In this paper, expressions are derived for the expected number of spurious peaks in a spectrum estimate, that is, crossings above a given significance level per frequency unit, as well as the expected width of these peaks. In numerous scientific applications, spectrum estimates are used for the purpose of identifying sinusoidal or modal components, often thinning large sets of candidate frequencies with coincidence detection. Because one always expects numerous false peaks in a spectrum estimate, knowing the expected rate of false peaks helps to decide whether the number observed is abnormal and hence determine the true nature of the process. An example using solar wind data from the Advanced Composition Explorer is given where spectra display pathological numbers of significant peaks, while temporally permuted versions of the data possess spectra with the number expected for a white, Gaussian process. The permutation test is a valuable diagnostic for processes suspected to contain many line components.

1. Introduction

Time series are often analysed using non-parametric spectrum estimates for the purpose of determining where harmonic components contribute significantly to signal power. In particular, a solution to the problem of the ‘Separation of Accidental from Real Periodicities’ [1, p. 25] has been sought since the dawn of spectrum estimation. Many important problems in physics and engineering (ranging from climate analysis to communications systems to surgical scheduling [2]) involve data which contain many harmonic components, from several hundreds to hundreds of thousands of lines in relatively narrow bands. For these applications, knowledge of the rate of false peaks is crucial. For example, the discovery of solar modal structure...
in charged particle flux in the interplanetary magnetic field [3] and subsequent discoveries [4]
implies the contribution of $10^7$ line components of varying amplitudes [5] on the 250–5100 µHz
band alone. Coherent modal structures also couple to atmospheric pressure on Earth, induced
currents on transatlantic ocean cables, communications failures and seismic oscillations [6]. Even
more pressing is the quest for experimental evidence of low-frequency gravity, or g-mode,
detections [7], whose frequencies are even more closely spaced.

When analysing processes which are suspected to contain many lines, the null hypothesis $H_0$ is
that the data do not contain periodic or other deterministic components. In this case, the spectrum
estimate follows a central $\chi^2$ distribution. The alternative, $H_1$, is that the process contains many,
perhaps several thousand, weak, approximately periodic components and the spectrum estimate
is non-central $\chi^2$ distributed at the modal frequencies. Known periodicities such as tides, annual
cycles, etc. are excluded from both hypotheses. When spectrum estimates are computed with
large datasets, the central $\chi^2$ assumption can be used to test for significant lines in a standardized
spectrum, e.g. at the 99% level. The hypothesis test thus becomes a question of how the significant
excursions of the standardized spectrum are bunched: either as (i) a relatively small number of
wide peaks above 99% or (ii) a relatively large number of very narrow ones. If the spectrum is
made up of thousands of genuine line components, as in case $H_1$, the spectrum estimate should
contain a large number of narrow peaks as in (ii). If $H_0$ is the case, all perceived lines are spurious,
so significant peaks in the estimate should be few—but how many peaks should be expected in
this case? In this paper, the rate of spurious peak detection is given. This expression also gives the
expected width of spurious peaks. Using a result of Blachman [8], the shape of spurious peaks
can be compared against the shape of genuine line components. These results depend on the
autocorrelation function of the spectrum estimate.

Certain objections to the initial findings in [3] have been raised. In particular: (1a) One-
dimensional simulations of the solar wind [9] implied that turbulence in the solar wind would
destroy such modes. (2a) When the logarithms of such spectra are plotted on a logarithmic
frequency scale, their overall shape is close to the $f^{-5/3}$ dependency characteristic of Kolmogorov
turbulence. (3a) Because the discovery of the solar modes was based on spectrum estimates
and the original time-series data were neither stationary nor Gaussian, some suggested that the
appearance of modes could be an anomaly due to a perverse distribution [10]. These objections
are gradually being overcome: (1b) Improved simulations of the solar wind using all three
spatial dimensions [11] have shown that discrete modes can coexist with turbulence. (2b) The
identification of the background shape with Kolmogorov turbulence is probably correct, but does
not imply that turbulence occurs in interplanetary space any more than the discrete modes occur
in interplanetary space; both occur in the Sun’s convection zone and imprint their signatures on
the solar wind [12]. (3b) In addition to the arguments given in [4], this paper uses a permutation
test along with the expected rate of spurious peaks to determine whether the perceived modal
structure is due to (i) the distribution of the data, (ii) an artefact of the spectrum estimation
procedure, or (iii) the correlation structure of the time series. When temporally permuted data
possess spectra with the number of peaks expected under $H_0$, one strongly suspects that the last
case is the real explanation.

(a) Outline

The paper begins with two background sections necessary to establish the new results in §4.
Section 2 is a brief introduction to the level crossing literature. Section 3a contains a note on
Slepian sequences followed by a concise summary in §3b of multitaper spectrum estimation,
with two examples in §3d showing why the multitaper method is highly preferred for scientific
purposes. The probability density functions of multitaper estimates are given in §3c.

The main results of this paper are in §4, with the basic formula for the upcrossing rate (the
upward crossing of a user-chosen significance level) and the ‘dwell bandwidth’ (the width of an
excursion above a given level). In §5, the crossing rates and the dwell bandwidth for a simple
class of ‘direct’ spectrum estimates are derived. However, smoothing is not the answer to the
problem of inconsistency and high crossing rates. The use of the multitaper estimate offers a
different approach to reduce the number of significant crossings. Next, §5b is devoted to ‘the
autocorrelation function of the multitaper spectrum estimate’, or the multitaper antecorrelation.
Section 6 invokes Blachman’s [8] analysis of stationary processes to show that spurious large
peaks in a multitaper spectrum estimate have a triangular shape. Section 7 notes a recent result
on extreme order statistics of Poisson random variables that is useful to answer the question of
how many peaks one might expect in a given band.

Section 8 is devoted to the spectrum of solar wind density. These data are strongly non-
Gaussian, and, as mentioned in the Introduction, results obtained from it have been controversial.
Finally, §9 gives a summary and conclusions.

2. Crossing problems in the time domain

Level crossing problems for zero-mean, stationary, continuous-time, Gaussian processes were
introduced by Rice [13], who gave the expected number of zeros per second as
\[
\hat{N}_z = \frac{1}{\pi} \left( -\frac{\psi''(0)}{\psi(0)} \right)^{1/2} = 2 \left[ \frac{\int f^2 S(f) \, df}{\int S(f) \, df} \right]^{1/2},
\]
(2.1)
where \( \psi(\tau) \) is the autocovariance function of the process at lag \( \tau \), \( \psi'' \) its second time derivative and
\( S(f) \) the spectrum of the process at frequency \( f \). The history of this formula is given in [14], and
the theory has expanded into a large literature with the surveys [15,16] which include 133 and 82
references, respectively.

The problem is to adapt these works to the frequency domain to study level crossing
rates of spectrum estimates. In addition to probability densities, this adaptation requires the
‘autocorrelation’ or ‘spectrum’ of the spectrum estimates plus subsidiary quantities such as their
moments. Fortunately, one does not encounter the main difficulty that occurs in time-domain
applications, namely differentiability of the process (e.g. [17]), because tapered Fourier transforms
of the data are entire functions of frequency.

It is well known that Fourier transforms of large data samples tend to a complex Gaussian
distribution [18,19] and experiments where the data ordering is randomly permuted have shown,
as in §8, that this assumption is surprisingly robust. Accordingly, spectrum estimates, as squared
linear combinations of these, can be analysed as \( \chi^2 \) processes. Then, under \( H_0 \) significance can be
assigned to peaks in spectrum estimates according to the \( \chi^2 \) distribution.

Rice’s method was extended to \( \chi \) or \( \chi^2 \) processes first by [20,21] and later to gamma-
distributed processes by [22–25]. Detections of single lines in spectra derived from short,
replicated datasets were considered in [26]. It is usually assumed that the individual Gaussian
processes making up the \( \chi^2 \) process are independent with identical autocovariances, but here
neither of these assumptions is true. The formula giving the number of crossings, specific to
spectrum estimates, is given in §4. The next sections give the necessary properties of direct and
multitaper spectrum estimates and establish notation.

3. Spectrum estimation

Following its nineteenth-century origins with Stokes [27], Kelvin [28], Schuster [1] and Rayleigh
[29], the computational effort associated with statistical spectrum estimation caused it to fissure
into three major classes: parametric, direct and indirect. Parametric models began with Yule
and Walker in the 1920s but have been shown to be unacceptable for many applications (see
[30] and fig. 16 of [31]). The remaining two major branches were formalized by Blackman &
Tukey’s [32] division of non-parametric spectrum estimates into ‘direct’ and ‘indirect’ classes.
Indirect estimates begin with sample autocovariances, tapers them with a ‘lag window’ and then
takes their Fourier transform to estimate the spectrum, producing a smoothed periodogram. The
periodogram was proved to be inconsistent by Rayleigh in 1903 [29]. However, our first example
in §3d shows that inconsistency is a minor problem when compared with bias. Furthermore, their advantage of reduced computation disappeared with the 1965 rediscovery of the fast Fourier transform (FFT), making indirect estimates obsolete.

By 1967, Tukey [33] was recommending the direct estimator. Given \( N \) data samples of a stationary series \( \{x(t)\} \) for \( t = 0, 1, \ldots, N - 1 \), one computes a direct estimate by multiplying \( x(t) \) by a data taper, \( D(t) \), computing the FFT and squaring

\[
\hat{S}_D(f) = \left| \sum_{t=0}^{N-1} x(t)D(t)\exp(-i2\pi ft) \right|^2,
\]

where the tapers are standardized by \( \sum_{t=0}^{N-1} D^2(t) = 1 \). By convention, the sampling interval \( \delta t = 1 \) is used, so frequency, \( f \), is restricted to the Nyquist band \(-1/2 \leq f < 1/2\). Commonly used data tapers are described in [34]. The idea of a data taper possibly originated with Rayleigh and certainly goes back to 1943: see the letter from Norbert Wiener to Tukey reprinted in [35]. Applying the boxcar (constant) data taper, \( 1/\sqrt{N} \), to (3.1), one obtains the periodogram defined by Schuster. Because the data tapers used in the 1960s and 1970s were largely ad hoc and often gave significantly different spectrum estimates with the same data, the idea of tapering met strong opposition.

Direct estimators have \( \chi^2 \) (exponential) distributions independent of the sample size and as such are inconsistent (i.e. the variance of the estimator does not decrease as \( N \) increases). Anticipating the notation of §3c, write the degrees of freedom (DoF) as \( 2\alpha \) so that for direct estimates \( \alpha = 1 \). Good data tapers can control the bias at the cost of increasing the bandwidth (the width of the central lobe of the spectral window, which is the magnitude-squared FFT of the data taper, denoted \( |\hat{D}(f)|^2 \)), but the direct estimator is still inconsistent and requires smoothing to produce consistent results. The smoothing further decreases resolution, trading bias to reduce the variance. This effect has been known since [36] and has since been investigated in numerous other papers. Smoothing also has the unwanted effect of making the calculation of the equivalent DoF of the estimate difficult and is undesirable for identification of peaks in the spectrum because both genuine and spurious peaks are smoothed indiscriminately while peak morphology is distorted.

(a) Slepian sequences

In a famous set of papers, ‘Prolate spheroidal wave functions, Fourier analysis, and uncertainty’, begun in 1961 [37], Slepian, Landau and Pollak established bounds on simultaneous time–frequency concentration, the uncertainty principle when boundaries are imposed. In part five of this series [38], Slepian described the discrete case where one has \( N \) time points at \( t = 0, 1, \ldots, N - 1 \) and continuous frequency, \(-1/2 \leq f < 1/2\). Given a unit norm sequence, \( v^{(k)}_t \), defined on \([0, N - 1]\), the integral of its magnitude-squared Fourier transform over the Nyquist band \([-1/2, 1/2]\) is, by Parseval’s theorem, also unity. Slepian found the sequences with the largest fraction of their energy, \( \lambda_k \), in the frequency band \([-W, W]\). They are the discrete prolate spheroidal sequences, now known as Slepian sequences. They are the real, orthonormal eigenvectors of the Toeplitz matrix eigenvalue equation

\[
\lambda_k(N, W) v^{(k)}_t(N, W) = \sum_{n=0}^{N-1} \frac{\sin 2\pi W(t-n)}{\pi(t-n)} v^{(k)}_n(N, W).
\]

They are ordered by their corresponding eigenvalues, or concentrations on \((-W, W)\), \( 1 > \lambda_0 > \lambda_1 > \cdots > \lambda_{N-1} > 0 \). Also, the \( k \)th sequence has \( k \) zero crossings in \([0, N - 1]\) and the corresponding Slepian functions (Fourier transformed Slepian sequences),

\[
V_k(N, W; f) = \sum_{n=0}^{N-1} v^{(k)}_n(N, W)\exp(-i2\pi nf),
\]

have \( k \) zeros in \((-W, W)\). The functions are doubly orthogonal; orthonormal on \([-1/2, 1/2]\) and orthogonal on \((-W, W)\). Omitting the explicit dependence on \( N \) and \( W \), the \( k \)th sequence
is denoted by \( v_i^{(k)} \) and the eigenvalues by \( \lambda_k \). Of these, the first \( K \lesssim [2NW] \) eigenvalues are extremely close to one. For example, with \( NW = 6 \), \( 1 - \lambda_0 \approx 1.31 \times 10^{-15} \) and with \( NW = 10 \), \( 1 - \lambda_0 \approx 3.05 \times 10^{-26} \).

Slepian’s analysis defines the dimensionality of the time–frequency region \([0,N-1] \times (-W,W)\). The \( K = 2NW \) lowest order sequences form a complete basis for this subspace. In spectrum estimation problems, \( W \) is the bandwidth of the estimate and is a factor of \( C_R = NW \), typically 2–10, of the Rayleigh resolution, \( R = 1/(N\delta t) \) or \( 1/N \) for unit sampling. Many of the properties of Slepian sequences depend on the time–bandwidth product \( C_R \) as opposed to \( N \) and \( W \) individually. As \( N \) increases, properties of the Slepian sequences rapidly approach those of the continuous-time prolate spheroidal wave functions. Details are given in \([38]\), and appendix A of \([39]\) gives a summary. Finally, note that while (3.2) is analytically convenient, it is not recommended for numerical use. To compute, the sequences use the tridiagonal form given in appendix B of \([39]\).

The choice of \( C_R \) determines the number of tapers, \( K \sim [2C_R] \); the DoF \( \approx 2K \lesssim 4C_R \); and bias properties of the estimate. The broad-band bias, i.e. the bias at \( f \) contributed by frequencies outside the band \( (f - W, f + W) \), decreases exponentially with \( C_R \).

(b) Multitaper estimates

Multitaper spectrum estimates arose \([40]\) as a local least-squares solution of the Fredholm integral equation that expresses the discrete Fourier transform of the data in terms of the spectral representation of the process, described by Parzen as ‘The fundamental equation of spectrum estimation’. Assume that the data \( \{x(t)\}_{t=0}^{N-1} \) are a sample from a zero-mean, stationary, random process with a smooth spectrum, as expected under \( H_0 \). Also assume that the baseline spectrum is smooth enough that, with application of a suitable pre-whitening filter, the spectrum is approximately white, so that broad-band bias is not a concern.

To compute a multitaper estimate, take the \( K \) windowed Fourier transforms, \( y_k(f) \), known as eigencoefficients \( k = 0, 1, \ldots, K - 1 \), where

\[
y_k(f) = \sum_{t=0}^{N-1} x(t) v_t^{(k)} \exp(-i2\pi ft),
\]

and combine them into a spectrum estimate by iterating

\[
\hat{S}(f) = \frac{\sum_{k=0}^{K-1} d_k^2(f) y_k(f)^2}{\sum_{k=0}^{K-1} d_k^2(f)} \quad \text{and} \quad d_k(f) = \frac{\sqrt{\lambda_k} S(f)}{\lambda_k S(f) + B_k(f)}
\]

until an appropriate convergence criterion is satisfied. The \( \{d_k^2(f)\} \) are positive, frequency-dependent weights defined in §V of \([40]\), and \( B_k(f) \) is the bias on the \( k \)th eigenspectrum, which is bounded by \( \sigma^2(1 - \lambda_k) \). As both \( S(f) \) and the bias \( B_k(f) \) are unknown, this pair of equations is solved recursively by replacing \( S(f) \) with \( \hat{S}(f) \) from the previous iteration. For a white spectrum, the weights, \( d_k^2(f) = \lambda_k \), are the integral eigenvalues of the Slepian sequences. The quantity

\[
2\alpha = 2\sum_{k=0}^{K-1} d_k^2(f)
\]

is useful as an estimate of the DoF of the estimate at frequency \( f \). It is helpful to think of \( \alpha \lesssim K \), while \( C_R = NW \) throughout. Equation (3.5) expresses the spectrum estimate as an average of terms of the form \( |y_k(f)|^2 \). The complex eigencoefficients, \( y_k(f) \), generally have uncorrelated real and imaginary parts, so \(|y_k(f)|^2 = \Re\{y_k(f)^2\} + \Im\{y_k(f)^2\}\) will have a distribution proportional to \( \chi^2_2 \). With a locally white spectrum, the orthogonality of the tapers shows that the eigencoefficients at a given frequency are uncorrelated, i.e. \( \mathbb{E}\{y_j(f)y_k^*(f)\} = 0 \) for \( j \neq k \). Thus, if the weights are uniform, \( \hat{S}(f) \) will have a \( \chi^2_{2K} \) distribution. However, in the spectrum estimation problem, the eigencoefficients are all obtained from the same \( x(t) \) via (3.4) and the different \( y_k(f) \)'s have...
different serial correlations in frequency (see (4.10)). Before studying their fluctuations, note the following:

— They are a weighted average of $K$ tapered direct spectrum estimates. Because $\psi_t^{(0)}$ has an approximately Gaussian shape, Tukey’s ‘direct’ spectrum estimates (with Wiener’s Gaussian taper) are approximately the leading term of the series (3.5). Here, however, the data tapers $\psi_t^{(k)}$ are the Slepian sequences.

— Because Slepian sequences are an efficient basis for narrow-band processes, multitaper methods are easily adapted to situations when the observations consist of a sum of deterministic narrow-band processes embedded in a non-deterministic background. This paper is concerned only with stationary noise-like processes.

— They are consistent, i.e. with $W$ fixed, the variance of (3.5), $\approx S^2(f)/(2NW)$, tends to zero as $N$ increases.

— They are more efficient than other spectrum estimates (e.g. [40–42]).

— They are approximately maximum-likelihood [43].

Further details and extensions are described in [6,39,44–47]. The emphasis here is on the fluctuations of the spectrum estimate, particularly on large extremes, not the overall shape of the spectrum. Under $H_0$, the spectrum is constant, giving $E[\hat{S}(f)] = S(f)$, and defining the standardized estimate as

$$z(f) = \frac{\hat{S}(f)}{S(f)} ,$$

which has an expected value of one. In practice, one does not know $S(f)$ but uses a combination of pre- and post-whitening operations as in §8 to produce an approximately constant baseline over a band of interest. Fluctuations of spectrum estimates depend on two basic quantities: first, the probability density function of the spectrum estimates, reviewed in §3c, and second, the covariance function, $Cov[z(f), z(f + \Delta)]$, of the standardized estimate. As in the time domain, this gives the expected crossing rate, as described in §2.

For this problem, the natural unit of frequency is the Rayleigh resolution $R = 1/T = 1/(N\delta t)$ Hz, where $\delta t$ is the length of the sampling interval of the series in seconds. Throughout this paper, frequencies expressed in units of the Rayleigh resolution will be referred to as ‘Rayleighs’, denoted by $R$.

We emphasize that we consider $N$ and $\delta t$ as fixed, so $T$ and $R$ are also constant. The major variable is the time–bandwidth product $CR$ and, subordinate to it, the number of tapers, $K$.

(c) Distribution of multitaper spectrum estimates

If $z$ denotes $z(f)$, (3.7), at some frequency, the distribution of $z$ under $H_0$ can be written as a scaled central $\chi^2$ with $2\alpha$ DoF, with $\alpha \lesssim K \lesssim 2CR = 2NW$, to allow for the weights $\{d_2^2(f)\}$. Scaling a conventional $\chi^2_{2\alpha}$ random variable with the factor $1/2\alpha$ results in the gamma($\alpha, 1/\alpha$) probability density function,

$$p(z; \alpha) = \frac{\alpha^\alpha}{\Gamma(\alpha)} z^{\alpha-1} \exp(-\alpha z),$$

which is appropriate for the standardized estimate (3.7) as it has unit mean and variance $1/\alpha$. The complementary cumulative distribution function (CCDF),

$$Q(z; \alpha) = \int_z^{\infty} p(t; \alpha) \, dt = \frac{\Gamma(\alpha, \alpha z)}{\Gamma(\alpha)},$$

or

$$Q(z; \alpha) = \frac{p(z; \alpha)}{\alpha} \left( 1 + \frac{\alpha - 1}{\alpha z} + \frac{(\alpha - 1)(\alpha - 2)}{(\alpha z)^2} + \cdots \right),$$

is obtained by repeated integration by parts.
In practice, the spectrum is standardized as described in §8, and the significance of peaks is assessed using this distribution. The rationale for standardizing the spectrum estimate is that (i) the residual spectrum after pre-whitening and normalization is comparable to that of unit variance white noise, (ii) standardized spectra can be visualized on a linear scale in the y-axis, and (iii) standardizing both the spectrum and the units of frequency allows for standard tables of expected numbers of upcrossings, e.g. in Table 1 and the electronic supplementary material.

### (d) Why multitapers?

In this section, two examples of physical time series are given and analysed using two different spectrum estimators. These examples illustrate why periodograms cannot be used for scientific purposes and highlight the distinct advantages of the multitaper estimate for practical purposes.

Figure 1 shows two estimates of the spectrum of 1000 samples of barometric pressure at the Black Forest Observatory (BFO) in Germany. The upper curve is a periodogram and the lower curve is a multitaper estimate. The multitaper spectrum has not been smoothed. The two estimates differ by a factor of at least 1000 at most frequencies, increasing to more than $10^7$ over the top third of the frequency range.

Because one can place bounds on the broad-band bias of the multitaper estimate (using nothing more than the properties of the tapers and the Cauchy inequality; see [40]), the multitaper estimate cannot be overly biased. In this example, $C_R = 8$ with $K = 12$ tapers were used, so combining $1 - \lambda_0 \approx 5.26 \times 10^{-21}$ with the variance approximately $1586 \text{ Pa}^2$ and the sampling rate $\delta t = 10.0 \text{ s}$ bounds the broad-band bias to be smaller than $\sigma^2 (1 - \lambda_0) \delta t \lesssim 8.3 \times 10^{-17} \text{ Pa}^2 \text{ Hz}^{-1}$, which is 10 decades below the observed minimum of the estimate, implying that the multitaper estimate is close to the true spectrum. The crippling bias seen in the periodogram in Figure 1 is because the high sidelobes of the Fejér kernel cannot resolve such a large range.

One could attempt to smooth the periodogram in Figure 1 with any of the conventional smoothers, wavelets or splines, which would result in a consistent estimate, but the extreme bias of the estimate makes impossible the extraction of any valid scientific conclusion from these data. Note that because the periodogram is the Fourier transform of the familiar Bartlett autocovariance sequence, all indirect and parametric estimates that depend on sample autocovariances are linear functions of this estimate. As such, these estimates all have similar bias.

Further examples of this are: (i) figs 7 and 8 of [6] are from the same record a few days earlier—most segments of these data give similar results; and (ii) fig. 18 of [48] shows a spectrum of engineering quality-control data (additionally complicated by some outliers) where the difference is even larger. The BFO pressure data are not pathological, but typical of high-precision data with frequent sampling.

### Table 1. Expected upcrossing rates $U(z; K)$ per $10^5$ Rayleigh resolutions for a few common choices of $C_R$ and $K$. More extensive tables are in the electronic supplementary material.

<table>
<thead>
<tr>
<th>$P$ (%)</th>
<th>$z$</th>
<th>$U(z; 7)$</th>
<th>$z$</th>
<th>$U(z; 8)$</th>
<th>$z$</th>
<th>$U(z; 9)$</th>
<th>$z$</th>
<th>$U(z; 10)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>50.0000</td>
<td>0.953</td>
<td>14.217</td>
<td>0.959</td>
<td>11.549</td>
<td>0.963</td>
<td>12.696</td>
<td>0.967</td>
<td>10.308</td>
</tr>
<tr>
<td>90.0000</td>
<td>1.505</td>
<td>5822</td>
<td>1.471</td>
<td>4750</td>
<td>1.444</td>
<td>5240</td>
<td>1.421</td>
<td>4267</td>
</tr>
<tr>
<td>99.0000</td>
<td>2.082</td>
<td>846</td>
<td>2.000</td>
<td>692</td>
<td>1.934</td>
<td>764</td>
<td>1.878</td>
<td>623</td>
</tr>
<tr>
<td>99.9000</td>
<td>2.580</td>
<td>104</td>
<td>2.453</td>
<td>85</td>
<td>2.351</td>
<td>94</td>
<td>2.266</td>
<td>77</td>
</tr>
<tr>
<td>99.9900</td>
<td>3.041</td>
<td>12</td>
<td>2.870</td>
<td>10</td>
<td>2.733</td>
<td>11</td>
<td>2.619</td>
<td>9</td>
</tr>
<tr>
<td>99.9990</td>
<td>3.480</td>
<td>1</td>
<td>3.265</td>
<td>1</td>
<td>3.094</td>
<td>1</td>
<td>2.952</td>
<td>1</td>
</tr>
</tbody>
</table>
Figure 1. (a) A comparison of a multitaper spectrum with a periodogram. The curve marked ‘P’ shows a periodogram of 1000 samples of barometric pressure data sampled at $\delta t = 10.0$ s during 26 April 2006. The superimposed black curve shows its average over $\pm 1$ Rayleigh resolution. The curve marked ‘MT’ is a multitaper estimate made with a time–bandwidth product $C_R = 8.0$ and $K = 12$ tapers. The two estimates differ by a factor of at least 1000 across most of the frequency range and by more than $10^7$ at higher frequencies. Note that the approximately flat part of the spectrum above 35 mHz is due to quantization noise. (b) Part of a higher resolution spectrum of barometric pressure from Piñon Flat Observatory (PFO), California, USA. The raw data at 1 min resolution were pre-whitened with an AR-1 prediction error filter, then a multitaper spectrum with $C_R = 5$ and $K = 8$ tapers computed using all 123 454 available samples. To keep the spectrum level, it has not been corrected for pre-whitening, so the scale is arbitrary. The average and significance levels (for a central $\chi^2_{15}$) are estimated from the lower 5% levels and checked against the median. The vertical dashed lines marked $0_{T7}$ and $0_{S7}$ are normal modes of the Earth, and those marked $P_{2,7}$, $P_{0,8}$ and $P_{3,7}$, the central frequency of solar $p$-modes. (Online version in colour.)

Figure 1b shows a small part of a spectrum made using 85.7 days of data from a similar microbarograph at PFO in California, USA, and it may be seen that it contains numerous discrete lines whose frequencies are closely related to those of the normal modes of the Earth and Sun. (Some of the others are splittings of the solar modes by $\pm 1$ cycle/day, 11.57 $\mu$Hz.) The frequencies in pressure are at slightly lower frequencies than those of the solar or seismic modes, possibly due to the Doppler effect. There are many solar $p$-modes on this range but, for simplicity, only those with spherical harmonic degree $l \leq 3$ plus two normal modes of the Earth are marked. The largest peak, corresponding to the solar $P_{2,7}$ mode at 1250.555 $\pm 0.003$ $\mu$Hz [49], is about 3.78 times the estimated background level. This spectrum was made using $K = 8$ tapers, so is approximately $\chi^2_{16}$ distributed. The probability of such a peak occurring by chance is $\lesssim 4 \times 10^{-7}$, very improbable in a sample from 123 000 points. There are 14 other peaks above the 99% level in this figure. Noting that the 120 $\mu$Hz frequency span of figure 1 corresponds to 890 Rayleighs and looking forward to figure 2 in §4, if the data were random one would expect an average of 5.3 peaks above the 99% level, not 15. This is one of many such examples. Fig. 11 of [6] and fig. 1 of [11] show even higher peaks in the BFO pressure spectrum and in that of solar wind density from the Advanced Composition Explorer (ACE) spacecraft, respectively. Both peaks correspond to the solar $P_{0,2}$ mode, so seeing solar mode peaks in terrestrial data is common.

The point of these examples is that multitaper estimates allow study of scientific phenomena that was not feasible with the estimates available in the 1970s. A major driver of spectrum estimation methodology has been the improvements in data quality, sampling rates and sample size. For example, in 1967 appendix 1 of Bogert et al. [50] compares hand-digitized data with three- and four-digit voltmeters (10–13 bits) in contrast to the 24-bit data used here.
4. Upcrossing rate of spectrum estimates

In time-domain crossing problems, the rates depend, as in (2.1), on the derivatives of the autocovariance function or, equivalently, the moments of the power spectrum. Here, the analogous functions are the antecovariance and antespectrum. In [46,51], the latter was defined as the spectrum of a spectrum estimate. To avoid confusion with the ordinary time-domain autocorrelation, we propose calling the autocorrelation of the spectrum estimate the ‘antecorrelation’.

The antecovariance of the estimate is defined to be

\[ \Upsilon(\Delta) = \text{Cov}\{z(f + \Delta), z(f)\} \quad (4.1) \]

and, by analogy, the antevariance is \( \Upsilon(0) = 1/\alpha \).

Now denote the upcrossing rate of a spectrum estimate of some level \( z \) by \( U(z; \alpha) \). Crossing rates for \( \chi^2 \) processes are found in at least five basic papers [20–22,24,52], all with slightly different assumptions and derivations. These appear to differ because, for example, Barakat’s formula gives the total number of level crossings and uses the covariance of the \( \chi^2 \) process, Sharpe’s gives upcrossings and uses autocovariance of the constituent Gaussian processes, etc., but they all agree. The average number of upcrossings of level \( z \) per Rayleigh resolution is

\[ U(z; \alpha) = \psi \sqrt{z^2/\pi \alpha} p(z; \alpha). \quad (4.2) \]

Here,

\[ \psi = \frac{1}{N} \left\{ -\left(\frac{d^2}{d\Delta^2}\Upsilon(\Delta)\right)|_0 \right\}^{1/2} \quad (4.3) \]

in which the factor of \( \Upsilon(0) \) in the denominator converts the covariances to correlations, and the \( N \) standardizes frequency to units of Rayleigh resolution. The term \(- (d^2/d\Delta^2)\Upsilon(\Delta)\)|\(_0\) is the variance of the derivative \( z' = (d/df)z(f) \).

Figure 2. (a) Expected number of upcrossings as a function of level for a periodogram (solid curve marked P), see equation (4.8), and for the multitaper estimate with \( C_R = 5 \) and \( K = 8 \) tapers (solid curve marked MT), see equation (4.23), and substitute the exact value of \( \psi \), (4.22). The dashed curves represent the result of simulations with \( N = 10^5 \) samples. Note that there is about a 40\% chance of an upcrossing of the mean level per Rayleigh resolution for the periodogram, thus partly explaining its grassy appearance seen, for example, in figure 1, which has dropped to about 10\% for the multitaper. The scale at the top of the plot represents the threshold used for the standardized spectrum for the multitaper (upper, marked MT) and periodogram (lower, marked P). (b) The average dwell band in Rayleighs, using the same scheme. Dotted lines marked \( W \) and \( 2W \) are added for reference. The overall impression is that the periodogram has lots of narrow spurious peaks, while the multitaper has fewer, wider peaks. (Online version in colour.)
$U(z; \alpha)$ gives the average number of upcrossings of the level $z$ by a spectrum estimate per Rayleigh resolution (i.e. the upcrossing rate). This formula also determines the average 'width' of random peaks. Define a dwell band, $D(z)$, as the average frequency difference from the point where the estimate crosses $z$ with a positive slope and the next frequency where the estimate returns to a level below $z$. For small $z$ such a dwell band often has more than one local maximum, but at high significance levels the dwell band usually contains only a single local maximum so peak bandwidth is used almost synonymously. The average number of such excursions multiplied by their mean bandwidth gives the total frequency range above $z$, the CCDF, so
\[
Q(z; \alpha) = U(z; \alpha)D(z).
\] (4.4)
This formula originated with Pratt [53, eqn (3.41)].

(a) Direct spectrum upcrossing rate

The antecorrelation between standardized direct estimates (3.1) offset by $\Delta$ in frequency was given in eqn (18) of [51] and [46, §4.4] as $\Upsilon_D(\Delta) = |\hat{D} \ast \hat{D}^*|^2$, where $\hat{D}$ is the Fourier transform of $D(t)$. For white, zero-mean, complex-valued processes, it becomes
\[
\Upsilon_D(\Delta) = \left| \sum_{t=0}^{N-1} D^2(t) \exp(-i2\pi t \Delta) \right|^2 .
\] (4.5)
Additional information on these correlations may be found in §5.6 of [54]. One must also remember that our theory assumes that frequency is a continuous parameter, whereas, in practice, spectra are computed on a discrete mesh. However, as mentioned in §2, finite discrete Fourier transforms of data are entire functions of frequency. If the transforms are computed with an FFT that has been zero-padded by more than a factor of $\pi$, Lagrange interpolation between the mesh points converges rapidly. Our simulations were zero-padded by a factor of 8, so there should be only slight differences between the continuous theory and discrete implementation.

To get the upcrossing rate, use (4.2) substituting $\alpha = 1$, so that $p(z; 1)$ is just $\exp(-z)$. The constant $\psi$ then becomes, upon application of Isserlis’s formula,
\[
\psi_D = \frac{2\pi}{N} \left[ \sum_{s,t=0}^{N-1} (s-t)^2 D^2(s)D^2(t) \right]^{1/2} .
\] (4.6)
and the upcrossing rate is
\[
U_D(z; 1) = \frac{\exp(-z)}{N} \left[ \frac{2\pi}{N} \sum_{s,t=0}^{N-1} (s-t)^2 D^2(s)D^2(t) \right]^{1/2} .
\] (4.7)

For example, straightforward substitution of the periodogram estimate discussed in §3 into (4.5) gives antecorrelation
\[
\Upsilon_P(\Delta) = \left| \frac{\sin N\pi \Delta}{N \sin \pi \Delta} \right|^2
\]
so the mean number of upcrossings of a level $z$ for a periodogram per Rayleigh resolution becomes simply
\[
U_P(z) = \left[ \frac{\pi z}{3} \right]^{1/2} \left( \frac{N^2 - 1}{N} \right)^{1/2} \exp(-z) \approx \left[ \frac{\pi z}{3} \right]^{1/2} \exp(-z).
\] (4.8)
Further, as in §4, the CCDF is just $\exp(-z)$. This implies that the average dwell band (in Rayleighs) of a peak crossing level $z$ is
\[
D(z) \approx \sqrt{\frac{3}{\pi} z^{-1/2}} .
\] (4.9)
(b) Multitaper antecorrelation

Evaluating \( \Upsilon(\Delta) \) for a multitaper estimate consists of expanding the equations for the spectrum estimates as in [46, §4.4] and applying Isserlis’s formula. (The simplest derivation found so far is for a white, complex-valued time series.) Because the Slepian functions are so concentrated in frequency, for slowly varying spectra, \( \Upsilon(\Delta) \) is indistinguishable from that of a globally white noise process ([51] or [45, §6.6]). The covariance between the eigencoefficients separated in frequency by \( \Delta \), called \( \Lambda_{jk}(\Delta) \) in [40], is

\[
\text{Cov}(y_j(f + \Delta), y_k^*(f)) = \sum_{n=0}^{N-1} v_n^{(j)} v_n^{(k)} \exp(i2\pi n\Delta), \tag{4.10}
\]

where the superscript * denotes complex conjugate. When \( \Delta = 0 \), this becomes the Kronecker delta, \( \delta_{jk} \), because the Slepian sequences are orthonormal. Next, using the expected weights, \( d_k^2(f) = \lambda_k \), \( \Upsilon(\Delta) \) becomes

\[
\Upsilon(\Delta) = \frac{1}{(2C_K)^2} \sum_{n,m=0}^{N-1} \left[ \sum_{j=0}^{K-1} \lambda_j v_n^{(j)} v_m^{(j)} \right]^2 e^{-i2\pi \Delta(n-m)}. \tag{4.11}
\]

For \( K \gg 2C_K \), the term in brackets approximates the matrix (3.2), so

\[
\Upsilon(\Delta) \approx \frac{1}{(2C_K)^2} \sum_{n,m=0}^{N-1} \left[ \frac{\sin 2\pi W(n-m)}{\pi(n-m)} \right]^2 e^{-i2\pi \Delta(n-m)}, \tag{4.12}
\]

and substituting \( n - m = \tau \) gives

\[
\Upsilon(\Delta) = \frac{1}{(2C_K)^2} \sum_{\tau=-N}^{N} (N - |\tau|) \left[ \frac{\sin 2\pi W\tau}{\pi \tau} \right]^2 e^{-i2\pi \Delta \tau}. \tag{4.13}
\]

Note that: first, the term in brackets in (4.11) is a truncated spectral decomposition of (3.2). Because the eigenvalues become very small for \( k > 2C_K \), (4.12) is accurate. Second, \( \Upsilon(\Delta) \) drops rapidly to almost zero at \( \pm 2W \) [46, fig. 3]. Third, real-valued processes have half the DoF within \( \pm 2W \) of the origin and Nyquist frequency than they have at other frequencies. Fourth, the standardizing factor \( 2C_K = 2NW \) is the trace of the sinc matrix in (3.2) and also the sum of the squared weights, here \( \lambda_k \). Fifth, observe that (4.13) is just the Fourier transform of a product of a triangular Bartlett window \( (N - |\tau|) \) for \( |\tau| \leq N \), and zero elsewhere) and a Fejér kernel. These are a Fourier transform pair and, keeping the same order, the convolution becomes

\[
\Upsilon(\Delta) = \frac{1}{(2C_K)^2} \int_{-2W}^{2W} \left( \frac{\sin N\pi(\Delta - \xi)}{\sin(\pi(\Delta - \xi))} \right)^2 [2W - |\xi|] \, d\xi. \tag{4.14}
\]

This shows that the antecorrelation is always positive. The Fejér kernel in (4.14) has its first zeros at \( \xi = \pm 1/N \), so when \( W \) is larger than this, say \( W \geq 3/N \), \( \Upsilon(\Delta) \) is a smoothed Bartlett window of total width \( 4W \). This argument is later reconsidered in §6 in connection with the shape of peaks in spectrum estimates. Sixth, recall from (2.1) that the basic crossing rate depends on the second
derivative of the antecorrelation and in this case the asymptotic form simplifies considerably. Evaluating the second derivative of (4.13), again with frequency in Rayleighs, the \((\pi \tau)^2\) terms cancel, and one obtains the variance of the derivative \(z'(f)\)

\[
\sigma^2_d = \frac{1}{N^2} \text{Var} \left\{ \frac{d}{df} z(f) \right\} = -\frac{1}{N^2} \frac{d^2}{d\Delta^2} \gamma(\Delta) \bigg|_0
\]

(4.15)

\[
= \frac{1}{(NC_R)^2} \sum_{\tau = -N}^N (N - |\tau|) \sin^2 2\pi W \tau
\]

(4.16)

\[
= \frac{1}{2C_R^2} \left\{ 1 - \left[ \frac{\sin 2\pi C_R}{N \sin 2\pi W} \right]^2 \right\}
\]

(4.17)

\[
\approx \frac{1}{2C_R^2} \approx \frac{2}{\alpha^2}
\]

(4.18)

using the Fejér kernel again to simplify (4.17). The second term in (4.17) is zero for non-zero positive integer and half-integer values of the time–bandwidth product, \(C_R\), and is otherwise \(\approx [\sin(2\pi C_R)/(2\pi C_R)]^2\). The second approximation in (4.18) holds exactly when \(\alpha = 2C_R\).

(c) Distribution of \((d/df)\hat{S}(f)\)

It is worthwhile to note that the variance of the derivative of the spectrum estimate can be derived by means of formula (4.18) for the standardized estimate \(z(f)\),

\[
\text{Var} \left\{ \frac{(d/df)\hat{S}(f)}{N^2} \right\} = S^2(f)\sigma^2_d \approx \frac{S^2(f)}{2C_R^2},
\]

(4.19)

where the \(1/N^2\) converts to Rayleighs. One may also evaluate the variance of \(z'(f)\) directly without resorting to the antecorrelation and, after some tedious algebra, one obtains the same expression. The joint distribution of \(z(f)\) and its derivative, \(z'(f)\), and the marginal distribution of \(z'(f)\), are given in [21,24]. Define

\[
\zeta = \frac{z'}{z} = \frac{1}{N} \frac{d}{df} \ln \hat{S}(f) = \frac{1}{N} \frac{\hat{S}'(f)}{\hat{S}(f)},
\]

(4.20)

so the marginal distribution of \(z'\) becomes

\[
p_z(\zeta; \alpha) = \frac{\alpha^2}{\sqrt{\pi} \Gamma(\alpha)} \left[ \frac{\alpha^2 |\zeta|}{2} \right]^{\alpha-1/2} K_{\alpha-1/2}(\alpha^2 |\zeta|),
\]

(4.21)

where \(K\) is MacDonald’s function [55, ch. 3, pp. 78–80]. This reduces to a Laplace distribution when \(\alpha = 1\). The conditional distribution of \(\zeta\), however, is Gaussian with zero mean (for a white spectrum) and variance \(\sigma^2_d/z\).

To conclude this section, note that, if one needs an estimate of the derivative of a spectrum, it is preferable to use a scaled version of the quadratic-inverse coefficient \(b_1(f)\) defined in [39] because the variance of this estimate is a factor of approximately \(2C_R/3\) lower than that of the direct derivative.

(d) Multitaper crossing rates

Direct substitution of the antecorrelations (4.17) and the antiveariance (4.1) into (4.3) gives

\[
\psi = \frac{2\pi}{N} \left[ \frac{\sum_{n,m=0}^{N-1} |(n - m) \sum_{j=0}^{K-1} \lambda_j v_n^{(j)} v_m^{(j)}|^2}{\sum_{k=0}^{K-1} \lambda_k^2} \right]^{1/2}
\]

(4.22)
Then, using \( K \approx 2C_R \) gives \( \psi \approx \sqrt{2\alpha} \) so that the upcrossing rate is approximately

\[
U(z; \alpha) \approx \sqrt{\frac{z}{\pi \alpha}} p(z; \alpha)
\]

(4.23)

per Rayleigh resolution. Using only the first term in the expansion (3.10) of \( Q(z; \alpha) \) shows that the number of spurious peaks is approximately \( \sqrt{z/\pi} Q(z; \alpha) \). Returning briefly to the standard antecovariance estimate (4.11) for smaller \( K \)'s or approximating \( d^2 = 1 \), the second derivative at \( \Delta = 0 \) can be inconvenient to evaluate when \( N \) is large. It can be accurately approximated using a moderately small \( N \), 100–500, as long as \( C_R \) is kept constant. Table 1 shows upcrossing levels \( z \) and the number of upcrossings \( U(z, K) \) per 100,000 Rayleighs for a few choices of \( C_R \) and \( K \). See also §8 and the electronic supplementary material.

It is common to average several multitaper estimates from different data segments. Assuming \( J \) non-overlapping segments, each with \( \alpha \approx K \), the crossing rate becomes

\[
U_J(z; \alpha) \approx \psi \sqrt{\frac{z}{2\pi \alpha J}} p(z; \alpha J) \approx \frac{1}{\alpha} \sqrt{\frac{z}{\pi J}} p(z; \alpha J),
\]

(4.24)

where the last form uses the asymptotic value of \( \psi \).

5. Crossing rate simulations

Combining the results of §4a, d for comparison, the approximations (4.8) and (4.9) are remarkably accurate. Figure 2 shows the level crossing rates of a periodogram of white Gaussian noise with \( N = 100,000 \) samples zero-padded to \( 8N \) samples superimposed on the prediction from (4.8). There are no adjustable parameters. The multitaper level crossing rates use the same pseudodata.

Common statistical lore says, correctly, that periodogram estimates are uncorrelated when separated by one Rayleigh resolution. An informal survey of about a dozen statisticians and graduate students having some experience with spectrum estimation shows that this is generally interpreted as implying that one should expect an average of one peak above the 99% level in 100\( R \). This, however, is incorrect. The 99% level occurs at \(-\ln 0.01 \approx 4.605\) (recalling that \( Q(z; 1) = \exp(-z) \)), so (4.9) predicts a mean width of 0.4554\( R \) for such peaks. The expected maximum of 100 independent exponential random variates \( \approx \ln 100 + \gamma + 0.01 \approx 5.192 \), where \( \gamma \approx 0.57721 \) is Euler’s constant, exceeds the 99% point of the estimate. Consequently, to make up the 1% of estimates that, on average, exceed the 99% level, one needs 2.20 upcrossings. Thus, one must have an upcrossing of the 99% level about every 45.5\( R \). This effect clearly becomes worse at higher significance levels. Naive intuition again suggests that there should be one peak above the 99.99% level in 10,000 Rayleighs but there are 3.1 times more than this. To conclude, a very large number of narrow spurious peaks should be expected when the periodogram is used.

The fact that the periodogram has more false peaks than commonly expected reinforces the opinion that it should never be used. In addition, it has an exponential distribution, known to have terrible sampling properties: in a sample of size \( N \), one expects the minimum to be \( 1/N \) and the maximum \( \approx \ln N + \gamma + 1/N \). Thus the range, ratio of maximum to minimum spectrum estimates, in a periodogram of a white noise process will be approximately \( N \ln N + \gamma N + 1 \). A periodogram with \( N = 10^5 \) from a ‘simple’ process where bias is not a problem would have a range of approximately \( 1.2 \times 10^6 \) beyond the true range.

Figure 3a shows multitaper crossing rates for \( C_R = 8 \) as the number of tapers varying from 2 to 14 and, with the default \( K \) of 14, the crossing rate is approximately \( 10^{-8} \) for \( z \approx 3.5 \).

Figure 3b begins to explain why the false discovery rate problem in spectrum estimates has remained so elusive. The standard work on false discovery rates [56] assumes that the number of possible hypotheses to be tested is known in advance but, with spectrum estimates, this does not seem possible. As we indicate in figure 2 and §3d the width of high peaks, the dwell band, decreases with significance level so that the effective number of frequency cells in a fixed bandwidth increases with \( z \). Here, for a bandwidth \( B = 1000R \), we show the expected number of false peaks \( p = B \cdot U(z; \alpha) \) and the approximate number of peak positions \( q = B/D(z) \). As a
first approximation, the number of arrangements of the false peaks is approximately \( \binom{q}{p} \). At the conventional 95% significance level, this simple example gives a non-testable approximately \( 10^{51} \) possible patterns. Looking ahead to the solar wind example in §8, suppose one is searching for \( 2l + 1 \) equally spaced modal singlets for, say, \( l > 3 \), and one sets the threshold \( z \sim 2.4 \), approximately the 99.9% level. One only expects a single false peak in a 1000\( \mathcal{R} \) band, so falsely detecting all peaks at 99.9% is extremely unlikely (e.g. fig. 6 of [12]).

6. Width and shape of peaks

While level crossing problems have been studied since Rice’s 1944 papers, the shape of extreme peaks in the spectrum of a random process has generally been assumed to be basically quadratic plus correction terms, that is, a ‘Slepian’ process. This appears to be reasonable for direct estimates of the form (3.1), where \( D(t) \) is a Hamming or similar taper, but not for multitaper estimates. The solid black line in figure 4 shows the average of 50 large peaks, all above the 99.9% level, occurring in a multitaper estimate and a quadratic is not an obvious approximation. The dotted black line in figure 4 shows the antecorrelation of the estimate. Blachman’s [8] analysis showed that the average shape of large excursions in Gaussian noise looks like the autocovariance function of the process.

Blachman’s derivation was done for continuous-time processes but, again, the methodology can be adapted to the frequency domain. His analysis is only strictly valid for Gaussian processes, but is asymptotically valid as one increases the DoF, and for spectra with \( \geq 20 \) DoF, his results apply reasonably well. By virtue of the comments in §4b, the antecorrelation looks superficially like a smoothed Bartlett window of width \( \pm 2W \) and is an entire function, which can be seen from (4.13), so the slope at the peak is continuous. Here, in contrast to Blachman’s relatively pessimistic conclusion ‘Unfortunately, the shape that we have found is identical with that of this filter’s response to a signal to which it is matched’, the shapes of random peaks in a multitaper spectrum are triangular with a width of \( \pm 2W \), whereas those from periodic signals are roughly a rectangle of width \( \pm W \) (corresponding to the shape of the spectral window) (figure 4b).
Figure 4. (a) The solid centre line shows the average shape of the 50 largest peaks in a multitaper spectrum estimate of white Gaussian noise data with $N = 100,000$, $C_R = 5$ and $K = 8$. The theoretical autocorrelation (dotted black line), scaled to match the average at the origin, is superimposed on the expected value, $z = 1$. In this figure, the light grey band shows the range, minimum to maximum around the 50 large peaks; the darker grey band $\pm 1$ s.d.; the dashed vertical lines $\pm W$ and $\pm 2W$ bands; and the horizontal dashed lines the mean, $z = 1$, $\pm 1\sigma = \pm \sqrt{\frac{1}{K}}$ and the 99% significance level. At this level, the average peak width is about $W$ Rayleighs wide, and the shape is that of the antecovariance function (4.11). (b) The average of 50 genuine peaks in a simulated process, $x_t = w_n + \frac{1}{50}\sum_i \cos(2\pi f_i n + \phi_n)$, where $w_n$ is a unit variance white noise process and random phase $\phi_n$, harmonics $f_i$, $i = 1, \ldots, 50$, are equally spaced in frequency. Expected weights ($d_i^2(f) = \lambda_i$) for a white noise process were used to compute the multitaper estimate. The overall shape is seen to be rectangular, with passband ripples corresponding to that of the multitaper spectral window. (Online version in colour.)

7. Distribution of peaks in a band

When attempting to identify periodic components in a spectrum, one generally plots the estimate over a moderate range. For example, for the solar wind data presented later on, one commonly generates between 50 and 200 plots for a single series, perhaps a few thousand Rayleigh resolutions per plot. The questions here are: (i) how many ‘large’ peaks should one expect to find in one such plot and (ii) given that one has, say, $M$ plots to examine, what is the most that one expects to find? Because one is dealing with high significance levels, large peaks should be rare and the number of such peaks Poisson. More formally, the arguments given in [57, §4.8] and elaborated on in [52, pp. 302–303] and heuristically in [10] imply that the number of upcrossings in an interval will be Poisson although it is probably wise, see §6, to exclude spacings less than the bandwidth $W$. By (4.2), the expected number of peaks per plot that exceed $z$ should be $\Lambda(z) \approx J \cdot U(z)$. The maximum over the $M$ plots is the extreme of $M$ independent Poisson trials. This appears to be a surprisingly difficult problem, but a satisfactory asymptotic expansion appears in [58],

$$x_0 = \frac{\ln M}{W(\ln M/\epsilon \Lambda(z))},$$

(7.1)

where, in this formula, $W(\cdot)$ is Lambert’s $W$-function.

8. A real data example

Returning to the space physics example mentioned in the Introduction, figure 5 shows an estimate of the spectrum of the proton density in the solar wind measured by the SWEPAM instrument [60] on the ACE spacecraft between 5 February 1998 and 23 November 2005, a span of 1899 days. (The ACE spacecraft is in a ‘halo’ orbit around the first Lagrange point, $L_1$, about $1.5 \times 10^6$ km sunward of the Earth.) These data were low-pass filtered and decimated to $\delta t = 16$ min and spectra computed on two slightly overlapping sections that, combined, have $\approx 22$ DoF. Because
Figure 5. Part of the solar wind proton density spectrum from the ACE spacecraft and its distribution. (a) An estimate of the standardized spectrum between 216 and 224 µHz, corresponding to periods of approximately 75 min. The horizontal dashed lines show estimated significance levels for a $\chi^2_{22}$ distribution standardized to unit noise power. The six highlighted vertical lines mark the ‘case 1’ candidate frequencies from [59, p. 463] from the Global Oscillations at Low Frequencies (GOLF) instrument on the Solar and Heliospheric Observatory (SOHO) spacecraft. Here, the first five all have detections above the 99.99% significance level within ±20 nHz. A histogram of the spectrum estimates is shown on (b) by the solid line. The best fitting mixture of central and non-central $\chi^2_{22}$ distributions is shown by the dotted line, with the central component shown by the short-dashed (marked C) and the non-central (marked NC) by the long-dashed curve. (Online version in colour.)

good frequency resolution is needed, $C_R = 3.5$ and $K = 6$ were chosen, giving a bandwidth $W = \pm 19.9$ nHz = ±3.5$R$. Bartlett’s $M$-test (applied across frequency) implies that the probability of this spectrum having a homogeneous distribution is approximately $3 \times 10^{-44}$.

Proceeding in the spirit of analysis first presented in §7 of [4], this spectrum is fitted using a mixture of central and non-central $\chi^2_{22}$ distributions, a histogram of which is shown in figure 5b. The motivation for this is that the data are assumed to be a mixture of ‘noise’ and discrete solar modes, so the distribution of the spectrum will be central $\chi^2$ at ‘noise’ frequencies and non-central at modes. The fit was made by assuming that the noise power is uniform across frequency and constraining the mean of the two distributions to match the average across frequency. This leaves two parameters to be estimated, the non-centrality parameter $\lambda$ and the fraction of the spectrum in modes, $\epsilon$. The non-centrality parameter $\lambda$ in standard notation is expressed as a signal-to-noise ratio $\rho = \lambda/\nu$, where $\nu = 22$, the DoF, is the mean of the central component. This fit was made by choosing $\rho$ and $\epsilon$ to minimize the Kolmogorov–Smirnov $D$ measure of goodness-of-fit between the empirical and mixture distributions (figure 6). The minimum, $D_{\text{min}} = 0.02216$, is close to the 60% point of the distribution, that is, about as good as one can expect, with $\hat{\rho} \approx 0.80 \pm 5\%$ and $\hat{\epsilon} \approx 0.50 \pm 10\%$. The corresponding probability density functions are shown in figure 5b and even the fit to the histogram appears to be reasonable. Incidentally, the best fitting central distribution is a very poor fit to the observations with $\text{Prob}(D) \lesssim 10^{-22}$. This decomposition implies that one cannot use the average spectrum or even the median as a reliable estimate of the noise level or significance levels of such spectra. Here, as in earlier work [4, app. B], the lower 5% point is used to standardize the spectrum. To produce the ‘standardized’ spectrum shown here, first remove a linear trend from $\ln(S(f))$ over the band, then scale so that theoretical and empirical 5% points match. Numerous tests of this procedure have been done and, at the large sample sizes used in such applications, this procedure has become our default process.

In addition to having spectra where approximately one-half of the frequencies are modal, the raw time-domain data are highly non-Gaussian. Specifically, the fourth moment $\mu^4/\sigma^4 \sim 120$ and the data contain frequent bursts of activity. Usually, there is no indication of instrument or communications failure, so these bursts cannot simply be dismissed as outliers. Because various competent physicists have asked whether the peaks in the spectrum could originate in the non-Gaussian character of the data as opposed to a solar source this must be examined carefully. Two approaches were attempted: first, computing the spectra both with the raw data and then after interpolating the largest 1% or 2% of the extremes: the interpolation does not seem to
The upcrossing rate, the observed rate divided by \( U_J \) of the signal-to-noise ratio, \( \rho \).

\[
\sigma \text{ at least 7 with theory.}
\]

\( \text{for Gauss noise (Gauss), randomly permuted (RP) and original data. These statistics are from the average of two spectrum estimates of the density data for two sections, each with} \]

\( \text{C}_R = 5 \) and \( K = 8 \). These data are mostly later than those used in figure 5, and run from January 2002 through December 2009, both in solar cycle 23. The top axis gives \( Q(z) \) in per cent.

The lower curve shows the same ratio when the order of the data has been randomly permuted. It terminates at \( z = 3.08 \), the largest peak in this spectrum.

\( \text{Figure 6. Panel (a), a continuation of figure 5, is a contour plot of the probabilities of the Kolmogorov–Smirnov D as a function of the signal-to-noise ratio, } \rho = \lambda / \nu, \text{ and mixing fraction, } \epsilon. \text{ The best fit is near the 50\% point. Panel (b) shows the relative upcrossing rate, the observed rate divided by } U_J(z) \text{ from (4.24) versus the level } z. \text{ This was done for the average of } J = 2 \text{ spectra of ACE density data on two non-overlapping segments made with } C_R = 5 \text{ and } K = 8. \text{ These data are mostly later than those used in figure 5, and run from January 2002 through December 2009, both in solar cycle 23. The top axis gives } Q(z) \text{ in per cent.} \]

\( \text{The lower curve shows the same ratio when the order of the data has been randomly permuted. It terminates at } z = 3.08, \text{ the largest peak in this spectrum.} \)

Table 2. Expected, \( U(z) \), and observed counts of upcrossings per \( 10^5 \) Rayleigh resolutions in the solar wind data. The column headed \( \sigma \) is \( \sqrt{U(z)} \), the standard deviation expected with Poisson statistics. The three right-hand columns give observed counts for Gaussian noise (Gauss), randomly permuted (RP) and original data. These statistics are from the average of two spectra made with \( C_R = 5 \) and \( K = 8 \) tapers. At and above the 99\% level, the number of peaks in the original data exceeds the expected by at least 7\( \sigma \).

<table>
<thead>
<tr>
<th>( P ) (%)</th>
<th>( Q ) (%)</th>
<th>( z )</th>
<th>( U(z; \alpha) )</th>
<th>( \sigma )</th>
<th>Gauss</th>
<th>RP</th>
<th>original</th>
</tr>
</thead>
<tbody>
<tr>
<td>50.000</td>
<td>50.000</td>
<td>0.959</td>
<td>13095</td>
<td>114</td>
<td>12840</td>
<td>11649</td>
<td>11428</td>
</tr>
<tr>
<td>90.000</td>
<td>10.000</td>
<td>1.471</td>
<td>5385</td>
<td>73</td>
<td>5297</td>
<td>4640</td>
<td>5163</td>
</tr>
<tr>
<td>99.000</td>
<td>1.000</td>
<td>2.000</td>
<td>784</td>
<td>28</td>
<td>781</td>
<td>743</td>
<td>1024</td>
</tr>
<tr>
<td>99.900</td>
<td>0.100</td>
<td>2.453</td>
<td>97</td>
<td>10</td>
<td>101</td>
<td>92</td>
<td>170</td>
</tr>
<tr>
<td>99.990</td>
<td>0.010</td>
<td>2.870</td>
<td>11</td>
<td>3</td>
<td>10</td>
<td>11</td>
<td>34</td>
</tr>
<tr>
<td>99.999</td>
<td>0.001</td>
<td>3.265</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>9</td>
</tr>
</tbody>
</table>

make much difference. \( \text{Second, randomly permuting the order of the data before computing the spectrum: this destroys the peaks. Table 2 shows the results of such an experiment. This table shows the statistics of two spectrum estimates of the density data for two sections, each with } \ N = 130 000 \text{ samples with } \delta t = 16 \text{ min. The first ran from 19 January 2002 to 2 January 2006, the second from 2 January 2006 to 17 December 2009. The spectra were computed with } \ N W = 5, \text{ and } K = 8 \text{ and deleting data below } 3.0 \mu \text{Hz and above } 470 \mu \text{Hz because the pre-whitening and anti-alias filter left some curvature at the low and high ends of the spectrum, respectively. This left a total bandwidth of } 116 563 \text{ Rayleighs. The first four columns show probability levels, } P \text{ (in per cent); the threshold, } z; \text{ } U(z; \alpha); \text{ and s.d. } \sqrt{U(z; \alpha)} \text{. The last three columns show the results of simulations with Gaussian pseudo-random white noise, with the raw solar wind data with their time order randomly permuted, and, finally, the original data. The crossing rates, given in units of upcrossings per } 10^5 \text{ Rayleighs, for the Gaussian and randomly permuted data agree well with theory.} \)
Numerous studies have shown that Fourier transforms tend to become complex Gaussian and, empirically, this appears to be the case here, despite the fact that the pre-whitened data have kurtosis of 161 and 259 for the two data sections. For the original data, however, the situation is different and, for example, at the 99.9% significance level, one expects 97 peaks and observes 170, a 7.4σ discrepancy from Poisson. Averaging the two sections gives a sample of bandwidth 58 280 Rayleighs and a nominal $\chi^2_2$ distribution. Here, $\max(z) = 3.77$, corresponding to $Q \approx 4.6 \times 10^{-7}$ and, as shown in figure 6b, above $z \sim 2$ the observed crossing rate is between 2 and 20 times expected. This increase in significance is easily explained if most of the peaks in the spectrum are not random, but real, so the peaks occur at the same frequencies in both sections.

Continuing, the data were pre-whitened with an AR-1 prediction error filter, the temporal order of the residuals randomly permuted, and spectra computed. In this spectrum, $\max(z) = 3.08$, just above the 99.99% level, as one would expect from samples of this size. Note that the pre-whitened residuals are extremely non-Gaussian, but repeating the experiment with Gaussian pseudo-random noise gives results within 1σ of expected at significance levels above 99%. At lower significance levels, there are often several peaks per upcrossing so the comparisons are not as good. We conclude that there are many systematic peaks in this spectrum and these peaks are a result of the temporal order of the data, not its distribution.

One of the referees asked whether ‘one could use an assumed modal propagation matrix and frequency domain regression to isolate the separate modes’ and the answer seems to be ‘maybe’. The multitaper harmonic $F$-test [40] is a frequency domain regression and was used in [3] followed by coincidence detection. This works reasonably for short series but with long time spans its effectiveness is limited by three factors: (i) detection is limited by the shifts in solar mode frequencies with solar activity; (ii) the spacecraft orbit and factors such as the inclination of the ecliptic on the Sun’s equator causes splitting on odd-parity modes; and (iii) the intrinsic propagation of modes in a plasma (see [11]), particularly the fact that modes in a plasma spiral [61], combined with changes in solar wind density and velocity and the heliographic range appear to impose limits on the time spans for coherent detection. Further comments on these problems are in [4,12] and [6, §IV, IX].

9. Summary and conclusion

In this paper, a new approach to the problem of classifying spurious peaks in spectrum estimates of random data has been introduced. By computing the expected level crossing rates of spectrum estimates for Gaussian data with a smooth spectrum, the rate, width and shape of purely spurious peaks in spectrum estimates is obtained. In examples of random processes which display many significant peaks in their estimated power spectra, it is crucial to justify those spectra which truly contain a large number of deterministic periodic components over those which contain mainly spurious ones. In an example in this study, line components in the multitaper spectrum of proton density in the solar wind are numerous and are expected to correspond with the normal modes of the Sun. By comparison of actual level crossing rates with expected crossing rates before and after random permutation of the data, it is found that random permutation greatly reduces the number of significant peaks and increases their width to those expected from a Gaussian noise-like process. As well, the empirical distribution justifies that there is a large amount of power in the modes. Though the data are highly non-Gaussian, this unlikely structure is due not to the distribution, but to the temporal order of the data, that is, the process truly contains ‘many lines’.

Expressions for the variance and distribution of the derivative of the spectrum estimate, $\hat{S}(f)$, are also given in this study. Both single taper and multitaper crossing rates are considered, and the crossing expressions obtained for both methods correspond well with the crossing rates obtained through simulations. The upcrossing rate of the multitaper estimate is shown to be far smaller than that obtained from the periodogram estimate, in the case of white noise, and further reinforces the opinion that the periodogram should never be used.

Given that this research has suggested a new approach to assessing significance of spectrum estimates, what are the next steps? Possibly, the easiest analytically are extensions to significance
tests such as the harmonic $F$-test for periodic components and its multi-variate $T^2$ counterpart. The next steps may be multi-variate spectra and coherence estimates and the mathematically similar cyclostationary problems. A similar analysis of the Loève spectrum for non-stationary Gaussian processes may be possible, but significantly more difficult, as well as similar analysis of bispectra and higher order spectrum estimates.

Finally, the problem of ‘processes with many lines’ deserves much more attention than it has received in the statistical literature. There has been a strong bias towards low-order parametric models that, even ignoring ‘chaoplexology’, usually lack a physical basis. We have analysed many time series from different fields and have yet to encounter a single example of real data that can be adequately described by a low-order parametric model—it is time to try a new approach.

Acknowledgements. Thanks to Dr Frank L. Vernon of UCSD and Dr Rudolf Widmer-Schnidrig of Black Forest Observatory for advice on the barometric pressure data, to Maja-Lisa Thomson for editorial help and to the referees for their questions and comments.

Data accessibility. The barometric pressure data used in this paper were collected by the IRIS Global Seismic Network, specifically by Project IDA, based at the Cecil and Ida Green Institute of Geophysics and Planetary Physics, Scripps Institution of Oceanography, University of California, San Diego, and the USGS Albuquerque Seismological Laboratory. The facilities of the IRIS Data Management System, and specifically the IRIS Data Management Center, http://www.iris.edu/dms/nodes/dmc/data/, were used for access to waveform and metadata required in this study. The ACE SWEPAM proton density data are publicly available at: nssdcftp.gsfc.nasa.gov/spacecraft_data/ace/4_min_merged_mag_plasma (see also [60]). Thanks to Dr D. J. McComas (PI) and the ACE SWEPAM team for these data.

Funding statement. This research has been supported by the Canada Research Chairs programme, the National Science and Engineering Research Council of Canada, a Killam Fellowship from the Canada Council for the Arts and an Ontario Graduate Scholarship. The IRIS DMS is financially supported through the National Science Foundation and specifically the GEO Directorate through the Instrumentation and Facilities Programme of the National Science Foundation under Cooperative Agreement EAR-0004370. As an affiliated IRIS/IDA station, the Black Forest Observatory is operated and funded by the universities of Karlsruhe and Stuttgart.

References


26. Davies RB. 1987 Hypothesis testing when a nuisance parameter is present only under the alternative. Biometrika 74, 33–43.


