THE METEOROLOGICAL MAGAZINE

HER MAJESTY'S
STATIONERY
OFFICE

April 1983

Met.O. 958 No. 1329 Vol. 112

METEOROLOGICAL OFFICE Met.O.958

THE METEOROLOGICAL MAGAZINE

No. 1329, April 1983, Vol. 112

551.501.45:519.23

Extreme value analysis in meteorology

By R. C. Tabony

(Meteorological Office, Bracknell)

Summary

The theory of extreme values assumes that maxima (or minima) are drawn from infinitely large samples of independent observations belonging to a single population. Failure to satisfy the theory, therefore, can be due to using too small a sample or to the inclusion of observations from more than one population. It is demonstrated that in meteorology these reasons are often alternative expressions of the same problem, namely lack of data.

A series of extremes may be regarded as belonging to the same population if a single forcing factor is responsible for the whole range of extremes encountered. This is seldom true of meteorological variables. It is shown that analyses of annual extremes are commonly to be preferred to those based on monthly data.

When observed extremes fall well short of a physically imposed upper limit it is suggested that they can appear to be unbounded above. For short duration rainfall this can be interpreted as being due to changes in the organizational structure of convective storms as we pass from the lesser to the greater extremes.

1. Introduction

A knowledge of the highest and lowest values which meteorological variables are likely to attain in a given number of years is important to many aspects of engineering design. The analysis of extreme values is therefore a topic of great importance in meteorology. A good introduction to the subject is given by Kendall and Stuart (1977) while comprehensive accounts are given by Gumbel (1958) and Galambos (1978).

Many extreme value analyses of meteorological variables have been undertaken in the past. In the United Kingdom, for instance, temperature has been analysed by Hopkins and Whyte (1975), wind by Hardman *et al.* (1973), and rainfall by Jenkinson in the *Flood Studies Report* (Natural Environment Research Council, 1975). The application of extreme value analysis to meteorological data is seldom without its problems. Hopkins and Whyte, for example, found that the predicted upper bound of temperature was too low, while Jenkinson found that rainfall extremes appeared unbounded above. Hardman *et al.* encountered problems with outliers, i.e. observations which, when plotted on extreme value probability paper, did not lie on the same general curve as the remainder of the data. An example of an outlier is shown in Fig. 1 which displays maximum temperature in June at Ivigtut on the south-west coast of Greenland. Most of the observations lie between 13 °C and 23 °C, but the highest recorded temperature is 30 °C.

In this paper the assumptions behind the theory of extreme values are examined and the difficulty of meteorological observations in meeting them is discussed. Suggestions are made as to how the various

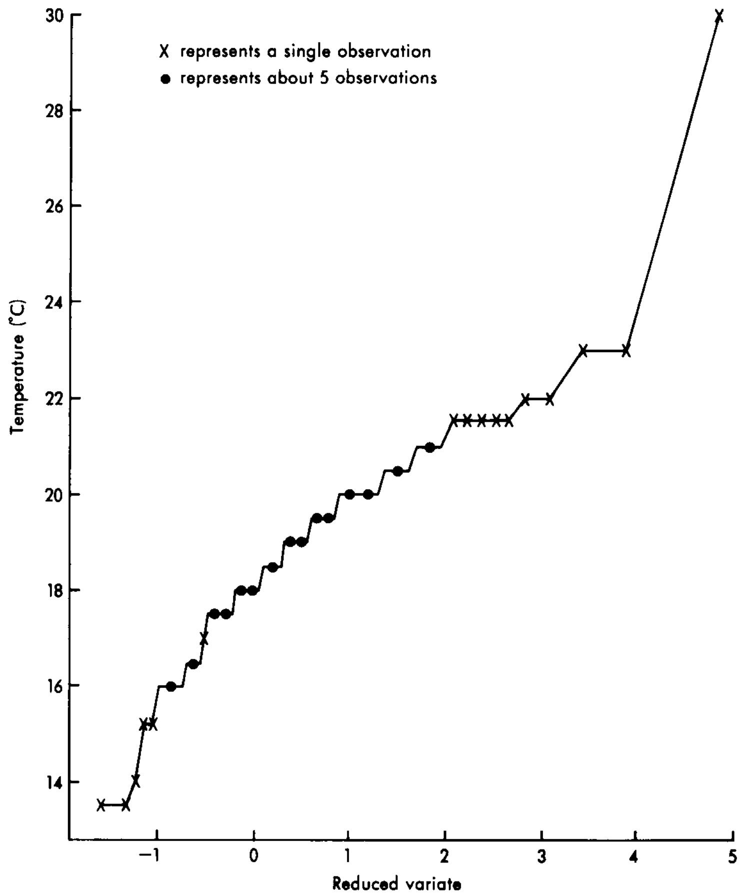


Figure 1. Maximum temperatures for June at Ivigtut, 1875–1960.

problems posed by analyses of meteorological extremes may best be interpreted. All the data used are tabulated in Appendix 2. They were either held in manuscript form within the Meteorological Office or extracted from the year books of the appropriate country.

2. Theory

Consider a series of independent observations belonging to the same population and divided into samples each containing N observations. The series of extreme values is constructed by selecting the highest (or lowest) observation from each sample. In the trivial case of $N = 1$, the sampling procedure would obviously result in the parent distribution itself. If $N = 2$, then each choice of maximum value in the sample will result in a bias towards higher values, and as N increases it is clear that the new distribution will progressively depart from the parent distribution. The differences are illustrated in Fig. 2 which displays schematically the probability density functions $f(x)$ of the parent and extreme value distributions. The theoretical extreme value distribution is approached asymptotically as N approaches infinity.

An extreme value distribution is usually expressed in terms of the cumulative distribution function F(x), and when this is plotted against x an S-shaped curve is obtained. It is usual to transform F(x) to a new variable y, known as the reduced variate, in which the cumulative probability distribution is represented by a straight line when plotted against x (see Fig. 3). The reduced variate can be related to the return period T. The value x which has the probability $1/T$ of being exceeded in any one sample is said to have a return period of T.

A general solution to the extreme value problem was obtained by Jenkinson (1955) in the form

$$x = x_0 + a\,\frac{(1 - e^{-ky})}{k}$$

where y is defined from the equation $F(x) = \exp(-e^{-y})$.

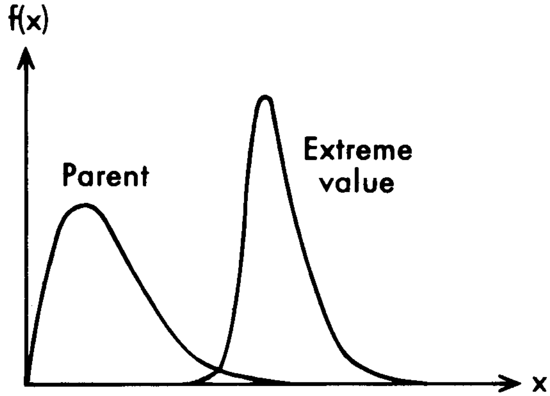


Figure 2. Probability density functions of parent and extreme value distributions.

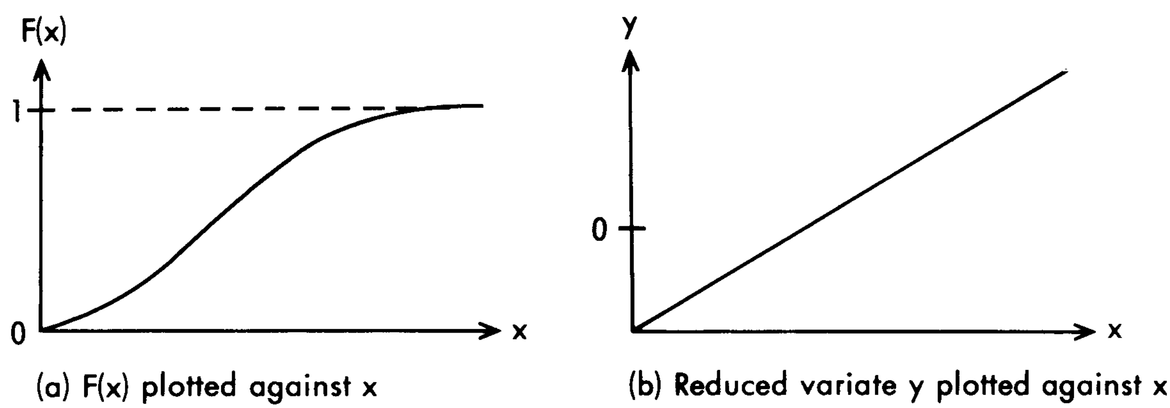


Figure 3. Cumulative probability function of extreme value distribution.

On a graph of x against y, x_0 is the value at $y = 0$ (which is exceeded by about two-thirds of the observations), α is the slope at $y = 0$, and k is a curvature parameter. The solution may be categorized into three types corresponding to separate solutions previously obtained by Fisher and Tippett (1928). They have come to be known as Fisher–Tippett types I, II, and III and are characterized by their different shapes when plotted on a graph of x against y (see Fig. 4).

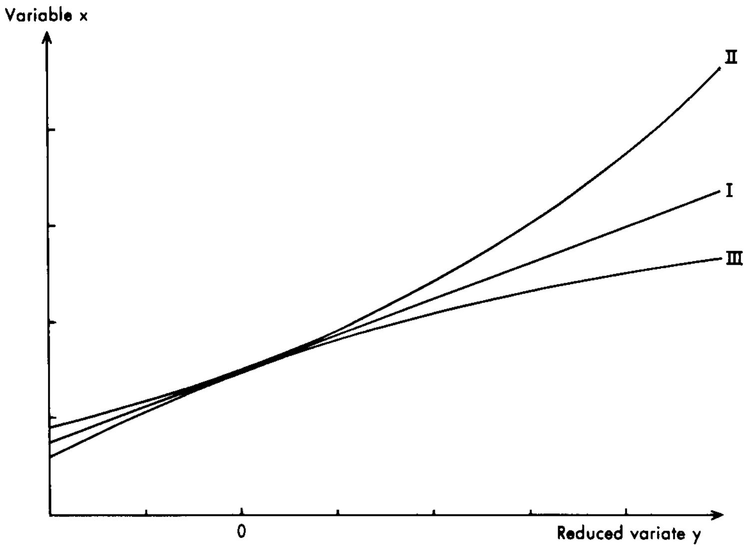


Figure 4. Fisher–Tippett distributions types I, II and III.

Type I corresponds to $k = 0$ and forms a straight line. It is the solution popularized by Gumbel (1958) and is unbounded above and below.

Type II corresponds to $k<0$ and is bounded below but not above.

Type III corresponds to $k>0$ and is bounded above but not below.

Fisher and Tippett (1928) obtained their stability postulate by assuming that the original data were independent and identically distributed (i.e. belonged to a single population). Galambos (1978) shows that asymptotic extreme distributions can exist if these conditions do not hold, but they will not necessarily be those of Fisher and Tippett. The application of her results, however, requires that the distribution of the original observations is known, and this is seldom the case in practice.

In meteorology, the problems caused by the lack of independence of the data appear to be limited to the associated reduction in the number of independent values. The problems caused by observations not being identically distributed, and by extremes not being drawn from infinite samples, are, however, considerable, and are discussed in the following sections.

3. Small samples

A series of independent and identically distributed observations will only yield a set of maxima that conform to an asymptotic extreme value distribution if the maxima have been drawn from infinitely large samples. In practice this is never achieved. The extent to which the asymptotic theory can be applied depends on how quickly the extreme value distributions approach their limiting form. Fisher and Tippett (1928) show that when the parent distribution is normal convergence is slow, while Cook (1982) demonstrates that an analysis of the square of wind speed converges more rapidly than that of wind speed itself.

Exactly how large N must be to satisfy extreme value theory within acceptable limits is an important but difficult question to answer and will vary from one application to another. Some useful guide-lines may, however, be given. The selection of maxima from a very large sample ensures that they are almost certainly drawn from the tail, which may be loosely defined as the top 10-15%, of the parent distribution. If N is so small that some of the maxima are not being drawn from the tail then this is an indication that the assumptions of extreme value theory are not being met.

For monthly maximum temperatures the number of observations from which extremes may be extracted is about 30, but serial correlation reduces the number of independent values N to about 10. This is clearly insufficient to ensure that all the maxima are drawn from the tail of the parent distribution. When N exceeds 100, however, experience indicates that observed extremes usually fit the asymptotic extreme value theory very well. For maximum temperatures this would mean taking, say, ten Januarys at a time.

In any extreme value analysis a failure to draw observations from the tail of the parent distribution will be most readily apparent in the less extreme observations. There the influence of the parent distribution may be expected. Since the type I distribution has a skewness of 1.14 these effects will be most evident when the parent distribution has negative or large positive skewness. This is illustrated in Fig. 5 for maximum temperature in January at Oxford. The general extreme value distribution has been fitted by simulating five year maxima using a computer program designed by Jenkinson (1977). Although the general curve is clearly bounded above, and may be fitted by a type III distribution, the lowest four points clearly reflect the negative skewness of the parent distribution.

The contrast with the effects of a positively skewed parent distribution is evident in Fig. 6 which displays maximum temperatures for August at Santander on the north coast of Spain.

Points drawn from a normal distribution are plotted on extreme value probability paper in Fig. 7. It can be seen that a sample of normally distributed observations could easily be accepted as belonging to an extreme value type III distribution. In practice, the only criterion for obtaining a set of data which displays linearity on extreme value probability paper is that it should have a skewness close to 1.14. Since positively skew distributions are common in meteorology, this explains why many sets of 'extreme values' appear to be well fitted by the type I distribution even though N falls short of that required by theory.

4. Mixed distributions

The discussion in section 3 assumed that the extremes were drawn from a single population. Where the observations are derived from several independent populations each may be treated separately. One area where this approach has been adopted is in the analysis of winds in regions affected by tropical storms (i.e. hurricanes, typhoons). The general methodology is well described by Gomes and Vickery (1977).

If the data contain samples from Q populations, and the distribution of extremes of the qth population are denoted by $\mathrm{F}_q(x)$, then the distribution of extremes associated with the mixed distribution is given by

$$\mathrm{F}(x) = \prod_{q=1}^{Q} \mathrm{F}_q(x).$$

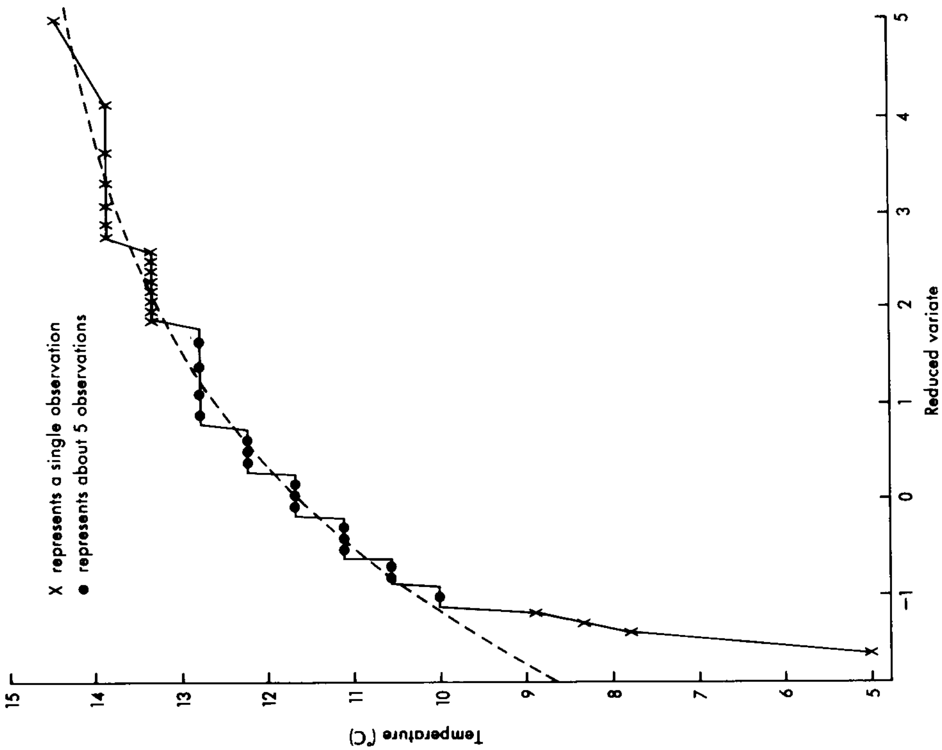


Figure 5. Maximum temperatures for January at Oxford for 1871–1970. The general extreme value distribution is represented by the dashed line.

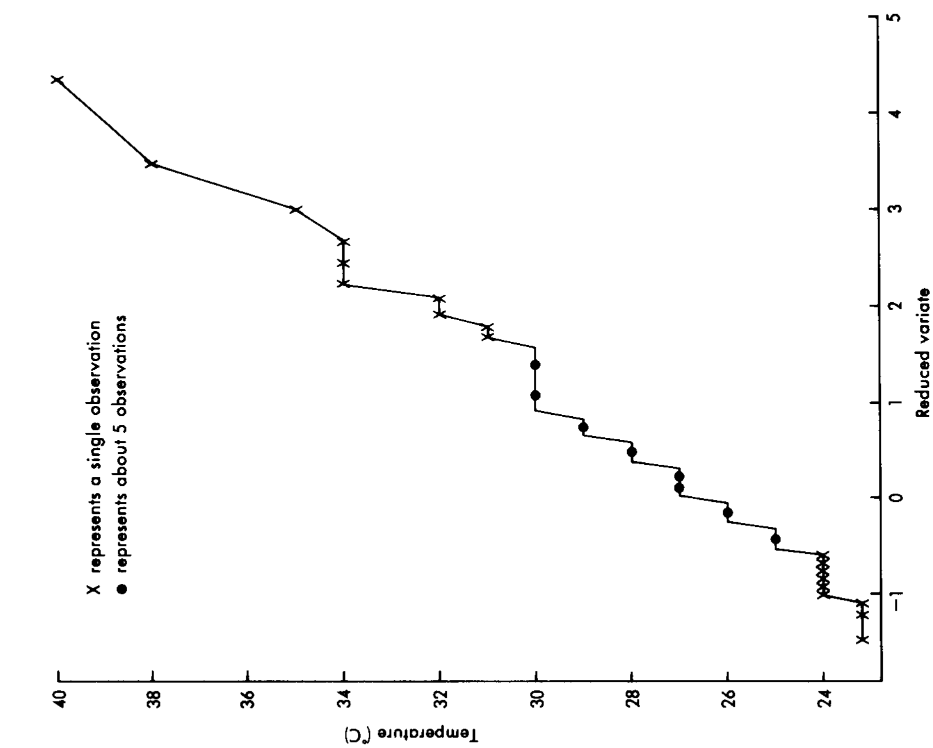


Figure 6. Maximum temperatures for August at Santander, 1927–80.

A simple example is illustrated schematically in Fig. 8. The extreme winds are assumed to belong to two populations, those due to hurricanes and those due to other causes. Each set of extremes is assumed to belong to a type I distribution. The combined probability distribution will then appear to be unbounded above and to be similar to a type II distribution. Fig. 9 presents an extreme value plot of winds for Progreso on the Yucatan peninsula of Mexico. In this example the discontinuity in the data is exceptionally well marked.

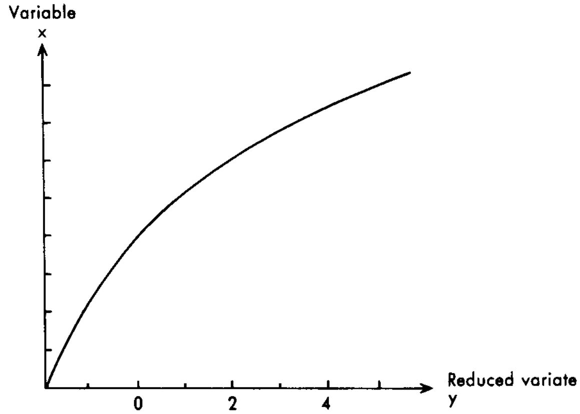


Figure 7. Normal distribution plotted on extreme value probability paper.

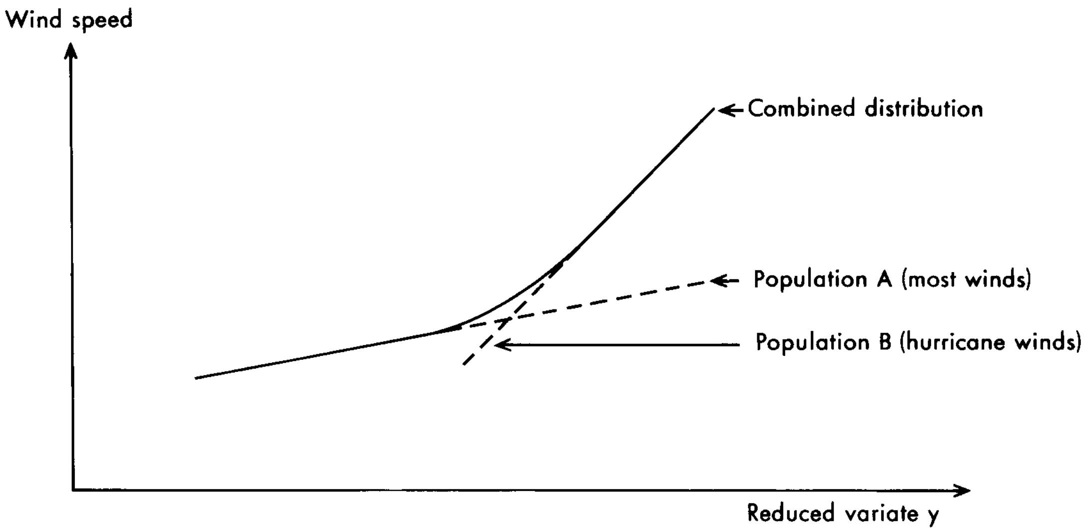


Figure 8. Extreme value analysis of observations drawn from two populations.

5. Seasonal variations

Most meteorological variables undergo a pronounced seasonal variation and consequently the observations cannot be regarded as coming from the same population. By extracting annual maxima, therefore, the theory of extreme values may well not be properly satisfied.

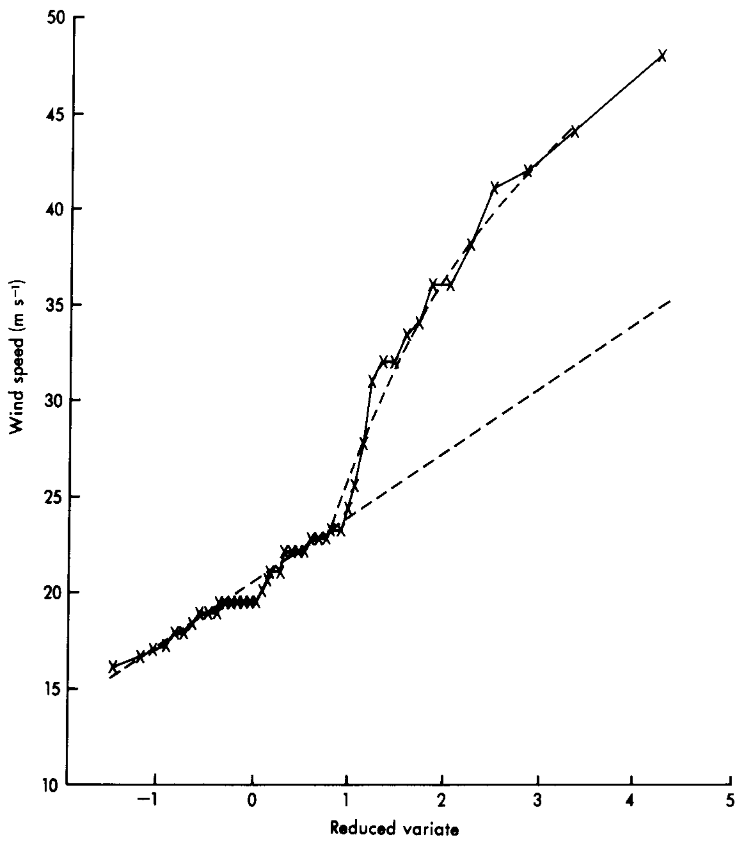


Figure 9. Annual maximum gusts at Progreso, 1921–1965.

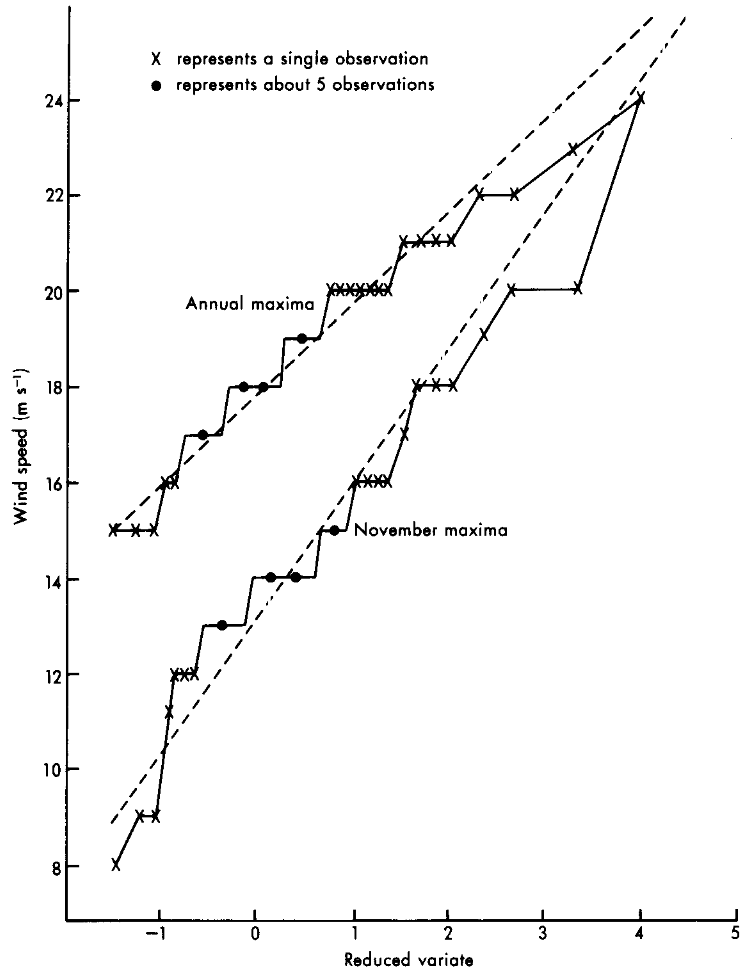


Figure 10. Annual and November maxima of mean hourly wind at Durham, 1938–78.

Consider a variable (e.g. wind) whose values rise to a seasonal maximum in (say) November. Suppose that the strongest wind was recorded in November and that a type I distribution is fitted to maximum values for November and the year. The results are illustrated using observations from Durham in Fig. 10. In general, the maxima for November are below that for the year and in some cases the differences are considerable. The highest maximum in November, however, is equal to that for the year and so the slope of the November maxima is greater than that for the year. As a result linear extrapolation of the November extremes will lead to higher estimates than those obtained from the annual analysis. Clearly linear extrapolation of both lines is not possible. Either the slope of the annual fit has to increase towards that of the monthly or the slope of the monthly relation has to decrease towards that of the annual.

Carter and Challenor (1981), analysing winds and wave height, obtain return values from linear extrapolation of the monthly extremes. The annual maxima are taken to represent a mixed distribution in which the data for each month are regarded as belonging to different populations. Carter and Challenor point out that, when choosing the number of intervals into which a year is divided, a balance has to be struck between reducing the variation within intervals and having sufficient data in each interval to maintain reasonable convergence to the asymptotic extreme value distribution. In meteorology, an interval of time as short as one month is seldom sufficient to enable the latter criterion to be satisfied.

To illustrate the point, consider the wind. One can easily imagine a month which is mainly anticyclonic with no deep depressions passing close to a station; a low maximum wind is then recorded for the month. Similarly, for short duration convective rainfall, one can easily imagine a summer month in which very little thunderstorm activity takes place; a low maximum rainfall is then recorded. At Kew, for instance, the highest daily rainfall in July 1977 was only 5 mm while in the Decembers of 1967 and 1976 the highest mean hourly wind speeds were only 17 knots. It is these low maximum values that are responsible for the steep slope of the plots of monthly maxima. These 'extremes' are clearly not being drawn from the tail of the parent distribution. In other words, for monthly maxima, N is too small.

The relative merits of a direct analysis of annual maxima and the recombination of monthly extremes can be illustrated by the example of maximum temperatures at Oxford. Assume that annual maxima always occur in June, July or August (this is nearly always the case). An analysis of annual maxima may then be equated to an analysis of maxima in summer, for which N is about 30. The average maximum temperature at Oxford ranges from 18.4 °C to 20.9 °C in June, 20.9 °C to 21.6 °C in July and 19.8°C to 21.5 °C in August (1941–70 figures). Average temperatures, therefore, show a range of 3.2 °C in the three month summer compared to 2.5 °C, 0.7 °C and 1.7 °C in the individual months. The difference between the methods, therefore, lies between a temperature range of 3.2 °C and a sample size N of about 30 for the annual analysis, and a temperature range up to 2.5 °C and a sample size of about 10 for the monthly analysis. The sample size of 10 is so small that the verdict clearly lies in favour of the analysis of annual extremes. Suppose, however, that a sufficiently long time series was available for the highest temperatures in a calendar month to be extracted once every M years. The fact that an annual analysis has a sample size about three times larger than a monthly analysis becomes less important as M increases and eventually the recombination of monthly maxima becomes the best technique.

For estimating values of environmental parameters associated with return periods beyond the length of record, i.e. for purposes of extrapolation, the relative merits of the two techniques can only be assessed by the kind of abstract arguments used above. However, for estimating the values corresponding to small return periods within the length of record, i.e. for purposes of interpolation, the relative merits of the two techniques can be tested against data. This was done by using monthly and annual maxima of mean hourly wind at Scilly from 1927–81. The results, displayed in Fig. 11, show that the estimates of annual maxima obtained by combining distributions of monthly maxima were all too high; similar findings were found when wind data from other stations were analysed. As the technique

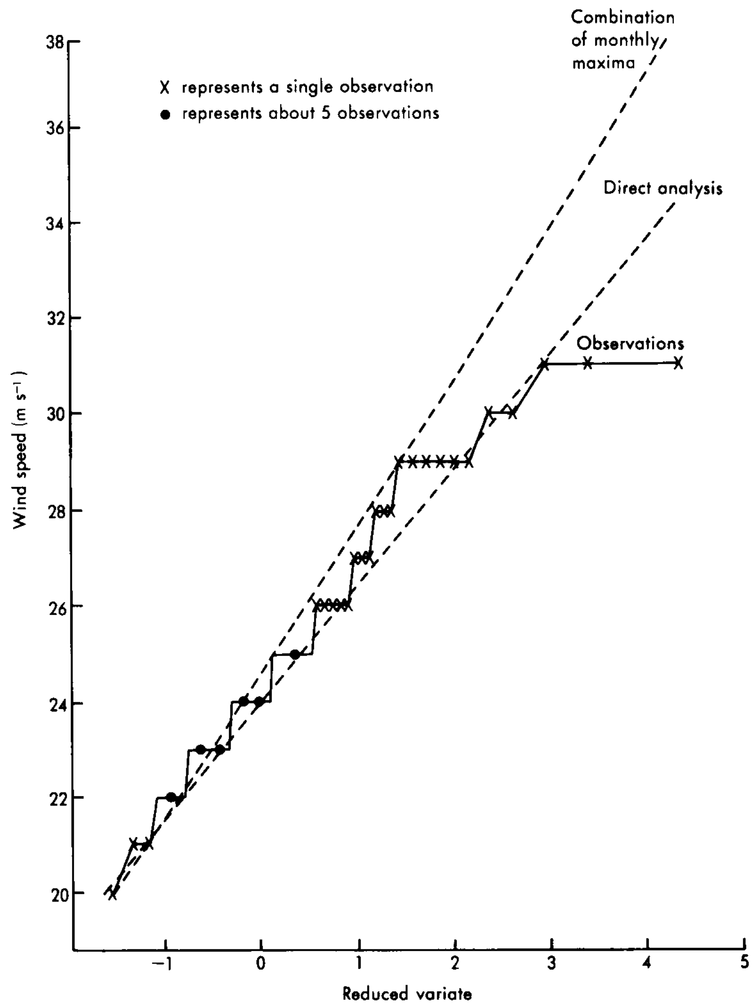


Figure 11. Annual maxima of mean hourly wind at Scilly, 1927–81.

recommended by Carter and Challenor (1981) fails to provide a good representation of observed annual maxima, one can have little confidence in its use for purposes of extrapolation.

6. Populations and forcing factors

The theory outlined in section 2 required that the extremes be drawn from a series of observations which belonged to a single population. This condition may be relaxed to enable the original observations to belong to several populations. As long as the extremes are drawn from just one of these populations, and the sample of original data belonging to that population is large enough, the theory will be satisfied.

Consider a station in a monsoon climate where the winds can be regarded as belonging to two populations associated with the NE and SW monsoons. If the strongest wind in a year is always associated with the SW monsoon, then the series of annual maximum winds can be regarded as belonging to one population, but the appropriate sample size will not be 365, but 180 (say).

Thus application of extreme value theory need not be restricted to cases where observations are drawn exclusively from a single population. Provided all the extremes belong to one population, and the associated sample size is large enough, the original data series may be comprised of observations from a large number of sources.

Meteorological observations may be assigned to different populations according to the external mechanism or forcing factor chiefly responsible for producing the observation. A series of extremes may then be regarded as belonging to the same population if a single forcing factor is responsible for the whole range of extremes examined. In meteorology there are so many degrees of freedom that this condition is rarely completely satisfied. In practice it is reasonable to regard the observations as belonging to the same population if one forcing factor is dominant.

Consider maximum temperatures in summer at a place like Oxford. The mean temperature (thickness) of the lower half of the troposphere may be regarded as the dominant forcing factor. Dynamical subsidence, sunshine, and state of ground are other relevant factors since they determine the mean lapse rate of temperature in the lower half of the atmosphere. If these additional factors vary widely on the hottest days of the year, conventional extreme value analysis may not fit annual maximum temperatures.

Suppose a location could be found that was almost permanently overcast, then the extremes of maximum temperature might follow a type III distribution with thickness as the dominant external factor. However, the observations would be drawn from cloudy days instead of sunny ones; if there was then one sunny day the maximum temperature would be higher than before and the type III curve would no longer provide a good fit to the observations. This would be because sunshine had become a second and important forcing factor.

In their analysis of annual maximum temperatures over the United Kingdom, Hopkins and Whyte (1975) found that a type III curve, fitted to all the data, produced an upper limit (37.5 °C) close to the highest recorded temperature. They considered this to be too low because, on the days which produced the lower maxima, a combination of high thickness and prolonged sunshine may not have been achieved and so these observations may be regarded as belonging to a different population from the majority of the maxima. This form of heterogeneity in annual extremes has previously been suggested by Jenkinson (1969). By combining a 1000–500 mb thickness of 5760 m (30 m above the highest observed in a 30-year record at Crawley) with a lapse rate observed at Cheltenham in July 1976, a maximum temperature of 43 °C seems possible in some places in southern England.

In some locations, especially near coasts, wind direction is another possible forcing factor. Imagine a coastal resort where the maximum temperature is almost always limited by a sea breeze. If on rare occasions a sea breeze fails to develop, much higher temperatures than usual would be observed. A type III curve would not provide a good fit to observations from these occasions.

In Britain, the best examples of the effect of a second forcing factor on temperature are found in places affected by the föhn. Fig. 12 displays a plot of maximum temperatures for January at Aber in North Wales. The majority of observations may be fitted by a type III curve, probably representing occasions when thickness is the dominant factor, but the more extreme points may be regarded as lying on another curve in which the föhn is a forcing factor. The föhn is quite capable of producing the highest temperatures observed (18 °C) since on those occasions temperatures at 900 mb were around 11 °C. As föhns are rare, the events plotted in Fig. 12 will not represent extremes selected from a large sample and there is no question of them satisfying the asymptotic theory of extreme values. The dotted lines sketched through the föhn events in Fig. 12 are therefore purely empirical; a very large amount of data would be needed before all the maxima could be drawn from a large sample of föhn events.

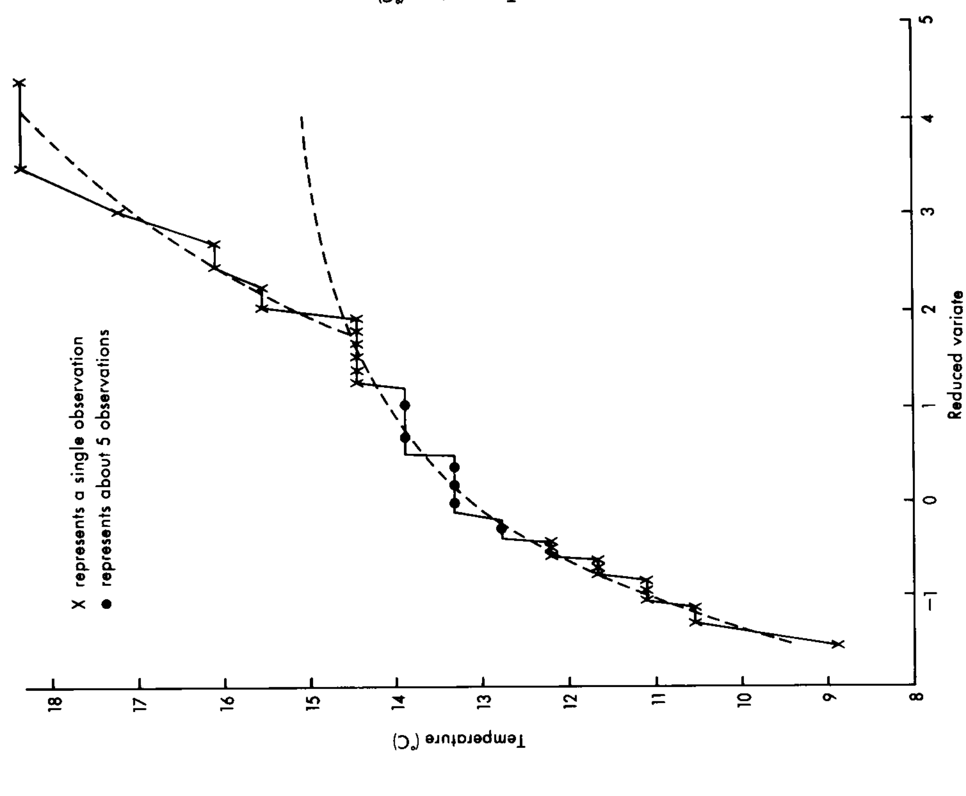


Figure 12. Maximum temperatures for January at Aber, 1925–78.

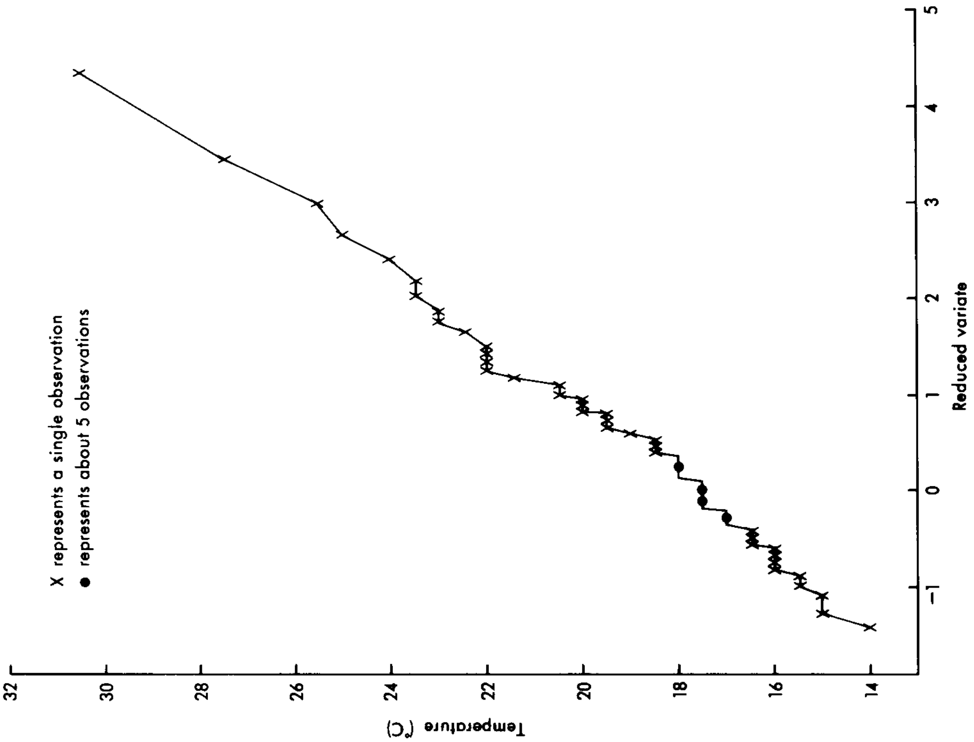


Figure 13. Maximum temperatures for June at Teigarhorn, 1922–75.

Another example of topography introducing additional forcing factors concerns the case of the Sheffield gales of 16 February 1962. These are described by Aanensen (1965) and were caused by standing waves set up by the Pennines. Much stronger winds were observed than if topographic effects *had been absent and so this event belongs to a different population from the majority of gales at* Sheffield.

The outlier in the data for Ivigtut, presented in Fig. 1, was probably caused by a föhn. Temperatures of 30 °C around southern Greenland are quite possible as is evident by a plot of maximum temperatures for June at Teigarhorn in south-east Iceland (Fig. 13). There the appearance of a type I curve is caused by the absence of a dominant forcing factor for many contiguous points, with thickness, wind direction, sunshine and föhn all playing a part.

7. Populations and sample sizes

In general, any set of data can be divided into a number (Q) of populations. As Q increases, the number N of independent cases from which the extremes are selected decreases. Thus while the aim of dividing data into separate populations is to provide a firmer foundation for extreme value analysis, this aim is negated by a decrease in N. The failure to satisfy extreme value theory is really due to one cause — lack of data. The two reasons advocated so far (mixed populations and insufficiently large N), are often different ways of expressing the same problem.

Consider the example of maximum temperatures for January at Oxford (Fig. 5). In section 3, the lack of homogeneity in the data was expressed by saying that N was too small. The lower values of the maxima were clearly not being drawn from the tail of the complete distribution. There is another way of looking at the problem. The majority of observations in Fig. 5 will be associated with incursions of tropical maritime air across the country. These can be regarded as constituting one population. The lower January maxima in Fig. 5 are clearly not associated with tropical maritime air and can, therefore, be regarded as belonging to a different population.

In section 6, Hopkins and Whyte's (1975) analysis of annual maximum temperature was discussed. The failure of the type III distribution to provide sensible extrapolations was attributed to the lower maxima belonging to a different population from the majority. An alternative way of expressing the problem is as follows. Hopkins and Whyte show that the lowest annual maximum at Oxford is 23 °C. Since the average maximum in July is around 22 °C, an observation of 23 °C is clearly not being drawn from the tail of the parent distribution. Hence the failure to obtain a good extreme value analysis can be ascribed to an insufficiently large N. By extracting maxima every five years, and so increasing N, the lowest maxima is raised to 29 °C and the maxima examined are much more representative of the tail of the parent distribution.

Similarly, consider a plot of extreme winds at a place like Progreso (Fig. 9) where hurricanes are a feature of the climate. The immediate reaction is to ascribe the lack of a good fit to a general extreme value distribution to the presence of two populations, i.e. those winds due to hurricanes, and those due to other causes. Suppose, however, that sufficient data were available for a long series of centennial (as opposed to annual) maxima to be extracted; then all the extreme events would be caused by hurricanes, and a good extreme value analysis for a single population could be obtained. It follows that the plot in Fig. 9 may be regarded as an inadequate sampling of hurricanes, i.e. the lack of fit to a general extreme value distribution is caused by a too small value of N.

The formation of a combined probability distribution from a number of populations (as indicated in section 4) is only valid if the populations are independent. Determining the independence of populations may not be easy and many of the different populations described above probably could not be considered independent. In any analysis, therefore, it is sensible to keep the number of populations to a minimum and, where possible, to regard the data as belonging to a single population. In general, this aim will be furthered by the choice of as large a value of N as possible (e.g. five-year maxima instead of annual maxima).

8. Rainfall and the type II distribution

Annual extremes of daily and hourly rainfall are frequently best fitted by a type II distribution; this has always caused problems of interpretation. In the *Flood Studies Report* (Natural Environment Research Council, 1975), Jenkinson groups observations according to the magnitude of the fall which has a return period of five years. He shows that the greatest departures from a type I distribution occur for five-year falls of 20 mm in England and Wales and 15 mm in Scotland and Northern Ireland. These falls correspond to a duration of about an hour, a typical duration for thunderstorms. The departure from a type I distribution is also greater in England and Wales than in Scotland or Northern Ireland. These facts suggest that the type II appearance of the observations may be related to the behaviour of individual convective storms.

Warrilow (1981) has shown that by taking the distribution of storm movement into account modest rainfall extremes which belong to a type III distribution following the storm (Lagrangian) are converted to a type II distribution when observed at a point (Eulerian). This is likely to account for most of the type II behaviour of observed rainfall extremes. Other possible contributory factors are as follows:

(i) The complete distribution of short-duration rainfall displays large positive skewness, so any failure to satisfy extreme value theory due to insufficiently large N will result in a concave upward distribution of the lesser extremes.

(ii) Most of the larger extremes will be due to thunderstorms, but in many places some of the lesser extremes may be due to frontal rainfall. The mixture of populations would then give rise to a type II appearance to the observations. As convection will be dominant in the heaviest frontal rainfalls as well as in thunderstorms, however, the distinction between the two may not be as great as at first appears. Some of the heaviest rainfalls have been frontal in origin, but have contained embedded thunderstorms. If consideration is restricted to convective storms, however, it is true that as we pass from the lesser to the greater extremes, the organizational structure of storms also changes, from single cell through multicell to supercell. Hence the increasing organizational structure of storms as we pass from small to large return periods is likely to contribute to the type II distribution of rainfall extremes.

(iii) In certain areas, topography may encourage the development of stationary storms and give rise to a distribution of storm movements different from that considered by Warrilow. In districts thus affected, the conversion from the Lagrangian to the Eulerian frame of reference will result in some spectacular type II curves, and very large point rainfalls may have relatively modest return periods. In the United Kingdom, some of the large storms that have occurred in south-west England, together with the Hampstead storm of 1975, may enter this category.

9. The upper bound

Many sets of meteorological extremes are well fitted by the type I distribution, but this is unbounded above. Where physical considerations impose bounds, the highest extremes may not belong to the same population as the more modest extremes and will be represented by a type III distribution.

For events related to the duration of a single physical entity, e.g. a thunderstorm, a gale, or an afternoon maximum temperature, it is clear than an upper limit to extremes must exist. For longer duration events, e.g. monthly rainfall, which involve a succession of physical entities, a realistic physically imposed upper limit is more difficult to visualize and evaluate. Most practical applications are concerned with short duration events for which the concept of an upper limit is valid, and consideration is now restricted to these cases.

The return period at which extremes approach the upper limit varies widely with element. Over the United Kingdom, for instance, maximum temperatures appear to approach their upper bound for return periods around 100 years, while for rainfall Jackson (1979) shows this does not happen until return periods of the order of a million years are reached. For a given element, the return period at which the upper bound is approached will also vary from place to place.

Consider maximum temperatures and compare typical inland and coastal sites. For any given return period, the maximum temperature on the coast will be lower than that inland. The upper limit on the coast, however, may be the same, or nearly the same, as inland. Although optimum conditions will be more rare, it is still possible to visualize a set of conditions in which the highest coastal temperatures would be nearly the same as those inland. The situation is illustrated schematically in Fig. 14. The inland station is represented by curve A, and the coastal resort by curve B. The upper limits at both locations are similar, but the return period at which this is approached is larger on the coast.

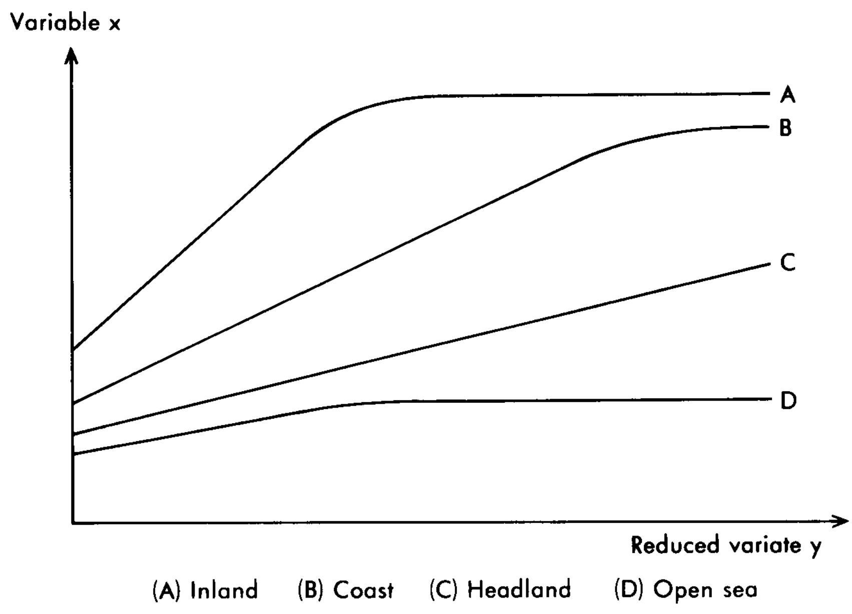


Figure 14. Schematic extreme value analysis of temperature.

Now consider the change from a linear stretch of coastline to a headland. For any given return period, the headland will experience lower temperatures than the remainder of the coast and the highest temperatures experienced inland may never be reached. The slope of the extreme value analysis, although smaller than that for other inland or coastal locations, may well remain linear for longer return periods than in the previous two cases. This is illustrated by curve C in Fig. 14. Over the open sea (curve D), however, the probable maximum temperature will be much lower than over the land and will probably be approached at a similar (modest) return period. These ideas are illustrated using real data (as far as is possible) in Fig. 15.

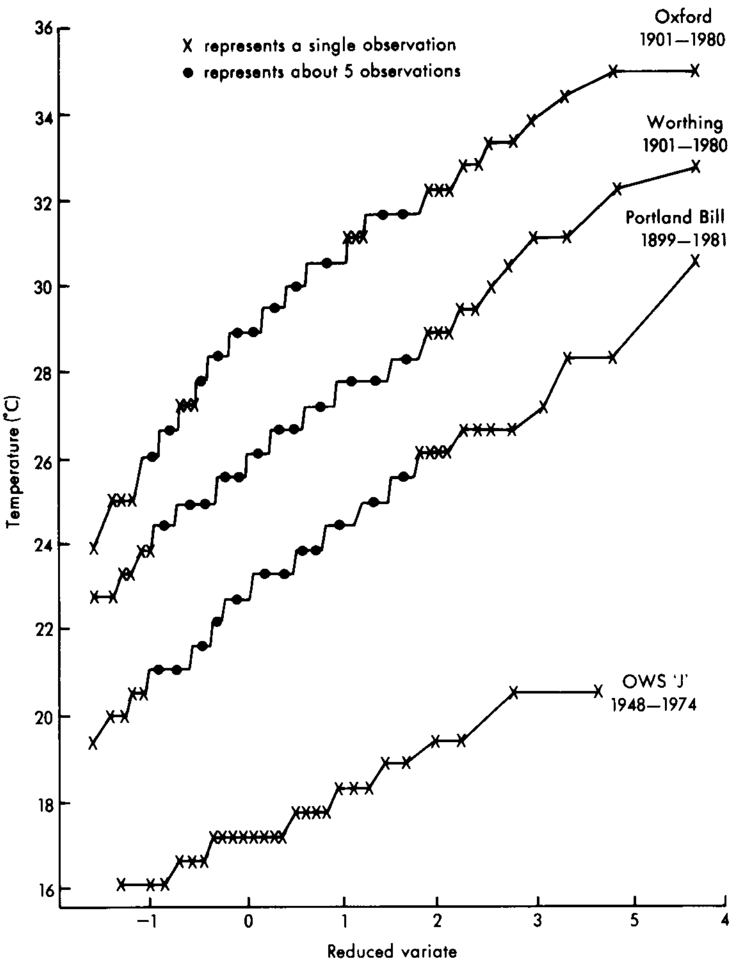


Figure 15. Annual maximum temperatures at Oxford, Worthing, Portland Bill and OWS'J'

The same arrangements may be extended to other elements. In the case of wind, for instance, curve A may represent places like the Faeroes, where intense depressions are frequent; curve B may then represent places further South, e.g. Valentia, where deep depressions are less frequent but still possible. In the case of rainfall, districts with frequent thunderstorms will be represented by curve A while curve B represents places where they are less frequent, but still possible.

When a set of extremes fall well short of their upper limit a highly skewed parent distribution and an inadequate sampling of limiting physical conditions are indicated. The lack of data for extreme value analysis is then acute. If the observed extremes are believed to belong to the same population as those near the upper limit, it can be argued that the observed extremes are not being drawn from the tail of the single population. Alternatively, if it is argued that the observed extremes are being drawn from the tail of one population, then there are grounds for thinking that the highest possible extremes belong to another population. Using either argument, the observed extremes are likely to display the appearance of being unbounded above.

10. Practical considerations

There are two main approaches to the estimation of extreme events:

(i) Analysis of the tail of the parent distribution.

(ii) Direct analysis of the extremes.

When all the observations belong to a single population, a mathematical distribution may provide an exact fit to the original observations and both techniques will then give the same results. An example is the representation of Brownian motion by the normal distribution. In these circumstances, the only reason for using extreme value theory is the convenience of data analysis. The fitting of a parent distribution involves handling many observations, not all of which may be available. If a series of extremes are available, it is much easier to analyse them.

When a set of observations are not identically distributed, direct analysis of the extremes will give different results from an analysis of the tail of the parent distribution. In these circumstances, a mathematical distribution is most unlikely to provide an exact fit to the original observations. The main body of observations may be fitted reasonably well, but the tails will be poorly represented. Under these conditions, a direct analysis of extremes is clearly indicated, but it is in just these circumstances that the theory of extreme values is not satisfied. The dependence of a series of observations on more than one forcing factor is therefore responsible both for providing a good reason for performing a direct analysis of the extremes, and for ensuring that the theoretical assumptions on which such an analysis is based are not satisfied.

It has been shown that in meteorology there are usually a large number of forcing factors operating on a given set of observations. Sometimes it is possible to divide the observations into two or more categories each belonging to separate populations, but it is more usual for there to be a gradual change in the forcing factors and their relative importance across the spectrum of observations. The advantage of extreme value analysis is that it restricts the range of observations under consideration and thereby limits the changes in the forcing factors involved. The greater the value of N that can be used, the more restricted the range of extremes considered and the more likely it is that those selected can be regarded as belonging to one population.

When an extreme value analysis is performed on data from mixed distributions, any extrapolation lacks theoretical justification. Any results based on interpolation, however, are likely to be the best obtainable and will certainly be superior to those derived from the tail of a fitted parent distribution. The larger the value of N that can be used, the more reliable will be any limited extrapolation that is attempted.

In any practical application, however, a balance has to be struck between the number of observations from which the extremes are selected and the number of those extremes contained in the available record. Although an increase in N may reduce the systematic error, the smaller number of points available for analysis will increase the random error and, with the record lengths commonly found in meteorology, this is likely to be important. Volume I of the *Flood Studies Report* (Natural Environment Research Council, 1975) gives the standard errors (SE) associated with fitting a type I distribution as,

$$\mathrm{SE}\,(x_0) = 1.05\alpha/\sqrt{M},$$
$$\mathrm{SE}\,(\alpha) = 0.78\,\alpha/\sqrt{M},$$
$$\text{and } \mathrm{SE}\,(X) = \sqrt{(1.11 + 0.52Y + 0.61Y^2)}\,\alpha/\sqrt{M},$$

where x_0 is the intercept, α the slope, M the number of extremes and Y the value of the reduced variate corresponding to an estimate X of the variable under analysis. The second of these equations shows that if M is as small as ten then the standard error of the slope is as much as 25% of the value of the slope. It is an inescapable fact that extreme value analysis is a technique which demands a large amount of data for its successful implementation.

11. Conclusions

The theory of extreme values assumes that the maxima (or minima) are drawn from infinitely large samples of independent observations that belong to a single population. Failure to satisfy the theory may therefore be due to selecting extremes from too small a sample or the inclusion of observations which belong to more than one population. These reasons, in meteorology, are alternative ways of expressing the same problem, namely lack of data.

Most meteorological variables undergo a pronounced seasonal variation and consequently the observations cannot be regarded as coming from the same population. The problem may be tackled by regarding monthly maxima as belonging to separate populations and then combining them to obtain a distribution of annual extremes. This approach, however, is likely to be compromised by the inclusion of insufficiently extreme monthly observations and it may be better to perform a straightforward analysis of annual maxima.

In meteorology there are so many physical processes involved in the creation of a series of observations that defining a single population is difficult. In practice, a set of data may be regarded as belonging to the same population if a single forcing factor is primarily responsible for the range of extremes encountered. Topography can often act as a second forcing factor. It can cause high temperatures or strong winds through föhns and standing waves and produce high point-rainfalls by encouraging the development of stationary storms.

The return period at which extremes approach a physically imposed upper limit varies widely with element and location. When a set of maxima lie close to the physically imposed upper bound, a type III distribution of the extremes may be expected, but when the observed extremes fall well short of their upper limit, they may appear to be unbounded above. For short duration rainfall this may be interpreted as being caused by changes in the structure of convective storms as we pass from the lesser to the greater extremes.

Acknowledgements

Acknowledgements are due to D. A. Warrilow, G. H. Ross and F. Singleton for constructive discussions.

References

Aanensen, C.J.M. 1965 Gales in Yorkshire in February 1962. *Geophys Mem, Meteorol Off*, No. 108.

Beard, L.R. 1943 Statistical analysis in hydrology. *Trans Am Soc Civ Eng*, **108**, 1110–1160.

Carter, D.J.T. and Challenor, P.G. 1981 Estimating return values of environmental parameters. *Q J R Meteorol Soc*, **107**, 259–266.

Cook, N.J. 1982 Towards better estimation of extreme winds. *J Wind Eng and Ind Aerodyn*, **9**, 295–323.

Cunnane, C. 1978 Unbiased plotting positions — A review. *J Hydrol*, **37**, 205–222.

Fisher, R.A. and Tippett, L.H.C. 1928 Limiting forms of the frequency distribution of the largest or smallest members of a sample. *Proc Cambridge Philos Soc*, **24**: 180–190.

Galambos, J. 1978 The asymptotic theory of extreme order statistics, New York, John Wiley and Sons Inc.

Gomes, L. and Vickery, B.J. 1977 Extreme wind speeds in mixed wind climates. *J. Indus Aerodyn*, **2**, 331–344.

Gumbel, E.J. 1958 Statistics of extremes. New York, Columbia Univ Press.

Hardman, Carol E, Helliwell, N.C. and Hopkins, J.S. 1973 Extreme winds over the United Kingdom for periods ending 1971. *Climatol Mem, Meteorol Off*, No. 50A.

Hopkins, J.S. and Whyte, K.W. 1975 Extreme temperatures over the United Kingdom for design purposes. *Meteorol Mag*, **104**, 94–102.

Jackson, M.C. 1979 The largest fall of rain possible in a few hours in Great Britain. *Weather*, **34**, 168–175.

Jenkinson, A.F. 1955 The frequency distribution of the annual maximum (or minimum) values of meteorological elements. *Q J R Meteorol Soc*, **81**, 158–171.

1969 Statistics of extremes. *In* Estimation of maximum floods, WMO *Tech Note* No. 98, Chapter 5, 183–257.

1977 The analysis of meteorological and other geophysical extremes. (Unpublished, copy available in National Meteorological Library, Bracknell.)

Kendall, M.G. and Stuart, A. 1977 The advanced theory of statistics, Vol. 1. London, Charles Griffin and Co. Ltd.

Natural Environment Research Council 1975 Flood Studies Report, Vols I and II, London.

Warrilow, D.A. 1981 The interpretation of rainfall data and its application to the design of hydrological structures. (Unpublished, copy available in National Meteorological Library, Bracknell.)

Weibull, W. 1939 A statistical theory of strength of materials. *Ing Vet Ak Handl* No. 151. Stockholm.

Appendix 1 — Plotting positions on extreme value probability paper

Intuitively, one expects a return period of about M to be associated with the largest of a series of M observations. This thinking is expressed by ascribing a cumulative probability p to the mth ranking observation of

$$p = \frac{m}{M+1} \qquad \text{(A1)}$$

This formula was first suggested by Weibull (1939) and was popularized by Gumbel (1958). The data used in this paper were plotted according to the relation

$$p = \frac{m - 0.31}{M + 0.38} \qquad \text{(A2)}$$

This equation was first proposed by Beard (1943) and has been widely used in the Meteorological Office following its adoption by Jenkinson (1969).

If 100 years of data are available then to the first ranking observation is attributed a return period of 101 years by equation (A1) but 145 years by equation (A2). The difference between the two is essentially the difference between the mean and the median.

If 1000 years of data are available then the event with a return period of 100 years may be approximated by the value of the 10th ranking observation. The distribution in time of these 10 largest events is not uniform. Their separation is bounded below by one and so a positively skew distribution emerges in which the median separation is less than the mean. Thus, while the mean separation of the 10 largest events will be 100 years the median separation will be less than this. Now the largest event in 100 years of data lies close to that whose median recurrence interval is 100 years. It can be shown that the mean return period of such an event is 145 years; this is the result given by equation (A2).

It can now be seen that it is the skewed distribution of the separation of the most extreme events that causes the largest observation in M to have a return period greater than M. The Weibull formula will only be correct if the largest events are uniformly distributed in time. An excellent review of plotting positions in general is given by Cunnane (1978) who recommends the use of the relation

$$p = \frac{m - 0.4}{M + 0.2}$$

The differences in the plotting positions given by the above equations are generally small except for the largest extreme. In the case of a set of extremes which fitted the type I distribution, use of the Weibull plotting positions result in a slight 'type II' appearance of the observations with the position of the largest event having the greatest error.

Appendix 2 — Data

A = Maximum temperatures in June at Ivigtut (°C) — Fig. 1.
B = Maximum temperatures in January at Oxford (°F) — Fig. 5.
C = Maximum temperatures in August at Santander (°C) — Fig. 6.
D = Annual maximum gusts at Progreso (m s^{-1}) — Fig. 9.
E = Maximum temperatures in January at Aber (°F) — Fig. 12.
F = Maximum temperatures in June at Teigarhorn (°C) — Fig. 13.
G = Annual maxima of mean hourly wind at Durham (m s^{-1}) — Fig. 10.
H = November maxima of mean hourly wind at Durham (m s^{-1}) — Fig. 10.
I = Annual maxima of mean hourly wind at Scilly (m s^{-1}) — Fig. 11.
J = Annual maxima of temperature at Oxford (°F) — Fig. 15.
K = Annual maxima of temperature at Worthing (°F) — Fig. 15.
L = Annual maxima of temperature at Portland Bill (°F) — Fig. 15.
M = Annual maxima of temperature at OWS 'J' (°C) — Fig. 15.

(Bracketed figures are estimates.)

	Year																			
	1871	1872	1873	1874	1875	1876	1877	1878	1879	1880	1881	1882	1883	1884	1885	1886	1887	1888	1889	1890
A					17.4	19.6	21.5	20.0	19.0	18.8	18.0	19.2	18.5	13.5	18.0	16.6	18.3	20.5	15.2	21.3
B	46	53	54	54	55	53	56	55	47	53	51	55	55	55	52	52	52	54	51	56

	1891	1892	1893	1894	1895	1896	1897	1898	1899	1900	1901	1902	1903	1904	1905	1906	1907	1908	1909	1910
A	21.2	20.7	17.4	16.1	21.9	18.0	23.0	20.0	16.9	19.9	21.0	20.6	20.6	19.1	18.4	19.4	17.4	17.8	22.2	17.7
B	52	51	53	54	51	53	50	55	54	53	52	53	54	54	55	56	52	55	50	53
J											89	84	84	85	81	92	79	84	85	79
K											81	78	78	77	77	79	76	80	81	74
L									74	75	79	73	78	75	73	71	74	74	76	72

	1911	1912	1913	1914	1915	1916	1917	1918	1919	1920	1921	1922	1923	1924	1925	1926	1927	1928	1929	1930
A	17.3	19.9	18.9	14.1	30.1	—	—	16.6	—	19.4	15.2	13.7	18.6	20.2	17.4	21.4	—	20.0	21.7	16.1
B	55	51	52	55	52	57	52	55	52	55	55	57	54	52	55	53	55	56	53	58
C																	30	31	27	38
D											17.2	25.6	17.8	24.4	19.4	19.4	16.1	19.4	18.9	23.3
E															60	60	54	56	63	61
F												16.8	20.3	17.8	23.3	25.2	19.9	17.5	21.5	20.2
I																	30	26	30	29
J	95	87	83	87	82	83	89	84	86	79	89	86	93	86	85	85	80	87	87	89
K	88	84	79	78	77	77	79	78	79	76	87	78	86	75	80	83	78	82	77	83
L	83	79	74	76	74	74	70	73	70	68	81	73	83	68	75	78	70	76	73	74

	1931	1932	1933	1934	1935	1936	1937	1938	1939	1940	1941	1942	1943	1944	1945	1946	1947	1948	1949	1950
A	19.4	20.2	18.2	16.5	19.4	21.1	16.0	17.4	18.4	16.1	18.6	17.3	17.6	18.0	19.3	16.2	23.1	19.6	19.2	21.0
B	52	55	53	54	54	57	53	54	55	51	48	50	55	56	50	57	56	57	53	54
C	26	32	28	26	26	27	35	23	24	25	24	23	40	30	26	30	31	29	24	27
D	19.4	22.2	27.8	23.3	22.2	22.2	22.2	33.3	20.6	22.8	22.8	19.4	21.1	20.1	18.9	22.2	21.1	34.0	32.0	48.0
E	51	57	56	55	53	57	58	57	57	56	51	52	57	57	52	55	56	56	56	61
F	20.3	23.6	22.1	23.9	20.6	27.8	20.9	22.7	30.5	22.2	25.7	18.7	19.7	23.0	18.0	17.3	17.0	18.1	22.0	16.0
G								18	15	20	18	17	20	18	19	17	21	22	20	20
H								14	14	20	13	12	16	16	9	13	14	15	13	14
I	24	23	22	26	29	27	23	25	29	24	23	21	27	25	29	31	26	22	25	26
J	75	95	89	86	86	83	87	84	84	85	89	88	92	87	84	84	90	90	89	86
K	77	80	81	82	84	81	77	83	82	80	83	83	80	81	80	76	90	88	81	81
L	74	75	78	80	80	71	74	72	75	73	70	74	67	76	69	(72)	(80)	(77)	—	—
M																		17.2	18.3	16.7

	1951	1952	1953	1954	1955	1956	1957	1958	1959	1960	1961	1962	1963	1964	1965	1966	1967	1968	1969	1970
A	21.7	18.6	20.0	16.5	20.5	19.8	19.6	17.8	18.0	18.8										
B	53	52	55	56	53	54	57	55	52	54	53	55	41	54	53	54	55	55	56	51
C	30	29	35	29	29	25	30	26	28	34	30	25	23	34	27	27	25	30	24	25
D	42.0	38.0	17.0	41.0	32.0	44.0	36.0	36.0	31.0	19.4	16.7	18.9	19.4	18.3	22.8					
E	53	53	54	57	57	56	58	65	55	58	55	55	48	57	55	57	57	55	58	58
F	17.5	16.6	18.6	18.7	18.1	19.2	17.7	16.2	17.5	24.1	17.0	17.9	22.2	18.0	17.0	14.3	16.5	18.0	15.2	16.2
G	17	16	18	19	18	21	20	20	17	—	—	22	15	19	21	19	21	23	15	24
H	14	16	15	18	8	18	12	9	14	—	11	12	14	14	18	16	13	15	14	24
I	28	23	25	31	28	25	22	26	25	23	23	29	29	21	27	28	23	22	24	24
J	82	89	90	83	87	82	86	80	91	82	87	77	81	87	81	80	83	89	88	88
K	76	82	80	75	84	77	82	77	82	80	79	73	80	78	73	81	79	82	80	78
L	—	(76)	76	71	76	70	74	70	77	78	71	70	73	77	70	77	71	75	73	75
M	17.2	18.3	16.8	17.2	20.5	17.3	18.2	17.8	19.5	17.7	17.2	17.8	16.0	18.7	19.2	18.7	16.9	20.5	16.4	17.5

	1971	1972	1973	1974	1975	1976	1977	1978	1979	1980	1981
C	28	28	32	27	28	30	24	24	27	30	
E	65	56	56	58	58	55	54	52			
F	15.5	19.5	15.0	16.2	15.5						
G	18	20	19	18	17	19	17	18	16		
H	15	20	19	14	13	13	13	17			
I	20	24	24	26	25	24	24	25	31	24	23
J	84	77	85	77	91	94	80	80	85	83	
K	82	79	85	78	85	91	82	74	77	76	
L	79	76	79	75	80	87	77	70	74	72	74
M	17.0	18.4	17.3	16.0							

551.509.323

Forecasting urban minimum temperatures from rural observations

By J. Roodenburg

(Royal Netherlands Meteorological Institute, de Bilt)

Summary

A regression formula has been derived from which the nocturnal minimum temperatures in an urban area may be computed from meteorological variables observed at a nearby airport.

1. Introduction

Many branches of economic activity take an interest in reliable minimum temperature forecasts, especially in an era of increasing energy costs.

For the heavily industrialized and densely populated conglomeration of Rotterdam a minimum temperature forecast is issued every late afternoon valid for the next night.

Until recently, the approach followed by the Regional Weather Office at nearby Zestienhoven ('Sixteen Farms') Airport was to adapt subjectively the official forecast from the Central Weather Office at de Bilt. Often this led to disappointing results, mainly for two reasons:

(a) in the Netherlands, as probably almost everywhere, minimum temperature forecasts traditionally have been verified against observations from well-exposed rural stations, and

(b) no objective or semi-objective method existed as to how the forecast issued centrally should be adapted in order to yield reliable results for an urban site.

The present paper presents a statistical method that includes several simple meteorological variables; as such it is an extension of earlier work done in England (Gordon *et al.*, 1969), in which noon temperatures only were taken into account.

2. Geographical description and data

Fig. 1 shows a sketch of the Rotterdam residential and business area (shaded), the industrial and harbour area (hatched) as well as the location of the urban site (encircled cross). Zestienhoven Airport is shown in the upper left corner.

The urban minimum temperatures were taken from a thermograph placed in a Stevenson screen in a 70 m^2 garden. The garden is enclosed by buildings on three sides, the fourth side faces the river Maas about 300 metres away.

There is some doubt as to the representativeness of these temperatures, but as no other data were available, representativeness had to be assumed.

All other meteorological information was taken from the records of Zestienhoven Airport; the period covered comprises January 1971 up to December 1973 inclusive. The period July 1980 – June 1981 was used as independent material.

All minimum temperatures in this paper refer to the period 1800 – 0600 GMT.

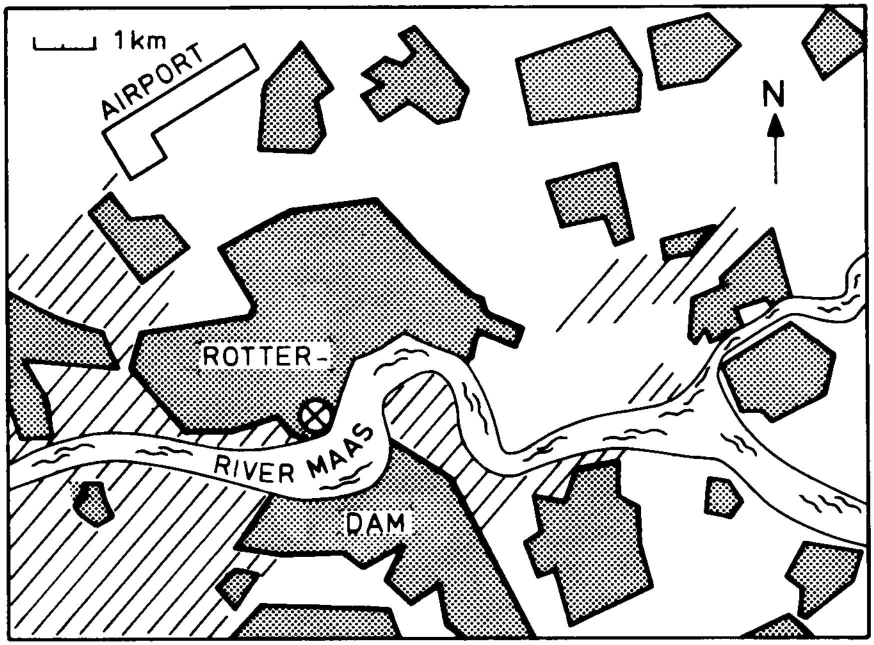


Figure 1. Plan of conglomeration of Rotterdam with urban site (encircled cross) and airport indicated. Shaded: residential and business area. Hatched: industrial and harbour area.

3. A 'first guess' urban minimum temperature

Without clouds, without advection and with neglect of the effect of heat absorbed during daytime, the minimum temperature would solely depend on the maximum temperature of the previous day, the period of time available for cooling and the atmosphere's transparency to long-wave radiation. The latter factor strongly depends on the atmosphere's water vapour content. It seemed a sensible first step, therefore, to link the minimum temperature to some combination of an afternoon temperature and a moisture parameter. This combination would at the same time reflect the length of the cooling period, since high afternoon temperatures and short nights go together. Gordon *et al.* (1969) essentially followed the same train of thought.

The sum of temperature and dew-point, observed at Zestienhoven Airport at 1500 GMT (henceforth referred to as *SUM*) was chosen on the following grounds:

(a) both temperatures are readily available in operational surroundings,

(b) 1500 GMT is normally close to the time at which the maximum temperature occurs,

(c) the dew-point gives an indication of the availability of moisture in the lower layers of the atmosphere, and

(d) if *SUM* is kept constant, various combinations of temperature and dew-point lead to approximately the same potential wet-bulb temperature (θ_w) as is easily demonstrated on a thermodynamic diagram.

As θ_w is conservative for adiabatic processes, it acts as an airmass identifier which thus, in a broad sense, also applies to *SUM*. Consequently, *SUM* should, under the assumptions mentioned above, be well correlated with the subsequent minimum temperature and thus provide a 'first guess'. The next step would be to apply corrections to this first guess in accordance with the actual (or forecast) deviations from these assumptions.

4. Systematic errors in the first guess minimum

A regression equation was derived to obtain the first guess urban minimum temperature:

$$\hat{T} = -0.33 + 0.439SUM \qquad \ldots \quad \ldots \quad \ldots \qquad (1)$$

where $\hat{T}$ is calculated temperature. The correlation coefficient (r) was 0.947, the root-mean-square (r.m.s.) error was 1.74 °C. This result is somewhat better than that achieved by Gordon *et al.*, who in rural surroundings obtained r = 0.89 and r.m.s. error = 2.3 °C. As the variability of minimum temperature is known to be much higher in rural than in urban areas (Böhm and Gabl, 1978) this is not surprising.

Application of equation (1) to independent material (July 1980 – June 1981, every tenth day) yielded Fig. 2. Scatter is considerable. This was to be expected because the ideal no advection and no cloud conditions, assumed in the preceding section, are seldom met in nature. Moreover, the thermal inertia of a large built-up area was not accounted for (Oke and Maxwell, 1975).

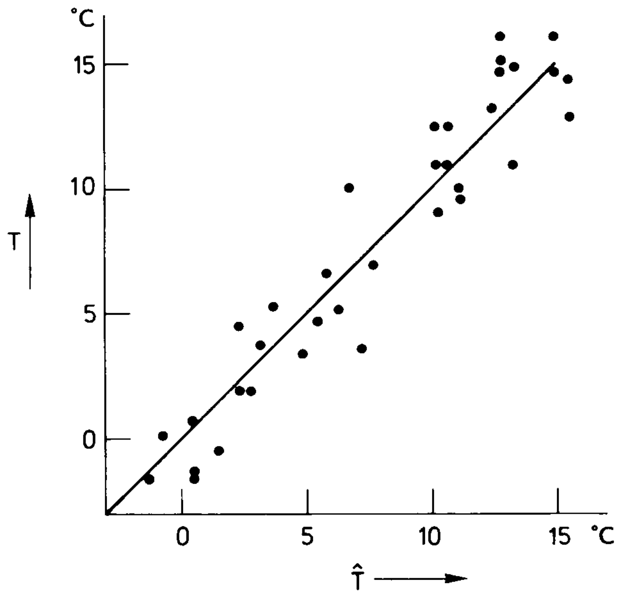


Figure 2. $\hat{T}$, minimum temperature calculated from equation (1), versus T, the observed minimum temperature.

In order to gain appreciation of the influence of cloud amount and wind speed Figs 3 and 4 were drawn up. They depict the average departures $< \hat{T} - T >$, in which T is the observed urban minimum temperature, as a function of cloud amount (oktas) and wind speed in five groups: ⩽3 kn, 4 – 6, 7 – 10, 11 – 16 and 17 – 21 kn respectively. Fig. 3 clearly shows that the calculated minimum temperature is on the average too warm on clear nights and too cold on cloudy nights. Cloud amounts were taken from the 0300 GMT observations made at Zestienhoven. Likewise Fig. 4 indicates that the calculated temperature is too high on nights with little wind and vice versa. Again the 0300 GMT Zestienhoven observations were used.

The influence of wind direction is demonstrated in Fig. 5. Here relative cumulative frequencies have been plotted of deviations from the calculated minimum temperature in excess of 1.5 °C. It can be seen that the number of too warm forecasts grows rapidly with wind directions between 010 and 100 degrees. The number of too cold forecasts shows a similar behaviour with winds between 200 and 290 degrees. Furthermore it was found that for any direction there was a seasonal variation as well. Fig. 6 gives average deviations for nine direction groups for February and August. Notwithstanding the fairly large scatter around the straight lines (obtained by linear regression), the seasonal influence is unmistakable.

5. Results of a multiple regression analysis

From the observations in the preceding section it was clear that more variables than *SUM* had to be considered as potential predictors. Apart from cloud cover and wind the seasonal effect, as illustrated by Fig. 6, would also have to be accounted for.

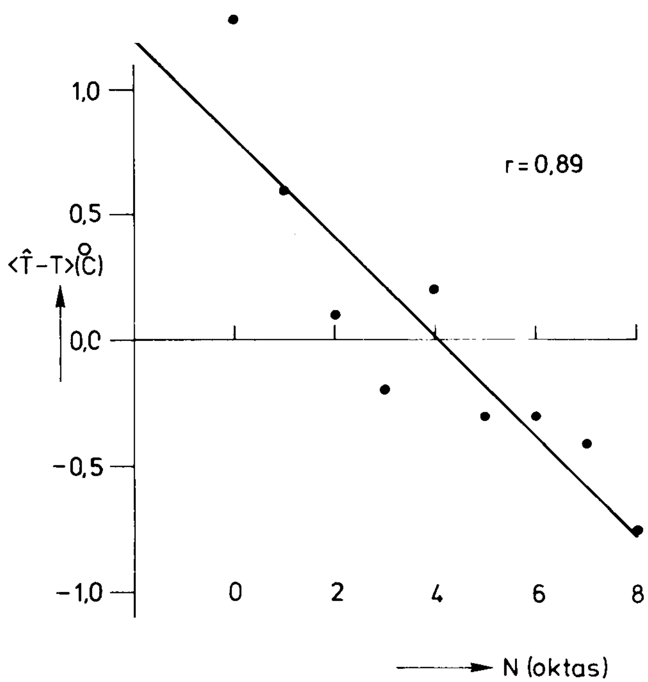


Figure 3. Average deviations from $\hat{T}$ as a function of cloud cover.

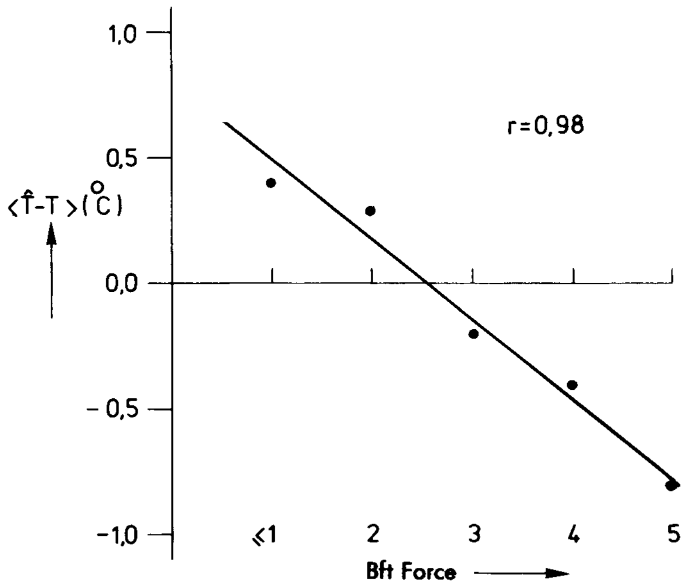


Figure 4. Average deviations from $\hat{T}$ as a function of wind speed.

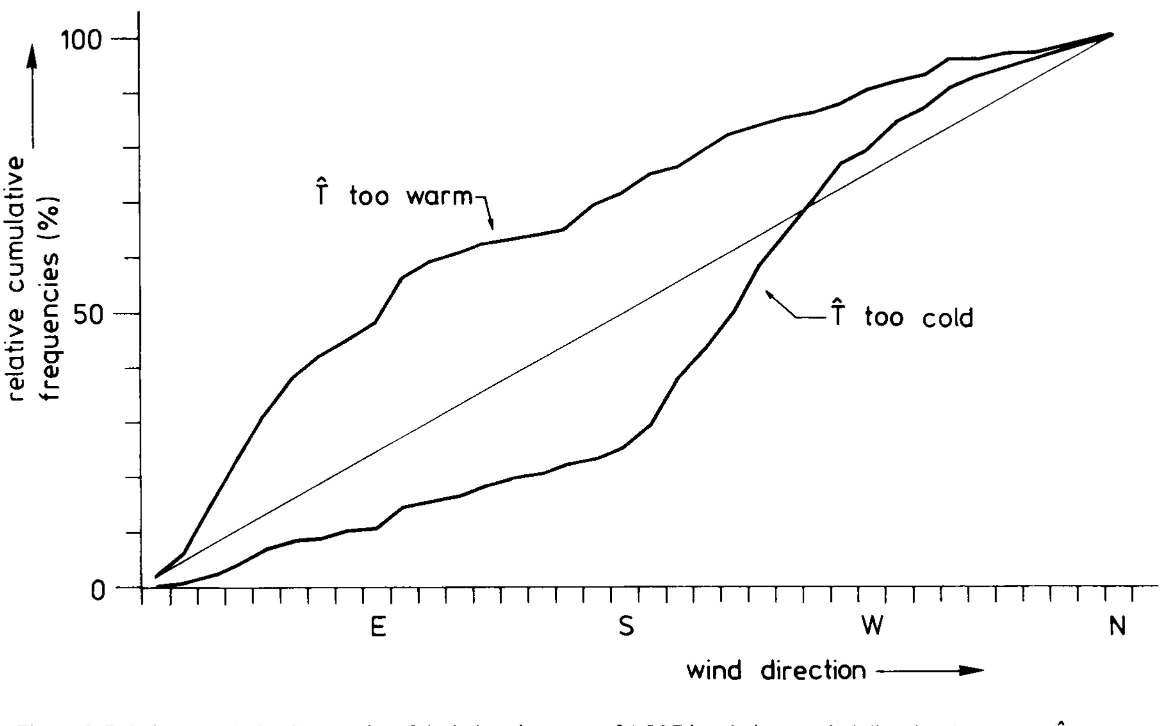


Figure 5. Relative cumulative frequencies of deviations in excess of 1.5 °C in relation to wind direction (top curve: $\hat{T}$ too warm, bottom curve: $\hat{T}$ too cold).

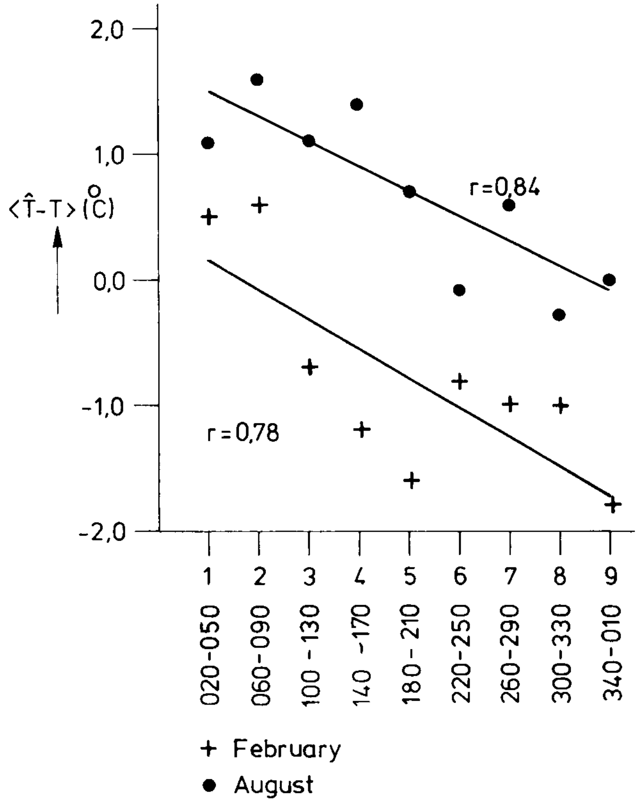


Figure 6. Average deviations from $\hat{T}$ for nine direction sectors for February and August.

The wind data were partitioned into three direction groups: 010–100, 200–290 degrees and remaining directions and into five speed groups: ≤6 kn, 7 – 10, 11 –16, 17 – 21 and 22 or more kn. If the wind speed was six knots or less, the direction was neglected. The groups into which the 0300 GMT wind fitted were assigned a value of 1, the remaining groups a value of 0.

To correct for the seasonal influence, after some experimenting, the best fit (in a least-squares sense) was obtained by a variable S, where

$$S = \sin\{(m - 4.5)2\pi/12\}$$

where m is the number of the month (January = 1, etc.)

Submission of the complete data-set (January 1971 – December 1973) to a forward stepwise regression scheme gave the results listed in Table I. It is clear from Table I that the first four variables noticeably contribute to the reduction of variance. Therefore only these variables were retained in the regression analysis. The following equation resulted:

$$\hat{T} = 0.62 + 0.36SUM + 1.73S + 0.17N - 1.06Y. \qquad (2)$$

(r = 0.964, r.m.s error = 1.43 °C; see Table I for the meaning of the various symbols.)

It is interesting that wind speed does not appear as an independent variable in equation (2) despite the effect demonstrated in Fig. 4; this is because wind speed and cloud cover are correlated to such an extent that only the latter is picked out by the regression analysis.

Table I. *Description and performance of variables*

Symbol	Description	Value assigned	Correlation coefficient	Total explained variance (%)
SUM	Temperature + dew-point (1500 GMT)	As observed	94.7	89.6
S	Sin $\{(m-4.5)2\pi/12\}$	Depending on month	85.2	90.9
N	Cloud amount (oktas)	As observed at 0300 GMT	-1.2	92.2
Y_2	Wind direction between 010° and 100°	*	-6.0	92.9
Y_3	Wind direction between 200° and 290°	*	8.4	93.0
B_6	Wind speed 22–26 kn	*	-3.9	93.1
B_2	Wind speed $\leqslant$ 6 kn	*	11.7	93.1
B_5	Wind speed 17–21 kn	*	-8.0	93.1
B_3	Wind speed 7–10 kn	*	2.2	93.1
B_4	Wind speed 11–16 kn	*	-11.0	93.1

* As observed at 0300 GMT; if yes: 1, if no: 0.

6. Performance of the equation

Equation (2) was applied to independent material (July 1980 – June 1981), i.e. observed values were used. The results have been listed in Table II. In operational practice, however, cloud amount as well as the sector from which the wind will blow at 0300 GMT has to be estimated some 12 hours earlier. After inserting forecast values into the equation, this effect proved to be quite small (bracketed figures in Table II).

Table II. *Monthly averaged errors* (°C) using observed and forecast values (bracketed) for N and Y_2.*

	Mean error	Mean absolute error	Root-mean-square error
January	0.05 (0.00)	1.18 (1.15)	1.38 (1.46)
February	-0.01 (-0.23)	1.19 (1.21)	1.40 (1.47)
March	-0.74 (-0.85)	1.41 (1.46)	1.77 (1.77)
April	0.35 (0.24)	0.98 (1.14)	1.23 (1.46)
May	0.25 (0.17)	1.28 (1.14)	1.56 (1.41)
June	0.32 (0.34)	0.97 (1.02)	1.29 (1.24)
July	0.00 (-0.05)	0.79 (0.82)	1.09 (1.14)
August	-0.67 (-0.57)	1.23 (1.27)	1.51 (1.55)
September	-0.14 (-0.23)	0.83 (0.84)	1.01 (1.06)
October	0.11 (0.21)	0.92 (1.05)	1.12 (1.23)
November	0.15 (0.02)	1.27 (1.18)	1.54 (1.46)
December	0.17 (0.00)	0.94 (0.95)	1.09 (1.12)

* Calculated from $\hat{T} - T$.

Verification of the forecasts issued during the months January–March in the years 1971–73, when no objective method was available, gave the average monthly errors displayed in Table III. Comparison of Tables II and III makes it clear that equation (2) performs satisfactorily.

Table III. *Monthly averaged forecast errors** (°C).

	Mean error	Mean absolute error	Root-mean-square error
January	−1.67	2.08	2.74
February	−1.72	2.05	2.55
March	−2.13	2.43	2.77

Calculated from T fcst - T.

7. Sources of errors

There are several sources of errors that affect the performance of equation (2). The rather heavy reliance on the representativeness of the afternoon temperature and dew-point certainly is a weak spot. A frontal passage after 1500 GMT may bring in an airmass with entirely different properties. Fortunately, this does not happen often in the temperate climate of western Europe. When it does, the forecaster may be able to adjust the equation by using, as input, temperature and dew-point of the new airmass.

Showers occurring shortly before 1500 GMT may temporarily alter temperature and dew-point considerably, thus leading to an erroneous outcome.

The data set that was available for the present investigation contained only three winters, all of which were rather mild. Therefore, the number of days with snow cover were far too few to warrant any conclusions. It is highly likely, however, that on such days the urban minimum temperature will be underestimated (Chandler, 1965).

Finally the water surface temperatures of the river Maas and the North Sea (at 25 km to the west) may exert an influence. It is believed, however, that only in cases of extremely high or low water surface temperatures will this influence be noticeable in too cold and too warm forecasts respectively.

8. Conclusions

It has been shown that there exists a strong relationship between the sum of temperature and dew-point, observed at Zestienhoven Airport at 1500 GMT, and the subsequent minimum temperature at a site in the Rotterdam urban area. Further refinement could be achieved by taking into account some simple meteorological variables, as well as a correction factor to eliminate seasonal influence. The resulting regression equation performs satisfactorily and may be regarded as a useful forecasting tool.

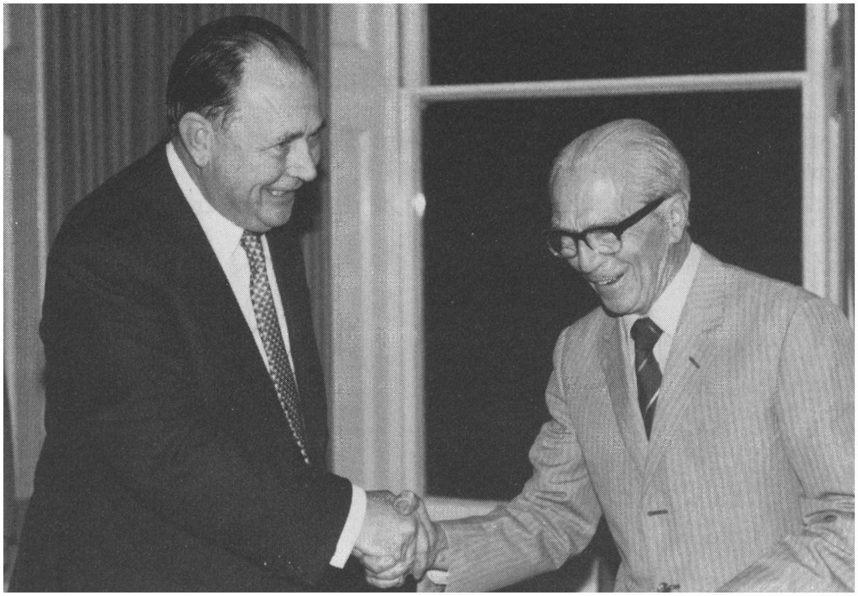
9. Acknowledgements

Thanks are due to Dr A.P.M. Baede for critically reading the manuscript, Mr S. Kruizinga who put a versatile computer program at the author's disposal, for which he is very grateful, and the forecasting staff of Zestienhoven Airport for helpful discussions and enthusiastic co-operation.

References

Böhm, R. Gabl, K.	1978	Die Wärmeinsel einer Grossstadt in Abhängigkeit von verschiedenen meteorologischen Parametern. *Arch Meteorol Geophys Bioklimatol,* Series B, **26**, 219–237.
Chandler, T.J.	1965	The climate of London, Hutchinson & Co.
Gordon, J., Perry, J.D. and Virgo, S.E.	1969	Forecasting night minimum air temperature by a regression equation. *Meteorol Mag,* **98**, 290–292.
Oke, T.R. and Maxwell, G.B.	1975	Urban heat island dynamics in Montreal and Vancouver. *Atmos Environ,* **9**, 191–200.

Notes and news



Mr T. Nagle receives his farewell presentation from the Director of Services, Mr F.H. Bushby.

The Principal, Mr S.G. Cornford, makes a presentation to Mr T. Nagle on his retirement.

Retirement of Mr T. Nagle

On 30 September the Office bade farewell to Tom Nagle at a gathering in the College at Shinfield Park. In his ten years as bar steward at the College Tom probably became better known personally to more meteorologists throughout the world than any meteorologist. Indefatigable, cheerful and with knowledge, skill and style born of his West End training and experience, he was absolutely the right man for the job.

The evening was a happy beginning to retirement after 60 years at work. On behalf of the whole Office a presentation was made by the Director of Services, Mr F. H. Bushby, and, on behalf of all past and present members of staff and courses, by the Principal, Mr S.G. Cornford.

THE METEOROLOGICAL MAGAZINE

No. 1329 April 1983 Vol. 112

CONTENTS

NOTICES

It is requested that all books for review and communications for the Editor be addressed to the Director-General, Meteorological Office, London Road, Bracknell, Berkshire RG12 2SZ and marked 'For Meteorological Magazine'.

The responsibility for facts and opinions expressed in the signed articles and letters published in this magazine rests with their respective authors.

Applications for postal subscriptions should be made to HMSO, PO Box 569, London SE1 9NH.

Complete volumes of 'Meteorological Magazine' beginning with Volume 54 are now available in microfilm form from University Microfilms International, 18 Bedford Row, London WC1R 4EJ, England.

Full-size reprints of out-of-print issues are obtainable from Johnson Reprint Co. Ltd, 24–28 Oval Road, London NW1 7DX, England.

Please write to Kraus Microfiche, Rte 100, Millwood, NY 10546, USA, for information concerning microfiche issues.

Printed in England by Robendene Ltd, Amersham, Bucks.
and published by
HER MAJESTY'S STATIONERY OFFICE

£2 monthly
Dd. 717701 K15 4/83

Annual subscription £26.50 including postage
ISBN 0 11 726933 6
ISSN 0026–1149