Quadratic–Inverse Estimates of Autocorrelation

David J. Thomson, Queen's University, Kingston, Ontario *

Abstract

We reconsider the classical problem of estimating the autocorrelation sequence of a stationary time–series using quadratic–inverse spectrum estimates. This collapses the free–parameter expansion ambiguity of quadratic–inverse spectrum estimates and results in estimates of autocorrelations that have simultaneously low bias and variance.

Key Words: multitaper, power spectrum, likelihood

1. History and Introduction

The problem of estimating autocorrelations of time series was first studied by Cave–Browne–Cave and Pearson (1902), shortly after Pearson introduced his estimate of correlation. Progress was slow, but by the middle of the twentieth century opinion had largely settled on Bartlett's (1948) positive–definite estimate of autocovariance. Assuming N samples of a zero–mean stationary process $x(t)$ with $t = 0, 1, \cdots, N-1$, Bartlett's estimate is defined by

$$\widehat{R_B}(\tau) = \frac{1}{N} \sum_{n=0}^{N-1-|\tau|} x_n \, x_{n+\tau} \; , \tag{1}$$

with modifications appropriate to the data. . This has become a fundamental tool in time series analysis and is still used in most texts. It is, nonetheless, a rather poor estimate.

In what was probably the first simulation study of time series, Kendall (1946) noted that, despite being biased towards zero, the estimates (1) decay slowly with τ. This was emphasized by Bartlett (1960, Page 272) who commented "It must be remembered that an observed correlogram always exhibits less damping than the theoretical, $\cdots$".

The variance of the sample autocorrelations was derived by Bartlett in 1946 and, citing this work, Kendall and Stuart (1963, Pg. 432, Vol. 3), remarked "This is awkward, for we cannot estimate them all directly from a finite series.". Somewhat earlier, Kendall (1954) had opined "We are inclined to think that the problem is most likely to yield to a new approach." Numerous improvements have been proposed since then, many with the goal of robustness, but the basic problem of finding a good estimate of the autocovariance for a stationary Gaussian process has remained open.

In §2, this paper reviews one "new approach", multitaper estimates and autocovariances. Following this summary §3 describes a simple motivating example; §4 describes basic quadratic–inverse expansions; and §4.1 discusses some problems. This is followed by some notes on likelihood for times series, §5; §6 on Karhunen–Loève expansions in the frequency domain; and the paper concludes, §7 with a short discussion.

2. Multitaper Estimates of Power Spectra and Autocovariances

Multitaper estimates of power spectra are becoming well known so only a short introduction to establish notation is given here. Assuming the same conditions given for Bartlett's

*This work has been supported by NSERC, The Killam Foundation, Bonneville Power Administration (TIP–290), and The Canada Research Chairs Program.

estimate, multitaper estimates begin by computing the *eigencoefficients* at a frequency f, Thomson (1982):

$$y_k(f) = \sum_{n=0}^{N-1} x(n)\ v_n^{(k)}(N,W)\, e^{-i2\pi nf}\ , \tag{2}$$

for $k = 0, 1, \cdots, K-1$ where $v_n^{(k)}(N,W)$ is the k^{th} Slepian sequence or discrete prolate spheroidal sequence as defined in Slepian (1978). They are the real, orthonormal eigenvectors of the Toeplitz matrix eigenvalue equation

$$\lambda_k(N,W)\, v_t^{(k)}(N,W) = \sum_{n=0}^{N-1} \frac{\sin 2\pi W(t-n)}{\pi(t-n)}\, v_n^{(k)}(N,W)\ . \tag{3}$$

that defines the sequences on $[0, N-1]$ with the maximum energy concentration in the frequency range $(-W, W)$[1]. These sequences are ordered by their corresponding eigenvalues, $1 > \lambda_0 > \lambda_1 > \cdots > \lambda_{N-1} > 0$. The λ_k decrease slowly with k (with a corresponding increase in sidelobe level) until $k \sim 2NW$ where they drop abruptly to almost zero, $K \approx 2NW$, is the number of samples required to represent the information in $(f-W, f+W)$, called the "local" or "inner" band. The notation $v_n^{(k)}(N,W)$ is shortened to $v_n^{(k)}$.

The raw eigencoefficients are usually adaptively weighted but here the weights appropriate to a low–range spectrum, the eigenvalues λ_k, are used. This gives the minimum MSE estimates of the eigencoefficients, $\hat{x}_k(f) = \lambda_k\, y_k(f)$ and results in a "high resolution" estimate, Thomson (1982, (3.6)),

$$\hat{S}_{hr}(f \ominus \xi) = \frac{1}{N} \left| \sum_{k=0}^{K-1} y_k(f) \lambda_k V_k(\xi) \right|^2\ , \tag{4}$$

where the notation $\ominus$ indicates the restriction $|\xi| < W$. Note, first, that this estimate has a χ^2_2 distribution and so is relatively unstable. However, it has the same frequency resolution as the periodogram but, when adaptive weighting is used, without the bias; and second, (4) is a form of free–parameter expansion for the spectrum because, for any given f_o and f within W of f_o, the pair f and $\xi = f - f_o$ give an estimate of $S(f_o)$. In Thomson (1982) this was resolved by the expedient of averaging over the band $(f_o - W, f_o + W)$ to get the usual multitaper estimate with a χ^2_{2K} distrbution. This multitaper spectrum estimate is

$$\widehat{S_{mt}}(f) = \frac{1}{2W} \int_{-W}^{W} \hat{S}_{hr}(f \ominus \xi)\, d\xi \tag{5}$$

$$= \frac{1}{2NW} \sum_{k=0}^{K} \lambda_k |y_k(f)|^2\ . \tag{6}$$

where the λ_k are the Slepian eigenvalues. This weighting gives the minimum mean–square–error for relatively white spectra and is used here instead of the more commonly used adaptive weights. Excluding frequencies near 0 and the Nyquist frequency, $f = \frac{1}{2}$, each of the terms, $|y_k(f)|^2$, in this sum is an ordinary direct spectrum estimate and has a chi–squared distribution with two degrees–of–freedom (DoF), χ^2_2. Because the tapers are orthonormal, the different terms are nearly uncorrelated so their average, $\hat{S}(f)$, has a chi–squared distribution with $2K$ DoF. Note that in most spectra, (5) is not used and the adaptively weighted form is used. Thus, instead of simply truncating the series at K terms, the $|y_k(f)|^2$'s are weighted to minimize their variance, see Thomson (1982).

[1] Appendix B of Thomson (1990a) describe an efficient method for computing Slepian sequences. Appendix A defines a convenient notation for their Fourier transforms, the Slepian functions, for data analysis purposes.

2.1 Multitaper Estimates of Autocovariance

One "new approach," but now over thirty years old, is a *multitaper* estimates of autocovariance. In Thomson (1982, 1990a) a multitaper spectrum estimate and used and the Einstein–Wiener–Khintchine (EWK) theorem[2] applied to obtain the corresponding autocovariance estimates. This was done without much more thought than that a better spectrum estimate should correspond to a better autocorrelation estimate following the comments in §9 of Thomson (1990b). This gives the multitaper estimate of autocovariance,

$$\widehat{R_{mt}}(\tau) = \int_{-\frac{1}{2}}^{\frac{1}{2}} \widehat{S_{mt}}(f)\, e^{i2\pi f\tau}\, df \ . \tag{7}$$

Expanding the multitaper spectrum estimate from (5) and (2) gives a triple sum over indices m, n, and k, plus the integral over frequency. The integral becomes a Kronecker δ, so $m = n - \tau$ and the expression simplifies to

$$\widehat{R_{mt}}(\tau) = \frac{1}{2NW} \sum_{k=0}^{K} \lambda_k \sum_{n=\tau}^{N-1} x(n)\, v_n^{(k)}\ x(n-\tau)\, v_{n-\tau}^{(k)} \tag{8}$$

as the "direct" form of the multitaper estimate of autocovariance.

The estimate is an average of K autocovariances of the *tapered* sequences, $\{\, x(n)\, v_n^{(k)} \,\}$. Taking expected values, one has

$$\mathbf{E}\{\widehat{R_{mt}}(\tau)\} = R(\tau) \cdot \bar{L}(\tau) \ , \tag{9}$$

where the "lag–window" is given by

$$\bar{L}(\tau) = \frac{1}{2NW} \sum_{k=0}^{K} \sum_{n=\tau}^{N-1} \lambda_k\ v_n^{(k)}\ v_{n-\tau}^{(k)} \ . \tag{10}$$

Note, however, that this does *not* imply that multitaper estimates are just another form of Blackman–Tukey estimates. Estimates made with sample autocovariances (1) and the lag–window (10) can fail to be positive definite; see Thomson *et al.* (2007, § **V**–E.). These "equivalent estimates" may have the same mean, but do not have the same distribution and, unlike (5) that is obviously positive, can be negative.

A critical step was McWhorter and Scharf's (1998) proof that the basic definition of autocovariance implies that acceptable estimates must be multitapers. This was followed by several papers, *e.g.*, Hanssen (2000); Erdöl and Günes (2005), and then by Thomson (2012) that derived sampling properties of multitaper autocovariance estimates. These work well for simple cases when there is enough data so that the corresponding spectrum estimate is reasonably resolved, but can be badly biased with short data series. This small–sample bias was the basic reason for developing "quadratic–inverse" spectrum estimates, introduced in Thomson (1990a). This theory showed that the resolution of a nonparametric spectrum estimate is effectively limited to one–half the Rayleigh resolution, $\mathcal{R} = 1/T = 1/(N\delta t)$. Typical multitaper spectra of large data sets have resolution of $3\mathcal{R}$ to $8\mathcal{R}$, but there are many cases where higher resolution is needed, but more data is simply unattainable.

[2]The basic theory of stationary processes was not well developed until the 1940's and even the relation between the power spectrum and autocovariance sequence was not understood until the work of Wiener (1930) and Khintchine (1934). It was later discovered that Einstein (1914) had preceded them in a short note, with Yaglom (1987) commenting that Einstein's derivation was more satisfactory than either Wiener's or Khintchine's.

3. An Example

The example in this paper consists of a sample of size $N = 100$ defined on $t = 0, 1, \cdots N-1 = 99$ assumed to be drawn from a real–valued, Gaussian, stationary process. The time step, $\delta t = 1$, so the Nyquist frequency is $1/(2\delta t) = \frac{1}{2}$, and cyclic frequency, f, is defined on $[-\frac{1}{2}, \frac{1}{2})$. Similarly the Rayleigh resolution $\mathcal{R} = 1/(N\delta t) = 0.01$. All frequencies are given in units of cycles per sample, denoted by "c/s". The covariance structure of the process is defined by its power spectrum, $S(f)$, a *Lorentzian* line, that is,

$$S(f) = \frac{A}{1 + \left[\frac{f - f_c}{\Delta_h}\right]^2} \; . \tag{11}$$

This form is commonly used to describe simple resonances. As an example, Table 1 of Masters and Widmer (1995) lists over 1100 seismic modes of the Earth giving their center frequencies and Q's, where Q is the ratio of center frequency to full–width at half–maximum (FWHM). Our example is for a stationary case, so is more appropriate, *e.g.*, for the study of seismic "hum", see Webb (2008), than for the transient case when the normal modes are excited by large earthquakes.

In this example, shown in Figure 1 the center frequency $f_c = 0.20$ c/s and the (FWHM) of the peak $2\Delta_h = 0.02$ c/s. The peak power A was chosen to be 100.0 to give a reasonable plotting range.

Because the process is real–valued the spectrum is even, $S(-f) = S(f)$. Data was generated using pseudo–random Gaussian variates using the spectral representation with a 2000–point FFT. The process autocovariance sequence, $R(\tau)$, was similarly obtained by taking the Fourier transform of $S(f)$.

A multitaper estimate with a time–bandwidth product $C_{\mathcal{R}} = NW = 5$ and $K = 10$ tapers was deliberately chosen because, with $N = 100$ samples, the full bandwidth $2W = 0.10$ c/s, is inadequate to resolve the 0.02 c/s width of the peak. The $K = 10$ tapers give an estimate with a nominal stability of 20 degrees–of–freedom. When the spectrum is "locally white," the 10% and 90% points for the spectrum estimate are 0.622 and 1.421 times the spectrum. However, when the slope of the spectrum is large the degrees–of–freedom are reduced. The time–bandwidth of 5.0 is typical of that used in applications where the range of the spectrum, the ratio of the maximum to minimum of $S(f)$, is moderately large.

A standard multitaper estimate of the spectrum of a single sample of such a process is shown in the left panel of Fig. 1, and it can be seen that, as expected, the peak is poorly resolved. Outside the band $(f_c - W, f_c + W)$, the estimate agrees reasonably with $S(f)$.

The right panel of this figure shows the QI test for "unresolved structure" defined in Thomson (1990a, Pg 555) using $L = 14$ basis functions. This test has an approximate χ^2_{L-1} distribution and clearly shows that the estimated spectrum, $\hat{S}(f)$, is not resolving details around the peak. This test does not assume knowledge of the true spectrum.

4. Basic Quadratic–Inverse Estimates

This section describes another "new approach"; *Quadratic–Inverse* (QI) estimates of autocovariance. Formal sampling properties have not yet been obtained, but here an example is shown. Empirical evidence implies that the maximum–likelihood extension of QI is almost unbiased and has a variance approaching that of the best parametric estimates.

QI estimates are extensions of multitaper estimates where, instead of simply finding the average power in a band $(f - W, f + W)$ centered on a frequency f, one expands the

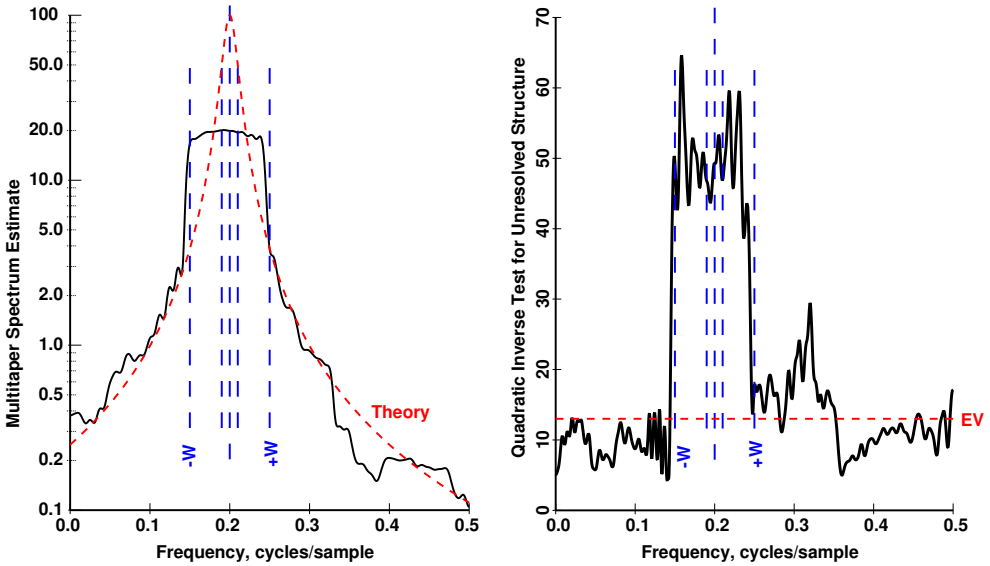

Figure 1: **Left panel** A multitaper spectrum estimate around an unresolved Lorentzian peak with a FWHM of $0.02 = 2\mathcal{R}$. The estimator had a time–bandwidth of $NW = 5$, $K = 10$ tapers, and $N = 100$ so is incapable of resolving the peak. The standard multitaper estimate is shown in solid black, and the theoretical spectrum in dashed red. The five vertical dashed blue lines show the center frequency of the peak, ± 1 $\mathcal{R}$, and $\pm W$.
The **Right panel** shows the quadratic–inverse test for unresolved structure in this estimate. Here $L = 14$ basis functions were used, so the test has an $\approx \chi^2_{13}$ distribution. The median across frequency is 12.0 and the maximum is 64.6, an event with probability $\sim 7.7 \times 10^{-9}$. The maxima of the unresolved structure test occur approximately where the squared derivatives $[\, \frac{d}{df} \ln S(f)\,]^2$ are high.

spectrum in this band on an orthogonal basis, that is,

$$\widehat{S_{QI}}(f \ominus \xi) = \sum_{l=0}^{L-1} \hat{b}_l(f) B_l(\xi) \tag{12}$$

where the $\ominus$ is a reminder that $|\xi| < W$, the $B_l(\xi)$'s are the basis functions, and the $\hat{b}_l(f)$'s are the QI expansion coefficients. The process of determining the $\hat{b}_l(f)$'s has close connections to maximum–likelihood and the Karhunen–Loève expansion. Because, for some given frequency f_o, there are a continuum of expansions with $f = f_o + \xi$, this is a form of free–parameter expansion. Formally, the EWK theorem is used to collapse these, obtaining the QI autocorrelations in the process. In practice, a tapered moving average is used because this reduces some of the Gibbs ripples.

The K eigencoefficients at each frequency are collected into a vector

$$\mathbf{X}(f) = [x_0(f), x_1(f), \cdots, x_{K-1}(f)]^T \,. \tag{13}$$

The covariance of the eigencoefficients can be estimated by $\widehat{\mathbf{C}}(f) = \mathbf{X}(f)\,\mathbf{X}^{\dagger}(f)$, where superscript † indicates conjugate–transpose. This estimate is rank 1. Assuming stationarity, the expected values of the elements of $\mathbf{C}$ are given by

$$C_{jk}(f) = \mathbf{E}\{x_j(f)\, x_k^*(f)\} \tag{14}$$

$$\approx \frac{1}{\sqrt{\lambda_j \lambda_k}} \int_{-W}^{W} V_j(\xi) V_k^*(\xi) S(f-\xi)\, d\xi \,. \tag{15}$$

If the spectrum is "locally white", that is, approximately constant across the inner band, $\mathbf{C}(f) \approx S(f) \cdot \mathbf{I}_K$ so the trace, the simple multitaper estimate, $\widehat{S_{mt}}(f)$, is efficient. However, at frequencies where this is not a good approximation, $\widehat{S_{mt}}(f)$ can have significant local bias.

Quadratic–inverse spectrum estimates were defined in Thomson (1990a) in a study of how well one could resolve the spectrum *within* the band $(f-W, f+W)$. One begins with a local expansion (12). After initial experiments attempting to express the local spectrum with Taylor's series and standard orthogonal expansions, it became obvious that there was a preferred basis set, the eigenfunctions of the Fejér kernel, defined by

$$g_l B_l(f) = \int_{-W}^{W} \left[\frac{\sin N\pi(f-\xi)}{\sin \pi(f-\xi)} \right]^2 B_l(\xi)\, d\xi \tag{16}$$

and standardized by

$$\frac{1}{2W} \int_{-W}^{W} B_l(f)\, B_m(f)\, df = \delta_{l,m} \; . \tag{17}$$

Note that the kernel in (16) is the absolute square of the Dirichlet kernel of the Slepian functions.

Given that one wants to use $\widehat{\mathbf{C}}(f)$ to estimate the coefficients $\hat{b}_l(f)$ in (12), define the sequence of $K \times K$ complex matrices $\mathbf{B}^{(l)}$ by

$$B_{jk}^{(l)} = \frac{1}{\sqrt{\lambda_j \lambda_k}} \int_{-W}^{W} V_j(\xi)\, V_k^*(\xi)\, B_l(\xi)\, d\xi \; . \tag{18}$$

These matrices are *trace orthogonal*, that is,

$$tr\{\mathbf{B}^{(l)} \mathbf{B}^{(m)\dagger}\} = g_l\, \delta_{l,m} \; . \tag{19}$$

Assuming that the true spectrum can be expanded in this basis and using (12) in (14) gives

$$\mathbf{C}(f) \approx \sum_{l=0}^{L-1} b_l(f) \mathbf{B}^{(l)} \; . \tag{20}$$

To estimate the coefficients in this expansion we substitute $\widehat{\mathbf{C}}(f)$ for $\mathbf{C}(f)$ and use (19) to obtain the quadratic form of the estimate,

$$\hat{b}_l(f) = \frac{1}{g_l} \mathbf{X}^\dagger(f) \mathbf{B}^{(l)\dagger} \mathbf{X}(f) \; . \tag{21}$$

It was shown in Thomson (1990a) that this estimate is unbiased.

It can now be shown that the multitaper "high resolution" estimate is biased. Expand (4) by expanding in the B_l basis to obtain

$$h_l(f) = \frac{1}{2W} \int_{-W}^{W} \widehat{S_{hr}}(f \ominus \xi) B_l(\xi) d\xi \; . \tag{22}$$

Expanding, and using (18) gives

$$h_l(f) = \frac{1}{2NW} \sum_{j,k=0}^{K-1} B_{j,k}^{(l)} x_j(f) x_k^*(f) \; . \tag{23}$$

Comparing this with (21) shows that $h_l(f)$ is biased, that is,

$$h_l(f) = \frac{g_l}{2NW} g_l(f) \; . \tag{24}$$

This, in a sense, parallels some of the problems surrounding the development of Bartlett's estimate; that is, should one keep the $1/N$ in (1) and have a positive–definite estimate, or, alternatively, use $1/(N-\tau)$, the "unbiased" form. In the following we resolve this quandry by using a maximum–likelihood form of the estimate.

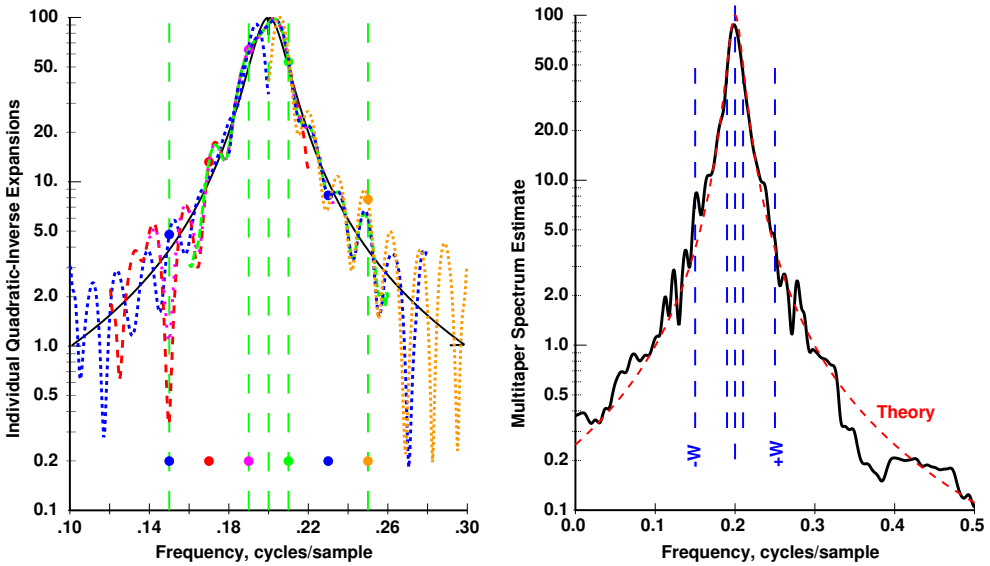

Figure 2: The **Left panel** shows six individual quadratic–inverse maximum–likelihood (QIML) expansions crossing the Lorentzian peak. These expansions have center frequencies spanning 0.15 to 0.25 cycle/sample and are offset by $2\mathcal{R} = 0.02$ c/s. Begining at the left, 0.15 c/s they are color coded as blue, red, magenta, green, blue, and orange. The centers of their domains, all $\pm 5\mathcal{R}$, are represented by filled dots of the same colour. The true spectrum is shown as the underlying solid black curve. The expansions on both sides of the peak have to cope with a range, S_{max}/S_{min} of 100 across their individual domains and so suffer worse Gibbs ripples than the expansions spanning the peak where the range drops to about 20. That is, the "blue" expansion on the left is over the frequency domain $(0.1, 0.2)$ corresponding to values of the true spectrum that range from 1.0 to 100. In contrast, the expansion centered on the peak at $f = 0.2$ covers the domain $(0.15, 0.25)$, where the range of the spectrum runs from 5.0 to 100.
Right panel The QIML averaged over the regions where the unresolved structure test is high, so a partial QIML, (PQIML) estimate. There are some remaining Gibbs ripples visible near the highest gradient parts of the frequency range. The dashed red curve shows the true spectrum.

4.1 Problems

This approach raises some questions:
First, the expansion assumed in (12) may not exist, but a goal of Thomson (1990a) was to establish bounds on nonparametric spectrum estimation. For an assumed spectral shape one can check (12) using the Bessel inequality.
Second, the eigenvalues $g_l \sim 2NW - l/2$ so the factor $1/g_l$ in (21) suggests that the $\hat{b}_l$'s for larger l's will be unreliable.
Third, the basic equation, (12) is again, in common with (4), a form of *free parameter expansion*. That is, for any specific frequency $f_o = f \ominus \xi$, all frequencies f within $\pm W$ of f_o give an estimate and these estimates differ. They are also expansions in a truncated set of orthogonal basis functions so one expects Gibbs ripples and this tends to be worse where the expansion must cover a large dynamic range. Here this occurs on the flanks of the peaks. This may be seen in Figure 2**L**. Near 0.12 c/s the red and blue curves mostly cancel, but near 0.275 c/s, the blue and orange curves have a common deep minimum.

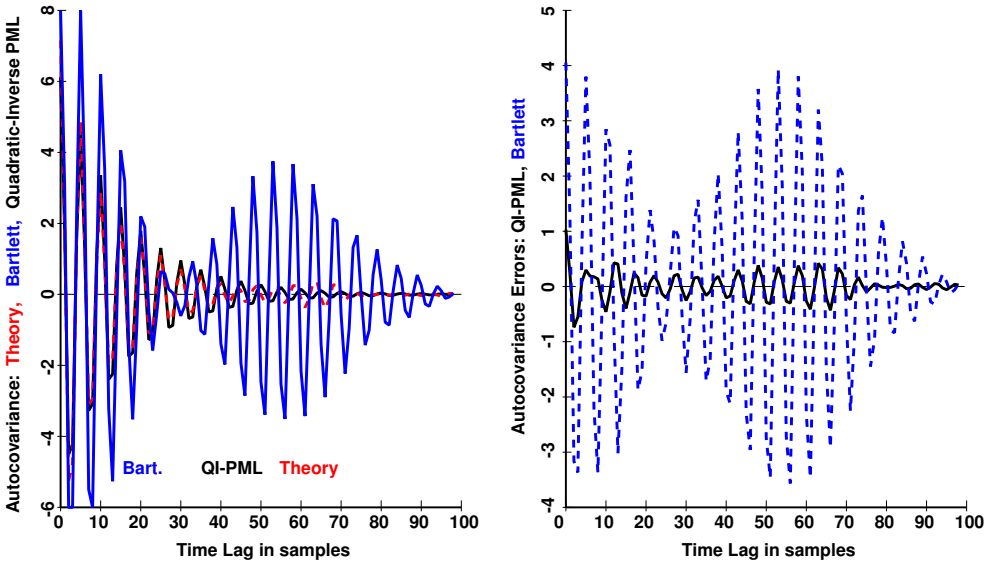

Figure 3: Autocovariance Estimates. **Left panel** The true autocovariance is shown in dashed red, the PQIML estimate in black (the two curves are nearly indistinguishable for lags $\lesssim$ 20), and the standard Bartlett estimate in blue. For this example, the MSE of the Bartlett estimate is 0.091, while that for the QIML is 0.00156, a factor of 58 lower. (Both have been divided by the process variance.) The **Right panel** shows the errors in the PQIML (solid black) and Bartlett estimate (dashed blue)

5. Likelihood for Time Series

Frequency–domain expressions for the likelihood of a Gaussian process go back to Whittle's (1952) papers. Defining the discrete Fourier transform of the series $x(t)$, with $t = 0, 1, \cdots, N-1$ at a frequency f, $-\frac{1}{2} \le f < \frac{1}{2}$, as

$$y(f) = \sum_{t=0}^{N-1} x(t)\, e^{-i2\pi f(n-\frac{N-1}{2})} \ . \tag{25}$$

Whittle's expansion depended on two ideas: *first*, that the coefficients of a Fourier expansion of the series, $y(j/N)$, are uncorrelated; and, *second*, that the periodogram, $P(f) = |y(f)|^2/N$, is asymptotically unbiased. The major problem is with the second idea. There are many statements in the statistical literature stating that the periodogram is asymptotically unbiased, but these should be ignored. Examples where this bias is so overwhelming that there is *no* existing sample size where this is true, range from quality control, see Figure 18 of Thomson (1977), to barometric pressure, see Figure 1 of Thomson and Haley (2014). (In the latter example a sample size of $N \gtrsim 10^8$ would be required, meaning one would need a continuous record at a 10 second sampling rate with a duration $\gtrsim$ 30 years.) The independence assumption in Whittle likelihood is also incorrect for such processes, but secondary to the bias problem. These problems with unwindowed Fourier transforms exclude Whittle's approach from consideration.

The problem has been rigorously considered; see Dzhaparidze (1986), the references therein, and the citations to this book (currently over 140). Most of these methods depend on a Karhunen–Loève expansion, that is given the autocovariance function of the process, $R(\tau)$, define the orthonormal eigenfunctions, $\phi_p(t)$, by

$$\theta_p\, \phi_p(t) = \sum_{n=0}^{N-1} R(t-n)\, \phi_p(n) \ , \tag{26}$$

for $p = 0, 1, \cdots, N-1$ and standardized by $\theta_0 \geq \theta_1 \cdots \geq \theta_{N-1} > 0$. Expanding the data $x(t)$ on this basis,

$$c_p = \sum_{t=0}^{N-1} \phi_p(t) x(t) \tag{27}$$

it is well–known that the coefficients are uncorrelated with variances $\mathbf{Var}\{c_p\} = \theta_p$. When the process is Gaussian the likelihood follows immediately. This expansion is transformed *into the frequency domain* in the following. This has two advantages: *First*, one does many small eigenvalue expansions instead of one large one. This is a major advantage because $R(\tau)$ is unknown, and the purpose is to estimate it. This might be considered a variant of an inverse eigenvalue problem. *Second*, Fourier transforms of reasonable data tend to be Gaussian, Mallows (1967), so approximate likelihoods are more accurate.

6. Frequency Domain Karhunen–Loève Expansions

Expanding the data inside $(-W, W)$ in a Karhunen–Loève expansion expresses that part of the data as an orthogonal series with uncorrelated coefficients. Because the eigencoefficients are often computed from large quantities of data, their distribution approaches complex Gaussian, allowing local likelihoods to be evaluated. Written in the frequency domain, the K-L eigenvalue equation is

$$\theta_p \tilde{\phi}_p(f) = \int_{-\frac{1}{2}}^{\frac{1}{2}} S(\xi) \frac{\sin N\pi(f-\xi)}{\sin \pi(f-\xi)} \tilde{\phi}_p(\xi)\, d\xi \ , \tag{28}$$

where θ_p is the eigenvalue, and $\tilde{\phi}_p(f)$ the frequency–domain eigenfunction. In this section the explicit dependence on the center frequency is omitted, and f is restricted to the domain $(-W, W)$. The Slepian functions, $V_j(f)$, are complete in $(-W, W)$, so $\tilde{\phi}_p(f)$ is expanded on this basis

$$\tilde{\phi}_p(f) = \sum_{j=0}^{K-1} A_{jp} V_j(f) \tag{29}$$

Substituting (29) into both sides of (28), multiplying by $V_k^*(f)$, and integrating over frequency, one obtains

$$\theta_p \, A_{jp} = \sum_{k=0}^{K-1} \int_{-W}^{W} V_j(\xi) V_k^*(\xi) S(f-\xi)\, d\xi \ A_{kp} \ . \tag{30}$$

Identifying the integral as $\mathbf{C}(f)$ defined in (14), and denoting the vector of A_{kp}'s by $\mathbf{A}_p(f)$

$$\theta_p(f) \mathbf{A}_p(f) = \mathbf{C}(f)\, \mathbf{A}_p(f) \tag{31}$$

where, as before, f denotes the center frequency of the band. *This shows that the eigenvalues of* $\mathbf{C}$ *are the same as the Karhunen–Loève eigenvalues.*

Substituting the expansion (20) for $\mathbf{C}(f)$ gives

$$\theta_p \, A_{kp} = \sum_{l=0}^{L-1} b_l \sum_{j=0}^{K-1} B_{k,j}^{(l)\,\dagger} \, A_{jp} \ . \tag{32}$$

A scheme to find a maximum–likelihood estimate of $\mathbf{C}$ was outlined in §4 of Thomson (1990a), but (32) is simpler because a perturbation on the b_l's enter as an additive perturbation on the spectrum. There are also $\sim 4NW$ of the b_l's, but only K eigenvalues so it

is feasible to find a set of b_l's at each frequency that maximizes the likelihood. The details of this perturbation expansion are too long to be included here. These estimates by are denoted $\tilde{b}_l(f)$ and turn to the resolution of the free–parameter expansion.

The Fourier transforms of the $B_l(\xi)$'s are index limited to $2N-1$ points, so formally, write the inverse of (7) as

$$\widehat{S}_q(f) = \sum_{\tau=1-N}^{N-1} \widehat{R}_q(\tau)\, e^{-i2\pi f\tau} \ . \tag{33}$$

Now minimize the squared difference between the global expansion (33) and the local expansions (12)

$$E^2 = \int_{-W}^{W} df \int_{-W}^{W} d\xi \ \left[\hat{S}_{QI}(f-\xi) - \widehat{S}_q(f)\right]^2 \tag{34}$$

with respect to the $\widehat{R}_q(\tau)$'s.

Practically, a tapered sum of the local QI expansions across the frequency range where the unresolved structure test, Figure 1**R**, is large was used. These expansions were stepped by a Rayleigh resolution. Moreover, the b_l coefficients were not chosen by the quadratic form (21) because these sometimes resulted in negative spectrum estimates. Instead, they were determined by directly maximizing the Karhunen–Loève likelihood numerically. A reasonably old optimizer, Gay (1983), was used for this test, and it would occasionally fail to converge, so a better optimization procedure is needed. Ironically, most of the failures occured in the more mundane parts of the spectrum, the range from 0.3 to 0.5 c/s while performance on the peak was usually excellent. In these regions, however, the unresolved structure test shown in Figure 1**R** is within an acceptable range for the "locally white" assumption to be satisfied. In these cases the ordinary multitaper estimate is known to be approximately maximum–likelihood, Stoica and Sundin (1999), implying that, at best, only a slight improvement could be obtained. However, as shown in Figure 3, the resulting estimates of autocovariance give a vast improvement on the conventional Bartlett estimate. In the figure, the relative mean-square error of the PQIML (Partial Quadratic–Inverse Maximum–Likelihood) estimate is a factor of 58 smaller than that of the Bartlett estimate. Here the Bartlett estimate "rings" for the duration of the sample while the QI estimate damps at about the same rate as the true autocovariance.

Looking at the averaged spectrum, around the peak, Figure 2**R**, the estimate is very close to the the true spectrum with the errors being competitive with parametric models. In this example, the peak of the estimate is about 15% lower than that of the true spectrum but the center frequency and peak width at the half–power point are almost exact so an estimate of Q would be accurate.

7. Discussion

This brief paper has outlined an approach to improving estimates of spectra and autocovariances in a common class of problems where frequency resolution is limited by sample size.

This work should be considered as a feasability study, rather than as a finished product. One can show that this overcomplete set of spectrum estimates can be reduced to a single estimate by expanding in terms of autocorrelations, and that this expansion largely eliminates the dependence on the nearly arbitrary choice of the bandwidth parameter W. However, how one characterizes the sampling properties of these estimates remains to be determined.

References

Bartlett, M. S. (1946). On the theoretical specification and sampling properties of autocorrelated time series. *J. Royal Statist. Soc. Suppl.*, **8**, 27–41.

Bartlett, M. S. (1948). Smoothing periodograms from time–series with continuous spectra. *Nature*, **161**.

Bartlett, M. S. (1960). *Stochastic Processes*. Cambridge University Press, Cambridge.

Cave–Browne–Cave, F. E. and Pearson, K. (1902). On the correlation between the barometric heights on eastern side of the atlantic. *Proc. Royal Soc. London*, **70**, 465–470.

Dzhaparidze, K. (1986). *Parameter Estimation and Hypothesis Testing in Spectral Analysis of Stationary Time Series*. Springer-Verlag, New York. S. Kotz, translator.

Einstein, A. (1914). Methode pour la determination de valeurs statistiques d'observations concernant des grandeurs soumises a des fluctuations irregulieres. *Archives des Sciences*, **37**, 254–256.

Erdöl, N. and Günes, T. (2005). Multitaper covariance estimation and spectral denoising. In *Circuits, Systems, and Computers*, pages 1144–1147. IEEE Press.

Gay, D. M. (1983). ALGORITHM 611 — subroutines for unconstrained minimization using a model/trust–region approach. *ACM Trans. Math. Software*, **9**, 503–524.

Hanssen, A. (2000). On multitaper estimators for correlation. In *Proc. Tenth IEEE Signal Processing Workshop*, pages 391–394, Pocono Manor, PA. IEEE Press.

Kendall, M. G. (1946). *Contributions to the study of oscillatory time–series*. Cambridge University Press, Cambridge.

Kendall, M. G. (1954). Note on bias in the estimation of autocorrelation. *Biometrika*, **41**, 403–404.

Kendall, M. G. and Stuart, A. (1963). *The Advanced Theory of Statistics*. Hafner, New York.

Khintchine, A. (1934). Korrelationstheorie der stationären stochastischen Prozesse. *Mathematische Annalen*, **109**(1), 604–615.

Mallows, C. L. (1967). Linear processes are nearly Gaussian. *J. Appl. Prob.*, **4**, 313–329.

Masters, T. G. and Widmer, R. (1995). Free oscillations: Frequencies and attenuations. In T. J. Ahrens, editor, *Global Earth Physics*, pages 104–125. Amer. Geophysical Union, Washington, DC.

McWhorter, L. T. and Scharf, L. L. (1998). Multiwindow estimators of correlation. *IEEE Trans. on Signal Processing*, **46**, 440–448.

Slepian, D. (1978). Prolate spheroidal wave functions, Fourier analysis, and uncertainty V: the discrete case. *Bell System Tech. J.*, **57**, 1371–1429.

Stoica, P. and Sundin, T. (1999). On nonparametric spectral estimation. *Circuits Systems Signal Process.*, **18**, 169–181.

Thomson, D. J. (1977). Spectrum estimation techniques for characterization and development of WT4 waveguide, *Part II. Bell System Tech. J.*, **56**, 1983–2005.

Thomson, D. J. (1982). Spectrum estimation and harmonic analysis. *Proceedings of the IEEE*, **70**, 1055–1096.

Thomson, D. J. (1990a). Quadratic-inverse spectrum estimates: applications to paleoclimatology. *Phil. Trans. R. Soc. Lond. A*, **332**, 539–597.

Thomson, D. J. (1990b). Time series analysis of Holocene climate data. *Phil. Trans. R. Soc. Lond. A*, **330**, 601–616.

Thomson, D. J. (2012). Some comments on multitaper estimates of autocorrelation. In *Proc. 2012 SSP Workshop on Spectrum Estimation and Modelling*, pages 656–659. IEEE.

Thomson, D. J. and Haley, C. L. (2014). Spacing and shape of random peaks in non–parametric spectrum estimates. *Royal Society Proceedings Series A*, **470**, 20140101.

Thomson, D. J., Lanzerotti, L. J., Vernon, III, F. L., Lessard, M. R., and Smith, L. T. P. (2007). Solar modal structure of the engineering environment. *Proceedings of the IEEE*, **95**, 1085–1132.

Webb, S. C. (2008). The Earth's hum: the excitation of Earth normal modes by ocean waves. *Geophys. J. Inter.*, **174**, 542–566.

Whittle, P. (1952). The simultaneous estimation of a time series harmonic components and covariance structure. *Trabajos de Estadistica*, **13**, 43–57.

Wiener, N. (1930). Generalized harmonic analysis. *Acta Mathematica*, **55**(1), 117–258.

Yaglom, A. M. (1987). Einstein's 1914 paper on the theory of irregularly fluctuating series of observations. *IEEE Signal Processing Magazine*, **4**(4), 7–11.