J. Wind Eng. Ind. Aerodyn. 97 (2009) 271–286

Contents lists available at ScienceDirect

Journal of Wind Engineering and Industrial Aerodynamics

journal homepage: www.elsevier.com/locate/jweia

XIMIS, a penultimate extreme value method suitable for all types of wind climate

R. Ian Harris *

RWDI Anemos Ltd, Lawrence Way, Dunstable LU6 1BD, UK

ARTICLE INFO

Article history:
Received 28 November 2007
Received in revised form
28 January 2009
Accepted 16 June 2009
Available online 17 July 2009

Keywords:
Extreme values
IMIS
Penultimate
Temperate storms
Thunderstorms
Cyclones

ABSTRACT

This paper introduces XIMIS, an extended version of the existing IMIS method. The new method is penultimate in that it does not rely on asymptotic results, which in turn depend on the rate parameter $rT \to \infty$. For input it requires a sample of a mutually independent data set drawn from the original parent data, but having the same annual maxima as that parent. Thus it can use independent storm data or m-day maxima from temperate storms, or thunderstorm or cyclone maxima.

In the paper, temperate storm data from Boscombe Down, UK, and cyclone and thunderstorm maxima, respectively, from Onslow, WA, and Brisbane, QD, in Australia are analysed. It is shown that derivation of standard 1:50 yr design values needs a mild extrapolation, which does not require any sort of probability model. A simple power law transformation is used to assist the extrapolation by linearising the plot. Derivation of 1:10,000 yr values does require a model and it is shown that if the relevant working variable is used, then there is no case for using any model except Type I. It is then argued that the transformation used for linearisation has good claims for validity for gross extrapolation, and the linearised plots are used to estimate 1:10,000 yr values.

It is concluded that XIMIS is a useful design tool.

1. Introduction

All countries that are members of the W.M.O. produce wind statistics in the form of mean wind speeds taken over an averaging time of between 10 min and 1 h. In the UK, an averaging period of an hour is used; so for convenience, in all that follows means will be described as hourly means, with the understanding that in other countries where a different averaging period is used, that average is implied.

Modern wind engineering design depends on knowledge of the probability distribution of the annual largest value of the hourly mean wind speed. This has to be deduced from available meteorological records.

If $P(V)$ is the parent distribution, i.e. the distribution of all the values in a year from which the annual maximum is drawn, then for data, like wind speed, which are correlated in time, the probability distribution of the annual maximum is

$$\Phi(V) = [P(V)]^{rT} \quad (1.1)$$

Here, r is the rate parameter. It is the number of independent values per annum drawn from the same parent as the (possibly correlated) data being analysed, which results in the same exact distribution of the annual maximum as the original data set. Note that in this context r is not, in general, an integer. T is the epoch for which the extremes are considered. (Since the concern here is with annual extremes always $T = 1$ yr, but will be left as a parameter in the development that follows.) Given a sample of annual maxima, estimation of $\Phi(V)$ involves fitting the data to $[P(V)]^{rT}$. Usually neither $P(V)$ nor the value of r is known with sufficient precision to make such a direct attack possible.

The alternative, known generically as extreme value methods, involves fitting the annual maximum data to an asymptotic form of $\Phi(V)$ to which it tends when $rT \to \infty$.

2. Review of extreme value methods applied to wind data

The pioneering paper by Fisher and Tippett (1928), which was considerably developed in the text by Gumbel (1958), showed that as $rT \to \infty$, $\Phi(V)$ tends to an asymptotic form $\Psi(V)$ belonging to one of three types called Types I, II and III by Gumbel (1958). Von Mises (1936) showed that all three asymptotes could be reduced to one common form:

$$\Psi(V) = \exp\{-[1 - k(V - \check{V})/a]^{1/k}\} \quad (2.2)$$

which has become known as the generalised extreme value distribution (GEV). In (2.2) $\check{V}$ and a are, respectively, location and dispersion parameters, and k is the shape factor that determines which of the three classical asymptotes the GEV represents. $k<0$ corresponds to Type II, which is not considered further here, since no genuine examples of Type II behaviour of extreme wind speeds

* Tel./fax: +44 1582661658.
E-mail address: rih.wind@ntlworld.com

0167-6105/$ - see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jweia.2009.06.011

appear to exist. The case $k>0$ corresponds to the Type III asymptote, known also as the Reverse Weibull distribution, and the limiting case $k\to 0$ gives

$$\Psi(V) = \exp\{-\exp[-(V - \check{V})/a]\} \tag{2.3}$$

which is Type I asymptote or Gumbel distribution.

Following the pioneering work by Shellard (1962), it has become customary to fit annual maximum data to the Gumbel distribution. With the limited data set, i.e. only one data value per year of record, the scatter due to sampling error is large, which limits the accuracy with which design values can be predicted. There is also a systematic convergence error, caused by the fact that the Type I asymptote is derived by assuming $rT\to\infty$, whereas accepted practical values of $r\sim 150$ or less, which leads to systematic error. The data usually forms a concave-down curve on the traditional Gumbel plot, rather than a straight line, which a perfect fit to the Type I asymptote would require, i.e. the Type I fit approximates the curve by the tangent at $V=\check{V}$. The redeeming feature of this is that the convergence error is always conservative, producing a degree of over-design, but always safe design values. Cook (1982), Harris (2004) and Cook and Harris (2004) have shown that this systematic error can be reduced or eliminated by pre-conditioning the data. In practice, this means that instead of analysing annual maxima of wind speed, the analysis is applied to values of pressure, i.e. (wind speed)^2. If, as in many cases, the relevant parent can be shown to have a Weibull distribution with index w, then plotting (wind speed)^w will eliminate the convergence error.

The curvature on the Gumbel plot has led some workers to try direct fitting of the GEV equation (2.2) to annual maximum data. Harris (2006) reviewed this procedure and found the following:

(i) There is a large convergence error, which in many cases is non-conservative, with the potential to produce unsafe design values.
(ii) Since $k=0$ is a singularity of (2.2), a Type I fit can never be obtained by fitting a real data set; thus to eliminate spurious non-zero values of k caused by sampling error, it is necessary to apply a significance test, with the null hypothesis $k=0$. With the limited sample size available, the null hypothesis will almost never be rejected, indicating that Type I re-analysis is required.
(iii) Very different results are obtained when the same procedure is applied to (a) wind speeds and (b) pressures, making the process indefensible as the basis of a design standard.

The reasons for objecting to this method as the basis of a standard are as follows. It is axiomatic that wind speed and pressure cannot have the same parent distribution. It is equally axiomatic that if $w>0$ the transformation $Z=V^w$ (of which pressure is the case $w=2$) cannot alter to which of the classical asymptote types, annual maxima drawn from these variables will eventually conform as $rT\to\infty$. Thus if V is Type III, so must be Z, and similarly for Type I. Yet analysis of wind speed data by this method usually indicates Type III, whilst analysis of the corresponding pressures indicates Type I, thus violating this axiom of extreme value theory.

Design standards have legal implications and therefore must be unambiguous. With this method, different design values (including the existence or not of an upper bound) may be obtained, depending on which order the operations of statistical analysis and squaring are carried out. This makes the process indefensible as the basis of a design standard.

On this basis, Harris (2006) concluded that this procedure should not be used.

Since sampling errors can be reduced only by increasing the size of the sample, this has led to an interest in methods that use more data from each year than just the annual maximum. There are two established methods and a third one has recently been suggested.

The first established method uses a result by Pickands (1975) that if the distribution of annual maxima tends to the GEV, then excesses of the data over a threshold u should tend to conform to the Generalised Pareto Distribution (GPD) as the choice of threshold $u\to\infty$. This method provides an alternative route to estimate the GEV parameter k in (2.2). Harris (2005) recently reviewed the application of this method to wind data, and showed that because it depends on conformation to the GEV, it possesses exactly the same convergence problems arising from too low values of rT. Additionally, there are extra convergence problems because the value of the threshold u has to be set too low in order to provide enough data to analyse, so that, in practice, the asymptotic condition $u\to\infty$ is not met either. It replicates the direct GEV problem by producing grossly different results for k depending on whether wind speed or pressures are analysed. Finally, because the increase in sample size is not enough to compensate for the other difficulties, the sampling errors are very large, so that again, when a significance test is applied, Type I re-analysis is indicated. Harris (2005) concluded that the method is sufficiently flawed that it should not be used as the basis of a design standard.

The r-LOS method is a new procedure devised by Coles (2001) which uses the r largest order statistics from each year of record. This has recently been tested on data from Canadian sites by An and Pandey (2007). The method again assumes convergence to the GEV, i.e. $rT\to\infty$, and hence is liable to the same systematic bias errors as other GEV methods, because practical values of rT are too low. This is clearly revealed by An and Pandey (2007) who processed both wind speed data and pressure data, (wind speed)^2, from each station. The results showed a systematic shift $\sim$0.12 in the GEV shape parameter towards a Type III fit, when wind speed was processed, as compared to an indication of a Type I fit when pressures were processed. This shift is roughly the same as that obtained with other GEV methods. Again, this discrepancy renders the method unsuitable in its present form as the basis of any design standard for wind actions.

The second established method is the "Method of Independent Storms" devised by Cook 1982, improved by Harris (1999) and known as IMIS. The sample consists of all wind storm maxima in each year ($\sim$100). These are detected by examining continuous records and identifying lulls, i.e. negative going crossings of a threshold. Between each pair of lulls is an independent storm from which the maximum is extracted. The data are pre-conditioned by squaring to obtain pressures and the top order-statistics from the complete set of data are fitted to the Type I asymptote. The method thus assumes a Type I fit and requires records of sufficient continuity to enable the independent storm maxima to be identified.

Cook and Harris (2004) showed that the IMIS method can also be used with a penultimate version of the Gumbel distribution, valid for finite rT, provided that the parent can be assumed to be of the Weibull form. The major advantage of this method is that it produces identical results, regardless of whether wind speeds or pressures are analysed.

To summarise this section, all extreme value methods that rely on convergence to an asymptotic form as $rT\to\infty$ will contain systematic errors because practical values of rT are not large enough. For a Type I fit, these errors can be reduced by analysing wind pressure instead of wind speed. Sampling error limits the accuracy of methods that use only annual maxima. Thus a method is required which uses more data from each year than the annual maximum and is penultimate, i.e. does not rely on $rT\to\infty$ for its