J. Wind Eng. Ind. Aerodyn. 125 (2014) 146–155

Contents lists available at ScienceDirect

Journal of Wind Engineering and Industrial Aerodynamics

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

A simulation method for the macro-meteorological wind speed and the implications for extreme value analysis

R. Ian Harris *

RWDI, Lawrence Way, Dunstable LU6 1BD, UK

ARTICLE INFO

Article history:
Received 6 August 2013
Received in revised form
4 December 2013
Accepted 5 December 2013
Available online 16 January 2014

Keywords:
Simulated time series
Correlation
Extreme value analysis
Penultimate methods
Annual maxima

ABSTRACT

This paper provides a contribution to the testing of existing methods of analysis of extreme wind speeds and to the development of better alternatives. A method is developed for synthesising a correlated random time series with a Rayleigh amplitude distribution and an arbitrary auto-correlation. The auto-correlation is selected to be the Von Karman model because the method is then used to generate 20,000 years of simulated hourly mean wind speeds. Annual maxima are extracted and exhibited on Gumbel plots. Familiar problems with convergence to asymptotic forms are confirmed and a new problem is revealed in that the annual rate parameter, previously believed to be constant, is found to vary significantly in the range of the measured data encountered in practical extreme value analyses. With the exception of newly developed penultimate methods, all the existing methods of analysis depend implicitly on either convergence to an asymptotic form, or invariance of the annual rate parameter, or both. This has serious implications for the accuracy of these methods, not only for the analysis of annual maxima, but also for extensions of these methods developed to use more data from each year.

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 values of the wind speeds that produce significant forces or stresses in the building or structure being designed. These have to be deduced from available meteorological records. The situation is complicated by the fact that at a given site several different physical mechanisms may each produce wind speeds that have to be considered. For instance, a site on the eastern seaboard of the United States is likely to be affected by winds derived from large scale depressions, tropical storms (hurricanes, typhoons) and thunderstorms. More rarely, the design wind speeds arise from only one such mechanism. Such a case is termed a “simple climate”. A well-known example is the UK, where strong winds relevant to design arise exclusively from Atlantic depressions.

All the available methods for extracting probabilities of annual maxima require that the data samples should be independent of one another. For some mechanisms, such as hurricanes, this is not a problem since these are easily identified well-separated events, and thus independent. The analysis is usually conducted by considering the maximum wind speed recorded in each hurricane. The same applies to thunderstorms, with the proviso that care is taken to eliminate the possibility of two storms in close succession occurring within one synoptic episode. [Easily done by imposing a minimum separation on the thunderstorms considered, as in Lombardo et al. (2009).] Conversely, for winds derived from extra-tropical depressions, the hourly mean data, being samples from a continuous physical variable, inevitably are mutually correlated, and thus successive samples are not statistically independent. Consequently, all present methods rely on extracting from the original sample an uncorrelated subset that has the same annual maxima as the original data. Inevitably, this raises a number of issues concerning the relationship between results based on these subsets, and the statistics of the original correlated data, which cannot be resolved since the correct answers for the original data set are unknown. Thus there appears to be considerable merit in devising a method for the computer simulation of a correlated data set of this type with a prescribed probability distribution and correlation. This paper describes such a method, and presents some results from applying existing methods to the resulting data.

There is a considerable body of evidence which suggests that a good model for the probability distribution of hourly mean wind speeds from depressions is the Forward Weibull form with an index lying between 1.8 and 2.2. [Even by 2004 there were well over 400 references to the use of the Weibull distribution for wind

* Tel./fax: +44 1582 661658.
E-mail address: harris-r1@virginmedia.com

0167-6105/$ - see front matter
http://dx.doi.org/10.1016/j.jweia.2013.12.003

speeds in the Scirus database of peer-reviewed journals.] The middle of this range is a Forward Weibull with index 2; otherwise known as a Rayleigh distribution. Recently Harris (2008) showed that the spectrum of the hourly mean wind speed, known for historical reasons as the macro-meteorological spectrum (Van der Hoven, 1957) consists of a broadband random component together with a number of discrete deterministic “lines” associated with the annual and daily cycles and their harmonics. Supplementary material supplied in discussion by Baker (2010) confirmed that the spectrum of the broadband component appears to conform to the Von Karman model (Von Karman, 1948). Hence this paper will describe a simulation of a correlated random variable with a Rayleigh first-order probability distribution, and a Von Karman spectrum. Harris (2008) also estimated the timescale of the broadband random component by integration of its auto-correlation and found a value T=22.15 h. Accordingly, in what follows, this value will be used for T when a numerical value is required.

2. Theoretical development

The starting point is a well-known result from probability theory. If $x(t)$ and $y(t)$ are identically distributed independent random variables, each having a Normal probability distribution with zero mean and unit standard deviation, then the joint probability density of x and y is given by:

$$p(x,\ y)dxdy=\frac{1}{2\pi}\exp\left(-\frac{x^2+y^2}{2}\right)dxdy \tag{2.1}$$

If x and y are regarded as Cartesian co-ordinates in a plane relative to some origin, it follows that the resultant $r(t)$ has a probability density:

$$p(r)dr=r\exp\left(\frac{-r^2}{2}\right)dr \tag{2.2}$$

That is $p(r)$ has a Rayleigh distribution. It then follows that the mean of r, $\overline{r}$, and the standard deviation of r, σ_r, are given by:

$$\overline{r}=\sqrt{\pi/2} \tag{2.3a}$$

$$\sigma_r=\sqrt{2-\pi/2} \tag{2.3b}$$

Suppose the variables $x(t)$ and $y(t)$ both have an auto-correlation function $\rho(\tau)$ and consider pairs of values $x_1(t)$, $x_2(t)$ and $y_1(t)$, $y_2(t)$ both separated by a time lag τ. Since $x(t)$ and $y(t)$ are independent, it follows that the joint fourfold probability density is given by:

$$p(x_1,\ x_2,\ y_1,\ y_2)=\frac{1}{4\pi^2(1-\rho^2)}\exp\left(\frac{-\{x_1^2+x_2^2+y_1^2+y_2^2-2\rho(x_1x_2+y_1y_2)\}}{2(1-\rho^2)}\right) \tag{2.4}$$

Now introduce the substitutions:

$$x_1=r_1\ \cos\ \theta_1;\ x_2=r_2\ \cos\ \theta_2;\ y_1=r_1\ \sin\ \theta_1;\ y_2=r_2\ \sin\ \theta_2 \tag{2.5}$$

Then the joint fourfold distribution becomes:

$$p(r_1,\ r_2,\ \theta_1,\ \theta_2)=\frac{r_1r_2}{4\pi^2(1-\rho^2)}\exp\left(\frac{-\{r_1^2+r_2^2-2r_1r_2\rho\ \cos(\theta_1-\theta_2)\}}{2(1-\rho^2)}\right) \tag{2.6}$$

To obtain $p(r_1, r_2)$, the joint marginal distribution of r_1 and r_2, it is necessary to integrate out the dependence on the angle variables θ_1 and θ_2, each over the range $-\pi$ to $+\pi$. The symmetry properties of the ranges of integration and of the trigonometric functions allow this double integration to be reduced to:

$$p(r_1,\ r_2)=\frac{r_1r_2}{2\pi^2(1-\rho^2)}\exp\left[\frac{-(r_1^2+r_2^2)}{2(1-\rho^2)}\right]\pi\int_{-\pi}^{+\pi}\exp\left[\frac{r_1r_2\rho\ \cos\ \varphi}{(1-\rho^2)}\right]d\phi \tag{2.7}$$

From Gradshteyn and Ryzhik (1994):

$$\int_{-\pi}^{+\pi}e^{z\ \cos\ x}dx=2\pi I_0(x) \tag{2.8}$$

where $I_0(x)$ is a modified Bessel Function of the first kind. Hence:

$$p(r_1,\ r_2)=\frac{r_1r_2}{(1-\rho^2)}\exp\left[\frac{-(r_1^2+r_2^2)}{2(1-\rho^2)}\right]I_0\left[\frac{r_1r_2\rho}{(1-\rho^2)}\right] \tag{2.9}$$

The auto-correlation of $r(t)$, denoted by $\zeta(\tau)$ is given by:

$$\zeta(\tau)=\frac{\overline{r_1r_2}-(\overline{r})^2}{\sigma_r^2}=\frac{(2\overline{r_1r_2}-\pi)}{(4-\pi)} \tag{2.10}$$

so that the evaluation of $\zeta(\tau)$ requires the evaluation of:

$$\overline{r_1r_2}=\int_0^{\infty}\int_0^{\infty}\frac{r_1^2r_2^2}{(1-\rho^2)}\exp\left[\frac{-(r_1^2+r_2^2)}{2(1-\rho^2)}\right]I_0\left[\frac{r_1r_2\rho}{(1-\rho^2)}\right]dr_1dr_2 \tag{2.11}$$

The calculation proceeds by the expansion of the Bessel Function into its power series (Gradshteyn and Ryzhik, 1994). Each term of the expansion then factorises into the product of an integral in r_1 and an identical integral in r_2. These integrals can be evaluated and the result is a power series in ρ^2. This series can be summed and further simplified with the aid of some identities given by Gradshteyn and Ryzhik (1994). The final result is:

$$\overline{r_1r_2}=\frac{\pi}{2}\,{}_2F_1\left(-\frac{1}{2},\ -\frac{1}{2};\ 1;\ \rho^2\right) \tag{2.12}$$

where the ${}_2F_1$ is the Gauss Hypergeometric Function. For arguments $|z|<1$ this is defined by the power series:

$${}_2F_1(a,\ b;\ c;\ z)=1+\frac{ab}{c1!}z+\frac{a(a+1)b(b+1)}{c(c+1)2!}z^2+\ldots \tag{2.13}$$

This series expansion is a reasonably efficient means of evaluating the ${}_2F_1$ function in (2.12).

From (2.10) it follows that:

$$\zeta(\tau)=\frac{{}_2F_1(-\frac{1}{2},\ -\frac{1}{2};\ 1;\ \rho^2)-1}{4/\pi-1} \tag{2.14}$$

As a check, putting τ=0 in this expression and using some further identities for the ${}_2F_1$ function (Gradshteyn and Ryzhik, 1994) gives $\zeta(0)=1$, as it must.

3. Computing the target input auto-correlation, $\rho(\tau)$

As already noted, the correct form for the auto-correlation of $r(t)$ is assumed to be the Von Karman formula, so that:

$$\zeta(\tau)=\frac{2}{\Gamma(1/3)}\left(\frac{\alpha\tau}{2T}\right)^{1/3}K_{1/3}\left(\frac{\alpha\tau}{T}\right) \tag{3.1}$$

where the constant $\alpha=\sqrt{\pi}\Gamma(5/6)/\Gamma(1/3)=0.746834$ (to 6DP).

Given that the data being simulated are hourly means and that the time-scale, T, is given in hours, it follows that the values of ζ