Journal of Wind Engineering & Industrial Aerodynamics 170 (2017) 274–282

Contents lists available at ScienceDirect

Journal of Wind Engineering & Industrial Aerodynamics

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

Short communication

Technical note: On extracting independent peak values from correlated time series for assessing extremes

Nicholas J. Cook

Highcliffe on Sea, Dorset, UK

ARTICLE INFO

Keywords:
Correlated time series
Independent peak values
Declustering
Poisson point process
Mean-crossing
XIMIS

ABSTRACT

The methods of extracting independent peaks from ergodic, but correlated, time series are usually called "declustering", and the principal method is to select the largest peak value between successive up- and down-crossings of the mean. It is shown that the independent peak values will follow the Poisson point process model if the declustering process uses an optimal low-pass filter to detect the crossings of the mean. When the correlated time series is a mixture of two, or more, processes a second, higher threshold may be applied that passes only those peaks from the dominant process, otherwise the resulting process will remain "mixed" and must be analysed accordingly.

1. Introduction

With the notable exception of the ACER method of Karpa and Naess (2013) which works directly from correlated and non-stationary time series, nearly all extreme-value theory and methods for estimating extremes of a given probability or return period require statistically independent peak values, with distribution P, that follow a homogeneous Poisson point process (PPP). The corresponding distribution of the extreme in a datum epoch, Φ, is then given by:

$$\Phi = P^r \tag{1}$$

where r is a rate parameter which may be estimated as the average number of independent peaks in the datum epoch. In practice, wind engineering time series, including wind speeds and pressure coefficients, are serially correlated. The methods of extracting independent peaks from ergodic, but correlated, time series are usually called "declustering", and the principal method is to select the largest peak value between successive up- and down-crossings of the mean. This approach was explicitly introduced for the assessment of extreme winds by Cook (1982) in the "method of independent storms" (MIS) but, long before this, was implicitly introduced by Davenport (1964) in his adaption of communication theory (Rice, 1945) which predicts that there is one, and only one, independent peak value between each successive up-crossing of the mean. Simiu and Heckert (1996) introduced the concept of m-day extremes, using $m = 4$, 8 and 16 days for assessing extreme winds in the USA but, although these values are statistically independent, they do not follow a PPP because the inter-arrival time is constrained by the value of m. However, if these m-day extremes are left-censored at a suitably high threshold value, the surviving peaks gradually converge towards a PPP.

The key indicative characteristic of a homogeneous PPP is that the inter-arrival time, t, the time between successive peaks, is random and is exponentially distributed:

$$P(t) = 1 - e^{-\lambda t} \tag{2}$$

where λ is the intensity function. This simple test will be used later to show that the method used to detect the up- and down-crossings is critical to whether the extracted peaks follow a PPP. In a homogeneous PPP $r = \lambda T$, where T is the length of the datum epoch and this gives a second estimate for r. The situation is more complex when the time series is a mixture of two or more processes with different λ values, or Poisson timescale $\mathbf{T} = 1/\lambda$. For example, Brabson and Palutikof (2000) reported three distinct timescales: $\mathbf{T} = 4$ h, 27 h and 160 h in wind gust data by fitting linear slopes to three different parts of the distribution.

2. Supporting data

The two sets of pressure coefficient (C_p) data used to illustrate this Note were used previously to support the discussion in Cook (2016b). They comprise very long time-series of C_p sampled at 40 Hz for 30 h full-scale (FS) at two locations, shown as taps A and B in Fig. 1, on the low-pitched multi-faceted roof of a 1:100 scale model in the University of Western Ontario (UWO) boundary layer wind tunnel. As these series are some 30 times longer than typical, they are useful for exploring extremes further into the upper tail that would be possible with the more typical

E-mail address: wind@njcook.uk.

http://dx.doi.org/10.1016/j.jweia.2017.09.003
Received 24 April 2017; Received in revised form 3 September 2017; Accepted 3 September 2017

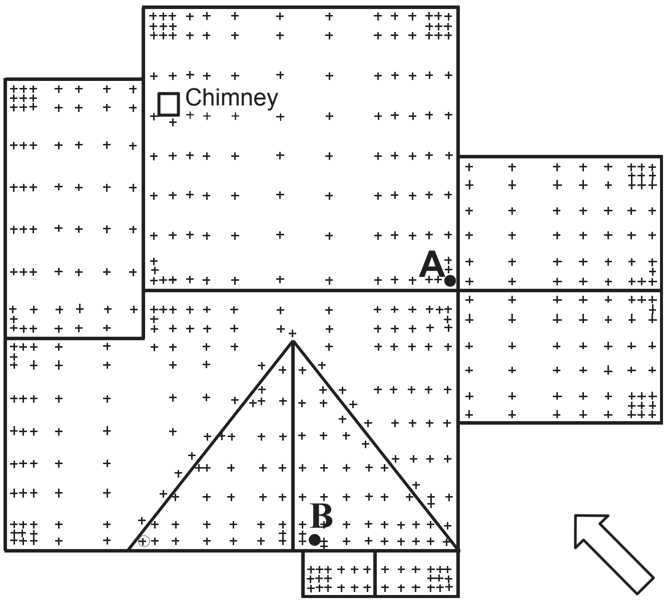


Fig. 1. Location of taps and wind direction.

data lengths of 1–3 h FS. The 40 Hz acquisition rate is also quite high in comparison with the gust load duration specified in codes of practice for design: e.g. 1 Hz is the datum duration for cladding loads in the UK wind code, BS6399-2. Hence the opportunity is taken to examine both shortened and time-averaged versions of these two sets of data. Applying a 1 Hz digital low-pass filter to these data produces time-averaged versions of the series that remain sampled at 40 Hz. The data series are designated in the later figures as: A40 & B40 for the original tap A & tap B series, and A1 & B1 for the time averaged versions.

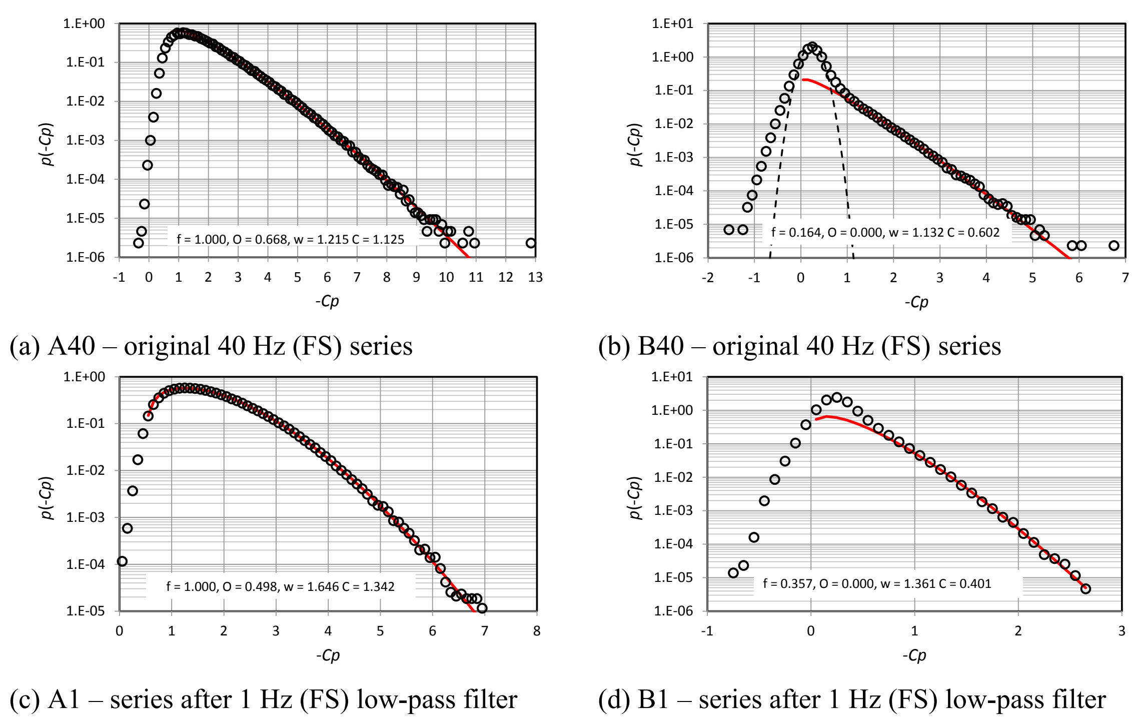
Both taps are in regions of separated flow, so that they experience large peak suctions. Here, each series has been negated to–C_p to permit the analysis of suction peaks (minima) using standard procedures for assessing maxima. Fig. 2 shows the probability density functions (pdf) which correspond to Fig. 4 of Cook (2016b), except they are mirrored due to the reversal of sign. The upper tails have been fitted by a Weibull distribution to demonstrate that the extremes all lie in the domain of attraction for the FT1 distribution. The Weibull parameters on the figure are: f the relative frequency of the component; O the origin; w the shape parameter; and C the scale parameter. As averaging is a summation process, the distributions of the time-averaged series diverge away from exponential towards Gaussian through the action of the Central Limit Theorem, and the Weibull shape parameter increases accordingly. Weibull-equivalent tails indicate that XIMIS, the penultimate FT1 methodology of Harris (2009), is appropriate for fitting the distribution of annual maxima of the extracted independent peaks.

The SGEMM model applied to B40 in Cook (2016a) revealed an additional Gaussian distributed component indicated in (b) by the dashed curve. This is the most frequent component ($f = 0.84$) forming a narrow peak near the origin in (b) which is considerably reduced for B1 in (d) by the 1 Hz filter. Whether this high-frequency "noise" component is due to shear-layer turbulence from intermittent flow reattachment or is merely intermittent noise from a faulty transducer is irrelevant here: the tap B series are useful in demonstrating the effectiveness of declustering in mixtures where the least-frequent component dominates in the extreme tail.

3. Method and results

The approach in this Note differs from Brabson and Palutikof (2000), who used a combination of peaks over threshold and "deadtime" between peaks, in that it takes a two-step approach. In the first step, a zero-shift low-pass filter is applied to data series to identify mean-crossings, as introduced in MIS (Cook, 1982), but here the filter time-constant is optimised so that the arrivals of the peak value between crossings follows Eqn (2) with the minimum residual variance. This first step is sufficient when it reveals a single PPP. If the first step reveals a mixture of processes, the second step rejects peaks below a second threshold, the value of which is optimised in the same manner, so that the surviving tail-dominant process is a PPP.

Wind engineering time-series data, whether wind speeds or pressure



(a) A40 – original 40 Hz (FS) series

(b) B40 – original 40 Hz (FS) series

(c) A1 – series after 1 Hz (FS) low-pass filter

(d) B1 – series after 1 Hz (FS) low-pass filter

Fig. 2. PDFs of example data series showing tail-equivalent Weibull parameters.