is constant, so enhancing the variance of the M-th highest extremes compared with the annual extremes. The sensitivity to variance errors can be demonstrated with a similar procedure to §5.3, by adding random errors to a set of "perfect" M-th highest extremes,[3] then comparing the analysis predictions to the original distribution. The effects of the same random errors as in §5.3 on the different methods are summarised in Table 4, listed in order of increasing regression coefficient, i.e. from "worst" to "best." As expected, the Gumbel method with M-th extremes gives a lower regression coefficient than the standard method, and ranks "worst." The linearly calibrated MJ&F method gives the highest regression coefficient combined with the closest estimate of q_{50} to the "perfect" value.

Year-to-year variations in windiness would be expected to increase this effect further. The substantial year-to-year variations in the Jersey data are illustrated by Table 5, which lists the number of exceedances of the threshold observed and expected for the rank, m, of each year. There are significantly more exceedances than expected and many occurred in middle-ranking years. As expected, the Gumbel method with M-th highest extremes gives a substantially lower regression coefficient for the Jersey observations ($R^2 = 0.702$).

[3]Note that the added random errors may change the order of the M-th extremes for each simulated year.

5.6. Acceptable ranges of parameters in the modified Jensen and Franck method and "Method of independent storms"

5.6.1. Annual rate, r

The annual rate of extremes, r, for peak-over-threshold data is set by the threshold and falls as the threshold is raised. The threshold must be set high enough that the rate of independent maxima, $r_i \approx 300$, exceeds r by an acceptable margin. In the UK, MIS gives $r \approx 100-150$, while the threshold of 29 m/s gives $r \approx 4$. But both give closely similar results because the Gumbel plot imposes a further threshold on the ranked extremes, corresponding to the lower limit of reduced variate, y. In the examples, this second threshold was set equal to the smallest annual maximum, equivalent to $y \approx -1$, so that the annual rate reduces further to $r \approx 3$. Harris (1996) recommended a minimum initial value of $r = 10$ based on the observation of Cook (1982), that changing from $r = 100$ to $r = 10$ makes less than 1% difference to predictions of 50-year return period values.

The minimum possible value would be $r = 1$, which would equate MJ&F directly to the standard Gumbel method, but this is not valid. For a period of $N = 20$ years, the population $n = 20$ in the MJ&F method would be expected to replace the seven lowest annual maxima with five 2nd and two 3rd highest maxima. This considerably increases the bias between the methods. The meteorological standard threshold of 29 m/s (56 kts, Beaufort Range 11) used for recording the maximum gusts, is expected in the UK to include all the annual maxima[4], so represents an acceptable minimum of $r = 4$.

5.6.2. Order of sub-annual extreme, M

In peak-over-threshold data, this parameter is linked to the annual rate, r, and the acceptable range is automatically met when r is within acceptable range. For $r = 4$, sub-annual extremes up to $M = 7$ are expected in the year of the highest annual maximum, but these higher orders will rank near or below the mode. For this reason, inclusion of sub-annual extremes up to order $M = 3$ was demonstrated in §5.4 to give results within 1% of the full set.

5.6.3. Observation period, N

Lieblein (1974) recommends $N = 10$ as minimum and $N = 16$ as optimal for analysis accuracy. Cook and Prior (1987, Fig. 7) demonstrate that MIS with $N = 11$ is comparable with the standard Gumbel method with $N = 50$, and part of this improvement is attributable to aggregating the year-to-year variations. The requirement for the observations to be representative of the long-term wind climate, not the analysis accuracy, determines the minimum observation period. In practice, the observation period is set by whatever record length is available, and the standard Gumbel analysis is frequently performed for periods as short as 10 years.

When the record length is shorter, Wieringa (1996) recommends aggregating data from adjacent stations with overlapping records to give "regional," rather than "station" estimates. However these stations, although adjacent, are likely to be differently exposed and it will be necessary to correct for this by correlating the data over the periods of overlap. Miller and Cook (1997) demonstrated a methodology for obtaining absolute exposure calibrations of anemograph

[4]But excludes two out of 21 annual maxima in the Jersey data.

stations. In some supposedly consistent long records they found significant changes in anemograph exposure calibration due to changes of the anemometer type, location or height, or due to changes in the ground roughness from development of the surrounding area. Effects of changes to anemometers at known times can be corrected in the data record, but gradual development is harder to address. The effects of exposure can be very significant (Miller and Cook 1997) and may be larger than the meteorological trends under investigation. For this reason, it is almost impossible to investigate the effect of climate change on extreme winds on the basis of individual station data.

6. Conclusions

1) Extreme value analysis should be performed on dynamic pressure, in preference to wind speed, because this parameter converges faster to the Gumbel (FT1) distribution.
2) The modified Jensen and Franck (MJ&F) method and the "Method of independent storms" (MIS) are significantly better than the standard Gumbel method because, by aggregating the data record, the methods reduce analysis variance and the effects of year-to-year variations of windiness.
3) Bias in the MJ&F method and MIS gives estimates of q_{50} typically 9% less than the standard Gumbel method. This bias may be removed by calibration during or after analysis.
4) The MJ&F method and the MIS require the annual rate of observations used, r, to be less than the annual rate of statistically independent observations, r_i. (Both r and r_I vary with the threshold used to select maxima.)
5) Extremes abstracted from physical correlated variables are statistically independent if separated by a period longer than the correlation time, longer than two days in the UK.
6) The annual rate of statistically independent observations, r_i, may be determined by iteration of Eq. (5) using independent estimates of the CDF of the parent, P_S, and the CDF of the extremes, P.
7) The minimum value of $r=4$ in the MJ&F method ensures that all extremes larger than the smallest annual extreme contribute to the analysis.
8) The acceptance threshold for MJ&F and MIS should be set to give an annual rate of extremes in the range $4 \leq r \leq 300$, but in practice little additional advantage is gained when $r>10$.
9) Extremes that lie outside the Gumbel confidence limits may do so because they:
   a) represent a return period longer than the observation period, or
   b) are members of a different population, or
   c) are erroneous.
   In all cases they should be rejected from the analysis, and the analysis repeated.
10) When abstracting annual maxima from peak-over-threshold data, count "missing" values from years when the threshold is not exceeded as being members of the population with the lowest ranks, in order to derive unbiased plotting positions, but exclude them from the fit.
11) Longer records are more likely to give lower estimates of dispersion and q_{50} owing to reduced analysis variance.
12) Changes in anemometer type, location or height often occur during supposedly consistent long records. It is advisable to request the anemograph siting, maintenance and inspection records before selecting long records of wind data for extreme value analysis.

References

BSI (1997a), *British Standard, Loading for Buildings, Part 2: Code of practice for wind loads, BS6399,*

Part 2, 1997. British Standards Institution, London.
BSI (1997b), *Eurocode 1: Basis of design and actions on structures, Part 2.4 Actions on structures – Wind actions, DD ENV 19910204:1997*. British Standards Institution, London.
Castillo, E. (1988), *Extreme Value Theory in Engineering*. Academic Press, New York.
Cook, N.J. (1982), "Towards better estimation of extreme winds," *Journal of Wind Engineering & Industrial Aerodynamics*, **9**, 295-323.
Cook, N.J. (1983), "Note on directional and seasonal assessment of extreme winds for design," *Journal of Wind Engineering & Industrial Aerodynamics*, **12**, 365-372.
Cook, N.J. (1985), *Designer's Guide to Wind Loading of Building Structures*, Part 1. Butterworth, London.
Cook, N.J. and Mayne, J.R. (1979), "A novel working approach to the assessment of wind loads for equivalent static design," *Journal of Wind Engineering & Industrial Aerodynamics*, **4**, 149-164.
Cook, N.J. and Mayne, J.R. (1980), "A refined working approach to the assessment of wind loads for equivalent static design," *Journal of Wind Engineering & Industrial Aerodynamics*, **5**, 125-137.
Cook, N.J. and Prior, M.J. (1987), "Extreme wind climate of the United Kingdom," *Journal of Wind Engineering & Industrial Aerodynamics*, **26**, 371-389.
Fisher, R.A. and Tippett L.H.C. (1928), "Limiting forms of the frequency distribution of the largest or smallest member of a sample." *Proc. Cambridge Philosophical Society*, **24**, 180-190.
Gomes, L. and Vickery, B.J. (1978), "Extreme wind speeds in mixed climates," *Journal of Industrial Aerodynamics*, **2**, 331-344.
Gumbel, E.J. (1958), *Statistics of Extremes*, Columbia University Press, New York.
Harris, R.I. (1966), "Gumbel re-visited – a new look at extreme values statistics applied to wind speeds," *Journal of Wind Engineering & Industrial Aerodynamics*, **59**, 1-22.
Harris, R.I. (1998), "Improvements to the method of independent storms," *Journal of Wind Engineering & Industrial Aerodynamics*, (in preparation).
Lieblein J. (1974), "Efficient methods of extreme-value methodology," Report NBSIR 74-602, National Bureau of Standards, Washington.
Jensen, M and Franck, N. (1970), *The Climate of Strong Winds in Denmark*, Danish Technical Press, Copenhagen.
Miller, C.A. and Cook, N.J. (1977), "Calibration of the exposure of UK anemometers", *Proc. 2nd European & African Conference on Wind Engineering*, Genoa, June 22-26, 171-178.
Shellard, H.C. (1963), "The estimation of design wind speed", *Wind Effects on Buildings & Structures, Teddington, 1963*, HMSO, London.
Wieringa, J. (1996), "Representativity of extreme wind data", *Hydrology of Disasters* (Ed. V P Singh), Kluwer, Amsterdam, 19-40.

Notations

a	= reciprocal of dispersion of extremes
M	= order of sub-annual extreme from annual maximum
m	= rank of extreme from lowest, $1 \leq m \leq n$
n	= population of extremes in set
N	= number of years in record
P	= probability cumulative distribution function (CDF), $0 \leq P \leq 1$
q	= dynamic pressure (Pa)
R^2	= regression coefficient of least-mean-squares fit
r	= average annual rate of extremes in set, $r = n/N$
r_i	= annual rate of independent extremes
U	= mode of extremes
V	= variate value
y	= reduced variate, $y = a(V-U)$
Π	= characteristic product, $\Pi = aU$