LONDON, METEOROLOGICAL OFFICE.

Met.O.3 Technical Note No.18.

Comments on some assumptions made in extreme value analysis. By GRAHAM, A.E.

London, Met.Off., Met.O.3 Tech.Note No.18, [1982], 31cm. Pp.[i]+12, 12 pls.+App.[10]. 4 Refs.

An unofficial document - not to be quoted in print.

MET O 3 TECHNICAL NOTE 18

COMMENTS ON SOME ASSUMPTIONS MADE IN EXTREME VALUE ANALYSIS

Anne E Graham

Contents

1. Five year extreme predicted by the Weibull Method

When considering the estimation of extremes it at first seems reasonable to assume that a better estimate may be made of the once in 5 year extreme value than the once in 50 year from a record containing, say, 20 years of data. However, the method used in estimating the extremes must also be considered. All of the data are used for the fit to a Weibull distribution which is defined by 3 parameters. These 3 parameters are then used in an expression to estimate all extremes regardless of return period, which is the only variable in the expression.

The three parameter Weibull distribution is given by:-

$$P(x) = 1 - \exp\left[-\left(\frac{x-x_o}{c}\right)^m\right]$$

where x , m and c are the three parameters derived from the data.

The expression used to give the extremes is:-

$$x = \exp\left[\frac{R + m\ln c}{m}\right] + x_o$$

Where R is the variable giving the return period in terms of the number of observations which would be expected during the chosen return period.

Some early work on extreme values compared extremes estimated using the Weibull distribution with those derived from an analysis of extremes using the Gumbel (or Fisher-Tippett Type I) distribution, which is generally considered to give much better estimates. This comparison seemed to confirm the assumption that Weibull estimates of extremes at the 5 year level were in good agreement with the 5 year extremes from the Gumbel analysis. It was also noted that for return periods less than 5 years the Weibull estimates of extreme value were overestimated and for return periods greater than 5 years were underestimated when compared with results from the Gumbel analysis.

Most of this earlier work was based on results of wind speed anslyses on data from coastal stations. Work on extremes using other variables, in particular air temperatures indicated that the assumption that Weibull gives the 'correct' extreme at 5 years did not apply. As a result wind speeds from 12 coastal stations were analysed to give extremes using both the Gumbel and Weibull methods of analysis. Figure 1.1 shows the results of this analysis. Extreme wind speeds in knots have been plotted against log return period for ease of prepresentation because this arrangement gives almost a straight line relationship. Table 1.1 lists the points of intersection for each station, ie where both methods estimate the same extreme for the same return period.

None of these stations confirms the assumption that the 5 year extreme given by the Weibull is 'correct', although 5 give 4 years as the point of intersection. The others show widely varying results for the point of intersection and the relationship between the results from Gumbel and Weibull for all of the return periods.

This study can be criticised on the grounds that different periods of data were used in the Gumbel and Weibull analyses. This was because the Gumbel extremes were those derived from the annual maxima of hourly mean values already available as standard Met O 3 results, whilst the data used in the Weibull analyses were hourly-mean anemograph winds which were only available for the period 1970 to 1981. The anemograph winds were used because the averaging time of each observation was known accurately and this is important when using the Weibull method. Gumbel analyses using the years 1970 to 1981 only must be considered unreliable because of the small number (12) of annual maxima involved. The extremes used are based on all of the data for each station and are the best estimates available.

The results clearly show that it cannot be assumed that Weibull will always give a 5 year extreme of comparable size to that given by a Gumbel analysis.

Table 1.1 Return periods at which both the Gumbel and Weibull methods of extreme value analysis estimate the same extreme value.

Station	Return period years
Tiree	4
Scilly	4
Manston	4
Squires Gate	1
Valley	8
Gorleston	550
Fraserburgh	12
Lerwick	2
Stornoway	4
Wick	3
Thorney Island	11
Portland Bill	4

FIGURE 11 a. Coastal Station Extreme Values.

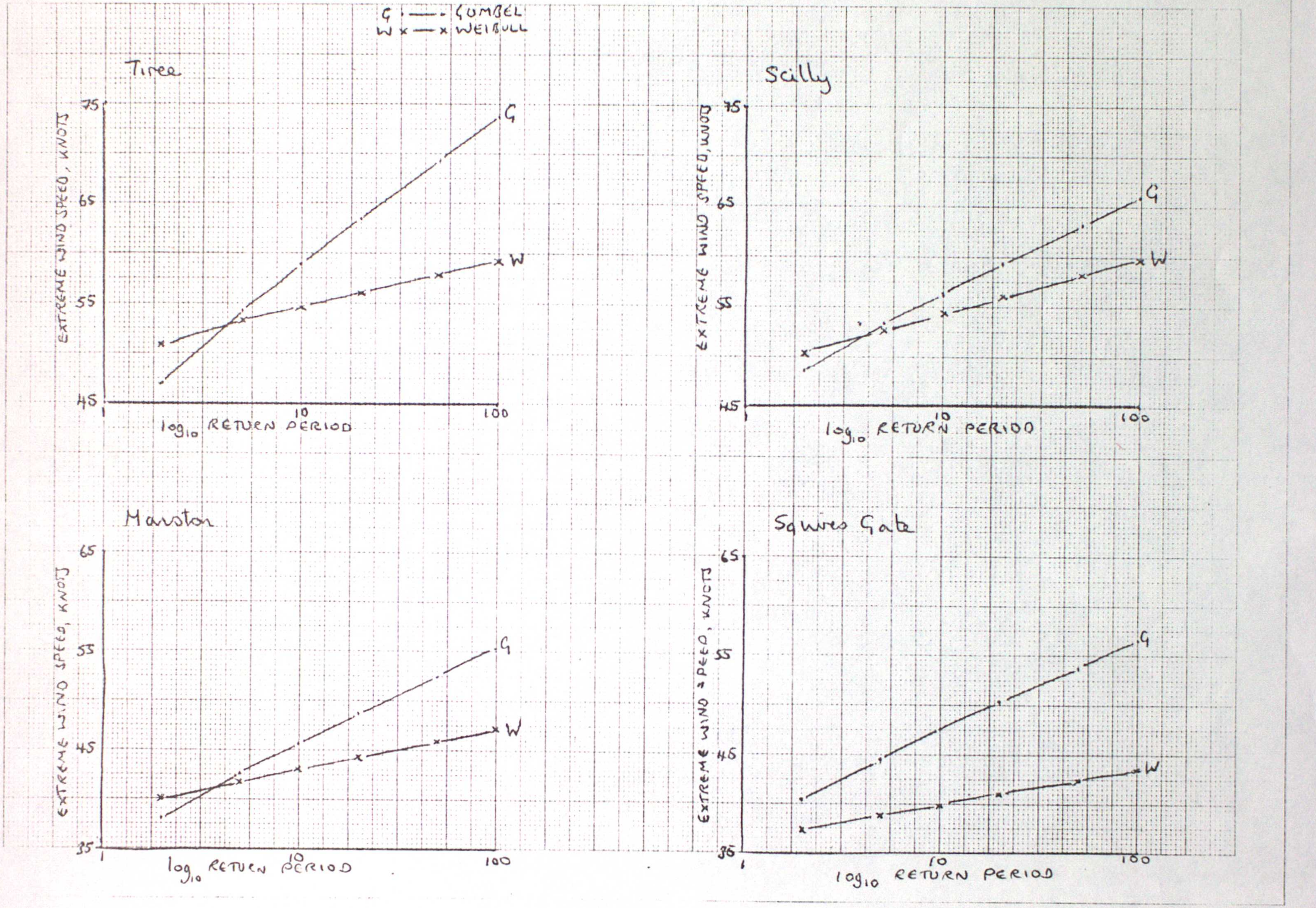


FIGURE 1.1.b Coastal Station Extreme values

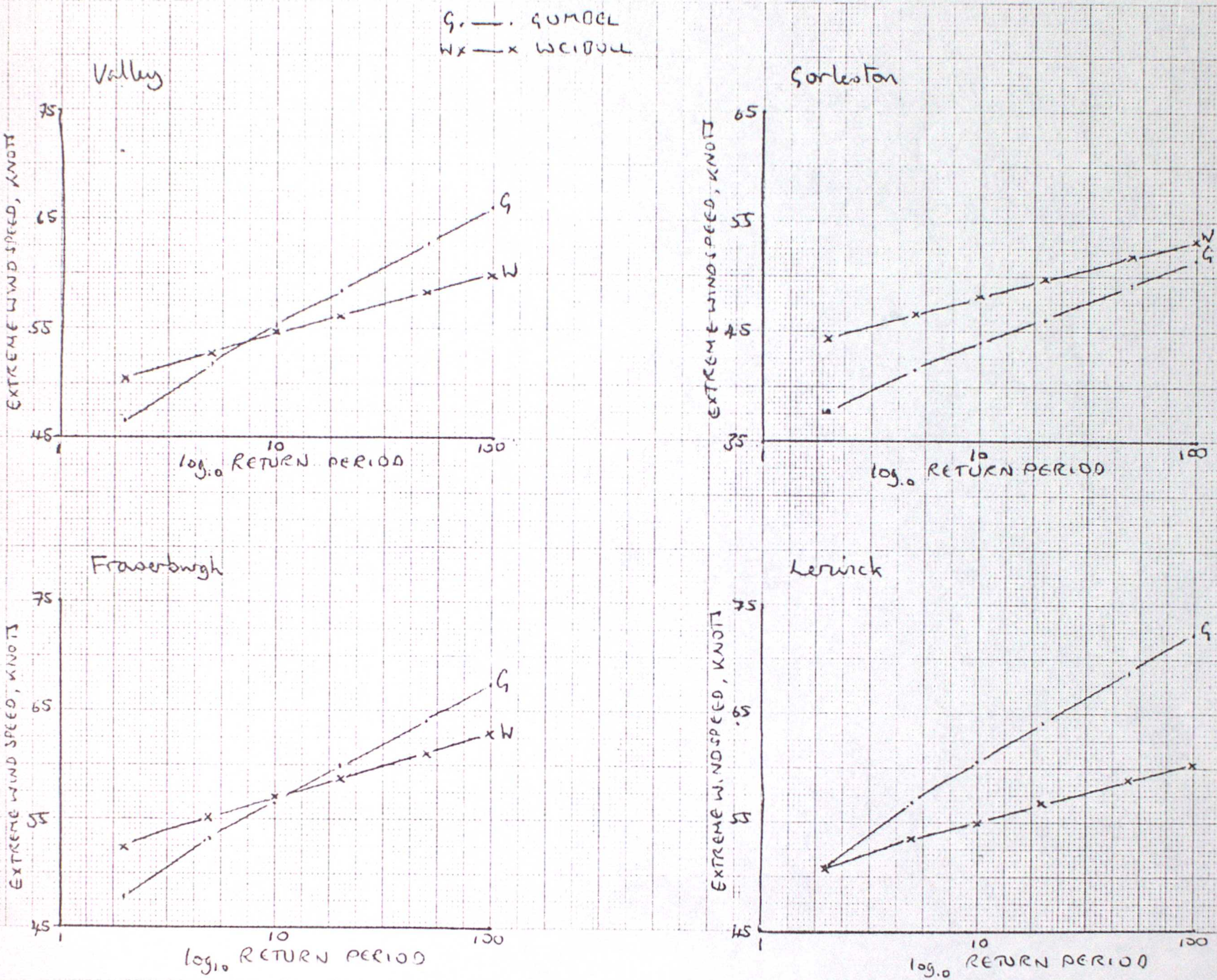
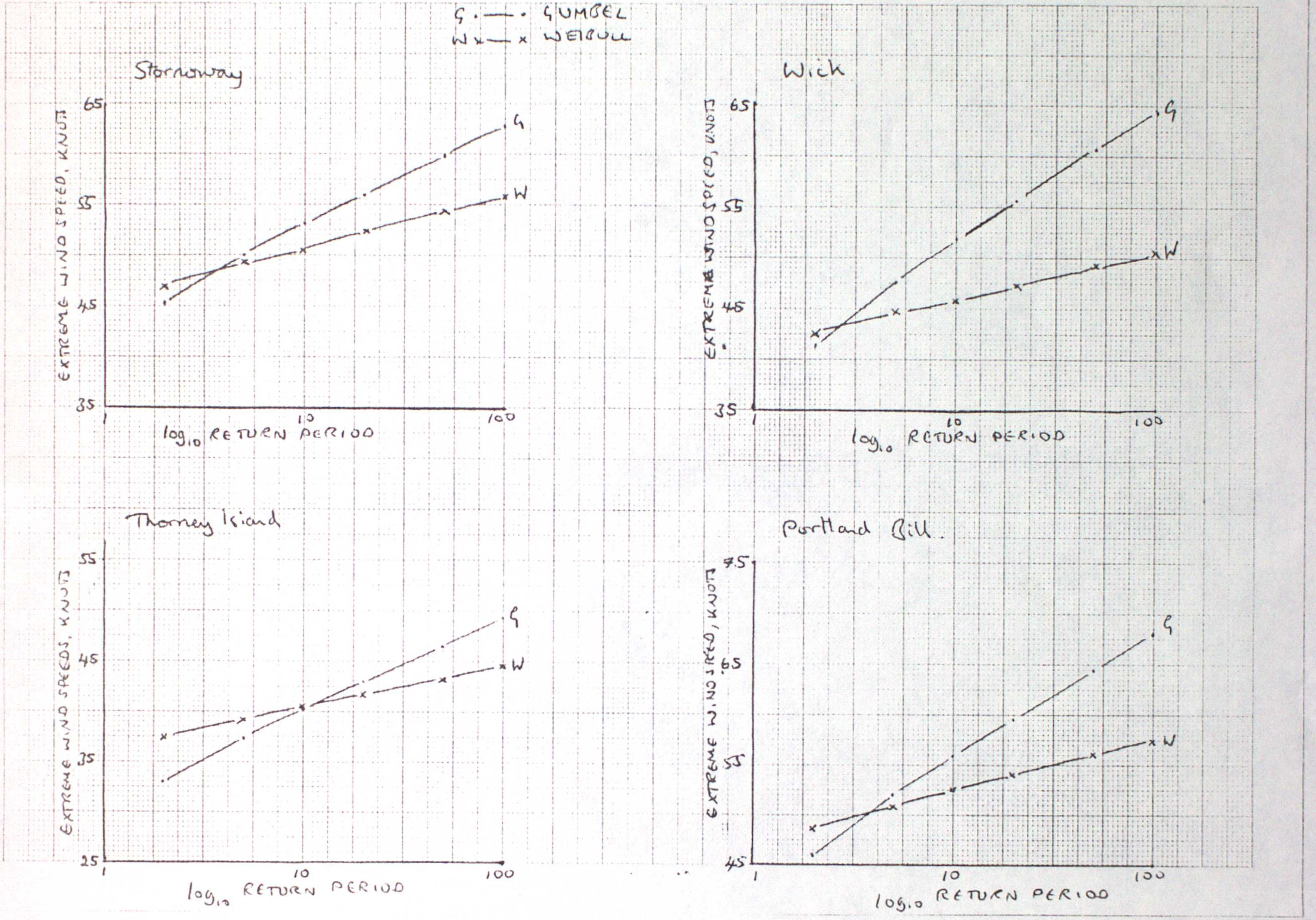


FIGURE 1.1c. Coastal station Extreme Values.



2. Ratios used in directional Weibull analyses

It is often necessary to give estimates of extreme wind speeds from different directions. Standard once in 50 year return period values of hourly mean wind speeds have been calculated for sea areas surrounding the British Isles. These are recorded in the Department of Energy Guidance Notes for reference.

In the past the Weibull method has been used for all directional analyses even when the data would normally be considered more appropriate for the Gumbel method, ie long period (at least 10 years) of regular observations. Work is being carried out to find out whether the Gumbel method can be used for directional analyses. However, in most cases the Gumbel method cannot be used even when an all direction extreme is required because the observations are not evenly distributed in time. This is the case for most observations at sea. These observations are made by the deck officers of merchant ships during the course of their voyage. The data are random in both space and time and add to the problems of directional analyses because it is not known whether the directional distribution of the observations is real or biassed due to poor sampling. The usual practice when estimating directional extremes has been to find the ratio of the once in 50 year extreme in each direction sector with that given by the all direction analysis. These ratios are then used to scale the guidance note value appropriate to the area or station used.

It has been suggested in view of earlier work, discussed in part 1 that a more accurate scaling of the guidance note value would be produced by using the ratio of the extremes having a return period of once in 5 years. Some work was carried out in this way but in view of the results of part 1, ie that the 5 year extreme value estimated by Weibull is not always comparable with the Gumbel 5 year extreme, it was necessary to investigate the magnitude of any errors which may have arisen by using the 5 year ratio as opposed to the 50 year ratio.

The method of scaling the guidance note 50 year extreme hourly mean wind speed value only produces an equivalent directional 50 year extreme and does not provide better estimates of events having different return periods, although this has been attempted by using wind speed growth curves (see Part 3).

Three test areas were chosen in the North Sea. The positions of these areas are given in Table 2.1. The period which the data covered in each case was Jan 1961 to April 1980. The data were divided into 12 directional groups of 30° and, for each area, the following ratios were calculated for each direction sector, using Weibull extremes:

<u>5 year ratio</u>

$$= \frac{\text{5 year directional extreme}}{\text{5 year all direction extreme}}$$

<u>50 year ratio</u>

$$= \frac{\text{50 year directional extreme}}{\text{50 year all direction extreme}}$$

The resulting ratios are shown in Table 2.2. There is very little difference between the ratios derived from the 5 year and 50 year values for each direction at each of the 3 sites. Figure 2.1 shows that, in general, there will be very little difference between the ratio calculated from the 50 year values and the ratios of the other return periods for each direction group.

The ratios derived for these three areas indicate that there will be little difference in the resulting 50 year extreme whether the guidance note value was scaled using the 5 year or 50 year extreme values.

Consequently, the work done using the 5 year ratio for scaling, based on the erroneous assumption that the Weibull method gives a 'correct' estimation of the 5 year extreme, will not be greatly in error. But, equally, there is no advantage in using the 5 year extremes.

Table 2.1 Positions of the three sea areas used

Area	Latitude		Longitude
		degrees	
1	60-62 N		1 -4 E
2	57-59 N		2W-1 E
3	51-53 N		2 -4 E

Table 2.2 Five and fifty year ratios for each area

Sector	Direction	Area 1		Area 2		Area 3	
number	degrees	5 year ratio	50 year ratio	5 year ratio	50 year ratio	5 year ratio	50 year ratio
1	360-10	0.93	0.93	0.97	0.98	0.96	0.96
2	20-40	0.95	0.95	0.90	0.92	0.97	0.98
3	50-70	0.81	0.81	0.77	0.78	0.88	0.87
4	80-100	0.79	0.81	1.01	1.02	0.90	0.90
5	110-130	1.13	1.17	1.15	1.16	0.98	1.01
6	140-160	1.06	1.06	1.09	1.10	0.98	1.00
7	170-190	1.05	1.05	0.94	0.94	0.97	0.98
8	200-220	1.04	1.05	0.96	0.96	0.94	0.93
9	230-250	0.97	0.97	0.99	0.99	0.95	0.94
10	260-280	1.06	1.08	0.95	0.95	1.01	1.00
11	290-310	1.10	1.12	1.06	1.07	0.98	0.98
12	320-340	0.98	0.98	1.06	1.07	1.05	1.05

FIGURE 2·1a EXTREME WINDS, AREA 1

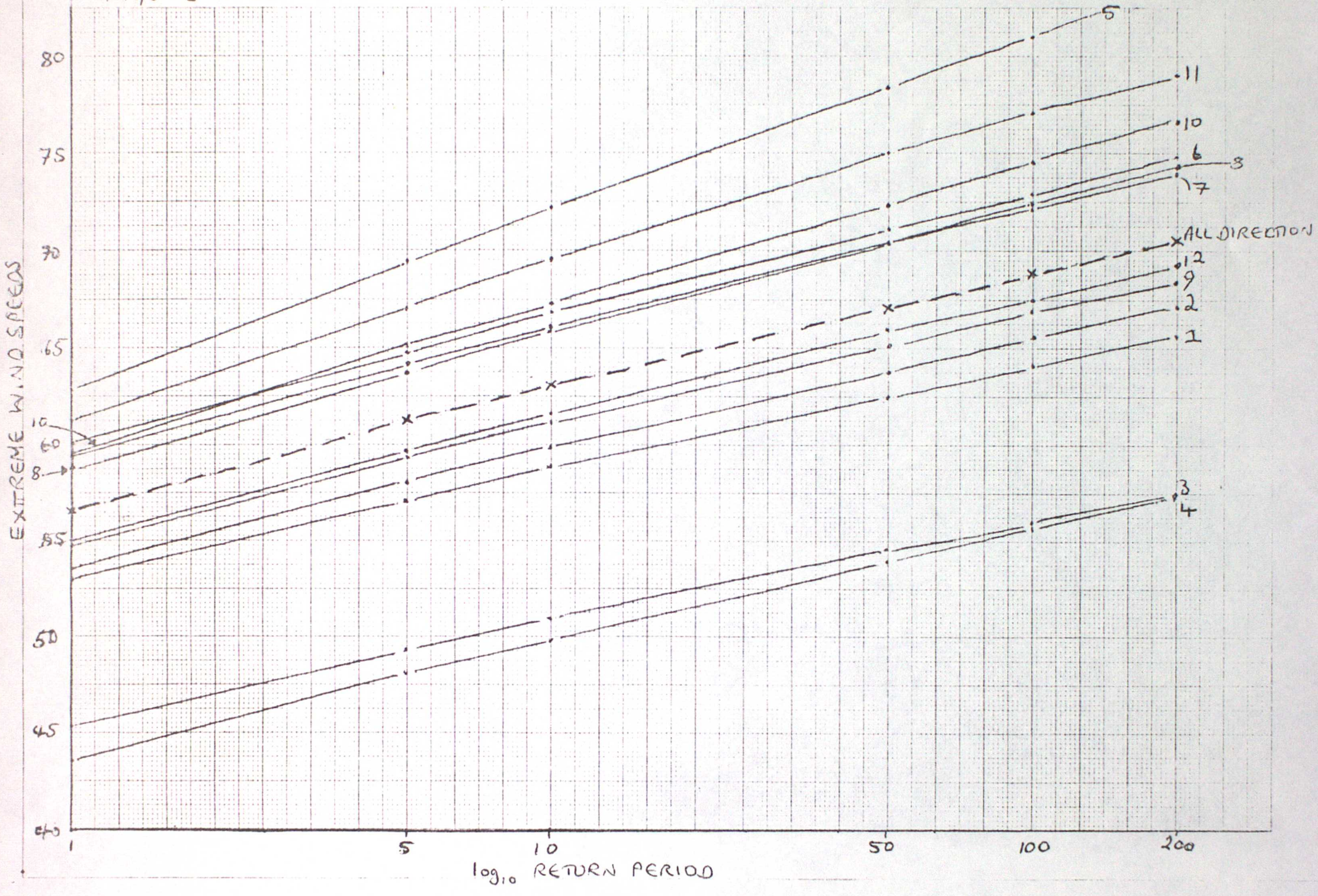


FIGURE 2·1b EXTREME WINDS, AREA 2.

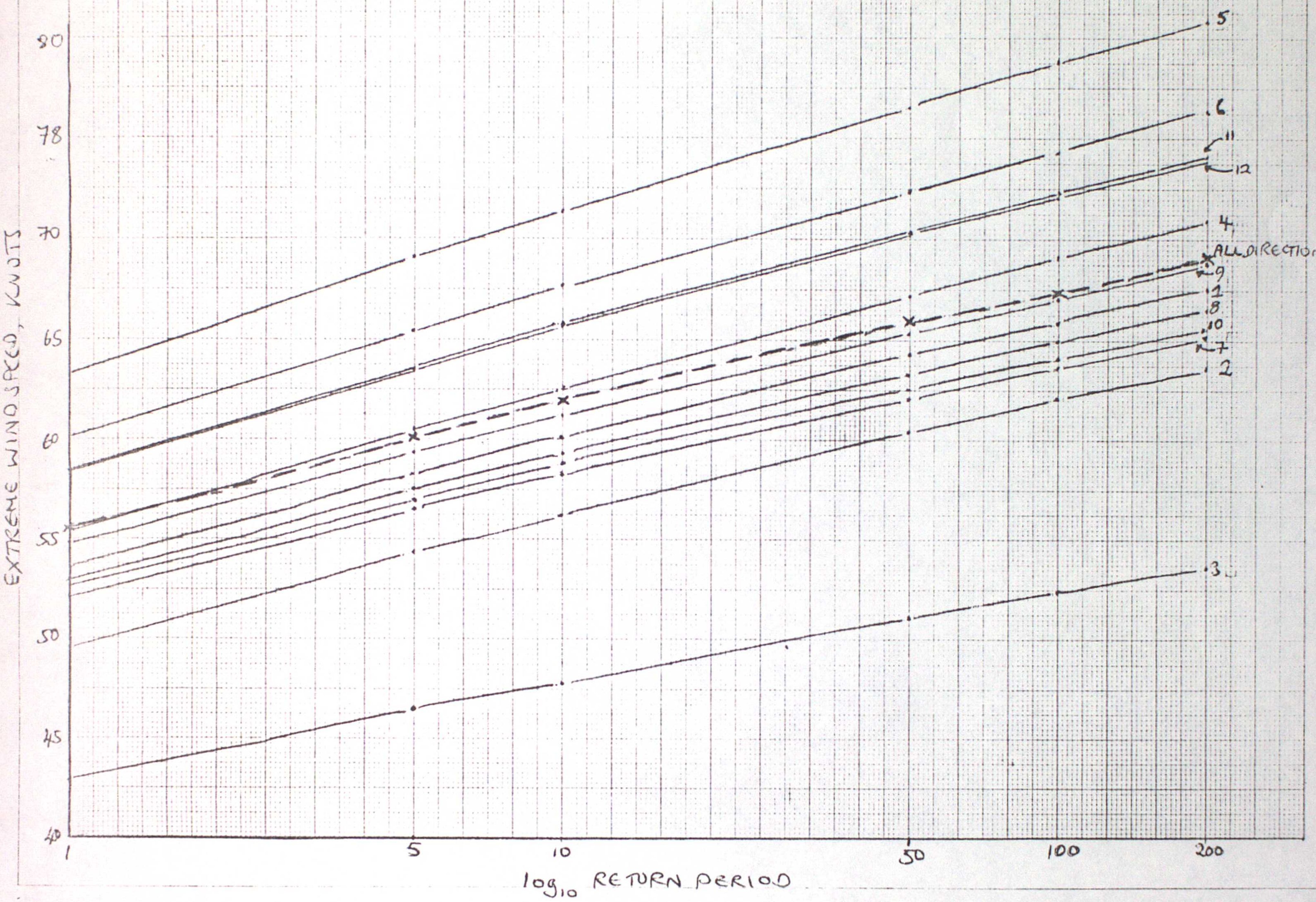
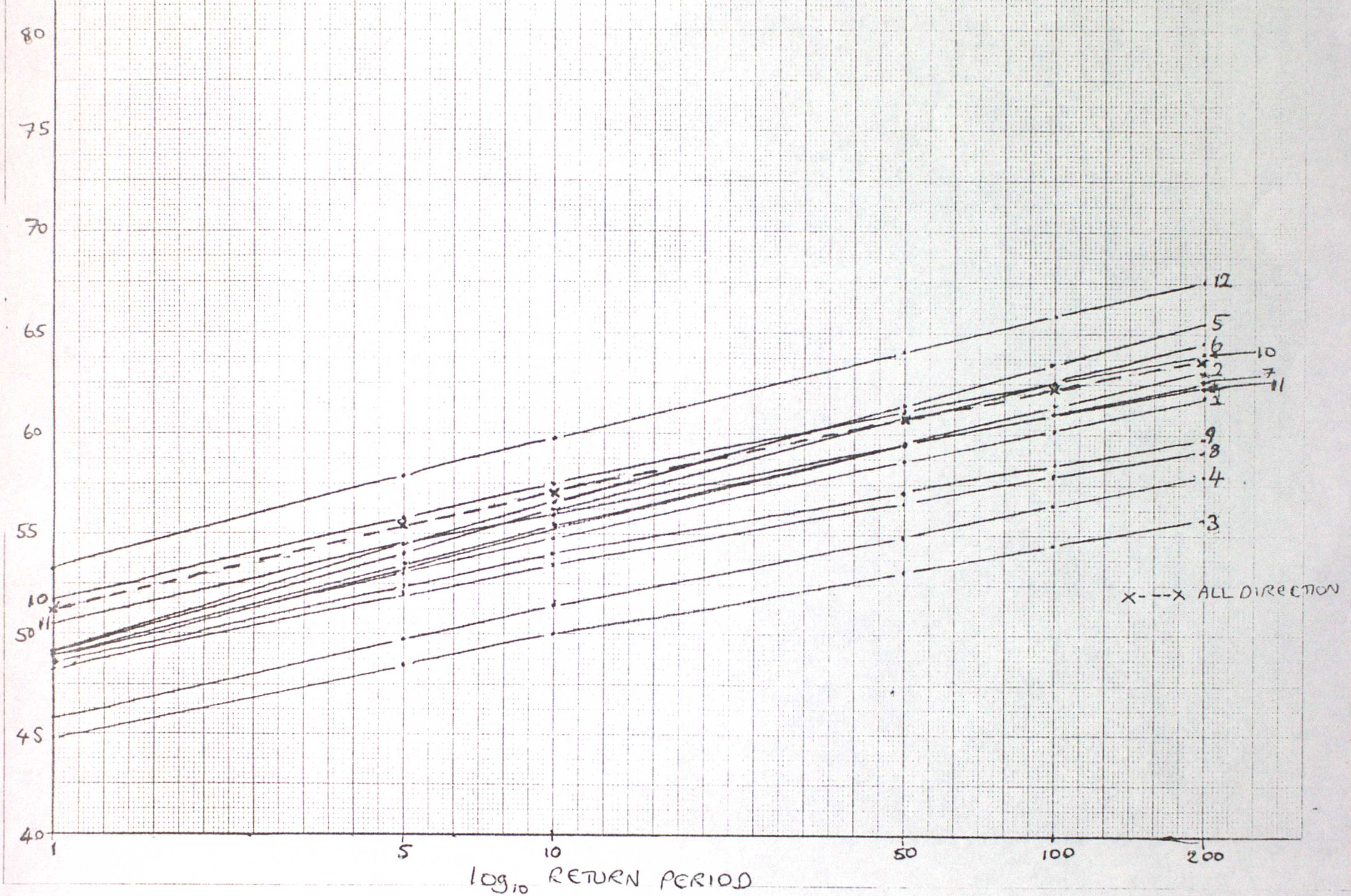


FIGURE 2·1c. EXTREME WINDS, AREA 3.



3. Wind speed growth curve for extreme values

In Met O 13 Branch Memorandum 55, A F Jenkinson records a standard growth curve for maximum annual wind speeds which is well defined up to a return period of 200 years. Table 3.1 lists the growth curve using the 2 year return period value as the base figure. Jenkinson used 45 coastal stations grouped by area and 6 different periods within the whole period covered by the data. He concluded that the maximum wind speed growth curve was essentially the same for all stations in all areas, and was as shown in Table 3.1.

Table 3.2 shows the individual growth curves for maximum annual wind speed calculated for 16 locations, of which 14 were land stations, mostly coastal, and 2 were ocean weather stations (OWS). Of these 16 stations only 7 have growth curves similar to the standard one. The growth curves were calculated from the results of Gumbel extreme value analyses. It would appear that, although Jenkinson used far more stations and looked more closely at the results, the standard growth curve cannot be taken as applying to all stations although for practical purposes it would often give very good results.

The original application of the wind growth curve was to relate wind speed with significant wave heights. It has since bee used to produce 'corrected' extreme values where the raw data have been analysed using the Weibull method of extreme value analysis. To use the growth curve in this way it is necessary to know one extreme value of a particular return period. For example, if the 2 year extreme was considered correct the growth curve values can be applied directly to that extreme to produce extremes having different return periods.

As discussed in part 1 it has been believed that Weibull extremes at the once in 5 year level were comparable with those from Gumbel analyses at the same level. Consequently, in some studies the Weibull method was used to estimate the once in 5 year extreme and the growth curve was applied to produce extremes for the other return periods required. This would have produced erroneous extremes at all return periods, since the 5 year extreme is unlikely to be correct. Even for those coastal stations where the two methods gave the same extreme at the once in 4 year level the application of the growth curve will have produced progressively poorer estimates of the 'correct' extreme.

An example of the application of these two assumptions; i) that Weibull gives a good estimate of the 5 year extreme, ii) the growth curve is applicable at all sites, is contained in the Stevenson Report (1980), where winds at the Lerwick land station were used for comparison with winds from the Stevenson station (positions shown in Figure 3.1). The winds at Stevenson covered the period February 1973 to February 1976 but contained many short periods where no observations were available. These winds were compared with a long period of wind data (41 years) from Lerwick and also with the corresponding short period.

Table 3.3 shows the results derived from the short periods of Lerwick winds, analysed using the Weibull distribution and then 'corrected' with the wind growth curve ratios, assuming that the 5 year extreme is a good estimate. The corrected extremes are shown in Figure 3.2a plotted against the $\log_{10}$ return period together with the extremes derived from the longer period of Lerwick winds (table 3.4) using annual maxima and the Gumbel distribution (from Met O 3 standard analyses). This figure shows that the 'gradient' is quite well matched with the Gumbel extreme line but the wind speed values are not in good agreement, though much better than the original Weibull extremes would have been. Figure 3.2b shows the results obtained with 'corrected' extremes (table 3.5) derived using the Weibull method and assuming that the 2 year extreme is a good estimate. It is known from the results in Part 1 that the 2 year extreme derived using the Weibull method is comparable with the Gumbel derived 2 year extreme. In this case the final results compare favourably with the Gumbel extremes though they are still slightly underestimated.

This example emphasises the need for great care when applying standard ratios to extreme value analysis, the extremes in the case of the 'corrected' Weibull values for Lerwick are underestimated by 2 ms^{-1} at the 50 year return period using the one in 5 year extreme as the original 'correct' value. These results were then used for comparison with the Stevenson winds and a similar technique applied to 'correct' extremes from this data, again assuming the one in 5 year extreme to be 'correct'. The technique has been shown incorrect for Lerwick and though nothing is known about the long term climate at Stevenson the same technique was applied. While it is reasonable to assume that the climate is similar to that at Lerwick, the differences which do exist may prove to invalidate the use of the technique at the Stevenson site and the extremes derived in this way may be in error.

Table 3.1 Jenkinson's maximum annual wind speed growth curve

Return period (years)	2	5	10	20	50	100	200
Growth curve	1.00	1.11	1.17	1.24	1.33	1.40	1.47

Table 3.2 Table of maximum annual wind speed growth curves

Return period (year)	2	5	10	20	50	100	200
Station							
Stornoway	1	1.10	1.17	1.24	1.32	1.39	1.45
Valley	1	1.11	1.18	1.25	1.35	1.41	1.48
Prestwick	1	1.11	1.19	1.26	1.35	1.42	1.49
Wick	1	1.15	1.25	1.34	1.46	1.56	1.65
Fraserburgh	1	1.11	1.18	1.25	1.34	1.40	1.47
Lerwick	1	1.11	1.18	1.25	1.35	1.41	1.48
Squires Gate	1	1.11	1.18	1.25	1.34	1.40	1.47
Manston	1	1.12	1.20	1.28	1.38	1.45	1.53
Dungeness	1	1.09	1.15	1.21	1.29	1.34	1.40
Portland Bill	1	1.13	1.21	1.29	1.39	1.47	1.55
Tiree	1	1.15	1.25	1.35	1.47	1.57	1.66
Thorney Island	1	1.13	1.22	1.30	1.41	1.49	1.57
Scilly	1	1.09	1.16	1.22	1.30	1.36	1.41
Gorleston	1	1.10	1.16	1.22	1.30	1.36	1.42
OWS 'J'	1	1.10	1.17	1.24	1.33	1.39	1.46
OWS 'I'	1	1.08	1.14	1.19	1.26	1.31	1.36

Table 3.3 Lerwick 'corrected' extreme wind speeds from the Stevenson Report.

Return period (years)	2	5	10	20	50	100	200
Extreme wind (m/s)	25.0	27.7	29.2	31.0	33.3	35.0	36.8

Table 3.4 'Gumbel' extremes for Lerwick, using Met O 3 standard results.

Return period (years)	2	5	10	20	50	100	200
Extreme wind (m/s)	26.3	29.3	31.2	33.1	35.5	37.3	39.1

Table 3.5 'Weibull' extremes for Lerwick, 1970 to 1981 hourly mean anemograph wind speeds. Corrected using Jenkinson's growth curve.

Return period (years)	2	5	10	20	50	100	200
Extreme wind (m/s)	26.2	27.5	28.3	29.2	30.3	31.1	31.9
'Corrected' extreme	26.2	29.1	30.7	32.5	34.9	36.7	38.6

Figure 3.1 Location of Lerwick and Stevenson stations.

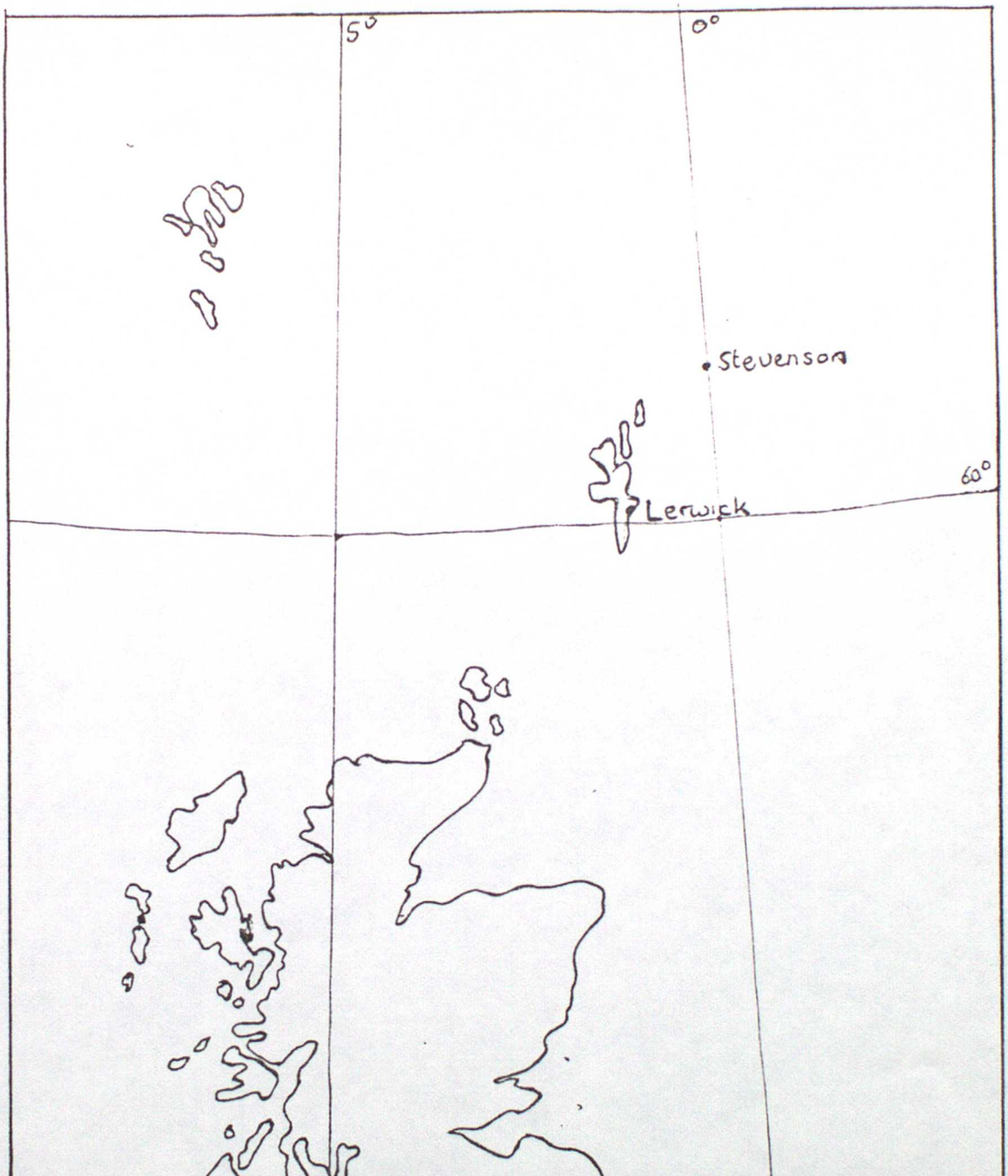
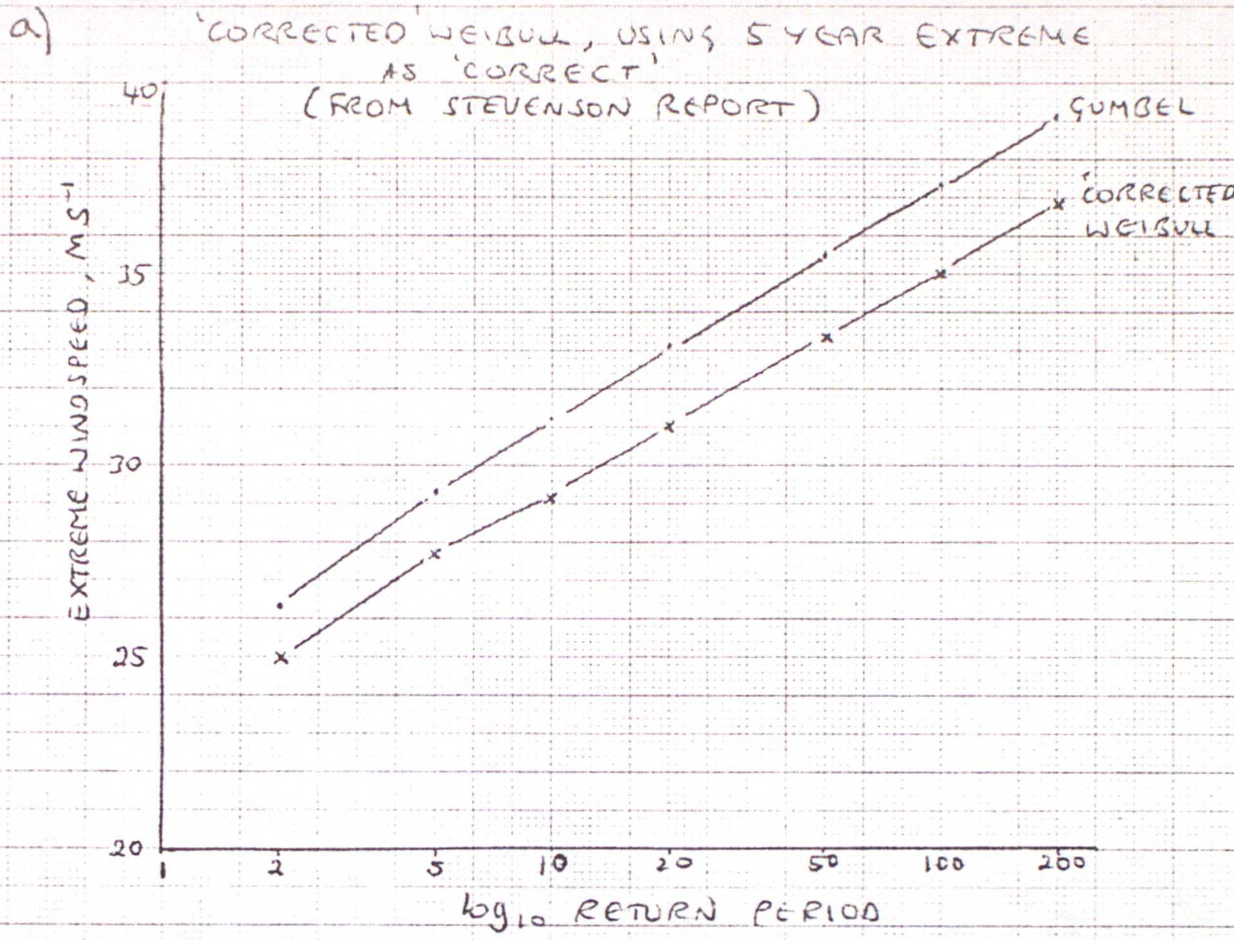
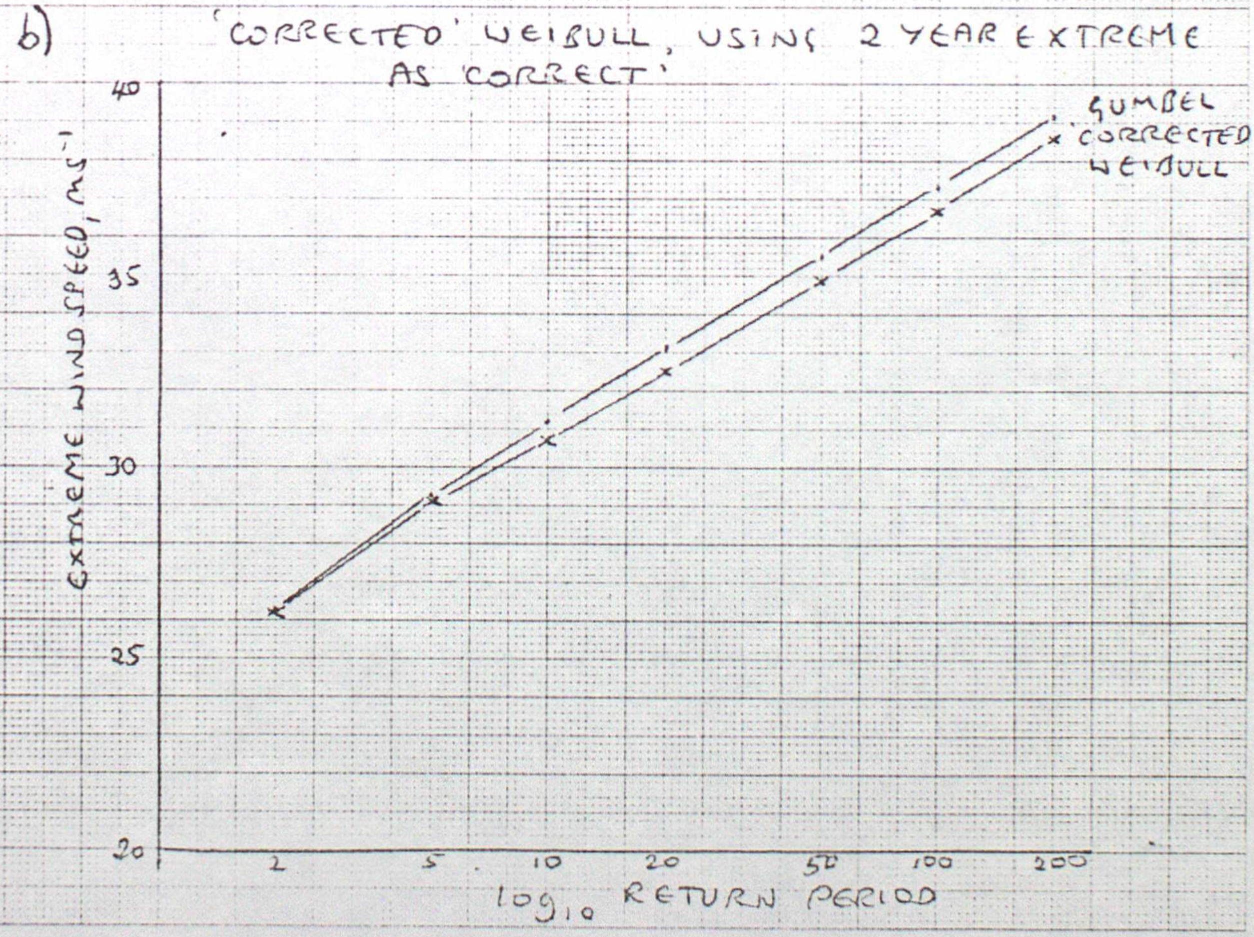


FIGURE 3.2 LERWICK EXTREME WINDS.





4. Expression to correct for persistence in data used in extreme value analysis

Meteorological variables are highly persistent and this persistence was believed to be the major cause of the differences between extremes derived from annual maxima using the Gumbel (or Fisher-Tippett Type 1) distribution and those derived using the whole spectrum of data with the Weibull distribution. The extremes estimated by the Weibull technique are for short return periods, overestimated when compared with those from the corresponding Gumbel analysis, and underestimated for longer return periods. For some time it was believed that 'short' return periods were those of up to 5 years and 'longer' return periods those greater than 5 years.

An exponential expression was derived by Painting (Appendix I) from studies of wind speeds at ocean weather station 'I', this expression took the form:

$$P\ (n) = 1 - \exp\ (-n/\lambda\ ) \qquad (1)$$

Where λ is the mean percentage of observations of winds $\geqslant$ 60 kt (the threshold value) and P (n) the proportion of years with winds $\geqslant$ 60 kt occurring in the range (0-n) per cent.

This expression was then assumed to apply for all threshold values at any site. It was redefined in general terms in an attempt to correct for persistence in data used in the Weibull analysis technique (Appendix II). The Weibull method was thought to be more susceptible to persistence in the data because the whole distribution of observations is used rather than the annual maxima alone.

The general form of the expression was given by:

$$Pv\ (x) = 1 - \exp\ (- x/\phi) \qquad (2)$$

Where v is the threshold value or extreme value concerned, and ϕ refers to the return value of v or the proportion of exceedances occurring in one year (eg for the once in one year, $\phi = 1$; once in 5 years, $\phi = 0.2$). The number of exceedances actually occurring in any year is given by x = 1, 2, 3 ..... and Pv (x) is the proportion of years with v exceeded on x occasions. Pv (x) is equivalent to r/Y, where Y is the total number of years involved and r the number of years with x exceedances of v.

Expression 2 defined in this way can be used to give the proportion of years which actually contain the exceedances of v. To illustrate this consider an example in which there is, on average, one exceedance per year. In Y years there will be E exceedances of v, the one year extreme, where E will be equal to Y. However, the occurrences will not be evenly distributed, there will be some years with only one exceedance of v_1 , some with 2 and some with 3 until all E exceedances are used up. Expression 2 shows that all E exceedances will be used up in 0.42 Y years. Painting then says that this can be used to give an 'equivalent' return period of 2.38 years (Y/0.42Y), for the once in one year event. This implies that at least one exceedance of v will take place every 2.38 years. However, there are E such exceedances which must occur in Y years, and these will not be distributed uniformly. Therefore, there are two choices when stating the return period, firstly to continue to use an average return period, secondly to give an indication of the likely distribution of the exceedances throughout the Y year period. The first option would give an average value of two exceedances every 2.38

years, or effectively 1 per year on average. The second option is not practical because the precise sequence of occurrence of exceedances in any period of years is not known.

The difference between the actual return period for the event which occurs, on average, once in 5 years, v_5 , and the 'equivalent' return period is negligible. This is because it is much less likely that v will be exceeded more than once in any year and so all E exceedances in a total period of Y years (where E = Y/5) will still occur in 0.2 Y years. The same is true for longer return periods.

Consequently, expression 2 cannot be used to give a general correction for persistence. It does indicate that for the more frequently occurring event, eg the once in one year case, the actual exceedances do not necessarily occur once each year but can occur up to 3 or 4 times in any one year provided that the average over a number of years is once per year.

Thus, Paintings method adds very little to the practical methodology, and unfortunately, closer examination reveals that the exponential expression devised is not reliable.

The use of expression 1 in its general form, 2, relies firstly upon the expression accurately describing the cumulative percentage of winds below the threshold value and secondly upon the proportion of years with repeated exceedances being the same in every case. The expression was tested (Appendix III) with several datasets for various threshold wind speeds and was found to be a poor fit to the data in almost every case. The main problem was P(0) which is always equal to zero in expression 1, but usually greater than zero for real data. A modified expression was proposed as follows:

$$P(n) = 1 - \exp\left[(-mn + c)\right]$$

where $m \equiv \lambda$ and $c = P(0)$. This expression gave a good fit in most cases but it cannot be used generally as was expression 1 because c can only be defined by real data, and is unique to the case under consideration.

In conclusion, expression 1 gives a poor fit to real data in most cases and the improved modified version cannot be used in general. The generalised version of expression 1 cannot be used to correct return periods affected by persistent data because it does not fit the data well. The method can only be used to give an average return period, and therefore, adds nothing to existing techniques.

It should be noted here that persistence is no longer believed to be the major cause of the differences between extremes derived using maxima and the whole spectrum of data as a result of work done in the Meteorological Office and by research workers elsewhere eg Grigoriu (1980).

APPENDIX I

Preliminary note on the Variability of Weather statistics as observed at OWS 'I'

By D J Painting

1. Introduction. To determine long period statistics for design and operational purposes in offshore areas a long and continuous series of reliable measurements are required, preferably made at or nearby the area of interest. In many cases such data do not exist and the problem may be tackled by establishing an observing station at the site for a relatively short period in order to characterize the local climate. The question then arises as to how long need such a station operate to enable reliable long term statistics (extreme and operational) to be deduced. To attempt to answer this question wind data at OWS 'I' have been studied over the period 1957-1973. This Ocean Weather Station, continuously manned by professional observers, can be considered to give the most reliable data available at such an ocean site.

2. Wind statistics. This preliminary note has involved the study of the annual variability of high wind speeds ($\geqslant$ 60 kt) since accurate knowledge of the long term proportion of wind speeds at or above this level is crucial to the determination of wind speeds at long return intervals.

3. Results. A time series of the proportion (%) of winds $\geqslant$ 60 kt each year (1957-1973) is shown in figure 1. This shows wide year to year variation with an extreme range of 0 to more than 0.13%. Equivalent return periods for winds equal to or greater than 60 kt are indicated showing a very wide variation based on the annual statistics. An important feature to note is the fact that more years show proportions below the average than above. This is characteristic of distributions of rare events. Also plotted are the 3 year running averages. These also show wide variability with equivalent return periods ranging from about 0.36 years to 0.9 years.

The annual data appear to fit a simple exponential model $P(n) = 1-e^{-n/\lambda}$

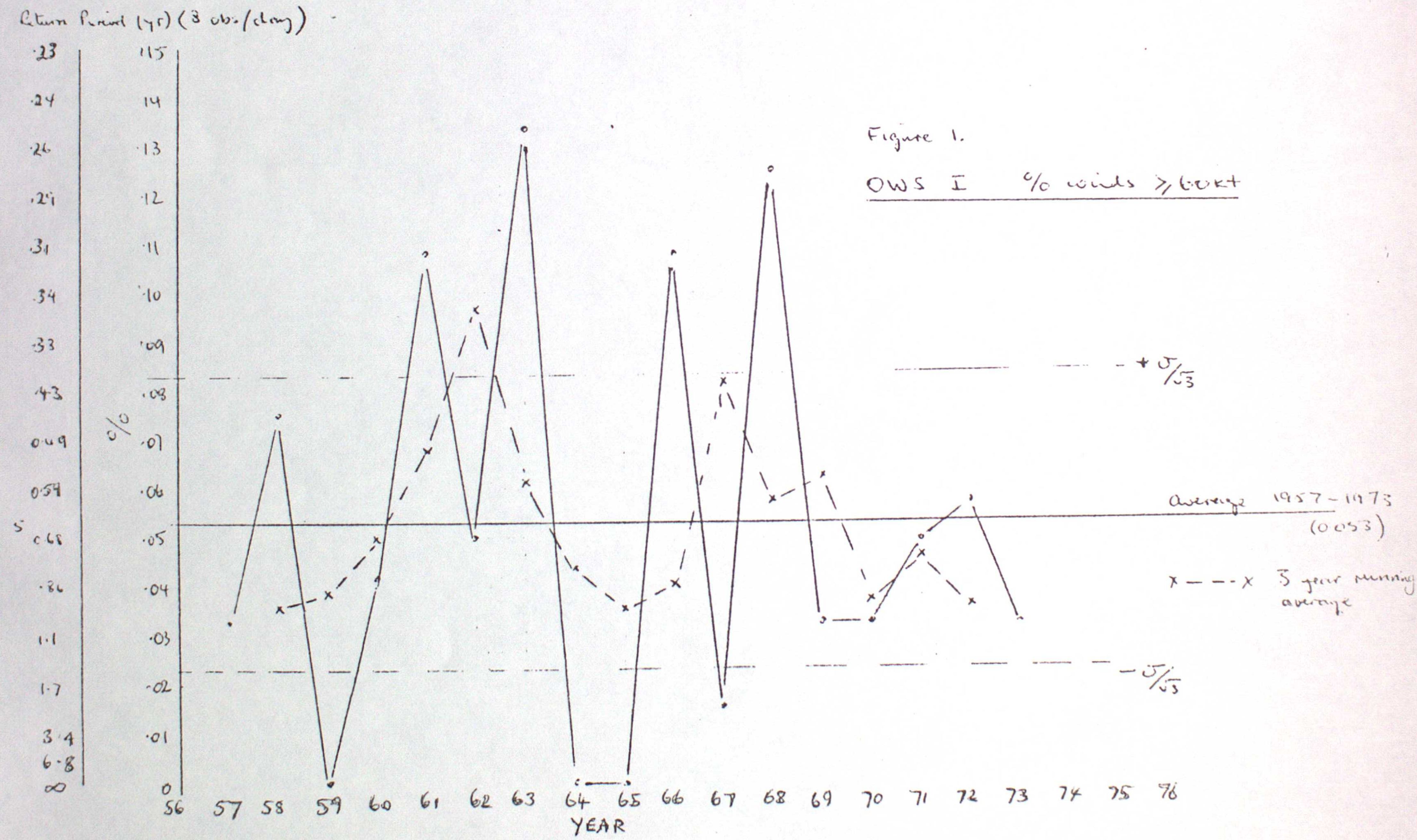
where λ is the mean percentage $\geqslant$ 60 kt and $P(n)$ is the proportion of events (years) with percentage winds $\geqslant$ 60 kt occurring in the range (0-n)%.

The fit of this model and the actual data are shown in figure 2.

4. Conclusions. Assuming the validity of the model described above very long averages will be required to reduce the uncertainty of the statistics to an acceptable level eg 30 years will produce a standard error of estimate of about 18% at the given probability level eg in the case considered $P(V \geqslant 60\text{ kt}) \simeq .053 \pm .010$. It is obvious that short period data in itself cannot reliably describe long term statistics; however if we have nearby a long period station which 'tracks' the short period station reliable estimates should be possible by adjustment of the long period station data. To establish relationships with such long period stations a period of measurement of some 3 years is probably required so that most of the possible synoptic situations will have been encountered.

Figure 1.

OWS I % winds ≥ 60 kt



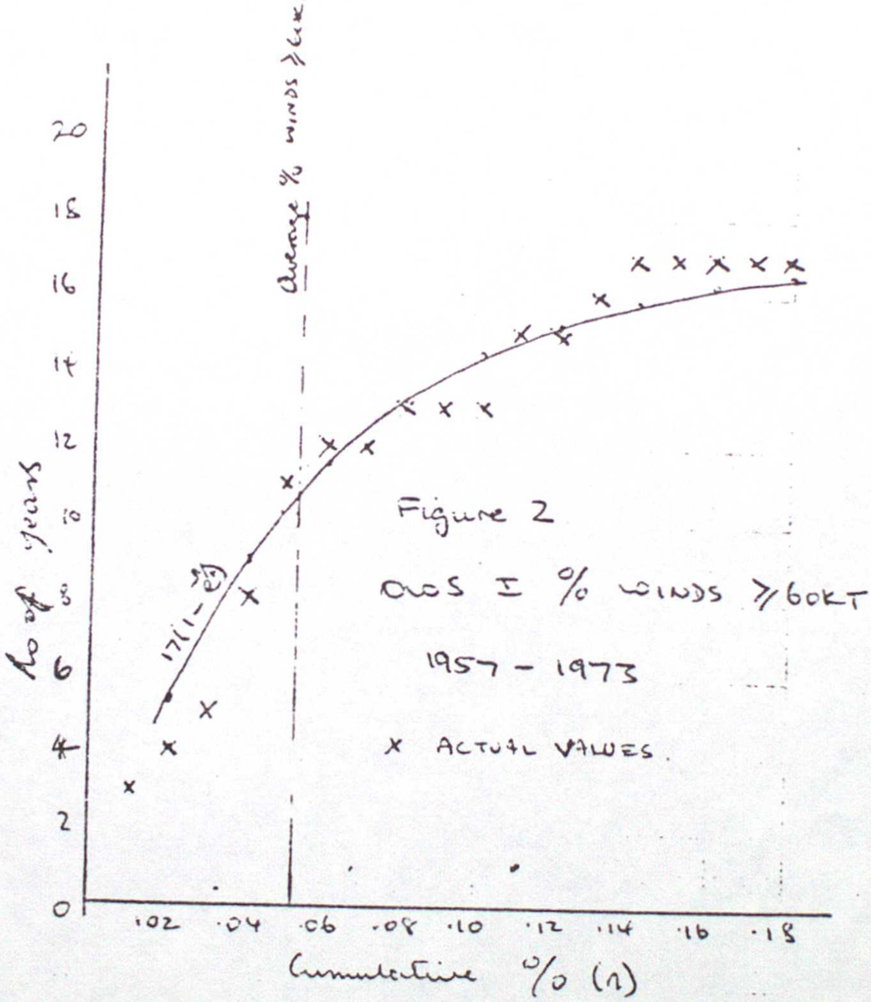


Figure 2

OWS I % WINDS ≥60KT

1957 - 1973

x ACTUAL VALUES.

APPENDIX II

EXTREME VALUE ESTIMATION - CORRECTING FOR PERISTENCE WHEN USING THE PARENT DISTRIBUTION

1. Introduction

Estimates of extreme values at long return intervals are often made by fitting annual maxima to a standard extreme value distribution eg the 'Gumbel' or Fisher Tippett type I extreme value distribution. In order to contain the errors of estimation within a reasonable range a long series of data is required, typically 30 years. In many cases long time series data are not available and estimates of extremes are made by fitting the frequency distributions of all available data to an empirical distribution (such as the Weibull distribution in the case of wind or wave data).

$$\text{eg} \quad P(v < V) = 1 - \exp\left[-\left(\frac{V-Vo}{Vc}\right)^M\right]$$

Where M, Vo and Vc are parameters of a curve fitted to the cumulative frequency distribution of v (the parent distribution).

2. The effect of persistence

Given a function which completely describes the parent distribution the exceedence probability at any threshold value can be calculated and related to the average return interval as follows:

$$Pv,y, = \frac{1}{Ny} \left[= \exp\left[-\left(\frac{v-Vo}{Vc}\right)^M\right]y \text{ using the Weibull equation}\right]$$

Where Pv,y is the value of v exceeded on average once in y years and Ny is the total number of <u>independent</u> values occurring in y years.

Thus for a return interval of 5 years given hourly mean winds (assumed independent) we have

$$Ny = 365.25 \times 24 \times 5$$

$$\text{and} \quad Py = .0000228$$

Wind and wave data, however, exhibit persistence; that is successive data samples are highly correlated. Thus, extreme values tend to group together in single storms. It is natural to consider the whole storm as a single extreme event and the highest value in the storms of a single year is equivalent to the value used in the 'Gumbel' type distribution of annual maxima. The overall effect of persistence, therefore, is to reduce the effective return interval associated with a particular exceedence level of the parent distribution by comparison with the equivalent interval from the extreme value distribution. As an example consider the following wind data collected at Tiree over the period 1957-1977.

Wind Speed	No of Hours	No of Storms	No of Years
≥ 60 Kt	6	3	3
≥ 58 Kt	11	4	4
≥ 50 Kt	68	24	11

It is noted that, at the 50 Kt threshold (for example), the equivalent return intervals are approximately as follows:

a. Based on annual exceedences - 1.9 years

b. Based on total exceedences - 0.31 years

It is clear that estimations based on the parent distribution underestimate the return interval for short return intervals compared those based on annual maxima.

3. Adjusting for the effect of persistence

It has been shown (eg Painting 1979) that the annual distributions of the occurrences of rare values are well described by an exponential relationship of the form

$$P_V \quad (0 < x < X) \quad = \quad 1 - \exp\left(-\frac{x}{n}\right)$$

Where n is the average number (or proportion) of values exceeding the value v.

Let us now consider the distributions of the values exceeded on average once in

1, 5 and 10 years respectively

a. $P_{v1} = 1 - \exp - x$

b. $P_{v5} = 1 - \exp - x/0.2$

c. $P_{v10} = 1 - \exp - x/0.1$

	P (v)		
n	1	0.2	0.1
x = 1	0.63	0.993	0.9999
x = 2	0.865	0.9999	
x = 3	0.95		
x = 4	0.98		
x = 5	0.993		

For the 'one year' value we have

$P\ (x>1) = 0.37$

$P\ (x>2) = 0.135$

$P\ (x>3) = 0.05$

$P\ (x>4) = 0.02$

$P\ (x>5) = 0.003$

Thus, on average, in a 100-year period (say) we would have 37 years with v_1 exceeded more than once, 13.5 years with v_1 exceeded more than twice etc. Assuming that x can only take the values 0, 1, 2 etc this implies that on average 22.5 years will have 2 exceedences, 8.5 years will have 3 exceedences etc. The total number of exceedences accumulated in the years having at least two is therefore 23.5 x 2 + 8.5 x 3 + 3 x 4 + 2 x 5 $\triangleq$ 95 in a total time of 37 years. This leaves 5 exceedences to occur in single years ie a total of 42 years will have v_1 exceeded at least once per year. The equivalent return interval for v_1, is thus 100/42 years = 2.38 years when considering annual extremes only.

For the '5-year' value we have

$$P(x > 1) = 0.003$$
$$P(x > 2) \triangleq 0$$

Thus in 1000 years we have 3 years on average with 2 exceedences ie a total of 6 in 3 years leaving 194 years with single occurrences. The equivalent return interval for v_5 is thus $\frac{1000}{197}$ years = 5.08 years.

For the '10-year' value the adjustment is insignificant.

4. Some practical considerations

A common procedure is to fit a Weibull curve to a frequency distribution. The nature of this process is to fit to the body of the parent distribution, ie the tail of the distribution (the rare events) cannot exert much weight on the resulting estimates of the Weibull parameters. If a correction for grouping is applied this has the effect of reducing the Weibull slope (m). In general this will cause the adjusted curve to cross the original curve which suggests the likelihood of underestimation of extremes by extrapolation of Weibull fitted curves at long return intervals together with overestimation of extremes at short return intervals.

5. Conclusions

We have shown how the persistence of data can affect the interpretation of extreme value estimates when using extrapolations from a parent distribution (such as the Weibull distribution). In particular it is noted that return periods are under-estimated up to return periods of near 5 years. A simple model predicts the order of this overestimation to be more than double when considering the '1-year' value from a 'Weibull' type estimate. This result is confirmed by data measured at Tiree. Extrapolations of Weibull curves beyond the data points are likely to yield under-estimates of extremes at long return intervals and thus this method should be avoided if possible. It is probable that 'Weibull' estimates at around the 5-year return interval are comparable with equivalent estimates from the distributions of annual maxima.

D J Painting

D J Painting
Head of Met O 16a

9 December 1980

Figure IIIa. Proportional exceedances of various threshold wind speeds for Valley (1964-1979)

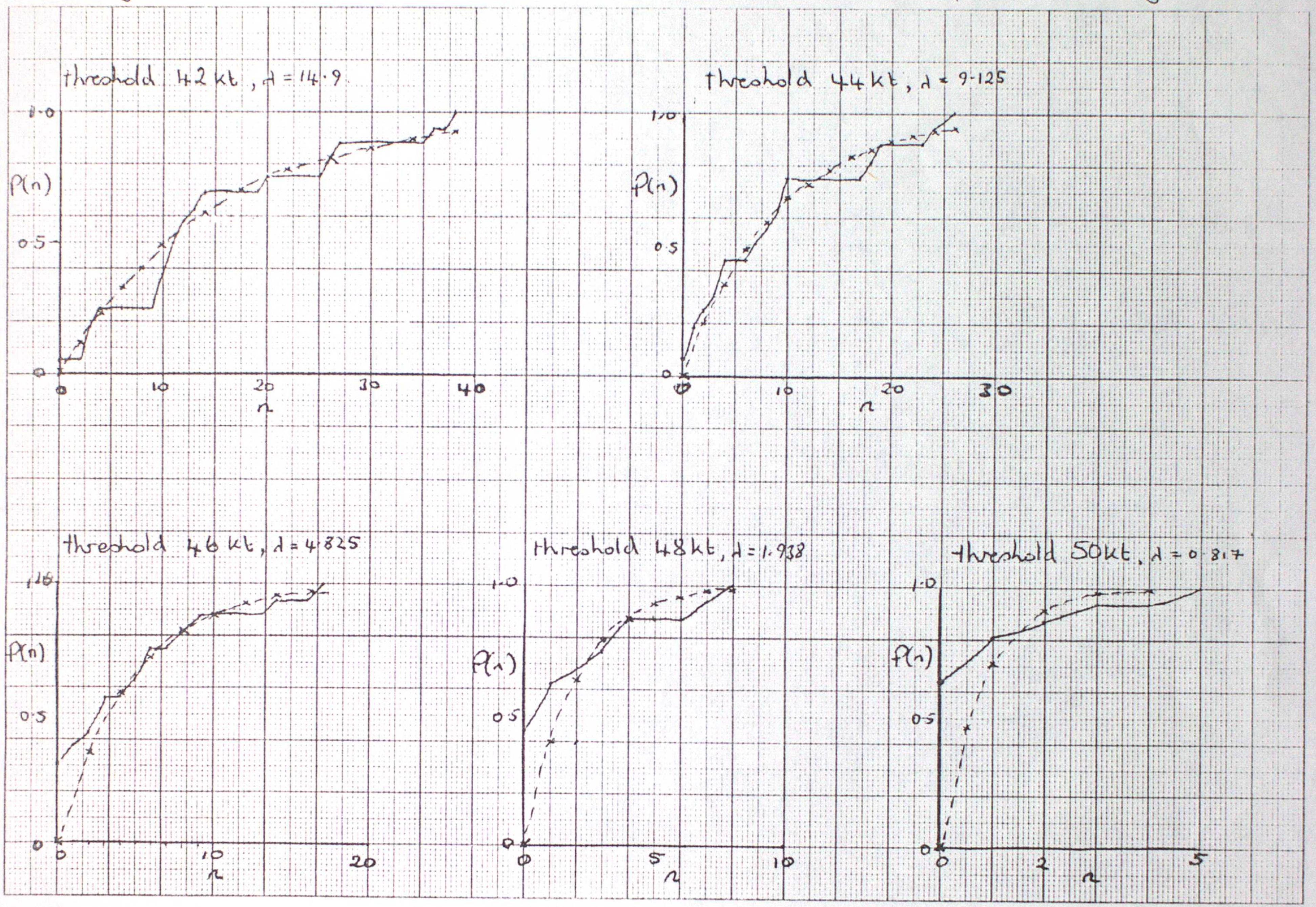
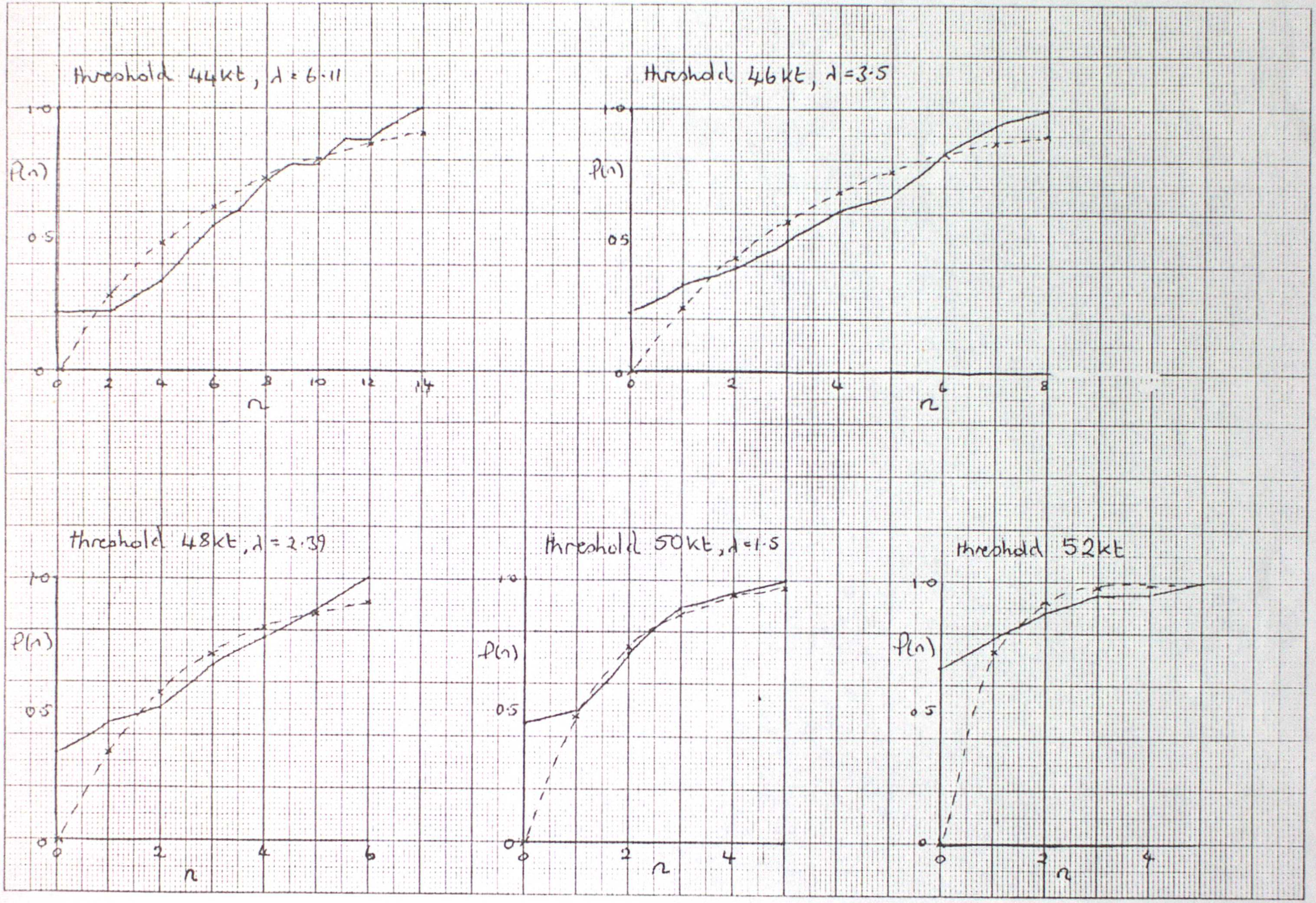


Figure IIIb. Proportional exceedances of various thresholds wind speeds for Lerwick (1962-1979)



APPENDIX III

Summary of the main results from a study by J Atkins to test the fit of the exponential expression derived in appendix I.

The figures IIIa and IIIb were taken from a study by J Atkins in Met O 3c to test the fit of the relationship:

$$P(n) = 1 - \exp(-n/\lambda) \qquad (1)$$

derived in appendix I, using wind speed data from different locations and for various thresholds. The results for the two land stations Valley and Lerwick are shown in figures IIIa and IIIb respectively. These clearly show that the fit of expression (1) to the data becomes progressively poorer for higher threshold wind speeds. This implies that expression (1) cannot be used for improving estimations of extreme conditions.

A modified expression suggested by Atkins is:

$$P(n) = 1 - \exp(-mn+c) \qquad (2)$$

where m and c are determined by a regression of $-\ln(1-P(n) = mn+c$. For Valley this expression provided a much better fit to the data, but for Lerwick the fit was generally poorer than that for expression (1). Atkins concludes that although neither of these expressions, (1) and (2), adequately fit all data, a two parameter relationship similar to (2) would be generally better than the simple one parameter expression.

Key to figures:-

$P(n)$, proportion of years with stated number of exceedances n, or fewer.

n, number of exceedances of a threshold wind speed.

___ actual proportion of years with n exceedances

x--x $P(n) = 1 - \exp(-n/\lambda)$, λ = average number of exceedances.

JB

5. References

Dept. of Energy, 1977 Offshore Installations: Guidance on design and construction, Part II. HMSO, London.

Jenkinson, A.F. 1975 Analysis of maximum significant wave height data for selected North Sea storms, Meteorological Office (Met O 13) Branch Memorandum No. 55.

Painting, D.J. 1980 Stevenson Station - Summary Report (Confidential to UKOOA) Meteorological Office, Bracknell.

Grigorui, M. 1980 Effects of correlation on extreme wind. N.Y. Amer. Soc. Civ. Eng., ASCE. Conv. Expo. Hollywood Flat, Oct 27-31, 1980, pre-print 80-677 pp (9).